

NISTIR 8045

Performance Evaluation Methods for Human Detection and Tracking Systems for Robotic Applications

Michael Shneier
Tsai Hong
Geraldine Cheok
Kamel Saidi
Will Shackleford

This publication is available free of charge from:
<http://dx.doi.org/10.6028/NIST.IR.8045>

NIST
**National Institute of
Standards and Technology**
U.S. Department of Commerce

NISTIR 8045

Performance Evaluation Methods for Human Detection and Tracking Systems for Robotic Applications

Michael Shneier

Tsai Hong

Geraldine Cheok

Kamel Saidi

Will Shackleford

Intelligent Systems Division

Engineering Laboratory

This publication is available free of charge from:
<http://dx.doi.org/10.6028/NIST.IR.8045>

March 2015



U.S. Department of Commerce
Penny Pritzker, Secretary

National Institute of Standards and Technology
Willie May, Acting Under Secretary of Commerce for Standards and Technology and Acting Director

Abstract

Detecting and tracking people is becoming more important in robotic applications because of the increasing demand for collaborative work in which people interact closely with and in the same workspace as robots. New safety standards allow people to work next to robots but require that they be protected from harm while they do so. Sensors that detect and track people are a natural way of implementing the necessary safety monitoring and have the added advantage that the information about where the people are and where they are going can be fed back into the application and used to give the robot greater situational awareness. This report describes work on defining and testing performance evaluation measures that provide quantitative information about how well a human detection and tracking system performs. The results should help users determine if such a system will provide sufficient protection for people to be able to work safely in collaborative applications with industrial robots.

Keywords: Human detection; human tracking; performance evaluation; performance metrics; robot safety; human-robot collaboration.

Disclaimer: Commercial equipment and materials are identified in order to adequately specify certain procedures. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

1. Introduction

For most of their history, industrial robots have been separated from people for reasons of safety. There are many disadvantages of this isolation including additional costs for fences and safety sensors, the need for extra floor space, and the difficulty of moving material through the facility. Until recently, however, both the safety standards and the capabilities of human detection sensors were barriers preventing people from working safely in close proximity to industrial robots. In the last few years the robot safety standards [1-3] have been modified to specify requirements to allow humans and robots to work together in limited circumstances.

Simultaneously, there has been a significant improvement in the capabilities of human detection and tracking systems, mainly due to their greatly increased use in surveillance. The needs of surveillance applications, however, are very different from those of safety. Losing track of a person for a short time is usually not a cause for concern in surveillance applications, whereas losing track in a safety situation is not acceptable. Similarly, reporting the presence of a person when there is nobody visible is usually not of major consequence for surveillance, whereas doing the same in a manufacturing environment may result in stopping production and causing unacceptable delays and losses.

The question that arises is whether or not current human-detection and tracking systems are reliable enough to be used in safety applications. The aim of the work described in this report is not to resolve the question, since the answer depends on the planned application. Rather, the goal is to provide a methodology and set of performance measures to enable potential users to decide for themselves if their application can be safely implemented using human detection sensors.

In most previous work on performance evaluation for human detection, ground truth (GT) was established by manual annotation of the imagery, either by drawing the outline of the target objects or by placing bounding boxes around them. The capabilities of the detection algorithms were then measured by the amount of overlap between the annotations and the algorithm being evaluated. The current work takes a different approach to obtaining ground truth. Two independent measurement systems are employed to provide the actual locations and motion tracks of the people and a number of performance measures are used to measure the Euclidean distances in three-dimensional space between the locations detected by the system under test (SUT) and the ground truth instrument. For brevity, the results are shown only for one of the ground truth systems.

Section 2 of this report refers to previous work on human and object detection and tracking. Some of the performance measures used in this report are taken from that work. Section 3 describes the ground truth systems and how the information is transformed into a common coordinate system for comparison. Sections 4 and 5 describe two separate sets of performance measures and their results. One of the sets of measures is designed specifically for safety applications. The report ends with a discussion and conclusions.

2. Related Work

Detecting and tracking people and objects has a long history, including a number of projects that focused on performance evaluation. Ogale [4] provides a survey of video-based human detection. The PETS (Performance Evaluation of Tracking and Surveillance) workshops (Ferryman and Crowley [5]) focused on algorithm development and performance evaluation of tasks such as multiple object detection, event detection, and recognition. Nascimento and Marques [6], proposed a way to evaluate the performance of object detection systems by comparing algorithm results to ground-truth data and calculating performance metrics such as correct detections, false alarms, detection failure, and splitting and merging errors. CLEAR (Classification of Events, Activities and Relationships) (Stiefelhagen and Garofolo [7]) provides performance evaluation of people, faces, cars, and object tracking and ETISEO, Nghiem, et al. [8], was a video understanding and evaluation project for tracking systems that used an event detection algorithm. The Image Library for Intelligent Detection Systems (i-LIDS) [9, 10] is a United Kingdom government initiative that conducts performance evaluations of vision-based detection systems to ensure that they meet Government requirements. Other papers specific to tracking-based metrics are Brown, et al. [11], who suggest a motion tracking evaluation framework that estimates the number of True Positive, False Positive and False Negative, Merged, and Split trajectories. Yin, et al. [12] proposed a large set of metrics to assess different aspects of the performance of motion tracking and to help identify shortcomings of motion trackers under specific conditions. Lazarevic-McManus, et al. [13], developed a tracking metric to enable evaluation of motion detection based on Receiver Operating Characteristic (ROC)-like curves and the F-measure (a combination of precision and recall). Bashir and Porikli [14], presented metrics based on the spatial intersection of ground-truth and system-generated bounding boxes and then calculated a number of performance metrics, which they then averaged for all the sampled frames. Black, et al. [15], used synthetic video to evaluate tracking performance. They varied the scene complexity of the tracking task by adding occlusions and clutter and increasing the number of objects and people in the scene and presented results based on a number of metrics. Several other performance evaluation metrics were developed and discussed in [16-21].

The National Institute of Standards and Technology (NIST) has helped to develop performance metrics for object and human detection in a number of different applications, ranging from videoconferences through surveillance to counting and tracking people in stores and commercial establishments. NIST has worked with the United States (US) Department of Homeland Security, with the British Home Office, and with the European CHIL program (Computers in the Human Interaction Loop), [13], and the CLEAR evaluations, [22]. NIST has also worked with the US Army Collaborative Technology Alliance (CTA) on Robotics to evaluate systems that locate and track human pedestrians from a moving vehicle, Bodt, et al. [23].

The biggest difference between the current report and previous work relates to how performance measures make use of ground truth and, particularly, the use of three-dimensional space instead

of the image domain to compute the performance measures. Instead of using windows or manually-outlined regions in the image as ground truth, the work described in this report explicitly equips each person with sensors that provide identity and location information, measured with substantially higher accuracy than the system under test. Further, both the ground truth and system under test provide pose information in the three-dimensional world rather than the image. This is important because, for safety applications, it is critical to know precisely where in space a person is located in order to guarantee that they will not be endangered by their environment.

3. Measuring Ground Truth

Godil, et al. [24], describe four types of ground truth data for object detection and tracking: annotation/label-based systems that rely on humans to outline regions in the data, fixture-based systems that use physical positioning constraints to locate items precisely, physics-based simulations that depend on models of the objects and their locations in the world, and sensor-based systems that use independent means to sense the world and locate the objects, ideally much more precisely than the system being evaluated. This report adopts the last of these methods, making use of two sensor-based ground truth systems, described in the following subsections.

3.1. Indoor Global Positioning System

The Indoor Global Positioning System (iGPS), [25], is a high-resolution measurement system, shown in Figure 1, that can locate objects in a large volume of space and provide full six degree-of-freedom pose for multiple objects at the same time. The system uses stationary laser transmitters (Figure 1a) and receivers (Figure 1b) mounted on moving or static objects to determine the poses of the objects. It requires line of sight to at least two transmitters to be able to make a measurement. iGPS systems have been used in aerospace, automotive, and industrial robot applications.

