THREE DIMENSIONAL LOCALIZATION AND TRACKING FOR SITE SAFETY USING FUSION OF COMPUTER VISION AND RFID

By

Rana Hammad Raza

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Electrical Engineering - Doctor of Philosophy

2013

ABSTRACT

THREE DIMENSIONAL LOCALIZATION AND TRACKING FOR SITE SAFETY USING FUSION OF COMPUTER VISION AND RFID

By

Rana Hammad Raza

We propose a state of the art fusion framework of Computer Vision (CV) and Radio Frequency Identification (RFID) to support object recognition and tracking in a three dimensional space. Fusion can significantly improve performance in applications of autonomous vision and navigation and site monitoring, especially in outdoor environments. Increasing safety in construction zones and enhancing security in airports are important problems that involve understanding interactions between objects, machines and material and can be solved using sensor fusion and activity analysis. Identifying objects solely via vision is computationally costly, error prone, limited by occlusion, and sometimes impossible in practice. RFID can reliably identify tagged objects and can even localize targets at coarse spatial resolution. Alternatively, CV can increase the performance of RFID by fine tuning the location information and providing fuzzy features to avoid cloning or deception.

We have implemented stereo using commodity cameras and have used a commercial RFID based Real Time Location System (RTLS) for our experiments and have achieved encouraging results. The performance of both modalities was evaluated separately and in fused mode. In our stereo experiments outdoors we obtained an RMS accuracy of within *~7.6 in (19.3 cm)* for objects up to *80 ft (24.4 m)* away from the cameras. For real time trajectories, RTLS provided *2 m to ~2.6 m* location accuracy for dynamic tagged objects in a cell of *40×40 m* with four readers. We propose a fusion based tracking algorithm and our research demonstrates benefits obtained

when most objects are cooperative and tagged. We abstract the information structures in order to support a Site Safety System (S-3) with diverse information sources and constraints and processes that may not have knowledge of each other. We have used relaxation to control the integration of information from CV, RFID, and naïve physics in tracking. The label elimination approach readily represents the ambiguity occurring in real-life applications. The key to reducing the computational requirements is to eliminate many labels at each filtering step while keeping those labels compatible with observation. As a post processing step to labeling, we have used total track smoothness for optimization to update computed tracks for increasing system tracking reliability. Work site analysis can proceed even when information from one sensor or information source is unavailable at some time instances.

We have shown with simulations and real data that fusion can greatly increase tracking performance and can reduce computational cost and combination search space up to *99%* in some cases. Test cases showed how fusion can solve some difficult tracking problems outdoors. We assessed performance of tracking using track error i.e fraction of wrong trajectory point assignments. For some object trajectories outdoors, the fused system reduced the track error from *0.53* to *0.13*. The likelihood of producing correct object trajectories in regions partially or fully occluded to CV is also increased. We conclude that significant real-time decision-making should be possible if the S-3 system can integrate information effectively between the sensor level and activity understanding level. Engineering faster RFID updates will likely reduce the number of objects that can be sensed; however, this should be a favorable tradeoff in a construction site. Employing knowledge based constraints and analyzing systematically object track initiation and termination are some of the possible research expansions to be worked upon in the near future.

DEDICATION

To Maheen, Hamna, Sadaf, my brother, my mother and father

ACKNOWLEDGEMENTS

I would like to extend my gratefulness to my research advisor George Stockman, for his years of guidance and support. His research insight and sheer persistence has made this a thoughtful and rewarding journey. I owe him debt of gratitude for his countless hours of reflecting, encouraging and guiding me through the entire process and believing in me. He has always been by my side like a father. I cannot find words grand enough to thank him.

Likewise, I want to acknowledge my other advisor Lalita Udpa. Her advice and support has continually been valuable to me over the years. I am also grateful to my committee members Mohamed El-Gafy and Subir Biswas for being available in times when I needed their guidance.

I would like to recognize the unwavering love, inspiration and constant support of my wife, Sadaf, through all stages of the process and beyond. She is my pillar, my joy and my guiding light. I really owe my beautiful daughters Hamna and Maheen for always cheering me up and even sometimes compromising on their daddy-daughter time. I am grateful and highly indebted to my mother and brother Hakim for their constant love, continued encouragement and prayers, when I needed them most. Cannot thank them enough for giving me the foundation of where I am today. I would like to keep in remembrance my late father, aunt Tasneem and grandparents who have always been with me in spirit. I would also like to show appreciation towards my brother's family, in laws and extended relatives and friends for their prayers and faith in me. Special thanks to Kathy and Adnan for making our stay at East Lansing an incredible experience.

Finally and most importantly praises and thanks to Allah for all his blessings.

TABLE OF CONTENTS

LIST OF TABLES xi

LIST OF FIGURES xii

CHAPTER 1 Introduction 1
1.1 Research problem and area of focus 2
1.2 Basic functionality requirement and significance of research 3
1.3 Motivating application: monitoring activities in kindergarten 5
1.4 Capabilities and limitations of CV 7
1.5 Capabilities and limitations of RFID 8
1.6 Dissertation outline 11

CHAPTER 2 Background 12
2.1 The functionality and limitations of RFID 13
2.1.1 RFID operation 14
2.1.2 Types of RFID tags 16
2.1.3 Radio frequency 18
2.1.4 Limitations and needed improvements 19
2.1.4.1 Standardization and cost 20
2.1.4.2 Read accuracy 20
2.1.4.3 Anti collision 20
2.1.4.4 Security 21
2.1.4.5 Size and power 21
2.1.4.6 Miscellaneous limitations 22
2.1.5 Positional information 22
2.1.6 Smart objects, networks, location-aware computing 23
2.2 Computer Vision functionality and limitations 24
2.2.1 High image processing and search cost 29
2.2.2 Optical sensing problems 30
2.2.3 Object modeling 30
2.2.4 Object tracking 31
2.2.4.1 Object recognition 32
2.2.4.2 Object location and pose estimation 33
2.2.4.3 Sensing and relating 3D object points 33
2.3 Research projects based on fusion of CV and RFID 35
2.3.1 Model based recognition 37
2.3.2 Human-object interaction and activity analysis 41
2.3.3 Mobile robot localization 47
2.3.3.1 Map based navigation 47
2.3.3.2 Obstacle recognition 49
2.3.4 Miscellaneous 50
2.3.5 Natural/outdoor site management 52

CHAPTER 3 Proposed solution and research methodology 55
3.1 Tokens code observations from images and RFID 56
3.2 Object tracks {<x, y, z, t, v, L>} 57
3.3 Obtaining 3D object location (x, y, z) 58
3.4 Heuristics from naïve physics 60
3.5 Fusion platform 61
3.5.1 Labeling with relational constraints 61
3.6 Sensor arrangement 62
3.6.1 Vision infrastructure 64
3.6.2 RFID infrastructure 66
3.7 Goals and related research parameters 69
3.7.1 Calibrating the cameras 69
3.7.2 Defining the ground truth 69
3.7.3 Structural stereo approach 70
3.7.4 Object sensing using multiple RFID readers 71
3.7.5 Integration of multiple CV sensors 71
3.7.6 Fusion of RFID and cameras 71
3.7.7 Smoothness of trajectories 72
3.7.8 Looming detection 72
3.7.9 Object inventory 73

CHAPTER 4 Localization of objects and scene points 74
4.1 Object detection and blob analysis using vision 75
4.1.1 Elliptical shape features for head detection 79
4.2 Stereo vision using ray-ray combination 83
4.2.1 Stereo configuration 84
4.2.2 Computing shortest line segment connecting two rays 86
4.3 3D location estimation results using stereo vision 89
4.3.1 Computing residual error using jig 89
4.3.2 Components of the stereo system used 93
4.3.3 Indoor stereo computation using a wireframe workspace 94
4.3.4 Indoor stereo computation using at surveyed lab area 98
4.3.5 Outdoor stereo computation 100
4.4 Active RFID based Real Time Location System 103
4.4.1 RTLS infrastructure 104
4.4.2 Cell architecture 107
4.5 RTLS based location estimation 107
4.5.1 Indoor location sensing 108
4.5.2 Outdoor location sensing 109
4.6 Summary discussion 113

CHAPTER 5 Fusion dynamics and analysis 115
5.1 The fusion process approach 116
5.2 Multiple sensor configuration 117
5.3 Building block for multiple sensor fusion 119
5.4 Benefits of fusing RFID and CV 121

5.5 Test cases to explain fusion and its analysis 125
5.5.1 Test case I - Same colored objects 125
5.5.2 Test case II - Different colored objects 126
5.5.3 Test case III - Different colored objects w/ intermittent RFID/CV feeds 127
5.6 Summary discussion 128

CHAPTER 6 Tracking using fusion of CV and RFID 129
6.1 Object tracking model using sensor fusion 129
6.2 Labeling via iterative processing 130
6.2.1 Sensor process 132
6.2.2 Combination process 132
6.2.3 Tracking process 132
6.2.4 Relaxation labeling algorithm 133
6.2.5 Test cases to analyze discrete relaxation labeling 136
6.2.5.1 Test case IV - Two same colored objects w/ simple dynamics 136
6.2.5.2 Test case V - Four same colored objects w/ increased complexity 137
6.3 Optimization process 141
6.3.1 Smoothness of trajectories 142
6.3.2 Tracking algorithm description 143
6.4 Summary discussion 149

CHAPTER 7 Experiments, results and analysis 150
7.1 Generating stereo trajectories 151
7.1.1 Real indoor trajectories from wireframe workspace 151
7.1.2 Mathematical trajectories 152
7.1.3 Real stereo trajectories from indoors lab area 154
7.1.4 Real stereo trajectories from outdoors 155
7.2 Real outdoor trajectories using RTLS 157
7.3 Metrics for evaluation of performance 158
7.4 Real-time tracking: indoor with RFID feed simulated using color 161
7.5 Indoor stereo live demo results 161
7.6 Stereo error analysis in x, y, z dimension versus distance from the cameras 164
7.7 Least squares analysis on real outdoor stereo trajectories 167
7.8 RTLS signal availability 175
7.9 Simulations of object tracking 176
7.10 Tracking efficiency using fusion 180
7.10.1 Simulated scenario: two persons and two briefcases 181
7.10.2 Real outdoor scenario 182
7.10.3 Outdoor scenarios with varying fusion information 187
7.11 Object color variations 189
7.12 Sensor Error and synchronization problems 193
7.13 Summary discussion 194

CHAPTER 8 Conclusions and future work 196
8.1 Background Survey 196

8.2 Evaluation of RFID, CV, and fused sensing 197
8.2.1 Demonstrated performance, potential and parameters for outdoor applications 198
8.3 Modeling fusion and its benefits 199
8.4 Data integration and filtering using relaxation labeling 199
8.5 3D object tracking algorithm 200
8.6 Future work and limitations 201

APPENDICIES 206
APPENDIX A Fundamental experiments on looming 207
APPENDIX B Stereo concepts and calibration procedure 216
APPENDIX C Site survey details 229
APPENDIX D Wireless location sensing 236

REFERENCES 251

LIST OF TABLES

Table 1.1	*Advantages and disadvantages of CV and RFID.*	10
Table 2.1	*RFID tag types.*	17
Table 2.2	*RFID typical frequencies in use and respective ranges.*	18
Table 2.3	*RFID frequency attributes.*	19
Table 2.4	*Obstacle recognition results. See Cerrada et al. [59].*	39
Table 2.5	*Activity and object recognition rates. See Jianxin et al. [69].*	43
Table 2.6	*Obstacle localization with different antenna settings. Data extracted from Songmin et al. [82].*	50
Table 4.1	*Left and right camera transformation matrices.*	91
Table 4.2	*2D residuals for left and right image of jig - scale is in pixels.*	92
Table 4.3	*3D stereo residuals for some points in wireframe workspace - scale is in inches.*	96
Table 4.4	*Comparison between ground and RFID observations in xy-plane- scale is in inches.*	112
Table 7.1	*3D ground truth and computed data for analyzing outdoor stereo error - scale is in inches.*	166
Table 7.2	*Ball projectile observed using stereo and fitted data along yz-plane - scale is in inches.*	172
Table 7.3	*Location error for RTLS tag trajectory in Figure 7.1 - scale is in inches.*	174
Table 7.4	*Track error with different points density and block length m. Track error is a fraction of wrong trajectory point assignments.*	180
Table C.1	*3D coordinates of some important landmarks in MSU Engineering courtyard - scale is in inches.*	235
Table D.1	*Location accuracies of cellular radiolocation technologies. See Kos et al. [138].*	244

LIST OF FIGURES

Figure 1.1	*Fusion based Kindergarten students' online observing system. "For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation."*	6
Figure 2.1	*General infrastructure of an RFID system. The RFID tag may be many meters distant from the reader.*	15
Figure 2.2	*Relationship of camera C with pyramid A and cube B in workspace W.*	26
Figure 2.3	*Work robot seeing a section of the Alaska Pipeline via a) an intensity image and b) its Laplacian. A few corner points are evident that can enable the robot to get oriented to inspect and operate on the pipeline. See Shapiro et al. [31].*	27
Figure 2.4	*The Perspective 3-Point problem: a camera computes its orientation relative to three points seen from a known object.*	28
Figure 2.5	*Still camera monitoring a workspace: a) workspaces can be monitored by staring cameras b) object motion can be detected by changes in region statistics.*	32
Figure 2.6	*Pose in 3D simplified by sensing 3D points - Two cameras C_1 and C_2 knowing their relation to the space W can triangulate to compute the coordinates of a point in that space. That point must lie on the intersection of the two imaging rays [31] Ch 13.*	35
Figure 2.7	*General architecture of RFID and vision fusion.*	36
Figure 2.8	*3D objection recognition algorithm. Recreated from Figures 3 and 5 of Cerrada et al. [57].*	39
Figure 2.9	*Map based navigation scheme. Recreated from Figures 1 and 3 of Weiguo et al. [79].*	49
Figure 3.1	*Sensor error volumes: (a) rays intersection with error cones (b) intersection of error cone with RF lobe.*	59
Figure 3.2	*Block diagram of the Site Safety System using fusion of RFID and CV.*	63
Figure 3.3	*Construction scene example [98] with aspect ratio of persons and the field of view.*	65

Figure 3.4	*Schematic of dense placement of reference RFID tags.*	67
Figure 3.5	*CSL RFID based Real Time Location System: (a) active tag (b) master reader.*	68
Figure 4.1	*Flow diagram of specific RGB color detection and connected components blob analysis.*	76
Figure 4.2	*Blob detection outdoors: (a) original image (b) blobs of blue and yellow balls.*	77
Figure 4.3	*Ellipse geometry showing basic parameters to define an ellipse.*	80
Figure 4.4	*Implementation steps for elliptical shape feature detection.*	81
Figure 4.5	*Results of head detection using elliptical shape features: (a) input image 640×480 (b) cropped image with best two ellipses (c) cropped edge image with best two ellipses.*	82
Figure 4.6	*Error cones obtained from projecting 2D imaging error back into 3D.*	84
Figure 4.7	*General stereo configuration with two cameras viewing a 3D object in a 3D workspace W.*	85
Figure 4.8	*Shortest line segment connecting the two skew rays.*	87
Figure 4.9	*Jig images with easily recognizable feature points: (a) left image (b) right image.*	91
Figure 4.10	*Procedure for calculating 2D residuals of jig images.*	93
Figure 4.11	*Wireframe workspace experimental setup for testing stereo localization indoors - red object with spiral trajectory.*	95
Figure 4.12	*Procedure for 3D stereo location estimation in wireframe workspace indoors - Flow diagram.*	97
Figure 4.13	*Computed sphere trajectory in wireframe workspace indoors.*	98
Figure 4.14	*3D survey data and model of indoors lab area with stereo cameras '●' (green). Calibration points are represented by '■' (yellow).*	99
Figure 4.15	*Indoors lab area trajectory for a person moving randomly with error estimation: (a) object tracks in 3D (b) histogram of error along z-axis.*	100

Figure 4.16 *Outdoor test site with calibration points shown by '■' (yellow): (a) left image (b) right image.* 101

Figure 4.17 *3D view of the outdoor test area with object tracks shown by '+' (red) using stereo.* 102

Figure 4.18 *Top view of the outdoor test area showing person locations computed using stereo and the ground truth '•'of outliers beyond x = 960 in (24.4 m).* 103

Figure 4.19 *CSL RFID based RTLS system: Tripods hold the readers.* 104

Figure 4.20 *RTLS location sensing indoors - setup.* 108

Figure 4.21 *RTLS location sensing indoors - Floor map.* 109

Figure 4.22 *RTLS location sensing outdoors - setup: (a) master reader on tripod (b) reader pointing towards test site center (c) reference tag placed in test site.* 110

Figure 4.23 *Top view of the outdoor test area showing person locations computed using RTLS.* 111

Figure 4.24 *Computed locations of the person using RTLS and error estimation - error circles represent difference between ground truth and RTLS.* 113

Figure 5.1 *Multi-sensor configurations: (a) cooperative (b) cooperative (c) complementary (d) competitive.* 118

Figure 5.2 *Basic fusion node architecture.* 120

Figure 5.3 *CV and RFID supplementing each other - Case I: (a) same colored object tracks (b) label assignments* 126

Figure 5.4 *CV and RFID supplementing each other - Case II: (a) different colored object tracks (b) label assignments.* 127

Figure 5.5 *CV and RFID supplementing each other - Case III: (a) considering visual occlusion and intermittent RFID, different colored object tracks (b) label assignments.* 128

Figure 6.1 *Schematic diagram of object tracking using sensor fusion and relaxation labeling.* 130

Figure 6.2 *Correct object tracks with possible compatible labels at each block of time frames.* 137

Figure 6.3 *Test case showing relaxation labeling: (a) correct trajectories of persons and balls. (b) correct balls and incorrect persons' trajectories (c) left matrix - General pattern of four relaxation constraint passes and final compatible label/s. Right matrix - RFID location information.* 138

Figure 6.4 *Label matrix updating steps for same colored objects at each time frame for Figure 6.3(a) tracks.* 140

Figure 6.5 *Block diagram of the smoothness algorithm.* 148

Figure 6.6 *Step by step results of an example with smoothness algorithm applied: (a) input data (b) nearest neighbor assignment (c) smoothed trajectories.* 149

Figure 7.1 *Example of stereo trajectories generated from wireframe workspace indoors.* 152

Figure 7.2 *Mathematically generated trajectories using dataset generator with T=11 and N=7.* 153

Figure 7.3 *Indoors lab area real stereo trajectory where a person moved over predefined points.* 154

Figure 7.4 *Five outdoor real stereo trajectories: (a) 3D display of site (b) zoomed in top view of trajectories.* 156

Figure 7.5 *Outdoors 2D RTLS trajectories of a tagged person (green paths).* 158

Figure 7.6 *3D stereo tracking in wireframe workspace indoors - live demo.* 163

Figure 7.7 *3D stereo tracking in lab area indoors - live demo: (a) orange cursive writing sample (b) heart shapes using orange and yellow color (c) yellow used as initializing marker to track orange.* 164

Figure 7.8 *Selected 2D corresponding points in left and right camera image for analyzing outdoor stereo error.* 165

Figure 7.9 *Stereo RMS error in x,y and z direction versus distance from the camera.* 167

Figure 7.10 *Left camera images showing projectile trajectory of a ball tossed upward.* 168

Figure 7.11 *Two different views of 3D points showing ball projectile trajectory computed using stereo.* 169

Figure 7.12 *Computed ball projectile trajectory (solid blue) with parabolic fitting (dotted magenta).* 170

Figure 7.13 *Actual trajectory (dashed blue) with linearly increasing error along y-axis and corrected trajectory (solid green) using parabolic curve fitting.* 171

Figure 7.14 *Piecewise line fitting (solid green) on RTLS tag trajectory (dotted blue).* 173

Figure 7.15 *RTLS tag location error circles.* 174

Figure 7.16 *RTLS trajectory analysis: (a) RTLS single tag trajectory (green path) (b) RTLS tag location signal availability over time.* 176

Figure 7.17 *Reduction in combination volume with probability of random ID information availability.* 178

Figure 7.18 *Possible combination volume with N objects and probability P of object ID in bursts of four tokens.* 179

Figure 7.19 *Testing fusion using simulated scenario of two persons exchanging briefcases: (a) wrong interpretation of trajectories with CV alone (b) correct interpretation of trajectories with CV & RFID fusion.* 182

Figure 7.20 *3D view of test site with calculated real trajectories.* 183

Figure 7.21 *Outdoor scenario to test fusion: (a) left camera view of test site (b) computed correct 3D ball trajectories using fusion.* 185

Figure 7.22 *Outdoor RTLS trajectories of two tagged persons with varying fusion information - persons split on sides at the center of test site.* 188

Figure 7.23 *Change in illumination observed in left camera video feed when: (a) sunny (b) shady. Color histograms: (c) blue ball in sunlight, (d) blue ball in shade, (e) yellow ball in sunlight, (f) yellow ball in shade.* 190

Figure 7.24 *Sample images for different weather and illumination conditions to study blue and yellow color consistency.* 191

Figure 7.25 *Analyzing blue and yellow ball color consistency in HSV color space under different weather and illumination conditions.* 192

Figure 8.1 *Occlusion management flow during stereo tracking.* 203

Figure A.1 *Looming image dataset at different distances: (a) object at ten feet (b) object at two feet.* 208

Figure A.2 *Graph of bounding box width and height relationship for training datase* 209

Figure A.3 *Graph of bounding box area vs looming distance relationship for training dataset.* 210

Figure A.4 *Graph of bounding box area vs looming distance relationship for two real-time datasets.* 211

Figure A.5 *Indoor lab platform consisting of NXT robotics kit and the iphone 3GS to test collision avoidance using optical flow.* 212

Figure A.6 *Motion vectors computation during real-time lab demo for collision avoidance:(a) test frame k-1 (b) test frame k (c) motion vectors using Horn-Schunck (d) motion vectors using Lucas-Kanade.* 213

Figure A.7 *Realtime indoor collision avoidance experiment: (a) robot approaching the obstacle (b) robot detected the obstacle.* 214

Figure B.1 *Loss of depth information in 2D - caused by projection of 3D points on same viewing line onto 2D image.* 217

Figure B.2 *Recovering 3D point coordinates using stereo vision.* 218

Figure B.3 *Stereo correspondence problem: which points in Image 1 actually correspond to points* ${}^{W}P$ *and* ${}^{W}Q$ *in Image 2?* 219

Figure B.4 *Epipolar geometry.* 220

Figure B.5 *3D reconstruction using 2D image points.* 221

Figure B.6 *Coordinate system used in camera calibration: (a) 3-D world (b) camera.* 222

Figure B.7 *Shortest line segment connecting the two skew rays.* 226

Figure C.1 *MSU Engineering building satellite view.* 229

Figure C.2 *Different views of the courtyard.* 230

Figure C.3 *Total station surveying equipment - Image extracted from [117].* 231

Figure C.4 *Top view of the outdoor test site with legend showing equipment position and the coordinate system.* 232

Figure C.5 *3D view of the outdoor test site with sensor configuration - scale is in inches.* 233

Figure C.6 *Outdoor test site with calibration shown by '■' (yellow): (a) left image (b) right image.* 234

Figure D.1 *Triangulation geometry.* 239

CHAPTER 1

Introduction

Site safety and security, which requires understanding the interaction of persons, objects and machines, is an important problem that can be solved using sensor fusion and activity analysis. Identifying objects solely via computer vision (CV) is computationally costly, error prone, limited by occlusion, and sometimes impossible in practice. Radio Frequency Identification (RFID) can reliably identify tagged objects and can even localize targets at coarse spatial resolution. Our research that focuses on construction sites demonstrates benefits obtained when most objects are "cooperative" by being RFID tagged. We do not assume a controlled environment, but do assume that a survey of the terrain exists, including benchmark locations. This partial control is needed since tracking, especially in an outdoor environment, presents difficulties with varying lighting, rain, smoke, dust, and noise, and occasional unexpected agents or objects. Real-time decision-making, which is needed for safety and security applications, should be possible if the overall system can integrate information effectively between the sensor level and activity understanding level.

Tracking multiple objects is a fundamental problem with wide application and a rich literature. We are interested in application problems in site monitoring, security, and

activity analysis. Examples are tracking workers, materials, and machines in construction sites; baggage and people in airports; or patients and care workers in medical care facilities. There are many other important applications that come under this domain such as analysis of social or workplace interactions, analysis of games or shopping, asset management, old age home monitoring, and assisting persons with disability etc.

1.1 Research problem and area of focus

In most of our experimental work we have focused on construction site safety in an off-highway outdoor environment. The approach we have presented is conceptually analogous and valid for other applications.

Construction sites are planned areas that consist of resources such as personnel, equipment and materials involved in active work tasks. These resources are continuously dynamic and possess uncontrolled and sometimes arbitrary trajectories during the construction work. Also, construction sites are generally constrained and crowded areas. Any spatial interference in such dynamic construction sites can cause accidents involving collisions. Each year, more than *100* workers are killed and over *20,000* are injured in the construction industry [1]. One of the distinct safety problems has been identified as the proximity of workers-on-foot to heavy construction equipment and other vehicles [2]. Fatigue, work pressure, repetition of work [1], lack of awareness of existing specific risk factors, along with blind spots [3] are among the major causes of such work fatalities. Also many commercially available types of proximity warning

sensors and systems can be rendered useless in the construction environment when covered with mud, ice, snow, ore, rock, and other material.

There are several proximity warning technologies that are available to help eliminate blind spots and associated accidents involving large off-highway equipment [4]. These include Radio Detection and Ranging (RADAR), Global Positioning System (GPS), RFID tags, cameras, 3D laser scanners and combinations of these technologies. Each of these comes with some limitations, such as operating range, signal availability, size and weight, cost, susceptibility to false alarms and applicability to construction environment. Successful implementation of these systems may be achieved if their shortcomings are realized properly and anticipated.

To avoid spatial conflicts of construction resources, time-lapse photography and cameras are often used to analyze daily safety procedures [5]. Due to computational complexity in an uncontrolled environment, vision based techniques lack capability for real-time alerts. Therefore using Computer Vision (CV) alone for object recognition, localization and tracking tasks is still challenging. Some of the limiting factors are blind spots, arbitrary trajectories, changes in pose, scale, lighting, occlusion, visual data quality, volume of data and uncertainty.

1.2 Basic functionality requirement and significance of research

Construction site safety and similar applications of interest require a Site Safety System (S-3) that provides some or all of the following basic functions.

a. **Detection** of the presence of objects of interest (persons, machines, materials, vehicles…)
b. **Identification** of the objects (by class or by a unique object instance)
c. Object **location** in workspace coordinates or by designated areas,
d. **Object track**, if the object is moving,
e. Important **object properties**, such as shape, color, weight, speed, ownership, supplier, etc.
f. A memory **representation** of space and time including location of objects, trajectories, and behaviors.
g. Application specific processes that manage objects and control their behavior (such as collision avoidance or creation)

Fusion of computer vision and RFID can provide this functionality in many cases. While humans rely on visual sensing for such problems, an automatic system will perform faster and more reliably with fused input compared to optical input alone. Higher level problem-specific analysis can then be applied on top of functionality *(a)* to *(e)* to create a dynamic inventory of a workspace, infer what agents or objects are doing (*function f*), manage interactions, define and summarize events etc. (*function g*).

In a construction site domain, fusion can be used to design a real-time three dimensional localization and tracking system incorporated to better automate work safety technologies. It is possible to accurately locate and track static and moving objects using three dimensional fused data obtained from RFID and CV. The S-3 system will have video cameras, RFID tags/readers, a spatial database, local and global warning systems, and wireless networking to synchronize and

transmit information to displays for remote control and alerting workers/operators. Such a system with effective management practices has sufficient potential to significantly improve safety at construction sites. In a construction site environment, it is safe to assume that the objects are mostly cooperative and tagged, wearing distinctive clothing, and that 3D survey data of the test site exists.

The significance of this research is to explore integration of S-3 system capabilities as follows:

a. Integrate cost effective sensors for fusion into a real-time construction site environment.
b. Enhancing system detection and recognition capabilities.
c. Understanding interactions between objects and materials in an outdoor environment.
d. Augmenting three dimensional localization and tracking that can enable safety spheres for construction personnel in the work envelope of construction vehicles and heavy equipment.
e. Creating dynamic inventory of a workspace for resource efficiency maximization. This will help improve construction site management by utilizing site information to better account for equipment, materials, personnel, and activities.

1.3 Motivating application: monitoring activities in kindergarten

Consider a motivating application of "analyzing" what is going on in a kindergarten: Figure 1.1 (see S. Nakagawa *et al.* [6]). The major output is video. Parents can view their child's activity on the Internet: they can see what their child does and what other children or toys he/she

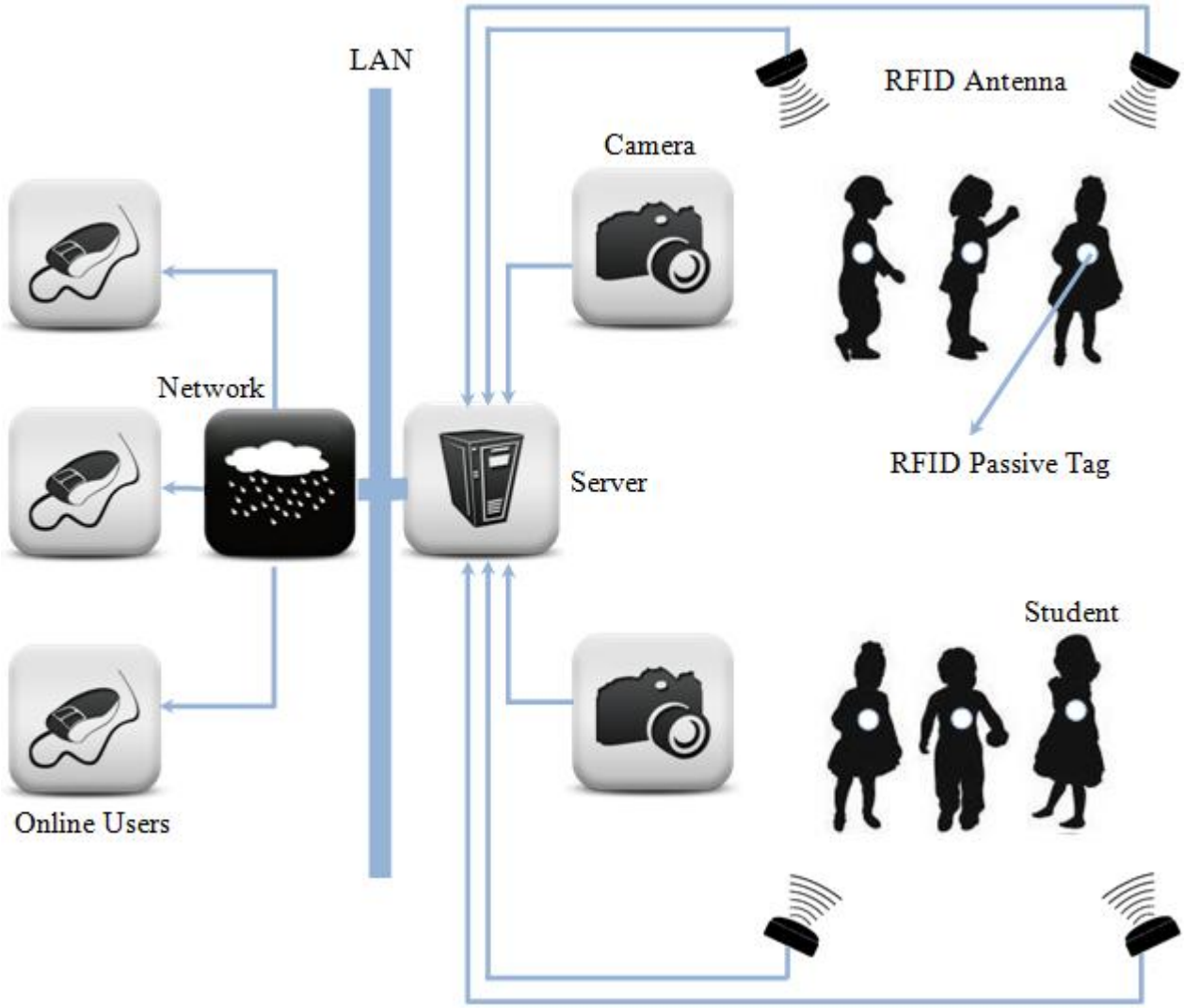


Figure 1.1

Fusion based Kindergarten students' online observing system.

"For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation."

plays with. Placing video cameras throughout the environment is easy, but selecting the right cameras and time slices is difficult. RFID tags can be placed on the play objects and on the children so that readers within the play space can locate and identify them as they move about.

Appropriate cameras can be selected for good views of selected children and/or objects. Alarms can be implemented for interaction between designated pairs of objects. Doing summarization automatically requires automatic identification of the children and toys, and perhaps their interactions. How much time does the child spend with toys versus other children? With which children did the child interact during the day? There is much current CV research on this, but use of RFID to identify children and toys can create a working system today. Motion and activity analysis can then be used to classify and select video segments. Although automatic activity analysis may not be robust, at least the parents will be watching the right child. There are other applications requiring similar functionality – for example, monitoring assisted living facilities or studying how shoppers examine items for sale in a store. Extension of this functionality to outdoor site management can enable some control of real time operations.

1.4 Capabilities and limitations of CV

Computer Vision has been very successful in controlled indoor environments, but challenged in uncontrolled outdoor environments. The CV literature contains thousands of reports on object detection and recognition, tracking, and motion analysis. Image sensors are passive, cheap, can be far-seeing, and can collect a good deal of information about a scene. The output images are conveniently interpreted by humans. Commodity cameras easily produce frame rates useable for most human motion analysis. Detections and relationships in a 2D image can often be mapped to the real 3D scene. Using multiple camera stereo, objects can be located in 3D. Or, special active sensors can yield range/depth images. Images are useful to sense the extent and pose of an object, its relationship to other objects, its motion trajectory and behavior. Vision requires a clear

line of sight and perhaps the object must be in a suitable *pose* (position and orientation) for detection and identification.

Object identification via CV, function *1.2(a)* above, is often difficult and is usually based on sensed features *1.2(e)*. Even accurate features may not precisely ID an object, e.g. who is that person or what year is that Chevy Cruz. ID by features, if possible and reliable, can be computationally costly given the necessary signal processing and the variation of appearance over many possible 3D poses. So, while person identification using carefully imaged biometrics might yield ID accuracies of over *98%*, more general object recognition via color camera image is far less accurate. Acquiring quality images under occlusion and variations in lighting also causes serious problems in CV applications. Uncontrolled outdoor environments might be dark, dusty, have rain or snow, and have both static and dynamic objects occluding the sensor view. Finally, CV may sometimes give too much information; for example, humans do not want to be imaged in private spaces, and may resent being watched in work places.

1.5 Capabilities and limitations of RFID

RFID applies generally in industry and business for automatic identification of items. Due to its improved functionality compared with barcodes, it is now replacing those ancestors wherever feasible. The major advantage of RFID is that its operation is independent of line of sight between reader and tag. The tags come with a read and write capability when there is onboard memory. RFID can also be combined with sensors to make sensory tags. RFID can easily and reliably provide a unique object ID by transmitting a digital signal to a reader. Such reliable

identification is often difficult using vision. With enough power, RFID can also transmit non-visual features of a person, such as name, weight, height etc. and in case of objects, their ownership, contents, and the full surface description of a 3D object etc. RFID can operate in smoke, and darkness, thus it is widely used in sales and inventory systems and is replacing bar coding when cost permits. Object ID approaches *100%* accuracy in commercial applications where objects are close to (presented to) a reader in a controlled environment. Objects with RFID tags can actually transmit their own physical description to an automated system or a security person. In robotics or material handling, the description might send a CAD model to a CV system to teach it how to recognize the object. RFID technology offers a wide variation in terms of cost, size, sensing distance, memory and processing power, and security [7]. With higher RFID frequency, higher data rate can be achieved. The higher data rate with appropriate anti-collision algorithm can enable a single reader to read a large population of tags.

RFID can even be used to locate objects. Although the inexpensive nature of passive RFID offers large scale utilization, cheap passive tags have the limitation that the tag can only provide its identity, but to acquire location information complex infrastructure is required. Some applications use active tags for localization. The Real Time Location System [8] that we have used for our RFID sensing is discussed in Chapter 4. RFID codes cannot be sensed by humans and hence can yield less overt ID and might be tolerated in private human spaces.

RFID requires that an object be physically tagged, thus changing the object itself and requiring that the object be “cooperative”. Although passive RFID tags can be tiny and do not require their own energy source, they are used for limited range and have limited memory.

Highly functional RFID tags require an energy source for communication, memory, and processing. An RFID tag is a proxy for an actual object so it or its communication can be counterfeit, thus making an object appear to be what it is not. Simple examples of this would be one driver stealing an EZ-Pass device from another driver, a shopper moving a tag from a cheap article to an expensive one, or two airline passengers swapping their RFID boarding passes. Thus RFID tags in critical security applications need to use encryption and secure operating system principles.

Table 1 highlights the relative strengths and weaknesses that will be discussed in Chapter 2 and referenced throughout this thesis.

Table 1.1

Advantages and disadvantages of CV and RFID.

CV versus RFID	*Computer Vision*	*RFID*
Advantages	▪ Passive feature-based sensing ▪ 2D array represents many properties of 3D world ▪ Cheap commodity sensors ▪ Similar to human vision ▪ Can recognize object and pose ▪ Yields human-useable displays	▪ Object oriented sensing ▪ Can provide arbitrary symbolic object properties ▪ Occlusion not a problem in several cases ▪ Object recognition not a problem ▪ Coarse object location possible
Disadvantages	▪ Occlusion prevents sensing ▪ Variations in lighting and object appearance ▪ Variations in 3D object pose ▪ Some object properties are not observable ▪ Object/background separation	▪ Active sensing recommended for distant location sensing ▪ Sensing distance and angle can be complex ▪ Greater distance requires greater power ▪ Tag content must be written ▪ Tags can be cloned or lost

1.6 Dissertation outline

The rest of this dissertation is organized as follows. Chapter 2 provides a summary of the related literature reviewed in developing the presented work. Chapter 3 provides the problem formulation and approach to its solution. Chapter 4 highlights location sensing with respect to stereo CV and RFID. Chapter 5 explains the fusion model. Chapter 6 provides details on tracking using fusion. Testing methodology and experimental results are given in Chapter 7 along with a discussion of limitations and problems. Finally, Chapter 8 presents the research conclusions and need for future work.

CHAPTER 2

Background

This chapter provides the necessary background material for understanding the work and discussion in the rest of this dissertation. Where required, a brief background about the topic under discussion will also appear in other chapters. This chapter discusses the functionality of computer vision and RFID and what are the constraints when these techniques are used separately. As the discussion progresses, the state-of-the-art projects are explained that have used fusion for detection, identification, location and tracking; however, some single mode applications are discussed to show how fusion can improve performance. We have also explained our work in the natural outdoor environment. The categories of these tag based vision projects have been made in the context of application defined functionalities.

Animal vision solves recognition and navigation problems in a 3D world. An organism needs to know what objects and activities are in its environment and what consequences these have. Visual sensing is often integrated with auditory or other sensing -- and also with memory -- for an organism to make decisions. A dog is attacking, a car is passing, a trash bin is heavy. Invention of radio frequency identification tags now provides the capability for an object to notify a nearby agent of its presence and properties; and, it may be that neither the presence

nor properties are observable via vision. For example, a buried steel drum can transmit a description of its contents; an unseen vehicle can warn of its approach, and a store item can tell that it has not been purchased. To start here the functionality and limitations of RFID is discussed.

2.1 The functionality and limitations of RFID

RFID technology is a mature and economically successful technology that can provide reliable symbolic identification of objects that are tagged, including features of the object that cannot be observed via vision. RFID technology has revolutionized manufacturing, distribution, sales, and transportation technologies by providing non contact, non optical object ID. Cheap low end RFID tags are substitutes for a printed bar code. Higher end RFID tags are communicating devices with a power source, memory and processing power, and are much like a wireless computing device. Moreover, RFID technology can not only provide object ID, but can also be used to provide object location. RFID enables all individual railroad cars [9] and many trucks in the US [10] to be recognized by a nearby reader and tracked via a network. The RFID based automatic vehicle identification system of Trans Core technologies is providing control and tracking of commercial ground transportation vehicles at busiest airports of US [11]. Other tracking solutions integrating RFID with GPS [12] and cell phones are currently available in world markets. In this chapter, we do not concentrate on how RFID can replace optical sensing or CV, but instead on how RFID together with CV can enable new or more capable systems.

Automatic Identification (Auto-ID) deals with information control and material flow problems. Contactless identification is an integral part of automatic identification. Contactless identification is an independent research area that combines varying fields such as telecommunication, semiconductor, cryptography, security, data protection and handling. Barcodes, magnetic inks, optical character recognition (OCR), smart cards, biometrics and Radio Frequency Identification (RFID) are automatic identification methods. RFID is a wireless technology since it communicates using Radio Waves. Want [13] describes the benefits of RFID technology, which include more reliable scanning, better tracking, integrated metadata management, reduced back-end communication, efficient label management, and wireless sensing. In addition to its obvious application in business architectures, RFID has also been integrated into robotics and artificial intelligence applications. In the subsections below, we consider several aspects of RFID, such as cost, wavelengths that are used, power that is needed, and distances and angles between object and reader. Different types of engineered solutions are available for different applications.

2.1.1 RFID operation

The major components of the RFID infrastructure are the tag, reader, database and a host application: Figure 2.1. The host application manages the RFID reader. It allows a reader, via radio frequency through its antenna, to activate and/or directly communicate with a transponder on a tag attached to an object. Access to the parking lot via an RFID tag presented a few inches from a reader at the gate allows the reader to remotely read and/or write data to the RFID tag.

The tag modifies the received signal and transmits back a modulated signal. The reader through its antenna receives this modulated signal and decodes the tag ID. The data related to the tag ID in the database is then used by the host application. For example, when a drum of antifreeze is loaded onto a truck, the tag on the drum and the tag on the truck are read in order to complete shipping and inventory records. RFID read distance ranges from few inches to more than *100 m* and RFID antenna read angle ranges from a pencil beam to 360^{o}.

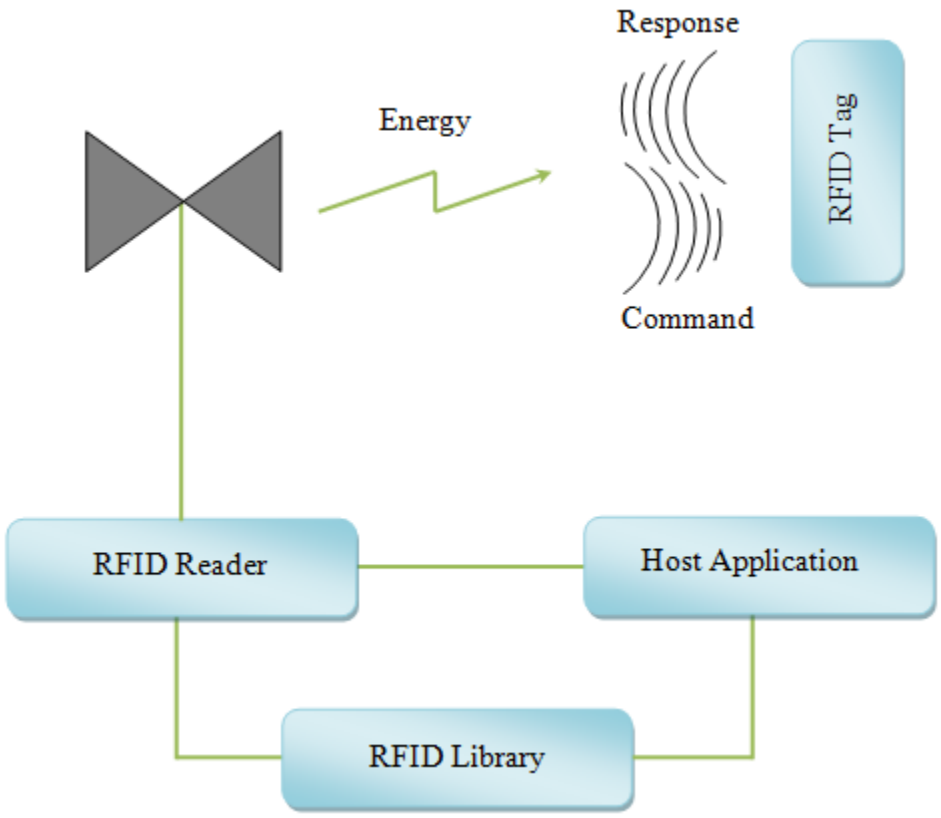


Figure 2.1

General infrastructure of an RFID system. The RFID tag may be many meters distant from the reader.

2.1.2 Types of RFID tags

RFID tags can be classified in different ways; however, we use the two groups commonly used in industry [14], – by power and by reading distance. First, depending upon the chip's power requirement, RFID tags can be categorized into passive, semi-passive (sometimes called as semi-active), and active tags. Moreover, the type of tag dictates the memory capacity of the tag. Passive tags consist of a semiconductor chip and an antenna. Electromagnetic energy received from the reader is used to power the response of the passive tag back to the reader. The read distance of passive tags is short due to limited induced voltage and reflected signal. Semi-passive tags contain a battery for powering only the logic in the tag; however, their communication principle is just like passive tags. Due to the onboard battery they can operate at long ranges compared with passive tags since they don't have to rely only on induced voltage. An active tag utilizes an onboard battery for communicating and for powering the tag logic and operates at ranges greater than passive and semi-passive tags. Unlike passive and semi-passive tags, the added benefit of an active tag is that it can initiate, as well as respond to communication with the reader. The broadly categorized RFID structure with their functionalities is listed in Table 2.1.

Table 2.1

RFID tag types.

Type	*Functionality*	*Power*	*Life span*	*Security*	*Communication*
Passive Tags	Purely passive and can vary from read only to read and write	Power from reader	Indefinite	Ranges from zero to low security	Response only
Semi-Passive Tags	Integrated sensing circuitry and onboard battery power to supplement received energy	Onboard battery	Depends upon battery life	Minimal to highly secured	Response only
Active Tags	Onboard battery, complex protocols and communication with active tags	Onboard battery	Depends upon battery life	Highly secured	Respond or initiate

A second classification of tags depends upon the field of operation, categorizing the tags as near field or far field. The near field tag's operating principle is based on Faraday's law of magnetic induction. The reader, through its antenna coil, induces voltage in the tag's coil. This induced voltage is then varied by the tag's onboard circuitry by changing the applied load, thereby encoding a unique tag ID. This varying induced voltage is then sensed at the reader. This method of sending data in the near field operation is called "load modulation". The far field tag operates on the principle of radio waves. The reader propagates electromagnetic waves using a dipole antenna. Some of this energy is reflected back from the tag dipole antenna due to impedance mismatch. Changing antenna impedance over time causes variation in the reflected signal strength. This pattern is used to encode the tag ID that is then decoded at the reader. This

way of sending data to the reader is called “back scattering”. The physics of these designs determine their cost and performance.

2.1.3 Radio frequency

RFID uses radio waves with frequencies from *125 KHz* to approximately *3.1 GHz*. Lack of worldwide uniformity of frequency regulation is hampering international standards of RFID systems. Though there has not been an international consensus on the frequency bands for RFID, typical RFID frequencies with read range comparison and respective costs are given in Table 2.2. A summary of the commonly used RFID frequencies with their different attributes is given in Table 2.3.

Table 2.2

RFID typical frequencies in use and respective ranges.

Tag class	*Band*	*Frequency Band*	*RFID Frequency*	*Read range*	*Cost*
Passive	LF	*30 - 300 KHz*	*125 - 134 KHz*	*0.2 to 0.5 m*	*$0.2-$0.8*
	HF	*3 - 30 MHz*	*13.56 MHz*	*1.5 to 3 m*	*$0.2-$2*
	UHF	*0.3 - 3 GHz*	*865 - 956 MHz*	*0.5 to 7 m*	*$0.2-$0.8*
Semi-Passive	Available in LF, HF and UHF band			*20* to *100 m*	*$4-$20*
Active	UHF	*0.3 - 3 GHz*	*433 MHz*	*≤100 m*	*$5-$50*
	MW	*2 - 30 GHz*	*2.45 GHz*	*≤100 m*	*$5-$50*
	UWB	*2 - 30 GHz*	*3.1-10.6 GHz*	*≥100 m*	*$5-$50*

Table 2.3

RFID frequency attributes.

Attributes	*Low Frequency*	*High Frequency*	*Ultra High Frequency*	*Microwave Frequency*
Frequency	*125 - 134 KHz*	*13.56 MHz*	*865 - 956 MHz*	*2.45 GHz*
Data Rate	Slow	Moderate	High	Very high
Range	Close contact	About *3 ft*	About *10* to *30 ft*	More than *100 ft*
Penetration	Penetrates water and tissue	Not good near metal and penetrates some materials	Does not penetrate metals	Does not penetrate metals
Tag size	Relatively bigger tag size	Thin construction but relatively bigger tag size	Small tag size	Small tag size
Moisture effect	No effect	No effect	Negative effect	Negative effect

2.1.4 Limitations and needed improvements

For maximum exploitation of RFID technology there is still a need for technical and security advancement. Following are some of the major limitations.

2.1.4.1 Standardization and cost

Business demands resulted in reduced RFID cost to replace other labeling technologies. However, there are two standards being generally followed, EPC governed by Auto-ID Center, and the ISO specified set of standards. Therefore, a definition of mutual commonality is required for global application and discerning of RFID coding.

2.1.4.2 Read accuracy

Read accuracy is important for a specific application and environment. It is affected by false-negative readings, i.e. missing a tag, and false-positive readings, i.e. detection of a tag that is not in range. Read accuracy is influenced by RFID reader design, reader and/or tag orientation and obstructions. A typical E-ZPASS installed on a highway toll station operates on *UHF (900 MHz)* with flawless detection of a low lying sports car to large trucks moving *5 MPH* [15].

2.1.4.3 Anti-collision

Anti-collision algorithms are applied to overcome signal collision and data loss during scanning of several tags at same time. The reader as well as tags can adopt an anti-collision algorithm [7]. Presently, anti-collision technologies allow simultaneous communication between a reader and up to *2000* tags in the reading area.

2.1.4.4 Security

An RFID system should ensure security at all interfaces to prevent unauthorized processes from reading or writing tag data. For example, a business must prevent a shopper from changing the price of an item and the toll authority must prevent driver A from charging an EZ-Pass toll to driver B's account. Higher level security features result in increased tag cost. Common security features are write lock, password protection, authentication, stream encryption and crypto processors [16]. Note that interfering with an RFID transaction is similar to disguising an object or incorrect feature detection in the optical domain.

2.1.4.5 Size and power

Nanotechnology is exploited to reduce the size of RFID tags. A few of the industry's lowest power microprocessor chips are reported by Phoenix as having *30 pW* of power requirement in sleep mode. For comparison, some passive tags power requirements range from $5.1\mu W$ to $25\ \mu W$ depending upon the frequency and read range. For passive tags, nano-particles are used to produce printable RFID transponders. Semi-passive and active tags incorporate a battery and thus the tag size and cost is increased; and, there is a need to consider the temperature limits as well as time of service. Toward a design without a battery, there is current research focusing on zero energy RFID tags being powered from thermal, vibration (piezoelectric), or solar energy.

2.1.4.6 Miscellaneous limitations

Having stated the main strengths of RFID towards fusion, we mention the weaknesses relative to our overall goals of object recognition and tracking. RFID tags do not reveal the appearance of an object – size, shape, color, etc. – unless it is symbolically encoded in the tag memory. CV here can also address this issue by providing visual analysis of the scene. RFID tags do not reveal object orientation, unless multiple antennas [17] or lattice of tags are affixed, which demands composite arrangements. In cases where initial estimate of the pose is predictable by RFID, CV can then be used for refined pose estimation and tracking. Finally, more power is required to supply more information to a reader and at greater distances – that is, active RFID is needed. Apart from indoor inventory management and similar long range settings, such a setup is generally required in outdoor environments where CV is mostly constrained due to lighting conditions and other weather effects.

2.1.5 Positional information

Within business applications the early focus for RFID use was for identification tasks only. However, with the expansion of RFID to location and tracking problems, position information comes into play; for example, where objects must be moved or grasped by robots or transfer systems. For positioning in outdoor environments GPS is often used, however its reliability in indoor scenes is poor. RFID identification systems generally lack positional information, thereby not providing direct information on the tagged object location. A network of RFID readers can be created, where the readers are used as artificial landmarks. An object can be located by being

near to a known reader at a known location. For example, using only RFID, it is easy to determine what cars are in a parking lot that has a reader at the gate, or in what hospital room is a tagged doctor, thus giving "symbolic" location. Coordinate information will be available according to the accuracy of the known reader location.

Methods that use signal strength or triangulation from multiple readers to compute general object coordinates are discussed more under Wireless Location Sensing (WLS) in Appendix D. For completeness here, we highlight a few reported results. In [18] the Average Error Distance using the active RFID tag infrastructure working at *433 MHz* in an outdoor environment is reported to be better than *7 m* within a range of *100~500 m*. In [19] the accuracy using *96-bit* UHF passive RFID infrastructure to localize objects in an indoor environment was *15 cm* within a 2×3 *m* area. In [8] CSL Technologies provide an economical off the shelf Real Time Location Systems (RTLS) using active RFID infrastructure at *2.4 GHz* in an outdoor environment with an accuracy of about *1~2 m* within a range of *200 m*. The system is in use for tracking of elephants in the Dallas zoo [20] and has been independently evaluated in [21]. RFID localization will depend on the number of objects and amount of environmental clutter in the application. Some other RTLS equipment providers are Ubisense [22] and AeroScout [23] .

2.1.6 Smart objects, networks, location-aware computing

The technology of wireless and mobile computing and communication is vast and changing rapidly. Functionally, a cell phone is much like an active RFID tag, the major difference being that the cell phone is designed to connect a human to a network rather than some other object.

Cell phones can have large memories and exchange arbitrary data across networks. They can act as GPS receivers and provide location information – hence a new field called "location aware computing". They also can contain accelerometers that can provide information on movement. Commodity pricing brings impressive power to cell phones at moderate cost. By using local multiple radio signals from known locations, a device or tagged object can locate itself within a few centimeters in a work area one kilometer across [24]. This supports precision agriculture, where precise location supports faster field operations, correspondence of work site points to aerial imagery or other sensor input, and more efficient application of fertilizer or pesticides.

Wireless location sensing has provided automation to indoor and outdoor systems. The outdoor location sensing systems are generally based on line of sight technologies e.g. GPS and cellular, while indoor systems use local positioning systems based on WLAN, bluetooth, sensor network, RFID, infrared and ultrasonic, etc. or combinations of these. There is a rich literature available on local sensing [25], [26], [27], [28], [29], [30]. We have briefly highlighted wireless location sensing in Appendix D.

2.2 Computer Vision functionality and limitations

Computer vision (CV) is concerned with extracting information about the real world from 2D images, or a retina of pixels. 2.5D and 3D "images" may be included, or can be considered derived from 2D images [31]. Human vision can provide all of the first five functions *a)* to *e)* given in Chapter 1 and much of the study of CV involves developing machines with such functionality. A recent survey of pedestrian detection from single video frames from an on board

automobile camera concluded that humans are skilled at detections whereas current algorithms perform poorly [32]. Since the vehicle safety application is of great importance, research will continue, and along multiple lines including both monocular imagery and fused input.

CV is related to both image processing and artificial intelligence, depending upon whether the ("lower level") image processing is emphasized or how sensed features are related to memory models in recognizing objects ("higher level") [33], [34]. Object recognition and visual motion analysis are two difficult problems for CV that generally involve several steps, each of which may be difficult. At the lower level, imagery needs to be normalized or interpreted relative to lighting and background so that image regions or boundaries corresponding to objects of interest can be extracted. At the higher level the features or extent of the object regions must be matched in some way to models of learned objects in memory. Matching of 3D objects can be complex due to unknown scale between the real object and sensed image, the large number of possible viewpoints (poses) and the effects of occlusion that prevents observation of some object features. Objects can be deformable or come in varied sizes and shapes, such as the human body. Moreover, even when there exists a workable CV solution, the computational costs can be high due to the large amount of image data processed and the large number of possible matches to memory. So, clearly there will be many applications that will benefit from using RFID for object recognition.

Figure 2.2 shows a camera *C* observing a workspace with coordinate system *W* containing two objects, pyramid *A* and cube *B*, each with their own local coordinate systems. There are several methods that *C* can use to calibrate its relationship to *W* by observing points of *W* [31] Ch

13, which would be necessary to understand the activities of *A* and *B* in the space. The shape of each object can be defined in terms of its own object-centered coordinate system. The relationship between objects *B* and *C* can be represented in terms of coordinate system *W*. The camera has its own 3D coordinate system, and its pose defines a projection from 3D space *W* to the 2D image plane *I*. The relationship between camera *C* and an object can be computed using three 3D points of that object and the 2D images of those points.

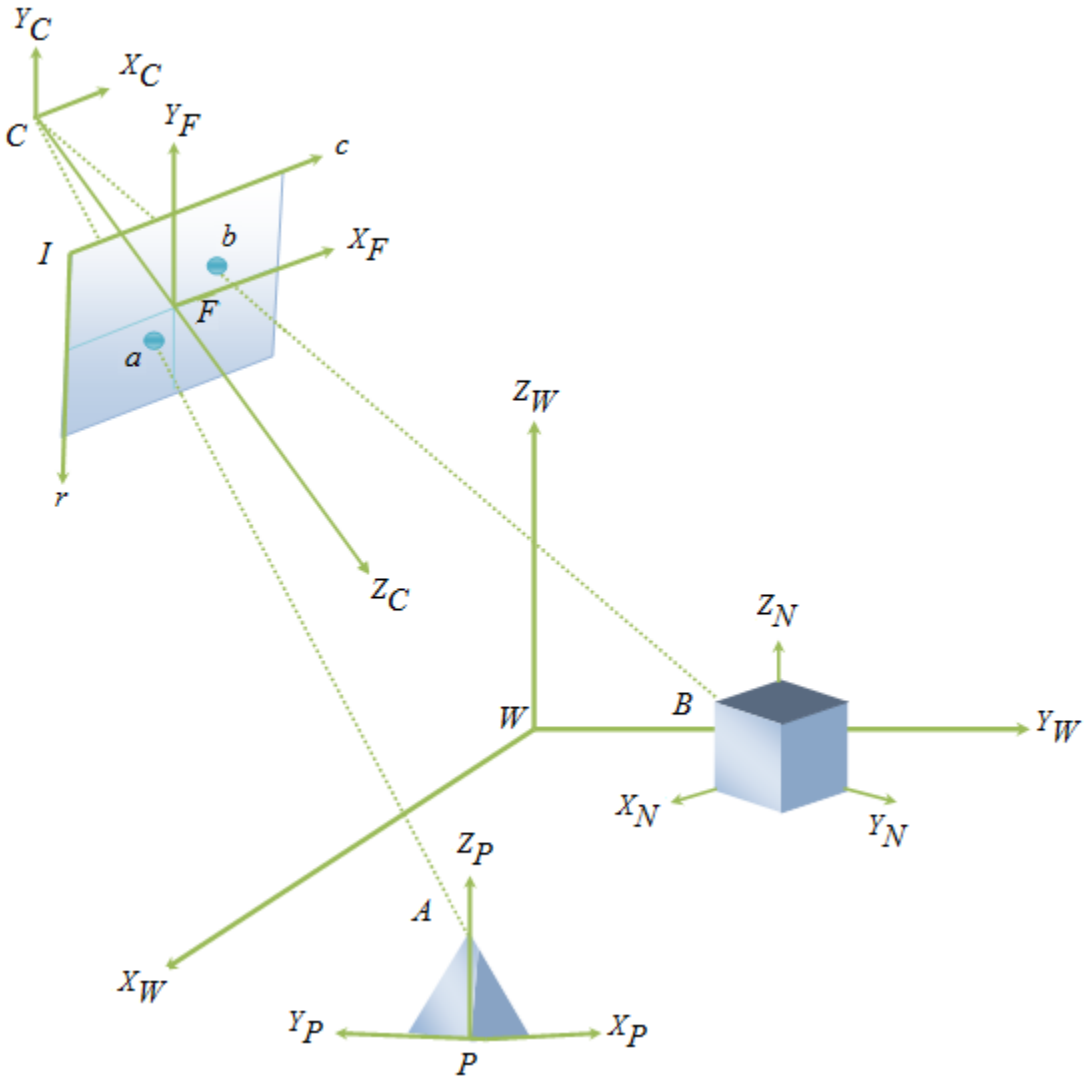
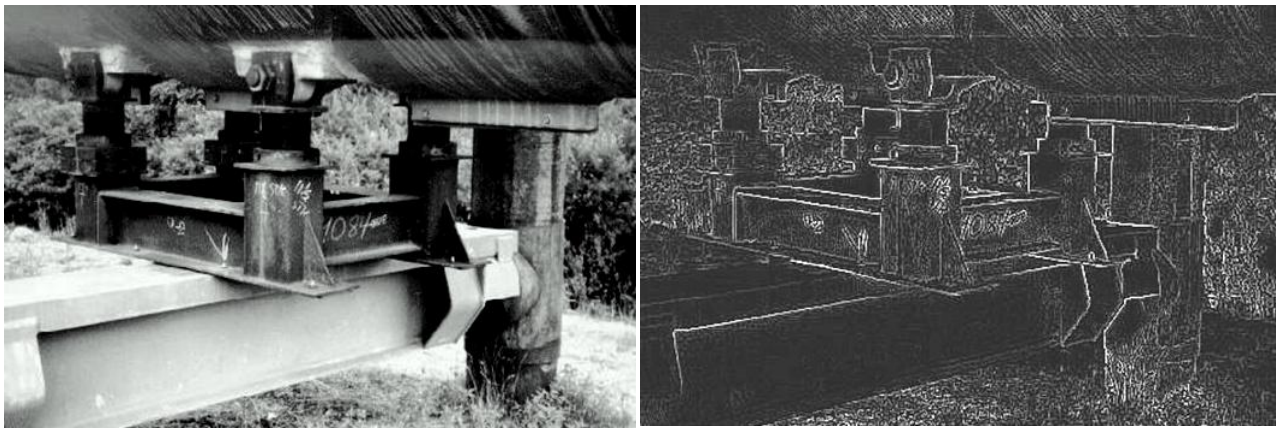


Figure 2.2

Relationship of camera C with pyramid A and cube B in workspace W.

Figure 2.3 shows an intensity image from underneath the Alaskan Pipeline. Because of the shadowing, the image intensities were stretched in order to show detail in the shadows. Meaningful image regions are difficult to identify; for example, the surface of the pipe appears striated by corrosion. By applying a Laplacian operator, points of high intensity change are highlighted. Some of these show object structures and corner points, which can be identified in the image using higher level image processing operations [31] Ch 10. If an inspection robot knows its global location and has some model of what it expects to see, it can compute its orientation relative to the pipeline structure in order to perform its work.



(a) (b)

Figure 2.3

Work robot seeing a section of the Alaska Pipeline via a) an intensity image and b) its Laplacian. A few corner points are evident that can enable the robot to get oriented to inspect and operate on the pipeline. See Shapiro et al. [31].

Figure 2.4 illustrates the "P3P problem": how does a known camera platform compute its orientation relative to a known object using the coordinates of three points of that object and a perspective projection of those three points. Fischler and Bolles [35] treat this problem, give a solution for P3P, and discuss more robust solutions when more points are known. The human head is a much-studied particular object that often has three observed points that can be used for computing head pose and possibly computing a normalized frontal view using that pose [36], [37].

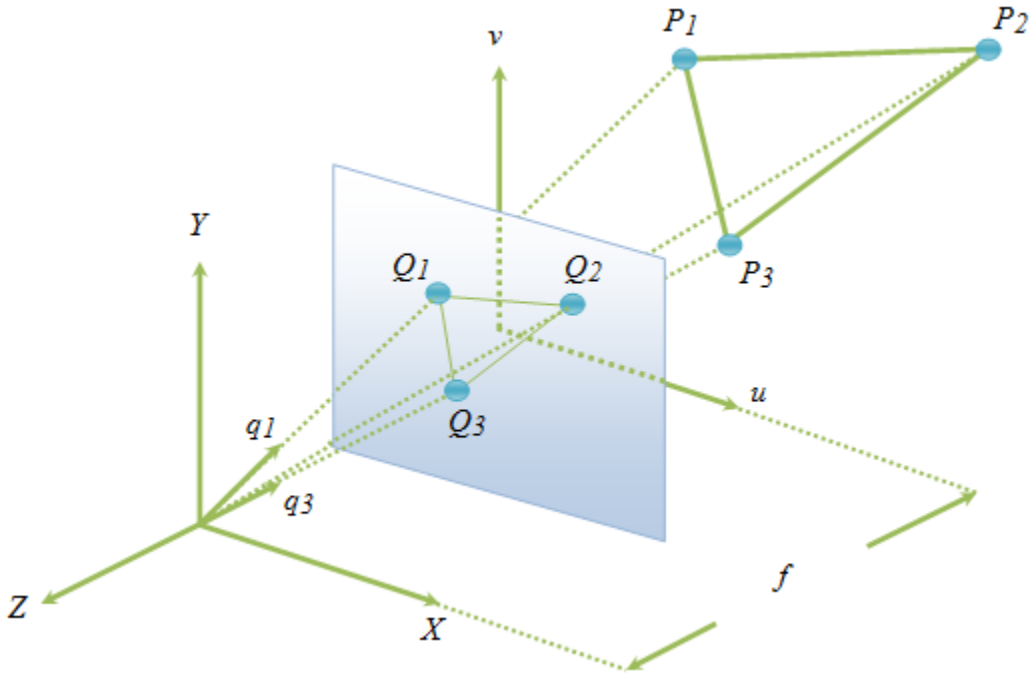


Figure 2.4

The Perspective 3-Point problem: a camera computes its orientation relative to three points seen from a known object.

Some areas of CV are not of concern here: for example, automated object inspection and measurement operate on precise representations of a known object and are not helped by RFID.

Image enhancement, restoration, coding, etc. whose outputs are again images are also of no concern. On the other hand, adding symbolic tags to raw video indicating what objects or actions occur within a video segment can indeed benefit from RFID. For image or scene understanding, an autonomous agent must know what the objects are and where they are in its environment. Some environments are controlled, meaning that possible objects, background, and lighting are known. Uncontrolled environments make trouble for autonomous vision, since objects, backgrounds, and lighting are uncontrolled or unknown. Environments can be "in between", for example, a soccer field or parking lot has mixed properties. There are major challenges to CV in uncontrolled scenes, some of which can be practically solved using RFID as noted below.

2.2.1 High image processing and search cost

A 2D image array records many relations in the 3D world. Although cameras can be cheap, processing an array of pixels can be costly. Even if viable object recognition algorithms exist, they may be expensive in time and memory due to both the large number of pixels and the large number of operations on those pixels. In addition, recognition implies stored object representations and a matching algorithm, which imply memory and computational cost [38]. Computational cost rises with the number of possible objects. RFID clearly can alleviate these problems by having objects declare their presence to a reader. The observer/reader can then request an object model from the tagged object, or from a network using the object ID as key.

2.2.2 Optical sensing problems

Several distortions are present in viewing the 3D world via a 2D image, which may interfere with object extraction or matching for object identification, e.g. lens distortion, lighting variation, digitization noise, and object surface variation. Computation may be required to restore proper object features. Extra computation is needed for partial matching when a perfect object representation cannot be extracted from an image. Excessive computation and the uncertainty inherent in matching partial representations can be avoided if RFID can reliably provide object identification, and possibly even object coarse location.

2.2.3 Object modeling

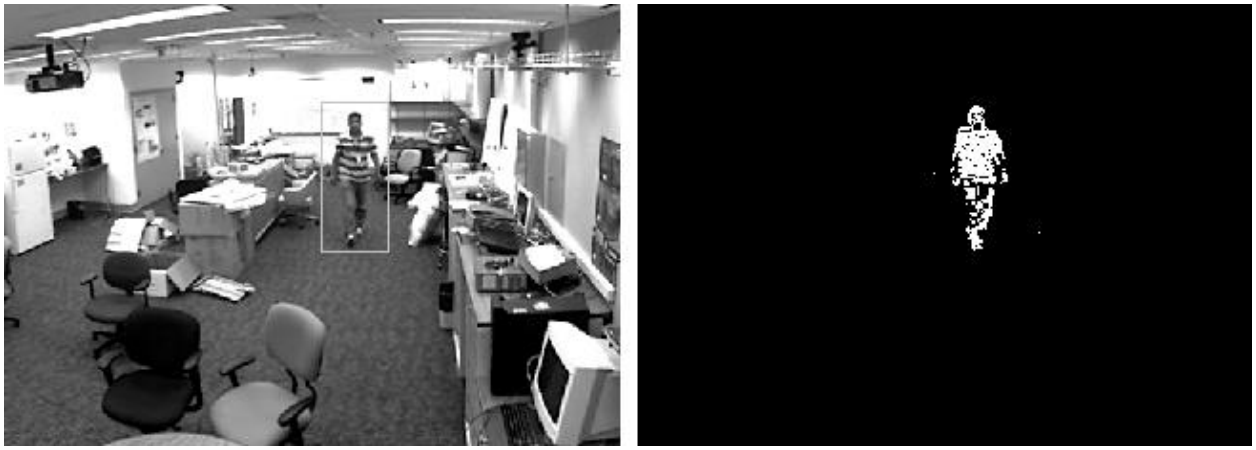
Irving Biederman [39] stated that a six-year old child might recognize *30,000* different objects while having a verbal vocabulary of only a few thousand words. The variety of objects, both man-made and natural, makes general object modeling extremely difficult. Imagine a junk yard robot tasked with sorting all the refuse of society! A grocery store or airline lobby is also very complex. Different types of object models have been proposed for different types of objects and applications. Learning or teaching of objects and the recognition algorithms vary with the type of object model [40]. Three common types of object models are appearance-based, feature-based, and geometric-based. Appearance-based models represent an object, or object part, based on the sensor representation of it [41], [42]. Feature-based models typically represent an object as a fixed length vector [list] of features computed from that object [43]. Geometric-based models typically represent an object as an aggregate of vertices, creases, surface patches, etc. in

an object coordinate system. The type of model determines how it is created or learned and how it is used in recognition. Some models may have to be changed even while in use – for example, an object tracker using an appearance-based model has to continuously update the model during tracking as the lighting and image shape changes. If an object can transmit its own model information via active RFID, both the search of the observer's model memory and the combinatorics of matching model to observations can be greatly reduced.

2.2.4 Object tracking

Recognition and tracking an object in an image sequence is one fundamental problem of computer vision. The goal is usually to recognize a moving object and analyze what it is doing. If the observer is in motion, objects that are stationary in 3D will yield apparent motion to the sensor complicating analysis, as in the case of a moving car and moving pedestrians. Figure 2.5(a) shows an image from a staring still camera monitoring a workspace. Entry of a person is detected by a change in region statistics over a few video frames. Simple change detection greatly simplifies segmentation; however, object detection and ID remain as problems. Related applications are motion-based recognition, automated surveillance, video indexing, human-computer interaction (HCI), traffic monitoring and vehicle navigation. Modeling object appearance and movement in vision-based methods are both computationally complex. Statistical and area processing methods of CV might be replaced by an engineered RFID solution. Khan and Shah [44] survey background on various image processing approaches to tracking objects and present a novel method for tracking people moving on a plane by combining information from multiple single image viewpoints. While their viewpoint combination method

performs well on the tests, it can be stymied by strong occlusion and by intersecting object tracks. They describe additional appearance-based methods that might help remove these ambiguities, but probably not as well as can fusion of RFID.



(a) (b)

Figure 2.5

Still camera monitoring a workspace: a) workspaces can be monitored by staring cameras. b) object motion can be detected by changes in region statistics.

2.2.4.1 Object recognition

Performance of recognition and tracking systems strongly depends on their ability to detect and identify objects in some environment. The motion of the object may be necessary for its identification. The detection of an object might be performed in the first frame or in all frames. The complexity increases due to false alarms and false dismissals and also due to objects actually entering or leaving the observed space. Some of the CV processes used in detecting objects include feature point detection, background subtraction, supervised learning, and segmentation

[32]. Due to the projection of 3D to 2D and due to articulation of some objects, whatever model that the observer is using has to change over time, adding more complexity to the tracking task.

2.2.4.2 Object location and pose estimation

Pose estimation is the process that estimates the position and orientation of an object in some coordinate system. Mathematically, we need to determine the three angles, or orientation parameters, and the three position parameters orienting and locating an object in the 3D coordinate space. Pose is used by a mobile platform for collision avoidance or interaction with the object. Some previous related work is given in [45], [46]. Figure 2.2 shows a camera viewing two objects in a workspace W. The application may need to compute the pose of each object relative to the workspace coordinates, as in the case of surveillance; relative to the observer, e.g. in the case of a robot operating on the objects; or relative to each other, as in the case of activity recognition. Computing the pose of the observer relative to an object has already been introduced and sketched in Figure 2.4.

2.2.4.3 Sensing and relating 3D object points

Pose in 3D is simplified by sensing 3D points on the object rather than just 2D points in an image. Stereo can be used to do this, as shown in Figure 2.6. If two cameras with known pose in 3D space *W* observe the same object point *P*, then *P* can be located in space *W* by intersecting the two camera rays in space. (Due to approximation errors, the closest approach of two rays is actually computed. See [31] Ch 13. If more than two cameras are used, then robust analysis can be used on a set of approximate ray intersections.) 3D sensors have been constructed by

packaging multiple cameras. Or, if the camera C_2 in Figure 2.6 is replaced by a laser beam or sheet of laser light, then a "structured-light" device is created. LIDAR scanners can compute range to a 3D point of an object surface by comparing the phase difference of a modulated light beam sent to and reflected from that surface. Thus there exist unit "range sensors" that can sense an entire scene as a set, perhaps a dense set, of 3D surface points [47], [48], [49], [50] and some current cars have these for collision avoidance [13].

Having 3D points greatly helps in scene segmentation and object shape analysis relative to having only 2D image features; however, it does not make segmentation and recognition easy. Once a set of 3D points is available from an object surface, they can be matched to a model surface using the general Iterated Closest Point Algorithm [51] to compute relative pose as well as quality of match. Approximate pose is required as a starting point. Point or surface matching is extended to a moving platform with SLAM (Simultaneous Localization and Mapping) [52]. By matching in 3D, we have seen that a sensor can compute its pose relative to 3D object/scene points. The sensor can then move and compute its new pose; moreover, it can compute the pose of newly observed scene points relative to formerly observed scene points and thus grow a map of a scene being explored.

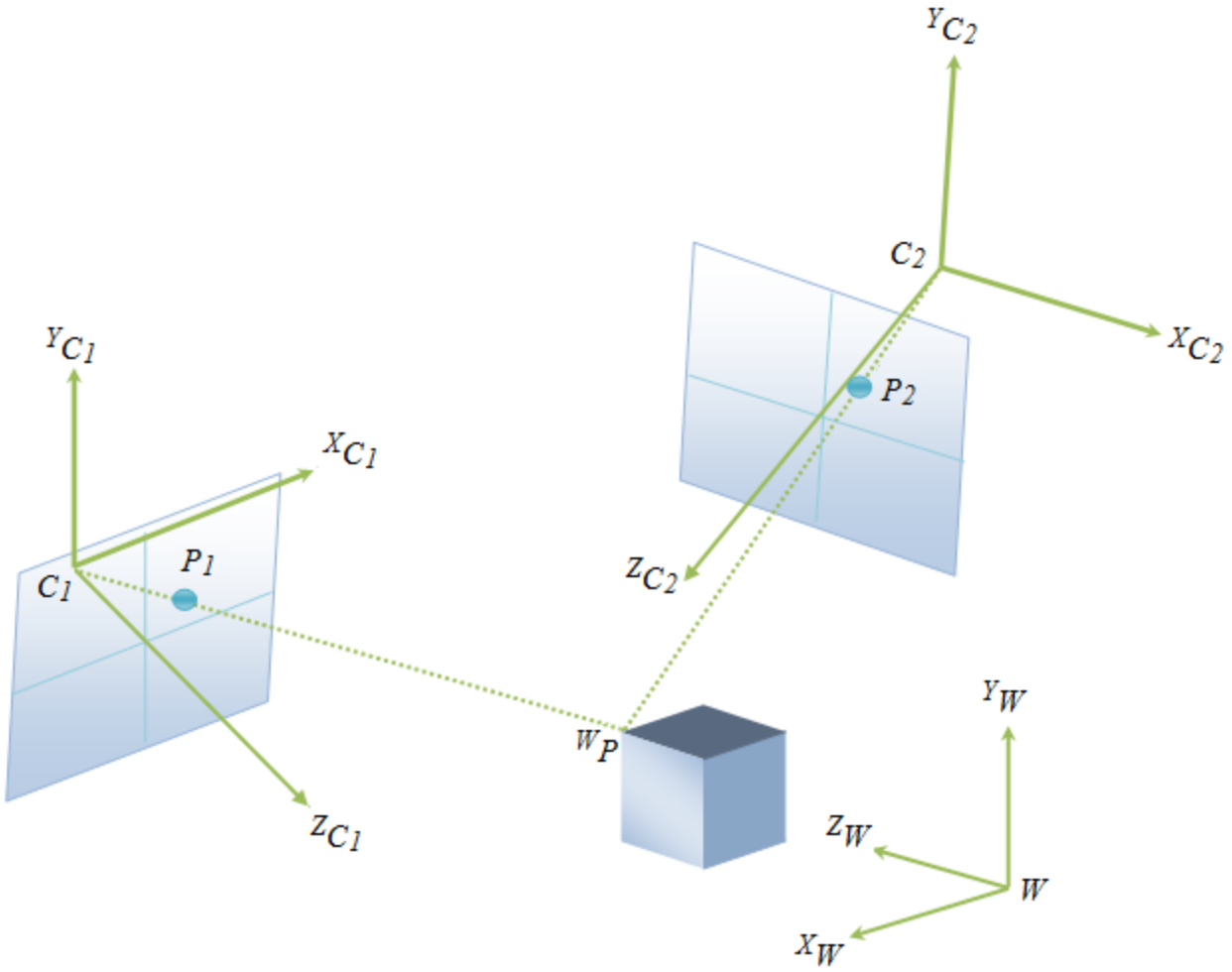


Figure 2.6

Pose in 3D simplified by sensing 3D points - Two cameras C_1 and C_2 knowing their relation to the space W can triangulate to compute the coordinates of a point in that space. That point must lie on the intersection of the two imaging rays [31] Ch 13.

2.3 Research projects based on fusion of CV and RFID

Various methods for fusing the visual and tag sensing data have been proposed. These revolve around the basic functions of detection, identification, location and tracking. The

categories have been made in the context of fusion-oriented applications. Within the context of identification and tracking, the general architecture of tag-based fusion is given in Figure 2.7.

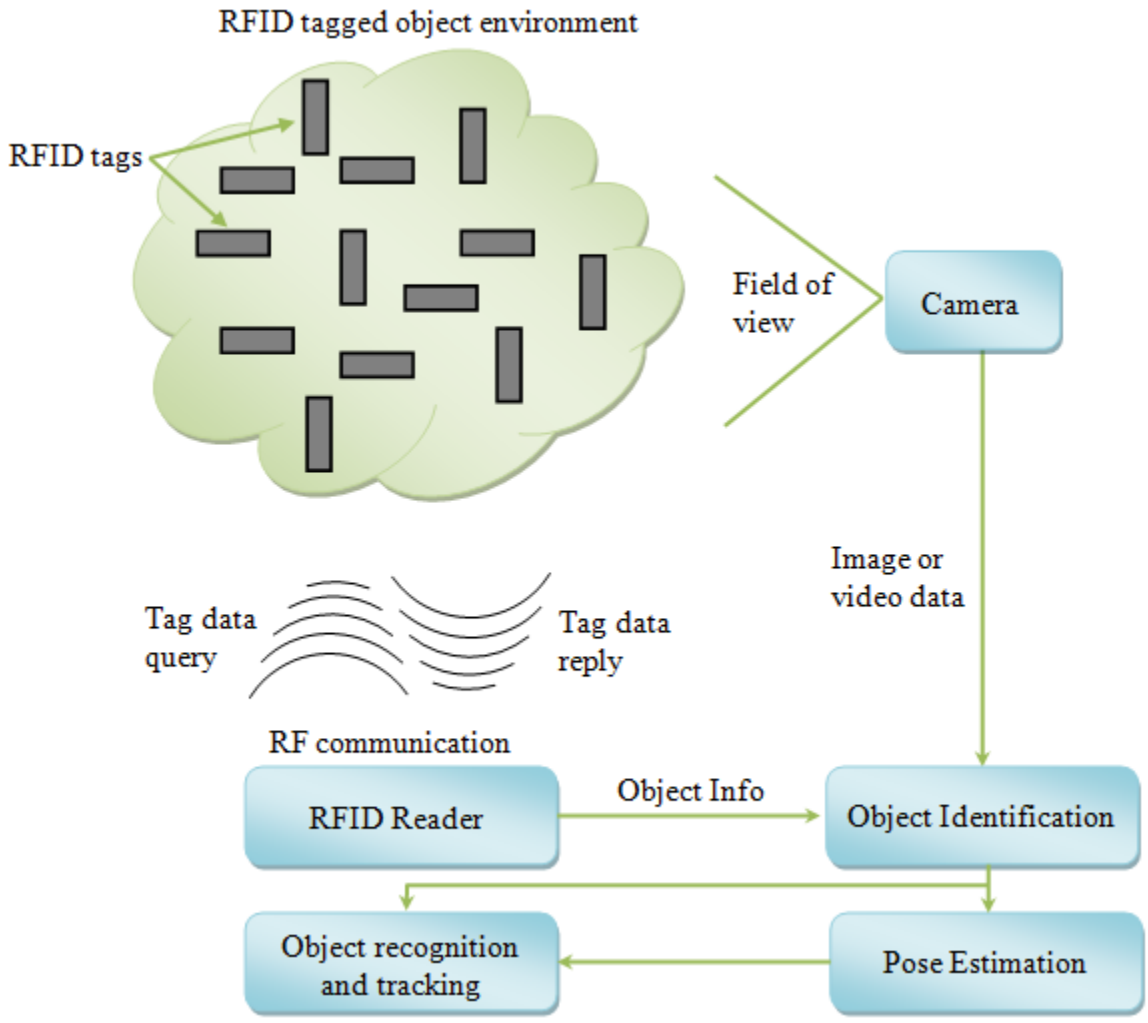


Figure 2.7

General architecture of RFID and vision fusion.

2.3.1 Model based recognition

One of the earliest approaches using model based recognition and RFID is proposed in [53]. The algorithm identifies an object using RFID and then recognizes it in the scene using an appearance model stored in the object tag. The algorithm compares the observed model with the stored model from the tag and recognizes the object if both models match. If the object is not recognized then it is considered to be an occurrence of a new object with no prior appearance model. The system acquires the object model and saves it in the tag. Using edge data, a model is generated and stored. To accumulate many models, Eigen space analysis is used. Eigen space is updated every time an object with model is observed. Only fixed rigid objects were used for the experiments. On the same lines Boukraa and Ando [54] have reported a 3D scene analysis architecture for polyhedral shaped objects. The object identification is performed using *2.54 GHz* passive tags with a read range of *1.2 m*. The unique tag ID is received from the tagged object. Using that tag ID, the object model is then located in a model database, through a network. The vision system then detects lines and edges and projective matching [55] is used for registration. Therefore, the object recognition task is reduced to registering the object model to the observed image and the recognition part is independent of the number of models in the database. In [54], [56] Boukraa and Ando used their knowledge-based recognition algorithm only for single object scenes with polyhedral shapes; whereas natural scenes are filled with free form objects.

Cerrada *et al.* [57] approached object recognition and localization for free form static objects in complex scenes using fusion. 3D information of the objects is generated using range sensors. For vision only based techniques, recognition and localization are costly computational

algorithms due to the uncertainty of the objects in the current complex scene. The fusion approach reduces the original database to a number of objects in the current view. Their scheme presents comparative results with and without RFID. The RFID reader identifies the objects in the tagged environment with a list of read tags in the read range but does not provide location information. The information is fed to the Weighted Cone Curvature [58] stage from which an initial partial view estimate is acquired. Reduction in the original database is achieved by carrying out comparison of principal components and partial views. Finally, for object recognition and localization, the difference between two clouds of points is minimized by the Iterative Closest Point algorithm [51]. Figure 2.8 shows the block diagram of the proposed 3D recognition method. The validity of the object recognition algorithm in [57] was constrained by having always the same number of objects in the scene. The authors further generalized the methodology in [59] by allowing the number of objects in the scene to range from *4* to *20*. The authors used RFID in the segmentation stage since RFID can easily give the number of objects in the scene along with their description. Their approach deals with the paradigm of object recognition for complex 3D scenes having medium and large databases. The statistical analysis provided in Table 2.4 estimates a linear regression model relating number of objects in the scene with the recognition time reduction. Recognition percentage increases by only *6%* using RFID, but the computation time is reduced tremendously by *74%* on average.

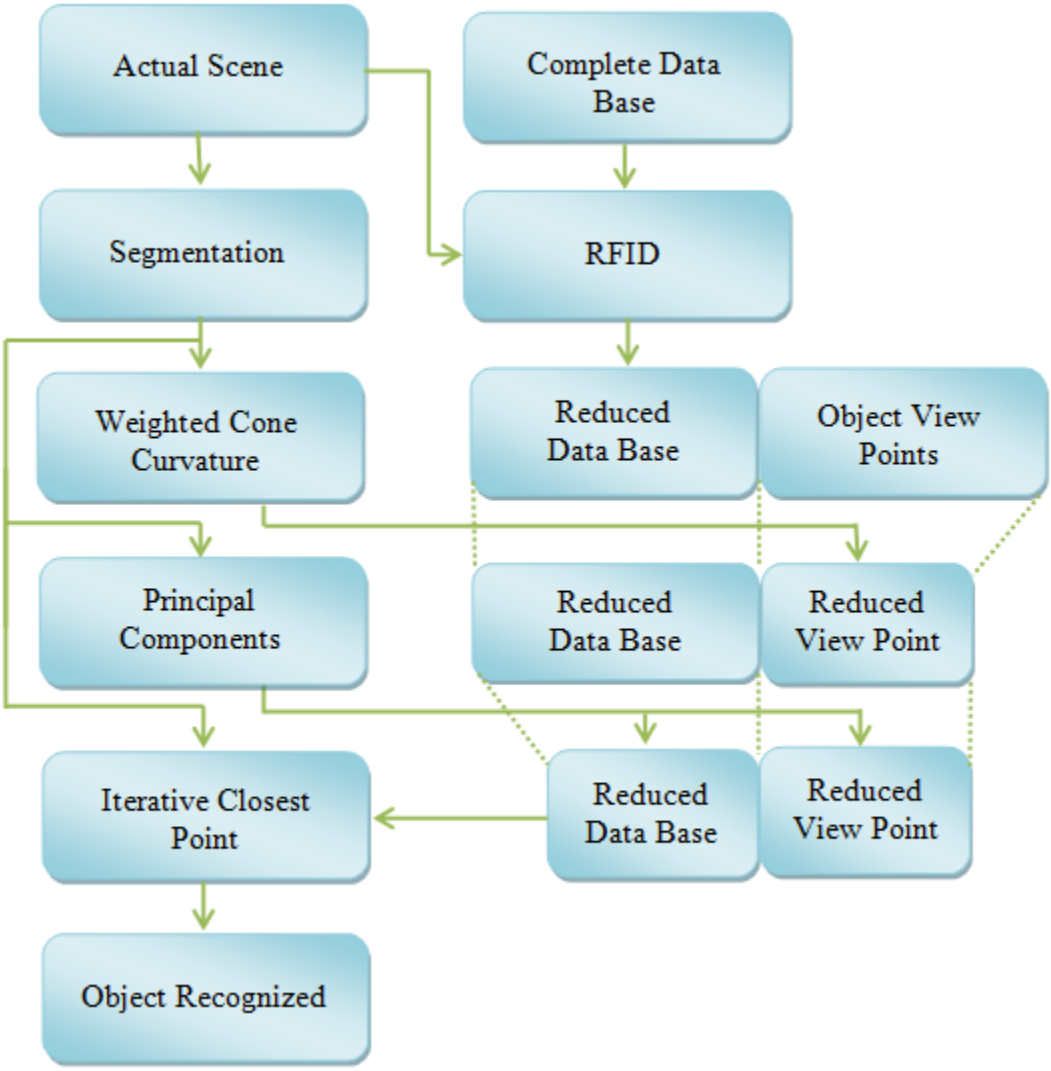


Figure 2.8

3D objection recognition algorithm. Recreated from Figures 3 and 5 of Cerrada et al. [57].

Table 2.4

Obstacle recognition results. See Cerrada et al. [59].

No of scenes	*No of objects in scene*	*% Recognition*		*Total time(sec)*	
		RFID	*w/o RFID*	*RFID*	*w/o RFID*
12	*4*	*91.7*	*86.1*	*2.45*	*9.44*
9	*3*	*86.1*	*83.3*	*2.96*	*9.19*
	Average	*89.3*	*84.9*	*2.67*	*9.33*
	Standard Deviation	*16.3*	*17.0*	*0.47*	*0.3*

The CAD model concept is also used by Hontani *et al.* [60], [61]. Some CAD based models for computer vision are summarized in [62]. In [60] the proposed system uses a combination of visual retro-reflector tags and RFID tags. After obtaining object ID from the RFID tag, the system gets an initial estimate of pose from the visual tag and then visually tracks objects by a model-based tracking method. The tracking algorithm updates object orientation and position by detecting edge movement in two consecutive frames. For building and learning the models in real time, vision algorithms for model-based recognition require user interaction. In [61] the system identifies the tagged object using RFID. The tagged object CAD model is then retrieved using the internet through a URL server. After determining a shape component in the captured image, the system selects an algorithm for initial estimate of pose relative to the camera. Thereafter the system starts tracking the object in front of the camera. The tracking is based on the difference between the visible edges and edges of the projected model.

Similarly, for a test-bed having a robotic manipulator arm to clear dishes from a table, the authors in [63] used RFID passive tags placed on static objects to identify and retrieve object model and information from the database. The reader is placed on the robotic arm. The tag system is used for object recognition, vision for object localization and the robotic arm for object handling. Using pre-stored template images the ceiling mounted vision system provides location information transformed into robot coordinates. This information is then used to position the robotic arm and execute predefined commands to interact with the objects. Here RFID increased the accuracy and speed of the vision system. The group is also working on calculating the orientation of the object based on received signal strength from the static tags on the objects. In another recent approach, Kim *et al.* [64] used a robot manipulator system for object recognition.

The proposed infrastructure uses self-fabricated smart tags having an active landmark (IRED) and a data structure consisting of geometrical, physical and semantic information. IRED is activated as soon as the tagged object comes in the read range. The robot then searches the shimmering light pattern produced by the IRED within the scene. Subsequently the object's depth, size and pose is calculated using model based vision from its stereovision on a pan-tilt mechanism. The manipulator can then interact with the object.

2.3.2 Human-object interaction and activity analysis

Within successful applications of vision-based approaches, human-object interaction and behavior and activity analysis are broadening. The problem domain ranges from tracking moving humans and objects to ubiquitous learning environments. However, based only on vision and image processing algorithms, the development of reliable solutions is still very difficult. A survey on human action recognition methods using vision is provided by Poppe [65]. The author has discussed limitations of vision state of art techniques. For data acquisition and to identify the human, biometric methods in speech recognition, image processing and computer vision are required. The RFID system, like other sensory devices, can provide such data. Like other fusion techniques in behavior and activity analysis, the key is to get the proper combination that can make better decisions and produce higher classification accuracy. Combinatorial fusion analysis is a growing research domain for analyzing data fusion methods from multiple scoring systems [66]; however, we focus only on fusion of RFID and vision. Hsu *et al.* [67] have proposed a layout for learning behavior monitoring. All the objects on a desk [books, stationary etc.] are tagged with an RFID reader under the table. A camera is used to monitor behaviors such as

away, sleeping, studying a particular subject, or doing homework. The system is designed to help the learner improve his/her learning efficiency compared to a planned schedule.

N. Krahnstoever *et al*. [68] proposed a robust real time tracking system for analyzing human and object interactions and did prototype experimentation on a shelf holding objects with varying size and shape. To get the important actions and interactions between the humans and the tagged objects, the system combines stereo based articulated motion tracking and RFID based tracking. The RFID module provides the presence and orientation information and the vision system tracks the human body parts such as head and hand. The object pose is estimated using tag orientation. The orientation and angle of the tag relative to the antenna can be approximated using received signal strength. With three RFID antennas tag orientation can be determined accurately. The vision system target model is a low dimensional approximation of the human upper body. The authors claim that the system can easily detect which item the user is interacting with, which would be a challenging task for a vision-only system. Also it can recognize that the user is probably reading the label on the item, which would likewise be difficult for an RFID-only system, since it can only estimate the orientation of items and has limited range. This research adds to the application areas that require user tracking and interactions.

Another related contribution in activity recognition is given by Wu *et al.* [69]. Most of the objects are tagged with the user wearing an RFID bracelet. Their proposed system uses a Dynamic Bayesian Network (DBN) framework for learning object models by modeling the correlation between events presented by the RFID and video data. Without any explicit human supervision, the method automatically acquires object models from video and provides the most

likely activity along with common-sense knowledge about which objects are likely to be involved. Additionally, the untagged objects in the vicinity are also identified intuitively. They have used skin color models and do segmentation using change detection. For object model representation, the Scale Invariant Feature Transform (SIFT) [70] is used to extract feature descriptors with maximum likelihood estimates for learning. The experimental setup handles 16 different household activities with *33* objects in the tracked environment. The unsupervised approach is used to learn the object models. The learned models are then used to infer the activity and object labels for the same video. Table 2.5 shows the experimental results from [69]. It is atypical to see comparable performance with vision only as compared with the outcome from both vision and RFID. This indicates that the Bayesian framework has utilized all the useful information in RFID and after the object models are learned, RFID is no longer useful. Their system lacks view independence due to a single camera and lack of learning motion information required for human-object tracking.

Table 2.5

Activity and object recognition rates. See Jianxin et al. [69].

Common sense used	*Testing sensors*	*Activity*	*Object*
Yes	RFID only	*64%*	*63%*
Yes	RFID+Vision	*81%*	*72%*
	Vision only	*81%*	*73%*
No	RFID+Vision	*61%*	*75%*
	Vision only	*63%*	*75%*

Park and Kautz [71] extended the approach in [69] and addressed above limitations. The proposed method deals with incorporating human and object models and also building a DBN

that recognized the human-object activities of daily living in a smart home. For obtaining view independent recognition, they used multiple cameras to attain a multi-view vision system. The vision system performs track-level and body-level analysis to attain a coarse and fine level recognition. The RFID module, having hand worn reader (iBracelet from Intel) and tags, is used for learning temporal segmentation of motion and object identification. The detection range of iBracelet is *10* to *15 cm*. As the person's hand approaches a tagged object, the reader detects the tag and transmits wirelessly the time stamped ID information to the PC-based activity recognition system. To generate the activity model, six coarse-level actions are coded for investigation. Reported activity recognition results show that different sensing modules better indicate different activities i.e activities such as "walking around" are better recognized with vision, while activities such as "preparing cereal, drinking water" are better recognized by RFID.

Continuing their work on human-object interaction in [17], [72], Deyle *et al.* [73] have recently presented an approach of constructing RF signal strength images from RFID to be used as a distinct sensing modality. Low cost UHF tags are used in their system. By measuring the signal strength at each bearing, RSSI images are generated by panning and tilting the readers. RSSI images provide ID specific features of the object in range. The RSSI image and the 3D point laser range data are transformed into the 2D camera images. By using a probabilistic framework, these RSSI images are fused with visual and laser data for generating a maximum likelihood 3D point estimate of the tagged object. This information is in turn used by the autonomous mobile manipulator to approach the identified object. The interaction with an object is then remotely specified by the user using the context aware remote user interface. While

performing several test trials, the algorithm is validated in an indoor unobstructed scene by achieving authentic localization.

Nemmaluri *et al.* [74] present a system named Sherlock to automatically recognize, locate and index tagged objects. This system is built on their previous work with Ferret in [75]. The primary difference is that Ferret used hand held readers and the Sherlock system infrastructure has fixed readers capable of controlling their own movement using steerable antennas and cameras. Sherlock scans in three different ways i.e fast, coarse and localize. While providing little position information, fast scan determines all objects in the environment and provides regions. Coarse scan gives a rough estimate of objects present in a particular region while localize scan provides position information of a desired individual object. The core component of the system is the fine grain RFID localization subsystem that can precisely recognize and locate objects. The overall cost of the equipment is high. The RFID reader used can control four independent antennas. In addition, the system has an integrated pan-tilt-zoom camera mounted at a fixed location. It provides an interface for query and visual system for user interaction and display. This system is used to help people interact with their moveable belongings in a realistic office environment. Sherlock can localize *90%* of objects in a volume of *0.55* m^3 and can localize *100* objects having passive tags in approximately *12 mins*. Antenna movement determines the worst case scan time.

Tracking humans in cluttered scenes by a mobile robot also requires effective interaction with the surrounding world. Some literature suggests utilizing multiple onboard sensors or visual measurements. In [76] Germa *et al.* provides initial adequate results for human tracking using a

mobile robot. They have modified a mobile platform with additional onboard capabilities such as a monocular pan tilt camera, WiFi, gyroscope and RFID reader with eight antennas for omni directionality. Their multi-modal tracking algorithm applies a particle filter for the heterogeneous data fusion in a stochastic framework. They have utilized vision to provide closed loop position control for a robot end effecter by designing three vision based PID controllers. This control strategy is helpful in providing required feedback control during visual information loss. The proposed infrastructure also roughly estimates collision detection in high-risk areas. Validation of the whole infrastructure is performed in an indoor environment with sporadic occlusions when tracking a tagged individual. As an extension to this work, the authors suggest future research for multi person tracking with better collision avoidance while focusing on overcrowded scenes and coarse measurements by RFID-based distance evaluation. For checking the system robustness to occlusions and target loss, the authors considered the ratio of the frames while the user is in view to the total number of frames. For determining this ratio for tracking one to four persons, the ratio decreased from *0.22* to *0.19* with a vision only system as compared with *0.93* to *0.85* in the fused system, thereby demonstrating the fusion effectiveness.

Blind people make use of tactile and haptic perception (the process of recognizing objects through touch and kinematics). The blind assistance devices available in the market require strategies for efficient understanding of the unseen environment. T. McDaniel *et al.* [77] have suggested a framework for integrating RFID and computer vision in enabling devices for remote object perception. Seeking a wearable device, they used vision and touch features, which can be classified at the perceptual level. The vision module provides users with only relevant information found through RFID identification. Their algorithm efficiently deals with RFID data

rate overload by particular content selection using vision and applies to untagged environments as well by gathering tactile features (shape, size, texture, material, etc.) from visual data. As future work the authors have suggested experimentation for usability of the proposed system in a real environment.

2.3.3 Mobile robot localization

For a mobile robot to accurately obtain a target pose it needs guided input from the navigation module, which in turn depends on a localization algorithm. The robot actual position and target position varies due to problems such as wheel slippage. These errors accumulate over time. Dead reckoning alone is inadequate in this scenario and additional sensory input is needed. Multi mechanical sensors, RFID, and vision techniques have been proposed to solve this problem. Some fusion based methods are discussed below.

2.3.3.1 Map- based navigation

Chae *et al.* [78] proposed a global to fine localization algorithm for a mobile robot in an indoor environment. They used *915 MHz* active RFID tags with *6 m* detection range as landmarks for achieving global localization. A mobile robot with camera onboard was used. The mobile robot movement area is divided into tagged regions and a visual map is built with known position of each RFID tag. For global localization, the algorithm assigns appropriate weights to each of the detected tags depending upon the distance of the found tag from the boundary of the region. To detect and describe local features in images, the authors have used SIFT features [70]

due to their stability against pose and lighting variation. This feature descriptor is used in a specific region to fine tune localization of the mobile robot, and its comparison with the visual map gives the current view angle of the robot. Given the tagged surrounding and the feature descriptors, the robot localization problem is narrowed down to feature-matching. In a work space of $6.2 \times 7.8\ m$ mean localization error of $0.23\ m$ is reported.

Weiguo *et al.* [79] have used RFID tags placed on the ceiling with a camera onboard the mobile platform to get relative positioning. RFID tags are used as visual landmarks. A topological map of indoor surroundings is built using adjacency relationships of the tags in the surroundings. The camera distance from the ceiling is kept constant thereby simplifying relative position and orientation calculation. The mobile platform traveling in the center uses direction and heading angle information from the identified node in the field of view of the camera as shown in Figure 2.9. The path planning is then post-processed as per the current heading and direction. Another object localization scheme for a mobile robot in a home environment using ceiling cameras is reported by Kamol *et al.* in [80]. The feature information for each object is also stored in each tag. The algorithm uses RFID to get rough location of the tagged object. For a precise estimate, the system recognizes the object features using a hue-color histogram with a subsequent location estimate using a particle filter.

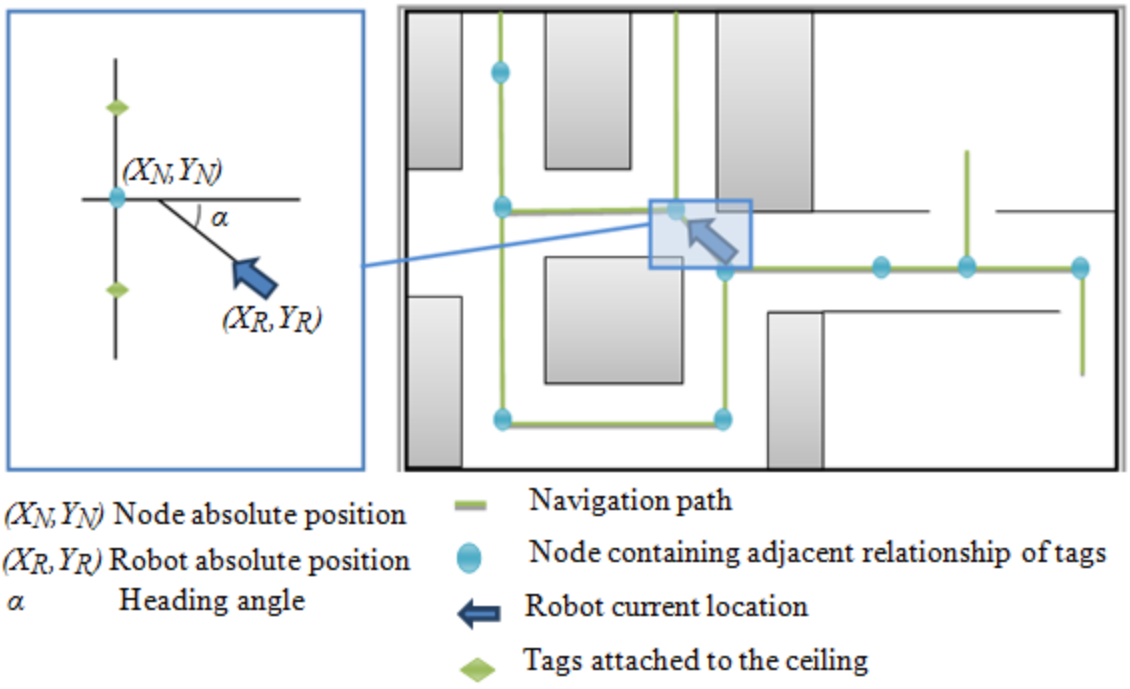


Figure 2.9

Map based navigation scheme. Recreated from Figures 1 and 3 of Weiguo et al. [79].

2.3.3.2 Obstacle recognition

S. Jia *et al.* [81] built a mobile cart, with RFID reader and a Bumblebee stereo vision camera. Passive tags are placed on the objects and the path. Obstacle detection and avoidance is handled by the RFID module. The presence of an identified object is further validated probabilistically using Bayes Rule. The obstacle/object direction with respect to the robot trajectory and the probability of the map in which that tag may exist is updated. The center position of the maximum probability is considered as the position of the tagged object. Experimental results achieved coarse object localization of $0.26\ m^2$ with RFID alone. The stereovision cameras are used to fine tune localization results. The camera platform recognizes the tagged objects as landmarks and gets pose estimates with the help of object (obstacle) tag information such as size,

color and shape. Once the pose of the tagged object is identified, the robot determines the avoidance route. As continuation of the ongoing work [82], the localization of obstacles was further enhanced by using three RFID antennas instead of one. The research has been extended by the same group [83], [84] for human recognition in which they have used the same infrastructure along with multi RFID antennas. Table 2.6 shows some of the results from experiments with different settings.

Table 2.6

Obstacle localization with different antenna settings. Data extracted from Songmin et al. [82].

Antennas configuration	*Actual position of obstacle (cm)*	*Simulation results (cm)*
Using *1* antenna	*(220, 400)*	*(268, 320)*
	(280, 400)	*(296, 276)*
	(360, 400)	*(316, 296)*
Using *3* antennas in parallel	*(240, 400)*	*(284, 304)*
	(300, 400)	*(300, 280)*
	(340, 400)	*(320, 284)*
Using *3* antennas with 45^{o} setting	*(160, 400)*	*(180, 400)*
	(320, 400)	*(324, 384)*
	(360, 400)	*(365, 404)*

2.3.4 Miscellaneous

Tracking in the domain of augmented and virtual reality has been researched for some time. The tracking devices are inbuilt components of virtual reality systems. Tracking with the camera in virtual reality incorporates sensor based or vision based techniques. Gear such as head mounts are used for tracking under prepared calibrated environments. Such tracking gear often uses active sensor based solutions, including electromagnetic, acoustic, optical, radio, mechanical and

inertial systems. However, due to certain issues such as wired power, jittering, and computational complexity, they sometimes do not provide viable solutions. In vision-based tracking, the camera orientation and location are tracked using pair wise fundamental matrices, fiducial markers and feature points. However, limitations such as high data rate, computational complexity and feature extraction make use of passive devices very complex. Thus, implementation of certain processes such as color keying or chroma key compositing become very challenging. Color keying generally involves segmenting objects from background by using color cues that is challenging for vision only modules. A common example is the meteorologist presenting weather updates with background weather clips.

The utility of RFID is being considered by some researchers in virtual environments to supplement some of the application specific vision limitations. Po *et al.* [85] proposed an RFID passive tag infrastructure for the 2D camera tracking problem in a virtual studio environment. Passive RFID tags were distributed randomly over the virtual studio area. An algorithm reads each RFID tag and then calculates orientation and velocity of the camera. The scheme reduces camera position estimation error by comparing actual position of the camera and estimated camera position info using RFID. Therefore estimation error is directly related to the distances between RFID tags. As a validation platform, simulations have been performed in Maya 3D view to generate an avatar. For the experiments, avatar position and camera position are known. By analyzing the experimental images, the authors suggest that while distributing RFID tags, a distance between tags over *12 cm* is not suitable for a virtual studio environment. Also, small tag distance of *5 cm* to *10 cm* is difficult for identification by the human eye. Moreover, use of a triangular tag distribution pattern reduces camera position error.

The amount of visual information available has been rapidly increasing across various fields, especially in video surveillance applications. There has been research in this area so that video information can be accessed more efficiently by retrieving important segments or highlights from lengthy video. This demands summarization of a large amount of visual data. Generating efficient video digests requires detecting interesting video portions, merging them into a digest and ignoring mundane parts. This research area can produce dependable outcomes by using fusion techniques. In [86] the digest generation method for Kindergarten surveillance uses a non-parametric approach to structure location information from the RFID channel and visual features from a video feed. (Parents are able to view what their children are doing during the day.) Essential video chunks are kept while discarding other portions. The technique computes pose estimation and object detection using background subtraction and motion information using inter frame differencing. Video is divided into different segments forming clusters. As the cluster members are continuous temporally, each cluster is coarsely treated as a single event. The experimental setup consists of two cameras and RFID system with active tags placed in the pockets of students. *63 hrs* of video with a resolution of 320×240 resulted in a digest of *20 mins* with a processing time of *2 hrs*.

2.3.5 Natural/outdoor site management

Resource and personnel tracking is a critical requirement in settings such as construction areas, hospitals and airports. This task is difficult due to a large number of sporadic interactions of objects and persons. Occlusions make tracking problems challenging. Radio frequency based tracking technologies have emerged as promising solutions in the market, including GPS, RFID,

Bluetooth and Wireless fidelity (Wi-Fi, Ultra-Wideband, etc). Several outdoor tracking and management approaches have been reported that use RFID and computer vision as separate entities. RFID alone is mainly used on sites for asset management [87], while CV has been used for personnel tracking [88]. The fusion of these technologies on sites and in similar applications such as airports [tagged boarding passes] and hospitals [tagged patient bracelets] should improve safety and security. The effectiveness of RFID on construction sites is examined in [89]. The authors checked performance of different tags in the lab and on site. The important findings of the paper show that passive tags do not perform well at long distances though they can be used for tracking tool loss and theft. On the other hand, active tags had *100%* read accuracy at any tag orientation with distances of *25 ft* or less. These results prove applicability of the fusion approach in an outdoor environment such as a construction site. Airport security and construction site safety and management require understanding the interaction of men, machines, materials, and terrain - other applications have similar requirements. For safety purposes the trajectory for desired objects on site should be updated in real time.

We have presented a method to examine the effect of partial object information, via RFID or special visual features, on the performance of object tracking, while solving the trajectory point correspondence problem in 3D space [90], [91]. In the initial work, the RFID feed is simulated and is used for reliable identification and locating target objects at coarse spatial resolution. Vision is used to provide finer spatial resolution for identified tagged objects. We extend geometry-based tracking so that intermittent information on object ID with location can be used in computing the overall quality of a set of paths of N objects over T time steps. We show that

partial object information can both reduce computation time and increase the likelihood of producing correct trajectories.

Location sensing based on GPS and local beacons is currently used in the outdoor environment for precision agriculture in farm management [24]. Highly precise results of one inch year-to-year and pass-to-pass accuracy also depend on real time kinematics. Such a real time location system relates to our discussion of natural outdoor environments. The ability to manage a large farm and register points of land to all kinds of maps and aerial images is analogous to managing and tracking a work site via real time RFID and CV tracking while recording state information in a database that supports other dynamic analysis.

The GRASP Lab from University of Pennsylvania and their colleagues at MIT reported research work on cooperative manipulation and transportation with multiple flying quad-rotors [92]. The quad-rotors perform a number of maneuvers with less than three inches of clearance on all sides. The angular velocity of the quad-rotors is measured with onboard Inertial Measurement Units (IMUs). For the dynamics and control they have used Vicon [93] motion capturing technology. Each quad-rotor is affixed with four passive optical markers that are tracked by multiple infrared cameras, which in turn gives the 3D position in a track volume of $5 \times 5 \times 5$ *m*. The group aims to have computer driven UAV flights that can be used for search and rescue in emergency situations such as earthquakes and fires. The published results have been reported in a lab environment and as future work they plan to validate these in natural outdoor settings.

CHAPTER 3

Proposed solution and research methodology

The introduction of fusion in the previous chapters showed that the problems of localization and tracking can be solved. Radio Frequency Identification (RFID) can be used to reliably identify target objects and can even locate targets at coarse spatial resolution, while CV provides fuzzy features for target ID at finer resolution. Our parameterization focuses on the site safety environment. We assume the agents are mostly cooperative and tagged, wearing distinctive clothing, and that 3D survey data of the test site exists. Fusion provides a method to simplify the correspondence problem in 3D space. A Site Safety System (S-3) can query for unique object ID as well as tag ID information, such as target height, texture, shape and color, which can greatly enhance scene analysis. We extend geometry-based tracking so that intermittent information on ID and location can be used in determining a set of trajectories of N targets over T time steps. Our model provides a design for stages of future improvements. The first section of this chapter formulates the problem and discusses the necessary steps. Next, an introduction of the sensor infrastructure used is provided. Finally, the goals and possible research issues are explained.

The research problem is to detect, identify, locate and generate real time tracks of *N* objects moving within a known 3D workspace within a global view. Observations from

diverse sensors are combined into object locations, and possible IDs, at discrete time steps, which must be aggregated into *N* trajectories. Motion analysis will be triggered by daemons that monitor conditions in the data – e.g. nearness of certain objects.

We abstract the information structures in order to support a system with diverse information sources and constraints and processes that may not have knowledge of each other. Without loss of generality, we sometimes ground our discussion using the site safety application. In order to study the global tracking problem and to provide a solution that is independent of a specific application, we abstract the problem as follows.

3.1 Tokens code observations from images and RFID

Consider a database that is to be built from observations from RFID readers and/or a sensor network infrastructure together with networked stereo vision sensors. Sensor observations and combination yield tokens $\tau = <x, y, z, t, v, L>$, each recording that an object with ID (name) L and feature vector v is at location (x,y,z) at time t. Some tokens will have incomplete or partial information: for example, ID L may be absent from CV observations and visual features may be absent from RFID observations. 3D coordinates may be absent for an observation from a single camera image or single RFID reader. Two or more of these tokens can be combined in the processing to get refined 3D coordinates. To keep the model simple at this point, we treat measurement accuracy and confidence values in a general heuristic manner and not as a component of a token. Higher level motion analysis will use this data and be triggered by

daemons that monitor conditions in the data – e.g. nearness of objects of class *C1* and class *C2*. Higher level activity analysis is thus based on the real-time object track data.

3.2 Object tracks {<x, y, z, t, v, L>}

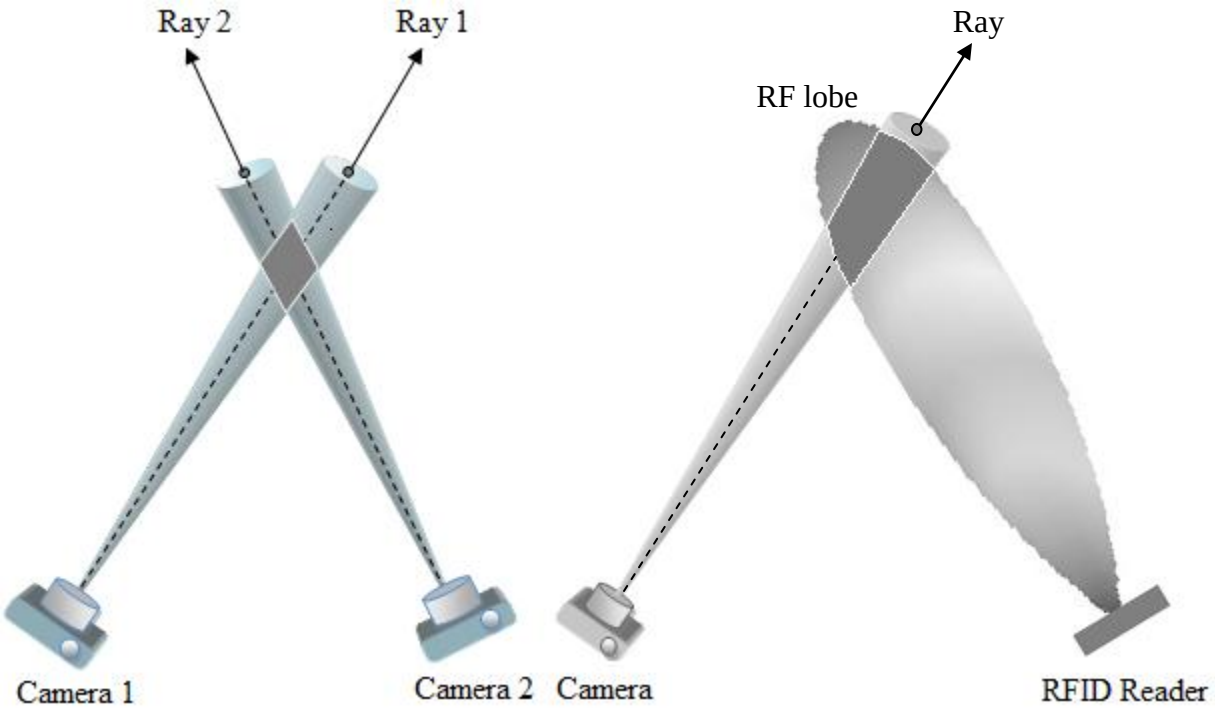
The Site Safety System (S-3) needs to identify and locate all significant objects in the workspace within a few frames *k* of real time observation. S-3 may know $L = f(<x,y,z,t>)$ from sensor subsystems that use RFID or visual features. When such information is unavailable, the system can use "tracking" to determine $L = f(<x,y,z,v,t>)$ using prior records $\{<x, y, z, t\text{-}k>\}$, or perhaps even forward records $\{<x, y, z, t+k>\}$. If object ID *L* is known, other object features $w = f(L)$ may be available from an RFID tag, such as object mass, or even a CAD model. Finally, we note that if sensors supply object speed or acceleration we consider these as components of *v* along with color, texture, elongation, etc. of its image.

An object track is *k* or more tokens in time sequence with consistent object ID and features that also satisfy constraints for motion in space. Tracking is an important concern of this dissertation, and is a low level of motion understanding that uses naïve physics to aggregate observations over time. Heuristics from naïve physics enable aggregation of individual tokens into a sequence or track, one for each moving object. As objects move through the workspace they may be occluded at any instant from either cameras or RFID readers so there may not be multiple tokens to fuse. Smoothness constraints, or motion applied over multiple time steps can be used to interpolate.

As we will see, it is not possible to assign unique object IDs to every token at every time instance. Consider, for example, the popular shell game where a bean is placed under one of three shells that look alike [94]. When the shells are shuffled quickly in space, most people cannot track the shell containing the bean. If the shells are of distinct colors, then the problem of picking the final shell is easy. If the shells are identical in appearance, but the bean is an RFID tag, RFID readers are unlikely to be able to distinguish the tagged shell in space when the shells are close to each other. Consider three workers with hard hats each with a tag and close together; if the hats are the same color, Real Time Location System (RTLS) cannot distinguish them due to read accuracy; if we know which colors contain which tags and the hats are of different colors, the system can solve the matching problem and locate each hat within the CV distance error. In order to model ambiguity, we will have to allow multiple labels L in the tokens of an object track: these labels record the ambiguity of ID at this point in time and space.

3.3 Obtaining 3D object location (x, y, z)

One fundamental sensing concept is that a sensor observes an object along a ray in the 3D space and all sensors are calibrated to the same 3D workspace. If we model sensor error, the object lies in a cone formed by projecting the error at the [2D] sensor into 3D as shown in Figure 3.1(a). Locating an object in 3D space is done by intersecting two (or more) rays [or error cones/lobes]. See [31] Ch 13. This can possibly be done by using two cameras as in the standard stereo solution or a structured light solution, or two RFID readers, or one RFID reader and one camera. Figure 3.1(b) shows the error volume in gray where the RFID reader directional antenna lobe intersects the camera error cone.



(a) (b)

Figure 3.1

Sensor error volumes: (a) rays intersection with error cones (b) intersection of error cone with RF lobe.

The sensor fusion algorithm we use computes the shortest line segment connecting two rays [95]. Given a site survey, it is simple to intersect a ray with the ground surface or with a lofted ground surface if we know the object height.

The underlying geometry is angle-side-angle, where the side is the known 3D baseline between the two sensors. A second fundamental sensing concept is where the sensor observes an object at some distance *d*. If the object transmission is observed by three such sensors it can be

located by trilateration, intersection of three spheres with radii equal to the sensed distances. An object can also be located by intersection of the ray/cone determined by an image observation and the spherical shell determined by distance *d* sensed by a single RFID reader. The commercial RTLS system encapsulates multiple RFID readers and yields a token with unique object ID and *(x, y)* coordinates on the ground plane of the workspace. The principle is similar if there is an encapsulated stereo vision system. Finally, fusion by ray intersection can alleviate the stereo correspondence problem since ID and features may be available from RFID-tagged objects.

3.4 Heuristics from naïve physics

Tracking is a lower level of motion understanding that uses naïve physics to aggregate observations of *N* objects moving over *T* time frames. Heuristics from naïve physics enable aggregation of individual observations into a sequence or track, one for each moving object. Naïve physics constraints are used to filter out unlikely labels for objects at time *t* based on the recent history of objects continuing from the *k* previous time steps. Our goal is to create a smart tracking algorithm based on the heuristics above, which will provide the means for safer activities and more efficient site management. Examples are as follows:

a. An object *n* must be at one and only one place at time *t*.
b. Location <*x,y,z*> can accommodate at most one object at time *t*.
c. Object *n* is likely to have consistent form and visual features.
d. Observations of object *n* must be consistent with its identity, if known.
e. The motion of object *n* is likely to have smooth direction.

f. The motion of object *n* is likely to have smooth velocity [makes problem more complex].

g. Constraints *e* and *f* are likely to be violated only when object *n* is in close proximity to another object *m*.

h. Known objects are likely to move in a known terrain in predictable ways.

j. Some objects are known at some locations and time instants.

k. Objects do not enter or exit the workspace [our assumption].

l. Noise may add in input trajectory points during stereo calculation.

These constraints are an extension of those used by Sethi-Jain [96] and Veenman *et al*. [97] and, unfortunately, none are *hard constraints*. For example, it may be that constraint *(b)* is violated as one object "consumes" another. Perhaps a machine consumes a worker - which S-3 should prevent!. Perhaps a driver enters a vehicle - should S-3 prevent this?

3.5 Fusion platform

We define fusion as the combination of different sensor tokens to obtain tokens containing information from the different sensors or with new information computed from the tokens from the different sensors. Most importantly, RFID and CV tokens will be fused to combine object ID with object features and to provide or to refine object location.

3.5.1 Labeling with relational constraints

To manage the complexity of the diverse information being fused and to provide a flexible experimental platform, we propose discrete relaxation to create the tracks of the *N* objects and to

update the time tokens comprising each track. Using relaxation, different sensors and sources of information can be turned on or off for experimentation or for practical reasons at a site. Fusion by relaxation is sketched as follows. Fusion processes operate on a blackboard containing the set of tokens.

/ discrete relaxation labeling for objects 1....N moving over time */*

a. Calibrate sensors to 3D site.
b. Initialize representation for N objects x T time steps x N labels.
c. For all time steps $t = 0.....T.$
    1. Run sensor processes to create [partial] tokens for detections.
    2. Run combination processes to merge and complete tokens.
    3. Run processes to eliminate impossible labels for object k at time t.
    4. Run tracking process to apply constraints and remove unlikely labels.
    5. Daemon processes possibly invoke higher level analyses processes.
d. Output object tracks as $N x T$ label matrix.
e. Store object tracks for further analysis.

3.6 Sensor arrangement

The sensor setup includes networked static cameras for visual coverage of the site. We have also considered installing cameras on the moving targets which in turn will be helpful for looming detection. The details of preliminary looming experiments are discussed in Appendix A.

The RFID readers are placed in known locations and most of the objects and personnel in the workspace are considered to be tagged. A GPS feed can also be used for validity. Figure 3.2 shows the Site Safety System (S-3) block diagram. Sections below explain the basic sensors infrastructure in detail.

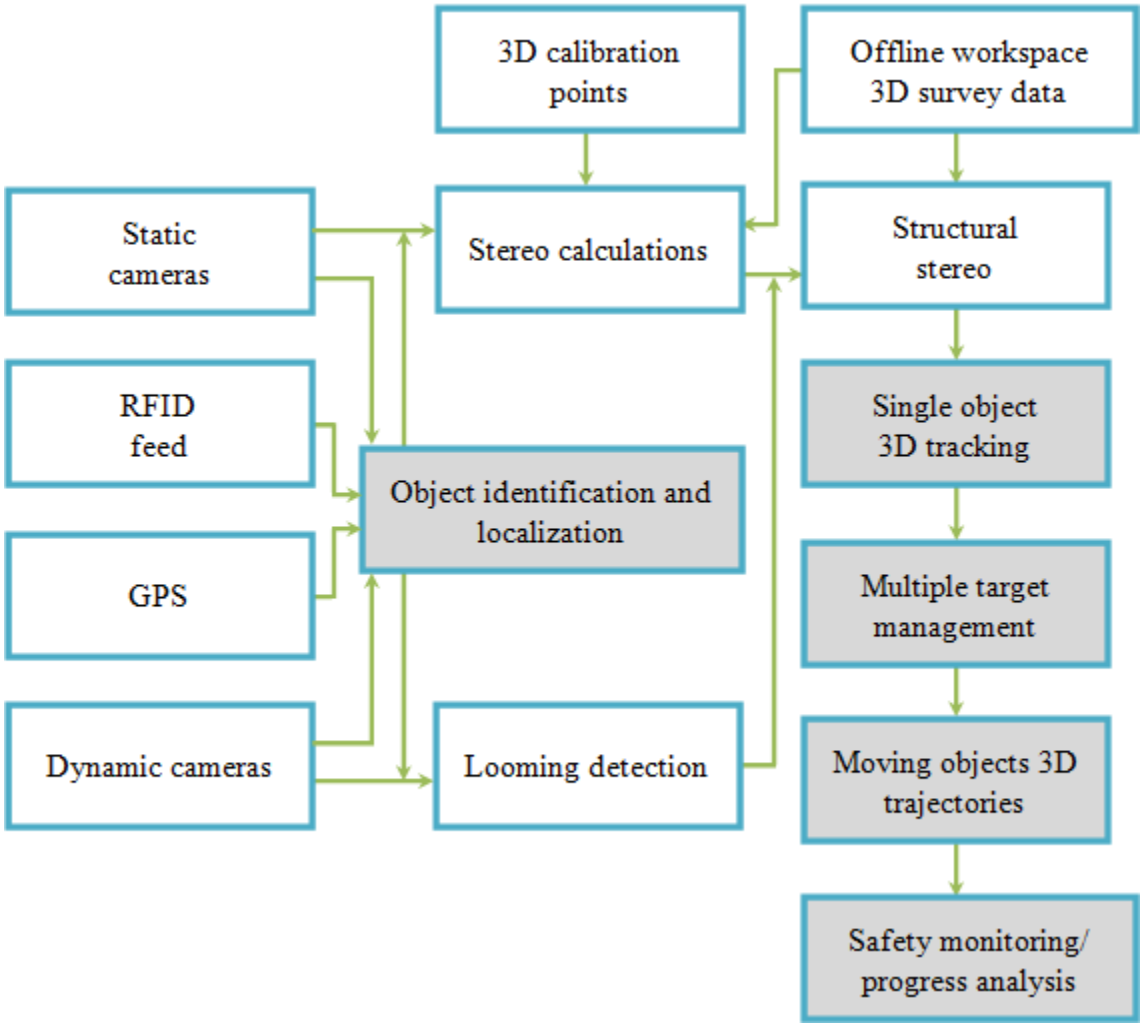


Figure 3.2

Block diagram of the Site Safety System using fusion of RFID and CV.