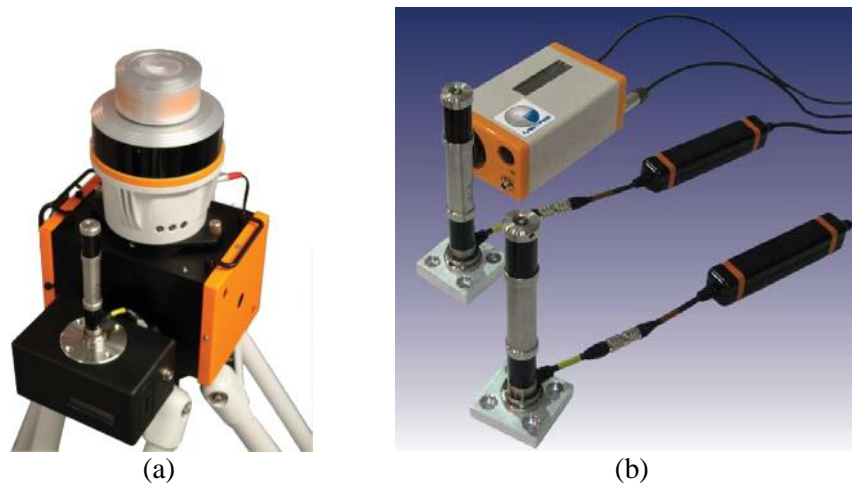


Figure 1. An iGPS transmitter (a) and two receivers (vector bars) (b) with cables and position computation engine (PCE).

The manufacturer specifies the accuracy of 3D position measurements of the iGPS as 0.25 mm, with a maximum measurement frequency of 40 Hz. A typical measurement area based on four to eight transmitters is 1200 m². Detailed analyses of the system are presented by Schmitt, et al. [26], and Mosqueira, et al. [27]. In [28], Wang et al. showed that the tracking accuracy of an object in motion is similar to the static accuracy for speeds below 10 cm/s. However, they found that as the speed of an object increases, the tracking accuracy decreases—at a speed of 1 m/s, the mean tracking error could be as high as 4 mm. In another study, Depenthal [29], showed that when tracking objects at velocities of 3 m/s, the 3D position deviation is less than 0.3 mm. She described the experimental comparison of the dynamic tracking performance between an iGPS and a laser tracker and showed that the iGPS performed well under dynamic conditions. In the human-tracking experiments described in this report, a pair of iGPS vector bars is attached to the top of a hard hat worn by each person, as shown in Figure 2. The center of a human head was tracked as a point relative to a frame, in global coordinates, defined by the two vector bars.

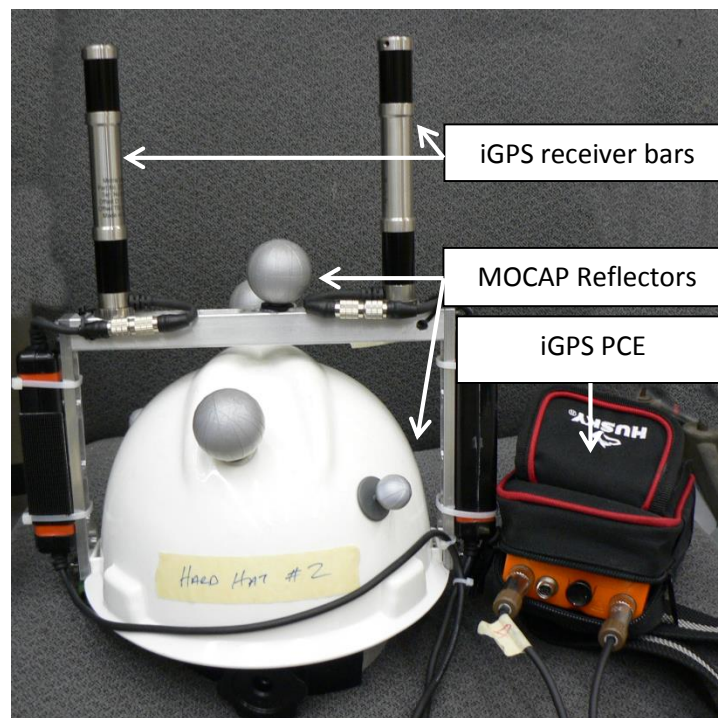


Figure 2. iGPS vector bars attached to hardhat to be worn by a human. The position calculation engine (PCE) is worn around the waist. The motion capture (MOCAP) reflectors form a unique pattern for each person.

The iGPS sensor has a fast enough update rate to track people moving at walking or running speeds and is accurate enough to provide an order of magnitude better pose measurement than most sensors used for human tracking. Its wide field of view allows a range of typical activities to be carried out by the people being tracked, and the need for only two transmitters to have line of sight to the sensors at any time ensures that the system can provide data even in scenes with significant clutter. A drawback of this system is the size of the vector bars, which are difficult to attach to small objects.

3.1.1. Determination of the head center in the iGPS coordinate frame

To determine the coordinates of the center of a “typical” human head and the uncertainty of those coordinates, an apparatus was constructed in which a person could place their head and have it remain static while it is measured with a laser scanner. The stability of the apparatus was measured using a laser tracker and was found to be an order of magnitude less than the uncertainty of the iGPS measurements. Six heads of varying sizes were measured using the apparatus and a handheld laser scanner. The basic procedure involved placing a person’s head in the rigid frame with their chin resting on a planar surface. Only the nose, the back of the head and the top of the head were measured¹. The surface on which the chin rested was also scanned and a plane was generated from those data points. The cloud of points measured on the head was then projected onto the chin plane. The X- and Y-coordinates of the head center were defined as the midpoint between the two farthest projected points. The Z coordinate was defined as the maximum z-value, z_{max} , of all the data points divided by two (since $z=0$ was the chin plane).

After scanning the head, a hardhat with the iGPS sensors mounted on it was placed on the person’s head while the head was still in the apparatus and in the same position. A measurement of the hardhat’s iGPS coordinates was made and the hardhat was then removed. This was repeated 21 times while the head was in the same position. These 21 measurements were then used to determine the XYZ coordinates of the center of the person’s head relative to the hardhat frame and the uncertainty associated with putting on and taking off the hardhat.

Based on the experimental data for six people, the coordinates of the center of the average head relative to the hardhat and the associated uncertainties (variability of people’s heads, putting on/taking off the hardhat, and instrument uncertainty) of the coordinates were:

$$X = (6.7 \pm 17.9) \text{ mm}$$

$$Y = (-24.9 \pm 10.5) \text{ mm}$$

$$Z = (-226.4 \pm 33.2) \text{ mm}$$

with a level of confidence of 95 %. The coordinates are given in the hardhat’s iGPS coordinate frame, which is shown in Figure 3. For the human tracking experiments, these coordinates were transformed into the iGPS World coordinate frame.

¹ The assumption was that the two points farthest apart were a point on the tip of the nose and a point on the back of the head.

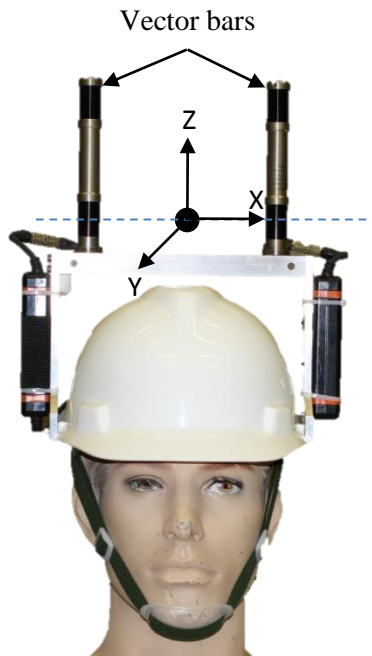


Figure 3. The iGPS hardhat coordinate frame for determining the center of a human head.

The results presented in this report make use of the ground truth provided by the iGPS system because it has been more rigorously characterized and it provided data with fewer dropouts in the experiments.

3.2. Motion Tracking System

The second ground truth system used in this study [30] is designed for motion capture (MOCAP) of people and objects, including articulated motion of joints. Systems like these are widely used to validate the performance of computer vision systems and in applications such as entertainment, sports, medicine, and robot control. The system uses a network of cameras, each of which has an infrared emitter surrounding the lens (Figure 4). The cameras track markers attached to the objects of interest (reflective spheres, such as those shown in Figure 2). Markers can be grouped and each set of grouped markers defines a rigid body which can be used to uniquely identify an object. When seen by multiple cameras, the positions of the markers or marker groups can be computed. At least three markers are required to form a rigid body; however, redundant markers are recommended because some markers may be occluded during motion and redundant markers will enable continued tracking of the object as long as at least three markers are visible to the system. Motion capture systems are easy to set up and can be equipped with varying numbers of cameras to cover larger or smaller areas and to ensure

adequate accuracy and tracking capability. They have a fast enough update rate to capture people walking or running and are a relatively low-cost option.