3.6.1 Vision infrastructure

Our system proposes the use of static cameras. As compared to costly cameras, these can be commodity good resolution cameras commercially available. The static cameras will be positioned in stereo pairs on fixed places to provide a global three dimensional field of view (FOV) of the work space. For looming detection moving objects and personnel are proposed to be equipped with wireless cameras for local operation. The site area is covered using a network of static cameras. The distance of the cameras from the tracking area can be governed by factors such as camera focal length, frame resolution and moving target desired size in images. Using low cost fixed focus equipment each camera can be positioned up to an approximate distance of about *100 ft (30.53 m)* from the site. In a general sense with a frame resolution of 800×600 pixels the minimal desired target size is approximately 50×30 pixels. This target size provides sufficient information (such as distinctive clothing eg. colorful hats, green and orange safety jackets etc.) for real-time video tracking. The scene extracted from [98] in Figure 3.3 shows the bounding boxes on some of the construction workers to give an idea of the minimal aspect ratio required of the moving objects with respect to a 800×600 pixels field of view. The image in Figure 3.3 extracted from [98] also explains the construction scenario where the proposed fusion technique will be well suited. The static cameras can be used to process stereo tracking of the moving objects whose presence in the view is also validated by the RFID feed. The preliminary experiments done on stereo tracking are explained in Chapter 6. Looming object detection can be sensed using the local dynamic cameras with motion detection techniques such as optical flow, background subtraction and frame differencing. We have performed some preliminary tests on looming detection to study its feasibility, see Appendix A. The 3D survey data of the site is

given as input for processing structural stereo. The overall system will monitor safety of multiple tracked targets and will generate a proximity warning about a possible collision threat to the tracked object; for example a worker-on-foot or vehicle backing up etc. It is known that processing is more complicated when the camera is on a moving platform where platform motion will cause optical flow even in the background. Though, this can be cancelled out as the object motion, direction and velocity information can be accessed from the trajectories with time stamping information, however, it is computationally expensive in real-time.

Figure 3.3

Construction scene example [98] with aspect ratio of persons and the field of view.

3.6.2 RFID infrastructure

Although GPS is widely used today for personal and commercial outdoor applications in open areas, it does not perform satisfactorily in indoor areas. Also it is not cost effective to equip every moving target with a GPS device. Since RFID works on wireless protocols, identification of the tagged object to be tracked can be conveyed to the visual feed for validation. The RFID location information, however coarse in nature, can supplement the visual info. RFID localization can be achieved using schemes such as lateration with distance estimation. The scene analysis can be enhanced with the deployment of extra reference tags, however, with RFID alone; target location in real time will not be as accurate as when using cameras. Refer to Section 2.1.5 for highlights on RFID positional accuracy. Each RFID localization approach and equipment has its own strengths and weaknesses.

Most targets to be tracked are considered to be equipped with an active tag, if possible. Keeping in view the outdoor dynamics the optimum locations of the readers can be analyzed by performing different trials. Also using readers with different read ranges will provide interesting results [99]. By properly placing the readers in known locations, the whole region can be divided into number of sub-regions called cells, where each sub-region can be uniquely identified by the subset of readers that cover that cell. Given an RFID tag, based on the subset of readers that can detect it, the system should be able to associate that tag with a known sub-region. The accuracy of this approach depends upon the number of readers, the placement of these readers, and the range and power level of each reader. In order to increase accuracy without placing more readers, the system might use extra fixed location reference tags to help location calibration.

These reference tags serve as reference points in the system (like landmarks in our daily life). This approach shown in Figure 3.4 helps offset many environmental factors that contribute to the variations in detected range because the reference tags are subject to the same effect in the environment as the tags to be located.

In our experiments we have used and evaluated the commercially available RFID-based Real Time Location System (RTLS) from CSL (Convergence Systems Ltd.) [8] for its accuracy and reliability of object detection and location. The RFID based RTLS development kit used has six readers (one master and five slaves). They can be used in different settings to form a cell. The system uses time-of-arrival (TOA) concept where the distance between the tag and the readers is calculated by the roundtrip time.

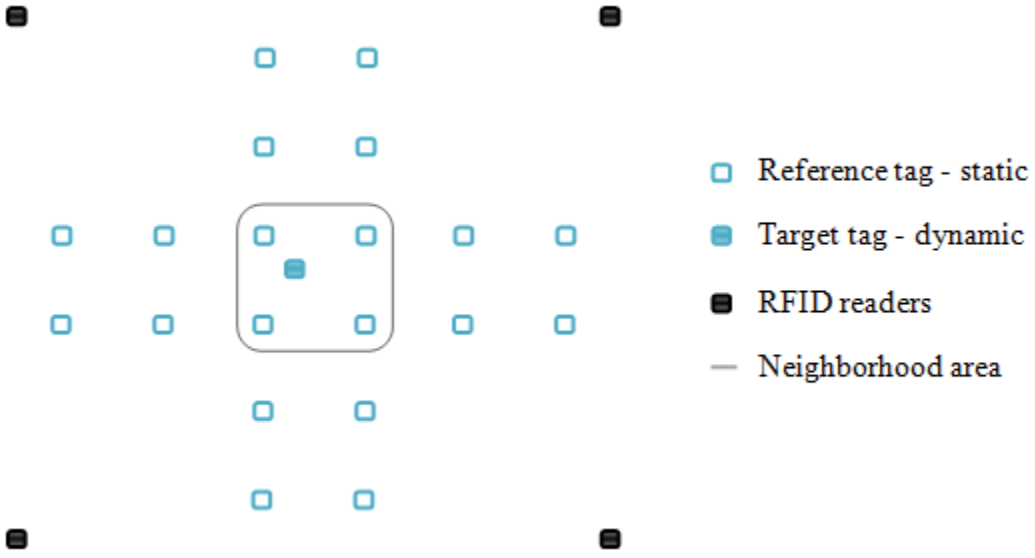


Figure 3.4

Schematic of dense placement of reference RFID tags.

The tags communicate with the readers using time-division-multiplexing (TDMA). Each reader has a beam width of 80^o in portrait and 30^o in landscape orientation. To perform location sensing each reader has to be pointed towards the center of the test site. To cover a larger area more readers can be installed thereby generating more cells to enhance the location accuracy in difficult configurations. The tags used are *2.4 GHz* active tags with up to *200 m* of read range in an open outdoor space. The tags run on $3 \times AAA$ batteries and are approximately of the size of an iphone 4S and weigh less than the phone with batteries installed. Figure 3.5 shows the readers and the tags used in the RFID RTLS system.

(a) (b)

Figure 3.5

CSL RFID based Real Time Location System: (a) active tag (b) master reader.

3.7 Goals and related research parameters

Using fused information in the multi-object tracking scenario, the approach envisions evolving according to the general steps and related research parameters discussed below.

3.7.1 Calibrating the cameras

The static cameras are to be attached at fixed positions in the work space so as to provide the global three dimensional Field of View (FOV). They will be used to provide stereo tracking of the moving objects. The cameras are calibrated using the affine calibration method. A 3D global workspace coordinate system *[X,Y,Z]* is created. Place some fixed visual markers or identify structural landmarks in the scene to provide calibration points. The system can have apriori survey information about the 3D world coordinates of these points. Synchronize these cameras in time and space. Finally a transformation matrix for each fixed camera is obtained for stereo calculation. Details of the calibration process is provided in Chapter 4 and Appendix B.

3.7.2 Defining the ground truth

Ground truth data is required to evaluate the system performance. To define ground truth measurements a mesh can be created which represents the surface of the ground upon which work will be done. Modern surveying instruments can be used for this task. In a lab environment this can be done by projecting a grid through a projector and later recording the surface detail. The ground can contain some fixed visual markers so that cameras can monitor and validate their

movement due to any undesirable factors, such as wind or vibration. Also the mobile objects are known to move on the ground, which provides constraints on their location. This helps formulate ground truth data by moving objects on predefined paths.

3.7.3 Structural stereo approach

Solving correspondence and camera calibration in stereo are key issues. This requires autonomous relative camera orientation and stereo matching that uses the relationship between the correspondence problem and the camera pose estimation problem. Each camera needs to be calibrated to the workspace to obtain its camera matrix. Each camera should be able to see some moving objects *A, B, C* and *a, b, c*, etc. Object matching can be done using color blob detection and/or object ID. The color detection performance can also be analyzed in different lighting conditions. In case of visual occlusion some objects might be identified by RFID alone which can help in the matching process. For each object region in the image of camera *1*, compatible matching regions are to be found in the image of camera *2*. The camera matrices obtained from the calibration process will be used to compute 3D location *[x,y,z]* for each possible object match. The consistency of the object can be checked with the ground plane. One to one mapping should be created for each object in camera *1* and camera *2*. IDs provided by the RFID system can be used to make mapping more efficient. Stereo using a single camera and a single RFID reader can also be examined. To analyze and evaluate the accuracy of the stereo system, it needs to be established how accurate can this matching be using visual and/or RFID information and how accurate motion trajectories are obtained after comparison with the ground truth.

3.7.4 Object sensing using multiple RFID readers

For the tagged objects, detection and spatial accuracy is to be analyzed in RFID sensing mode. The object detection and read accuracy will vary in different settings. Similarly the object location accuracy will change if the object is directly visible to more readers or otherwise or the readers configuration gets changed.

3.7.5 Integration of multiple CV sensors

To cover the view of the workspace from all directions multiple cameras can be networked together. However, selecting the right cameras and time slices is difficult. The ID of each object by RFID can help choose the right camera for further calculations.

3.7.6 Fusion of RFID and cameras

The fusion algorithm is the back bone of the S-3 system. Also the RFID and vision data being diverse in nature needs to be transformed in an appropriate format so that it can be compared and fused into refined tokens. All the sensors are to be calibrated to the same coordinate system. Synchronization within cameras or between cameras and RFID will be a complicated task. The number of RFID readers required to fully cover the desired area and their distribution needs to be worked out. Provision of any additional information through the database that can be used by daemon processes will help reduce runtime complexity.

3.7.7 Smoothness of trajectories

The dynamic scene can contain multiple independently moving objects. The objects are permanent; except for new objects entering or old objects leaving the workspace. Track initiation and termination are to be carefully examined as it might be that a tracked object reappeared after occlusion for a short time or a new object has entered the workspace. Smoothness is a global operation which can help high level processes to define and/or correct computed object tracks. Smoothness of trajectories requires a burst of time frames for reliable results. Defining appropriate naive physics constraints will help identify correct objects. It is safe to assume here that typically objects move along the surveyed ground plane.

3.7.8 Looming detection

Motion and looming detection will apply toward the cameras placed on the head gear and moving vehicles. An optical flow algorithm can be used to detect the motion and looming phenomenon. Lucas-Kanade and Horn-Schunck are widely used methods for optical flow estimation. If the object is not fully available in the Field of View (FOV) then RFID input can help. Motion vectors can be extracted (possibly 3D) for one or more moving objects in the FOV. Looming object may or may not be a “smart object”. It can be analyzed how accurately the distance from smart object to sensing object (camera platform) can be computed. Depth measurement accuracy and motion measured parallel to the image plane can also be studied.

3.7.9 Object Inventory

The work safety system monitors 6-tuples in real-time and outputs safety controls. Other applications can analyze these attributes offline for work efficiency, materials inventory and person tracking, etc.

CHAPTER 4

Localization of objects and scene points

Localization involves computing object location with respect to some external frame using sensory data. Pose is an umbrella term that defines object location and orientation relative to the global reference frame. Visual image data has embedded information such as color, texture, and shape etc. which can be used to ascertain object pose. However, the errors due to vision system constraints and lighting conditions may propagate during computation. Data fusion from other active sensors such as RFID can help with object detection, identification and coarse location estimation. We introduce and discuss methods and configurations used for location estimation by both RFID and stereo vision and show accuracy that could be achieved by each of these modalities. In outdoor experiments stereo provided location accuracy within 7.6 in (19.3 cm) whereas for RFID was up to 1.5 m for a tagged person being stationary for few seconds at predefined points.

The ability to detect, identify and locate objects in an environment are some tasks that determine the performance of a tracking system. Object appearance, features, orientation and motion are some characteristics that are used in the recognition and tracking processes. Over the past decade, fusion of RFID and CV has also being used in indoor mobile and industrial robotics to support tasks such as autonomous recognition, localization and tracking. RFID alone

has also been researched widely in this quarter. Passive stereo vision can locate detected objects in a 3D volume provided the image of the same object can be identified in two or more cameras. An RFID reader can be used to ID an object observed in some 2D image, thus aiding stereo; or, a network of RFID readers can provide coarse 3D location without cameras. Thus RFID can help with object localization in multiple ways. RFID technology also enables smart objects to communicate information about themselves not available to optical sensors; for example object weight, container content, etc. A tagged rigid object can even help provide an optical observer with a network downloaded CAD model of itself to be used for pose computation by the observer. The focus of this chapter is to analyze object location procedures and accuracy using vision and RFID as single modalities.

4.1 Object detection and blob analysis using vision

The site safety system acquires video frames from cameras observing the work site. Simultaneous frames from two cameras can be used as a stereo pair for detecting desired objects and locating them in 3D. To locate an object, detection and identification are initial tasks that are complex processes using vision as a single modality. RFID supports vision in this step. In our approach we locate objects in all frames. Working toward an automatic system, we currently have some manual steps in the research methods. Our vision based detection stage uses color and blob analysis for real-time processing and also allows user interactivity at times for outdoor data to avoid detection failure. We have also explored Hough transform based elliptical shape features for head detection and have used it for offline processing. We have used *MATLAB® 2009a* to acquire video from the cameras. The video frames are extracted using the image acquisition

toolbox. For color based detection, the desired color is extracted from the RGB image. Figure 4.1 shows steps involved in the RGB color detection process. We have also used HSV space for detecting colors, which will be covered in the coming paragraphs. HSV space is capable of separating color components from intensity and is more robust towards lighting changes and shaded regions. The blob analysis steps in our approach are the same in RGB and HSV based color detection.

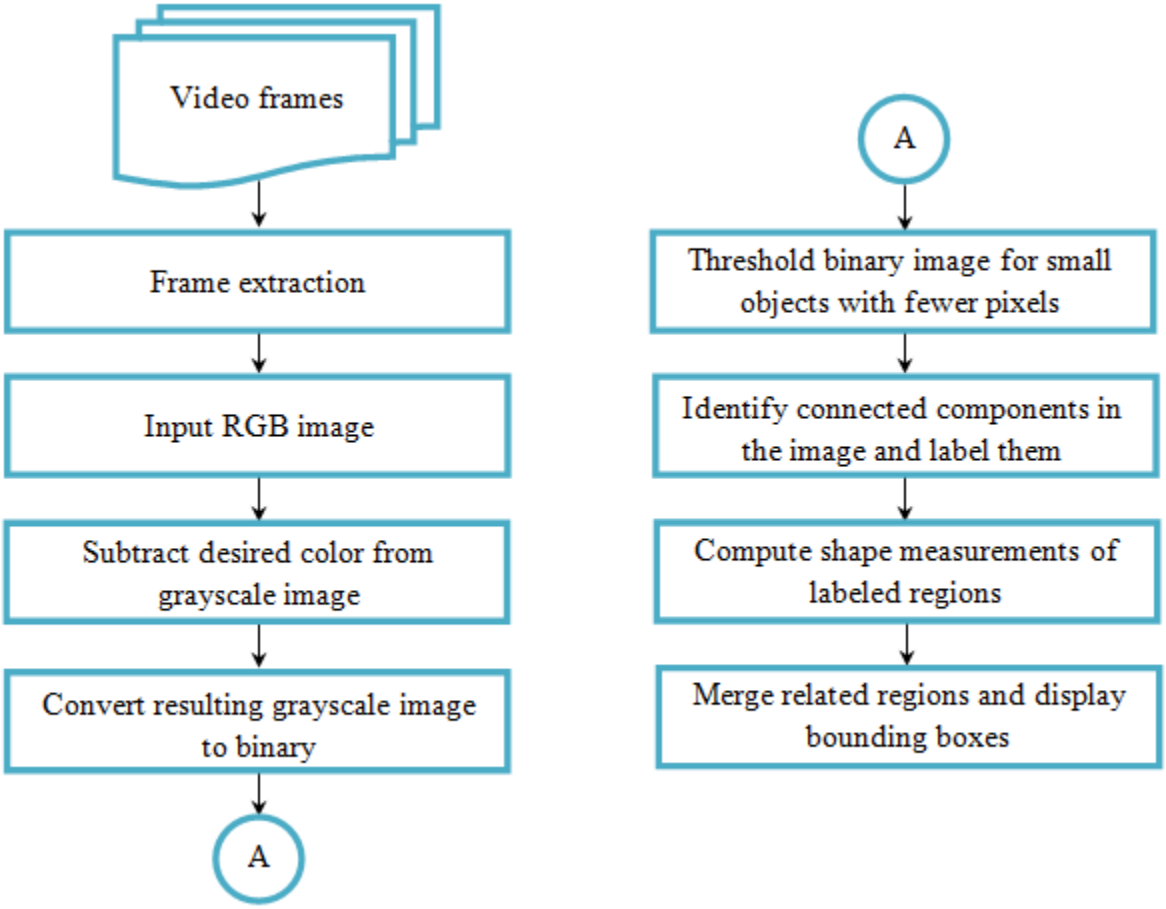
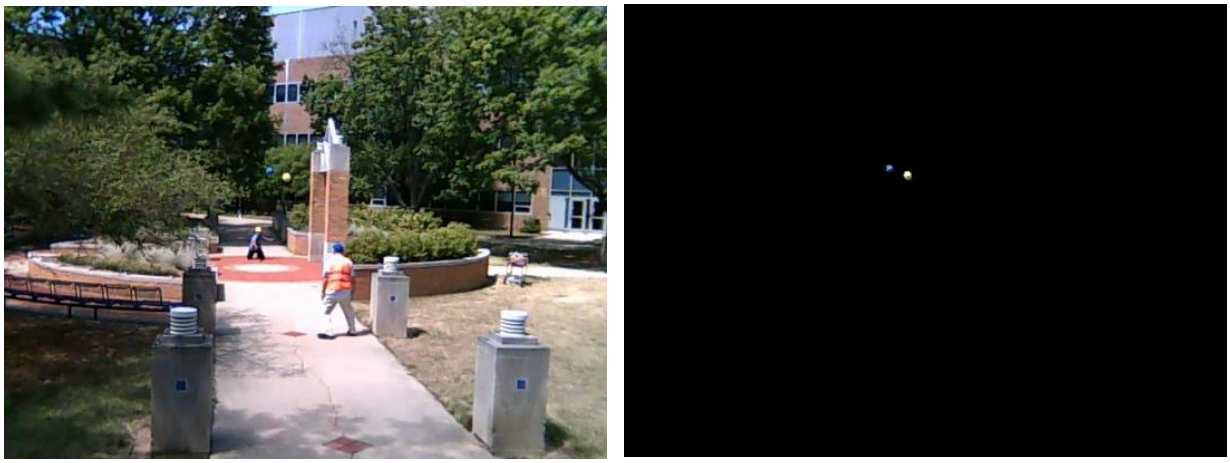


Figure 4.1

Flow diagram of specific RGB color detection and connected components blob analysis.

The input image is converted to a grayscale image. The desired color is then subtracted from the grayscale image and the resultant image is converted to binary. Through connected components analysis, the algorithm merges object pixels that are close to each other so as to create blobs. For example, pixels that represent yellow are a portion of a person's head gear and are grouped together. Next, in the blob analysis the properties of the region are extracted and bounding boxes of these blobs are calculated. Based on these blob properties the individual bounding boxes maybe merged together so that each tracked object-part (eg. safety helmet etc.) is enclosed by a single bounding box. The center of each bounding box in both images is then considered to be the object center point for performing stereo correspondence. An object will have little displacement between two consecutive frames, therefore, the center of the blob provides a strong and useful feature for locating and tracking objects. Figure 4.2 shows desired blobs detected in input image in an outdoor environment.



(a) (b)

Figure 4.2

Blob detection outdoors: (a) original image (b) blobs of blue and yellow balls.

For HSV based color detection the input RGB image is converted to the HSV space and H, S and V images are extracted individually. A histogram of these individual color bands is then calculated. Based on the color to be detected the minimum and maximum threshold is defined for hue, saturation and value image. For example, we have used the following threshold values for detecting the yellow color in our indoor experiments:

Hue threshold low = *0.11*

Hue threshold high = *0.19*

Saturation threshold low = *0.39*

Saturation threshold high = *1*

Value threshold low = *0.39*

Value threshold high = *1*

The defined threshold is then applied to each color band and individual masks are generated. The masks are then combined together to find where all of these are true for the color to be detected. With a little iteration the threshold values of H, S and V can be adjusted according to test environment. Based upon the size of the detected object the smaller objects are then filtered out. Holes are then filed in individual color band images. The desired color mask is then obtained to mask out the desired color from the RGB image.

.

4.1.1 Elliptical shape features for head detection

The human head can be approximated well with elliptical shape features. We have used Hough transform based ellipse shape detection given in [100]. This method takes advantage of the major axis of an ellipse to find ellipse parameters fast and efficiently.

For an arbitrary ellipse, there are five unknown parameters as shown in Figure 4.3. These are orientation α, center (p_0, q_0), major axis $2m$ and minor axis $2n$. Their relationship is shown in equations below. The algorithm is initiated by giving a range of major axis [user specified] which is then used to find the minor axis of the ellipse. Since only a one-dimensional accumulator array is required to accumulate the length of the minor axis this step of the transformation is very efficient.

$$p_0 = \frac{(p_1 + p_2)}{2} \tag{4.1}$$

$$q_0 = \frac{(q_1 + q_2)}{2} \tag{4.2}$$

$$m = \sqrt{\frac{(p_2 - p_1)^2 + (q_2 - q_1)^2}{2}} \tag{4.3}$$

$$\alpha = a\tan\left(\frac{q_2 - q_1}{p_2 - p_1}\right) \tag{4.4}$$

$$n = \frac{m^2 b^2 \sin^2 \beta}{m^2 - b^2 \cos^2 \beta} \tag{4.5}$$

$$\cos\beta = \frac{m^2 + b^2 - \mathrm{d}^2}{2mb} \tag{4.6}$$

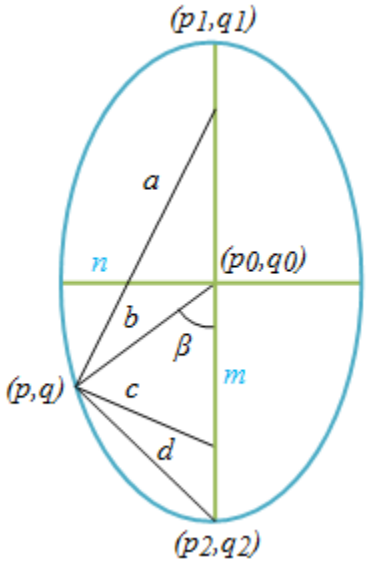


Figure 4.3

Ellipse geometry showing basic parameters to define an ellipse.

The background subtraction technique is applied on the incoming frame and the cropped image of the desired person is generated. This step helps reduce the edge pixels space required in the next steps. Edge detection is performed on the R, G and B channel of the cropped image and a union edge image is acquired. Binary image dilation using linear structuring elements is then performed to acquire boundary pixels. Each pair of edge pixels is considered as candidate for two vertices on the major axis of an ellipse. Using these two candidate pixels the four parameters are calculated. Another arbitrary point is used to find the half-length of the minor axis *n*. A voting process is then initiated to acquire the desired *n* using a one-dimensional accumulator array. Figure 4.4 shows the implementation steps.

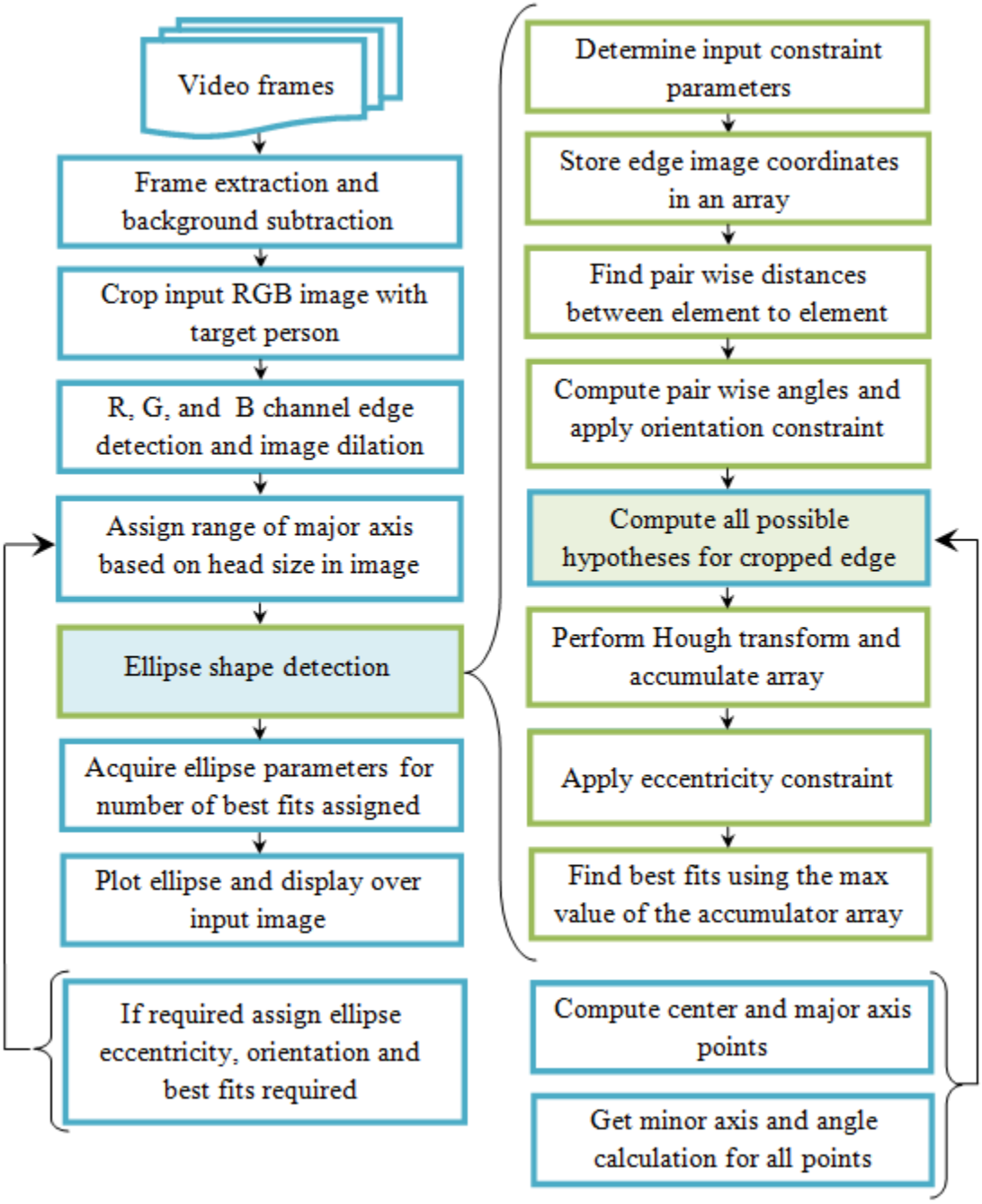
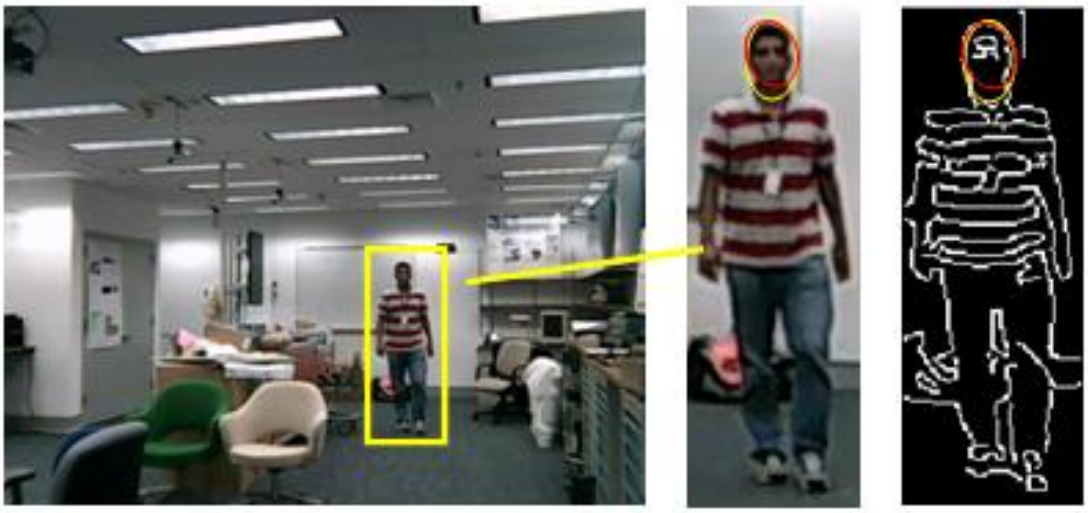


Figure 4.4

Implementation steps for elliptical shape feature detection.

Figure 4.5 shows the results of head detection using elliptical shape features. Figure 4.5(a) shows the input *640×480* image. The cropped image with best two ellipses [red and yellow color] is shown in Figure 4.5(b). Figure 4.5(c) shows the cropped edge image with the same two best fit ellipses. For this case, eccentricity ratio of *0.4* with orientation angle range between 50^o to 90^o was used.



(a) (b) (c)

Figure 4.5

Results of head detection using elliptical shape features: (a) input image 640×480 (b) cropped image with best two ellipses (c) cropped edge image with best two ellipses.

4.2 Stereo vision using ray-ray combination

Stereo vision based on the principle of ray-ray intersection uses cameras that have slightly different pose in space. Unlike various other stereo approaches, the cameras do not need to be specially configured relative to each other, however they should have an effective and common field of view. Each camera is calibrated to the 3D workspace. The 3D point is calculated by solving the correspondence problem, that entails observing the same feature point in two or more 2D images from these cameras. Some basic concepts about stereo vision and 3D reconstruction are provided in Appendix B.

Due to factors such as lens distortion, digitization noise, small camera vibrations and sub-pixel difference in correlating corresponding points, the errors generate and propagate in the stereo system. If we model camera errors and project the error at the cameras into 3D, then due to the propagating nature of these errors, rays are transformed to cones. The object in 3D lies in the overlapping volume of these cones referred to as the error volume. Apart from the factors mentioned, the error volume also varies with the selected pose of the cameras. The variation in the error volume with change in camera pose is demonstrated in Figure 4.6.

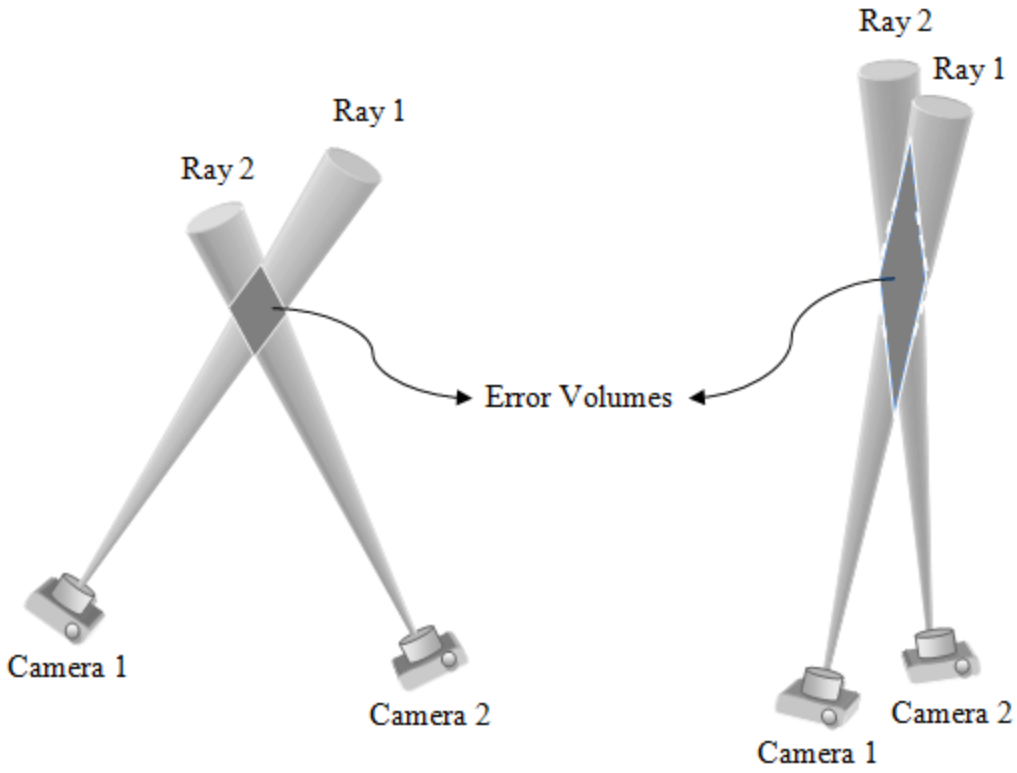


Figure 4.6

Error cones obtained from projecting 2D imaging error back into 3D.

4.2.1 Stereo configuration

In our setup, the vision system consists of two cameras separately calibrated to the 3D work space. The known 3D points in the work space are used for camera calibration. Figure 4.7 shows the perspective model having two cameras viewing the same 3D workspace. A right hand coordinate system is used. The points farther from the camera have more positive depth coordinates in the camera coordinate system. The site area is considered as the 3D world with its own global coordinate system. A 3D world point is represented as ${}^{W}P = [\, {}^{W}P^{x}, {}^{W}P^{y}, {}^{W}P^{z} \,]^{t}$. The intersection of the two imaging rays, ${}^{W}PO_1$ and ${}^{W}PO_2$ determines the location of the 3D

point ${}^{W}P$. We have adopted a general stereo approach [31], [101], [102] where the same feature point ${}^{I}P_i$ in two or more calibrated cameras is used to calculate the 3D world point ${}^{W}P$.

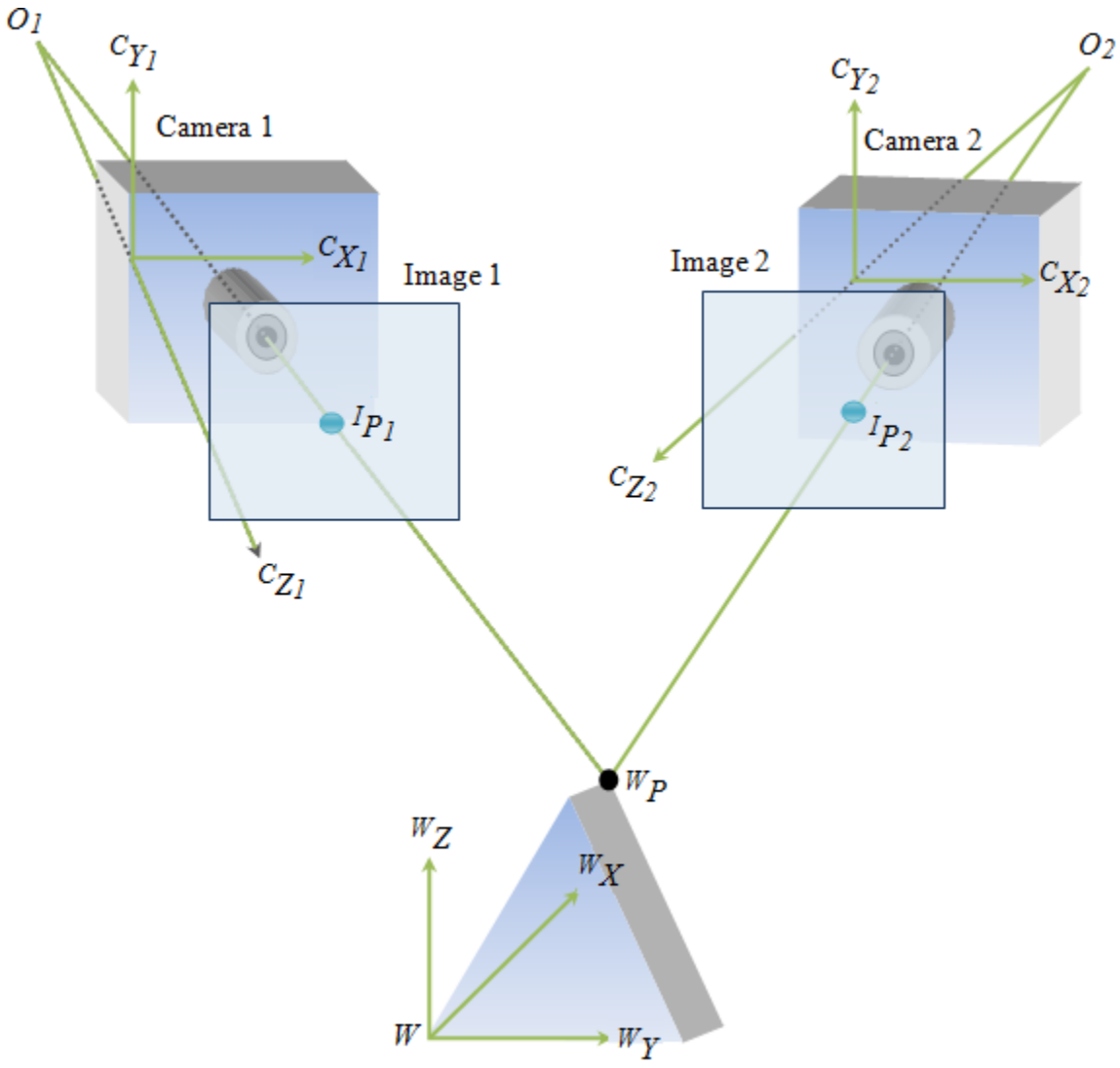


Figure 4.7

General stereo configuration with two cameras viewing a 3D object in a 3D workspace *W.*

For any required computation, the pose of camera *1* and camera *2* in the 3D world coordinate system *W* and camera intrinsic parameters such as the focal length etc. shall be known through calibration. This information is defined in the camera matrix obtained by calibration, known as the affine method; see [31] Ch 13 and [103] Ch 12. The calibration procedure does not model radial distortion and we do not rectify the images. This method provides a more general form of camera parameterization and the exterior and interior parameters are combined in the elements C_{ij} of the camera matrix. Fewer parameters means fewer required calibration points. The affine camera matrix calibration procedure used is explained in Appendix B. For stereo processing, the correspondence between a set of 2D and 3D points needs to be recognized.

4.2.2 Computing shortest line segment connecting two rays

In practice, two camera rays will not intersect in 3D space. The main cause of this can be due to the approximation errors in camera models and due to errors in image point location. Such errors can occur even due to sub-pixel inaccuracy in the image points. Once generated, this error amplifies as the ray propagates in space. To get a reasonable 3D location estimate, the approach of shortest line segment connecting two rays [31] Ch 13, [95] Ch 10 is used and is shown in Figure 4.8. The coordinate system symbols are dropped from the notation hereafter. The center of this line segment will represent the 3D point. So the smaller the segment, the better is the correspondence of image points and vice versa. We have also used this segment length criterion as a constraint to solve the correspondence problem. Epipolar constraints are also used in conjunction for robustness. Refer to Appendix B for background on epipolar constraint.

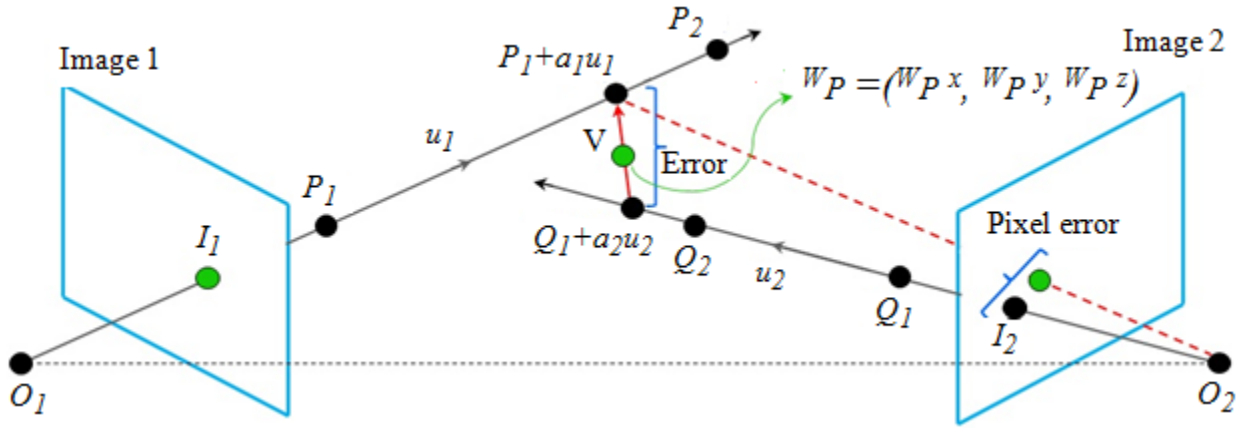


Figure 4.8

Shortest line segment connecting the two skew rays.

P_1 and P_2 are the points on the ray originating from camera optical center O_1 and passing through image point I_1 while Q_1 and Q_2 are the points on the ray originating from camera optical center O_2 passing through image point I_2. If the optical center of the cameras is not known then camera *1* ray points can be computed using the two equations in Equation A.9 while choosing an arbitrary value of ${}^W P^z = z$. If the computed ray is parallel with the *z-axis* then the same procedure can be repeated for *y* and *z* while ${}^W P^z = x$ and so on. u_1 and u_2 are the unit vectors along these rays respectively. The shortest line segment is represented by vector V and is orthogonal to both u_1 and u_2 and is given as:

$$V = (P_1 + a_1u_1) - (Q_1 + a_2u_2) \tag{4.7}$$

The variables a_1 and a_2 can be computed using the following set of linear equations. Here '$\odot$' represents dot product:

$$\begin{aligned}[(P_1 + a_1u_1) - (Q_1 + a_2u_2)] \odot u_1 = 0 \\ [(P_1 + a_1u_1) - (Q_1 + a_2u_2)] \odot u_2 = 0\end{aligned} \tag{4.8}$$

Rearranging Equations 4.8:

$$\begin{aligned}[(P_1 - Q_1) + (a_1u_1 - a_2u_2)] \odot u_1 = 0 \\ [(P_1 - Q_1) + (a_1u_1 - a_2u_2)] \odot u_2 = 0\end{aligned} \tag{4.9}$$

$$\begin{aligned}[(P_1 - Q_1)] \odot u_1 + [(a_1u_1 - a_2u_2)] \odot u_1 = 0 \\ [(P_1 - Q_1)] \odot u_2 + [(a_1u_1 - a_2u_2)] \odot u_2 = 0\end{aligned} \tag{4.10}$$

$$\begin{aligned}[(P_1 - Q_1)] \odot u_1 + [(a_1.1)] - [(a_2u_2)] \odot u_1 = 0 \\ [(P_1 - Q_1)] \odot u_2 + [(a_1u_1)] \odot u_2 - [(a_2.1)] = 0\end{aligned} \tag{4.11}$$

$$a_1 - a_2(u_1 \odot u_2) = -[(P_1 - Q_1)] \odot u_1 \tag{4.12}$$

$$-a_2 + a_1(u_1 \odot u_2) = -[(P_1 - Q_1)] \odot u_2 \tag{4.13}$$

Solving Equation 4.12 and 4.13 further to get a_1 and a_2. Multiply Equation 4.13 by $(u_1 \odot u_2)$ and subtract from Equation 4.12:

$$a_1[1 - (u_1 \odot u_2)^2] = [(Q_1 - P_1)] \odot u_1 - [(Q_1 - P_1) \odot u_2](u_1 \odot u_2) \tag{4.14}$$

$$a_1 = \frac{[(Q_1 - P_1)] \odot u_1 - [(Q_1 - P_1) \odot u_2](u_1 \odot u_2)}{[1 - (u_1 \odot u_2)^2]} \tag{4.15}$$

Multiply Equation 4.12 by $(u_1 \odot u_2)$ and subtract Equation 4.13 from Equation 4.12:

$$a_2[(u_1 \odot u_2)^2 - 1] = [(Q_1 - P_1)] \odot u_2 - [(Q_1 - P_1) \odot u_1](u_1 \odot u_2) \tag{4.16}$$

$$a_2 = \frac{[(Q_1 - P_1) \odot u_1](u_1 \odot u_2) - [(Q_1 - P_1)] \odot u_2}{[1 - (u_1 \odot u_2)^2]} \tag{4.17}$$

If the magnitude of vector *V* is less than a desired threshold then the 3D world coordinates *x, y, z* of the point ${}^{W}P$ are given as the midpoint of V:

$$W_P = \frac{1}{2}[(P_1 + a_1 u_1) + (Q_1 + a_2 u_2)] \tag{4.18}$$

4.3 3D location estimation results using stereo vision

We provide here the details and results obtained from our stereo experiments. We have used commodity cameras in an indoor and outdoor scenario. The indoor test site was surveyed using laser range finder and tape measurements. The outdoor test site was surveyed using a total station [surveying equipment], laser range finder and tape measurements. Details about site survey procedures are provided in Appendix C.

4.3.1 Computing residual error using jig

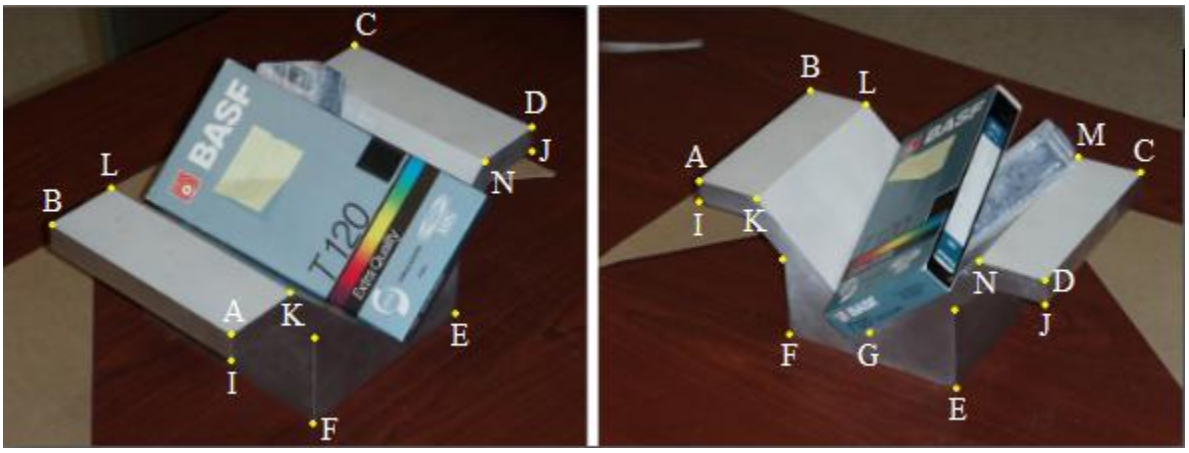
The cameras were calibrated using the *affine transformation* procedure in Appendix B. The test stereo image pairs of a *jig* were used for further analysis. The *jig* is a physical object with precise and easily recognizable feature points (yellow) as shown in Figure 4.9. The feature points/corners are then used as calibration points to compute the transformation matrix of each camera. For ground truth the 3D dimensions of the *jig* assembly are known. A typical camera matrix *C* is represented as follows:

$$C = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & 1 \end{bmatrix} \tag{4.19}$$

To do a fundamental experiment we used test images. These images contain the jig with a video tape box. We have used ten corresponding calibration points i.e $i \geq 10$ in Equation 4.20 to calculate the transformation matrix.

$$\begin{bmatrix} W_{P_i}^{x}, W_{P_i}^{y}, W_{P_i}^{z}, 1, 0, 0, 0, 0, -W_{P_i}^{x} \times I_{P_i}^{r}, -W_{P_i}^{y} \times I_{P_i}^{r}, -W_{P_i}^{z} \times I_{P_i}^{r} \\ 0, 0, 0, 0, W_{P_i}^{x}, W_{P_i}^{y}, W_{P_i}^{z}, 1, -W_{P_i}^{y} \times I_{P_i}^{c}, -W_{P_i}^{y} \times I_{P_i}^{c}, -W_{P_i}^{z} \times I_{P_i}^{c} \end{bmatrix} \begin{bmatrix} c_{11} \\ c_{12} \\ c_{13} \\ c_{14} \\ c_{21} \\ c_{22} \\ c_{23} \\ c_{24} \\ c_{31} \\ c_{32} \\ c_{33} \end{bmatrix} = \begin{bmatrix} I_{P_i}^{r} \\ I_{P_i}^{c} \end{bmatrix} \quad (4.20)$$

The calibration points *(A,B,C,D,E,F,K,L,N,P)* are shown in Figure 4.9(a) and (b). The lower bound on the number of calibration points to be used is eight.



(a) (b)

Figure 4.9

Jig images with easily recognizable feature points: (a) left image (b) right image.

The camera matrices obtained from the right and left images are shown in Table 4.1.

Table 4.1

Left and right camera transformation matrices.

Right camera	*146.1*	*114.2*	*-24.9*	*525.7*
	35.6	*-73.6*	*-168.4*	*911.3*
	-0.008	*0.01*	*-0.01*	*1*
Left camera	*183.6*	*-154.9*	*-22.7*	*1206.6*
	-89.9	*-80.9*	*-206.9*	*1721.2*
	0.01	*0.01*	*-0.01*	*1*

For quantitative analysis Table 4.2 shows the 2D residuals (absolute difference between original and computed 2D points) in pixels for the left and right image of the *jig with tape*. The

underlined readings show an error greater than *one pixel*. The *RMS error* of the left image is ${}^{L}X_{rms} = 0.9$ *pixel*, ${}^{L}Y_{rms} = 0.9$ *pixel* and of the right image is ${}^{R}X_{rms} = 0.7$ *pixel*, ${}^{R}Y_{rms} = 1$ *pixel*.

Table 4.2

2D residuals for left and right image of jig - scale is in pixels.

Points	*World 3D points*	*Left Image original 2D points*	*Left Image computed 2D points*	*Left image 2D Residuals*	*Right Image original 2D points*	*Right Image computed 2D points*	*Right image 2D Residuals*
A	*0,0,0*	*1208*	*1206.6*	*1.4*	*527*	*525.7*	*1.3*
		1721	*1721.2*	*0.2*	*910*	*911.2*	*1.2*
B	*0,6,0*	*254*	*256.3*	*2.3*	*1121*	*1120.9*	*0.1*
		1142	*1143.2*	*1.2*	*434*	*434.9*	*0.9*
C	*11,6,0*	*1859*	*1858.8*	*0.2*	*2854*	*2853.7*	*0.3*
		200	*199.7*	*0.3*	*871*	*872.3*	*1.3*
D	*11,0,0*	*2792*	*2791.5*	*0.5*	*2349*	*2350.9*	*1.9*
		635	*633.6*	*1.4*	*1437*	*1436.1*	*0.9*
E	*8.25,0,-4.5*	*2389*	*2390.2*	*1.2*	*1878*	*1878.5*	*0.5*
		1617	*1617.3*	*0.3*	*2002*	*2000.6*	*1.4*
F	*2.75,0,-4.5*	*1644*	*1643.2*	*0.8*	*1012*	*1011.6*	*0.4*
		2179	*2179.1*	*0.1*	*1717*	*1718.7*	*1.7*
K	*2,0,0*	*1528*	*1530.8*	*2.8*	*833*	*831.9*	*1.1*
		1500	*1499.1*	*0.9*	*1000*	*999.4*	*0.6*
L	*2,6,0*	*583*	*581.9*	*1.1*	*1413*	*1413.4*	*0.4*
		952	*951.9*	*0.1*	*511*	*508.8*	*2.2*
N	*9,0,0*	*2537*	*2537.8*	*0.8*	*1992*	*1991.8*	*0.2*
		810	*809.1*	*0.9*	*1333*	*1332.9*	*0.1*
P	*2.75,0,-1.81*	*1646*	*1645.9*	*0.1*	*976*	*975.3*	*0.7*
		1740	*1738.3*	*1.7*	*1323*	*1320.1*	*2.9*

Figure 4.10 shows the procedure to compute 2D points for the jig that are not used in the calibration. The 2D points are recalculated from 3D projected into 2D using the transformation matrices.

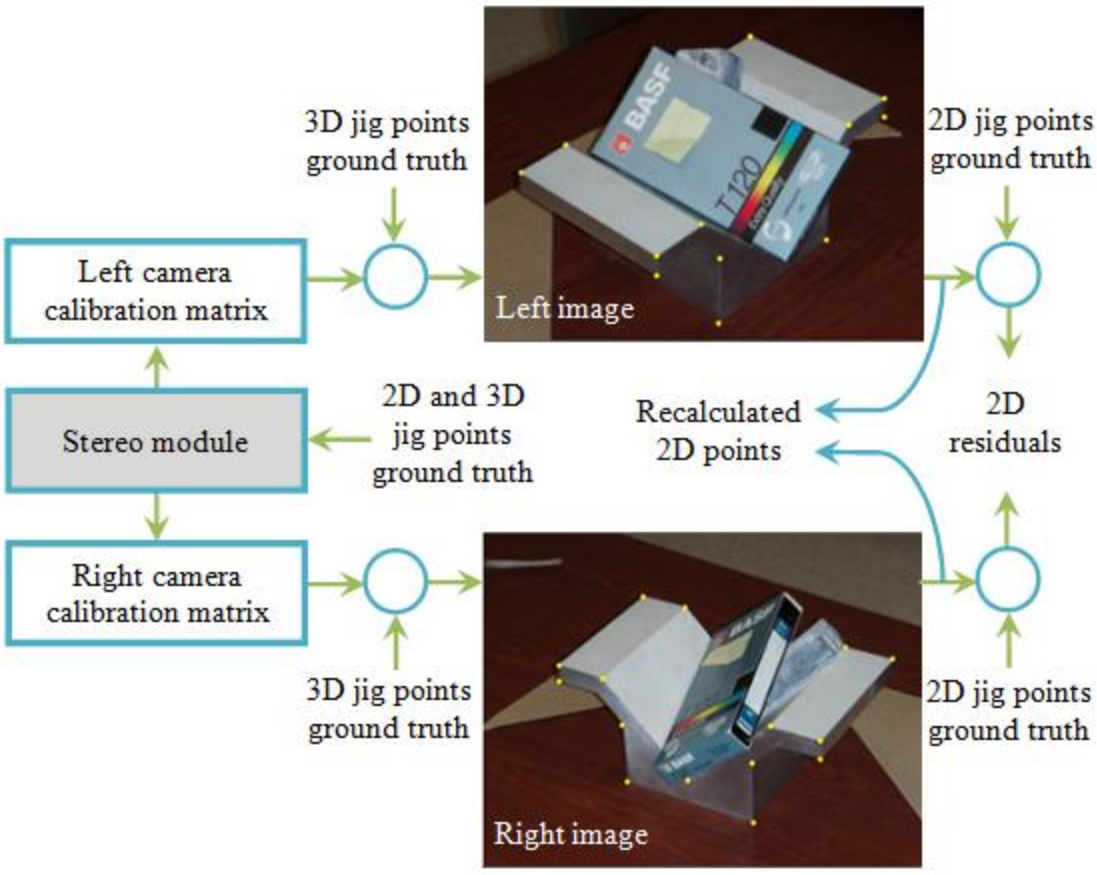


Figure 4.10

Procedure for calculating 2D residuals of jig images.

4.3.2 Components of the stereo system used

To perform the stereo experiments the platform used is a *core i5 580M* with *4GB RAM*. Two *Logitech C210* fixed focus cameras were used for the stereo trials. The focal length of these cameras is *~4 mm*. The frame rate is *15 fps* with *640×480* resolution. All the computation is done using *MATLAB® 2009a*.

4.3.3 Indoor stereo computation using a wireframe workspace

Next for our stereo experiments in an indoor lab environment, we have used a wireframe workspace as our test volume. The object detection and identification is achieved using symbolic color [representing RFID]. To avoid the illumination and pose problems, the experiment is performed in a controlled indoor scenario using a red ball. Since the ball has a spherical shape, it will have the same projection regardless of view point/pose. Both the cameras were connected to the same laptop. Since the cameras acquire image frames one at a time, there is a very small lag in synchronization, which can be disregarded at this stage.

To calculate the relative accuracy of the stereo results compared to the ground truth, we initially did experiments based on the dataset of images with known 3D points, instead of live feed frames. The cameras are calibrated using the affine transformation calibration procedure explained in Appendix B. The wireframe workspace volume $x \times y \times z$ is $27 \times 31.75 \times 24$ *in (68.6* $\times$ *80.6* $\times$ *61 cm).* The camera pair used were *27.5 in (69.9 cm)* apart. As explained before our stereo approach doesn't require the camera poses to be specially configured relative to each other. The distance between them is provided here just to present the layout of the test environment. The cameras were *59.25 in (150.5 cm)* away from the wireframe. Figure 4.11 shows the experimental setup for stereo testing and the spiral track for the red sphere. For reference, point *1* in is considered to be the origin. Both the cameras locate the red sphere in *RGB* color space. Runs have also been conducted using the *HSV* space. The color detection module gives the center of the sphere to the stereo module for every frame captured by the two cameras. The system

computes the 3D world coordinates by observing the same feature point (center of the red sphere) in two 2D images from cameras that are calibrated to the 3D workspace.

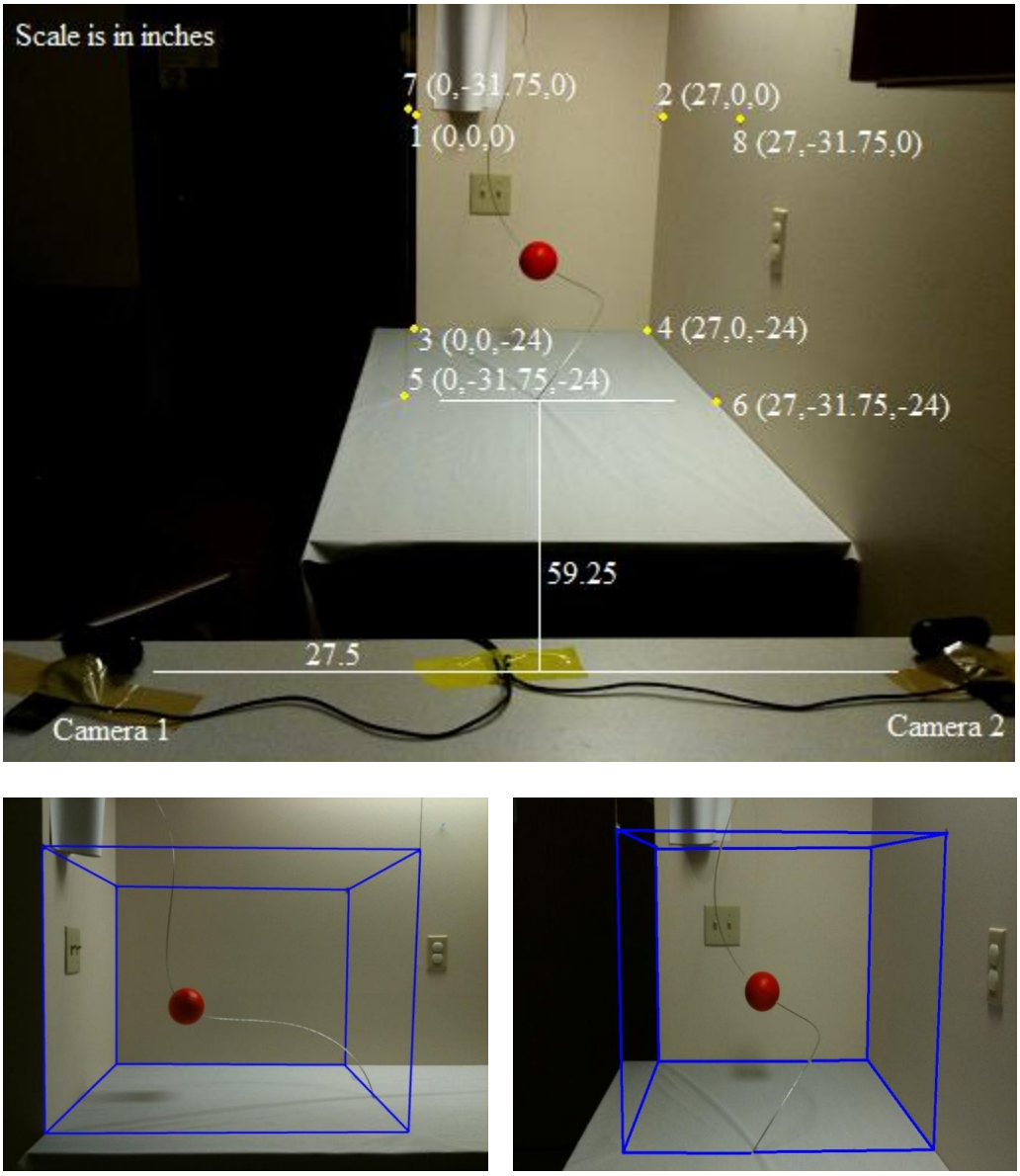


Figure 4.11

Wireframe workspace experimental setup for testing stereo localization indoors - red object with spiral trajectory.

With this procedure, whenever two or more cameras are calibrated, the user can then use the camera models to compute the 3D locations of any identifiable 2D feature point set that has not been used for calibration.

Ground truth data for several 3D points in the wireframe workspace were acquired. Some of these points were used for calibration and the rest were used to test the 3D sensing accuracy. Experiments show that generally eight calibration points in this setup can provide sufficient accuracy for further location estimation. These eight points can be chosen in a way to create a volumetric space of interest as shown in Figure 4.11 (b) and (c).

During the error analysis, the 3D residuals are computed. For points whose ground truth coordinates are known, the residuals are the difference between the ground truth coordinates and those computed via stereo. For other points, the residuals are an estimate of the standard deviation of the computed estimate of the coordinates. Table 4.3 shows the 3D residuals for some points in wireframe workspace. The maximum error is only *0.34 in (8.6 mm)* in a tracking space volume of *27×31.75×24 in (68.6×80.6×61 cm).*

Table 4.3

3D stereo residuals for some points in wireframe workspace - scale is in inches.

Actual 3D points			*Calculated 3D points*			*3D residuals*		
${}^{W}_{r}P^{x}$	${}^{W}_{r}P^{y}$	${}^{W}_{r}P^{z}$	${}^{W}P^{x}$	${}^{W}P^{y}$	${}^{W}P^{z}$	${}_{r}X$	${}_{r}Y$	${}_{r}Z$
13.2	*-30.7*	*-20.7*	*13.54*	*-30.61*	*-20.99*	*0.34*	*0.09*	*0.29*
15.9	*-14.6*	*-15.9*	*16.18*	*-14.29*	*-15.81*	*0.28*	*0.31*	*0.09*
8.2	*-10.3*	*-0.5*	*8.46*	*-10.04*	*-0.35*	*0.26*	*0.26*	*0.15*

The RMS error in three directions is calculated as X_{RMS} = *0.3 in (0.76 cm),* Y_{RMS} = *0.24 in (0.61 cm),* Z_{RMS} = *0.2 in (0.51 cm).* The maximum error is *~1.3%* of the *X-axis* of the calibrated volume. This error scalable to a construction space of $40 \times 40 \times 40$ *m* turns out to be X_{RMS} = *44 cm,* Y_{RMS} = *30 cm,* Z_{RMS} = *33 cm.* Keeping in view a *5 m* radius circle of safety for moving personnel to avoid any collision in the construction environment the localization error of *less than an arm length* in the 3D space using a vision only method validates its real-time applicability. Figure 4.12 shows the 3D stereo location estimation procedure. Figure 4.13 shows the computed trajectory of the sphere.

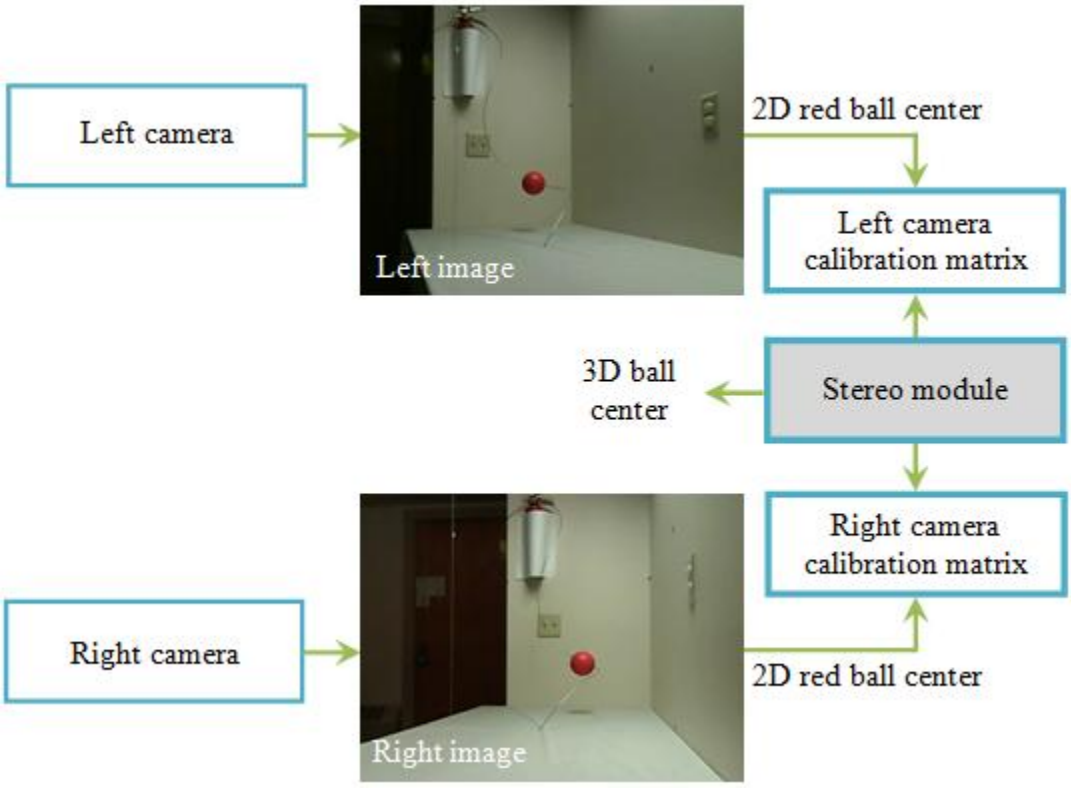


Figure 4.12

Procedure for 3D stereo location estimation in wireframe workspace indoors - Flow diagram.

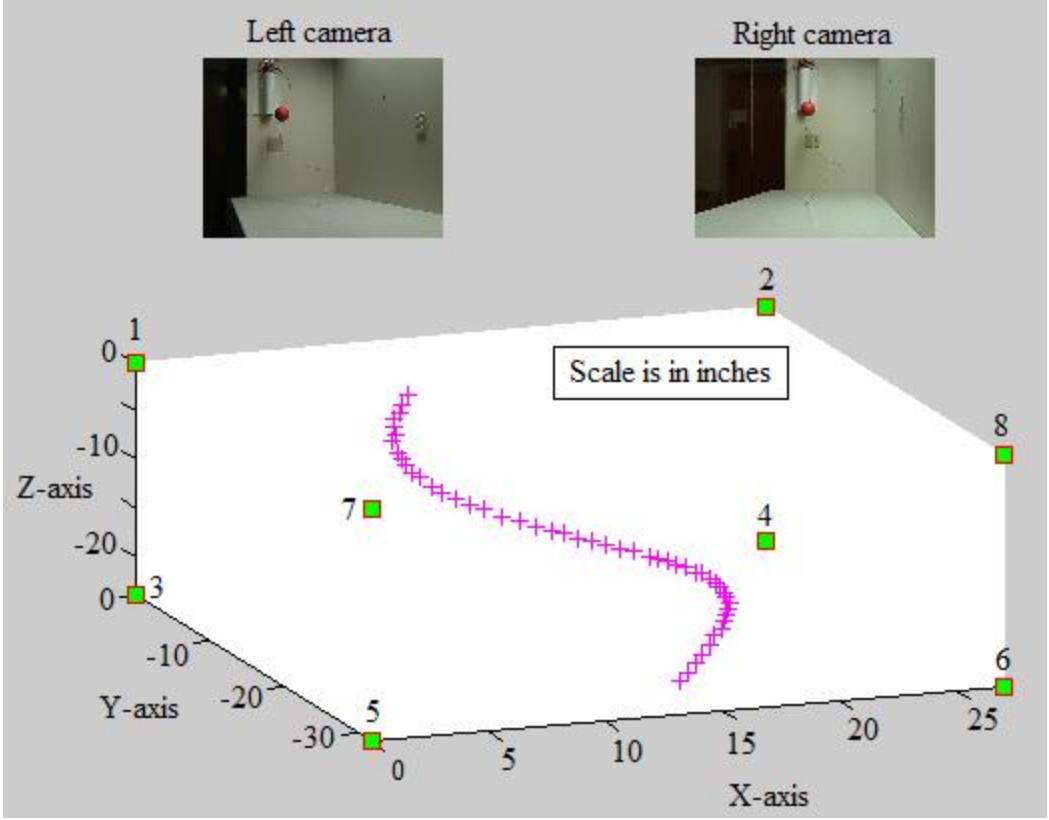


Figure 4.13

Computed sphere trajectory in wireframe workspace indoors.

4.3.4 Indoor stereo computation at surveyed lab area

To test stereo performance at a room scale, we surveyed an indoor lab area *280×476×105 in (7.1×12.1×2.7 m)*. The cameras were calibrated using affine calibration explained in Appendix B. The *11* calibration points chosen are marked with '■' (yellow) in Figure 4.14. The lab area and the 3D model generated from the survey data along with the position of the stereo cameras represented with '●' (green) are also shown in Figure 4.14.

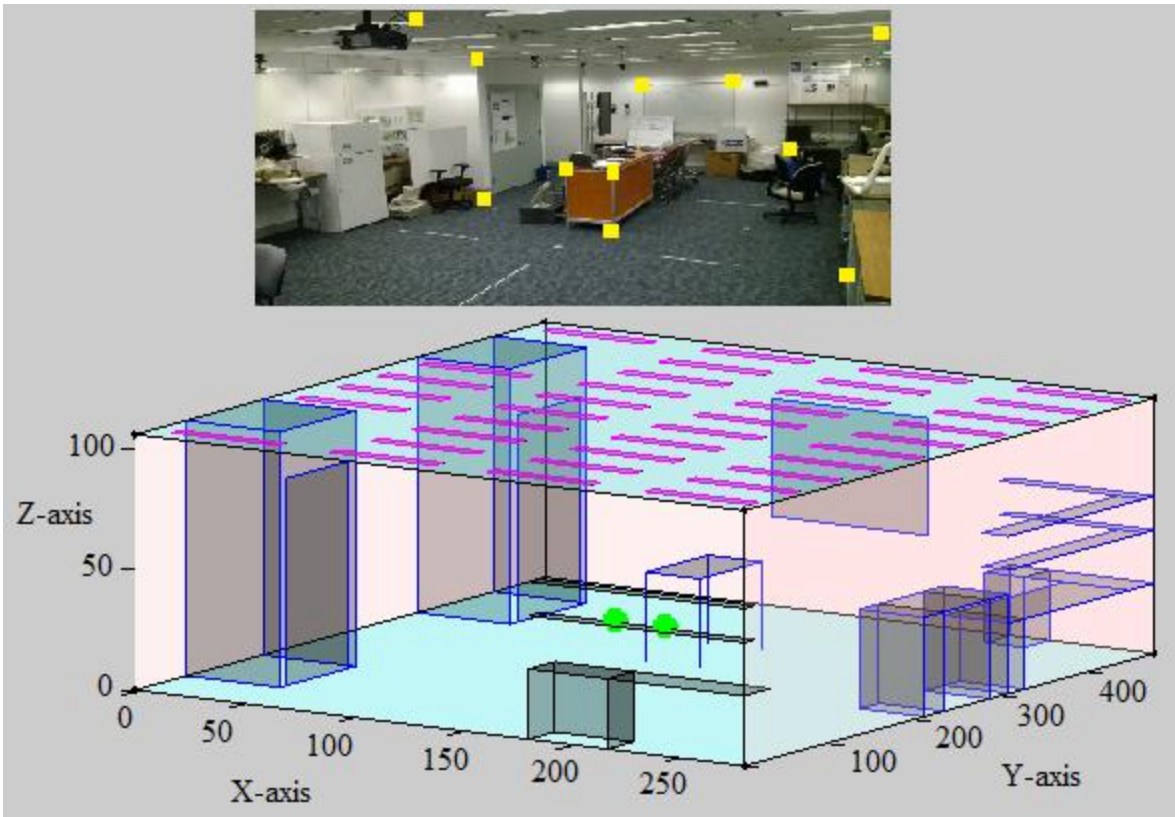


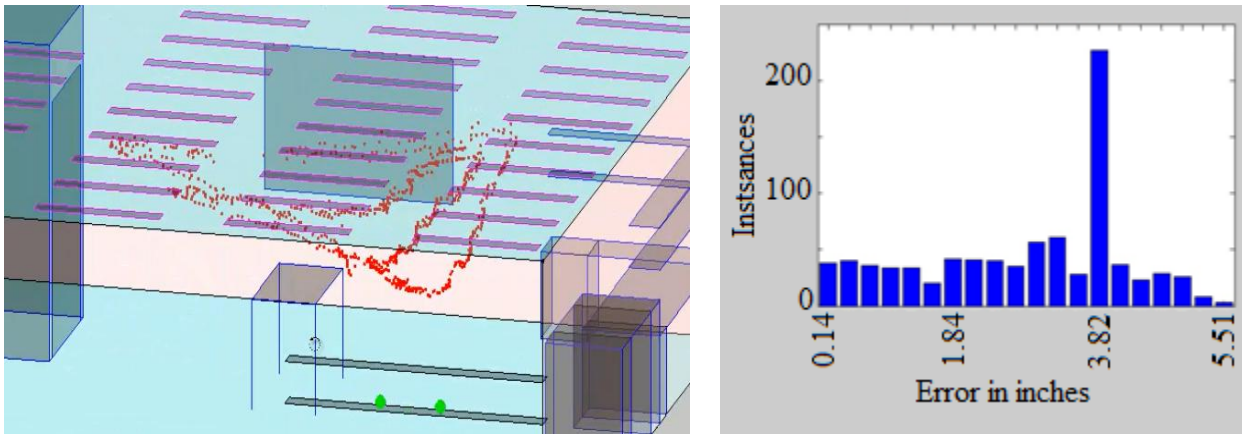
Figure 4.14

3D survey data and model of indoors lab area with stereo cameras '●' (green). Calibration points are represented by '■' (yellow).

The stereo cameras were placed *22.5 in (57.2 cm)* apart. The area was divided into *4×2* grid with center points marked as shown in Figure 4.14 top image. Each grid cell measured *119×140 in (3.02×3.6 m)*. The person moved in the common view area of both cameras while wearing a colored hat. The center of the hat was located in both cameras to perform stereo. Color detection was performed in both RGB and HSV space. To calculate the relative accuracy of the stereo results as compared to the ground truth, the experiments were performed with the person standing at the center of each grid cell *~15 ft (4.6 m)* to *34 ft (10.4 m)* from the cameras. Various other trajectories were also observed to analyze the stereo error. The RMS error observed was

X_{RMS} *=4.1 in (10.4 cm),* Y_{RMS} *= 6.4 in (16.3 cm),* Z_{RMS} *= 2.7 in (6.9 cm).* It is to be noted that the *Y-axis* here is along the camera viewing direction.

Another example of a trajectory of a person moving randomly is shown in Figure 4.15(a). Since the height of the person is known i.e *~72 in (1.8 m),* therefore it is easy to compare the z-dimension error here. Figure 4.15(b) shows the histogram of error in *z-direction* with error mostly accumulated in *3.82 in* histogram bin.



(a) (b)

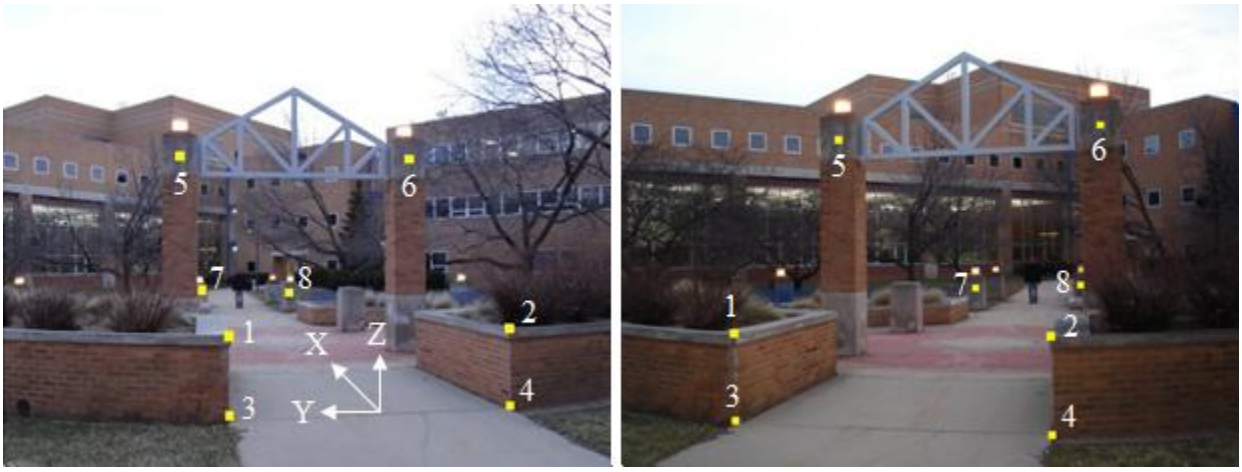
Figure 4.15

Indoors lab area trajectory for a person moving randomly with error estimation: (a) object tracks in 3D (b) histogram of error along z-axis.

4.3.5 Outdoor stereo computation

We installed our cameras in a *40×40 m* outdoor test area. The cameras were placed on *4.5 ft (1.37 m)* high pillars. The separation between the pillars is *~10 ft (3.05 m).* The persons were wearing distinctive color clothing and head gear. We located the center of the head of the

persons who moved on predefined points without stopping or, in some runs the 3D data was obtained with the persons being stationary for few seconds at those points. The analysis was done offline over several trajectories to compare with the ground truth. Figure 4.16 shows the left and right image with eight 2D calibration points shown by '■' (yellow). These points were selected in nearby and distant parts of the image. Choosing appropriate calibration points covering the near and far field of the scene is necessary for results with minimal error.



(a) (b)

Figure 4.16

Outdoor test site with 2D calibration points shown by '■' (yellow): (a) left image (b) right image.

The 3D view of the outdoor test site with object tracks is shown in Figure 4.17. '●' (red) shows the center of the test area. The 3D location of the person's head is represented by '+'. Analyzing individually two persons' trajectories moving within *10 ft (3.05 m)* to *80 ft (24.4 m)* distance from the stereo system, the RMS error observed was X_{RMS} *=7.6 in (19.3 cm),* Y_{RMS} =

5.2 in (13.2 cm), Z_{RMS} *= 3.8 in (9.7 cm)* as compared to the scaled error of X_{RMS} *= 44 cm,* Y_{RMS} *= 30 cm,* Z_{RMS} *= 33 cm,* from our lab experiments indoors.

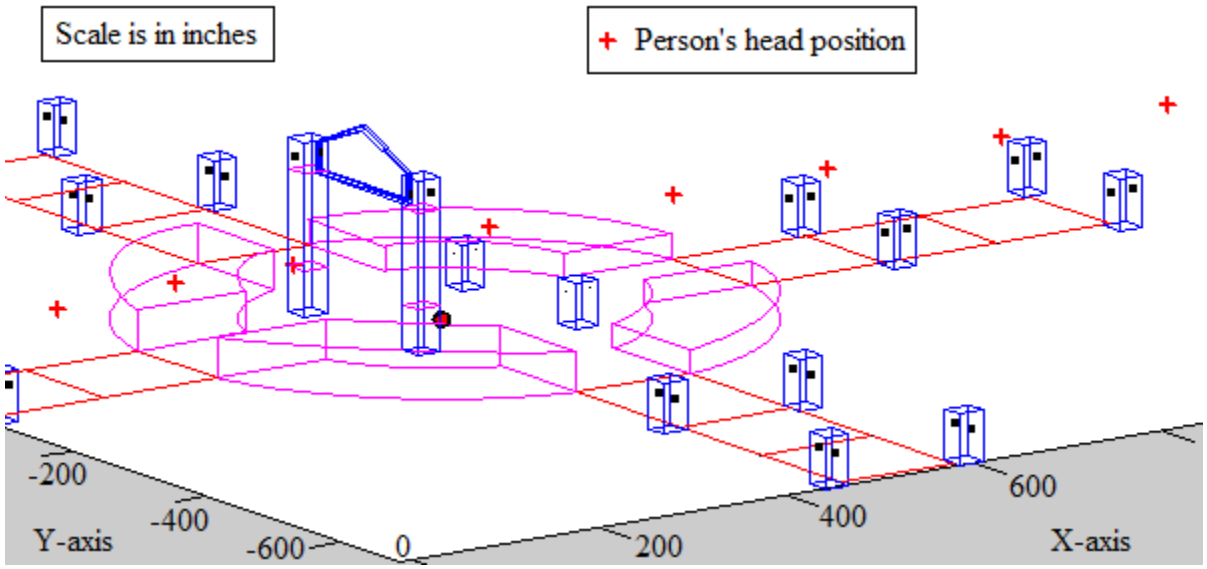


Figure 4.17

3D view of the outdoor test area with object tracks shown by '+' (red) using stereo.

As shown in Figure 4.18 the observations beyond *80 ft (24.4 m)* distance show the outliers with '•' (black) representing the ground. The stereo cameras are represented by '▲' (grey). The actual height of the person in Figure 4.18 is *72 in (1.83 m)*. The detected height of the person on average within *10 ft (3.05 m)* to *80 ft (24.4 m)* from the stereo system is *70 ± 4 in (10.16 cm).*

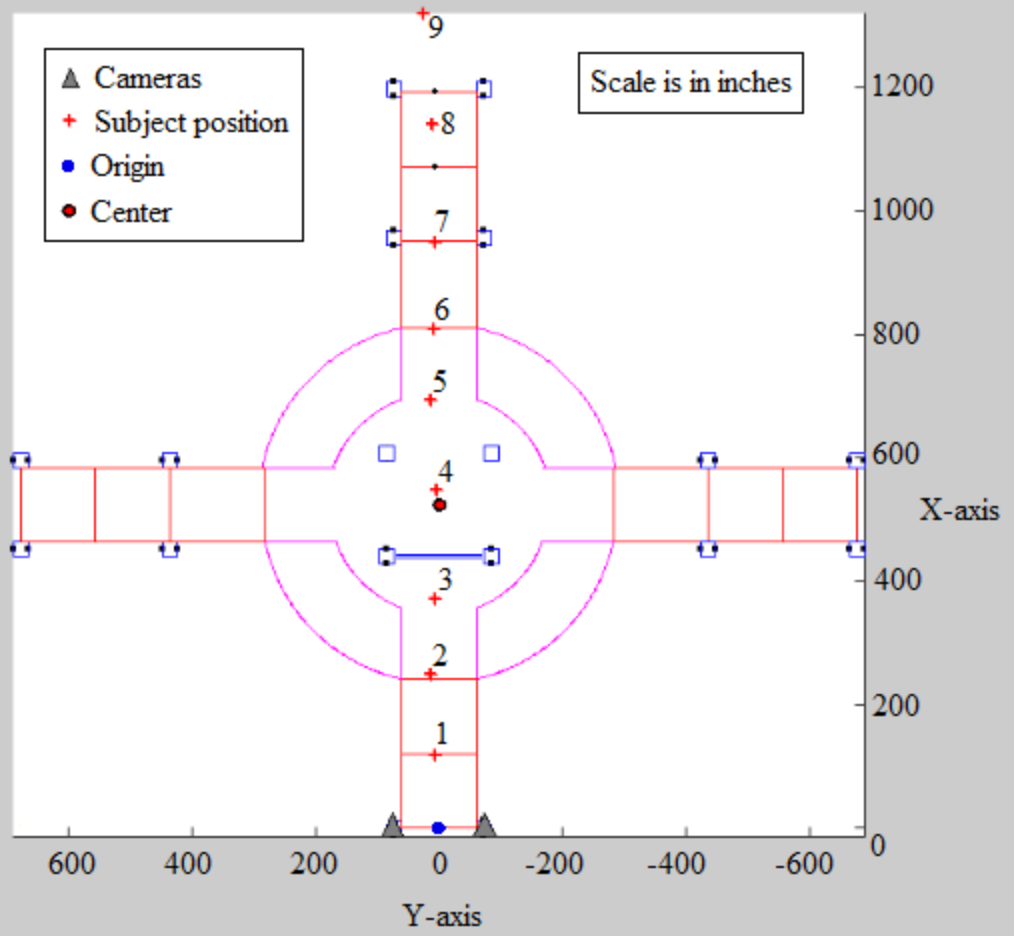


Figure 4.18

Top view of the outdoor test area showing person locations computed using stereo and the ground truth '•'of outliers beyond x = 960 in (24.4 m).

4.4 Active RFID based Real Time Location System

A Real Time Location Systems (RTLS) typically refers to a collection of sensors that work together to automatically identify and track the location of objects (including people) in real time.

4.4.1 RTLS infrastructure

We have used the Convergence Systems Limited RTLS [8] development kit. The kit includes one narrow beam width master RFID reader (CS5113TD) with ethernet support, five narrow beam width slave readers (CS5111TD) and ten active RFID tags (CS3151TC). Figure 4.19 shows four readers and tripods. The master and slave readers come in different beam width configurations. Due to our test area size and narrow beam width readers we have used one master and three slave readers as specified by OEM.

Figure 4.19

CSL RFID based RTLS system: Tripods hold the readers.

The equipment operating frequency is *2.4 GHz* and it uses Time-Of-Arrival (TOA) for location determination, where the distance between the tag and the readers is calculated by the roundtrip time. The equipment works on Non Line of Sight (NLoS) communication. As compared to Received Signal Strength Indicator (RSSI) methods the TOA based location estimation and *2.4 GHz* frequency makes the system more robust towards RF energy absorption by water and dynamic environments. Each reader has a beam width of 80^o in portrait and 30^o in landscape orientation. As per OEM instructions, for a cell with a square or near square shape, all the readers should be set up in a portrait manner. For a cell with a highly rectangular shape where one dimension is much longer than the other, all the readers should be set up in a landscape orientation. The active tags run on *3×AAA* batteries and are of the size of an iphone 4S and weigh less than the phone with batteries installed. The tag read range is up to *200 m* in an open space outdoors.

Triangulation is based on the geometric principle of triangles: if one side and two angles of a triangle are known then the other two sides of the triangle can be calculated . Trilateration determines the object position in 2D by measuring its distance simultaneously from three known locations using the geometry of circles. With the TOA scheme, location can be estimated using triangulation or trilateration. Triangulation can be used by antennas that search over a range of angles for best signal strength. Trilateration requires raw data from at least three readers. Intersection of the three circles around the readers can yield the location of the tagged object on the ground plane. However, the problem becomes more complex with issues such as clock synchronization, software delays and multiple paths that results in degraded position accuracy. Beep-Beep is another localization technique reported in [104] which avoids sources of

inaccuracy found in typical TOA schemes. The proposed beep-beep system uses cell phones commercially available and provides around two centimeters accuracy within a range of ten meters. However, sound level in the environment can limit the range of the beep-beep system accuracy.

Like all other electromagnetic waves, radio waves travel at the speed of light. The CSL RTLS basic principle of operation is that the total time of radio waves travel between the tag and the reader is multiplied by the speed of light to calculate the total distance of travel [21]. This number is divided by two to determine one way distance between the reader and tag. The CSL system further refines the estimate by accounting for hardware latency and uses probabilistic positioning algorithms that apply Bayesian statistics. As per CSL policy, details of location estimation algorithm were not provided, therefore analysis at the algorithm level is not possible.

The overall location accuracy varies depending on how often the tag transmits, the number of readers used, whether the tag is stationary or moving and the type of structures in the environment. The OEM reports that the system can provide average accuracy of one meter outdoors and two meters indoors for stationary objects.

The RTLS tag trajectories acquired in real time provide two dimensional data in the 3D workspace *xy-plane* due to limitation in CSL software provided with the equipment. Subsequently the readers placement are in the *xy-plane* of the workspace. The *z-plane* information is however considered constant and requires user interaction by defining tag height during the run. This assists in estimating correct location of the tag in the 3D workspace. The

RTLS location in the *xy-plane* of the 3D workspace carries sufficient information for supporting the 3D stereo location data and should not be confused with the 2D imagery data. Acquiring real time height data from RTLS is not limited theoretically and can be obtained by updating the software and deploying the readers in a 3D space.

4.4.2 Cell architecture

The system allows defining geographical areas, called cells, where tagged objects are localized and tracked. Each cell is made up of at least four readers, in most cases six, and up to eight readers in challenging environments such as an indoor warehouse. The maximum cell size can be 100×100 *m*. Larger areas can be segregated into multiple cells. To achieve good outdoor location accuracy, the tags should preferably be in the cell area where antenna beams of at least three readers intersect without any solid obstruction that can reflect RF energy. Moreover to increase system accuracy, software optimization tools can also be used.

4.5 RTLS based location estimation

We deployed the active RFID based location sensing equipment in indoors and outdoors respectively to analyze its location performance. We provide here the details and results obtained from our experiments.

4.5.1 Indoor location sensing

We placed our four RTLS readers (one master and three slaves) indoors in a hallway *27×12 ft (8.23×3.66 m)*. The readers were placed on the four corners to form a single cell. The reference tag was placed at the center of the cell and other points to check for the location accuracy. Figure 4.20 shows the indoor setup of RTLS. The average location accuracy for static tags was *~1.9 m*. The performance for dynamic tags varied a lot due to multipath effects in a compact space.

Figure 4.20

RTLS location sensing indoors - setup.

Figure 4.21 shows the floor map of RTLS indoors. It shows one active cell with the location of four readers and one tag in active state. The master reader is represented here as M1, and the slave readers are represented as S1, S2 and S3.

Figure 4.21

RTLS location sensing indoors - Floor map.

4.5.2 Outdoor location sensing

To check RTLS outdoors we installed the four readers in the four corners of the same test site (*40×40 m)* where we previously tested the stereo system. A single cell was configured using the four readers. Each reader was placed in a portrait orientation to provide beam width of 80^{o}. Figure 4.22 shows the outdoor setup of RTLS.

(a) (b) (c)

Figure 4.22

RTLS location sensing outdoors - setup: (a) master reader on tripod (b) reader pointing towards test site center (c) reference tag placed in test site.

As per the operator manual a 15^{o} spatial offset was applied to the readers to direct the antenna beam towards the center of the test area. Reference tags were also placed at specific positions to estimate location accuracy. A tagged person followed exactly the same predefined eight path points used in stereo runs with the person being stationary for few seconds at those points. The tag location was recorded when the person reached the desired points. This approach provided RTLS system accuracy in the workable area of the test site. Figure 4.23 shows the person locations ('♦') computed using RTLS at eight points. Four green squares ('■') show the location of the four readers.

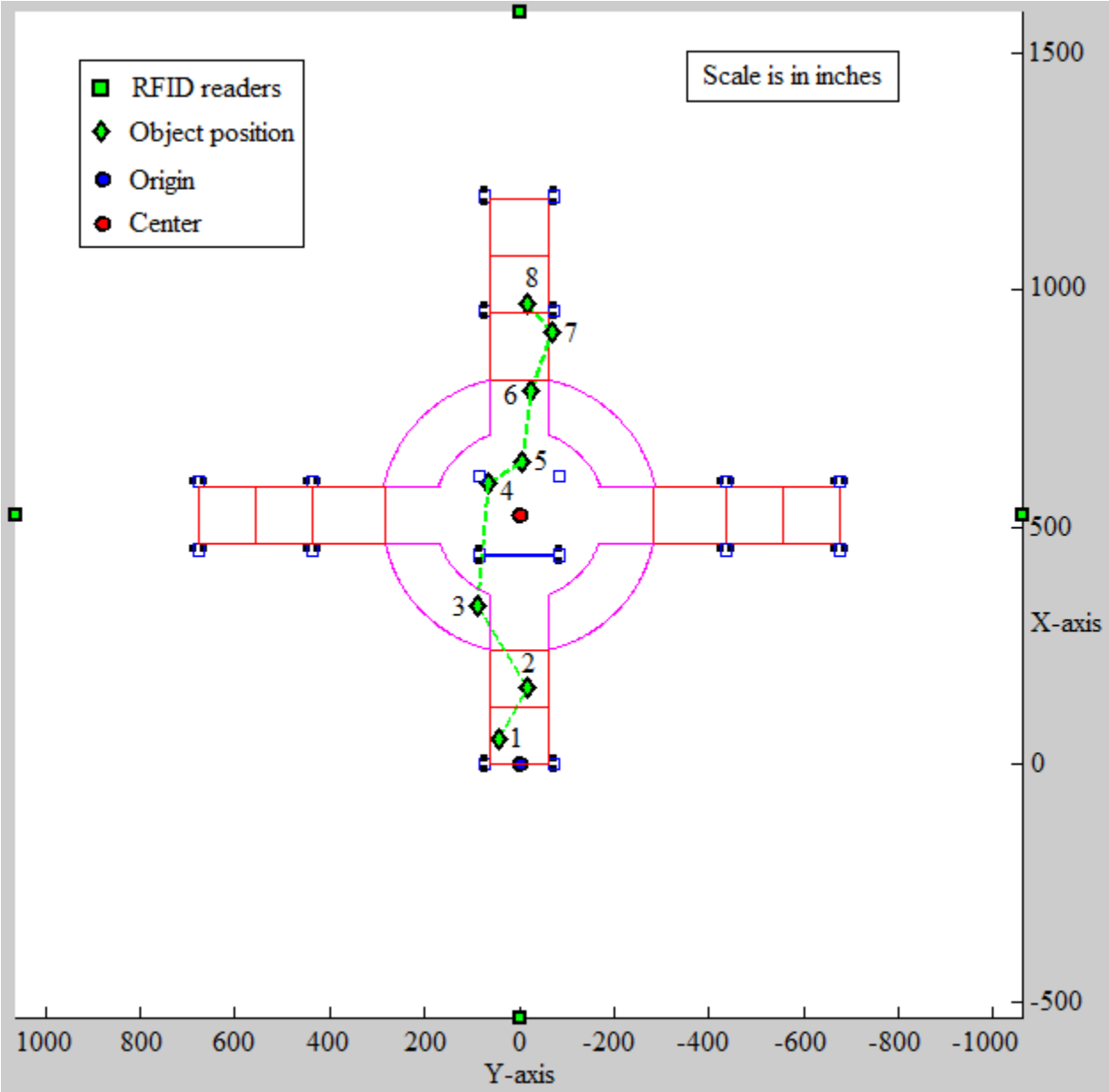


Figure 4.23

Top view of the outdoor test area showing person locations computed using RTLS.

Table 4.4 shows location data in the *xy-plane* and the difference between the ground truth observations and the location obtained using RTLS. The average error obtained along *x-axis* is $X_{avg\ diff}$ *= 60.1 in (1.53 m)* and similarly along *y-axis is* $Y_{avg\ diff}$ *= 54.1 in (1.37 m).*

Table 4.4

Comparison between ground and RFID observations in xy-plane- scale is in inches.

${}^{W}P^{x}_{ground}$	${}^{W}P^{y}_{ground}$	${}^{W}P^{x}_{RFID}$	${}^{W}P^{y}_{RFID}$	X_{diff}	Y_{diff}
120	*-1.2*	*53.02*	*50.4*	*67.0*	*51.6*
250	*0.7*	*162.5*	*-36.5*	*87.5*	*37.2*
370	*-2.9*	*333.1*	*89.2*	*36.9*	*92.1*
534	*-4.1*	*473.9*	*63.7*	*60.1*	*67.8*
700	*0.4*	*636.4*	*-45.4*	*63.6*	*45.8*
810	*-1.3*	*781*	*53.4*	*29*	*54.7*
950	*-3.2*	*912.8*	*-22.9*	*37.2*	*26.1*
1070	*2.1*	*970.3*	*-55.2*	*99.7*	*57.3*

Figure 4.24 shows the tagged person's locations computed using RFID ('♦') and the ground truth (○). The radius of the circles [shown in dotted blue line] represents the location error between the ground truth and the RFID observations. The center of the circle is at the ground truth observations. The circles are shown in different colors for easy representation. Since the test site center has solid structures, due to factors such as multipath and partial occlusion the location error increases as the tagged person moves towards the center of the test site. In the next chapters we will explain how this coarse location can help fusion.

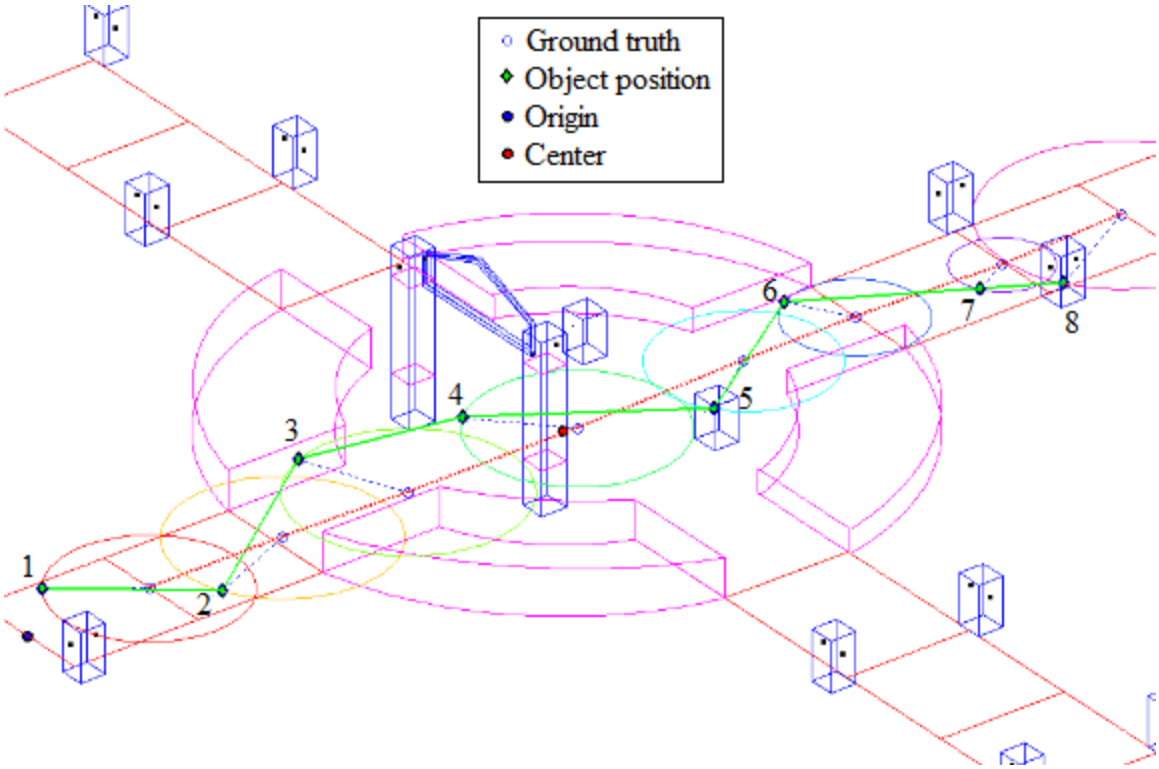


Figure 4.24

Computed locations of the person using RTLS and error estimation - error circles represent difference between ground truth and RTLS.

The average accuracy for the person location while being stationary at desired points was observed to be ~ *1.5 m*. The readers' locations needs to be accurately provided in the software otherwise the tag location accuracy will be affected. It is also noted that the RTLS system has at the minimum ~ *+2 sec* refresh rate with four tags in the area. The more tags in the cell the greater will be the time required to update the next location of each tag.

4.6 Summary discussion

In this chapter we have evaluated the localization performance of stereo vision and RFID as single modalities. Working towards an automatic system, we currently have some steps in the

research methods that require user interactivity. We have studied and implemented the ray-ray stereo scheme [31] for 3D localization in an indoor lab environment using commodity cameras and have reported an RMS accuracy of *~0.34 in (8.64 mm)*. Later the efficacy of the stereo approach was tested outdoors in a surveyed test site. In our analysis we have obtained RMS location accuracy within *~7.6 in (19.3 cm)* which offers potential for future researchers to examine automated three dimensional tracking outdoors using economical vision sensors. Later an RFID based location system was used for analyzing location performance. We have assessed that the system can achieve *~1.5 m* accuracy outdoors for tagged persons following predefined paths and being stationary for few seconds at selected points. The acquired RTLS system presently provides real time location in the *xy-plane* of the 3D workspace with approximately two seconds minimum system latency due to OEM software constraint and no provision of copyrighted location processing details to the customers. The *z* dimension, though available comes with planar restriction defined by the tag height. Getting interpolated location estimates and varying tag height data is however, not limited in theory.

In general we have shown how proper sensor placement can support localization and tracking. This enables the system to spend more resources on tracking anomalies -- "unauthorized" objects, machines, or materials in a site safety environment.

CHAPTER 5

Fusion dynamics and analysis

In performing data fusion, our aim is to combine and enhance the sensor information - so that it is better than would be possible if the sensor observations were used individually. In this chapter we define and characterize our sensor relationship w.r.t recognized standards and explain the basic building block of fusion. Next, general fusion benefits in a tracking environment are listed. Finally, particular examples of fusion using CV and RFID are explained to highlight how fusion may address the weaknesses of single sensing modalities.

With a single sensing modality the sensor will cover a limited region of the environment and provide measurement data of only local events, aspects, or attributes. Frequency of measurement or the refresh rate and the accuracy/precision of the basic sensing element in the sensor are some other constraints worth considering while choosing sensors for a tracking application.

Fusion of sensor data is a dynamic process that involves association, correlation, and combination of data derived from multiple sensors resulting in a fused product with a common representational format, which is more complete and accurate. Employing more than one sensor

in the workspace may enhance the synergistic effect in several ways, including: increased spatial and temporal coverage, increased robustness to sensor and algorithmic failures, better noise suppression and increased estimation accuracy. Data from multiple sensors could be of the same type or of different types. For the fusion process this data has to be represented in a common format that is meaningful in order to estimate or predict some aspect of an observed scene. In the paragraphs below we characterize our sensor setup and fusion approach in light of recognized standards.

5.1 The fusion process approach

Boudjemaa *et al.* [105] categorized fusion as *across sensors*, *across attributes*, *across domains* or *across time*. In fusion *across sensors* a same property is measured by a number of sensors versus the *across attributes* category where sensors measure different properties associated within the same workspace. In *across domain* the sensors measure the same attribute over varying domains. Lastly the fusion is categorized as *across time* when current measurements are fused with prior information.

Our sensor infrastructure as single modalities exploit a fusion scheme across sensors. This is because two or more cameras are used to observe the same workspace for stereo vision; however independent observations can be utilized when calculating depth information from single camera and single RFID reader. Also four or more RFID readers provide object ID and location information. Common representation of the location information *x,y,z* from RFID and vision as single modalities, into tokens $\tau = <x, y, z, t, v, L>$ depicts fusion across sensors. However, the

different feature set *v* of these modalities provided to resolve ambiguities between objects exemplifies the use of fusion across attributes. Our S-3 system can also use prior records *{<x, y, z, t-k>}*, or perhaps even forward records *{<x, y, z, t+k>}* to determine $L = f(<x, y, z, v, t>)$ to update the object tracks presenting fusion across time.

5.2 Multiple sensor configuration

Durrant-Whyte [106] classifies a multiple sensor data fusion system according to three basic sensor configurations. They are described as *complementary*, *competitive* and *cooperative*. In complementary configuration the sensors may not have a direct dependency relationship, however they can be combined to provide a comprehensive image of the phenomenon under observation. It is exemplified by the use of multiple cameras, each observing different parts of a workspace, to provide a complete view of the scene. Complementary sensors help resolve the problem of incompleteness. Fusion of complementary data is relatively easy because the data from independent sensors can be appended to each other. A competitive relationship among sensors is described as independent measurement of the same property by each sensor. This configuration allows combining competing sensors that are not necessarily identical. Calculating refined object location by RFID and vision location estimate uses competitive relationship. It provides robustness and fault-tolerance because comparison with another competitive sensor can be used to reduce the effects of uncertain and erroneous observations. Cooperative type data provided by two independent sensors are used to derive information that would not be available from a single sensor. The resulting data will be sensitive to the inaccuracies in all the individual

sensors. Our stereo approach where two independent cameras are used to compute the 3D pose of an object comes under cooperative sensor configuration.

In terms of usage, these sensor configurations are not mutually exclusive because more than one of these categories can be used in most cases. Figure 5.1(c) illustrates complementary configuration where cameras are networked to cover most of the workspace area and increase the workspace visual coverage. Figure 5.1(d) explains competitive sensor configuration where both RFID are CV are used to obtain ID and refined 3D location of the tagged object.

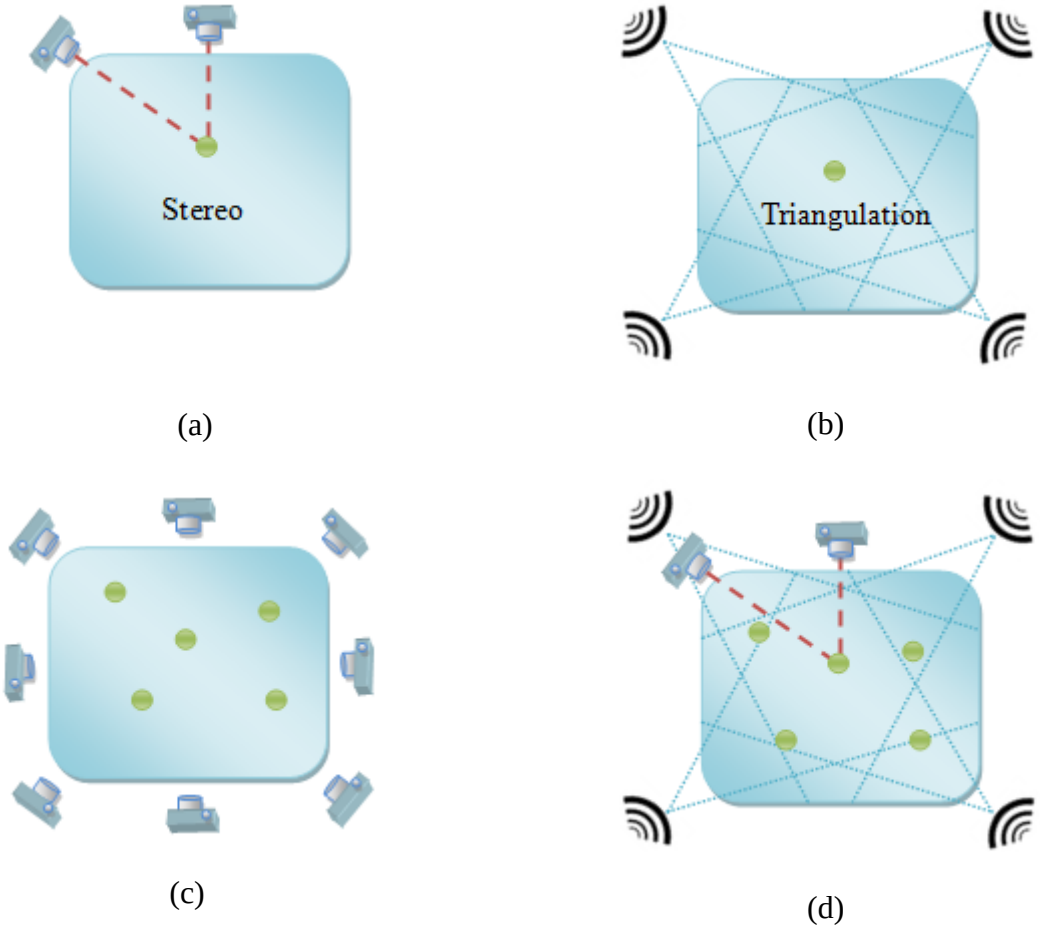


Figure 5.1

Multi-sensor configurations: (a) cooperative (b) cooperative (c) complementary (d) competitive.

5.3 Building block for multiple sensor fusion

The basic building block for multiple sensor fusion is called a *fusion node*. An overall system can have a distributed network of these nodes. The sensor observations ${}_sO_i$ are received as single or group inputs to the fusion node:

$$ {}_sO_{i,j} =< {}_sx_{i,j},\ {}_sy_{i,j},\ {}_sz_{i,j},\ {}_st_{i,j} > \qquad \begin{matrix} s \in \{CV, RFID, group\} \\ i \in \{\varnothing, 1, 2, 3, ...\} \\ j \in \{\varnothing, 1, 2, 3, ...\} \end{matrix} \qquad (5.1) $$

The single inputs, for example, can be 2D image points in case of stereo calculation. The sensor group input is provided when a 3D location of the object is provided by the cameras or RFID. The ${}_sZ_{i,j}$ is not applicable for 2D calculations. The feature set ${}_sv_{i,j}$ is provided to the node as auxiliary information from the sensor. This includes information such as color, texture, ID, height, dimensions, predefined labels, etc., that are generated dynamically or accessed from the database. The sensor process in the node processes the observations and forms tokens ${}_s\tau_{i,j}$.

$$ {}_s\tau_{i,j} =< {}_sx_{i,j},\ {}_sy_{i,j},\ {}_sz_{i,j},\ {}_st_{i,j},\ {}_sv_{i,j},\ {}_sL_{i,j} > \qquad (5.2) $$

The node receives data in common representational format and initiates data association, estimation and filtering processes. Data association and estimation can be based on hard decision methods, such as nearest neighbor, Euclidean distance etc., whereas Kalman filtering or particle filtering can be used for a probabilistic approach. We have used a hard decision method to

correlate sensor observations, which involves discrete relaxation labeling. All the data received by the node is combined together to produce a fused token. The output token represented as $\tilde{\tau}$ may, or may not, differ in position, time, value and uncertainty from the input observations. The fused output is also provided to the node for fusion across time which might be required to generate the object tracks. Figure 5.2 shows a graphical representation of a fusion node [107] which integrates in applications such as temporal tracking of objects in a given environment.

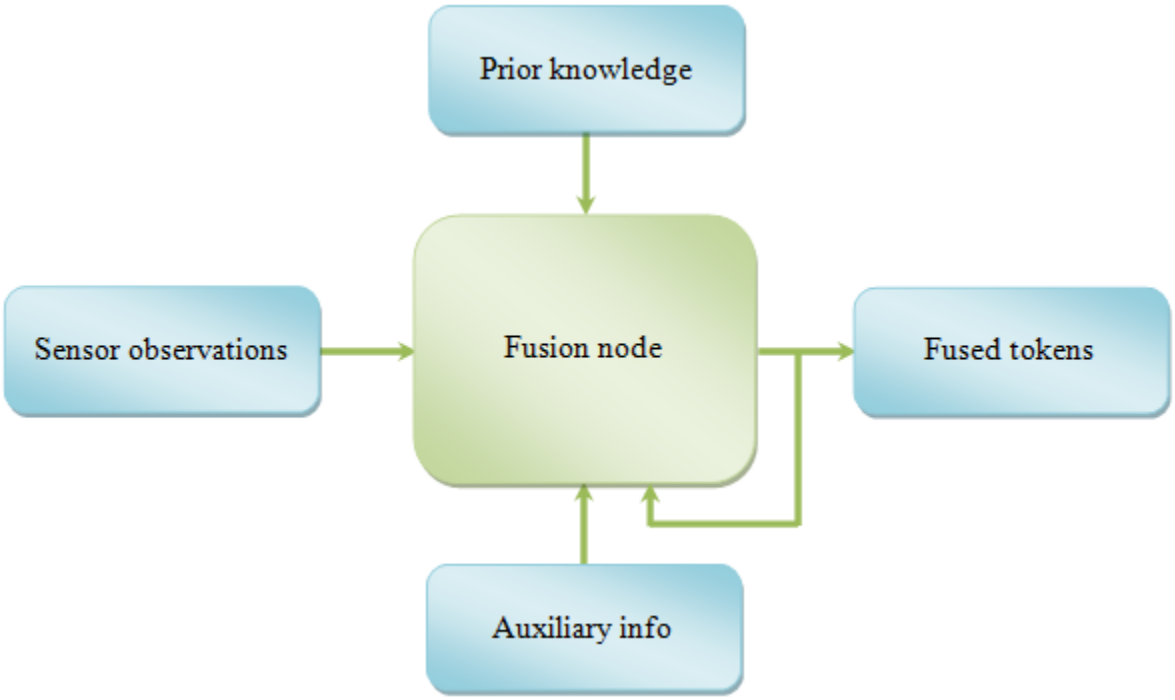


Figure 5.2

Basic fusion node architecture.

5.4 Benefits of fusing RFID and CV

Although object recognition and scene understanding using a purely CV approach is advancing, performance lags what many applications require. Lai *et al*. [108] report on object modeling and recognition experiments with *300* common objects using color, shape, and depth features. Coarsely summarized, they show *80%* precision at *60%* recall. This work involved stationary objects in indoor environments and mostly manufactured surfaces. NIST reports on a large base of biometrics experiments for person identification [109]. Some cases are reported where automatic systems perform better than humans and where there are viable commercial applications. In general, excellent performance is achieved only when high quality images are available. In the Advantage I-75 project, Walton *et al*. reports [110] optical license plate reading performance of *35%* to *45%* while RFID systems routinely achieve read rates exceeding *99%*. Many of the difficult cases for CV are due to poor image quality such as caused by dirty or bent license plates.

In many applications, engineering with RFID technology can avoid the problems of a purely CV recognition approach and can yield reliable object recognition leading to better object tracking and motion analysis as we will see in the next sections. In addition, some purely RFID solutions can benefit from adding some CV. Many retail stores use either bar code or RFID tags on items for machine reading. A customer, or "sweetheart" clerk, can cheat by changing the tag from one item to another, or perhaps using a counterfeit tag. Some CV capability at the checkout station can guard against this: visual features of the item can be sensed and compared to symbolic features stored on the tag. IBM's Veggie Vision system [43] can recognize *350*

produce items using color, texture, shape, and size features -- produce items are usually not tagged with bar codes unless packaged. This technology can be extended to thousands of other supermarket items as a check on bar code or RFID recognition. A related application of RFID is the EZ Pass highway toll-collection system [15]. Typically, a video camera is included in the system to take frames of cars that pass through without a legal transaction so that fines may be levied based on license plate ID. CV could be used here as a check that the car with the EZ Pass transponder has visual features corresponding to those stored in the active RFID tag. This adds security against theft or cloning. CV can be used similarly for person ID when credit cards or "smart cards" are used. Symbolic features on the card can be compared to an image of the live person using the card - either by a clerk or automatically. The state of the art of person verification is good enough to support this operation.

There are many possible applications based on fusion of RFID and vision. They can range from reliable object and person recognition, to assisting persons with disabilities, making shipping more efficient, and enhancing construction site safety. Specific examples are already provided in Chapter 2. Some important general characteristics of the fusion are as follows:

a. *Improved object recognition accuracy* - Uncertainty depends on the object being observed and arises in case of occlusions and limited sensor measurement accuracy. Since the primary purpose of RFID is object identification, object detection and identification/recognition will be much more accurate in the fused infrastructure compared to a CV only approach thereby decreasing uncertainty.

b. *View independent tracking* - The fused system has the main benefit of view independent tracking. This is only possible due to the powerful tool of wireless identification in RFID.

c. *Easy object representation* - Representation of the objects and/or humans is a complicated operation when using only 3D CV. The fusion approach provides direct symbolic representation.

d. *Robust segmentation* - With the fusion approach, object identification in a scene becomes easy and accurate since object model information can be retrieved from the object tag. Fusion can provide robust segmentation of images in real time.

e. *Training free learning* - The fused system can train itself on the fly for the tagged objects in the environment. By directly sensing object ID, the system can efficiently learn object appearance models using the CV component without human intervention. New objects teach the system when they arrive.

f. *Compatibility of RFID passive tags and CV* - RFID passive tags are of very small size. Applying them to objects does not change their appearance or pose. Therefore their use does not hamper operation of any vision based techniques. Active tags though come in bigger sizes.

g. *Efficient handling of occlusion* - The fused system can be used for handling the occlusion problem without increasing the computational and hardware complexity if multiple cameras are used.

h. *Supplemental object information* - The onboard memory of an RFID tag can carry a variety of information about the tagged object. For example, an object can transmit its color and size, which supplements the visual feed. This element is essential in designing autonomous systems that require real time interaction. Moreover, RFID tags can transmit important non-visual information, such as object weight or chemical composition.

i. *Increased spatial and temporal coverage* - The networked RFID and cameras can increase the workspace spatial and visual coverage.

j. *Improved resolution* - When multiple independent observations from similar sensors of the same property are fused, the resolution of the resulting value can be better than a single sensor's observation

To benefit from sensor fusion, there is a need to combine the strengths of RFID and CV to avoid problems of each mode. Fusion must be designed so that problems such as failure in un-tagged environments, high data rate, and lack of positional information do not defeat its principles.

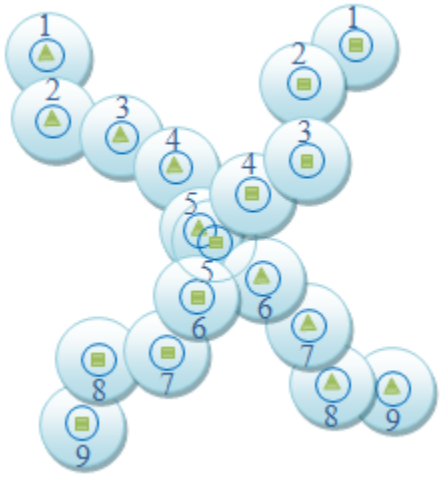
5.5 Test cases to explain fusion and its analysis

We illustrate in this section how CV and RFID in a competitive configuration supplement each other in critical test cases. For clarity we first assume that for case I and II, both CV and RFID feeds are continuously available and the objects are not occluded by each other or by the background and are moving with approximately the same velocities. We briefly mention relaxation labeling based filtering which eliminates the incompatible labels iteratively. To maintain sequential flow of concepts, relaxation labeling will be explained in Chapter 6. The cases mentioned below are symbolic versions of our outdoor test trials.

5.5.1 Test case I - Same colored objects

For case I, consider two objects represented as '▲' and '■' with 3D data at each instance over nine time frames. For better visualization the tracks in Figure 5.3(a) are displayed in the *xy-plane*. The objects are converging from north to south towards each other, and intersect at time frame *5* and thereafter follow their direction of motion. Even if the objects were moving in a straight line, the points would appear to be scattered along the true path due to propagating location errors and distortions in a 3D space. CV and RFID location accuracies are shown with circles. The inner circle around every point shows the localization error of CV and the outer circle represents that of RFID. Figure 5.3(a) and (b) shows case I where both the objects are of the same color. Figure 5.3(a) shows the object tracks and Figure 5.3(b) represents label assignments. The CV system can correctly assign labels to '▲' as label *1* and '■' as label *2* up until time frame $t = 4$. Thereafter there is a probability of no ID assignment, which based on

relaxation labeling means no wrong label elimination and is represented here as keeping both possible labels for both points. On the other hand, RFID provides correct label assignments other then at $t = 5$ due to fully overlapping localization error of point '▲' and '■'. In this case RFID helps CV to generate correct object tracks. However, no label elimination in the intersection area is possible. CV contributes by refining location.



(a)

CV alone									
Time	1	2	3	4	5	6	7	8	9
Label 1	▲	▲	▲	▲■	▲■	▲■	▲■	▲■	▲■
Label 2	■	■	■	▲■	▲■	▲■	▲■	▲■	▲■
RFID alone									
Label 1	▲	▲	▲	▲■	▲	▲	▲	▲	▲
Label 2	■	■	■	▲■	■	■	■	■	■
CV + RFID									
Label 1	▲	▲	▲	▲■	▲	▲	▲	▲	▲
Label 2	■	■	■	▲■	■	■	■	■	■

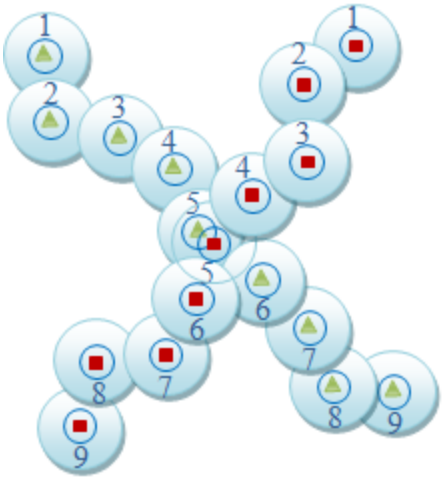
(b)

Figure 5.3

CV and RFID supplementing each other - Case I: (a) same colored object tracks (b) label assignments.

5.5.2 Test case II - Different colored objects

Figure 5.4(a) and (b) shows case II where objects are of different color. Due to no occlusion CV will be able to provide correct label assignments. However, RFID will have no label assignments at $t = 5$. Fusing both feeds, CV supplements RFID here and the label assignment at $t = 5$ is obtained.



CV alone									
Time	1	2	3	4	5	6	7	8	9
Label 1	▲	▲	▲	▲	▲	▲	▲	▲	▲
Label 2	■	■	■	■	■	■	■	■	■
RFID alone									
Label 1	▲	▲	▲	▲	▲■	▲	▲	▲	▲
Label 2	■	■	■	■	▲■	■	■	■	■
CV + RFID									
Label 1	▲	▲	▲	▲	▲	▲	▲	▲	▲
Label 2	■	■	■	■	■	■	■	■	■

(a) (b)

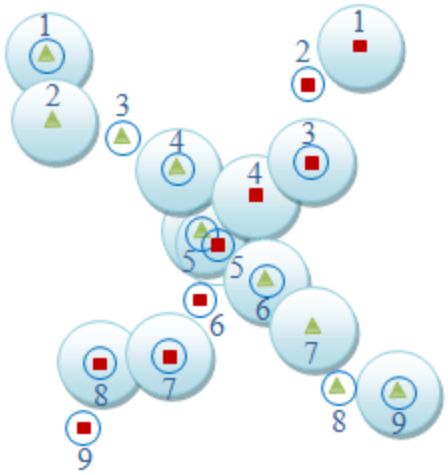
Figure 5.4

CV and RFID supplementing each other - Case II: (a) different colored object tracks (b) label assignments.

For both case I and II, CV support can also be clearly appreciated when RFID location error is maximum (i.e on outer circle boundary) for two points having overlapping localization error in consecutive frames.

5.5.3 Test case III - Different colored objects with intermittent RFID/CV feeds

Figure 5.5(a) and (b) shows case III where some of the objects are occluded by the background and the RFID feed is intermittent. This is represented here as missing vision and/or RFID location accuracy circles. The dash symbol shows non-availability of observation for that time instance. Comparing Figure 5.5(a) and (b) it is obvious that CV and RFID supplement each other at missing spots and fusion of these generates correct label assignments.



(a)

CV alone									
Time	1	2	3	4	5	6	7	8	9
Label 1	▲	–	▲	▲	▲	▲	–	▲	▲
Label 2	–	■	■	–	■	■	■	■	■
RFID alone									
Label 1	▲	▲	–	▲	▲■	▲	▲	–	▲
Label 2	■	–	■	■	▲■	–	■	■	–
CV + RFID									
Label 1	▲	▲	▲	▲	▲	▲	▲	▲	▲
Label 2	■	■	■	■	■	■	■	■	■

(b)

Figure 5.5

CV and RFID supplementing each other - Case III: (a) considering visual occlusion and intermittent RFID, different colored object tracks (b) label assignments.

5.6 Summary discussion

This chapter presented a multi-sensor fusion configuration that combines information from vision and RFID and generates tokens. We have presented the attributes of our fusion scheme to explain its adaptability towards established fusion standards. It explains the competitive and cooperative relationships of our sensors and the fusion building block needed to develop the fused system. The potential benefits of fusion focusing on localization and tracking tasks are also highlighted. To practically elaborate the potential in fusion, we have demonstrated test cases where fusion disambiguates object tracks and combines the strengths of RFID and vision to avoid problems of each mode.

CHAPTER 6

Tracking using fusion of CV and RFID

Object tracking comprises estimation of the current position and orientation of a tracked object and its motion, usually based on noisy measurements that generate uncertainties especially for dynamic objects. This chapter covers the algorithm development procedure for object tracking using fusion. We introduce a relaxation labeling scheme that can implement object tracking. The constraint satisfaction process is based on fusion of CV and RFID. Further, we have used smoothing for optimization, which is a high level technique that operates globally, to update computed tracks for increasing tracking reliability.

An object track is *k* or more tokens in time sequence with consistent object ID and features that also satisfy constraints for motion in space. The complexity of the tracking problem increases for multiple object tracks. Observations from multiple sensors helps decrease the uncertainty.