Motion tracking systems have been used in a number of previous performance evaluation studies. The synchronized video and motion capture dataset developed by Sigal, et al. [31] has been widely used. In it, actions of a single person were captured by multi-camera video together with marker-based MOCAP data. The main drawback of the dataset is that there is only one person in the environment at a time, so there is no person-to-person occlusion. Other datasets include multiple people. One is the Utrecht Multi-Person Motion (UMPM) Benchmark, by Van der Aa, et al. [32]. It includes a collection of multi-person video recordings together with ground-truth based on MOCAP data. The dataset is intended to be used for assessing the quality of methods for pose estimation and articulated motion analysis of multiple people using video data from single or multiple cameras. Balan, et al. [33] evaluated the 3D pose of human motion obtained from synchronized, multi-camera video against 3D ground-truth poses acquired with a MOCAP system.



Figure 4. One of the MOCAP cameras.

3.3. Determination of the head center in the motion capture coordinate frame

Once the MOCAP system was calibrated, eight 12.7 mm (0.5 in) spherical MOCAP targets were placed throughout the work volume within which the experiments were to take place. The targets were held in kinematic nests that were mounted on metrology stands. The 3D coordinates of the centers of the targets were then measured with the MOCAP system and then the targets were removed from the kinematic nests. The 3D coordinates of the kinematic nests were then also measured with the iGPS probe (Figure 5) using a 12.7 mm (0.5 in) ball tip so that the iGPS measured the same coordinates as the MOCAP system, i.e., the centers of the MOCAP targets. A least-squares fitting algorithm was then used to calculate the homogeneous transformation matrix

between the corresponding sets of eight 3D coordinates so that any iGPS coordinate could be transformed into a MOCAP system coordinate and vice versa.



Figure 5. iGPS probe with ball tip.

After the transformation matrix was determined, six MOCAP targets were affixed to each hardhat (as in Figure 2) and the hardhat was placed on a stationary (and rigid) mannequin, within the MOCAP system's measurement volume, so that all six targets were visible to the MOCAP system. The hardhat was then defined as a rigid body, within the MOCAP system's software, using the measured coordinates of all six targets. The origin of the rigid body was defined as the centroid of the six targets. Simultaneously, the coordinates of the head center were also measured with the iGPS (as described in section Determination of the head center in the iGPS coordinate frame).

Since the iGPS measures the center of the average human head on which the hardhat is placed, while the MOCAP system measures the centroid of the six targets affixed to the hardhat itself, the two systems do not measure the same thing. Therefore, the iGPS head center coordinates were transformed into the MOCAP system's coordinate system and the translation vector was calculated between the transformed iGPS coordinates and the coordinates of the origin of the rigid body. This translation vector was then used to permanently move the origin of the rigid body to the location of the center of the average human head. Subsequently, any measurement of the coordinates of the origin of the rigid body resulted in the MOCAP system's estimation of the average human head center for that hardhat.

The above process was repeated for the other hardhats that were used in the experiments described in this report.

4. Methodology

A series of experiments was conducted with human subjects to collect data to evaluate two different sets of performance metrics. One set of measures has been used in earlier studies, while the other is novel and is particularly focused on applications where safety in human-robot interaction is the main concern. A human detection and tracking algorithm was adopted based on a literature search and was used to analyze the data, although the aim was not to evaluate that particular algorithm's performance. The algorithm and its implementation are briefly described in the next subsection. This is followed by a description of how the data were collected and what information was gathered. The two sets of performance metrics are then described and their performance is evaluated.

4.1. The Human Detection and Tracking Algorithm

Because one of the applications of human detection is for safety, an algorithm was needed that has very low false-negative detections. That is, the algorithm should not miss seeing a person if there is one in the field of view of its sensors. On the other hand, for manufacturing applications there is also a low tolerance for false-positive detections in which a human is reported when none is present. This can result in reduced productivity if the system has to take evasive action for no reason. Thus, an algorithm with very high accuracy was needed. Over the past several years, human detection has been the subject of significant advances, and one of the motivations for this study was to determine if algorithms had progressed to the point at which the two requirements could be met simultaneously.

The algorithm that was selected is that of Chambers, et al. [34], because it claims a very high accuracy: 99 % recall with less than 10^{-6} false positives per window on the INRIA person dataset (Dalal [35]). The algorithm operates on disparity data acquired from either a stereo camera system or an RGB-D camera (in this case a ASUS Xtion Pro Live sensor). RGB-D sensors provide a color image plus the depth of each point in the image. The approach taken by Chambers et al. is to start with computationally non-intensive feature detectors that can rule out image regions as locations potentially containing people. More computationally-intensive detectors are then applied to the remaining regions in a cascade that continues to eliminate candidates until, following the application of the final detectors, the remaining candidates have high confidence.

The sequence of detectors starts by using Haar-like features (Papageorgiou, et al. [36]) classified first using adaBoost (Freund and Schapire [37]), then a support-vector machine (SVM) (Cortes and Vapnik [38]) on a subset of the features and finally, an SVM on all the features. This is followed by a Haar-like feature detector applied to the disparity image using an adaBoost

classifier on a subset of the features, but without the subsequent SVM classifier. Windows that have passed the preliminary classifiers are processed with a histogram of oriented gradients (HOG) (Dalal and Triggs [39]), customized to reflect the expected silhouette of a person. Two additional constraints are applied to the data, made possible because disparity is available. The first is a size constraint on the regions that may contain people, while the second is an estimate of the ground plane on which people are expected to be standing. The algorithm outputs an estimate of the position in 3D space of each detected person and an identifier for each person. Over time, the same identifier should be given to the same person, and the system can tolerate short-time occlusions or intervals when people move out of the field of view of the sensors and then come back into view.

Each of the classifiers in the algorithm needs to be trained. A set of training data was collected independently of the data used for evaluation. Because the HOG classifier is heavily dependent on the appearance of the silhouettes of the people, the training data were collected with people wearing the hard hats with ground truth markers, even though no ground truth was needed for training. It was felt that this would give a fairer evaluation of the algorithm.

4.2. Data Collection

A series of experiments was designed in which human subjects were asked to stand, walk, or jog in the field of view of the sensors². The people wore hard hats which were instrumented with markers from the MOCAP system and detectors from the iGPS system so that their actual locations and trajectories would be known, and data from all sensors were collected with timestamps to enable the data to be compared. Each of the sensors was calibrated, and the ground truth sensors and system being tested were registered to a common coordinate system. The experiments started out very simply with stationary people in simple scenes (no obstructions, simple backgrounds). The next experiments involved a single moving person in simple scenes and with obstacles. After that, multiple people moved in scenes with some human-to-human or human-to-obstacle occlusions. In some of the experiments, an object moved along a straight line through the scene on a motorized sled. The obstacles used were empty cardboard boxes. Figure 6 shows a view of an experiment in progress. Note the hard hat with reflectors for ground truth measurements.

² The NIST Institutional Review Board approved the experimental protocol before the experiments were conducted. Due to the possible presence of personally identifiable information, they determined that the data collected during the experiments could not be made available for use outside NIST.

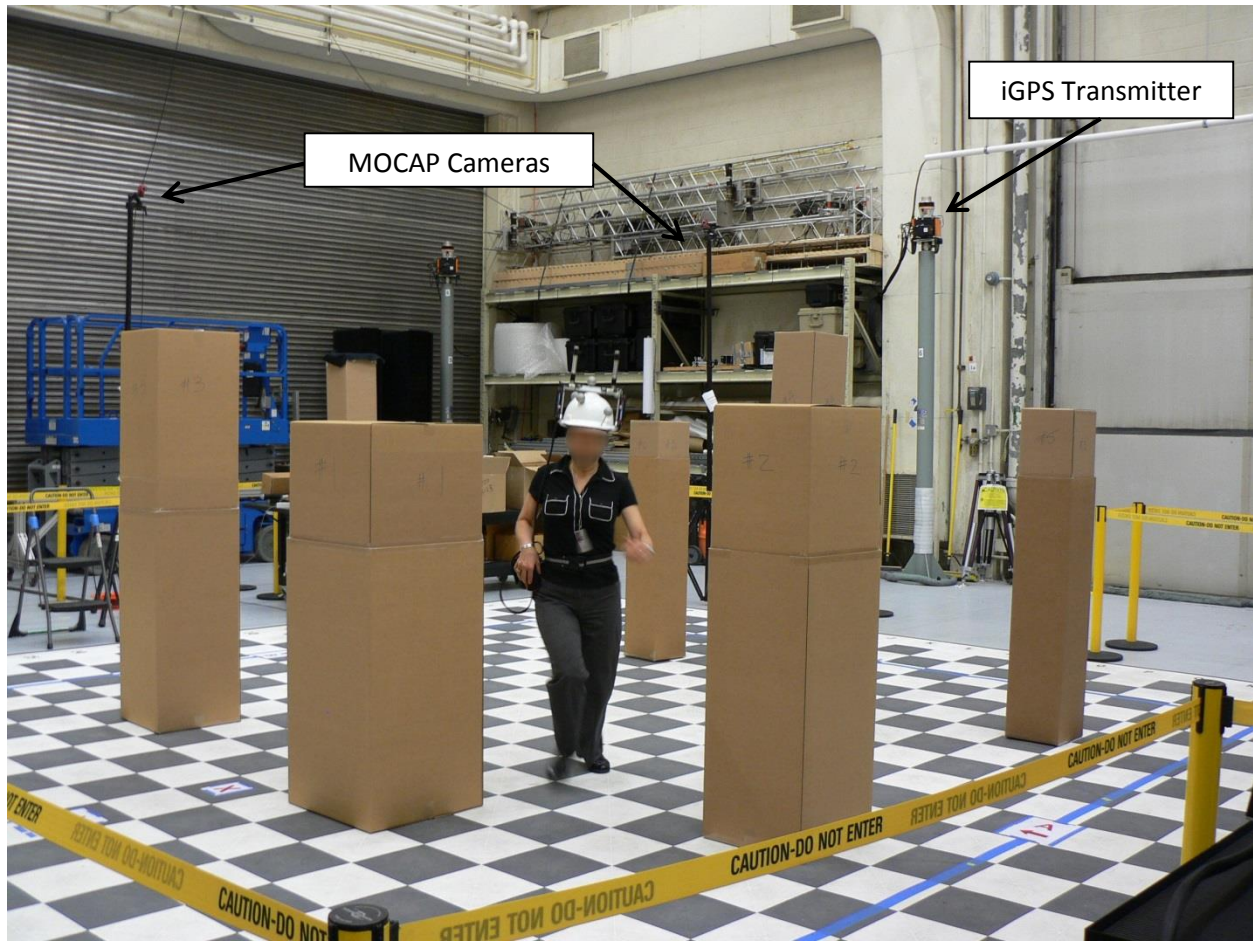


Figure 6. A view of the test area showing an experiment in progress with obstacles in place.

Due to the limited field of view of the sensors being tested, the active area used for testing was small, only 5 m by 5 m, with a 1 m buffer area surrounding it. The floor was tiled to reduce the risk of slipping and to enable the paths to be described to participants in terms of coordinates. One side of the tiled area was labeled with letters and the other with numbers. This meant that a person could be asked to stand in a particular location (e.g., J5) or to move from one location to another without researchers having to mark the paths, which changed with different experiments.

There were a total of 34 different experiments (described in Appendix A), each of which was repeated five times with different participants. Fifteen different subjects participated. The experiments were designed to include factors such as distance from the system under test, occlusions, clutter, speed of motion, different types of trajectories, and different human shapes and features. To give a flavor of the experiments, four are briefly described below. The data that were collected for each experiment included a sequence containing, for each instant (about 20 frames per second), the position of each person (x, y, z, roll, pitch, yaw), an identifier for the person, and a time stamp. This was collected for both ground truth systems and for the sensors being evaluated (stereo and RGB-D). Each sensor collected data at its own rate, but the time stamp made it possible to align the data for analysis.

Although data were collected for both ground truth sensors and both stereo and RGB-D sensors being tested, the results are presented only for the iGPS ground truth and the RGB-D sensor under test. This is because the goal of the study is to develop and describe the performance evaluation methods, not to evaluate the human detection algorithm.

4.2.1. Stationary People

Figure 7 shows a schematic of an experiment where people were asked to stand at given locations and remain still for a few seconds while the sensors try to locate them. The experiment was repeated with obstacles in the field of view and also with the people posing as if gesturing. The other stationary experiments were similar, although fewer people were used. The algorithm being investigated was implemented to find at most five people, and this was the maximum number used in any experiment.

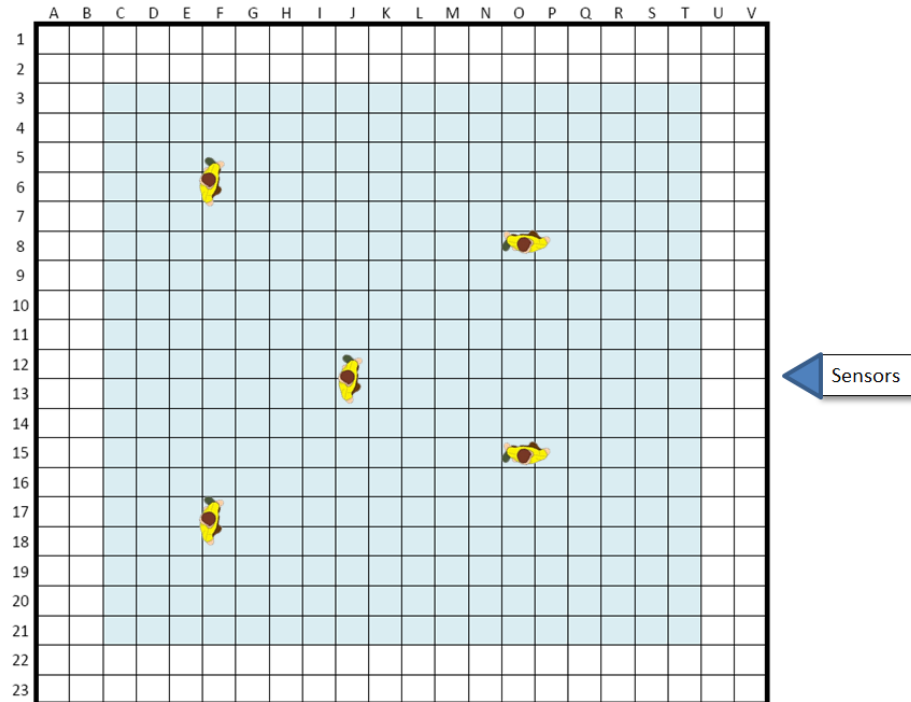


Figure 7. A schematic of an experiment in which five people stand still and are viewed by the systems under test. The people showed different views to the sensors, that is, front, back, or side views.

4.2.2. A Single Person

In the experiments when people were moving, the same trajectory was covered for four different cases: walking without obstacles, jogging without obstacles, walking with obstacles, and jogging with obstacles. The person was given a trajectory at the start of the experiment and told to walk

or jog at a comfortable speed (Figure 8). Data were collected simultaneously from all the sensors for the duration of the experiment.

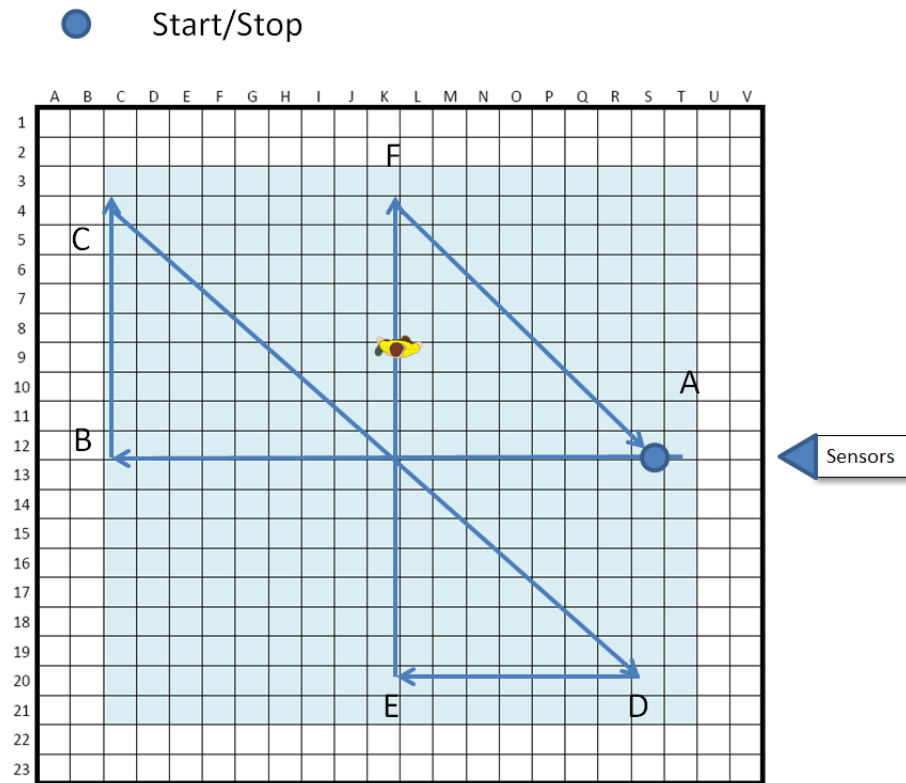


Figure 8. A single person walks along a pre-defined path starting at A.