6.1 Object tracking model using sensor fusion

In the process of object tracking, object tracks are updated by correlating sensor tokens with the existing tracks or by initiating new tracks using tokens from different sensors. An object

track here can be defined as a temporal sequence of assigned tokens with consistent features and label that also satisfy constraints for motion in space. Token association, which provides token-token or token-track correlation, facilitates the constraint satisfaction iterative steps. The process of relaxation labeling filter out incompatible objects leaving behind compatible candidates for track updates by the optimization process. We have used smoothness as an optimization process. Details of the smoothness algorithm will be provided towards the end of this chapter. Figure 6.1 describes tracking using sensor fusion.

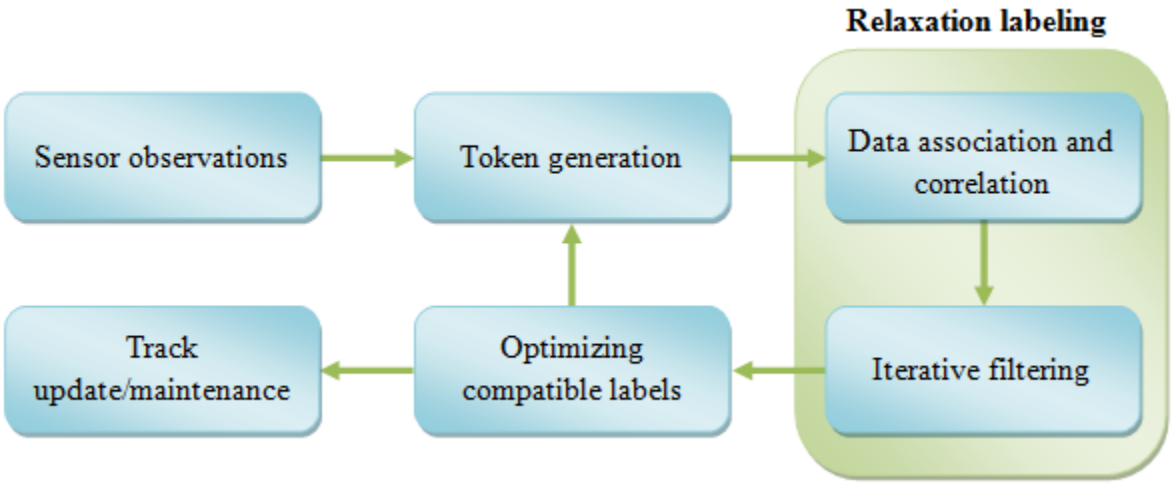


Figure 6.1

Schematic diagram of object tracking using sensor fusion and relaxation labeling.

6.2 Labeling via iterative processing

Relaxation labeling is an attractive technique because it is highly parallel, involving the propagation of local information via iterative processing. Suppose there are *N* objects detected at

a particular time instance *t*. We use discrete relaxation to create the tracks of these *N* objects and to update the time tokens comprising each track. Using relaxation, different sensors and sources of information can be turned on or off for experimentation or for practical reasons at a site. Fusion processes operate on a blackboard containing the set of tokens. When an observation is made, its initial label set is the set of all possible *N* known objects. Filtering processes are then applied to eliminate labels inconsistent with constraints. Sensing continues over the *T* time steps and naïve physics processes aggregate object consistent tracks.

For clarity, suppose that *N* objects are detected at time *t=1* and that we arbitrarily label these objects *1,…,N*. At time *t=2*, we have another *N* observations and we want to label each of those with the labels from time *t=1*. A label possible for a token at time *t=2* will be consistent in color, motion, and RFID ID with the tokens at time *t=1*. Initially, a new observation may detect any of the known objects, so all labels *L* are possible. A totally new object entering the site could be given a new unknown label. Most of these labels are filtered out quickly by failing constraint satisfaction criteria. For example, suppose five orange hard hats are detected at *t=1* and these have initial labels *3,4,6,8,9*. For any token for time *t=2* that is not orange, labels *3,4,6,8,9* will be deleted from its possible label set. Filtering can be done by space as well as by color. If any token at time *t=2* is unreasonably far from a token *m* at time *t=1*, then label *m* should be deleted from its label set.

6.2.1 Sensor process

A CV sensor takes a video frame, segments it into special regions of color, and provides two points on each of the imaging rays that are presented as sensor observations. Object region features are provided in the feature vector *v*. Using this information, the sensor process then generates tokens, one for each segment detected. Object label *L* is initially unknown. An RFID reading produces a similar token, except that an object label *L* is known in almost all cases.

6.2.2 Combination process

Fusion processes take the sensor observations and generate tokens and possibly merge information using ray intersection, ray-surface intersection, etc., whichever applies, and outputs a token with refined 3D location or label information. Filtering processes eliminate unlikely token labels by comparing tokens and by looking at feature vectors over time. The current software implementation of relaxation inputs combined tokens that have been pre-computed from stereo correspondence. Similarly, RFID tokens have 3D information from the encapsulated RTLS system.

6.2.3 Tracking process

Naïve physics constraints are used to filter out highly unlikely labels for objects at time *t* based on the recent history of objects continuing from the *k* previous time steps. Our current results have used the current and two previous time steps.

6.2.4 Relaxation labeling algorithm

The processes and constraints described above are now formalized into an algorithm:

Output: Object Labels at L_k and 3D location $XYZ_{refined} \epsilon R^3$ for each object

Input: Object Labels at L_{k-2} and L_{k-1} with color, RFID and XYZ_{RFID} and $XYZ_{stereo} \epsilon R^3$

FOR $t = k : K_{frames}$ */*Process all time frames */*

- Obtain color information if any for XYZ_{stereo} observations from 2D histogram matching
- Sort colors into groups */* How many colored hats and balls and which colors*/*

Detect

- n number of XYZ_{stereo} observations detected */* Motion detection and color detection*/*
- m number of XYZ_{RFID} observations detected */* Active RFID*/*
- Merge tokens and generate empty label matrices for p observations */* $p = max(n,m)$*/*
- Assign p labels to all p observations and proceed to next pass

Identify */* Binary relationship criteria*/*

- Identify XYZ_{stereo} observations based on color information
- Identify XYZ_{RFID} observations based on ID
- Correlate identity information

IF Only one color group */*All XYZ_{stereo} observations have same color*/*

No label elimination and proceed to next pass

ELSEIF Different color groups */*Some XYZ_{stereo} observations have different color*/*

Eliminate labels from *p* label matrices based on respective color groups and proceed to next pass

END IF

Locate */* Binary relationship criteria*/*

- Set stereo and RFID location threshold values */*Thresholds are defined based on sensor location accuracy and object speed*/*
- Locate XYZ_{stereo} observations
- Locate XYZ_{RFID} observations with ID and location
- Correlate location information

FOR $i = 1 : p$

StereoObservationSet(i) = Starting at $t=k-1$ use stereo threshold and find near neighbors at $t=k$ for XYZ_{stereo} observation

IF Label is in *StereoObservationSet(i)*

Keep label

ELSE

Eliminate label and proceed to next pass

END IF

END

Smooth

- Calculate direction of flow/velocity for every XYZ_{stereo} observation at $t=k$ relative to $k-2$ and $k-1$ */* z dimension gives valuable information here */*
- Correlate with RFID label/s ID and location information from XYZ_{RFID}

IF No difference in flow detected

Labels kept

ELSEIF Difference in flow detected */* Only compatible labels remaining.*/*

Eliminate unlikely labels

END IF

Compatible label/s obtained */*All possible labels for specific object*/*

- Compatible label/s provided to optimization process to obtain $XYZ_{refined}$

END FOR

6.2.5 Test cases to analyze discrete relaxation labeling

To realize how the dynamics of relaxation labeling can fuse information we describe here some related critical test cases.

6.2.5.1 Test cases IV - Two same colored objects with simple dynamics

We start with a simpler dynamics as shown in Figure 6.2 where two objects having the same color features moving from west to east first converge, move side by side for some time, and then diverge. The possible compatible labels after fusion are shown below each block of time frames. Assuming there is no occlusion, the labeling algorithm generates unique compatible labels '▲', '■' before the object tracks intersect at *t=4*. Thereafter there will be no label elimination until *t=7* due to overlapping location errors of CV and RFID. The applied constraints can then categorize inconsistent labels at *t=8* and onwards. Moving towards a more complex scenario in the next test case, the structure and efficacy of the relaxation scheme is explained step by step.

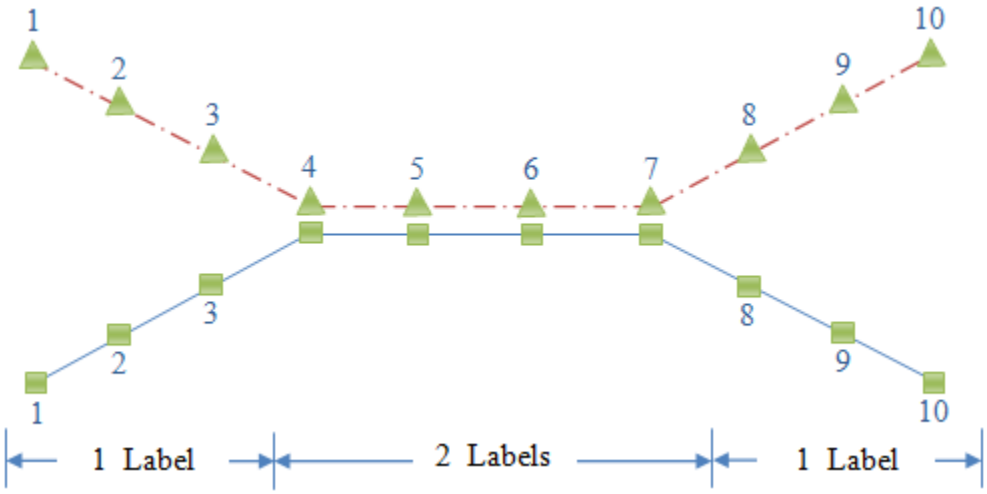


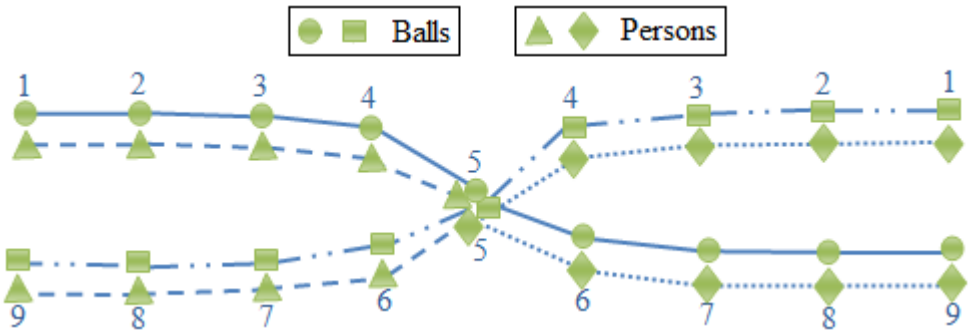
Figure 6.2

Correct object tracks with possible compatible labels at each block of time frames.

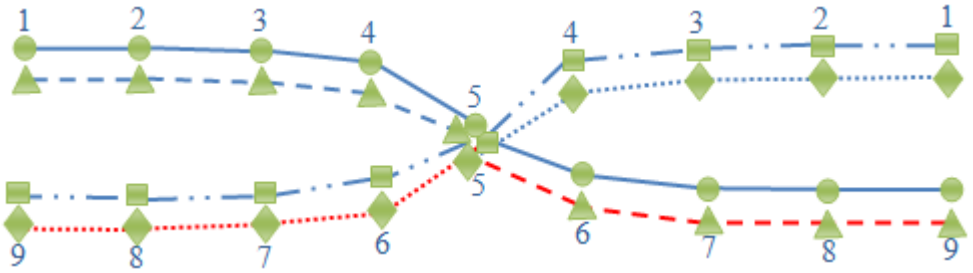
6.2.5.2 Test cases V - Four same colored objects with increased complexity

Two persons '▲', '♦' carrying two balls '●', '■' move towards each other and meet at the center of the test area. They then exchange the balls and move back towards the direction of their starting positions. The return paths are separated for better illustration. Both persons and both balls are tagged. To test algorithm robustness and increase complexity we consider that the color of the balls and the persons head gear are the same. Figure 6.3(a) and (b) show the 3D points over consecutive time frames.

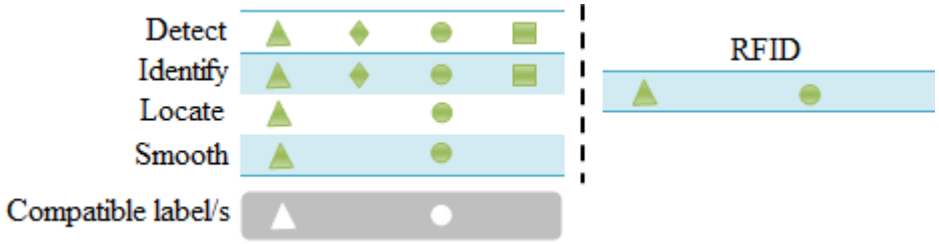
Here Figure 6.3(a) represents the correct trajectories of the persons and the balls. If we assume that there is no occlusion and the stereo feed is continuously available then Figure 6.3(b) shows incorrect trajectories of the persons calculated by the stereo feed alone.



(a)



(b)



(c)

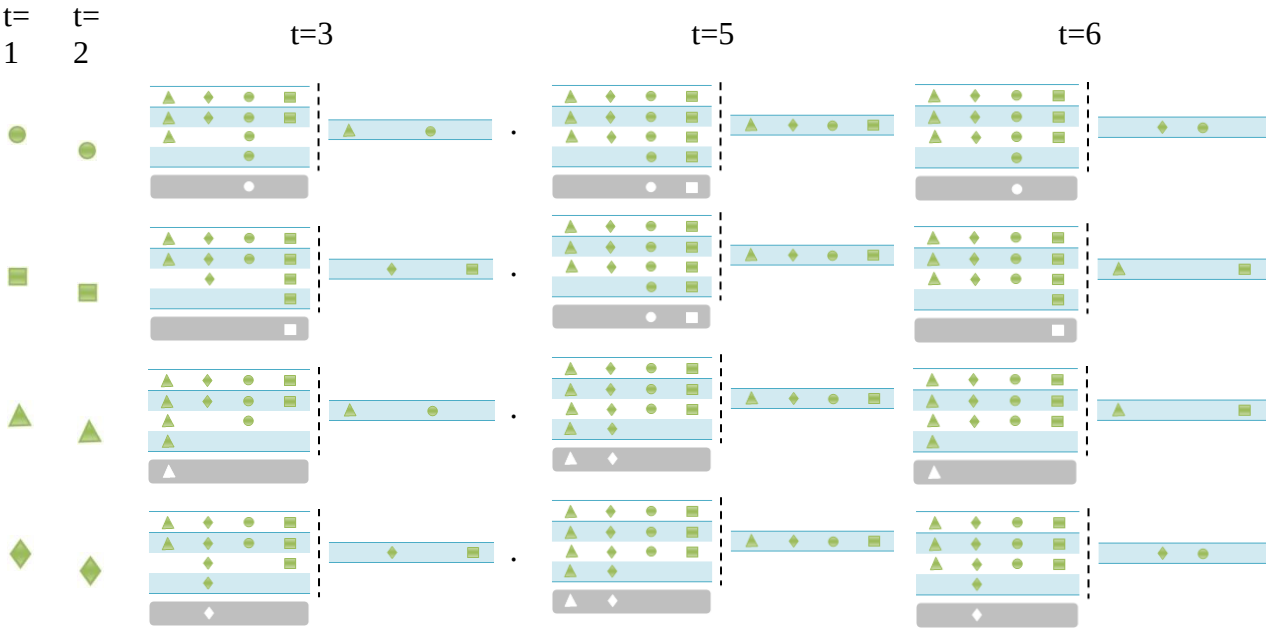
Figure 6.3

Test case showing relaxation labeling: (a) correct trajectories of persons and balls. (b) correct balls and incorrect persons' trajectories (c) left matrix - General pattern of four relaxation constraint passes and final compatible label/s. Right matrix - RFID location information.

It is assumed that we have information about the feature set and 3D location of the object labels at time frame *t=1* and *2*. For subsequent time frames we correlate CV and RFID information and apply constraints in *detect, identify, locate* and *smooth* passes. Constraint based label elimination by these filtering passes update the label matrix for every observed point at every time instance *t>2*. Once all impossible labels are removed and no further elimination is possible then the remaining label/s is/are considered as compatible label/s. The label/s is/are then passed on to the post-processing optimization process for updating fused token feature vector *v* and the refined location *XYZ* and, where required, determining a possible unique label amongst the four compatible labels sets. The optimized labels acquired are then assigned to the observed points respectively. Figure 6.3(c) demonstrates a typical label matrix on the left that shows all four passes with the remaining compatible labels at the end. The RFID location information on the right is shown with each label matrix to provide evidence of objects presence.

For the observations in Figure 6.3(a), a step-by-step explanation is provided on how the label matrices in Figure 6.4 are updated. For time frame *t=3* and *4* in each label matrix the objects are detected in the *detect* pass based on motion, color and ID and subsequently all the possible labels are assigned to all the observed object points. In the *identify* pass the system identifies objects based on color groups and ID from RFID. Since the color information for the observed points is the same no label is eliminated at this pass. The color histogram similarity measures can be used for color based sub-grouping. In the *locate* pass the labels are eliminated based on near neighbors where thresholding is done using sensor location accuracy and object speed. This helps identify '▲','●' and '♦', '■' as consistent label pairs. The two inconsistent labels are then eliminated from the respective label matrices.

The *smooth* pass correlates labels with RFID and deletes one further label with unlikely motion according to local (*3 point* or *2 point*) smoothness and object height constraints. This completes relaxation labeling. Global tracking is done as post processing. Note that at $t = 5$ the process of label elimination is complex due to the overlapping location errors of stereo and RFID and therefore label elimination is not possible in *detect*, *identify* and *locate* passes. During the *smooth* pass, RFID provides no label elimination information showing all four labels '▲','♦','●','■' as valid; however, the system identifies '●','■' and '▲','♦' as possible label set pairs based on object height and velocity constraint and subsequently outputs two compatible labels.



(d)

Figure 6.4

Label matrix updating steps for same colored objects at each time frame for Figure 6.3(a) tracks.

The compatible labels from relaxation are fed to the post optimization process to identify the optimal label for each observation. It is to be noted here that the system keeps one extra label as

part of the possible compatible label set. This explains a tradeoff between increased post processing computation for keeping a wrong label and the cost of eliminating a correct label. Since the objects have the same color and are assumed to be moving with the same velocity, at *t=6* the color and near neighbor constraints will not provide valuable information for label elimination. In the *smooth* pass based on height and direction of flow relative to previous velocity vector direction, CV identifies '▲','●', and '♦','■' as compatible label sets for the respective observations. These two label pairs for each observation represent correct trajectories of the balls; but incorrect trajectories of the persons as shown in Figure 6.3(b). However RFID provides '♦','●', and '▲','■' as possible label pairs. Correlating this information helps obtain one correct compatible label for each observation.

6.3 Optimization process

An optimization process is applied on the compatible labels obtained after relaxation labeling so that the tracks can be optimized. We explain here how smoothness, which is a global operator, can help optimize object tracks. The process only optimizes some token parameters without violating specified constraints. Smoothness of trajectories requires a burst of time frames. Our tracking version used a burst of four consecutive time frames to compute track smoothness, curvature, and acceleration.

6.3.1 Smoothness of trajectories

Our smoothing algorithm is motivated by the Sethi-Jain [96] and Veenman *et al.* [97] algorithms. It can work in either 3D or 2D and includes more information on some objects at some time instants [as is available from the RFID or vision sensors] than those previous algorithms. If a 2D algorithm is used, the constraint that two objects cannot be in the same location at the same time should be relaxed since it may just be that one object is occluding the other at some instant. The general algorithm will have different specializations depending on the application and how much sensor information and object constraints are available. For example, in the S-3 system, our cameras are calibrated to a surveyed 3D terrain, so if an image object is known, then an approximate 3D object location can be computed using the image from a single calibrated camera (we can just intersect a camera ray, or cone, with a surface in the 3D space).

Our goal is to optimize the tracking based on the heuristics explained in Section 3.4, which will provide the means for safer activities and more efficient site management. A set of T vectors of information for each time frame $t=1,2,3,\ldots T$ is provided as input. Each of these time frames represent "frame vectors" that contain N tuples $<x,y,z,v,L>$, where label L may identify a known object $(L=1,2,\ldots,q \text{ or } N)$ or it may be unknown $(L=0)$. The purpose of the algorithm is to assign (discover) labels $L = 1,2,3 \ldots q \text{ or } N,$ to each position tuple at each time t.

6.3.2 Tracking algorithm description

The algorithm takes input tokens containing *N* observations of 3D points over *T* time instants, thus *NT* tuples total are grouped into *T* time frames. It extracts a smoothest set of paths through these points, observed in frames *1* to *T*; all tuples are now grouped into *N* tracks. The object ID and location provide labeling information with the 3D points when available *i.e L = 1,2,3,……,N.* The number of such "tagged points" must be less than or equal to *T* for any of the objects *n*. For object track *n*, the trajectory consists of 3D points at each time frame *t=1,2,3, ……..,T*. The trajectory with label *L* is represented as:

$$C_L = [P_{n1}, P_{n2}, ..., P_{nT}] \;;\; P_{n,t} = < x, y, z, t, v, L > \tag{5.3}$$

As in Sethi-Jain [96], the path difference between two consecutive 3D points is defined as:

$$D_{n,t} = P_{n,i} - P_{n,j}; i \neq j \in t \tag{5.4}$$

Smoothness at a current point $P_{n,t}$ is calculated using the previous point $P_{n,t-1}$ and next point $P_{n,t+1}$. $D_{n,t-1}$ is the path difference between the current and previous point and $D_{n,t+1}$ is the path difference between current and future point. Smoothness value $S_{n,t}$ of a 3D point is then defined as follows:

$$S_{n,t} = w\left(\frac{D_{n,t-1}.D_{n,t+1}}{|D_{n,t-1}||D_{n,t+1}|}\right) + (1-w)\left(\frac{2\sqrt{D_{n,t-1}.D_{n,t+1}}}{|D_{n,t-1}||D_{n,t+1}|}\right) \quad (5.5)$$

To yield $0 < S_{n,t} \leq 1$ a weight factor w is used such that $0 < w \leq 1$. The initial points of N object tracks are assigned arbitrarily. The total sum of smoothness over T time frames for a single object track n with assigned label L is then given as:

$$S^{L}_{Total} = \sum_{t=2}^{T-1} S_{n,t} \quad (5.6)$$

One can then define total smoothness for all tracks by summing smoothness over all N tracks. For efficient implementation, the algorithm uses a *burst* or *block set* concept. A real time algorithm must make decisions within, say, a fifth of a second, or six video frames. This limits the amount of look ahead that can be used. A *burst* or *block set* denoted by B is defined as a sequence of N-tuples $<x,y,z,t,v,L>$ for a fixed length of time frames m. The size of B is then $N \times m$. From RFID properties, it is reasonable to assume that object IDs and their respective locations persist or are absent for multiple frames. In reality, these bursts can have arbitrary length and start time; however, in our simulations and analysis we assume more regularity.

Step by step implementation of the algorithm is as follows:

Input: N tokens of 3D points over T time instants */*Each time frame having N*

number of rows is considered as a frame vector/*

Output: Smoothed trajectories C_N.

- At *t=1*, for all *N* object tracks, assign labels *n =1,2,3,……,N* arbitrarily to the frame vector. */* Initialize labeling */*
- Define burst length *m* */* m = 3,4,5,6 */*
- Assign *k = 0.* */* Initialize k */*

FOR *t = 2 : m-1 : T-1* */* Loop over T-1 time frames with increment of m-1 */*

- *k = k+1;* */* Increment k */*
- Label *L* may identify a known object *(L=1,2, …,q or N)* or it may be unknown *(L=0).* */* Availability of partial label information will reduce number of possible combinations */*
- *Optional* - Use nearest neighbor assignments for *t-1:t* frames with *N* tracks. */* If selected, this helps reduce combination volume in next step */*
- Consider *t-1:t-1+m* time frames with *N* object tracks and form frame vector block set B_k having c_N^k sub trajectories. */* Generate frame vector block set of length m starting t=1*/ */* If nearest neighbor assignment selected then consider t:t+m-1 time frames */*

- Compute all possible combinations $|U|$ of the elements of block set B_k.

/$|U|$ is $r \times m+1$ */*

/$|U|$ is $r \times m$ with nearest neighbor option */*

/ r is the product of the number of elements of N trajectories in block set B_k */*

FOR *j=1:r*

FOR *d=2:m*

/ d=2:m-1 for nearest neighbor option */*

- Calculate smoothness $S_{total}^{j,d}$ at every instance in *m-1* time frames for *r* combinations.

END

- Calculate total smoothness S_{total}^{r} over *m-1* time frames for each combination of *r*.

END

- Sort end total smoothness for each combination in descending order.
- While indexing, choose highest total smoothness of *N* pairs with combinations having different elements in each frame vector in an instance.

*/*Point $P_{n,k}$ in frame vector is assigned once to a trajectory in one time instance and cannot be*

reassigned elsewhere/*

- Exchange points and assign c_N^k smooth trajectories in B_k as a subset of final smoothed trajectories.
- Save c_N^k.
- Increment *t.*

END */*When last positive integer value of t is reached */*

- Correlate similar end points of c_N^k smoothed trajectories in B_k with similar initial points of c_N^{k+1} smoothed trajectories in B_{k+1}. */* Based on this similarity measure, rearrange the order/label of c_N^{k+1} smoothed trajectories */*
- Combine similar label subset trajectories from processed frame blocks and generate final smoothed trajectories C_N.

Figure 6.5 shows the block diagram of the smoothness algorithm. Tracking algorithms sometimes lose correct object tracks at ambiguous intersecting points of trajectories. Optimizing object tracks here using trajectory smoothness helps the system in such ambiguous areas to interpret correct trajectories. Real-time implementation is possible as availability of ID from RFID could reduce the search space up to *99%*.

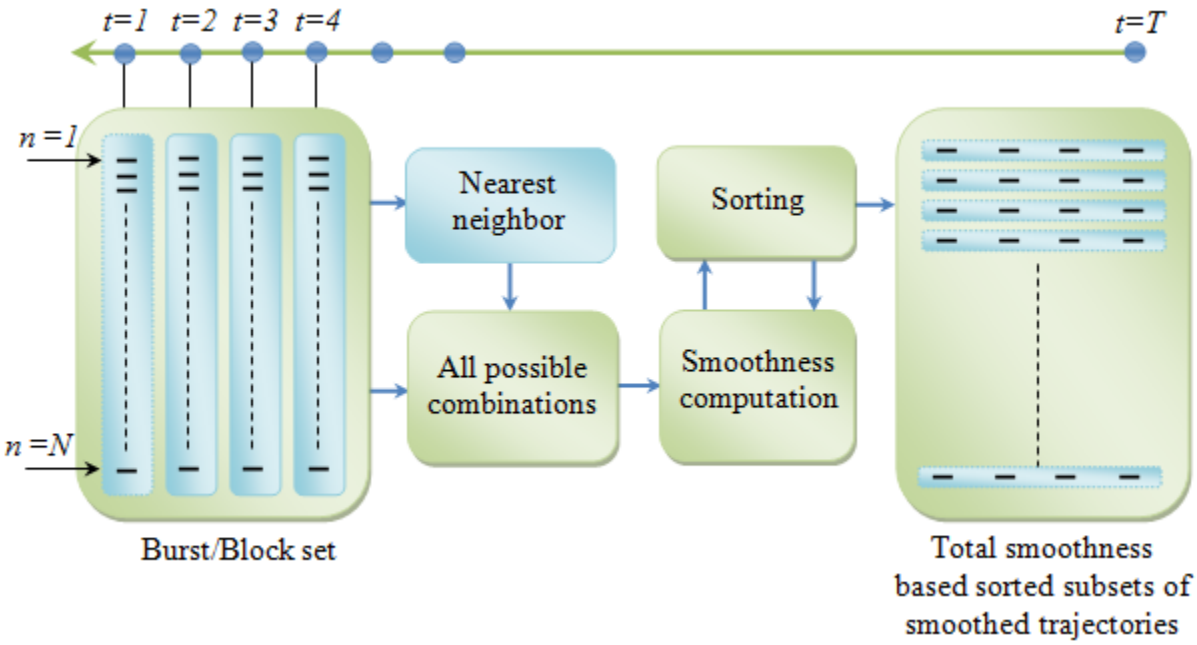
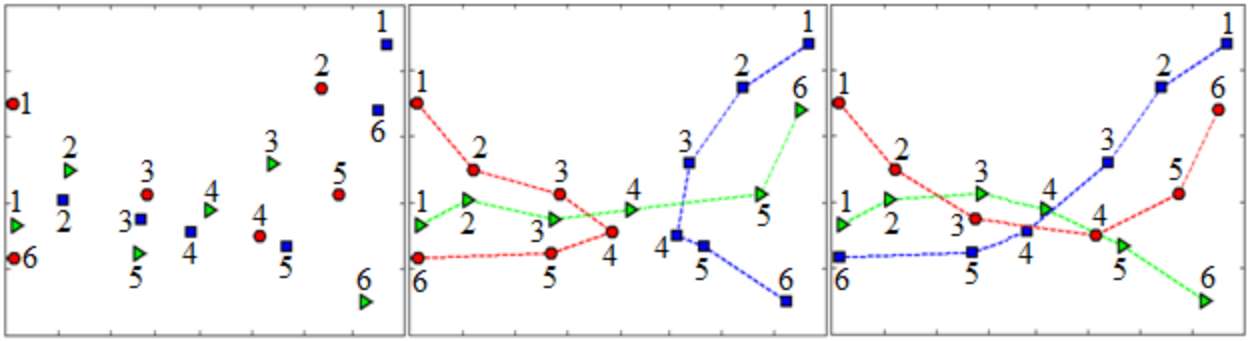


Figure 6.5

Block diagram of the smoothness algorithm.

To show the effectiveness of the smoothness algorithm we demonstrate step by step results of an example. The input data is displayed in Figure 6.6(a) and consists of a sequence of six time frames with three trajectories having 3D data at each point. Each point of the trajectories is symbolized by '■', '●' or '►'. For better visualization the *z* dimension of all the input data is fixed. Figure 6.6(b) shows the trajectory assignments after nearest neighbor linking. In the next step the exchange candidates are then decided using total smoothness. For this example block set B_k of length $m=6$ is used. Figure 6.6(c) shows the smoothed trajectories with '■' as label *1*, '●' as label *2* and '►' as label *3*. The algorithm took only *0.102 sec* with no assignment error.



(a) (b) (c)

Figure 6.6

Step by step results of an example with smoothness algorithm applied: (a) input data (b) nearest neighbor assignment (c) smoothed trajectories.

6.4 Summary discussion

We have proposed a three dimensional object tracking scheme using fusion of vision and RFID. Data integration and filtering are important tasks in tracking. We used discrete relaxation to control the integration of information from CV, RFID, and naïve physics. We have provided the theoretical and practical understanding of our proposed relaxation filtering technique that is based on the constraint satisfaction. The label elimination approach easily represents the ambiguity occurring in real-life applications. As a post processing step to labeling we have used smoothness for optimization to update computed tracks for increasing tracking reliability. We have shown how fusion can greatly increase tracking performance.

CHAPTER 7

Experiments, results and analysis

To study the value of fused sensor information in localizing and tracking multiple objects we report here experiments, results and analysis. Based on the defined performance metrics, analysis for CV and RFID as single and fused modalities is reported. First, we have evaluated use of stereo vision both indoors and outdoors for 3D accuracy and reliability of object location. Secondly, we have evaluated the commercially available RFID-based Real Time Location System (RTLS) for its accuracy and reliability of object detection and location. Finally, we have explored via simulations and also with the real time data how fusion can reduce the combinatorics of tracking.

To validate the localization and tracking approach provided in Chapters 4 and 6 we provide in this chapter detailed experiments, results and analysis. For stereo experiments we have used *Logitech C210* fixed focus commodity cameras. The focal length of these cameras is *~4mm* with a frame rate of *15 fps* at *640×480* resolution. For RFID based experiments we have used the CSL RTLS kit for localization and tracking. The computing platform used to run the simulations is a *core i5 580M* with *4GB RAM*. All the simulations were done using *MATLAB® 2009a*.

7.1 Generating stereo trajectories

To evaluate use of stereo vision both indoors and outdoors for 3D accuracy and reliability of object location we have generated three types of trajectories. The first type of stereo data was extracted from an indoor lab bench using stereo vision. The second type of stereo data was generated artificially using mathematical curves and the third type was the real stereo trajectories obtained indoors and outdoors. Simulations allowed us to construct interesting test cases and to control the ground truth, however, gaps between simulation and real-time may occur with respect to the assumption about the sensor capabilities, natural environment and the tracked objects attributes and dynamics.

7.1.1 Real indoor trajectories from wireframe workspace

To generate real trajectories indoors, we used 3D stereo. A colored sphere on a stick was moved by hand along a specified trajectory within a wireframe workspace of $27 \times 31.75 \times 24$ *in* $(68.6 \times 80.6 \times 61$ *cm)*. The structure was used for calibrating the cameras. The trajectory of the sphere yielded T records $<x,y,z,t,L>$ for object track L at times $1,2, \ldots, T$. The experimenter then repeated this using the stereo system to generate more trajectories until there were N of them, one for each object moving in the workspace: $L = 1,2,3 \ldots N$. Each of these N sequences was an "object track". If we had N object tracks, then there were 2^N subsets of these to choose for study. We had generated multiple tracks by varying the path and velocity through the workspace and also took care to create some near collisions. Figure 7.1 shows a set of a few trajectories generated using our stereo setup.

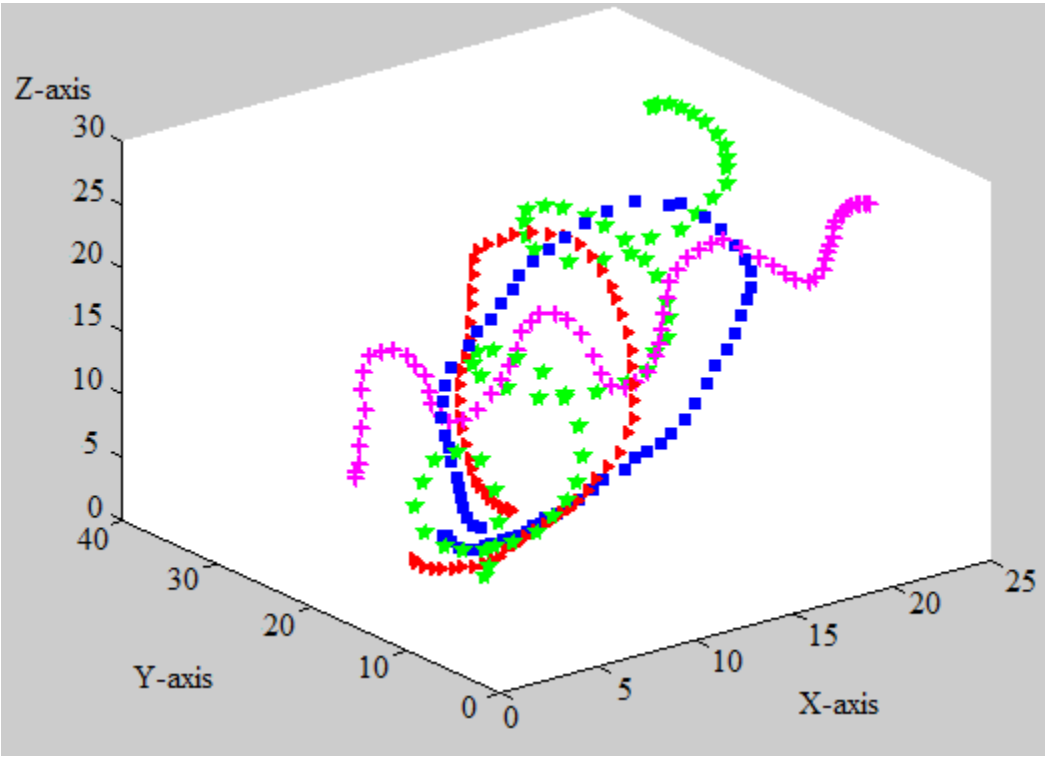


Figure 7.1

Example of stereo trajectories generated from wireframe workspace indoors.

7.1.2 Mathematical trajectories

We created a dataset generator that can randomly create smooth object tracks with various speeds and densities without collision. We generated N smooth paths for T time frames each in 3D space using a helix structure, which was randomly spread out for a selected number of time frames using pseudorandom values as shown in Figure 7.2. The circular helix of radius a and pitch $2\pi b$ in 3D space can be parameterized with Cartesian coordinates as follows:

$$x(t) = a\cos(t)$$

$$y(t) = a\sin(t)$$

$$z(t) = bt$$

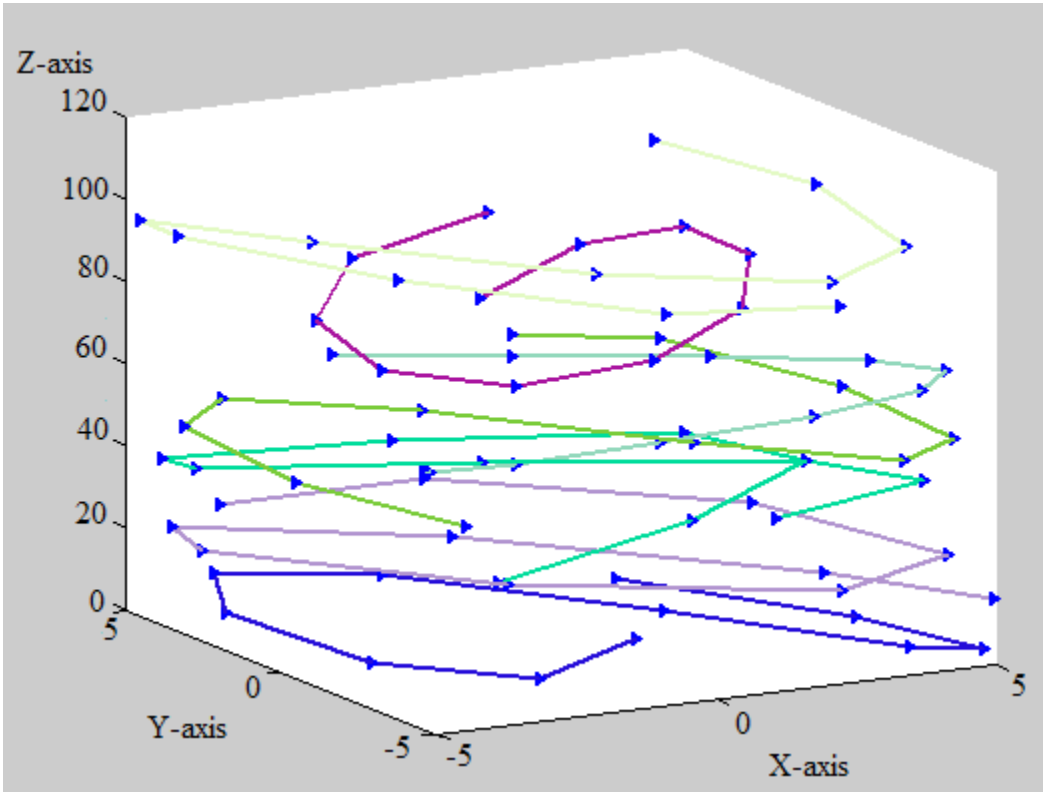


Figure 7.2

Mathematically generated trajectories using dataset generator with T=11 and N=7.

To meet the constraints in Section 3.4, the generated data has the following parameters by default:

a. Object tracks *N=10*.

b. Time frames *T =11*.

c. Smooth velocity vectors.

d. Unique trajectory directions.

e. No chance of collision.

f. Randomly spread out trajectories in a 3D space of *1* m^3.

7.1.3 Real stereo trajectories from indoors lab area

We generated several trajectories using colored hats i.e orange and yellow in our lab area. We acquired random as well as ground truth trajectories where persons wearing colored hats moved over specified paths in real-time or while stopping at known points for few seconds. Figure 7.3 shows the trajectory generated while the person moved on a predefined path.

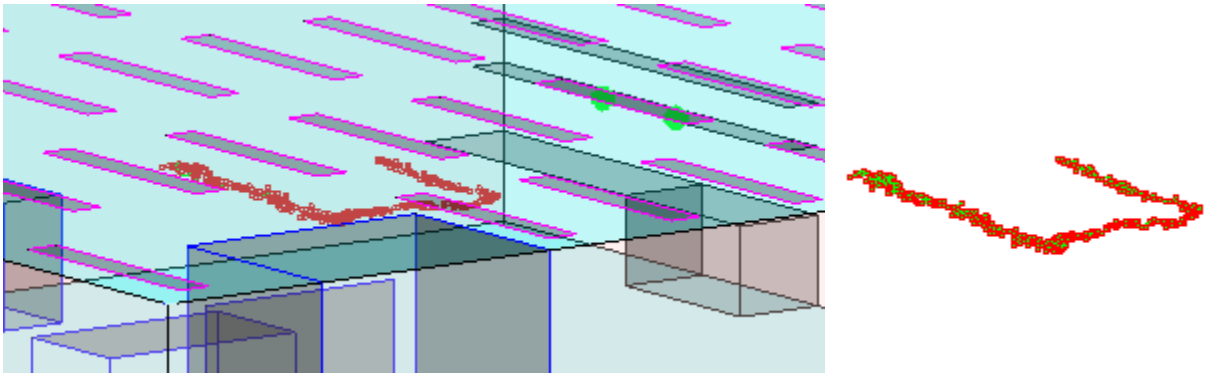
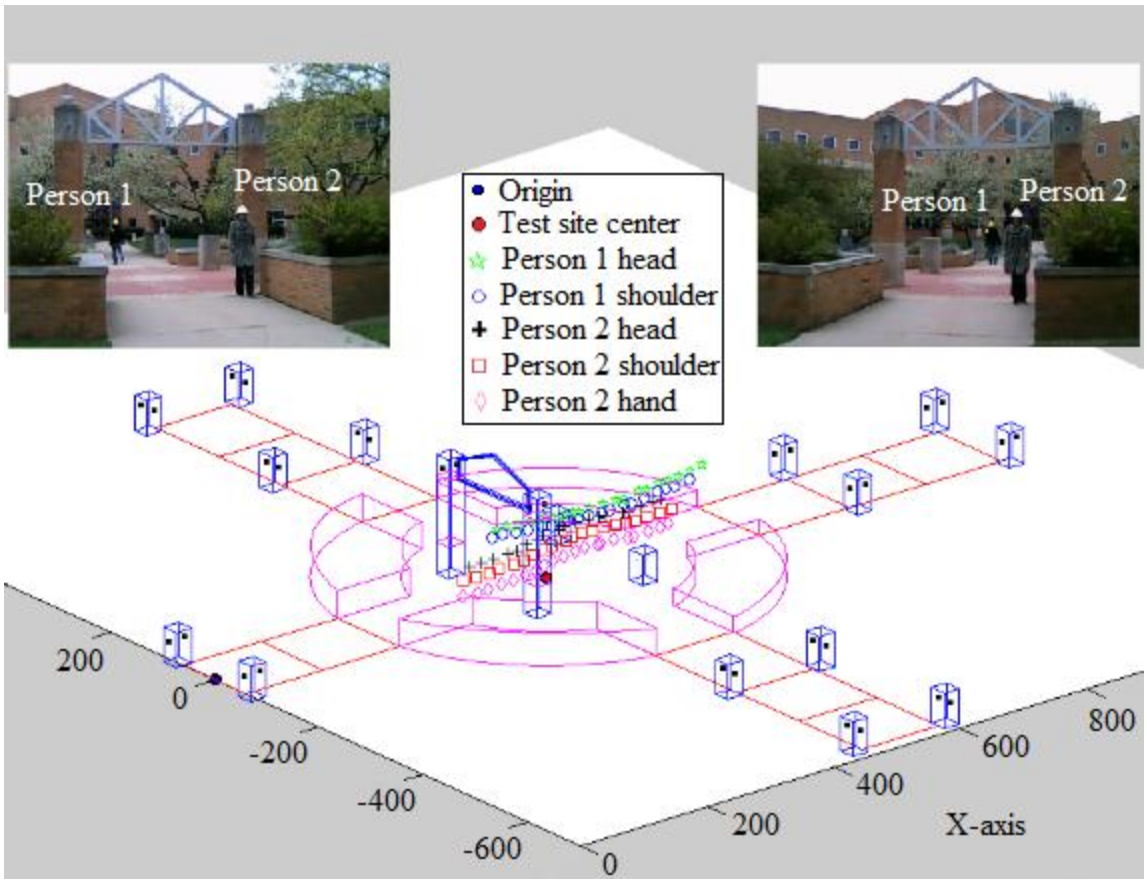


Figure 7.3

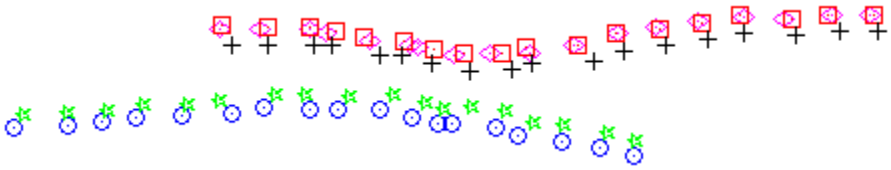
Indoors lab area real stereo trajectory where a person moved over predefined points.

7.1.4 Real stereo trajectories from outdoors

We conducted several experiments to test the stereo setup outdoors. The persons moved over predefined measured points without stopping or, in some runs the 3D data was obtained with the persons being stationary for few seconds at those points. We tracked head, shoulder and hands of two persons. The ground truth points were carefully generated so that the persons were not occluded by each other or by the background. The tracker solves stereo correspondence and generates object tracks automatically. With cheap commodity cameras placed *10 ft* apart and *4.5 ft* high we were able to track five same color coded points in real time with *6 fps*. The RMS error was within *7.6 in (19.3 cm)*. Some of the real trajectories where the persons moved continuously are shown in Figure 7.4.



(a)



(b)

Figure 7.4

Five outdoor real stereo trajectories: (a) 3D display of site (b) zoomed in top view of trajectories.

The system was able to track and distinguish between the head, shoulder and hands of the persons during the run time. This was achievable by applying epipolar geometry along with the

threshold defined for the shortest line segment constraint while solving the correspondence problem in 2D.

7.2 Real outdoor trajectories using RTLS

Using RTLS outdoors we also generated data sets of ground truth trajectories for the tagged persons and objects. The *z* dimension requires user interaction for defining the tag height that is helpful in estimating tag location. The readings were taken while the persons moved in real time and didn't change tag height. We also placed some stationary reference tags to help analyze RTLS location performance. Figure 7.5 shows sample trajectories of a tagged person. The CSL software generates trajectories in XML format which are then imported to MATLAB. The XML file contains date, time and location information for each observation. Below is a sample XML parent-child node structure generated by the RTLS. The parent node is accessed by the tag name. The child nodes under the parent tree then contains the desired location information for each time instance.

<id>EE4CBB6A6223</id>
<name>EE4CBB6A6223</name>
- <position_list>
  - <Position>
    <x>14.504</x>
    <y>9.519</y>
    <z>48</z>

<time>2012-07-30T19:22:51.7626037-04:00</time>

</Position>

</position_list>

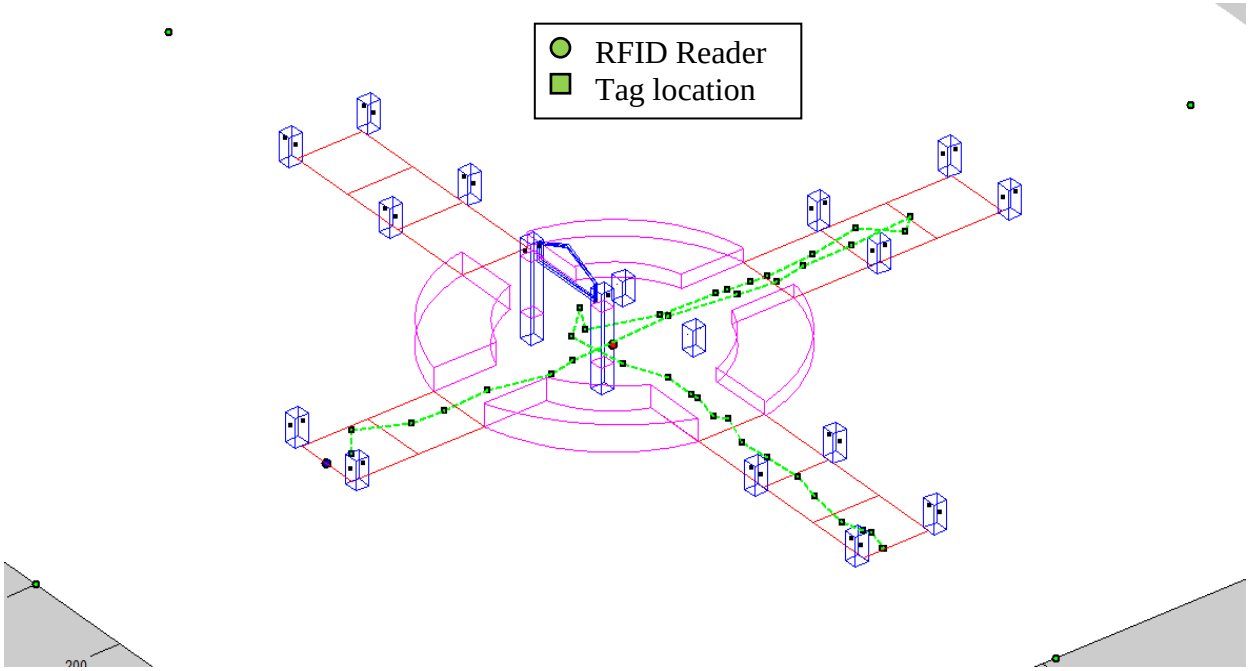


Figure 7.5

Outdoors 2D RTLS trajectories of a tagged person (green paths).

7.3 Metrics for evaluation of performance

In this section we provide information on performance evaluation metrics used. We have evaluated fusion of CV and RFID as well as their effectiveness as single modalities for localization and tracking. Following are the details of the evaluation metrics used:

a. *Location accuracy* - The location accuracy is defined as the difference between location observations obtained by sensing and the corresponding ground truth data. We have primarily used statistical measures such as root mean square (RMS) error to express the location error in *x*,*y* and *z* direction.

b. *Least squares error* - We have used least squares and polynomial fitting schemes for evaluating the error in the real data obtained in our experiments for which acquiring ground truth data was complex. We have also utilized a piecewise line fitting scheme to analyze the RTLS trajectories to have more meaningful results.

c. *Probability and percentage of observation availability* - The RTLS readers' communication with the tags varies from one place to another. Also the system refresh rate changes with different number of tags present in the test site. We have measured probability of tag location-signal availability in respective runs. This information is useful in analyzing the reduction in the combination search space of the tracking algorithm.

For real-time tracking the object recognition and tracking needs to be automated. There is a fair chance of missed observations due to lack of object detection/identification, object exiting the field of view of the sensor, or occlusion caused by another object or background. We express missed observation quantity in terms of percentage.

d. *Track error* - We assessed performance of tracking using smoothness constraints in terms of track error. Track error is defined as a fraction of wrong trajectory point assignments by the tracking algorithm. At present we have assumed that the objects do not enter or exit; therefore a pair of wrong label assignments between points $P_{j,t}$ and $P_{k,t}$ (where $j \neq k$) is considered as one error. The final track error is then averaged over the number of simulation runs. Alternatively, track error can be more fairly defined in terms of point sensing tolerance; an object label assigned to a sensed point is considered correct if the sensed point is within measurement tolerance of the ground truth sensed point. This alternative does not penalize switching the labels of observations that are very close together in space.

We assessed performance of tracking using track error for different burst/block set lengths *'m'* defined in Section 6.3.2. Varying the burst length and the density of points over time affects the track error performance. Increasing the block length decreases the error but increases the number of combinations. Fusing ID information from another source such as RTLS helps reduce this combination space.

e. *Similarity of object color* - For studying color based object detection outdoors we have correlated histograms of the segmented objects under different illumination conditions, i.e sun and shade. For comparing these we have used Euclidean distance to characterize color histogram variations.

7.4 Real-time tracking performance: indoor with RFID feed simulated using color

The stereo approach was tested in the real time indoor environment while tracking two balls in the wireframe workspace of *27×31.75×24 in (68.6×80.6×61 cm).* Since it is impossible for us to gather the number of cases needed using real data, RFID is simulated using same and different color (red and blue) -- in extracting trajectories, object location and ID are sometimes randomly provided to the tracking algorithm for some of the observations where possible. The tracking algorithm was applied to the observations to segment them into separate object trajectories. Later, to assess the performance real-time trajectories were then compared with ground truth data were available. The tracks were generated and displayed in parallel. We show that recognition of some objects during some time intervals can greatly speed up and make more reliable the organization of time frame information into the tracks of separate objects. For one of our tests the system took *58.6 sec* to acquire *1000* frames from both cameras while executing and displaying the input and the output. The tracking algorithm had less than *1.4%* missed observations with zero track error when both the colored balls were moving and being tracked. It was established that the proposed approach has an ability to track the object while generating its tracks on the fly.

7.5 Indoor stereo live demo results

To analyze the stereo live demo results visually we generated trajectories that can be interpreted easily. Details of the stereo tracking live demo are provided below and shown in Figure 7.6.

a. *Blue object kept stationary while red object is moving.*

The red cursive writing sample in the *yz-plane* while a red smiling face in the *xz-plane* is the red ball trajectory. The blue object was stationary and is shown by blue dots.

b. Blue and red object moving.

The heart shapes in blue and red are the respective trajectories of blue and red balls being tracked. The shapes were made in the *yz-plane* and its view from the *xz-plane* is shown for better understanding.

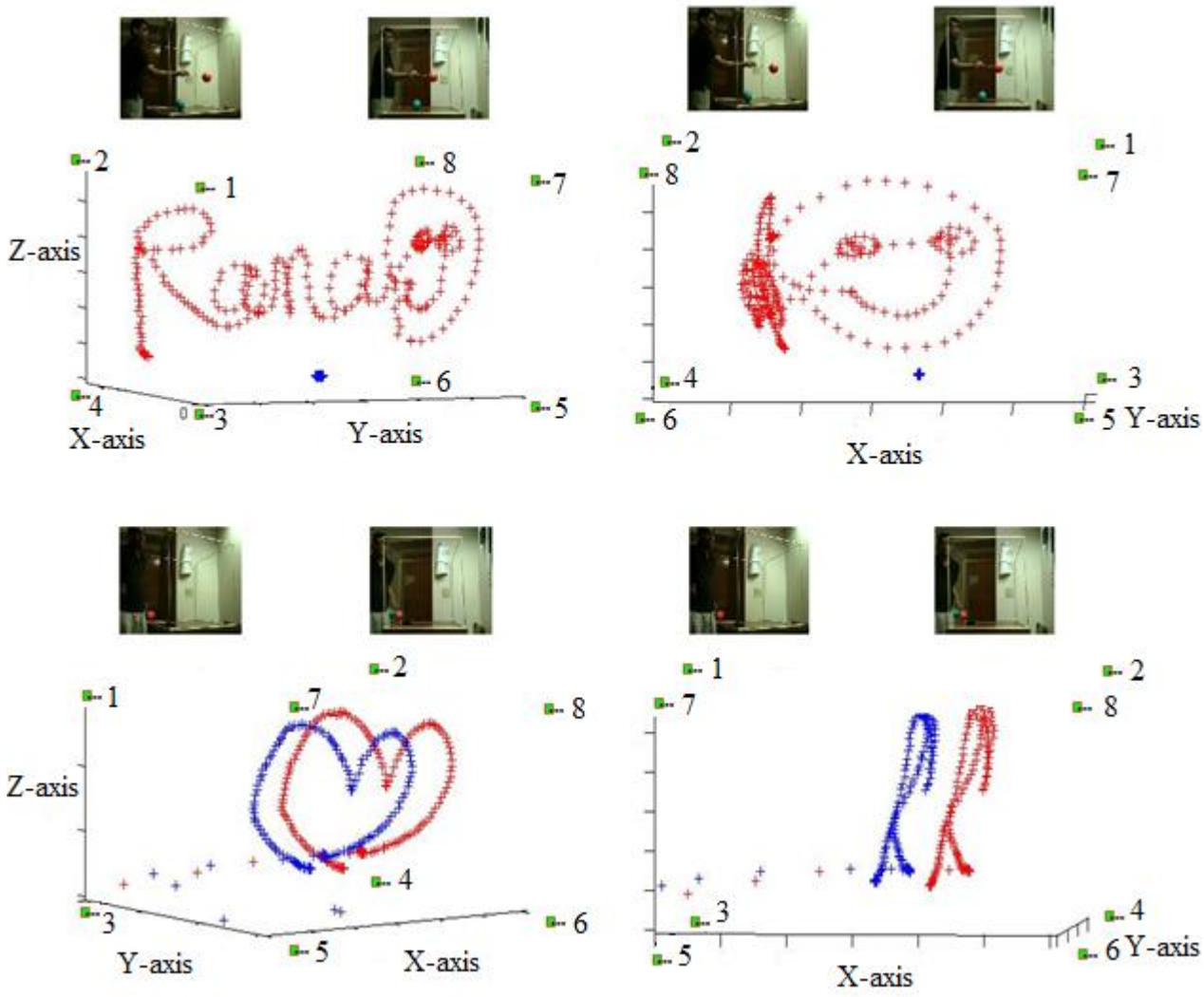
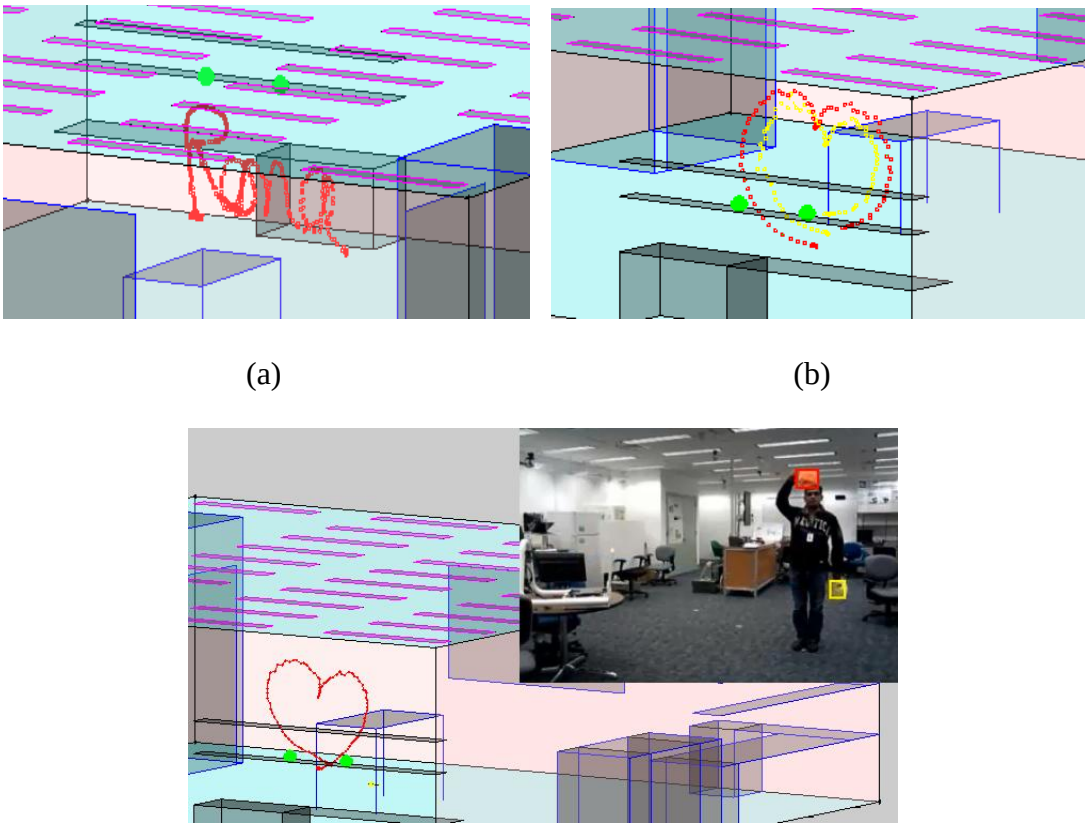


Figure 7.6

3D stereo tracking in wireframe workspace indoors - live demo.

Next the system was tested at room scale and real-time trajectories were generated in the lab *280×476×105 in (7.1×12.1×2.7 m)* using one (orange) and two color (orange and yellow) combination. As above for better visual analysis Figure 7.7(a) and (b) shows cursive text and heart shape trajectories using single and two colors tracking respectively. Figure 7.7(c) shows the

two color detection where yellow color is used as an initializing marker for tracking orange. The image also shows 2D bounding boxes over colored hats.



(a) (b)

(c)

Figure 7.7

3D stereo tracking in lab area indoors - live demo: (a) orange cursive writing sample (b) heart shapes using orange and yellow color (c) yellow used as initializing marker to track orange.

7.6 Stereo error analysis in x,y,z dimensions versus distance from the cameras

Slight inaccuracies in selecting the calibration points in the image generates error in the camera matrix while error in image point location of objects will yield error in the imaging rays

used in stereo computations. Also factors such as lens distortion, lighting variation, digitization noise, and object surface variation contribute to image point error and hence the stereo error. The stereo error in *x,y,*z dimensions is observed by computing 3D location for six selected ground truth points not chosen for the calibration procedure. The 3D points were selected on the basis of their distances from the cameras. The run was repeated eight times. Figure 7.8 shows the selected 2D corresponding points in the image pair.

Figure 7.8

Selected 2D corresponding points in left and right camera image for analyzing outdoor stereo error.

We have used RMS error to represent the stereo location accuracy in all the three dimensions. Table 7.1 tabulates the 3D ground truth and the 3D computed data from eight runs.

Table 7.1

3D ground truth and computed data for analyzing outdoor stereo error - scale is in inches.