4.2.3. Multiple People

In order to determine how well the algorithm can keep track of many people, a number of tests were conducted with up to five subjects (Figure 9), the maximum number for which the algorithm was developed. In the limited space available, because of the field of view of the system being evaluated, there is a lot of crowding and with obstacles in place there is also a lot of person-to-person and person-to-obstacle occlusion. This is not atypical of a workspace such as an assembly line where there will often be equipment and other people in the area. As in the single person case, each experiment with multiple people was conducted four times with different speeds and with and without obstacles.

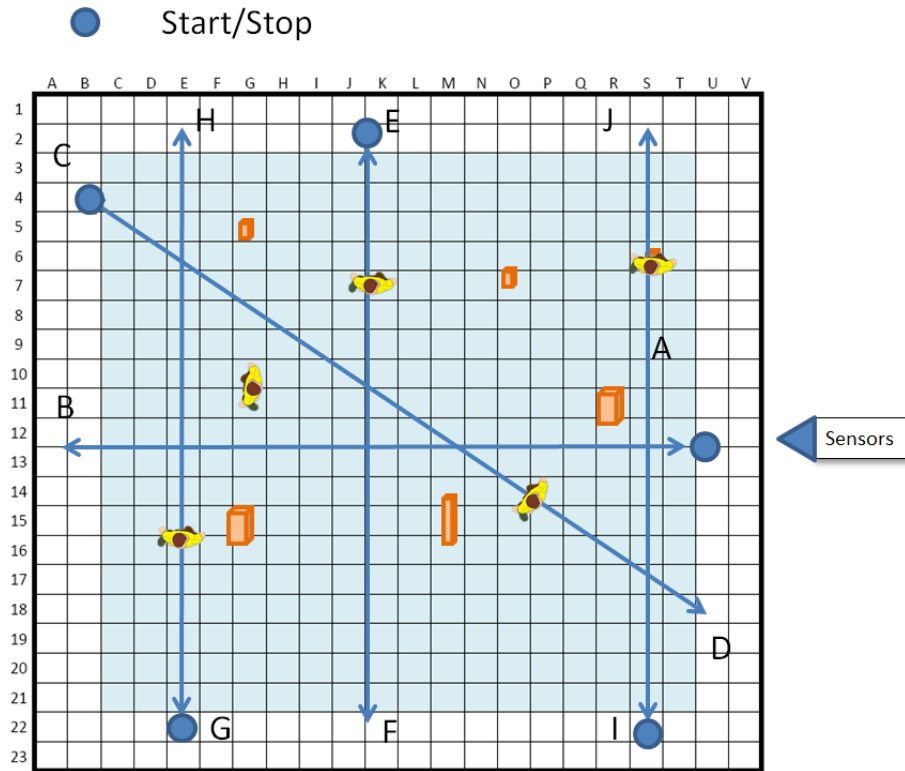


Figure 9. Five people moving through the space simultaneously with obstacles.

4.2.4. People and Sled with Obstacle

A motorized sled was used to pull a box of roughly human size through the field of view of the camera while people were walking or jogging at the same time (Figure 10). This was done with and without obstacles. The aim was to see if the algorithm would be able to distinguish between a moving object that is not a person and the real people in the scene.

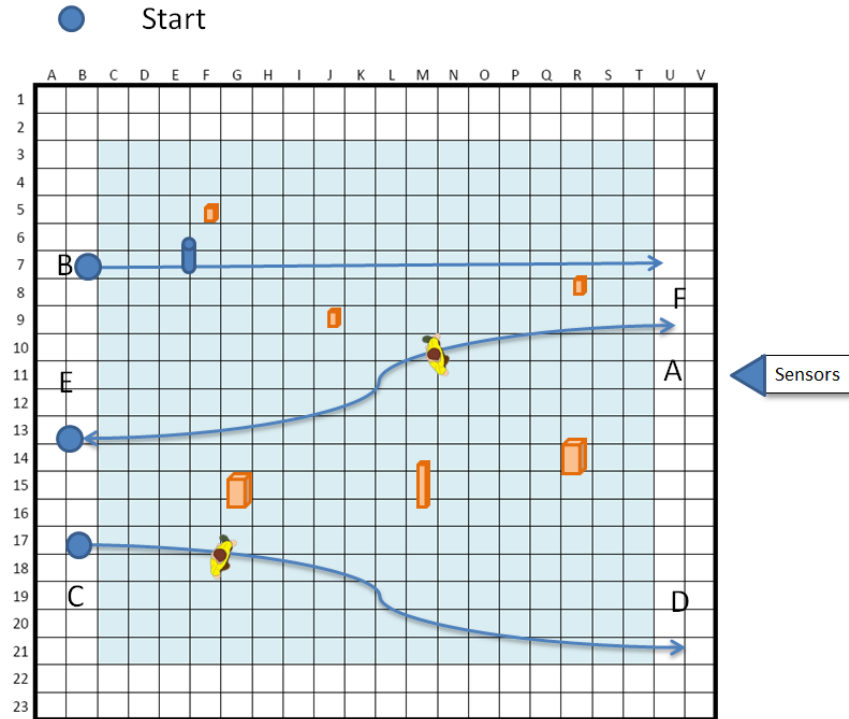


Figure 10. An experiment in which a motorized sled and two people move through a field of obstacles.

5. Performance Measures

Two sets of performance measures were developed for the human detection evaluation. One is based on measures previously used for evaluating detection in a range of applications including pedestrian location by autonomous vehicles (Section 5.1), while the other is a new approach developed specifically for applications in which safety is critical (Section 5.2). Both sets of performance measures can be used for human-robot collaboration and situation awareness applications and they provide complementary information. Before describing the performance measures, we discuss the data obtained from the tests, registration of data from the iGPS ground truth (GT) sensors and the systems under test (SUT), and linear interpolation of the ground truth data for evaluating the data from the SUT.

Input data: The data obtained from the tests form the input to the analysis system. There are three types of input files. All of the files are in Comma-Separated-Variable (CSV) format.

○ Ground-Truth Data Files

- Separate files for each sensor and for each test (i.e., two ground-truth sensors * 100 tests = 200 files).
- Each row contains the following required fields:
 - Timestamp, ID, X,Y, Z

- Timestamp = a floating point number representing the time in seconds since Jan 1, 1970 UTC
 - ID = unique string corresponding to a person or physical target
 - X,Y,Z = Cartesian coordinates of the target
- Each row may contain an optional radius field, which is used by the analysis system if present. It describes the distance from the center point of a detected person sufficient to contain the entire person
 - Other optional fields are present but are ignored in the analysis.
- System-Under-Test Data Files
 - Separate files for each sensor and for each test. (i.e., two sensors * 100 tests = 200 files)
 - Each row contains the following required fields:
 - Timestamp, ID, X,Y, Z
 - Each row may contain optional fields used by the analysis system if present:
 - Velocity in X direction, Velocity in Y direction, Velocity in Z direction, Radius, Confidence
 - Other optional fields are ignored in the analysis.
- Position and Orientation Transform Files
 - A single transform file is used for each sensor for all tests (i.e., a total of two transforms) for all 100 tests because the sensors were not moved between tests. Each transform file contains a 4x4 matrix (homogenous transform) that encodes the registration between a sensor and the ground truth obtained from a calibration and registration procedure.

Data interpolation: The timestamps and distances may have different scales for different sensors. In this case, the system must be provided with scale factors to convert to the units used in the analysis: seconds and meters. The timestamps must also be based on the same epoch, which for our tests was January 1, 1970. Computers were synchronized using a combination of the Network Time Protocol (NTP) and a trigger signal.

The GT system and SUT rarely produce data at the same time, so points from one system cannot be compared with the other without some implicit time shifting. In addition to running at different frame rates and starting at different initial times, both the GT systems and the SUTs occasionally fail to produce data. The goal of the interpolation in the case of the GT system is to provide the best estimate of the position of each person at each time given all data, including data collected after that point in time. This is used in the conventional measurements for evaluating the SUT, described in Section 5.1. In addition to interpolating the GT data, the new grid based measurements described in Section 5.2 also require interpolation of the SUT data to provide the most reasonable estimate of the position that would have been available to the control and planning software at the time of each frame. For the SUT data, this assumes only data produced prior to that time.

A number of interpolation mechanisms were implemented, including simply using the previous point or the closest point as well as linear or cubic interpolation, or using the previous point plus the reported velocity multiplied by the time since the last reported point. A linear interpolation method between two points was implemented for the results shown in Section 6.1, while for the results reported in Section 6.2 ground truth measurements were interpolated using a linear method for the GT data and the previous point method for data from the SUT. It is unlikely that the interpolation choice had a significant effect on the results.

The GT system and the SUT are started and stopped manually, so there may be a period of time during which the GT system was collecting data but the SUT had not been started and a period after the SUT was stopped when the GT system continued to collect data. With at least one system, there was a period of time after the SUT was started when it missed detections at a far greater rate than the rest of the test period. For this reason there is an additional configurable start-up delay to ignore data before the SUT was truly ready. The evaluation software analyzes the timestamps in the logs to determine for which period of time the SUT was expected to provide accurate positions and limits analysis to this period.

Sensor Registration: The GT system and SUT typically produce data in different coordinate systems. The evaluation software includes a capability to visually align the data in order to determine a transform that can be applied to all of the data from one system so that it can be compared with data in the other. The results are shown in Appendix B. Automatic methods were also developed, but could not be used due to large and currently unexplained systematic biases in the SUT.

Correspondence between Ground Truth System and System Under Test: The method used for determining correspondences between humans detected in the SUT data and the GT data significantly affects the values of the performance measures. Matching using centroids of regions labeled as people is based on measuring the Euclidean distance between the person's centroid as reported by the SUT and the GT data at each frame in the system under test data, with a threshold for a successful match. Normalization based on the size of the bounding box is often also used (Bodt and Hong [40]). However, the method described in Section 5.2 does not need to establish correspondence because it does not explicitly track people.

5.1. Human Tracking and Detection Measurements

In this section, we present several performance evaluation metrics that have been used for human detection and tracking. The basis for comparing the strengths and weaknesses of different human detection and tracking methods is to evaluate their results on a set of experiments with known ground-truth data using the same performance metrics. The human detection metrics that we used include:

- False Positive (FP): A human is present in the SUT data but not in the GT data.

- False Negative (FN): A human is present in the GT data, but not in the SUT data.
- True Positive (TP): A human is present in the GT data and the SUT data.
- False Positive Rate (FPR): The number of false positives divided by the sum of the number of true positives and false positives. FPR is a measure of how well the system correctly rejects false positives.
- Detection Rate (DR): The number of true positives divided by the sum of true positives and false negatives. DR is the percentage of true targets detected.
- False Negative Rate (FNR): FNR is the number of false negatives divided by the sum of true positives and false negatives. FNR is the likelihood that a target will be missed given the total number of actual targets.
- Detection Precision (DP): DP is the number of true positives divided by the sum of the true positives and false positives. That is, precision is the fraction of detected items that are correct.
- Localization: The average distance between the position of the person detected by the SUT and the position in the corresponding GT data over the course of an experiment.

The human tracking metrics measure the ability of a SUT to track humans over time. The tracking metrics consider the identity and the complete trajectory of each object separately over the experiments and compare the GT tracks with the SUT tracks based on best correspondence. Based on these correspondences, various error rates and performance metrics are computed.

Since the GT track(s) could correspond to more than one SUT track, a correspondence first has to be established, as described above. Two measures are used to express the performance of the tracker. The first is the tracking precision, which expresses how well the tracker estimates the positions of objects or people. The second is the tracking accuracy, which measures how well the system keeps track of people or objects and how many mistakes are made in terms of misses, false positives, mismatches, failures to recover tracks, etc. Two main human tracking metrics were used:

- Human Tracking Precision (HTP): is the precision of the tracker in determining the position of a tracked person or object. HTP is calculated as:

$$HTP = \frac{\sum_t d_t^i}{c_t},$$

where d_t^i is the Euclidian distance error between the matched GT location and the matched SUT location and c_t is the total number of matches made. The HTP is a Euclidian distance error for matched GT-SUT pairs over all frames, averaged over the total number of matches made. It shows how well positions of persons or objects are estimated. HTP is reported in units of length (e.g., m).

- Human Tracking Accuracy (HTA): is the accuracy of the tracker in keeping correct correspondences over time, estimating the number of humans, recovering tracks, etc.

$$HTA = 1 - \frac{\sum_t (FN_t + FP_t)}{\sum_t GT_t},$$

where FN_t and FP_t are the number of false negatives and false positives in the SUT for time t . HTA is the sum of all errors made by the tracker over all frames, averaged over the total number of humans detected by the GT sensor. HTA is unitless.

5.2. Grid Based Measurements for Safe Human-Robot Interaction

In addition to the traditional performance metrics for human detection defined in Section 5.1, a new set was defined with the following goals in mind:

- Provide safety guarantees: No area is falsely considered to be clear. An area is falsely considered to be clear if a person may be in the area but the system says no human is in the area. Details are provided in section 5.2.3.
- Compare systems as configured by the vendor with safety guarantees met.
- Provide fair comparisons between systems with different frame rates, detection mechanisms, and different numbers of outputs
- Make it easy to combine multiple detection systems.
- Separate responsibility between human detection and robot path planning and execution.

All of the software used to compute these measures and to perform the analysis in Section 6.2 is available from Shackleford [41]. Full source code is provided for a complete implementation to ensure that the algorithms are unambiguously specified and the results are reproducible.

There are two types of mistakes a human-detection system could make. It could report that an area contains a human when it is actually safe for the robot to traverse or it could report that an area does not contain a human when in fact it does. The first kind of error, a false positive, reduces productivity by leading the robot to stop, slow, or take a longer path than necessary, while the other, a false negative, reduces safety by potentially allowing the robot to collide with a human. The guarantee we expect human detection systems to provide is that for all tests conducted, for every point in time during the test, and for all locations within the test area, the detection system never makes the second, and far more dangerous, kind of error. To infer that the system is safe beyond what is reported by the tests, two additional requirements must be met:

- The tests thoroughly cover all possible cases the system will encounter. The tests described in Section 4 provide a bare-minimum set, but are not intended to be sufficient for this purpose. A more complete set would require designing the test with much more specific information about the application and intended environment.
- The robot's planning and path execution must meet its own set of requirements to be briefly described later.

5.2.1. Detection Area vs Safety Coverage Area

The system under test is only expected to report the positions of people within its field of view and between the minimum and maximum ranges. The safety coverage area is the area the system could potentially tell was safe to move through and is smaller than the area where detections could occur. The reasons for this are: 1) Part of a person centered just outside the detection area could extend into the detection area; 2) If the robot has non-zero reaction time, then a person outside the detection area could move into the detection area before a robot could react [42]; 3) Part of the area where detections could normally occur may be occluded from the system; and 4) At the farthest and nearest ranges of the detection area there may be an area where detections sometimes occur but are not sufficiently reliable for safety applications. The evaluation software allows the user to specify a polygon as the outer edge of the safety coverage area.

5.2.2. Ground-Truth Radius Estimate

Both of the ground-truth systems used here provide only a single position at a time for a target attached to the helmet. The positions of the person's shoulders, hands, feet and other body parts can only be inferred to be located somewhere within a circle centered at the helmet target position (plus optional offset). This radius is configurable and could be expanded to include an additional safety margin. If increased accuracy is required, then a ground truth system that could provide positions and possibly shapes of each body part would be required.

5.2.3. False Clear Area/ False Occupied Area

The two primary metrics are computed using a similar procedure except for a change in the order of operations.

To compute the false clear area:

1. Create an image covering the area of interest at the desired resolution.
2. Paint the entire image black.
3. Paint white circles surrounding each ground-truth location of a person (See Figure 11). The sizes of the circles are as described in Section 5.2.2.
4. Paint black circles surrounding each location where a person was found by the system under test.
5. Count the number of pixels that are white (See Figure 12) to give the false clear area.