	3D ground truth	*Computed 3D data from eight runs*								*RMS error*
X_1	*242*	*237.9*	*237.9*	*238.6*	*237.1*	*237.9*	*237.9*	*237.9*	*238.6*	*4.06*
Y_1	*60*	*61.2*	*61.2*	*61.5*	*61.2*	*61.2*	*61.2*	*61.2*	*61.5*	*0.17*
Z_1	*36*	*31.6*	*31.3*	*31.3*	*31.7*	*31.6*	*31.6*	*31.3*	*31.3*	*0.19*
X_2	*357*	*355.5*	*357.1*	*357.1*	*356.9*	*356.9*	*357.1*	*358.5*	*358.5*	*0.92*
Y_2	*60*	*60*	*60*	*60*	*60.5*	*60.5*	*60*	*60.5*	*60.5*	*0.33*
Z_2	*36*	*36.7*	*36.3*	*36.6*	*36.6*	*36.6*	*36.6*	*36.6*	*36.6*	*0.59*
X_3	*429*	*434*	*433.9*	*433.9*	*431.7*	*436*	*433.7*	*431.7*	*433.8*	*4.77*
Y_3	*72*	*73.6*	*73.6*	*73.6*	*73.6*	*74.2*	*74.2*	*73.6*	*74.2*	*1.84*
Z_3	*41*	*43.5*	*43.9*	*43.7*	*44*	*43.5*	*43.5*	*43.5*	*43.3*	*2.63*
X_4	*540*	*542.2*	*542.1*	*538.9*	*545.3*	*542.2*	*545.4*	*542.1*	*538.8*	*3.13*
Y_4	*0*	*1*	*1.8*	*0.6*	*1.4*	*1*	*1.4*	*1*	*1.4*	*1.27*
Z_4	*0*	*2.4*	*2.5*	*2.9*	*2.4*	*2.4*	*2.2*	*2.1*	*2.4*	*2.43*
X_5	*953*	*960.4*	*965.4*	*951*	*960.8*	*960*	*959.6*	*951*	*960.4*	*7.28*
Y_5	*0*	*1.6*	*1.5*	*1.5*	*1.5*	*0.9*	*0.9*	*1.5*	*1.5*	*1.39*
Z_5	*0*	*2*	*3*	*3.9*	*2.6*	*2.9*	*2*	*3.9*	*3*	*2.99*
X_6	*1190*	*1210*	*1194.8*	*1210.5*	*1194.9*	*1195.6*	*1196.1*	*1195.6*	*1194.6*	*11.09*
Y_6	*0*	*3.4*	*2.5*	*3.4*	*2.5*	*2.5*	*4.2*	*2.5*	*4.2*	*3.25*
Z_6	*0*	*4.8*	*6.5*	*4.3*	*6*	*4.8*	*3.2*	*4.8*	*6.1*	*5.17*

The RMS error in all three dimensions was within *7.6 in (19.3 cm)* for 3D points *20 ft* to *80 ft* distance from the cameras. The RMS error has nearly linear behavior in relation to the distance from the cameras and is shown graphically in Figure 7.9 and supports the error cone concept. The magnitude of error in *x* dimension is more compared to *y* and *z* dimension. The *x* dimension here is in line with camera viewing direction.

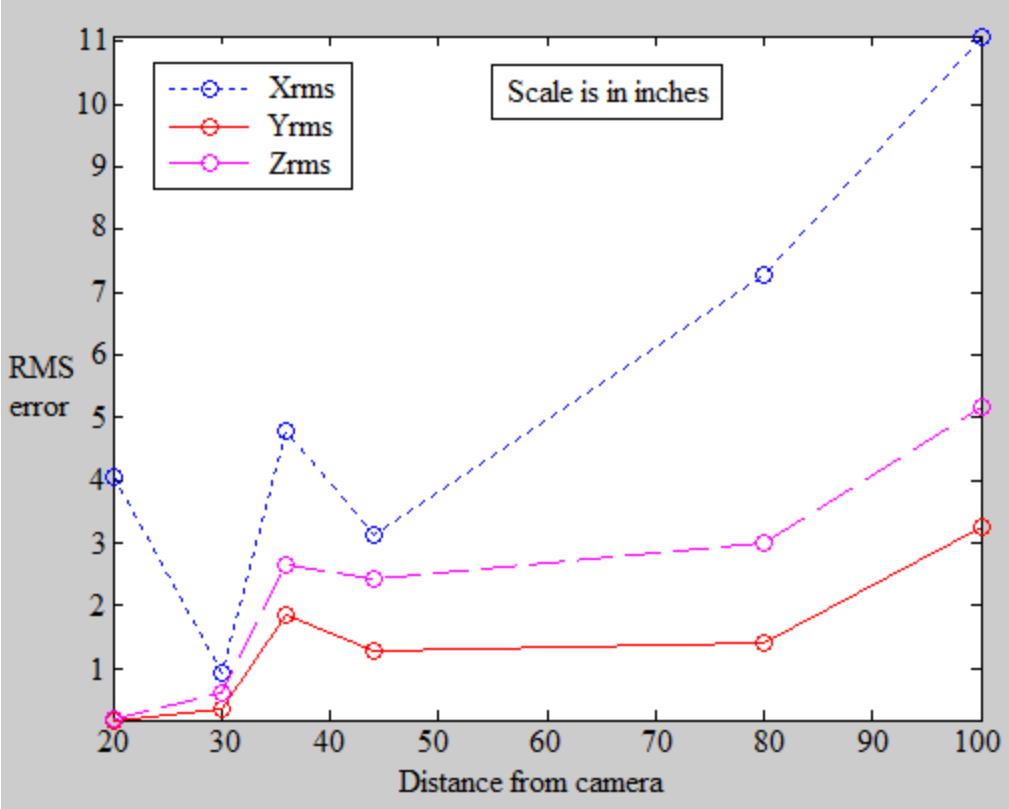


Figure 7.9

Stereo RMS error in x, y and z direction versus distance from the camera.

7.7 Least squares analysis on real outdoor stereo trajectories

Figure 7.10 shows some of the left camera frames that were used to compute the trajectory of a ball [tossed upward] using stereo alone. The ball is circled in the images for easy representation.

Figure 7.10

Left camera images showing projectile trajectory of a ball tossed upward.

The projectile motion of the five 3D points corresponding to the five images in Figure 7.10 is shown in Figure 7.11. The ball's trajectory can be seen in two different views. The 3D points are represented by '+' in green.

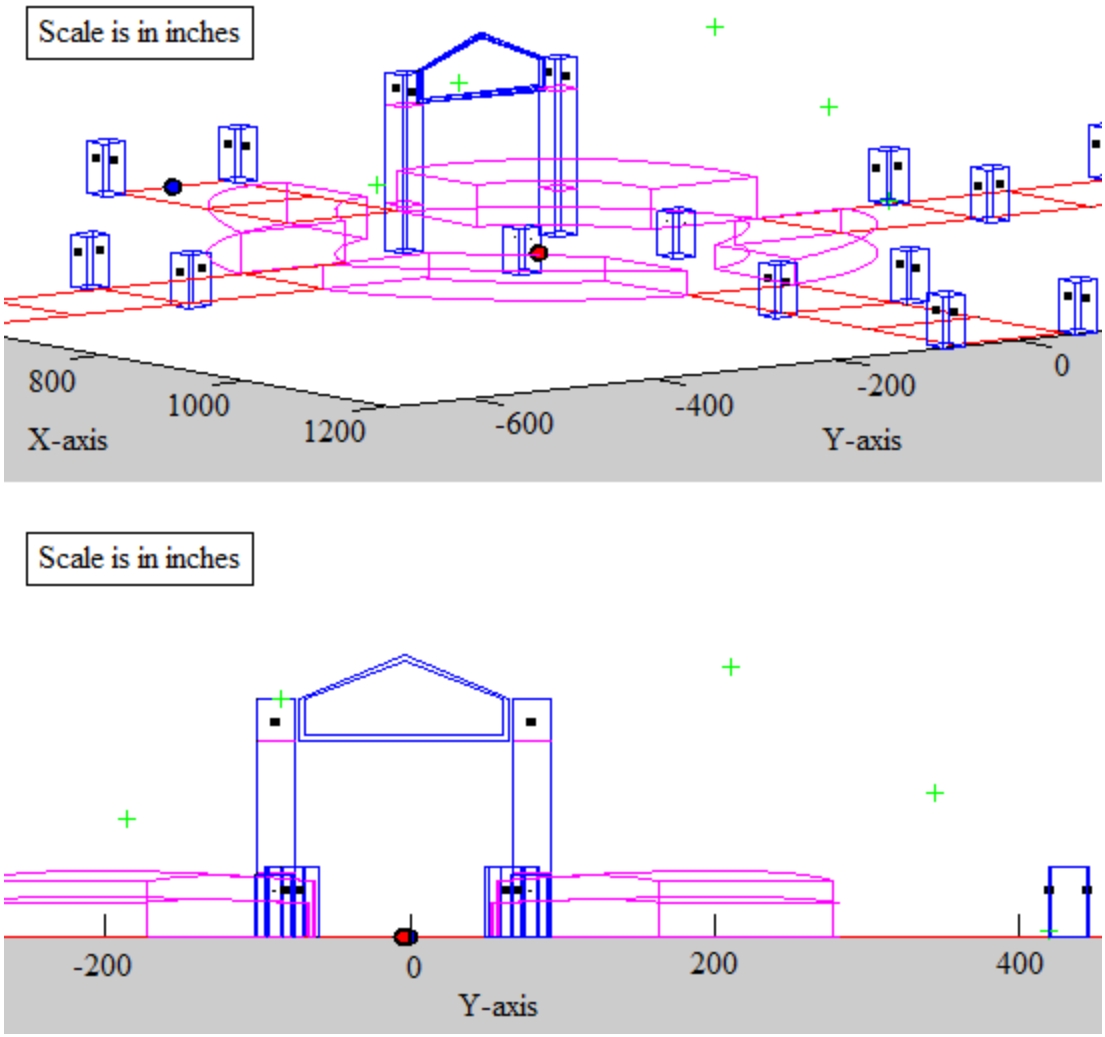


Figure 7.11

Two different views of 3D points showing ball projectile trajectory computed using stereo.

A solid (blue) line in Figure 7.12 shows the ball's computed trajectory in the *yz-plane*. To analyze stereo performance here we used least squares fitting. The dotted (magenta) line shows the parabolic curve fitted on the computed data. We took note of the estimated start [position 1 *~98 ft* from cameras] of the ball during the experiment which is different by *~15.3 in* from the one observed. This supports our analysis in the previous section about stereo RMS error

increasing linearly especially at a greater rate along the axis in line with camera viewing direction i.e *y-axis* in this case.

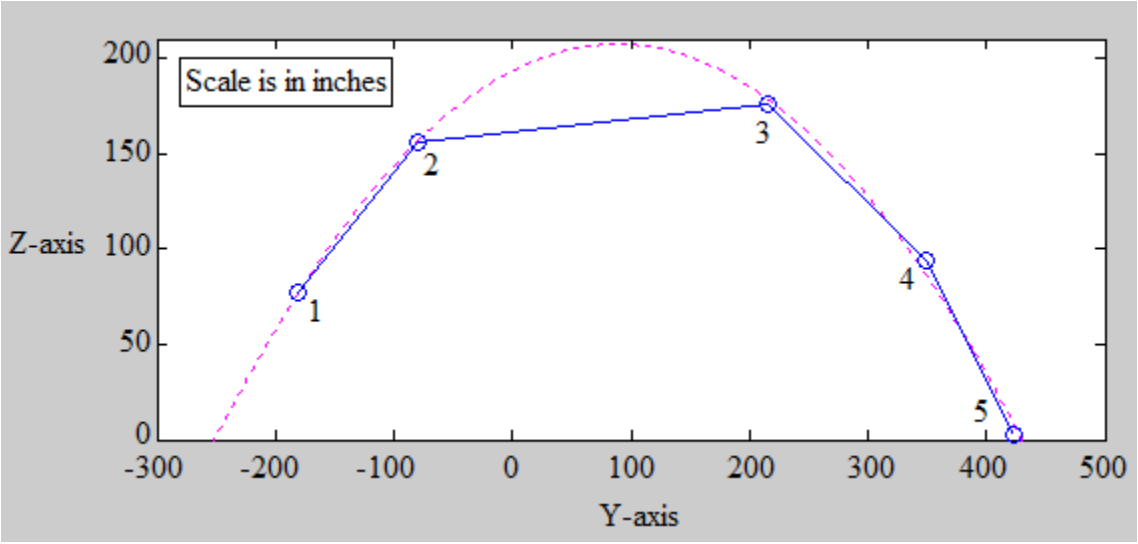


Figure 7.12

Computed ball projectile trajectory (solid blue) with parabolic fitting (dotted magenta).

We use three position estimates here that includes initial estimated thrower position, ball position in frame *3*, ball position while touching the ground. These points can define the projectile trajectory of the ball. Using a three parameter model the parabolic curve can be fitted to these three points as shown by solid (green) line in Figure 7.13.

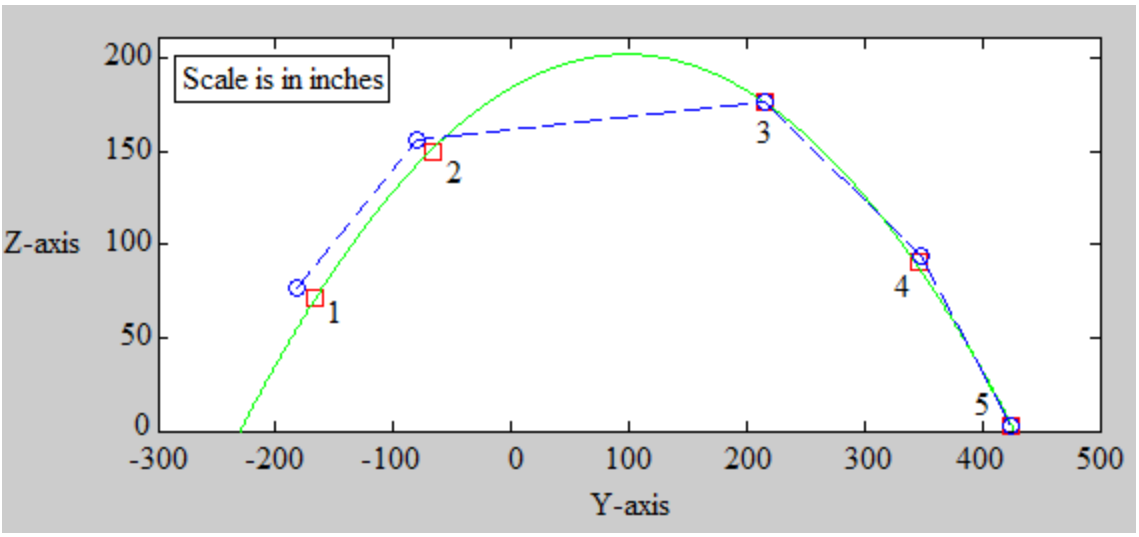


Figure 7.13

Actual trajectory (dashed blue) with linearly increasing error along y-axis and corrected trajectory (solid green) using parabolic curve fitting.

It is noted that the ball's trajectory has gradual increase in stereo error (right to left) along the *y-axis* away from the cameras and beyond the site center (*~84 ft* from cameras). Table 7.2 shows comparison of the observed and corrected [after curve fitting] observations. This started decreasing as the projectile crossed the site center towards the cameras. Modeling error estimates along *y-axis* with corrected curve fitting can help calculate intermediate trajectory points.

Table 7.2

Ball projectile observed using stereo and fitted data along yz-plane - scale is in inches.

Point	$Y_{observed}$	$Z_{observed}$	$Y_{fitting}$	$Z_{fitting}$	Y_{diff}	Z_{diff}
1	*-181.6*	*76.7*	*-166.3*	*71.5*	*15.3*	*5.2*
2	*-79.7*	*155.9*	*-67.1*	*149.4*	*12.6*	*6.5*
4	*348.4*	*93.9*	*345.8*	*90.2*	*2.6*	*3.7*

The problem of increasing stereo error was explained here to give an idea about the upper bound using *Logitech C210* cameras. This can be addressed in practice by using multiple cameras in the test site and naive physics constraints. The combined 3D observations from two or more stereo pairs can be smartly combined to cover an overlapping area. For example as soon as an object trajectory enters the *70+ ft* range from one stereo pair then another set of stereo pairs can be engaged for which the object is within the desired tracking range.

To analyze RTLS location performance we consider the tagged person trajectory (without stopping) given in Figure 7.5. Piecewise line fitting on the tag trajectory is shown in Figure 7.14. To find the least squares error, we have evaluated the sum of the squares of the differences between the line fit and the tag trajectory. The least squares error here is *22.1 in*.

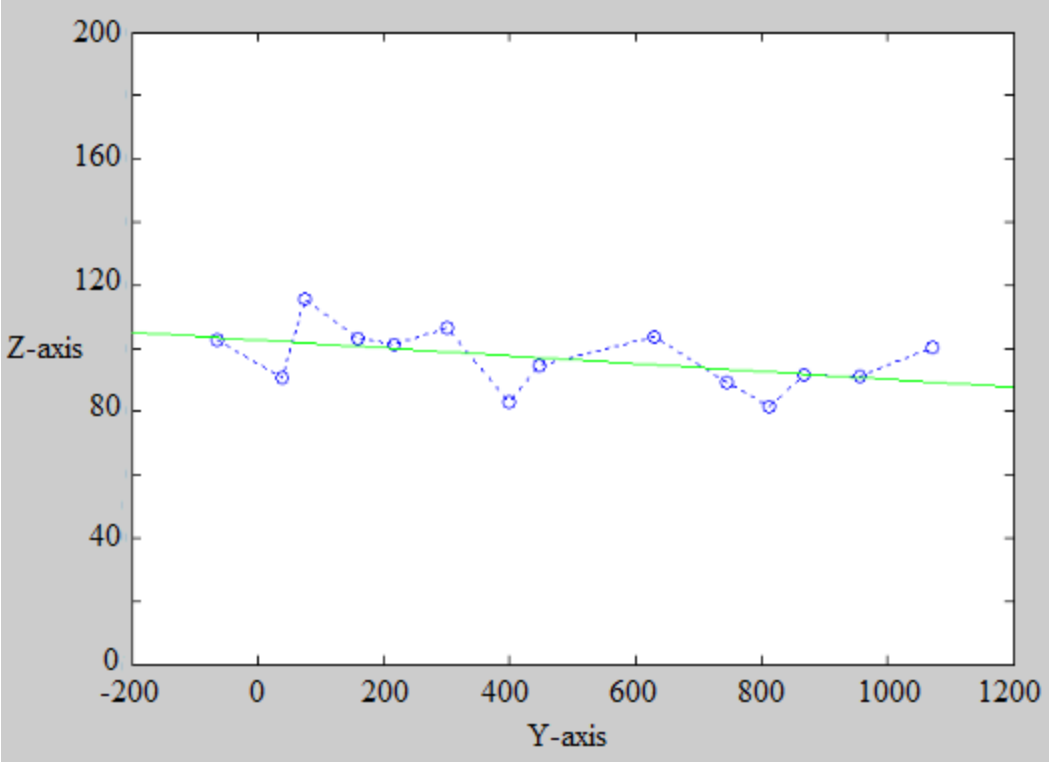


Figure 7.14

Piecewise line fitting (solid green) on RTLS tag trajectory (dotted blue).

Next we compared this RTLS tag linear trajectory [dashed green] with the ground truth [dotted red] for checking location accuracy. For reader understanding, the location error circles are drawn at every instant around the ground truth observations as shown in Figure 7.15. The radius of the circle represents the spatial difference between the ground truth and RTLS location observations. The tag observations are represented by '■' [green] and that of ground truth by '♦' [111].

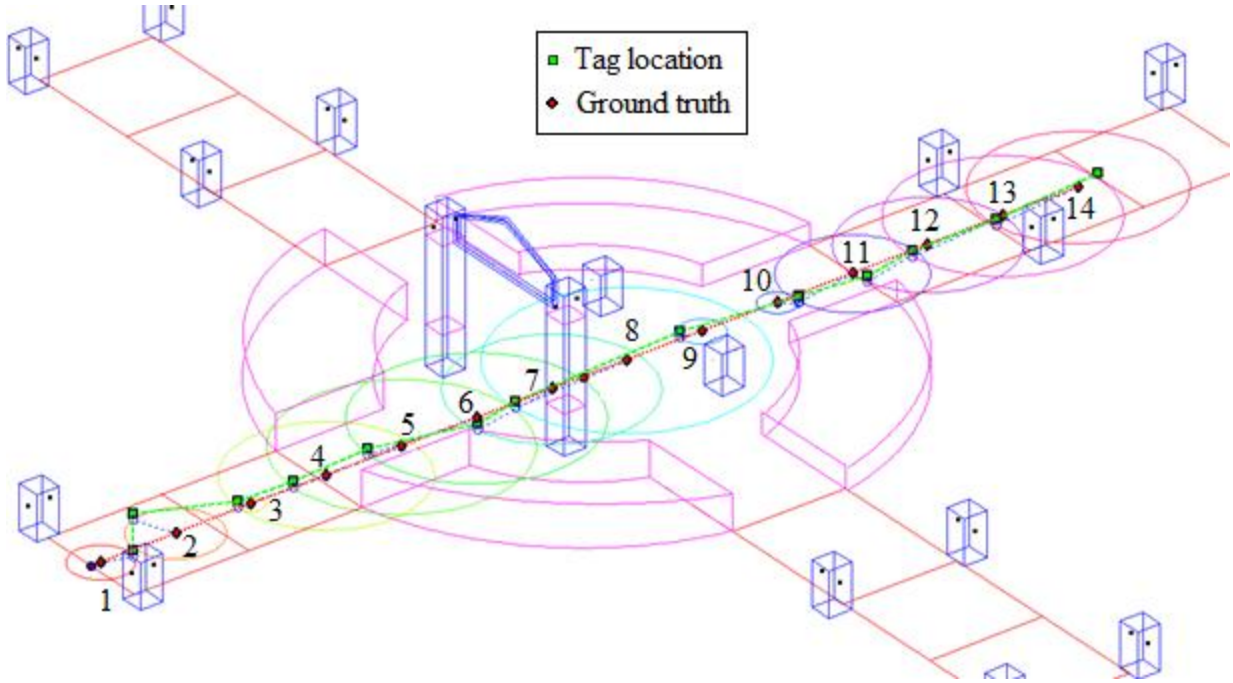


Figure 7.15

RTLS tag location error circles.

As shown in Table 7.3 the location error increases as the tag moves towards the arch structure. This is likely due to the multipath effect as our test site represents a semi-indoor environment. Similarly the location error decreases as the tag moves away. The RMS location error for the tag trajectory is *80.7 in* (*2.05 m*).

Table 7.3

Location error for RTLS tag trajectory in Figure 7.1 - scale is in inches.

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
29.4	*43.1*	*11.9*	*91.5*	*113.8*	*110.6*	*92.9*	*123.3*
9	*10*	*11*	*12*	*13*	*14*	*RMS*	
21.4	*17.9*	*65.6*	*80.7*	*102.2*	*94.6*	*80.7*	

7.8 RTLS signal availability

Having RFID feed available significantly reduces CV tasks for tagged object detection and identification. The RFID feed availability depends on the refresh cycle and the number of tags. Also, missing tag information is another key factor to be considered. Figure 7.16(a) shows the RTLS real trajectory of a single dynamic tag obtained when there were three other tags actively transmitting and present in the test site. The trajectory was designed to cover most of the test area. The RFID signal for a single tag was available on average after every *2.5 sec*.

Note that in Figure 7.16 there are some visible gaps between two consecutive tag locations, which represent missing observations. Missed observation means non-availability of location information when the RTLS signal is expected to be there. This can occur due to tag orientation, miss reads by the reader, or some direct occlusions which resulted in read failure. For quantitative analysis there were *~25* missed observations in addition to *136* times signal was availabe i.e $\left(\frac{25}{136+25}\times 100\right)$ *15.5%* missed observations.

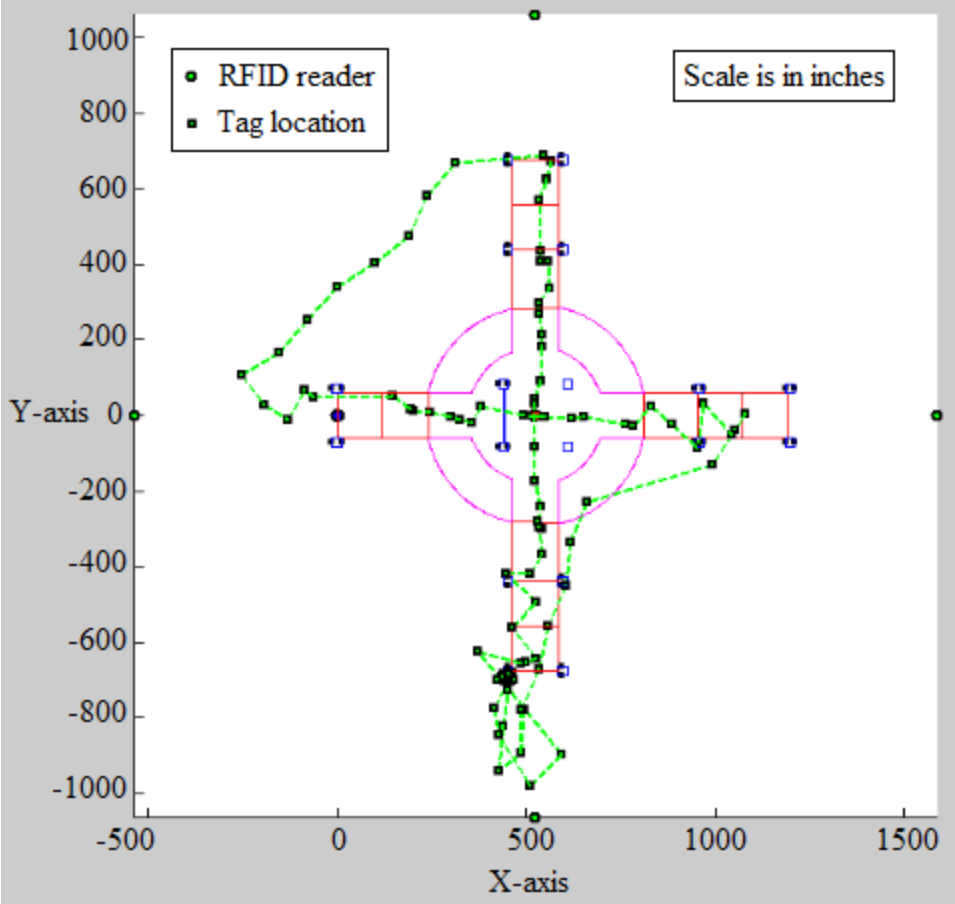


Figure 7.16

RTLS trajectory analysis: (a) RTLS single tag trajectory (green path) (b) RTLS tag location signal availability over time.

7.9 Simulations of object tracking

Prior to working on real fused outdoor data, we performed many simulations in order to assess how effective RFID labels could be in tracking under smoothness constraints – using observations of location but not color. For simulations we used ten subsets of real indoor stereo trajectories explained in Section 7.1.1. *N* observations over the time steps *1...T* were selected and presented to our tracking algorithm to see what tracks would be aggregated using the naïve

physics constraints. Smoothness of trajectories requires a burst of time frames for reliable results. Below, we have used burst length of four time frames to compute track smoothness, curvature, and acceleration.

If we consider n objects and a burst of m time frames then the number of possible paths will be $(n)^m$. Assume that T is divisible by m. If there is no ID information available then the number of combinations for T time frames will be $(T/m)\times(n)^m$. Depending upon the probability P for an observation ID being available for the burst, the combination volume may be reduced accordingly. It is considered that the ID when present is available for the whole burst. For example $n=3$, $m=4$ and $T = 60$ the total number of track combinations will be *1215*. Different frequency of ID availability across bursts will have different impact in reduction of combinations. As shown in Figure 7.17, for $P=0.267$, ID availability settings represented by a solid red line reduces the possible combinations to *435*; however for $P=0.267$ the dotted blue line setting reduces it to *631*. This shows that for the same probability, an object ID can be available in many configurations. In this case *435* also explains the upper bound in combination reduction with the lower bound ranging to *891*. Therefore, the more the RFID signal availability is spread across time, more volume reduction is possible.

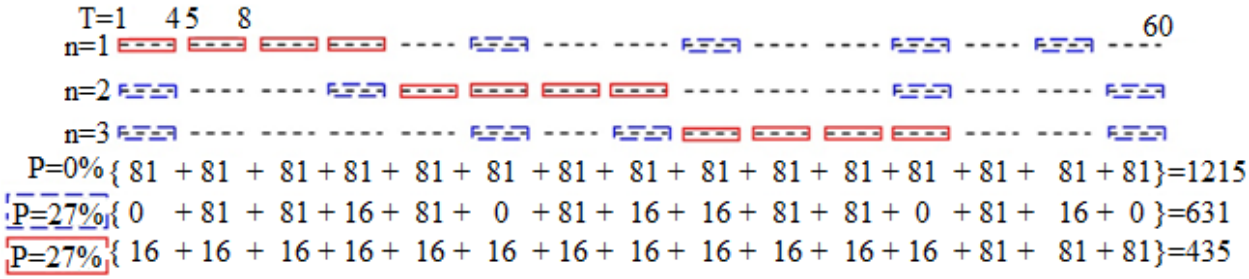


Figure 7.17

Reduction in combination volume - with probability of random ID information availability.

The simulations were done using *MATLAB®2009* on a *Core i5 M580 2.67 GHz* platform. Simulations were conducted for *N=5, 6* and *10* object tracks and *T=60* time steps. Using probability *P*, the ground truth ID was provided in the token. Figure 7.18 shows results for possible reduction in combination volume with increase in probability *P* of object ID being in the token. Computation time is also shown at marked places to realize the reduction in volume. With respect to our outdoor experiments, the probability *P* represents the time percentage for which the RFID feed for a tag was available for each object. The algorithm was run with frame burst length, *m=4*. ID was assumed to be randomly available [across bursts] over time steps *T*. Figure 7.18 shows that while tracking ten objects the combination volume can be decreased up to *99.9%* with the partial ID feed thereby significantly reducing computation time. The effect of having some ID in the tokens increases as the number of object tracks *N* increases. Also, object location and ID info increase the accuracy of calculated trajectories. This data shows the difficulty faced by tracking algorithms that only use motion of image points to aggregate object tracks. Without any object ID, quantifying motion over several time steps leads to too many possible tracks. Although color, shape and texture features can be used by a passive CV system, the reliability of

unique labels from RFID can yield correct tracks with far less computation. Thus we were motivated to implement an actual Site Safety System using fusion of CV and RFID.

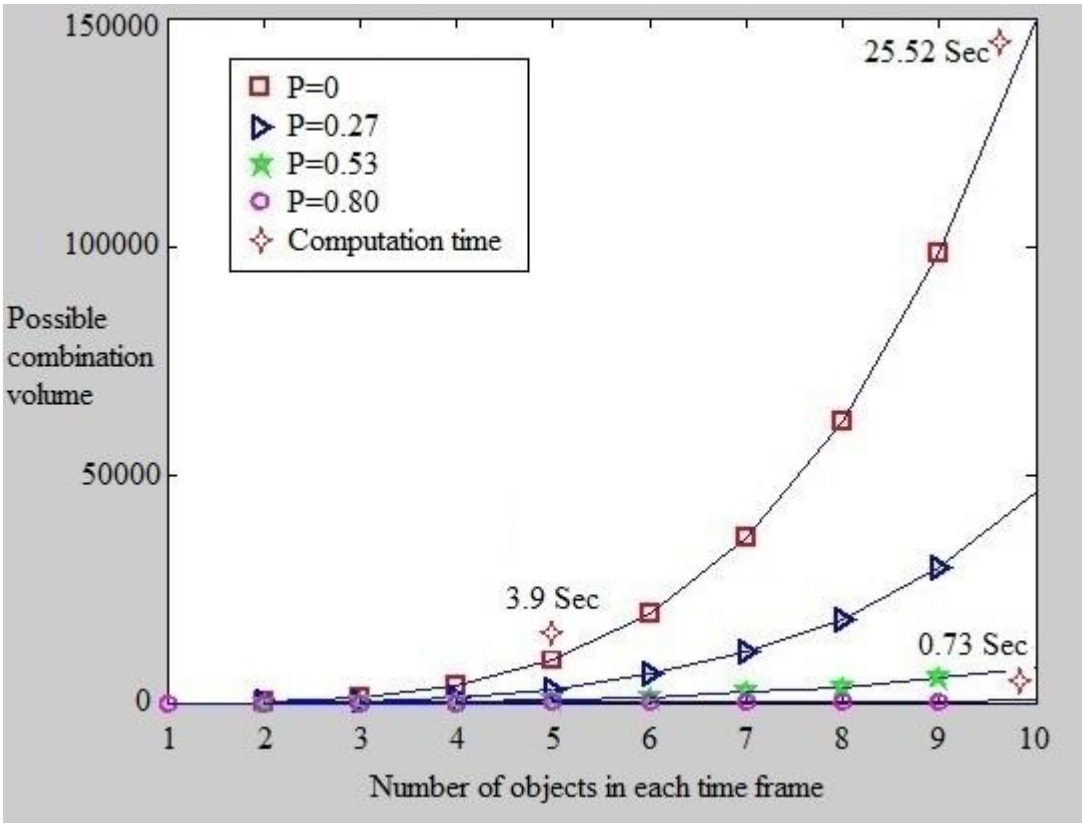


Figure 7.18

Possible combination volume with N objects and probability P of object ID in bursts of four tokens.

Table 7.4 shows the behavior of the algorithm in terms of track error performance. The experiments were conducted while randomizing observations of three real time 3D trajectories acquired from the stereo system. The simulations were run twenty times with different burst length *m.* Different values of time frames *T* were used. The outputs were then averaged to generate the results. The results were also compared to the ground truth. It is clear from Table 7.4 that varying *m* and density of points over time affects the track error performance. Since the

indoor stereo system readings generate an error of up to *0.34 in (~8.64 mm),* this error tolerance can be used in comparison to ground truth in determining track error. The last column of Table 7.4 shows the results with error tolerance applied while using *m=6*. Increasing *m* decreases the error, however, the number of possible combinations also increases so it affects computation time. These combinations as explained above can be reduced if we have partial knowledge of the trajectory points.

Table 7.4

Track error with different points density and block length m. Track error is a fraction of wrong trajectory point assignments.

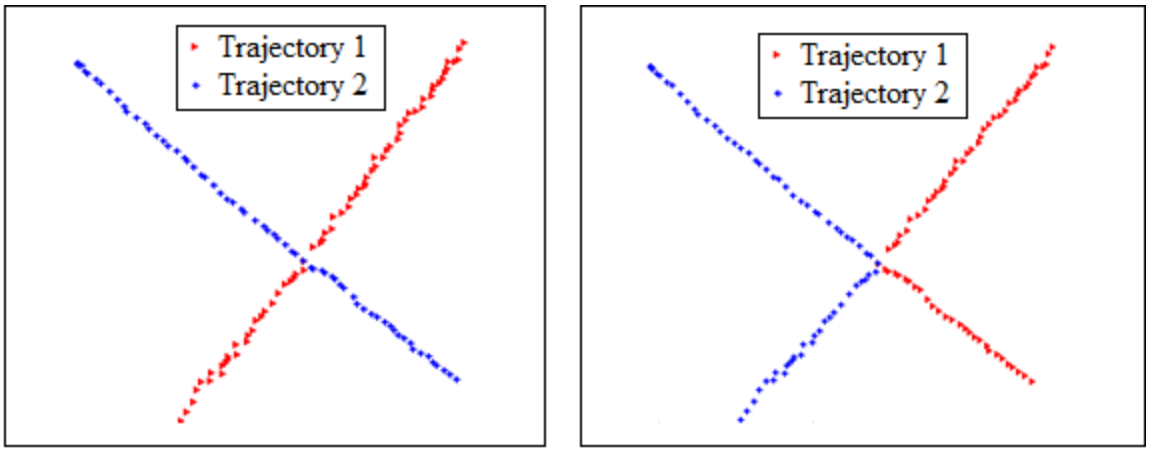
T	*m=3*	*m=4*	*m=5*	*m=6*	*m=6 w/ error tolerance*
10+	*0.121*	*0.095*	*0.043*	*0.027*	*0.0*
20+	*0.196*	*0.096*	*0.054*	*0.042*	*0.004*
30+	*0.239*	*0.103*	*0.067*	*0.058*	*0.021*
40+	*0.251*	*0.134*	*0.083*	*0.066*	*0.023*
50+	*0.262*	*0.159*	*0.091*	*0.071*	*0.036*
60+	*0.266*	*0.167*	*0.098*	*0.075*	*0.038*

7.10 Tracking efficiency using fusion

We have explored via simulations and real scenarios how fusion can reduce the combinatorics of tracking. These cases can help reveal our fusion approach behavior. Although we have acquired both RFID and CV data from an outdoor site, it is impossible to explore the many possible parameterizations using real data. Moreover, simulations allow us to construct interesting test cases and to control their ground truth.

7.10.1 Simulated scenario: two persons and two briefcases

We consider two persons who walk toward each other, exchange brief cases, and then move to different final positions. Due to smoothness constraints, the geometric data will produce incorrect tracks with the persons continuing with their briefcases to different final positions. However, reliable location and ID of whichever person using either RFID or CV enables the correct interpretations to be extracted. This case is simulated by generating two trajectories *(N=2* for *T=70* time frames) with ID info randomly provided and simulated using color. The point of intersection occurs at frame *t=43*. The mean velocity of both the object tracks is kept the same. As shown in Figure 7.19(a) using the smoothness criteria alone with no labeling information, produced wrong trajectories with a track error of *0.39*. However, once ID labels with location information are provided near the intersection, the tracking algorithm interprets correct object tracks as shown in Figure 7.19(b). Therefore to avoid ambiguities in the vicinity of collision points, some localization and object ID information is necessary outside the area of collision.



(a) (b)

Figure 7.19

Testing fusion using simulated scenario of two persons exchanging briefcases: (a) wrong interpretation of trajectories with CV alone (b) correct interpretation of trajectories with CV & RFID fusion.

7.10.2 Real outdoor scenario

We generated a scenario to track the activity in the test site outdoors. The cameras were placed on tripods at a height of *9 ft* with a baseline of *10 ft*. The center of the test site was *84 ft* from the cameras. The area of activity was *47 ft to 93 ft* from the cameras. The persons wore bright colored clothing, which was helpful to generate a good quality feature vector. Four RTLS readers (one master and three slaves) were installed in the test site of *40×40 m* to generate a single cell.

Two tagged persons wearing distinctive color clothing and helmets slowly move forward towards each other and meet at the center of the test site. They exchange RFID tagged colored

balls and then backtrack to their starting positions. All the movement was done on predefined paths to compare the acquired data with the ground truth trajectories. The run was carefully conducted so that the exchange interaction over some of the time frames is either fully or partially not visible to the cameras. If a ball's feature vector does not contain color, then CV detection using shape/size might show wrong 3D labels of the ball's track. This might result as if the persons took them back towards their starting position. Even if the color info were available there still would be uncertainty or loss of trajectory points as the interaction was occluded from the cameras at the center of the test site. This might again result in wrong labeling or lost tracks. Additionally, in case of person tracking, the smoothness constraints would fail to provide correct trajectories (see Section 3.4 for details). Figure 7.20 shows the outdoor arrangement as well as the 3D map of the site with computed real trajectories.

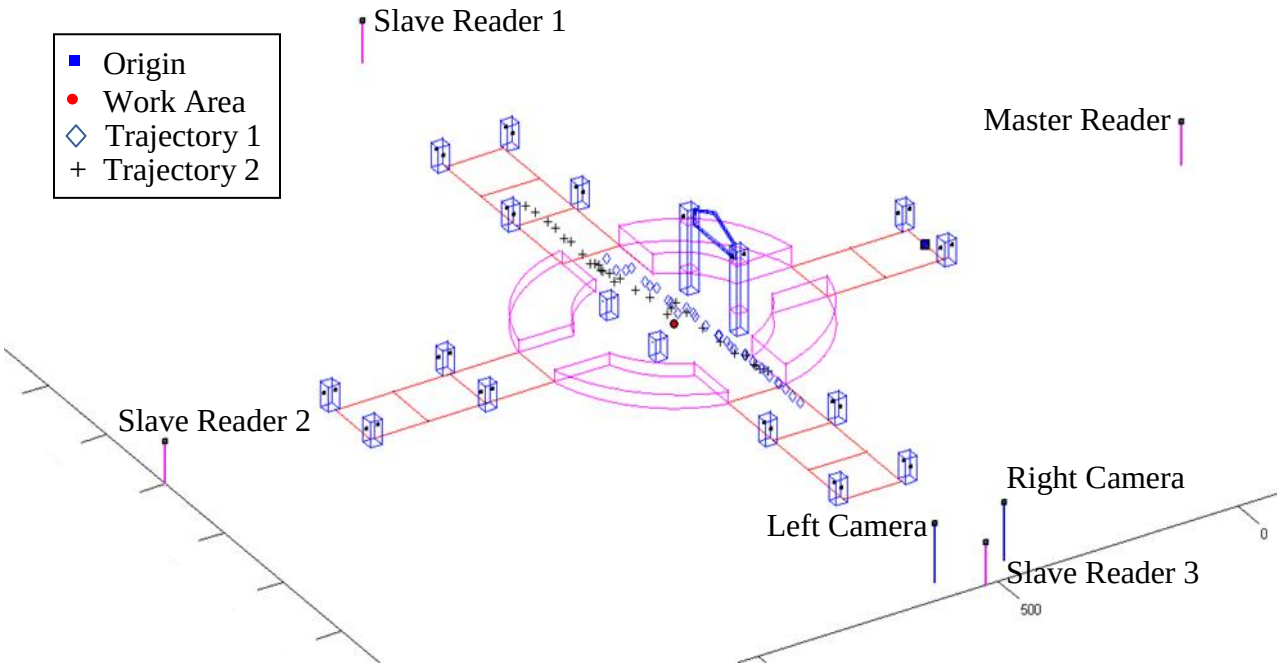
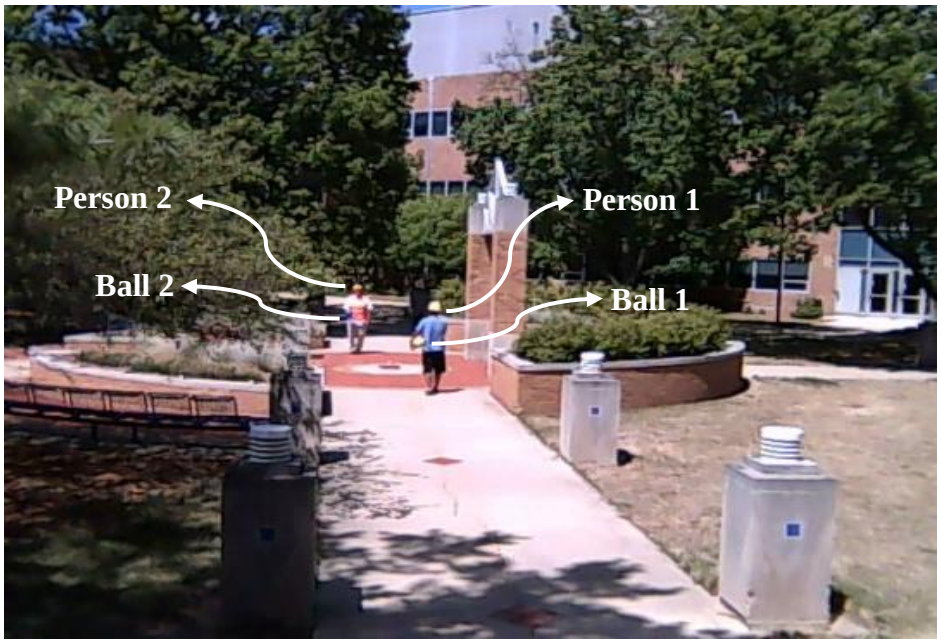


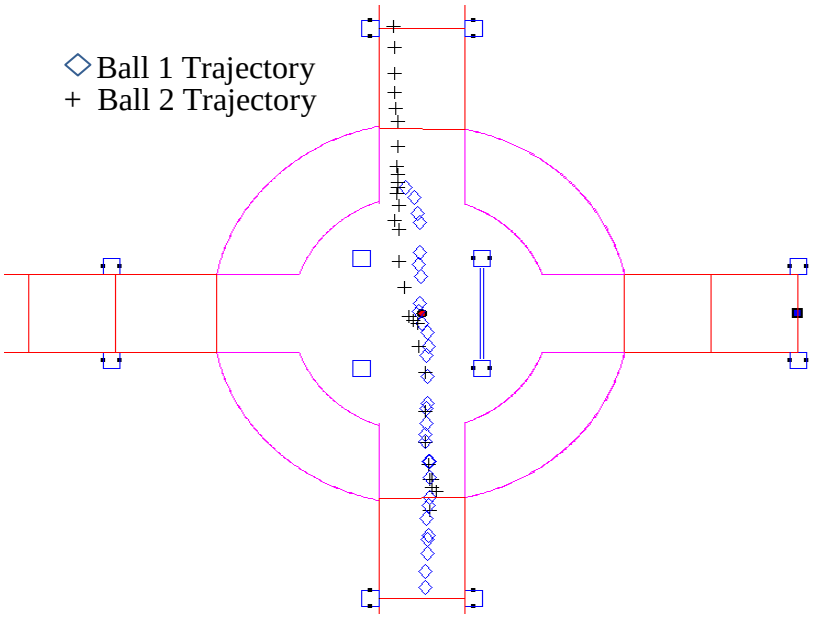
Figure 7.20

3D view of test site with calculated real trajectories.

The experimental runs were conducted while tags were placed on the objects or in the pockets of the persons. The reported experiment was conducted on a partly cloudy day with considerable variation in illumination; moreover, *30 mph* gusts of wind typically shook cameras and RFID readers at some point during each data collection trial. We also placed some reference tags in the area during the experiments. It is observed that the location accuracy of the stationary tags, when visible by all four readers, was within *1.5 m*. This ranged up to *4.3 m* at some points where the tags were visible only to two readers. The test site selected is such that it has some indoor properties -- brick structures and trees -- that cause obstruction and generate a multipath effect. These obstacles also provided occlusion for CV, which helped our study. We also note that the RFID system software initially lost track of all the tags due to the operating system protection scheme to counter hacking attacks. Once registered, the system on average kept track of five RFID tags *79%* of the time. Figure 7.21 shows the left camera view of the test site and correct 3D trajectories using fusion.



(a)



(b)

Figure 7.21

Outdoor scenario to test fusion: (a) left camera view of test site (b) computed correct 3D ball trajectories using fusion.

For this scenario we assessed performance of tracking using track error. To sync video frames and RFID refresh cycle the test was conducted by tracking the balls and the persons with thirty non-consecutive time frames *(i.e T=30)* approximately three seconds apart and burst length of four *(i.e m=4).* While person *1* was holding ball *1* the stereo track error performance for ball *1* and person *1* trajectory was *0.06*. For the time person *2* was holding ball *2*, the stereo track error for ball *2* and person *2* trajectory was *0.53*. The track error for ball *2* and person *2* was much larger due to fewer distant camera calibration points in the image, which resulted in larger 3D stereo error. Also due to strong winds the cameras position was not stable which likely increased stereo error. Linearly increasing stereo error is an another reason. Therefore the trajectories beyond the site center point (*84 ft* from the cameras) had more stereo location accuracy error, which subsequently resulted in greater track error. The greater stereo error however, generated favorable conditions to investigate fusion efficacy.

Due to occlusion, the vision system often lost track of the balls at the center of the test site (*84 ft* from cameras) while the persons were exchanging the balls. The tracking algorithm assigned correct labels to both person trajectories up to the site center. Thereafter the labels were assigned incorrectly due to change in direction of the persons as they backtracked to their starting position. The interaction at the test site center was clearly detected by the readers since the tags were directly visible to all four readers in that area. The RFID location information, when fused around these points, helped reassign correct tokens resulting in correct labels and trajectories of both persons and balls. When RFID location information was applied, the computation time in this area was reduced due to reduction in the combination space. Also the track error for the initial trajectory of ball *2* and person *2* decreased from *53%* to *13%*. Discarding the outliers, the

RFID location accuracy for dynamic tags averaged *2.6 m*, which is nearly comparable to the one mentioned by the OEM [21]. To achieve this location performance it is noted that the tags should be in the cell area where antenna beams of at least three readers intersect without any solid obstruction.

7.10.3 Outdoor scenarios with varying fusion information

To analyze further we generated other ground truth trajectories with varying sensor information in different configurations. The persons are assumed to be moving with the same mean velocities. We generated two other cases where two tagged persons start from opposite directions towards each other and:

a. Keep going without any direction change.
b. Split on sides at the center of test site as shown in Figure 7.22.

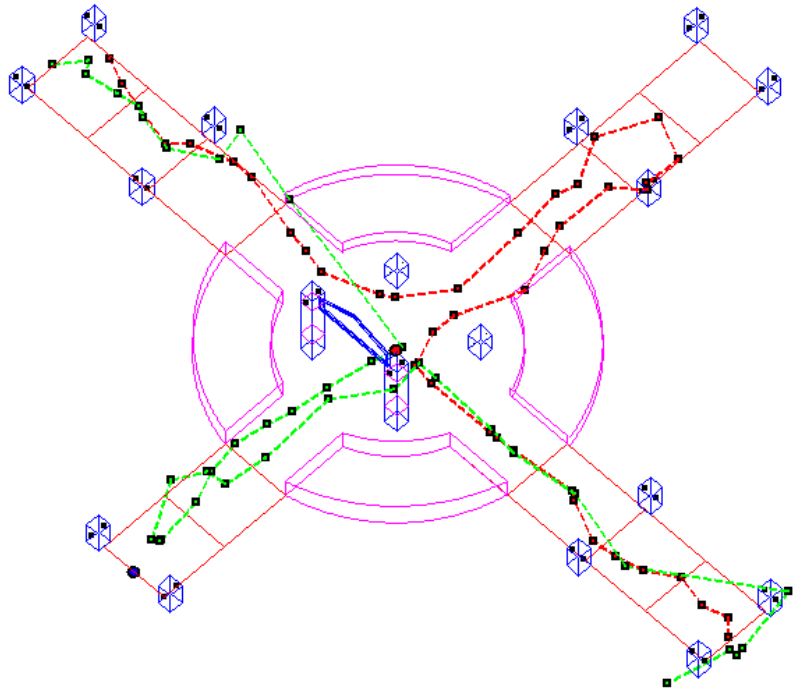


Figure 7.22

Outdoor RTLS trajectories of two tagged persons with varying fusion information - persons split on sides at the center of test site.

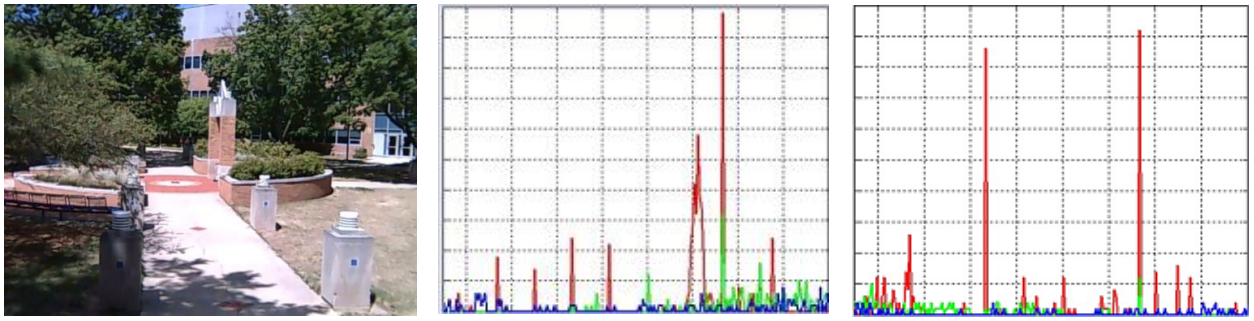
For site safety and security, we can assume that a high percentage of objects are tagged and cooperative. We have shown how proper sensor placement can support tracking. This enables the system to spend more resources on tracking anomalies -- "unauthorized" animals, machines, or materials. Location accuracy and compute time of a central algorithm, although good, is insufficient to handle collision avoidance, so we recommend that moving objects use local collision avoidance -- perhaps based on looming (CV) or locally shared kinematics. Fundamental tests on looming detection are reported in Appendix A. Cell phone and sensor network technology are advancing rapidly and probably will soon provide such functions [104].

7.11 Object color variations

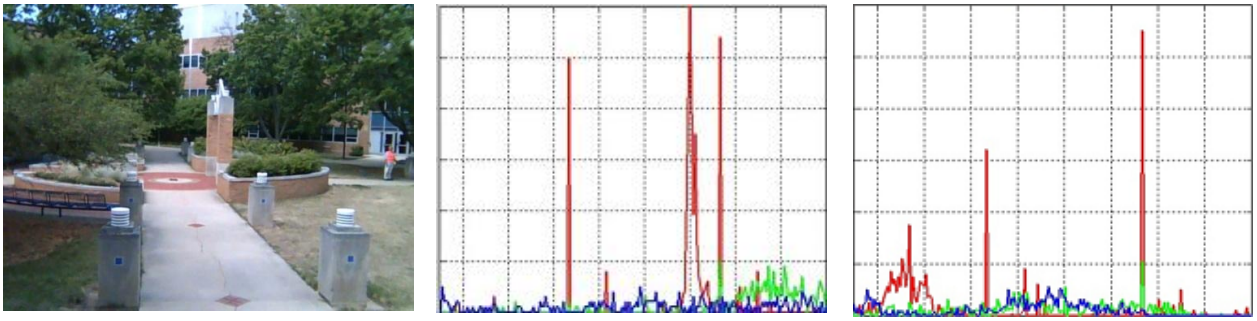
Figure 7.23 shows four color histograms of segmented objects extracted from the left camera video feed under different outdoor conditions at different time instances. The histograms were computed using HSV color space. There are two objects, a blue ball and a yellow ball, and two different illumination conditions, sun and shade.

The irregular outdoor illumination variations and abrupt changes of brightness is evident in Figure 7.23(a) and (b). If color is to be used by CV to help tag and distinguish objects, then the objects must be for the most part distinguishable in the video images. In many cases workers will be wearing hard hats or vests of special coloring. The S-3 should be able to take advantage of these distinctive colors by exploiting color consistency for reliable color clustering.

Even though there is variation in illumination however the histograms (in sun and shade) of the balls (blue or yellow) in Figure 7.23 show noticeable association within each color group. The experiments reported in the previous sections of this chapter did not use automatic color similarity computations to distinguish the class of object color: instead, a symbolic color was assigned to the token.



(a) (c) (e)



(b) (d) (f)

Figure 7.23

Change in illumination observed in left camera video feed when: (a) sunny (b) shady. Color histograms: (c) blue ball in sunlight, (d) blue ball in shade, (e) yellow ball in sunlight, (f) yellow ball in shade.

We collected various samples and analyzed HSV color space consistency for the blue and yellow balls in different weather (winter and summer) and illumination (sun and shade) conditions as shown in Figure 7.24.

Figure 7.24

Sample images for different weather and illumination conditions to study blue and yellow color consistency.

The results shown in Figure 7.25 show the color consistency for blue and yellow balls for reliable color clustering. The points shown are the average pixel values of the ball area taken from different frames. The yellow marker 'o' represents the yellow ball HSV value and the blue marker '+' represents the blue ball HSV value. The color clusters are clearly separated along the hue axis which supports usefulness of CV to help distinguish objects based on color in an outdoor environment.

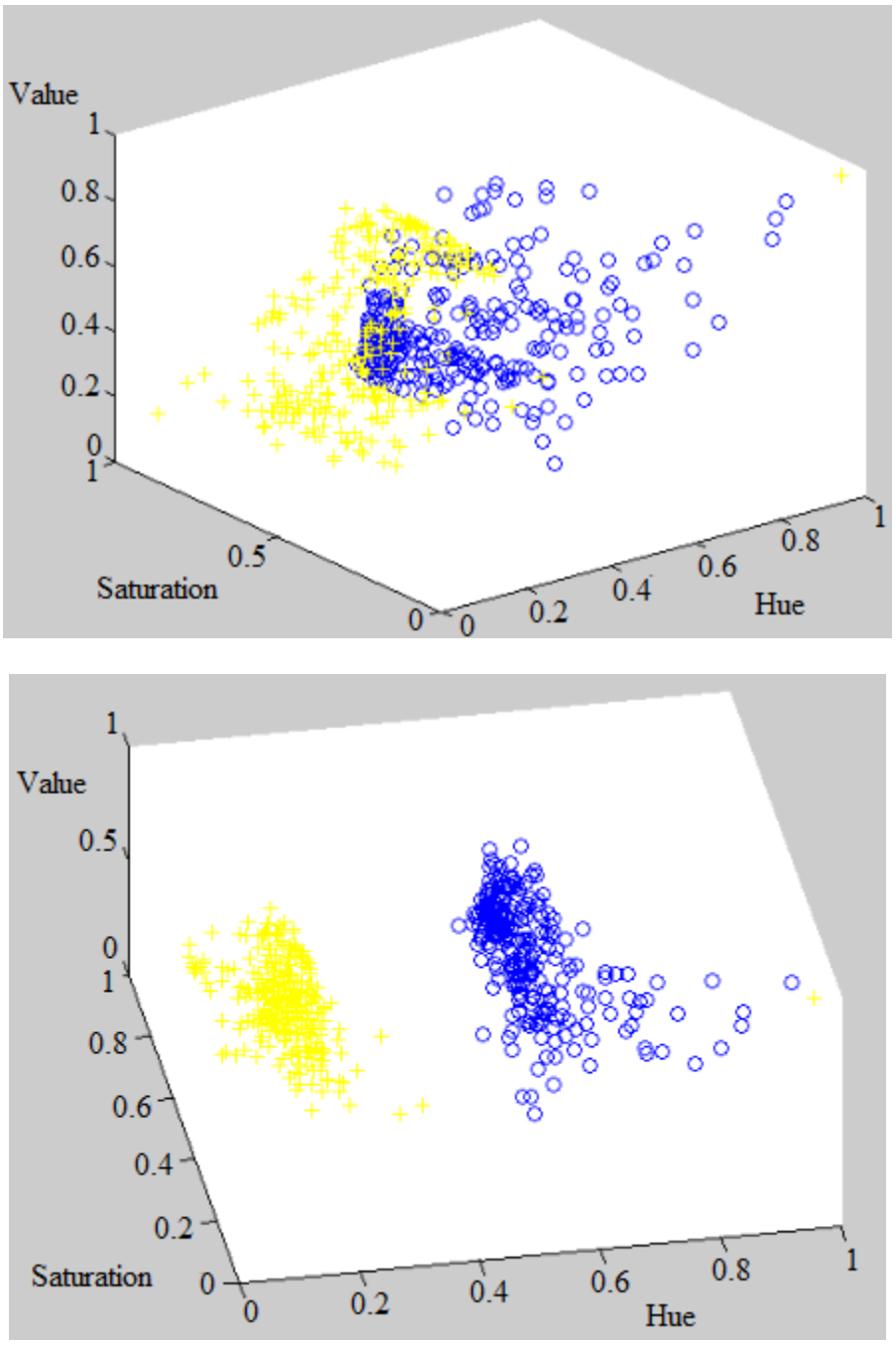


Figure 7.25

Analyzing blue and yellow ball color consistency in HSV color space under different weather and illumination conditions.

7.12 Sensor Error and synchronization problems

In an outdoor environment we evaluated positional accuracy and reliability for the RFID based Real Time Location System (RTLS) and the stereo computation obtained using commodity cameras. The calibration of stereo system and RTLS system were done on the same test site. An adequate number of distant points (in the background) and nearby points (in the foreground) were acquired to serve as calibration markers for stereo computation. The stereo infrastructure provided RMS positional accuracy within *7.6 in* (*19.3 cm)* for *x, y* and *z* directions. The reported location accuracy for the RTLS system for static tags is *~1.5 m* and that of dynamic tags is *~2.6 m*. RMS error does not include the occasional outliers that are possible from incorrect stereo correspondence or multiple path effects in RFID. The RTLS location accuracy however can be increased by deploying further readers in the test site.

One significant practical problem for fusion is the different sampling rates of the sensors or the extended time needed to smooth data or to make decisions about the motion of an object. Due to time division multiplexing, our RFID system provides data on all objects every two to three seconds, while our stereo implementation could produce ten updates per second for a few objects. In our experiments, we typically force a common sampling time for RFID and CV and look back two time samples to estimate motion. The uncertainty of location for RFID is much larger than for CV for static objects and even larger for moving objects due to under-sampling. Interpolation using CV locations can be used with sparse RFID samples with reliable ID – another benefit of fusion. Finally, it is possible that an object is invisible at some time steps to

either or both CV and RFID due to occlusion and higher level processes are left to interpret what is happening.

7.13 Summary discussion

In this chapter we have reported experiments and results that evaluate use of stereo vision and the commercially available RTLS system in single and fused modes. To test the 3D accuracy and reliability of object location using stereo we have generated three types of trajectories that includes mathematical, real indoors and real outdoors. We acquired ground truth data for various predefined points in the surveyed test site with which the RTLS and stereo data can be compared. We have discussed the performance metrics criteria used to evaluate localization and tracking schemes. We assessed performance of tracking using smoothness constraints in terms of track error. The stereo approach with RFID simulated using color was tested in a real time indoor lab bench. The tracking algorithm there had less than *1.4%* missed observations with zero track error. For visual analysis we showed trajectories both in the wireframe volume and lab area. We analyzed that the stereo error has a linear behavior when the distance between the observed object and stereo setup varies. Least squares analysis was also performed to assess the error in location using vision and RFID. The stereo provided within *7.6 in (19.3 cm)* accuracy and for RTLS it varies between *~2 m to ~2.6 m* for moving tagged objects. To achieve this location performance it is noted that the tags should be in the RTLS cell area where antenna beams of at least three readers intersect without any solid obstruction. With a refresh rate of two to three seconds the RTLS hardware provided *79%* to *84.5%* signal availability which can significantly reduce the combination space. Also in the fused system, the RFID information availability at ambiguous instants in tracking could reduce runtime up to *99%*.

The likelihood of producing correct object trajectories in regions partially or fully occluded to CV is also increased. Lastly we have studied object color variation in different illumination conditions and found noticeable association within each object color group that can help object detection in outdoors.

We have shown how fusion can improve identification, localization and tracking results while also reducing computational cost. In general fusion of RFID and CV is better than using only one mode alone and, where costs are justified, will produce systems that are better than those using only one modality.

CHAPTER 8

Conclusions and future work

In this dissertation we have presented, a generalized framework for the fusion of Computer Vision (CV) and Radio Frequency Identification (RFID) that can produce more accurate object localization and tracking in a three dimensional space and do so using more efficient computation. The important components of a fused system have been implemented and tested and the results obtained support the premise that fusion can improve performance in various applications over use of RFID or CV alone. The basic rationale is that RFID can provide highly reliable unique object identification, although with coarse object location, while CV can provide more accurate object location along with confirming visual features and can also avoid cloning of tags and decrease counterfeiting. Below we provide the concluding discussion highlighting our contributions and the research expansions that are possible as a future work.

8.1 Background Survey

Research and development using fusion of RFID and computer vision has been thriving over the past decade. Almost all work has been done in indoor environments. During our background research we have presented the collection of these schemes and have related the research and

actual applications and installations. Dozens of publications were found that directly used the fusion approach. The work had been generally from the area of recognition, localization, and tracking where RFID was mostly being used at the initial stages of object detection and/or identification. Moreover, a few dozen more publications in either RFID or CV showed clear potential for improvement via fusion. Most reported work was done in an indoor controlled environment at small scales and using passive tags which require close read range. We also learned that RFID can be very useful outdoors in a number of applications, such as construction site safety, when tagging is possible. Moreover, formal linkage of RFID based RTLS with vision is a new and expanding research area with great potential. All these factors gave us motivation towards exploring these modalities for real time localization and tracking. We believe that the survey of the fusion approaches that we have provided in this dissertation will offer great support to the researchers interested in this area.