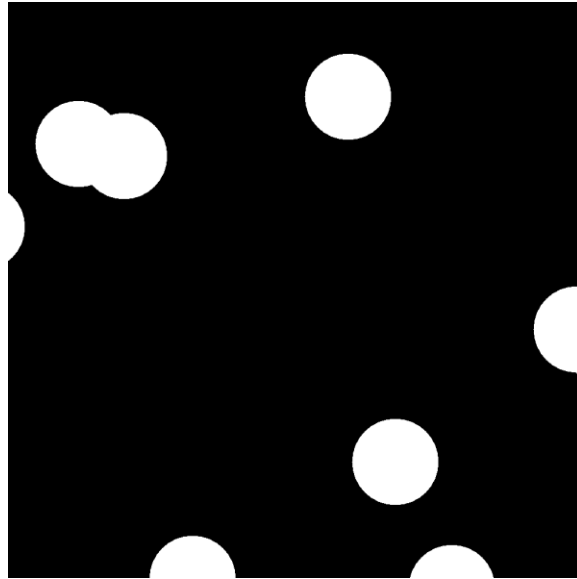


Figure 11. Computing the False Clear Area after step 3, while ignoring occlusions.

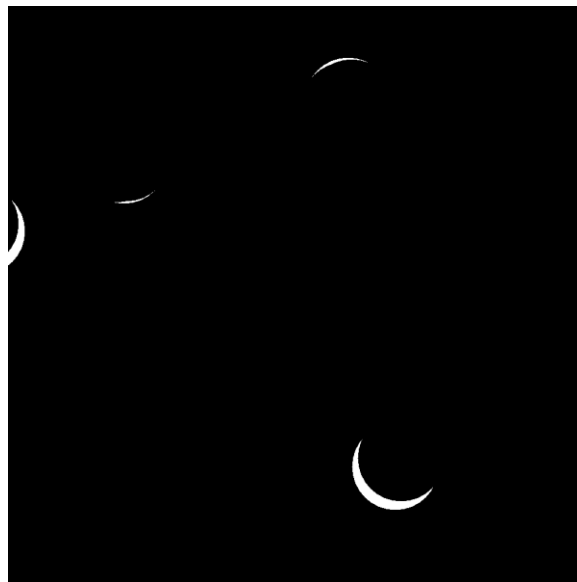


Figure 12. Computing the false clear area after step 5, while ignoring occlusions.

To compute the false occupied area:

1. Create an image covering the area of interest at the desired resolution.
2. Paint the entire image black.
3. Paint white circles surrounding each location where the system under test detected a person.
4. Paint black circles surrounding the locations where the ground truth detected people.
5. The number of pixels that are white following this procedure gives the area that is falsely determined to be occupied.

Note: It would be incorrect to assume that false clear or false occupied regions less than the size that a person could occupy can or should be ignored. A false clear area less than the size that could entirely contain a person cannot be ignored. If the area were to be ignored a robot could hit a person who was partially in that area and partially in an area that was correctly labeled or was outside the detection area. Similarly, false occupied areas of any size may require the robot to stop or take a longer path and thus reduce productivity.

5.2.4. Occlusions

Some areas of the coverage area will be occluded from the sensor by dynamic or static obstacles. Neither of the systems tested provides an estimate of where these occluded areas are. Any planner that intends to produce a safe plan must treat areas that are occluded as if they contain a person. The procedure above is therefore modified by painting an additional polygon behind each static obstacle and behind the positions where people were detected, as provided by the system under test. Painting these polygons requires that the evaluation system know the location of the sensor, which was not needed in the evaluation described in Section 5.1. In some cases the system under test will detect people within this occluded area either because the person is only partially occluded or because the detection makes an extrapolation from previous data when that person was not occluded. However, because these detections cannot be relied upon when safety is at stake, the consequences to the planner are the same. If an area is even partially occluded a person may be in this area and therefore the only safe assumption is to treat the area as if it was known to contain a person. This results in an estimate of the occupied areas that appears something like Figure 13.

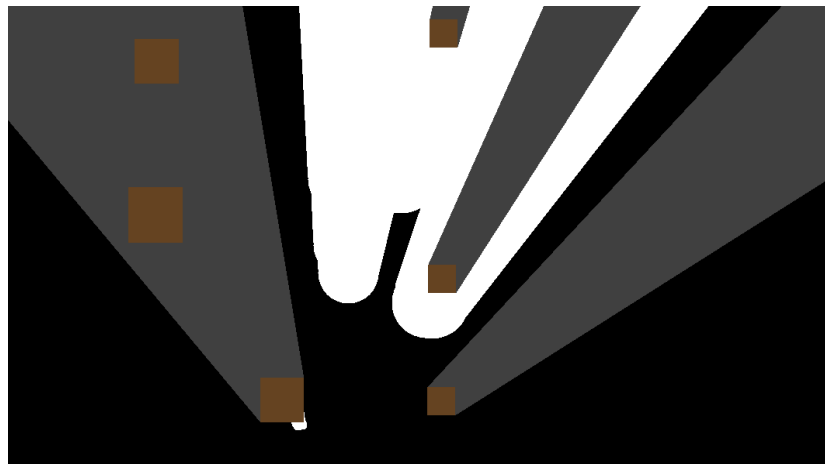


Figure 13. Top view of occupied areas after occlusions are accounted for. Grey regions are occluded by static obstacles. White areas are areas occupied or occluded according to the system under test. Brown boxes indicate the locations of static obstacles. The sensor is located below the image.

5.2.5. Time Projection for Assumed Reaction Time

No system can react instantaneously. The planning and control system must have time to respond to any new detection or change in the estimated position of a detected person. This reaction time must include communication and computation delays as well as delays caused by the physical system, e.g., due to inertia. What really matters is a guarantee of where people cannot be over some time interval into the future. False clear and false occupied regions are computed just as in Section 5.2.3 except now from each point in time we integrate over a period into the future. This means that the system under test must not only estimate the current position of each person but also predict all potential future positions measured from that point until the reaction time has expired (Figure 14). The simplest method by which this prediction can be accomplished is to increase the radius around each detected person until it contains the entire area that the person can travel during the required reaction time. Systems that can accurately measure a person's velocity may be able to predict their location through short reaction time periods without increasing the radius or by increasing it by a smaller amount. The evaluation software is currently limited to projecting along a single vector at constant speed since this is all the information the systems to be tested could provide. To test systems capable of more sophisticated predictions, the evaluation software would have to be updated.

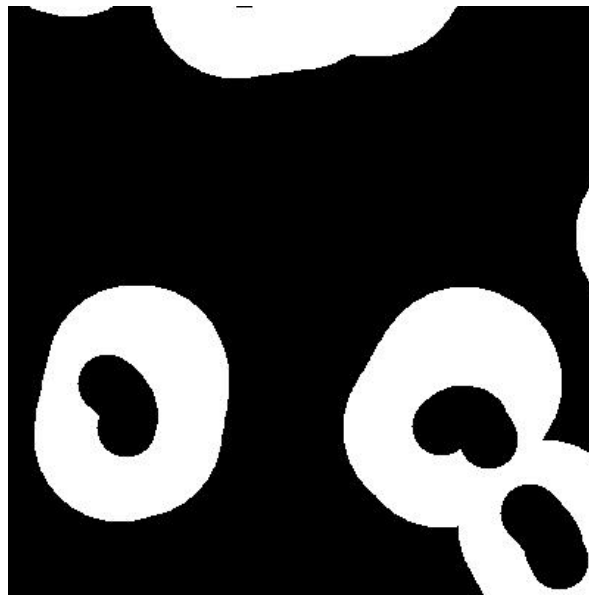


Figure 14. Display of False Occupied regions as computed by integrating over the reaction time. SUT Occupied area is white stretched circle. GT Occupied area is black region within the SUT Occupied area.

6. Performance Analysis

6.1. Conventional Measures

In this section, we present the analysis of the metrics described in section 5.1 and details of factors potentially contributing to the results. These factors include the speeds at which the people moved, the amount of environmental clutter, and the number of people in the scenes.

Table 1. Performance statistics for the four classes of experimental runs .

Cluttered Environment	Detection Rate(DR)	False Negative Rate(FNR)	Localization Mean/Std (m)	False Positive Rate (FPR)	Detection Precision(DP)
Stationary Humans	0.21	0.79	0.39/0.28	0.75	0.25
Single Human	0.11	0.89	0.24 /0.14	0.97	0.03
Multiple Humans	0.20	0.80	0.29 /0.16	0.84	0.16
Sled & Humans	0.12	0.88	0.40 /0.28	0.94	0.06

Table 2. Performance statistics for the four classes.