8.2 Evaluation of RFID, CV, and fused sensing

For RFID we have used a commercially available Real Time Location System [8] and we developed our own stereo system with a laptop, MATLAB, and two commodity color cameras. Our total hardware cost was only about *US$5500*, for both the RTLS and only one stereo pair of cameras. High level performance would require more cameras and more RFID readers in the workspace than we have used. We have defined our performance metrics and have done error analysis for location estimation by these modalities. We have evaluated the use of stereo vision as a single modality both indoors and outdoors for 3D accuracy and reliability of object location. We have studied and implemented the ray-ray stereo scheme [31] for 3D localization in an

indoor environment and have reported an RMS accuracy of *~0.34 in* in a wireframe workspace and *~6.4 in (16.3 cm)* at room level. The average RTLS location accuracy indoors for static tags was *~1.9 m*. The performance for dynamic tags varied a lot due to multipath effects in a compact space. However, using optimization tools RTLS accuracy indoors can be improved.

8.2.1 Demonstrated performance, potential and parameters for outdoor applications

We analyzed the efficacy of the stereo approach outdoors in a surveyed test. In our analysis we have obtained RMS location accuracy within *~7.6 in (19.3 cm)* in *x, y*, and *z* for trajectories within range of *30 ft* to *70 ft* from the cameras in a workspace that is *40×40 m.* Choosing appropriate calibration points covering the near and far field of the scene is necessary for results with minimal error. This offers potential for future researchers to examine automated three dimensional tracking outdoors using economical vision sensors. We established that the location accuracy for RFID in the same outdoor test area was *~1.5 m* in *x* and *y* ground coordinates for static objects, but *~2 m* to *~2.6 m* for dynamic objects, which we attribute to the location update frequency i.e one RTLS observation approximately every two seconds. Deploying more readers in the area can substantially improve the location accuracy for moving objects. The *z* dimension, though available, comes with a planar restriction defined by the tag height. Getting interpolated location estimates and varying tag height data is, however, not limited in theory. We have shown with simulations and real data that fused sensing increases the likelihood of producing correct object trajectories in regions partially or fully occluded to CV. As the stereo system readings generate an error of up to *7.6 in (19.3 cm)*, this error tolerance can be helpful in comparison to ground truth in determining track error. We have also studied object

color variation in different illumination conditions and found noticeable association within each object color group that can help object detection in outdoors. In general we have shown how proper sensor placement can support localization and tracking. This enables the system to spend more resources on tracking anomalies -- "unauthorized" objects, machines, or materials in a site safety environment.

8.3 Modeling fusion and its benefits

We have presented our fusion model and have described the competitive and cooperative relationship of our sensors and the fusion building block needed to develop the fused system. The features and characteristics of our fusion scheme are also provided to explain its adaptability towards established fusion standards. With a focus towards localization and tracking applications we have provided the potential benefits achievable from the fused system and have documented examples where fusion disambiguates object tracks and combines the strengths of RFID and vision to avoid problems of each mode.

8.4 Data integration and filtering using relaxation labeling

To manage the complexity and integration of the diverse information being fused and to provide a flexible experimental platform we proposed and demonstrated an algorithm based on discrete relaxation. Discrete relaxation was chosen to control tracking so that we could easily experiment by switching on or off sources of information and develop our software in a modular way. Moreover, the label elimination approach easily represents the ambiguity occurring in real-

life applications. If there are N objects and N labels, the computational complexity of tracking is potentially of the order N^2 across just two time steps. The key to reducing the computational requirements by using relaxation is to eliminate many labels at each filtering step while keeping those labels compatible with observation. The output labels from the relaxation labeling process can be optimized further by the post processing operation.

8.5 3D object tracking algorithm

We have proposed a three dimensional object tracking scheme using fusion of vision and RFID. As explained above, relaxation was used for filtering out incompatible labels in the tracking algorithm. As a post processing step to relaxation labeling we have used total track smoothness for optimization to update computed tracks for increasing system tracking reliability. We assessed performance of tracking using smoothness constraints in terms of track error. With simulations we have shown how fusion can greatly increase tracking performance while also reducing computational cost and combination search space up to *99%* in some cases. Test cases show how fusion can solve some difficult tracking problems outdoors. For some object trajectories outdoors, the fused system reduced the track error from *0.53* to *0.13*. We have demonstrated cases where fusion disambiguates object tracks and we have also given cases where disambiguation is impossible, as in the well known shell game. We demonstrated how uncooperative objects can cheat the system. However, in general, fusion of RFID and CV is better than using only one mode alone and, where costs are justified, will produce systems that are better than those using only one modality. Moreover, an automatic system detects the ambiguities and can cue the attention of higher level processes or longer lived processes,

including the attention of human security personnel. Simulations of tracking over many ground truth paths demonstrates how knowledge of unique object ID for some time instances can significantly improve correct tracking as well as reduce computation time in producing the tracks. Thus, many more objects can be tracked in practice if fused sensing is available compared to tracking by CV alone. A fast tracking implementation would be active – it could plan more efficient work, warn of possible collisions, or detect illegal operations. Finally, it is clear that the global workspace view we have used is too imprecise for detailed object interactions, such as cooperation compared to collision, or handing off carried objects. Object born touch or looming sensors would be needed for some applications. Our current work shows that pursuit of these extensions should be fruitful.

8.6 Future work and limitations

One significant problem in fusing the RFID and CV feeds is the difference in sensing frequency. Commodity cameras are designed to represent human motion well and produce upwards of ten video images per second, whereas our RTLS system produced tokens for all tags at approximately two second intervals. Engineering faster RFID updates will likely reduce the number of objects that can be sensed; however, this should be a favorable tradeoff in a construction site. It may also be good design to have a hierarchy of RFID sensing with a slow system for asset/material inventory and a fast system for critical objects such as workers and moving machinery.

We need to continue to develop our system to perform the lower level token combination and to test it fully using a set of objects with some typical behavior. We will also make the revisions that model objects that appear and disappear from the surveyed workspace. Also the constraints and heuristics used in the tracking algorithm should be further studied and improved. There are many knowledge based constraints that we have not yet applied.

Much of what has been discussed assumed objects were single independently tracked points. Clearly, some objects would be a rigid aggregate of points. For example, a truck might have a single RFID tag and perhaps four or eight visual markers that would reduce combinatorics and enable rigid motion analysis. Such planar rigid structures and symmetries are also helpful to track moving objects over wide variations in position and orientation of the objects.

Considering the fact that most construction sites involve collaborative work, interactions between workers, and workers-machines will happen frequently. The interactions will introduce static or dynamic occlusion, which causes difficulties for visual tracking. In case of short-duration, partial *object-object* and *background-object* occlusions, the vision system should continue to track the object. RFID though can help CV here, but it is important to deal with occurrences of an occlusion while having an RFID feed failure scenario in rare cases. Knowledge based information input can be used to inform the tracking processes to handle occlusion. Also by tracking each object, it is possible to use global 3D information to tackle occlusion with predictive trajectories in our optimization process. The occlusion management framework to be worked on is shown in Figure 8.1.

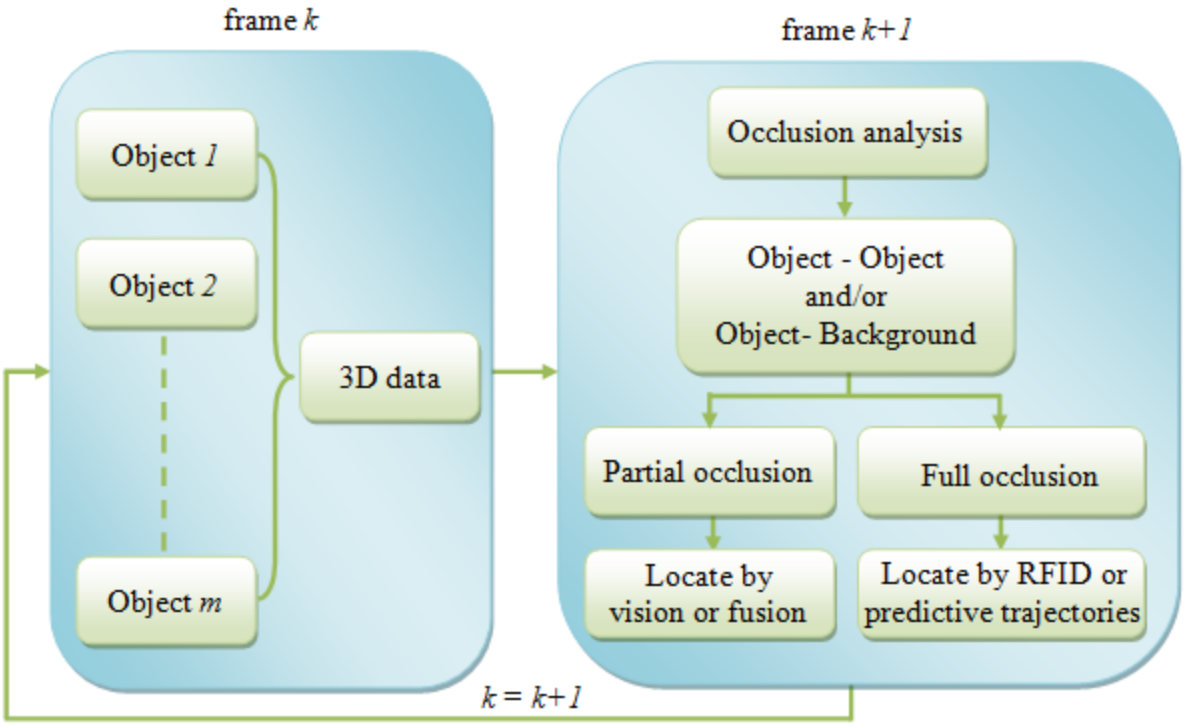


Figure 8.1

Occlusion management flow during stereo tracking.

Location accuracy and computing time of a central algorithm, although good, is insufficient to handle collision avoidance. So there is need to have a local object-object communication. We recommend that moving objects use local collision avoidance -- perhaps based on looming (CV) or locally shared kinematics. Cell phone and sensor network technology are advancing rapidly and probably will soon provide such functions [104]. As part of our future work we have presented some of the fundamental experiments on looming in Appendix A.

The results we have presented apply to tracking in either 3D space or in a 2D image of that space. Our conceptual model is ahead of our current implementation, and thus provides a design for stages of future improvements. Incorporating object track initiation and termination, more

cameras and readers, more constraints, and more sensed features, such as object velocity are on our list for future work.

Historically, most security and surveillance systems have used video input fed to human monitors. Automated methods in CV have been developed to replace or augment the human recognition duties with varying success. In controlled areas, such as airports, hospitals, workplaces, and construction sites, many objects can be tagged, including cooperative humans. Thus RFID based detection and location can be available for integration with the video data. Tracking of tagged objects using RFID can drastically reduce the computational load of a vision only approach as well as increase its performance. The CV component would only need to process exceptions and might be able to pass some of them to a human monitor. This fused tracking information will increase the safety or security of those being tracked and their activities. Current applications include individual and group recreational and commercial sports.

We finish this dissertation with a conviction that the work on fusion of RFID and CV would further the cause of mankind; e.g. more secure and safer public transportation and human management [such as at airports], while even more efficient and economical resource management. Looking in the future, it can be applied for disaster/rescue management as in an air crash. Currently, a downed plane can be localized but the search and rescue of unfortunate passengers is still dependent on human sight. Imagine the possibility of human movement tracking through cell phones or medium tolerant RFID bracelets worn by passengers. This could be achieved by deploying preprogrammed flying robots equipped with cameras and RFID tag

readers and tag libraries, which would relay real time and processed crash site info to the main rescue vehicle for timely actions and making the difference between living or otherwise.

APPENDICES

APPENDIX A

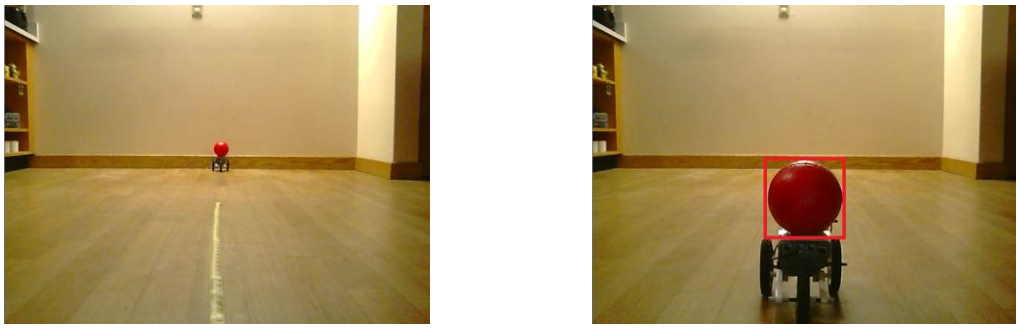
Fundamental experiments on looming

Looming is present in animal vision and is vital for collision avoidance and alighting. In a Site Safety System (S-3) realizing collision avoidance requires understanding local interactions between workers and machines. For understanding looming concepts, fundamental experiments were carried out. We have studied the relationship between the rigid object area versus the looming distance. Later we employ and analyze significance of looming information for collision avoidance in real time. The details of our indoor test platform are also provided that uses optical flow algorithm for object and looming detection. Possible future developments using other sensors onboard the smart phones are also discussed at the end.

A.1 Generating looming dataset

The initial experiment was conducted offline to generate a training dataset for looming in a controlled indoor environment. It included a ball placed on a *LEGO NXT* robot as shown in Figure A.1. A red colored ball was used due to a rigid spherical shape and easy detection. A lightly textured background also helped in object detection. The assembly start position was at ten feet and the stop position was at two feet from the camera. The distance markers were placed

on the floor after every two inches for providing distance info. The robot was programmed to move towards the camera in a straight line while stopping for five seconds on the predefined points. Images were captured when the robot was stationary and a dataset of images was generated with distance stamps. The camera used was a low cost *Logitech C210* at a resolution of 640×480 *pixels*. The simulation was done using *MATLAB* © *2009a*.



(a) (b)

Figure A.1

Looming image dataset at different distances: (a) object at ten feet (b) object at two feet.

As explained in Section 4.1 we used color and blob analysis to detect the ball. The algorithm generated an [approximately square] bounding box around the ball as shown in Figure A.1(b). Ideally, after object detection the algorithm should be able to produce an approximate square around the ball. However, in practice it becomes challenging to obtain precise object edge information due to factors such as low camera resolution and varying illumination conditions. As our further analysis depends upon area of the bounding box, therefore we considered comparing the bounding box height and width to test their degree of uniformity. Figure A.2 shows the linear relationship between the bounding box parameters [width and height in pixels] relative to

the distance from the camera . The maximum error observed due to the noise was within *4%* of the linear theoretical range.

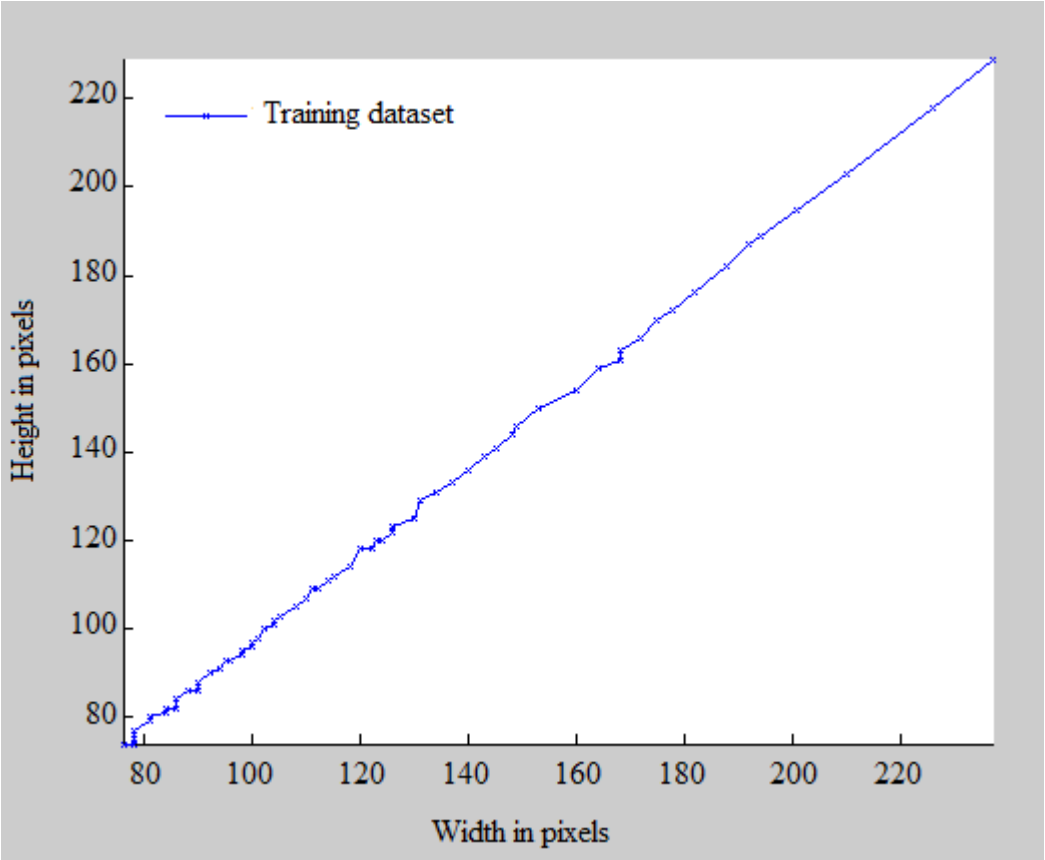


Figure A.2

Graph of bounding box width and height relationship for training dataset.

Figure A.3 shows the bounding box normalized area versus the distance relationship. The graph provides information that the area of the rigid square object changes quadratically with the change in looming distance from the camera.

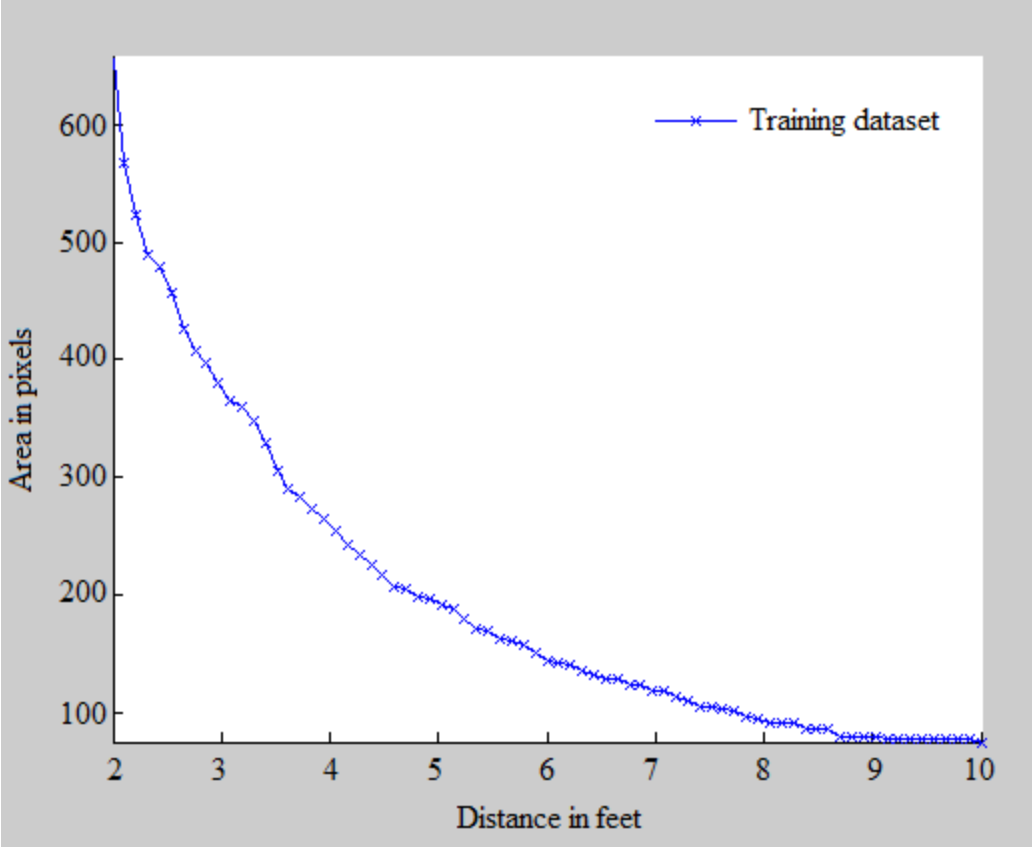


Figure A.3

Graph of bounding box area vs looming distance relationship for training dataset.

A.2 Object distance measurement in real-time

Next the same run was conducted in real-time for ten trials in a controlled environment. Following the supervised learning approach by having a training dataset, the looming algorithm computed the normalized area of the ball [in pixels] and estimated the distance of the ball from the camera. Figure A.4 shows the results for two real-time trials compared to the offline results generated from the training dataset in Section A.1. The variation in the graph represents variation

in the bounding box width and height. This mainly occurred due to pixelation effects, varying illumination, *jpg* compressed version of the acquired images from the camera, and the system inherent noise which in turn affect the color and blob detection output. The RMS error in the area for the dataset *1* compared to the training set was *22.6 pixel square* and for dataset *2* was *68.7 pixel square*. This experiment presented a basic understanding on how local workspace agents can learn and acquire knowledge about looming and distance for rigid objects in a controlled setting.

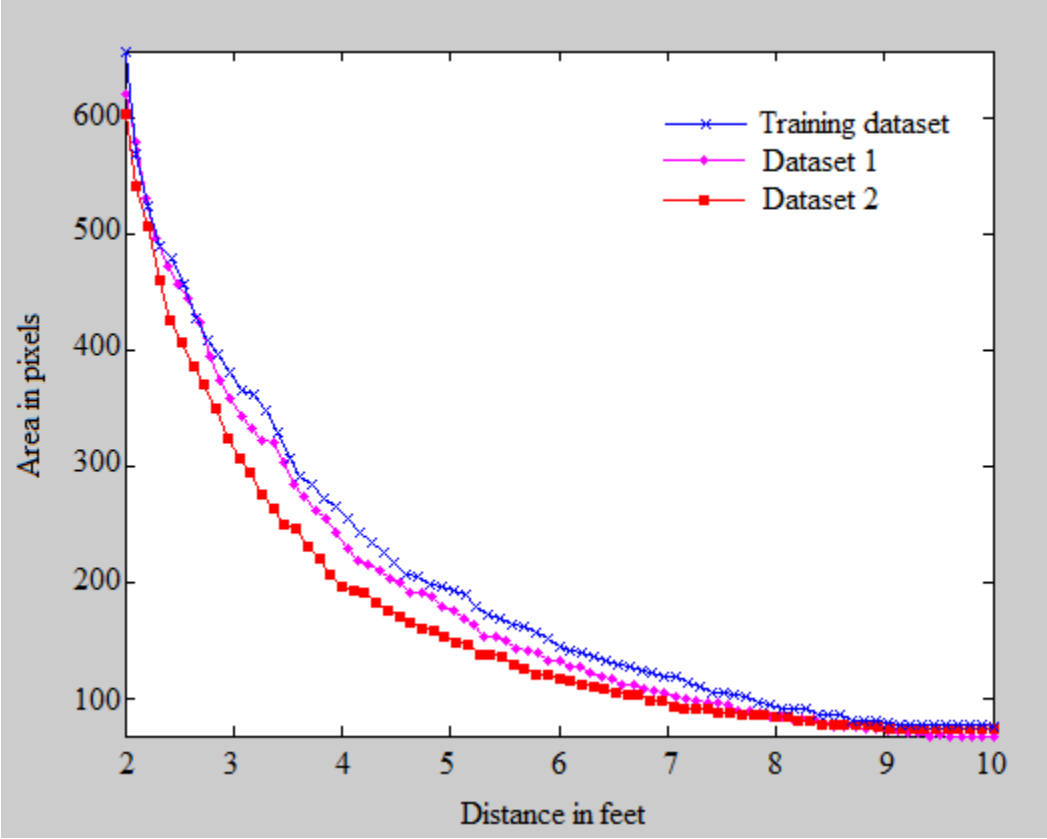


Figure A.4

Graph of bounding box area vs looming distance relationship for two real time datasets.

A.3 Real-time lab demo for collision avoidance

Next for studying looming detection for collision avoidance in an indoor lab environment we designed our own test platform with a LEGO NXT robotics kit with an onboard wireless camera. The collision avoidance algorithm is based on the optical flow. The wireless video was obtained by installing the iPhone 3GS on the NXT robot as shown in Figure A.5.

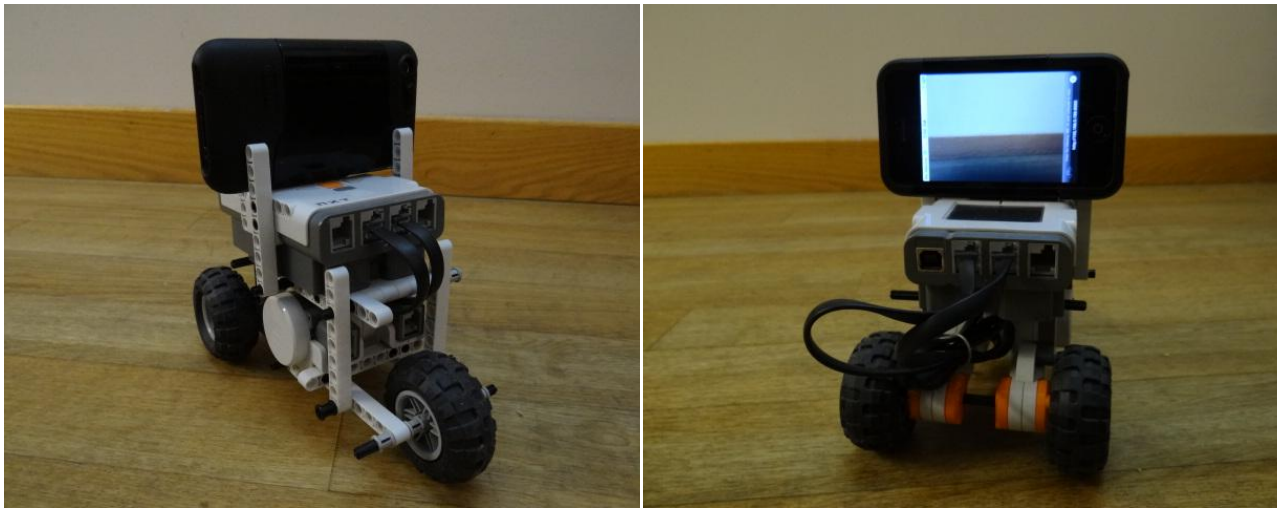
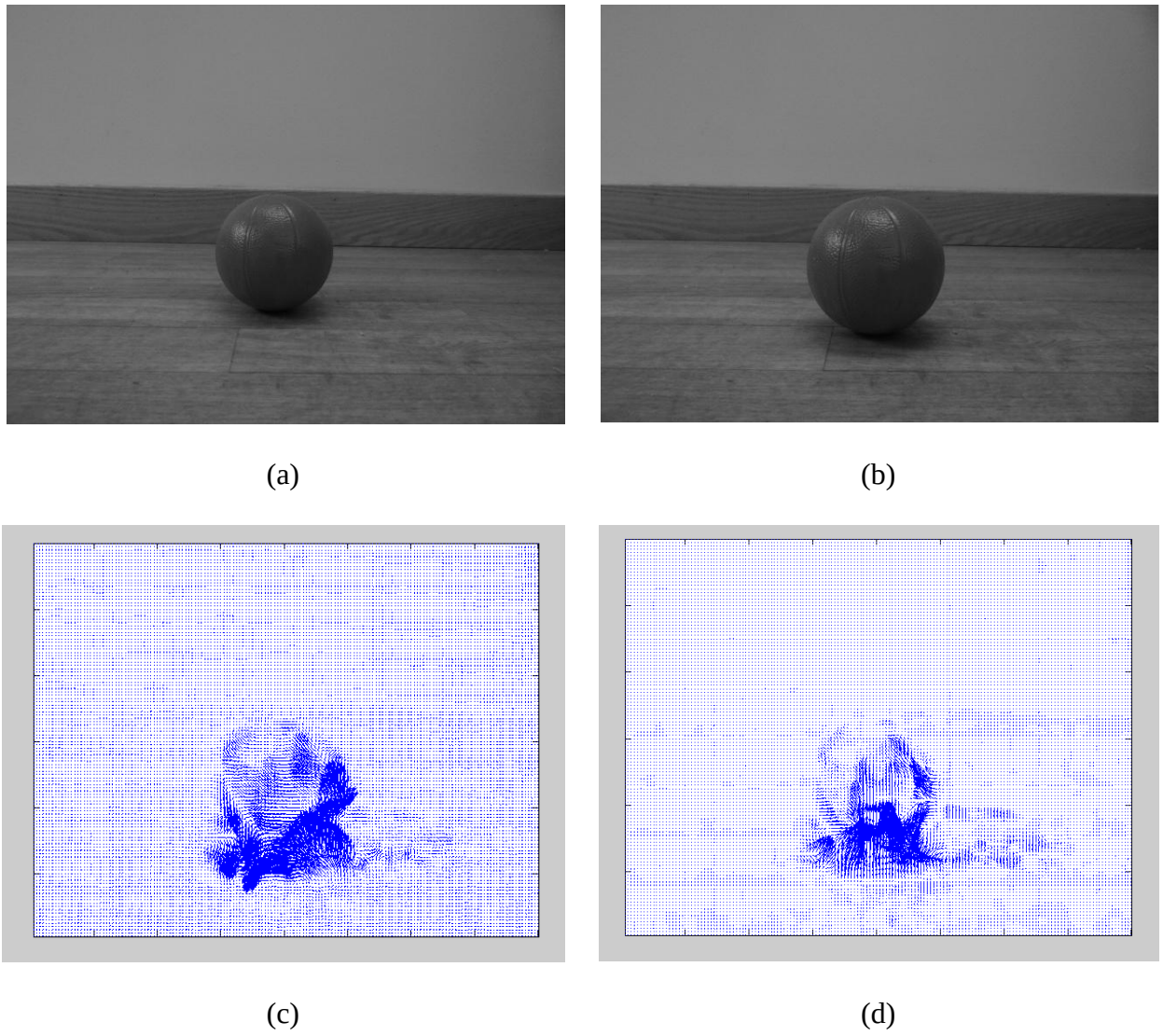


Figure A.5

Indoor lab platform consisting of NXT robotics kit and the iphone 3GS to test collision avoidance using optical flow.

The iPhone camera video was accessed as an IP camera over a local WiFi network, using *IP Cam application* [112]. Since all the simulations were conducted in *MATLAB*©*2009a,* an *m-file* routine was written to acquire iPhone video feed using the *MATLAB* image acquisition toolbox. For acquiring motion vectors using optical flow we have used both *Horn-Schunck* [113] and *Lucas-Kanade* [114] methods separately. The optical flow algorithm converts the acquired RGB

image into a gray scale image. Figure A.6 shows the results of motion vectors computed between a test image pair [gray scale] using *Horn-Schunck* and *Lucas-Kanade* algorithm. The performance of both the algorithms varied by changing their controlling parameters. A shadow effect at the base of the ball is visible in both the motion vector images.



(a) (b)

(c) (d)

Figure A.6

Motion vectors computation during real-time lab demo for collision avoidance:(a) test frame k-1 (b) test frame k (c) motion vectors using Horn-Schunck (d) motion vectors using Lucas-Kanade.

We have used the RWTH Aachen University's NXT MATLAB toolbox [115] to interface NXT with MATLAB. The NXT wirelessly communicates with the PC using the bluetooth protocol. The optical flow algorithm was used to calculate the motion vector magnitudes in both halves of each consecutive frame acquired. If the sum of the magnitudes at a particular time instance reaches a certain threshold then it was considered to be an obstacle and the robot changed its path. The direction in which the robot turned was again governed by the motion vector magnitude sum of the image halves. If the sum of the magnitude of the left half was smaller than that of the right half, then the robot turned left and vice versa. Figure A.7(a) shows a motion vector frame when the robot was approaching the object. The clip on the left shows the actual image acquired by the system. Figure A.7(b) shows the motion vector image when the robot detected the object and changed its trajectory. Due to inherent noise and factors explained above, there was a small number of motion vectors detected when the robot was stationary.

(a) (b)

Figure A.7

Realtime indoor collision avoidance experiment: (a) robot approaching the obstacle (b) robot detected the obstacle.

We are also interested in accessing the sensors in the iPhone using the *User Datagram Protocol* (UDP). These can be useful in acquiring the object pose and trajectory in real-time that will support collision avoidance structure. The iPhone 3GS along with the camera has digital compass, accelerometers and GPS onboard. The latest smart-phone versions also carry a gyroscope and secondary cameras which can be of additional value. Presently we have been able to access the sensor data using the *SensorData application* [116]. We have written our *m-file* routine to access this buffered sensor data in *MATLAB* through the UDP port. Typical data accessed through the iPhone carries information in the following format:

Timestamp,Accel_X,Accel_Y,Accel_Z,MagHeading,TrueHeading,HeadingAccuracy,MagX,MagY ,MagZ,Lat,Long,LocAccuracy,Course,Speed,Altitude

Though GPS information cannot be utilized indoors, digital compass and accelerometers can be used to calculate a machine's or person's course, speed, altitude and pose. Such data from workspace agents indoors as well as outdoors will also be helpful for the local as well as global processes in the Site Safety System (S-3) to make appropriate decisions and generate system alarms.

APPENDIX B

Stereo concepts and calibration procedure

We have provided basic stereo concepts here that are helpful to understand the stereo approach used in this dissertation. It is explained how stereo can be used for recovering 3D information from 2D and what is a correspondence problem. Later we have provided the camera calibration procedure used in our experiments. The method of computing 3D from 2D using the shortest line segment approach is also highlighted.

B.1 Basic stereo vision principles

3D world points on the same viewing line have the same 2D point on the image. Therefore the inverse process in general will be unable to recover all 3D point coordinates from 2D image coordinates, which results in depth information loss as shown in Figure B.1.

^{W}X, ^{W}Y, ^{W}Z represents world coordinates and camera coordinates are represented by ^{C}X, ^{C}Y, ^{C}Z. Both 3D world points $^{W}P(^{W}P^{x}, ^{W}P^{y}, ^{W}P^{z})$ and $^{W}Q(^{W}Q^{x}, ^{W}Q^{y}, ^{W}Q^{z})$ project into

the same image point ${}^{I}P({}^{I}P^{r}, {}^{I}P^{c}) = {}^{I}Q({}^{I}Q^{r}, {}^{I}Q^{c})$ which makes it impossible to recover ${}^{W}P$ and ${}^{W}Q$ from ${}^{I}P = {}^{I}Q$.

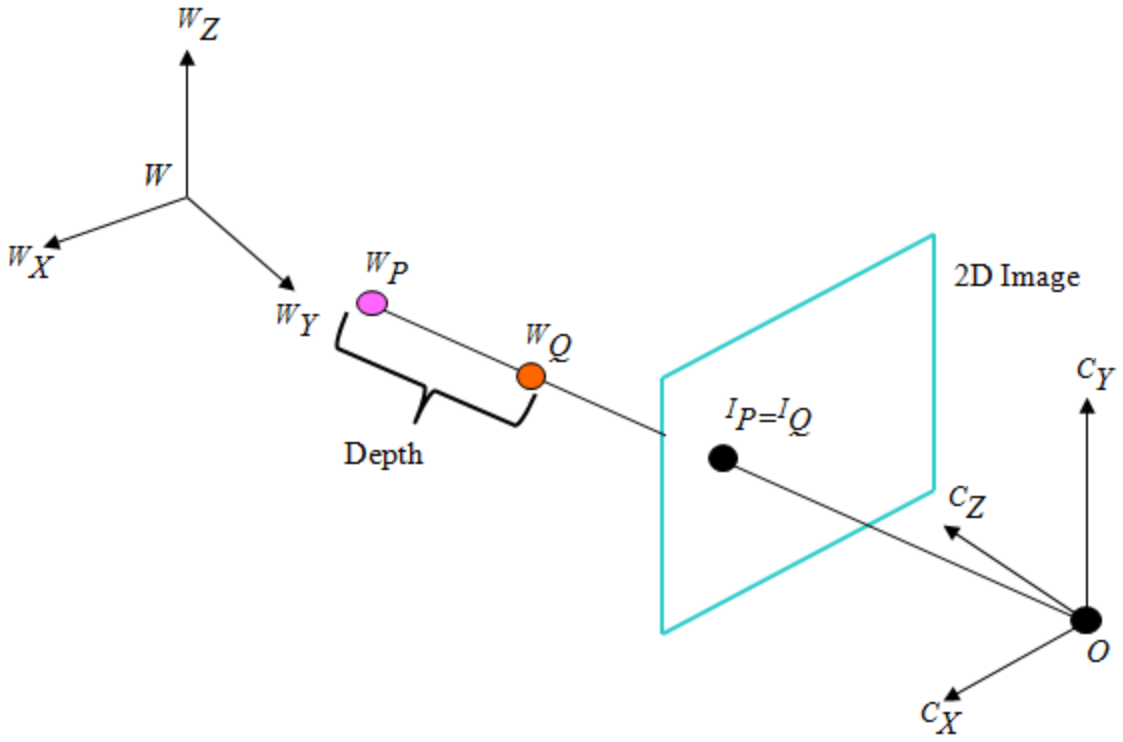


Figure B.1

Loss of depth information in 2D - Caused by projection of 3D points on same viewing line onto 2D image.

The 3D information can be fully recovered using two 2D images of the same scene with slightly different views. Figure B.2 shows how ${}^{W}P$ and ${}^{W}Q$ can be recovered when the same points in Figure B.1 are viewed by another camera at a slightly different position. The setting

represents stereo vision. O_1 and O_2 are the optical centers of the two cameras and the relative pose of both cameras is independent of each other.

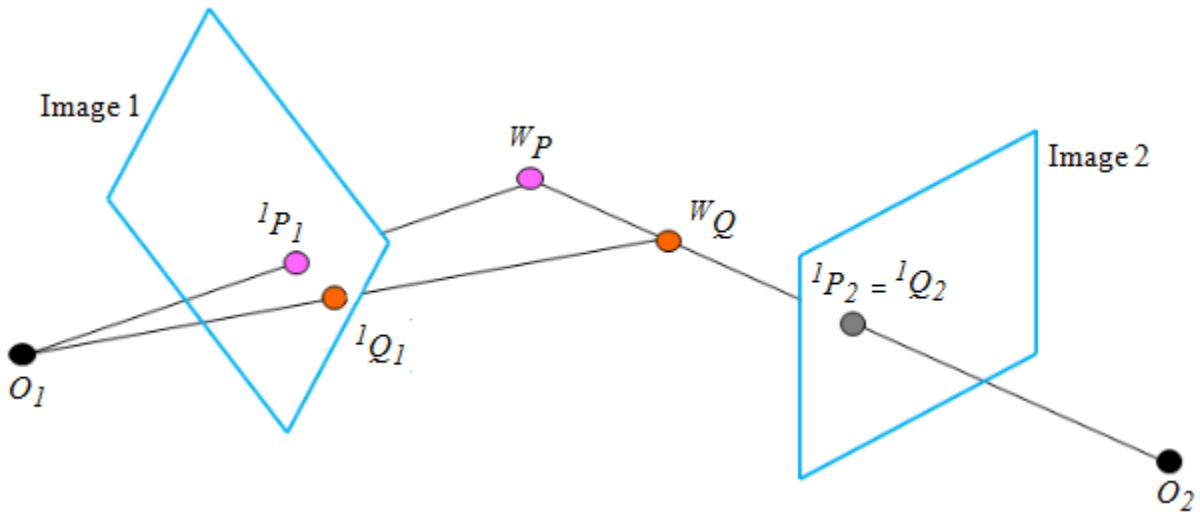


Figure B.2

Recovering 3D point coordinates using stereo vision.

Now consider that both points ^{W}P and ^{W}Q are not on the same viewing line of camera *1* and camera *2*. To perform stereo computation there is a need of identifying 2D projections of ^{W}P and ^{W}Q in image *1* ($^{I}P_1$, $^{I}Q_1$) which can be identified as the same points in image *2* ($^{I}P_2$, $^{I}Q_2$). This is known as the correspondence problem as shown in Figure B.3. The geometric relationship between the 3D world points and the 2D projections is known as the epipolar geometry and is explained below. Epipolar constraints can be used to solve the correspondence problem.

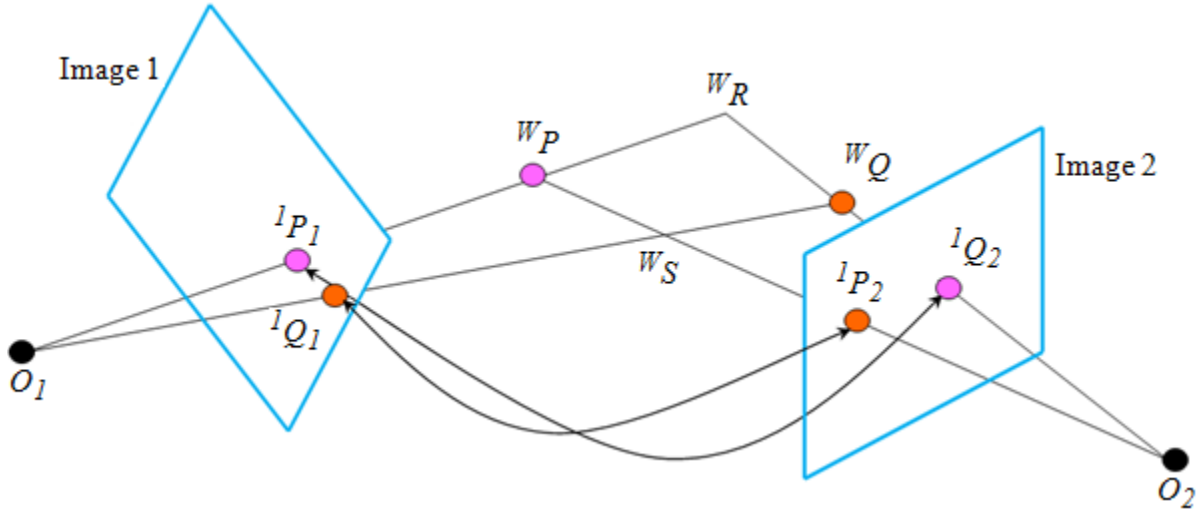


Figure B.3

Stereo correspondence problem: which points in Image 1 actually correspond to points ^{W}P *and* ^{W}Q *in Image 2?*

The projection of one camera's optical center into the image of the other camera is called the *epipole*. Figure B.4 shows the epipolar geometry. ^{W}P here represents the point of interest in both cameras. Points $^{I}P_1$ and $^{I}P_2$ are the projections of point ^{W}P onto the left and right image planes respectively. The projection of O_2 on the image *1* plane is the left epipole e_1, similarly projection of O_1 on the image *2* is the right epipole e_2. The plane defined by ^{W}P, O_1, O_2 is known as the epipolar plane. The ray O_1 ^{W}P is seen by the camera *1* as a point because it is directly in line with the camera's optical center. However, camera *2* sees this ray as a line in its image plane. That line e_2 $^{I}P_2$ in camera *2* is called an epipolar line. In other words intersection

of the epipolar plane with the image plane represents the epipolar line. It is the property of the system that all epipolar lines should go through the camera's epipole.

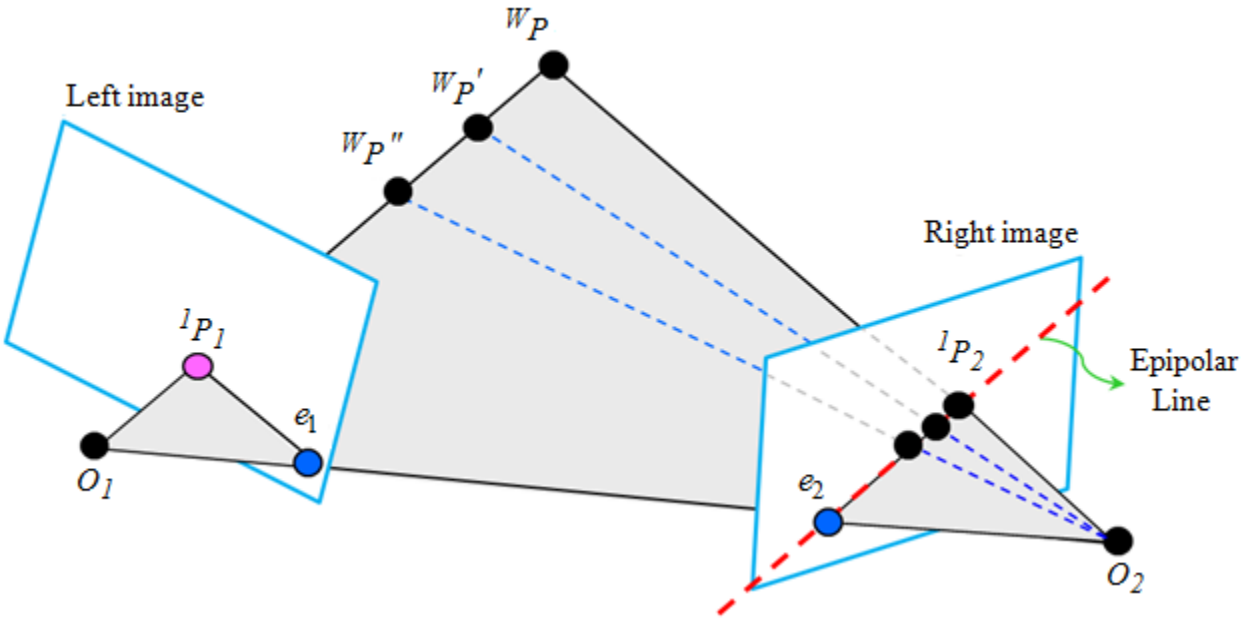


Figure B.4

Epipolar geometry.

Given an image point $^{I}P_1$, ^{W}P can lie anywhere on the ray from O_1 through ^{W}P. To establish the epipolar constraint the correct match of $^{I}P_1$ must lie on the epipolar line on right image. The search for correspondences is reduced to a 1D problem. This makes the epipolar constraint effective in rejecting false matches due to occlusion. Conjugate points along corresponding epipolar lines have the same order in each image. However, ordering is not a hard constraint because corresponding points may not have the same order if they lie on the same epipolar plane and imaged from different sides. Once the correspondence problem is solved then using cameras

transformation matrices the 3D coordinates can be recovered and the object model can be reconstructed as show in Figure B.5.

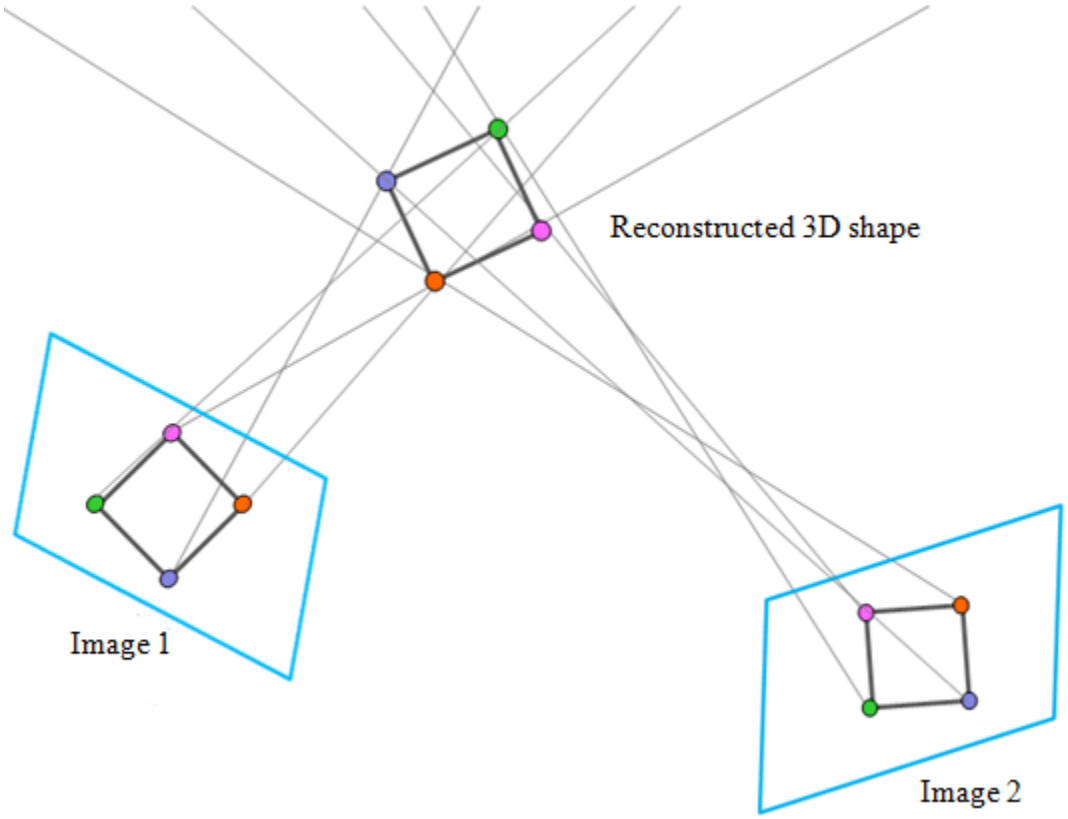


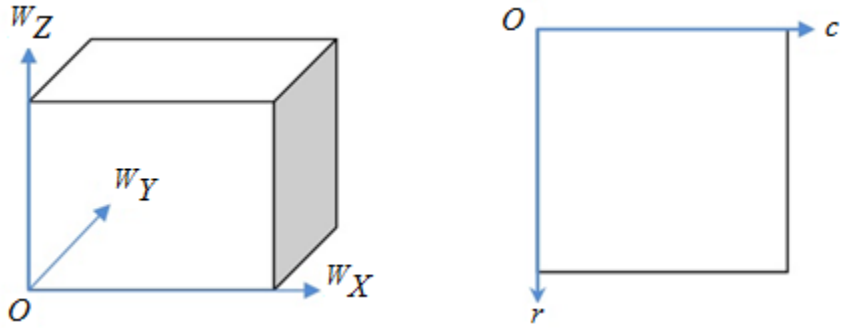
Figure B.5

3D reconstruction using 2D image points.

The camera transformation matrices are obtained by calibrating the cameras by the 3D world. The section below explains the camera calibration procedure that we have used.

B.2 Camera calibration

The coordinate system used in the affine camera calibration procedure [31] is shown below in Figure B.6.



(a) *(b)*

Figure B.6

Coordinate system used in camera calibration: (a) 3-D world (b) camera.

We define here the transformation formula to project every point of 3D model to camera image coordinates. ${}^{I}P$ represents here the image coordinates, C is the calibration matrix and ${}^{W}P$ is the world point.

$$ {}^{I}P = {}^{I}_{W}C \; {}^{W}P \tag{B.1} $$

$$ \begin{bmatrix} s\,{}^{I}P^{r} \\ s\,{}^{I}P^{c} \\ s \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & 1 \end{bmatrix} \begin{bmatrix} {}^{W}P^{x} \\ {}^{W}P^{y} \\ {}^{W}P^{z} \\ 1 \end{bmatrix} \tag{B.2} $$

In the matrix above, the parameter *s* is the scale factor which is used to adjust the pixel position according to the unit difference. To calculate the *11* parameters in the transformation matrix following derived formula need to be used, and then input the set of corresponding points from 3D world coordinates and the camera images from the *stereo pair*. Repeated experiments show that at least eight (*i.e* $i \geq 8$) 3D calibration points in most cases are required to provide good results.

$$\begin{bmatrix} W_{P_i}^{x}, W_{P_i}^{y}, W_{P_i}^{z}, 1, 0, 0, 0, 0, -W_{P_i}^{x} \times I_{P_i}^{r}, -W_{P_i}^{y} \times I_{P_i}^{r}, -W_{P_i}^{z} \times I_{P_i}^{r} \\ 0, 0, 0, 0, W_{P_i}^{x}, W_{P_i}^{y}, W_{P_i}^{z}, 1, -W_{P_i}^{y} \times I_{P_i}^{c}, -W_{P_i}^{y} \times I_{P_i}^{c}, -W_{P_i}^{z} \times I_{P_i}^{c} \end{bmatrix} \begin{bmatrix} c_{11} \\ c_{12} \\ c_{13} \\ c_{14} \\ c_{21} \\ c_{22} \\ c_{23} \\ c_{24} \\ c_{31} \\ c_{32} \\ c_{33} \end{bmatrix} = \begin{bmatrix} I_{P_i}^{r} \\ I_{P_i}^{c} \end{bmatrix} \quad \text{(B.3)}$$

Each input pair of points has two equations in the left matrix, and the size of the left matrix is *2×11*, so *least squares fit* method to calculate the *11* parameters can be used:

$$A_{2n \times 11} \; X_{11 \times 1} = B_{2n \times 1} \quad \text{(B.4)}$$

$$X_{11 \times 1} = \left(A^{T} \times A \right) \backslash \left(A^{T} \times B \right) \quad \text{(B.5)}$$

The '*X*' transform matrix, once calculated, can be recomposed as follows to get the camera calibration matrix for each camera:

$$\begin{bmatrix} c_{11} \\ c_{12} \\ c_{13} \\ c_{14} \\ c_{21} \\ c_{22} \\ c_{23} \\ c_{24} \\ c_{31} \\ c_{32} \\ c_{33} \end{bmatrix} \Rightarrow \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & 1 \end{bmatrix} \tag{B.6}$$

B.3 Computing 3D from 2D

Using the above camera model the real world 3D point $\left[{}^{W}P^{x}, {}^{W}P^{y}, {}^{W}P^{z} \right]$ can be calculated from the two images $\left[{}^{I}P_1^{r}, {}^{I}P_1^{c} \right]$ and $\left[{}^{I}P_2^{r}, {}^{I}P_2^{c} \right]$. This yields the following two camera models.

$$\begin{bmatrix} s\ {}^{I}P_1^{r} \\ s\ {}^{I}P_1^{c} \\ s \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & 1 \end{bmatrix} \begin{bmatrix} {}^{W}P^{x} \\ {}^{W}P^{y} \\ {}^{W}P^{z} \\ 1 \end{bmatrix} \tag{B.7}$$

$$\begin{bmatrix} t\ {}^{I}P_2^r \\ t\ {}^{I}P_2^c \\ t \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & 1 \end{bmatrix} \begin{bmatrix} {}^{W}P^x \\ {}^{W}P^y \\ {}^{W}P^z \\ 1 \end{bmatrix} \tag{B.8}$$

Eliminating the homogeneous coordinates *s* and *t* following 4 equations and 3 unknowns are obtained:

$$\begin{aligned} {}^{I}P_1^r &= \left(b_{11} - b_{31}\ {}^{I}P_1^r\right){}^{W}P^x + \left(b_{12} - b_{32}\ {}^{I}P_1^r\right){}^{W}P^y + \left(b_{13} - b_{33}\ {}^{I}P_1^r\right){}^{W}P^z + b_{14} \\ {}^{I}P_1^c &= \left(b_{21} - b_{31}\ {}^{I}P_1^c\right){}^{W}P^x + \left(b_{22} - b_{32}\ {}^{I}P_1^c\right){}^{W}P^y + \left(b_{23} - b_{33}\ {}^{I}P_1^c\right){}^{W}P^z + b_{24} \\ {}^{I}P_2^r &= \left(c_{11} - c_{31}\ {}^{I}P_2^r\right){}^{W}P^x + \left(c_{12} - c_{32}\ {}^{I}P_2^r\right){}^{W}P^y + \left(c_{13} - c_{33}\ {}^{I}P_2^r\right){}^{W}P^z + c_{14} \\ {}^{I}P_2^c &= \left(c_{21} - c_{31}\ {}^{I}P_2^c\right){}^{W}P^x + \left(c_{22} - c_{32}\ {}^{I}P_2^c\right){}^{W}P^y + \left(c_{23} - c_{33}\ {}^{I}P_2^c\right){}^{W}P^z + c_{24} \end{aligned} \tag{B.9}$$

$$\begin{bmatrix} {}^{I}P_1^r - b_{14} \\ {}^{I}P_1^c - b_{24} \\ {}^{I}P_2^r - c_{14} \\ {}^{I}P_2^c - c_{24} \end{bmatrix} = \begin{bmatrix} \left(b_{11} - b_{31}\ {}^{I}P_1^r\right) & \left(b_{12} - b_{32}\ {}^{I}P_1^r\right) & \left(b_{13} - b_{33}\ {}^{I}P_1^r\right) \\ \left(b_{21} - b_{31}\ {}^{I}P_1^c\right) & \left(b_{22} - b_{32}\ {}^{I}P_1^c\right) & \left(b_{23} - b_{33}\ {}^{I}P_1^c\right) \\ \left(c_{11} - c_{31}\ {}^{I}P_2^r\right) & \left(c_{12} - c_{32}\ {}^{I}P_2^r\right) & \left(c_{13} - c_{33}\ {}^{I}P_2^r\right) \\ \left(c_{21} - c_{31}\ {}^{I}P_2^c\right) & \left(c_{22} - c_{32}\ {}^{I}P_2^c\right) & \left(c_{23} - c_{33}\ {}^{I}P_2^c\right) \end{bmatrix} \begin{bmatrix} {}^{W}P^x \\ {}^{W}P^y \\ {}^{W}P^z \end{bmatrix} \tag{B.10}$$

$$\begin{bmatrix} {}^{W}P^x \\ {}^{W}P^y \\ {}^{W}P^z \end{bmatrix} = \begin{bmatrix} \left(b_{11} - b_{31}\ {}^{I}P_1^r\right) & \left(b_{12} - b_{32}\ {}^{I}P_1^r\right) & \left(b_{13} - b_{33}\ {}^{I}P_1^r\right) \\ \left(b_{21} - b_{31}\ {}^{I}P_1^c\right) & \left(b_{22} - b_{32}\ {}^{I}P_1^c\right) & \left(b_{23} - b_{33}\ {}^{I}P_1^c\right) \\ \left(c_{11} - c_{31}\ {}^{I}P_2^r\right) & \left(c_{12} - c_{32}\ {}^{I}P_2^r\right) & \left(c_{13} - c_{33}\ {}^{I}P_2^r\right) \\ \left(c_{21} - c_{31}\ {}^{I}P_2^c\right) & \left(c_{22} - c_{32}\ {}^{I}P_2^c\right) & \left(c_{23} - c_{33}\ {}^{I}P_2^c\right) \end{bmatrix} \backslash \begin{bmatrix} {}^{I}P_1^r - b_{14} \\ {}^{I}P_1^c - b_{24} \\ {}^{I}P_2^r - c_{14} \\ {}^{I}P_2^c - c_{24} \end{bmatrix} \tag{B.11}$$

Any *3* of these *4* equations can be solved to obtain the 3D world point $\left[{}^{W}P^{x}, {}^{W}P^{y}, {}^{W}P^{z}\right]$, however, due to approximation errors in the camera model and image points, each subset of three equations will yield slightly different coordinates for ${}^{W}P$.

These inaccuracies explained above once generated, amplifies as the ray propagates in space. Therefore we have used a more robust *shortest line segment approach* [95] as shown in Figure B.7. We have dropped the coordinate system symbols from the notation. The center of this line segment will represent the 3D point. So the smaller the segment better is the correspondence of image points and vice versa. We have also used this segment length criterion as a constraint to solve the correspondence problem. Epipolar constraints are also used in conjunction for robustness.

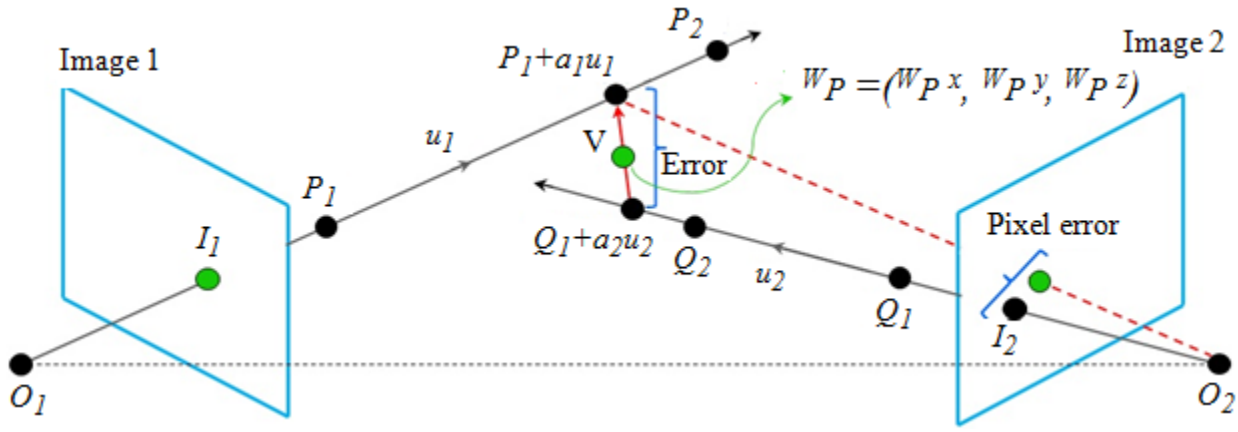


Figure B.7

Shortest line segment connecting the two skew rays.

P_1 and P_2 are the points on the ray originating from camera optical center O_1 and passing through image point I_1 while Q_1 and Q_2 are the points on the ray originating from camera optical center O_2 passing through image point I_2. If the optical center of the cameras is not known then camera *1* ray points can be computed using the two equations in Equation B.13 while choosing an arbitrary value of ${}^{W}P^{z} = z$. If the computed ray is parallel with the *z-axis* then the same procedure can be repeated for *y* and *z* while ${}^{W}P^{z} = x$ and so on. u_1 and u_2 are the unit vectors along these rays respectively. The shortest line segment is represented by vector V and is orthogonal to both u_1 and u_2 and is given as:

$$V = (P_1 + a_1u_1) - (Q_1 + a_2u_2) \tag{B.12}$$

The variables a_1 and a_2 can be computed using the following set of linear equations. Here '$\odot$' represents dot product:

$$\begin{aligned} [(P_1 + a_1u_1) - (Q_1 + a_2u_2)] \odot u_1 &= 0 \\ [(P_1 + a_1u_1) - (Q_1 + a_2u_2)] \odot u_2 &= 0 \end{aligned} \tag{B.13}$$

Rearranging Equations B.13:

$$\begin{aligned} [(P_1 - Q_1) + (a_1u_1 - a_2u_2)] \odot u_1 &= 0 \\ [(P_1 - Q_1) + (a_1u_1 - a_2u_2)] \odot u_2 &= 0 \end{aligned} \tag{B.14}$$

$$\begin{aligned} [(P_1 - Q_1)] \odot u_1 + [(a_1u_1 - a_2u_2)] \odot u_1 &= 0 \\ [(P_1 - Q_1)] \odot u_2 + [(a_1u_1 - a_2u_2)] \odot u_2 &= 0 \end{aligned} \tag{B.15}$$

$$\begin{aligned} [(P_1 - Q_1)] \odot u_1 + [(a_1.1)] - [(a_2u_2)] \odot u_1 &= 0 \\ [(P_1 - Q_1)] \odot u_2 + [(a_1u_1)] \odot u_2 - [(a_2.1)] &= 0 \end{aligned} \tag{B.16}$$

$$a_1 - a_2(u_1 \odot u_2) = -[(P_1 - Q_1)] \odot u_1 \tag{B.17}$$

$$-a_2 + a_1(u_1 \odot u_2) = -[(P_1 - Q_1)] \odot u_2 \tag{B.18}$$

Solving Equation B.17 and B.18 further to get a_1 and a_2. Multiply Equation B.18 by $(u_1 \odot u_2)$ and subtract from Equation B.17:

$$a_1[1 - (u_1 \odot u_2)^2] = [(Q_1 - P_1)] \odot u_1 - [(Q_1 - P_1) \odot u_2](u_1 \odot u_2) \tag{B.19}$$

$$a_1 = \frac{[(Q_1 - P_1)] \odot u_1 - [(Q_1 - P_1) \odot u_2](u_1 \odot u_2)}{[1 - (u_1 \odot u_2)^2]} \tag{B.20}$$

Multiply Equation B.17 by $(u_1 \odot u_2)$ and subtract Equation B.18 from Equation B.17:

$$a_2[(u_1 \odot u_2)^2 - 1] = [(Q_1 - P_1)] \odot u_2 - [(Q_1 - P_1) \odot u_1](u_1 \odot u_2) \tag{B.21}$$

$$a_2 = \frac{[(Q_1 - P_1) \odot u_1](u_1 \odot u_2) - [(Q_1 - P_1)] \odot u_2}{[1 - (u_1 \odot u_2)^2]} \tag{B.22}$$

If the magnitude of vector *V* is less than a desired threshold then the 3D world coordinates *x, y, z* of the point ${}^{W}P$ are given as the midpoint of V:

$${}^{W}P = \frac{1}{2}[(P_1 + a_1 u_1) + (Q_1 + a_2 u_2)] \tag{B.23}$$

APPENDIX C

Site survey details

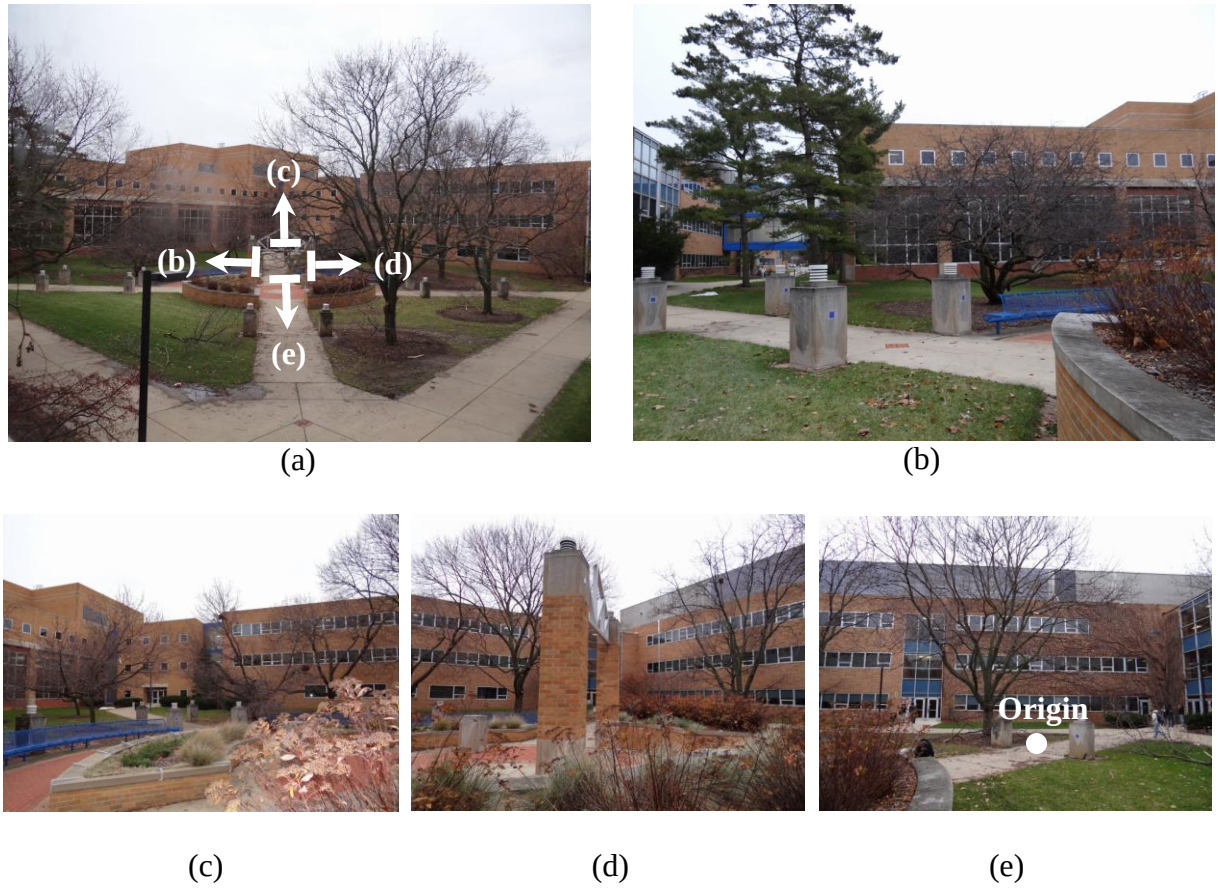
This appendix provides information about the outdoor test site that we have used in our experiments. Some extra pictures of the test site are also provided to elaborate the local dynamics of the test environment. We have also briefly explained the survey procedure used to acquire the survey data.

We have selected the MSU Engineering building courtyard as our outdoor test site. Figure C.1 shows the satellite view of the Engineering building obtained by *Google Earth*.

Figure C.1

MSU Engineering building satellite view.

The Engineering building latitude and longitude are 42.72477, -84.481594 respectively. For object localization and tracking the courtyard possesses complex semi-indoor features due to surrounding walls, trees, different sized pillars and arch structure in the middle of the courtyard. For our experiments the ground was considered to be leveled. Figure C.2(a) shows the aerial view of the courtyard. To have better insight of the local structures Figure C.2(b) to Figure C.2(e) provide local images. The approximate position and direction where these pictures were taken is highlighted in Figure C.2(a).



(a) (b)

(c) (d) (e)

Figure C.2

Different views of the courtyard.

We started survey of the site using the *total station* provided courtesy of the MSU Civil Engineering department. Figure C.3 shows the similar equipment used in the survey. The location where we placed the total station was selected carefully so that most of the points are in direct line of sight of the total station. The equipment location was selected as the origin of the 3D coordinate system.

Figure C.3

Total station surveying equipment - Image extracted from [117].

We have used a right hand side coordinate system as shown in Figure C.4. We started the survey by leveling the equipment. The height of the equipment once leveled was adjusted at *4.8 ft.* The goal was to obtain the coordinate information for all the possible corners that will be helpful to design a simulated 3D model of the test environment. The angles and distances of the predefined points from the total station were acquired. The relative position of the points from the origin was then calculated using trigonometry. We also used laser meters, tape measurements and ranging poles to obtain detail for the points not visible to the total station. This was also helpful to validate data obtained from the total station. Later the acquired survey data was

imported to MATLAB for making scaled model of the test site that were then used in simulations and experiments. The data was obtained in feet. To remain in sync with the unit system used in the lab experiments we later converted it to inches during the simulation.

Figure C.4

Top view of the outdoor test site with legend showing equipment position and the coordinate system.

The RFID readers were placed at a *40×40 m* space. In our experiments, to assess the performance of stereo system we selected two different positions which provided varying choice of near and far field calibration points. The landmarks for camera positions, RFID reader

positions, origin, test site center and the coordinate system are mentioned in Figure C.4. Figure C.5 shows the simulated 3D view of the outdoor test site in same orientation of Figure C.4. The location of the RFID readers and the cameras in position *2* are also shown.

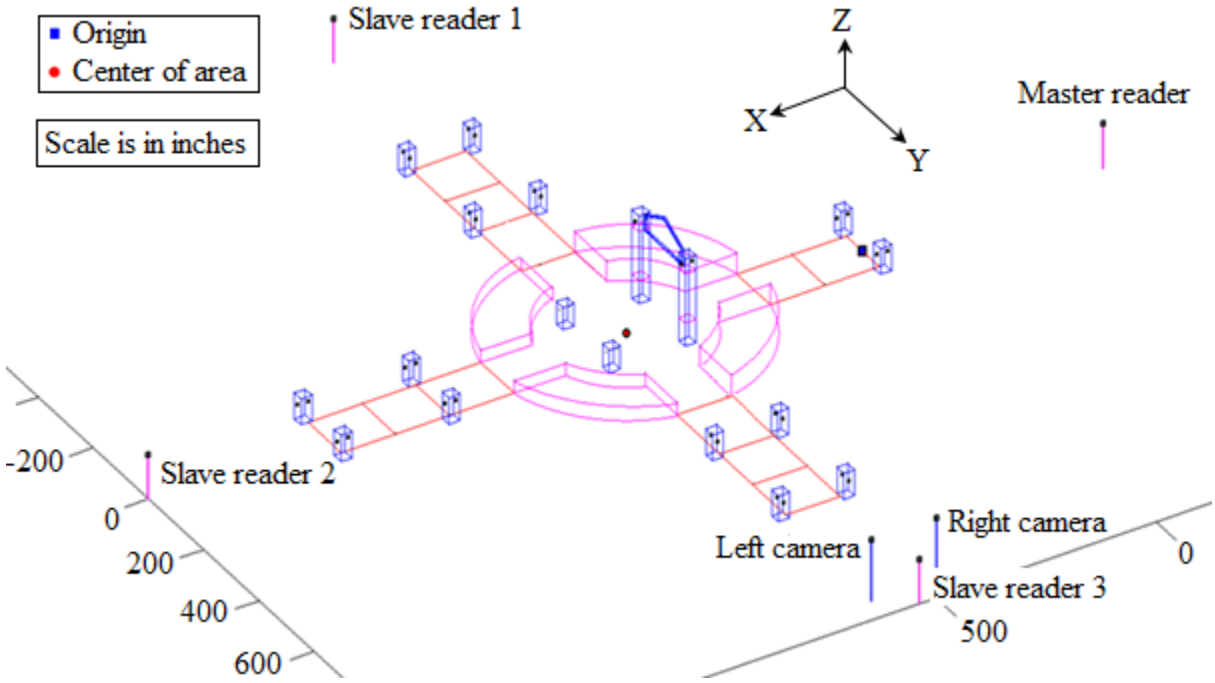
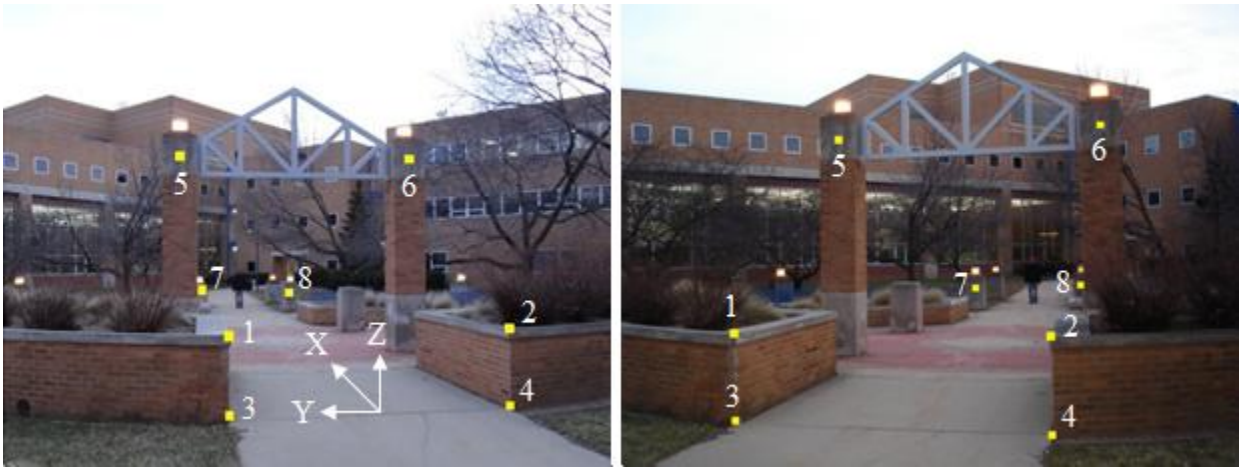


Figure C.5

3D view of the outdoor test site with sensor configuration - scale is in inches.

Figure C.6 shows labeled calibration points (shown by '■' in yellow) used during camera calibration.



(a) (b)

Figure C.6

Outdoor test site with 2D calibration points shown by '■' (yellow): (a) left image (b) right image.

Table C.1 provides coordinates of some of the important 3D landmarks in inches. The 3D ground truth points with respect to the 2D calibration points shown in Figure C.6 are also mentioned.

Table C.1

3D coordinates of calibration points and some important landmarks in MSU Engineering courtyard - scale is in inches.

Landmarks	*X*	*Y*	*Z*
Origin	*0*	*0*	*0*
Site center	*524*	*0*	*0*
Master reader	*-534*	*0*	*0*
Slave reader *1*	*524*	*-1061*	*0*
Slave reader *2*	*1585*	*0*	*0*
Slave reader *3*	*524*	*1061*	*0*
Left camera position *2*	*0*	*72*	*54*
Right camera position *2*	*0*	*-72*	*54*
Point *1*	*242*	*60*	*36*
Point *2*	*242*	*-60*	*36*
Point *3*	*242*	*60*	*0*
Point *4*	*242*	*-60*	*0*
Point *5*	*429*	*84*	*140*
Point *6*	*429*	*-84*	*140*
Point *7*	*945.5*	*72*	*30*
Point *8*	*945.5*	*-72*	*30*

APPENDIX D

Wireless location sensing

Wireless location sensing (WLS) methods have provided a new layer of automation to many indoor and outdoor location systems that need to know the physical location of objects and persons either relative to a known location or within a coordinate system. The process of WLS estimates the node location, where a node can be a smart phone, GPS receiver, wireless sensor, tag or a cellular base station. In this appendix we have briefly explained location system topologies, principles and methods with a summary of recent updates in the location sensing technology.

Wireless location sensing is an umbrella term used for many type of indoor and outdoor location providing schemes and is interchangeably used with wireless positioning systems (WPS). The variety of location systems can broadly be categorized into global positioning system (GPS) and local positioning systems (LPS) [30]. GPS provides global position information whereas LPS provides relative position information. Some of the outdoor systems use cellular or satellite based positioning which are mostly line of sight based, while the indoor schemes use local positioning technologies such as Wireless Local Area Network (WLAN), cameras, bluetooth, sensor networks, Radio Frequency Identification (RFID), infrared

and ultrasonic etc. The wireless assisted Global Positioning System (GPS) and cellular base station linked with indoor mobile client can also be used for indoor localization. For example, locating an airport might best be done using GPS, but analyzing the behavior of a group of people waiting at an airport gate can be done in an LPS for just that gate or for just that airport. Several surveys covering broader local sensing aspects are available on the topic [25], [28], [29], [118], [119] and a handbook of location estimation is recently published by Zekavat and Buehrer [30]. In the ensuing paragraphs we have provided the basic information about WLS systems with recent advances and trends.

D.1 Location system topologies

The LPS can have following four different system topologies [120].

a. *Self positioning* - Such a system has receiver and measuring unit onboard the mobile object. It communicates with several geographically distributed transmitters with known locations and calculates its position accordingly. An inertial navigation system (INS) works on this phenomenon.

b. *Indirect remote positioning* - A self-positioning system sending position information to a mobile unit via a wireless link.

c. *Remote positioning* - The mobile transmitter onboard the tracked object communicates with fixed receivers. Based on the received signals the position of the mobile object is measured in a central unit.

d. *Indirect self positioning* - A remote positioning system sending position information to a mobile unit via a wireless link.

D.2 Location system principles

In a broader spectrum, location sensing has two general principles i.e triangulation and trilateration.

a. *Triangulation* - With triangulation multiple sensor nodes observe some other node. Triangulation can be done in 2D if two network nodes fixed in space can compute the heading to the moving receiver node (angle-side-angle). Figure D.1 shows the triangulation concept. Same concept applies in 3D, where three fixed nodes form a tetrahedron with the mobile object.

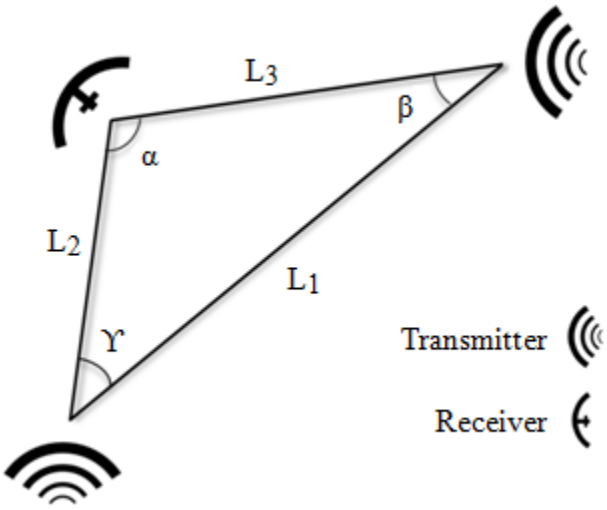


Figure D.1

Triangulation geometry.

The location of the receiver in Figure D.1 can be calculated as:

$$L_1^2 = L_2^2 + L_3^2 - 2L_2L_3\cos\alpha$$
$$\alpha = 180 - \beta - \gamma$$

b. Trilateration - Using trilateration a mobile node locates itself relative to other transmitting base nodes that are in known locations. Distance from each base node is computed from signal strength or timing. In a 2D space, the mobile node locates itself at the intersection of three circles whose radii are the sensed distances, and in 3D at the intersection of four spheres. Using more than the minimum number of base nodes enables more robust computation of location in real noisy environments.

D.3 Location methods

The location methods utilized in WLS are either geometric or time based. The geometric based techniques are *angle of arrival* (AOA), also called *direction of arrival* (DOA), *received signal strength indicator* (RSSI), or *phase of arrival* (POA); while propagation time based systems are *time of arrival* (TOA), also called *time of flight* (TOF), *time difference of arrival* (TDOA), and *round trip time of flight* (RTOF) [25], [118].

AOA estimates signal direction/angle at the desired point from at least two known reference points. RSSI calculates signal strength by comparing transmitted and received signal. This signal attenuation factor is then used to estimate range. The POA method estimates the received signal phase difference to get range estimates. The TOA/TOF measures one way signal travel time from a transmitter to receiver. For distance calculation the time measurements should at least be from three points of reference. TDOA on the other hand estimates difference in time at which the same signal arrives at multiple receivers instead of absolute arrival time. RTOF measures the two way signal travel time between the transmitter and receiver. Propagation time based systems are sensitive to the availability of line of sight (LOS) [121] and doesn't work well in mountainous terrain or around skyscrapers. However, non line of sight (NLOS) method such as RSSI is affected only slightly with lack of LOS. To improve position and tracking performance, location sensing technologies also use parameter estimators such as Kalman, Particle filter and Bayesian estimation.

D.4 Some location sensing system descriptions

There are a number of different implementation approaches that exist for the above systems. Some of them are given below.

D.4.1 Received Signal Strength based localization

An electronic fingerprint makes it possible to identify a wireless device by its unique radio transmission characteristics. Using spectrum analyzers, the RF location fingerprints of the scene are initially calculated. To estimate the position of an object, the observed measurements are compared with the fingerprint database. In a Wi-Fi environment, the RSS based algorithms mostly use Wireless Local Area Networks (WLAN) signatures for indoor localization. RADAR [122] was the first indoor system location and tracking system. Depending on the environment and application, RADAR and similar WLAN based location sensing systems provide tracking accuracy of one to three meters [123]. The accuracy of other typical WLAN positioning systems is approximately *3* to *30 m* [25]. Some of the recent work on WLAN based localization is presented in [124], [125], [126]. RSS based localization is classified as *RF fingerprinting*, *model based* and *kernel based*.

a. *RSS fingerprinting* - These methods [127] work in *two steps* i.e offline and online. RF signatures are captured in the off line mode and the database is generated. In online mode the location is estimated based on the database matching. Fingerprinting location techniques do not rely on LOS geometric assumptions.

b. *RSS Model based* - Model based RSS use a statistical model to generate the relationship between the RSS and distance [128], [129].

c. *RSS Kernel based* - Kernel based methods are statistical algorithms which provide the relationship between RSS and physical location using kernel functions [130].

D.4.2 Radiolocation using cellular signals

Position can also be estimated by cellular phones using measurements for the signal between different signal towers and the phone. Location sensing using cellular signals has a benefit that mobile existing hardware can be used and the system can potentially provide location estimates anywhere wireless service is available. Radiolocation through cellular telephony mainly includes techniques, such as cell identification, AOA, TOA/TDOA, Assisted GPS (AGPS) [131] and Enhanced Observed Time Difference (E-OTD). AGPS combines mobile technology and GPS. E-OTD measures the signal arrival time difference at handset, transmitted from minimum three synchronized base towers. United States Federal Communications Commission (FCC) for reliability, subscriber safety and quicker response, directed wireless carriers to provide automatic location identification [132] for the 911 emergency calls. Consequently there has been a wave of exploration in this area by the cellular services. For instance 2G GSM (Global System for Mobile communication) using E-OTD and 3G HSDPA (High Speed Downlink Packet Access) based wireless providers have integrated FCC positioning accuracy requirements in their systems. Universal Mobile Telecommunication System (UMTS) is a third generation technology for GSM

networks. The observed TDOA (O-TDOA) is considered as the UMTS version of E-OTD. CDMA (Code Division Multiple Access) based networks are utilizing TDOA and AGPS techniques for location based services.

There are several solutions reported for positioning using cellular phones [133], [134], [135], [136], [137], [138]. A large variety of smartphones in the market has also played an important role. Radiolocation from the cellular infrastructure can be achieved by handset based methods (upgraded handsets with GPS-based technology), network based methods, SIM based methods or hybrid.

a. *Handset based* - These methods require a client software running on the phones. Development of such client software with multi OS interface and cooperative mobile subscriber are some of main concerns in this approach.

b. *Network based* - Network based cellular localization method requires additions only in the provider's infrastructure. Its accuracy varies with the concentration of the signal towers and the timing method being used.

c. *SIM based* - Using the SIM it is possible to get cell ID, RSS and the RTOF measurements.

d. *Hybrid* - Hybrid systems use mixture of techniques, for example, network based technique can use GPS feed for validating location information. As cell sizes vary from

tens of meters in crowded urban areas to thousands of meters in rural area (having clear LOS), therefore, the location accuracy using cell ID varies. Fusing techniques such as TOA and TDOA with cell IDs can increase the accuracy. 2G GSM networks mostly use TDOA techniques however, for more accuracy AT&T now also utilize GPS feed for position estimation just like CDMA based networks.

FCC has specified that position estimation for *67%* of emergency calls should be within *50 m* for handset based and *100 m* for network based methods. Table D.1 from [138] compares the location accuracies in cellular phones using above mentioned technologies.

Table D.1

Location accuracies of cellular radiolocation technologies. See Kos et al. [138].

Type	*Rural*	*Suburban*	*Urban*	*Indoor*
Cell ID	*1-35 Km*	*1-10 Km*	*150-500 m*	*10-50 m*
E-OTD	-	*50-150 m*	*50-150 m*	good
AGPS	*10m*	*10-20 m*	*10-100 m*	variable

D.4.3 Localization using smart phone sensors

Increasing the number of embedded sensors such as Wi-Fi radio, cellular radio, accelerometer, gyroscope, compass, cameras, magnetometer, microphone, speakers and GPS in the cell phones presents new opportunities for logical localization. Using phone embedded hardware to determine RSSI fingerprints, Martin *et al.* has reported localization accuracy of *1.5*

m [139]. The accuracy is better in regions having more Wi-Fi radios in range. The authors in [136] have used smart phone compasses and accelerometers for localization without relying on Wi-Fi wireless networks with reported accuracy of around *11 m*. Location estimation is also done using photo-acoustic signatures such as sound, light and color (from microphone and camera) and user motion (from accelerometers) [140]. Overview of how AGPS provides better accuracy and cost is given in [131]. Peng *et al.* [104] provides an acoustic based ranging system using only the phone's microphone and speaker. The software based algorithm relies on two-way sensing, self recording and sample counting to estimate the location. Their system provides one to two centimeter accuracy in an area of about *10 m*. A similar approach in 3D without any infrastructure support has been reported in [141]. Their acoustic signatures are based on time of arrival and power level. Their system can provide localization accuracy of *13.9 cm* for *90%* of estimates when the phones are several meters apart. Kessel and Werner [142] evaluated location based services using deterministic 802.11 RSS fingerprinting and a digital compass on a smart phone. The reported position accuracy is *2.74 m* over an area of $250\ m^2$.

D.4.4 Sensor networks

Advancement in micro electro mechanical systems (MEMs) has enabled small size, low cost, low power wireless sensors possible. Sensor networks are generally used to monitor the environment but location based services and GPS positional accuracy can be combined. Solving location estimation using sensor networks faces challenges such as lack of central control system, computational capability, limited wireless bandwidth and high data traffic. In [143] the RSSI based location-tracking of an object in sensor networks was simulated by cooperation of

sensors through an election process and initiation of a mobile tracking agent. A mobile software agent is an intelligent program that follows an automated sequence of actions to track the target object. The system has prior knowledge of global and relative position information of each sensor. The mobile agent monitors the object by choosing the sensor closest to the object; i.e inviting nearby sensors and inhibiting irrelevant sensors. Each object is marked with its unique ID code by interpreting signal strengths from different sensors. The data overload issue was addressed by forwarding tracking histories to a location server. Recent research indicates that using low cost wireless sensors is an acceptable approach to scalable target tracking applications such as smart homes, fleet monitoring, air traffic control and security. In an indoor sensor network setup, an RMS location error of *1.2 m* for TOA and *2.2 m* RSSI are reported [144]. Some of the background work on sensor network localization methods is given in [145], [146], [147], [148].

D.4.5 Infrared positioning

The infrared positioning systems do not have reflection problems and are widely used for high accuracy applications such as virtual reality, games and computer graphics in the movie industry. One popular infrared camera based motion tracking system is provided by Vicon [93]. Its results, operating range and accuracy varies over different applications and environments, mainly due to camera placement, lighting conditions and volume location effects etc. The Vicon system is also being used by the group at University of Pennsylvania for controlling highly accurate maneuvers of cooperating flying robots [149]. Also, IR motion trackers are used for medical applications such as surgical navigation [150].

D.4.6 Ultrasonic trackers

Due to ultrasonic noise interference, ultrasonic trackers are more suitable for sound controlled areas such as indoor environments (offices, hospitals, labs etc.). Their low cost and good accuracy for small distances leverage their use in human movement analysis [151] and for robot collision avoidance and distance measurement [152], [153].

D.4.7 Laser range finders

Laser range finders (LRF) are also being used in position estimation systems especially in the field of robot navigation. They provide the estimate of how far is the closest obstruction from the robot. Wall mounted LRFs are also used in human tracking system with people wearing infrared tags [154]. The system merges multi-sensor information using Bayesian filter and perform identity estimation. Efficiency of LRFs is independent of the lighting conditions and provides accuracy within centimeters in controlled indoor environments [155], [156].

D.4.8 Magnetic motion trackers

Position and orientation information can also be obtained by magnetic motion trackers [157]. These systems generate magnetic pulses by the transmitter, which are then observed and reported by the magnetic receiver mounted. These sensors do not require LOS and are small and lightweight but have high cost and small range of operation (within $\sim 3\ m$ of the transmitter). Since the sensors can be affected by ferrous material and electricity [158], [159], the highest

accuracy is ensured in a controlled indoor environment where there is minimal magnetic distortion. The wide range trakSTAR system estimates *X, Y, Z* positional coordinates and orientation angles within *2.1 m* range from the transmitter with a single sensor static accuracy of *3.8 mm*. The system is used for human motion and activity capturing and analysis [160], biomechanics, simulations and computer graphics. The typical accuracy of a magnetic tracking system is less than *10 mm* [160]. Unlike other postion sensors, its permeability through human tissue allows tracking objects inside the human body and therefore they are used to track surgical equipment and drug delivery inside the human body [161].

D.4.9 Ultra Wide Band

Another localization technique uses Ultra Wide Band (UWB) signals. UWB wireless technology uses frequency spectrum larger than *500 MHz*. The UWB trackers have wall penetration capability, typical accuracy between *30-50 cm* in *10 m* working range (better than RF) and require low transmission power. However the system itself is costly. A commercially available UWB based tracking system is provided by Ubisense [22]. The system has tens of meter range and estimates 3D location of UWB moving tags. The company claims that the system provides *15 cm* accuracy *95%* of the time. UWB use in military applications and systems is given in [162].

D.4.10 Bluetooth

Bluetooth wireless networking technology can be used for location sensing; however, due to fewer transmitters and low scan/refresh rate it does not make an ideal choice for real time

location systems (RTLS). If sufficient transmitting beacons are available then typically Bluetooth can provide up to *10 m* accuracy [163]. Authors in [163] used a combination of WLAN and Bluetooth technologies to improve the location accuracy. Purely Bluetooth RSSI based indoor position estimation is reported in [164].

D.4.11 Inertial Measuring Units (INS)

For outdoor localization and tracking INS is being used in airplanes, submarines, shuttles, spacecrafts and unarmed vehicles (UAVs). By virtue of micro-electro-mechanical systems (MEMs) the smaller version of these systems have now made their place as a position and orientation estimator in object location and tracking. *Inertial Measuring Units* (IMU) are the main component of INS. The IMUs consist of gyroscopes and accelerometers and they provide position accuracy of around *10 m* without the requirement of LOS. An autonomous positioning system having IMU as a system component is explained in [165], which can locate and track the firefighter's position during rescue operations. An IMU integrated with GPS allows the GPS feed to continue in case of GPS signal loss. Integrating information from IMUs and marker based video tracking, a system for 3D indoor location tracking is provided in [166].

D.4.12 Miscellaneous

One of the system examples for RTOF based position estimation is Siemens local positioning radar [167]. The system is claimed to provide an accuracy of a few centimeters. To locate the position of the object earlier systems such as the Active badge system [168], Cyberguide [169]

used infrared and CricketNav [170], ActiveBat [171] used ultrasound. The active badge system and CricketNav estimate the room or portion of a room where the device is located. The ActiveBat system provides accuracy of *9 cm 95%* of the time in a *100* m^2 area. These technologies however, suffer from LOS restriction and require large amount of extra hardware to be installed.

REFERENCES

REFERENCES

[1] S. G. Pratt, D. E. Frosbroke, and S. M. Marsh, "Building safer highway work zones: measures to prevent worker injuries from vehicles and equipment," Center for Disease Control and Prevention 2001.

[2] J. Teizer, M. Venugopal, and A. Walia, "Ultrawideband for Automated Real-Time Three-Dimensional Location Sensing for Workforce, Equipment, and Material Positioning and Tracking," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2081, pp. 56-64, 2008.

[3] D. E. Fosbroke, "Studies on heavy equipment blind spots and internal traffic control," in *Roadway Work Zone Safety & Health Conference*, Baltimore, MD, 2004.

[4] T. M. Ruff, "Monitoring blind spots - a major concern for haul trucks," *Engineering and Mining Journal,* vol. 202, pp. 17–26, 2001.

[5] J. Bohn and J. Teizer, "Benefits and Barriers of Construction Project Monitoring Using High-Resolution Automated Cameras," *Journal of Construction Engineering and Management,* vol. 136, pp. 632-640, 2010/06/01 2009.

[6] S. I. Nakagawa, K. I. Soh, S. i. Mine, and H. Saito, "Image systems using RFID tag positioning information," *NTT Technical Review Journal*, vol. 1, pp. 79-83, 2003.

[7] J. Banks, *RFID applied*: Wiley. com, 2007.

[8] "Convergence System Ltd. RTLS development kit," ed, 2013.

[9] RFID and Rail: Advanced Tracking Technology; An interview with RFID pioneer J. Landt [Online]. Available: http://www.railway-technology.com/features/feature1684/

[10] J. Crabtree. (1995) Advantage I-75 - Electronic Clearance Test Project. *Public Roads*.

[11] *Airport RFID Services*. Available: http://www.transcore.com/rfid#963767e05df8b6f90c45c200b5ef4fde

[12] *Track and locate* Available: http://www.rfidc.com/

[13] R. Want, *RFID Explained: A Primer on Radio Frequency Identification Technologies*, 2008.

[14] R. Want, "An introduction to RFID technology," *Pervasive Computing, IEEE,* vol. 5, pp. 25-33, 2006.

[15] K. Bonsor. *How E-Zpass works*. Available: http://auto.howstuffworks.com/e-zpass1.htm

[16] S. Ahson and M. Ilyas, *RFID handbook : applications, technology, security, and privacy*. Boca Raton: CRC Press, 2008.

[17] T. Deyle, C. C. Kemp, and M. S. Reynolds, "Probabilistic UHF RFID tag pose estimation with multiple antennas and a multipath RF propagation model," in *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, 2008, pp. 1379-1384.

[18] H. Xin, R. Janaswamy, and A. Ganz, "Scout: Outdoor Localization Using Active RFID Technology," in *Broadband Communications, Networks and Systems, 2006. BROADNETS 2006. 3rd International Conference on*, 2006, pp. 1-10.

[19] K. Chawla, G. Robins, and Z. Liuyi, "Object localization using RFID," in *Wireless Pervasive Computing (ISWPC), 2010 5th IEEE International Symposium on*, 2010, pp. 301-306.

[20] *Dallas Zoo Tracks Elephants Using CSL Real Time Location System*. Available: http://rfid.net/news/399-dallas-zoo-track-elephants-real-time-location-system

[21] *How to Install a Real Time Location System – RTLS* Available: http://rfid.net/basics/rtls/241-how-to-install-a-real-time-location-system-rtls

[22] *Ubisense research and development packages*. Available: http://www.ubisense.net/en/rtls-solutions/research-packages.html

[23] *AeroScout: Technology overview*. Available: http://www.aeroscout.com/technology

[24] R. Buik, "Gps guidance and automated steering renew interest in precision farming technique," *Trimble Navigation Limited. July*, pp. 1-10, 2006.

[25] M. Vossiek, L. Wiebking, P. Gulden, J. Wieghardt, C. Hoffmann, and P. Heide, "Wireless local positioning," *Microwave Magazine, IEEE*, vol. 4, pp. 77-86, 2003.

[26] S. Gezici, "A survey on wireless position estimation," *Wireless Personal Communications*, vol. 44, pp. 263-282, 2008.

[27] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 37, pp. 1067-1080, 2007.

[28] T. Teixeira, G. Dublon, and A. Savvides, "A survey of human-sensing: Methods for detecting presence, count, location, track, and identity," *ACM Computing Surveys*, vol. 5, 2010.

[29] J. Raper, G. Gartner, H. Karimi, and C. Rizos, "Applications of location–based services: a selected review," *Journal of Location Based Services,* vol. 1, pp. 89-111, 2007.

[30] R. Zekavat and R. M. Buehrer, *Handbook of position location: Theory, practice and advances* vol. 27: Wiley. com, 2011.

[31] L. G. Shapiro and G. Stockman, *Computer Vision*. Upper Saddle River, NJ: Prentice Hall PTR, 2001.

[32] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian Detection: An Evaluation of the State of the Art," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 34, pp. 743-761, 2012.

[33] E. Charniak, *Introduction to artificial intelligence*: Pearson Education India, 1985.

[34] S. L. Tanimoto, *The elements of artificial intelligence using Common Lisp*: WH Freeman & Co., 1993.

[35] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM,* vol. 24, pp. 381-395, 1981.

[36] K. Ohmura, A. Tomono, and Y. Kobayashi, "Method Of Detecting Face Direction Using Image Processing For Human Interface," 1988, pp. 625-632.

[37] D. Colbry, G. Stockman, and A. Jain, "Detection of Anchor Points for 3D Face Verification," in *Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, 2005, pp. 118-118.

[38] G. Stockman, "Object representation for recognition-by-alignment," in *Object Representation in Computer Vision*, ed: Springer, 1995, pp. 77-87.

[39] I. Biederman, "Recognition by components: a theory of human image understanding," *Psychological review,* vol. 94, p. 115, 1987.

[40] G. Stockman, "Object Recognition, in Interpretation of Range Images," ed: R. Jain and A. Jain (Eds), 1989.

[41] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-D objects from appearance," *International journal of computer vision,* vol. 14, pp. 5-24, 1995.

[42] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on*, 1991, pp. 586-591.

[43] R. M. Bolle, J. H. Connell, N. Haas, R. Mohan, and G. Taubin, "VeggieVision: a produce recognition system," in *Applications of Computer Vision, 1996. WACV '96., Proceedings 3rd IEEE Workshop on*, 1996, pp. 244-251.

[44] S. M. Khan and M. Shah, "Tracking Multiple Occluding People by Localizing on Multiple Scene Planes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, pp. 505-519, 2009.

[45] C. Jin-Long and G. C. Stockman, "Determining pose of 3D objects with curved surfaces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 18, pp. 52-57, 1996.

[46] G. Stockman, "Object recognition and localization via pose clustering," *Computer Vision, Graphics, and Image Processing*, vol. 40, pp. 361-387, 1987.

[47] D. F. Huber and M. Hebert, "A new approach to 3-D terrain mapping," in *Intelligent Robots and Systems, 1999. IROS '99. Proceedings. 1999 IEEE/RSJ International Conference on*, 1999, pp. 1121-1127 vol.2.

[48] *Flying robots equipped with 3D gear*. Available: http://www.homelandsecuritynewswire.com/dr20120507-flying-robots-equipped-with-3d-gear-better-surveillance-on-the-cheap

[49] F. Goulette, F. Nashashibi, I. Abuhadrous, S. Ammoun, and C. Laurgeau, "An integrated on-board laser range sensing system for on-the-way city and road modelling," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 34, 2006.

[50] J. Teizer, "3D range imaging camera sensing for active safety in construction," *Electron. J. Inf. Technol. Constr.*, vol. 13, pp. 103-17, 2008.

[51] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes," in *Robotics-DL tentative*, 1992, pp. 586-606.

[52] S. Thrun, Y. Liu, D. Koller, A. Y. Ng, Z. Ghahramani, and H. Durrant-Whyte, "Simultaneous localization and mapping with sparse extended information filters," *The International Journal of Robotics Research*, vol. 23, pp. 693-716, 2004.

[53] Y. Mae, T. Umetani, T. Arai, and K. Inoue, "Object recognition using appearance models accumulated into environment," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, 2000, pp. 845-848 vol.4.

[54] M. Boukraa and S. Ando, "Tag-based vision: assisting 3D scene analysis with radio-frequency tags," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, 2002, pp. I-269-I-272 vol.1.

[55] I. Weiss and M. Ray, "Model-based recognition of 3D objects from single images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 23, pp. 116-128, 2001.

[56] M. Boukraa and S. Ando, "A computer vision system for knowledge-based 3D scene analysis using radio-frequency tags," in *Multimedia and Expo, 2002. ICME '02. Proceedings. 2002 IEEE International Conference on*, 2002, pp. 245-248 vol.2.

[57] C. Cerrada, S. Salamanca, E. Perez, J. A. Cerrada, and I. Abad, "Fusion of 3D Vision Techniques and RFID Technology for Object Recognition in Complex Scenes," in *Intelligent Signal Processing, 2007. WISP 2007. IEEE International Symposium on*, 2007, pp. 1-6.

[58] M. Adan, A. Adan, C. Cerrada, P. Merchan, and S. Salamanca, "Weighted cone-curvature: applications for 3D shapes similarity," in *3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. Fourth International Conference on*, 2003, pp. 458-465.

[59] C. Cerrada, S. Salamanca, A. Adan, E. Perez, J. A. Cerrada, and I. Abad, "Improved Method for Object Recognition in Complex Scenes by Fusioning 3-D Information and RFID Technology," *Instrumentation and Measurement, IEEE Transactions on,* vol. 58, pp. 3473-3480, 2009.

[60] H. Hontani, K. Baba, T. Kugimiya, K. Sato, and M. Nakagawa, "Visual tracking system using an ID-tag and the network," in *SICE 2003 Annual Conference*, 2003, pp. 2375-2380 Vol.3.

[61] H. Hontani, M. Nakagawa, T. Kugimiya, K. Baba, and M. Sato, "A visual tracking system using an RFID-tag," in *SICE 2004 Annual Conference*, 2004, pp. 2720-2723 vol. 3.

[62] O. Camps, P. J. Flynn, and G. C. Stockman, "Recent progress in CAD-based computer vision: an introduction to the special issue," *Computer Vision and Image Understanding,* vol. 69, pp. 251-252, 1998.

[63] C. Nak Young, H. Hongu, M. Miyazaki, K. Takemura, K. Ohara, K. Ohba, *et al.*, "Robots on self-organizing knowledge networks," in *Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on*, 2004, pp. 3494-3499 Vol.4.

[64] J.-Y. Kim, C.-J. Im, S.-W. Lee, and H.-G. Lee, "Object recognition using smart tag and stereo vision system on pan-tilt mechanism," in *Proceedings of International Conference on Computer Applications in Shipbuilding*, 2005, pp. 2379-2384.

[65] R. Poppe, "A survey on vision-based human action recognition," *Image and Vision Computing,* vol. 28, pp. 976-990, 2010.

[66] D. F. Hsu, Y.-S. Chung, and B. S. Kristal, "Combinatorial Fusion Analysis: Methods and Practices of Combining Multiple Scoring Systems," in *Advanced Data Mining Technologies in Bioinformatics*, ed: IGI Global, 2006, pp. 32-62.

[67] H.-H. Hsu, Z. Cheng, T. Huang, and Q. Han, "Behavior analysis with combined RFID and video information," presented at the Proceedings of the Third international conference on Ubiquitous Intelligence and Computing, Wuhan, China, 2006.

[68] N. Krahnstoever, J. Rittscher, P. Tu, K. Chean, and T. Tomlinson, "Activity Recognition using Visual Tracking and RFID," in *Application of Computer Vision, 2005. WACV/MOTIONS '05 Volume 1. Seventh IEEE Workshops on*, 2005, pp. 494-500.

[69] W. Jianxin, A. Osuntogun, T. Choudhury, M. Philipose, and J. M. Rehg, "A Scalable Approach to Activity Recognition based on Object Use," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, 2007, pp. 1-8.

[70] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, pp. 91-110, 2004.

[71] P. Sangho and H. Kautz, "Hierarchical recognition of activities of daily living using multi-scale, multi-perspective vision and RFID," in *Intelligent Environments, 2008 IET 4th International Conference on*, 2008, pp. 1-4.

[72] T. Deyle, C. Anderson, C. C. Kemp, and M. S. Reynolds, "A foveated passive UHF RFID system for mobile manipulation," in *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, 2008, pp. 3711-3716.

[73] T. Deyle, N. Hai, M. Reynolds, and C. C. Kemp, "RF vision: RFID receive signal strength indicator (RSSI) images for sensor fusion and mobile manipulation," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, 2009, pp. 5553-5560.

[74] A. Nemmaluri, M. D. Corner, and P. Shenoy, "Sherlock: automatically locating objects for humans," presented at the Proceedings of the 6th international conference on Mobile systems, applications, and services, Breckenridge, CO, USA, 2008.

[75] X. Liu, M. D. Corner, and P. Shenoy, "Ferret: RFID localization for pervasive multimedia," presented at the Proceedings of the 8th international conference on Ubiquitous Computing, Orange County, CA, 2006.

[76] T. Germa, F. Lerasle, N. Ouadah, V. Cadenat, and M. Devy, "Vision and RFID-based person tracking in crowds from a mobile robot," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, 2009, pp. 5591-5596.

[77] T. L. McDaniel, K. Kahol, D. Villanueva, and S. Panchanathan, "Integration of RFID and computer vision for remote object perception for individuals who are blind," presented at

the Proceedings of the 2008 Ambi-Sys workshop on Haptic user interfaces in ambient media systems, Quebec City, Canada, 2008.

[78] C. Heesung and H. Kyuseo, "Combination of RFID and Vision for Mobile Robot Localization," in *Intelligent Sensors, Sensor Networks and Information Processing Conference, 2005. Proceedings of the 2005 International Conference on*, 2005, pp. 75-80.

[79] L. Weiguo, J. Songmin, Y. Fei, and K. Takase, "Topological navigation of mobile robot using ID tag and WEB camera," in *Intelligent Mechatronics and Automation, 2004. Proceedings. 2004 International Conference on*, 2004, pp. 644-649.

[80] P. Kamol, S. Nikolaidis, R. Ueda, and T. Arai, "RFID Based Object Localization System Using Ceiling Cameras with Particle Filter," in *Future Generation Communication and Networking (FGCN 2007)*, 2007, pp. 37-42.

[81] J. Songmin, S. Erzhe, T. Abe, and K. Takase, "Localization of Mobile Robot with RFID Technology and Stereo Vision," in *Mechatronics and Automation, Proceedings of the 2006 IEEE International Conference on*, 2006, pp. 508-513.

[82] J. Songmin, S. Jinbuo, and K. Takase, "Obstacle recognition for a service mobile robot based on RFID with multi-antenna and stereo vision," in *Information and Automation, 2008. ICIA 2008. International Conference on*, 2008, pp. 125-130.

[83] J. Songmin, S. Jibuo, D. Chugo, and K. Takase, "Human recognition using RFID technology and sterero vision," in *Robotics and Biomimetics, 2007. ROBIO 2007. IEEE International Conference on*, 2007, pp. 1488-1493.

[84] J. Songmin, S. Jinbuo, and K. Takase, "Human recognition using RFID system with multi-antenna," in *Advanced Intelligent Mechatronics, 2008. AIM 2008. IEEE/ASME International Conference on*, 2008, pp. 1213-1218.

[85] Y. Po, W. Wenyan, M. Moniri, and C. C. Chibelushi, "RFID tag infrastructures for camera tracking in virtual studio environment," in *Visual Media Production, 2007. IETCVMP. 4th European Conference on*, 2007, pp. 1-8.

[86] W. Yu, J. Kato, Z. Wei, and S. Yokoi, "Digest Generation of Kindergarten Surveillance Video with Location Information and Visual Features," in *Innovative Computing, Information and Control (ICICIC), 2009 Fourth International Conference on*, 2009, pp. 768-771.

[87] F. Zoega, "Review of the Current State of Radio Frequency Identification (RFID) Technology, Its Use and Potential Future Use in Construction," 2006.

[88] J. Yang, O. Arif, P. A. Vela, J. Teizer, and Z. Shi, "Tracking multiple workers on construction sites using video cameras," *Advanced Engineering Informatics*, vol. 24, pp. 428-434, 2010.

[89] E. C. Jones, K. Kopocis, T. Wentz, R. Franca, and T. L. Stentz, "Measuring the Effectiveness of RFID on Mechanical Contracting Jobsites: A Practical Evaluation," University of Nebraska, LincolnNovember 28, 2007.

[90] R. H. Raza and G. C. Stockman, "Target tracking and surveillance by fusing stereo and RFID information," in *Proc. of SPIE Vol*, 2012, pp. 83921J-1.

[91] R. H. Raza and G. C. Stockman, "Fusion of stereo vision and RFID for site safety," in *Proceedings of 25th International conference on Computer Applications in Industry and Engineering*, New Orleans, Louisiana USA, 2012.

[92] N. Michael, J. Fink, and V. Kumar, "Cooperative manipulation and transportation with aerial robots," *Autonomous Robots*, pp. 1-14, 2009.

[93] *Vicon systems* Available: http://www.vicon.com

[94] *Shell game*. Available: http://en.wikipedia.org/wiki/Shell_game

[95] R. O. Duda and P. E. Hart, *Pattern classification and scene analysis*. New York,: Wiley, 1973.

[96] I. K. Sethi and R. Jain, "Finding Trajectories of Feature Points in a Monocular Image Sequence," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-9, pp. 56-73, 1987.

[97] C. J. Veenman, M. J. Reinders, and E. Backer, "Motion tracking as a constrained optimization problem," *Pattern Recognition*, vol. 36, pp. 2049-2067, 2003.

[98] *Queen Victoria building construction site, Melbourne*. Available: http://commons.wikimedia.org/wiki/File:QV_Building_construction_site,_Melbourne_-_March_2002.jpg

[99] N. Vaidya and S. R. Das, "Rfid-based networks: exploiting diversity and redundancy," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 12, pp. 2-14, 2008.

[100] X. Yonghong and J. Qiang, "A new efficient ellipse detection method," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, 2002, pp. 957-960 vol.2.

[101] S. T. Barnard and M. A. Fischler, "Computational stereo," *ACM Computing Surveys (CSUR)*, vol. 14, pp. 553-572, 1982.

[102] R. M. Haralock and L. G. Shapiro, *Computer and robot vision*: Addison-Wesley Longman Publishing Co., Inc., 1991.

[103] R. Jain, R. Kasturi, and B. G. Schunck, *Machine vision*. New York: McGraw-Hill, 1995.

[104] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan, "Beepbeep: a high accuracy acoustic ranging system using cots mobile devices," in *Proceedings of the 5th international conference on Embedded networked sensor systems*, 2007, pp. 1-14.

[105] R. Boudjemaa and A. B. Forbes, *Parameter Estimation Methods for Data Fusion*: National Physical Laboratory.Great Britain, Centre for Mathematics and Scientific Computing, 2004.

[106] H. F. Durrant-Whyte, "Sensor models and multisensor integration," *The International Journal of Robotics Research*, vol. 7, pp. 97-113, 1988.

[107] S. Houzelle and G. Giraudon, "Contribution to multisensor fusion formalization," *Robotics and autonomous systems*, vol. 13, pp. 69-85, 1994.

[108] K. Lai, B. Liefeng, R. Xiaofeng, and D. Fox, "A large-scale hierarchical multi-view RGB-D object dataset," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, 2011, pp. 1817-1824.

[109] *Detection, Inspection, and Enforcement*. Available: http://www.nist.gov/mml/mmsd/security_technologies/detection.cfm

[110] J. Walton and J. Crabtree, "A needs assessment and technology evaluation for roadside identification of commercial vehicles," *SAE transactions*, vol. 108, pp. 516-522, 1999.

[111] M. Maeterlinck, A. L. Teixeira de Mattos, and A. Sutro, *Joyzelle*. New York: Dodd, Mead and Company, 1907.

[112] *IP Cam Viewer Pro Application*. Available: https://itunes.apple.com/us/app/ip-cam-viewer-pro/id402656416

[113] B. K. P. Horn and B. G. Schunck, "Determining Optical Flow," Massachusetts Institute of Technology1980.

[114] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," presented at the Proceedings of the 7th international joint conference on Artificial intelligence - Volume 2, Vancouver, BC, Canada, 1981.

[115] *RWTH - Mindstorms NXT Toolbox for MATLAB*. Available: http://www.mindstorms.rwth-aachen.de/

[116] *Sensor Data Application*. Available: https://itunes.apple.com/us/app/sensor-data/id397619802

[117] *Total Stations*. Available: http://www.brandt.ca/SiteCollectionImages/Total-Stations/GTS-240NW/GTS-240NW.jpg

[118] L. Hui, H. Darabi, P. Banerjee, and L. Jing, "Survey of Wireless Indoor Positioning Techniques and Systems," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on,* vol. 37, pp. 1067-1080, 2007.

[119] S. Gezici, "A Survey on Wireless Position Estimation," *Wireless Personal Communications,* vol. 44, pp. 263-282-282, 2008.

[120] M. Vossiek, L. Wiebking, P. Gulden, J. Weighardt, and C. Hoffmann, "Wireless local positioning-concepts, solutions, applications," in *Radio and Wireless Conference, 2003. RAWCON'03. Proceedings,* 2003, pp. 219-224.

[121] W. Xu and S. Zekavat, "Spatially correlated multi-user channels: LOS vs. NLOS," in *Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop, 2009. DSP/SPE 2009. IEEE 13th,* 2009, pp. 308-313.

[122] P. Bahl and V. N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE,* 2000, pp. 775-784 vol.2.

[123] A. M. Ladd, K. E. Bekris, A. Rudys, L. E. Kavraki, and D. S. Wallach, "Robotics-based location sensing using wireless ethernet," *Wireless Networks,* vol. 11, pp. 189-204, 2005.

[124] F. Shih-Hau, L. Tsung-Nan, and L. Kun-Chou, "A Novel Algorithm for Multipath Fingerprinting in Indoor WLAN Environments," *Wireless Communications, IEEE Transactions on,* vol. 7, pp. 3579-3588, 2008.

[125] U. Grossmann, M. Schauch, and S. Hakobyan, "RSSI based WLAN Indoor Positioning with Personal Digital Assistants," in *Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, 2007. IDAACS 2007. 4th IEEE Workshop on,* 2007, pp. 653-656.

[126] X. Yubin, W. Yong, and M. Lin, "A Novel WLAN Indoor Positioning Algorithm Based on Positioning Characteristics Extraction," in *Genetic and Evolutionary Computing (ICGEC), 2010 Fourth International Conference on,* 2010, pp. 134-137.

[127] M. Brunato and C. Kiss Kallo, "Transparent location fingerprinting for wireless services," 2002.

[128] A. LaMarca, J. Hightower, I. Smith, and S. Consolvo, "Self-mapping in 802.11 location systems," in *UbiComp 2005: Ubiquitous Computing*, ed: Springer, 2005, pp. 87-104.

[129] Y. Ji, S. Biaz, S. Pandey, and P. Agrawal, "ARIADNE: a dynamic indoor signal map construction and localization system," in *Proceedings of the 4th international conference on Mobile systems, applications and services*, 2006, pp. 151-164.

[130] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, "LANDMARC: indoor location sensing using active RFID," *Wireless Networks*, vol. 10, pp. 701-710, 2004.

[131] G. M. Djuknic and R. E. Richton, "Geolocation and assisted GPS," *Computer*, vol. 34, pp. 123-125, 2001.

[132] *E911 Phase II Decision: Fact Sheet of FCC Wireless 911 Requirements*. Available: http://transition.fcc.gov/pshs/services/911-services/enhanced911/archives/factsheet_requirements_012001.pdf

[133] J. J. Caffery and G. L. Stuber, "Overview of radiolocation in CDMA cellular systems," *Communications Magazine, IEEE*, vol. 36, pp. 38-45, 1998.

[134] B. Ludden and L. Lopes, "Cellular based location technologies for UMTS: a comparison between IPDL and TA-IPDL," in *Vehicular Technology Conference Proceedings, 2000. VTC 2000-Spring Tokyo. 2000 IEEE 51st*, 2000, pp. 1348-1353 vol.2.

[135] I. K. Adusei, K. Kyamakya, and K. Jobmann, "Mobile positioning technologies in cellular networks: an evaluation of their performance metrics," in *MILCOM 2002. Proceedings*, 2002, pp. 1239-1244 vol.2.

[136] I. Constandache, R. R. Choudhury, and I. Rhee, "Towards Mobile Phone Localization without War-Driving," in *INFOCOM, 2010 Proceedings IEEE*, 2010, pp. 1-9.

[137] I. Constandache, X. Bao, M. Azizyan, and R. R. Choudhury, "Did you see Bob?: human localization using mobile phones," in *Proceedings of the sixteenth annual international conference on Mobile computing and networking*, 2010, pp. 149-160.

[138] T. Kos, M. Grgic, and J. Kitarovic, "Location Technologies for Mobile Networks," in *Systems, Signals and Image Processing, 2007 and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services. 14th International Workshop on*, 2007, pp. 319-322.

[139] E. Martin, O. Vinyals, G. Friedland, and R. Bajcsy, "Precise indoor localization using smart phones," in *Proceedings of the international conference on Multimedia*, 2010, pp. 787-790.

[140] M. Azizyan, I. Constandache, and R. Roy Choudhury, "SurroundSense: mobile phone localization via ambience fingerprinting," in *Proceedings of the 15th annual international conference on Mobile computing and networking*, 2009, pp. 261-272.

[141] J. Qiu, D. Chu, X. Meng, and T. Moscibroda, "On the feasibility of real-time phone-to-phone 3d localization," in *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*, 2011, pp. 190-203.

[142] M. Kessel and M. Werner, "SMARTPOS: Accurate and precise indoor positioning on mobile phones," in *MOBILITY 2011, The First International Conference on Mobile Services, Resources, and Users*, 2011, pp. 158-163.

[143] Y.-C. Tseng, S.-P. Kuo, H.-W. Lee, and C.-F. Huang, "Location tracking in a wireless sensor network by mobile agents and its data fusion strategies," *The Computer Journal*, vol. 47, pp. 448-460, 2004.

[144] N. Patwari and A. O. Hero, "Location estimation accuracy in wireless sensor networks," in *Signals, Systems and Computers, 2002. Conference Record of the Thirty-Sixth Asilomar Conference on*, 2002, pp. 1523-1527.

[145] S. Meguerdichian, F. Koushanfar, G. Qu, and M. Potkonjak, "Exposure in wireless ad-hoc sensor networks," in *Proceedings of the 7th annual international conference on Mobile computing and networking*, 2001, pp. 139-150.

[146] M. Rudafshani and S. Datta, "Localization in wireless sensor networks," presented at the Proceedings of the 6th international conference on Information processing in sensor networks, Cambridge, Massachusetts, USA, 2007.

[147] J. Ash and L. Potter, "Sensor network localization via received signal strength measurements with directional antennas," in *Proceedings of the 2004 Allerton Conference on Communication, Control, and Computing*, 2004, pp. 1861-1870.

[148] A. Boukerche, H. A. B. Oliveira, E. F. Nakamura, and A. A. F. Loureiro, "Localization systems for wireless sensor networks," *Wireless Communications, IEEE*, vol. 14, pp. 6-12, 2007.

[149] N. Michael, D. Mellinger, Q. Lindsey, and V. Kumar, "The GRASP Multiple Micro-UAV Testbed," *Robotics & Automation Magazine, IEEE*, vol. 17, pp. 56-65, 2010.

[150] Z. Ping, L. Yue, and W. Yongtian, "Multiple infrared markers based real-time stereo vision positioning system for surgical navigation," in *Instrumentation and Measurement Technology Conference, 2009. I2MTC '09. IEEE*, 2009, pp. 692-696.

[151] R. B. Huitema, A. L. Hof, and K. Postema, "Ultrasonic motion analysis system—measurement of temporal and spatial gait parameters," *Journal of biomechanics*, vol. 35, pp. 837-842, 2002.

[152] L. Choon-Young, C. Ho-Gun, P. Jun-Sik, P. Keun-Young, and L. Sang-Ryong, "Collision Avoidance by the Fusion of Different Beam-width Ultrasonic Sensors," in *Sensors, 2007 IEEE*, 2007, pp. 985-988.

[153] G. Hueber, T. Ostermann, T. Bauernfeind, R. Raschhofer, and R. Hagelauer, "New approach of ultrasonic distance measurement technique in robot applications," in *Signal Processing Proceedings, 2000. WCCC-ICSP 2000. 5th International Conference on*, 2000, pp. 2066-2069 vol.3.

[154] D. Fox, J. Hightower, L. Lin, D. Schulz, and G. Borriello, "Bayesian filtering for location estimation," *Pervasive Computing, IEEE,* vol. 2, pp. 24-33, 2003.

[155] D. Fox, W. Burgard, and S. Thrun, "Active markov localization for mobile robots," *Robotics and autonomous systems,* vol. 25, pp. 195-207, 1998.

[156] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "FastSLAM: A factored solution to the simultaneous localization and mapping problem," in *AAAI/IAAI,* 2002, pp. 593-598.

[157] *trakSTAR by Ascension Technology Corporation*. Available: http://www.ascension-tech.com/realtime/RTtrakSTAR.php

[158] E. R. Bachmann, X. Yun, and A. Brumfield, "Limitations of Attitude Estimnation Algorithms for Inertial/Magnetic Sensor Modules," *Robotics & Automation Magazine, IEEE,* vol. 14, pp. 76-87, 2007.

[159] J. Hummel, M. Figl, C. Kollmann, H. Bergmann, and W. Birkfellner, "Evaluation of a miniature electromagnetic position tracker," *Medical physics,* vol. 29, p. 2205, 2002.

[160] J. F. O'Brien, R. E. Bodenheimer Jr, G. J. Brostow, and J. K. Hodgins, "Automatic joint parameter estimation from magnetic motion capture data," 1999.

[161] C. Tercero, S. Ikeda, T. Uchiyama, T. Fukuda, F. Arai, Y. Okada, *et al.*, "Autonomous catheter insertion system using magnetic motion capture sensor for endovascular surgery," *The International Journal of Medical Robotics and Computer Assisted Surgery,* vol. 3, pp. 52-58, 2007.

[162] R. J. Fontana, "Recent applications of ultra wideband radar and communications systems," in *Ultra-Wideband, Short-Pulse Electromagnetics 5*, ed: Springer, 2002, pp. 225-234.

[163] A. LaMarca, Y. Chawathe, S. Consolvo, J. Hightower, I. Smith, J. Scott, *et al.*, "Place lab: Device positioning using radio beacons in the wild," in *Pervasive Computing*, ed: Springer, 2005, pp. 116-133.

[164] S. Feldmann, K. Kyamakya, A. Zapater, and Z. Lue, "An Indoor Bluetooth-Based Positioning System: Concept, Implementation and Experimental Evaluation," in *International Conference on Wireless Networks*, 2003, pp. 109-113.

[165] Y. Suh, "Development of an INS integrated autonomous positioning system for assisting effective fire-fighting activity," *KSCE Journal of Civil Engineering,* vol. 8, pp. 569-574, 2004/09/01 2004.

[166] B. Hartmann, N. Link, and G. F. Trommer, "Indoor 3D position estimation using low-cost inertial sensors and marker-based video-tracking," in *Position Location and Navigation Symposium (PLANS), 2010 IEEE/ION*, 2010, pp. 319-326.

[167] L. Wiebking, M. Glanzer, D. Mastela, M. Christmann, and M. Vossiek, "Remote local positioning radar," in *Radio and Wireless Conference, 2004 IEEE*, 2004, pp. 191-194.

[168] R. Want, A. Hopper, V. Falcão, and J. Gibbons, "The active badge location system," *ACM Transactions on Information Systems (TOIS),* vol. 10, pp. 91-102, 1992.

[169] G. D. Abowd, C. G. Atkeson, J. Hong, S. Long, R. Kooper, and M. Pinkerton, "Cyberguide: A mobile context-aware tour guide," *Wireless Networks,* vol. 3, pp. 421-433, 1997.

[170] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The cricket location-support system," in *Proceedings of the 6th annual international conference on Mobile computing and networking*, 2000, pp. 32-43.

[171] A. Harter, A. Hopper, P. Steggles, A. Ward, and P. Webster, "The anatomy of a context-aware application," *Wireless Networks,* vol. 8, pp. 187-197, 2002.