Uncluttered Environment	Detection Rate(DR)	False Negative Rate (FNR)	Localization Mean/Std (m)	False Positive Rate (FPR)	Detection Precision(DP)
Stationary Humans	0.57	0.43	0.13 /0.07	0.24	0.76
Single Human in Motion	0.59	0.41	0.21 /0.13	0.55	0.45
Multiple Humans in Motion	0.26	0.74	0.29 /0.13	0.75	0.25
Sled & humans in Motion	0.23	0.77	0.38 /0.30	0.82	0.18

Table 3. Performance statistics of the tracking measures.

Cluttered Environment	Human Tracking Precision (HTP) Mean/Std (m)	Human Tracking Accuracy (HTA)
Stationary Humans	0.39 /0.28	-0.26
Single Human in Motion	0.24 /0.14	-2.91
Multiple Humans in Motion	0.29/0.16	-0.82
Sled & humans in Motion	0.40 /0.28	-1.72

Table 4. Performance statistics of the tracking measures.

Uncluttered Environment	Human Tracking Precision (HTP) Mean/Std (m)	Human Tracking Accuracy (HTA)
Stationary Humans	0.13 /0.07	0.44
Single Human in motion	0.21 /0.13	-0.44
Multiple Humans in motion	0.29 /0.13	-0.50
Sled & Humans in motion	0.38 /0.30	-0.86

Table 1 and Table 3 show the performance measurements of the human detection and tracking system using the detection metrics and the tracking metrics respectively in a cluttered environment. Table 2 and Table 4 show the performance of the human and tracking detection system using the detection metrics and the tracking metrics respectively in environments without clutter. The results indicate, as expected, that detection performance is degraded when the environment is cluttered.

Another factor considered in these experiments is the speed at which the people moved. Three different speeds were considered (stationary, walk, and jog). Table 5 shows the detection metrics when the human was stationary, walking, or jogging in a cluttered environment. Table 6 shows the tracking metrics when the human was stationary, walking, or jogging in an uncluttered environment. The results indicate that the system may be sensitive to the speed of the person.

Table 5. Performance statistics of detection at different speeds.

Run Type	Speed	Detection Rate (DR)	False Negative Rate (FNR)	Localization Mean/Std (m)	False Positive Rate (FPR)	Detection Precision (DP)
Stationary Humans	stationary	0.43	0.57	0.18 /0.08	0.45	0.55
Single Human	walk	0.33	0.67	0.22/0.15	0.78	0.22
Single Human	jog	0.37	0.63	0.22 /0.15	0.74	0.26
Multiple Humans	walk	0.26	0.74	0.28 /0.13	0.76	0.24
Multiple Humans	jog	0.21	0.79	0.30 /0.16	0.81	0.19
Sled & Humans	walk	0.25	0.75	0.37 /0.28	0.82	0.18
Sled & Humans	jog	0.10	0.90	0.42 /0.30	0.94	0.06

Table 6. Performance statistics of tracking at different speeds.

Run Type	Speed	Human Tracking Precision (HTP) Mean/Std (m)	Human Tracking Accuracy (HTA)
Stationary	stationary	0.18 /0.08	0.14
Single Human	walk	0.22 /0.15	-1.67
Single Human	jog	0.22 /0.15	-1.83
Multiple Humans	walk	0.28 /0.13	-0.56
Multiple Humans	jog	0.30 /0.16	-0.68
Sled & Humans	walk	0.37 /0.28	-1.04
Sled & Humans	jog	0.42 /0.30	-1.53

6.2. Grid Based Measures for Safe Human-Robot Interaction

Neither of the systems tested met the minimum safety requirement of producing no false clear areas. Even with the ground-truth radius set to 0.1 m, which is probably too small to provide a reasonable safety offset, a 0.5 s startup time to eliminate startup problems and an assumed 0 s reaction time, and a very small coverage area to consider only detections in the center of its range limits and field of view, the system under test had false clear areas for 97 out of 169 tests (or 57 % of tests). One diagnostic method is to plot the false clear area versus time and examine traces at the point in time corresponding to the beginning of the largest peak. Figure 15 shows an example. One theory for the poor behavior is that the static obstacles confused the system. In fact the six tests with the highest false clear values all contained obstacles. However, even some tests that contained no obstacles had high false clear values. In Figure 16, a person walks directly in front of the sensor without ever being detected despite there being no other people or boxes in front of or even near the person. There are, however, people in the background and this suggests that the biggest problem with the system under test may be that it is unable to track people in the foreground when there are people or objects that may be mistaken as people in the background. In fact, of those tests with only a single person and no static obstacles, there were only two cases using the RGB-D sensor with a false clear area.

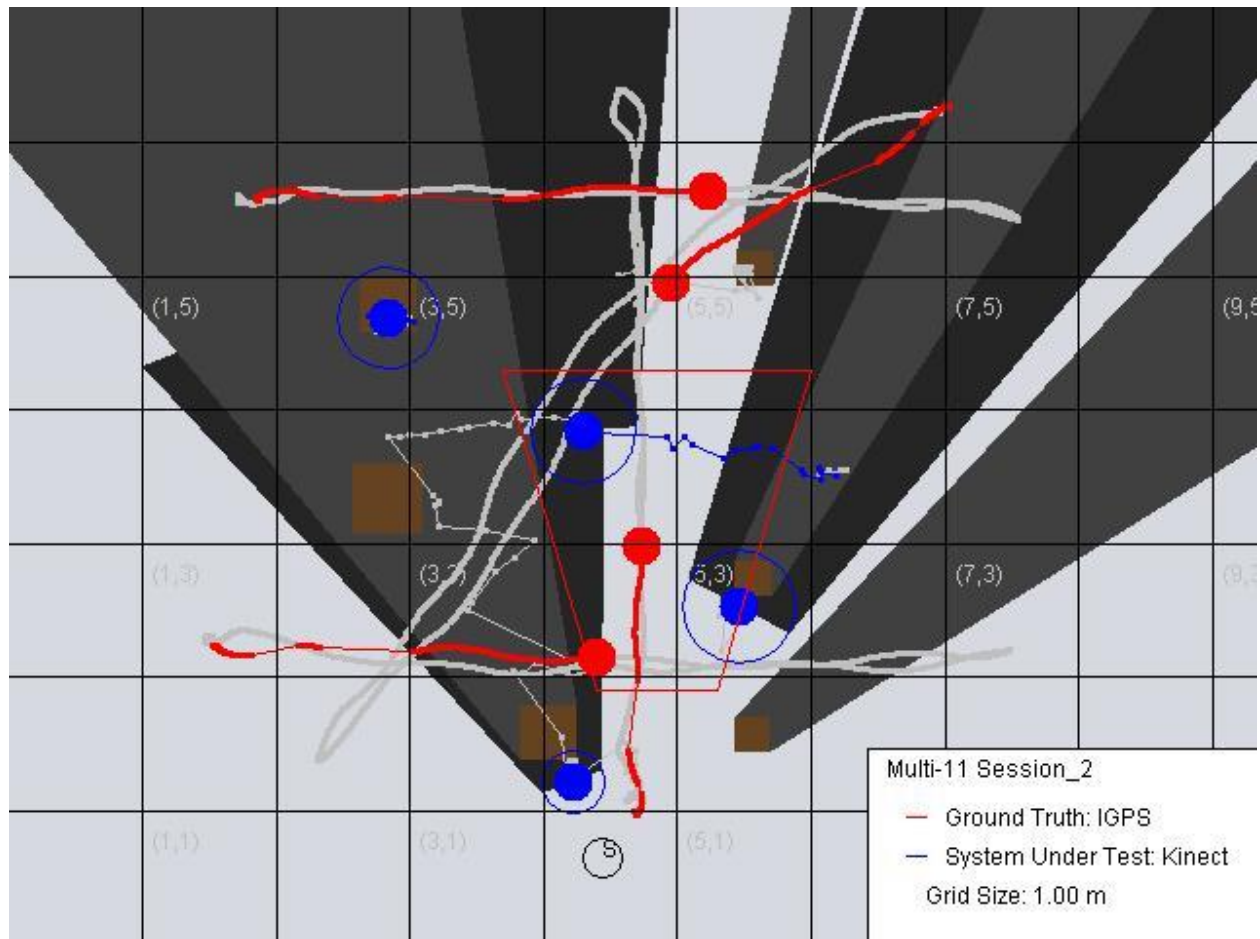


Figure 15. The tracks when the worst false clear area was detected. Sensor is located where the S in a circle is marked. None of the four ground-truth tracks (red) are matched by the tracks from the system under test (blue). The plot shows a single snapshot in time of the positions of each person as determined by each system (circles) colored (blue/red) tracks behind them show their tracks since the beginning of the test to this point and grey tracks show the tracks from this point in time until the end of the test.

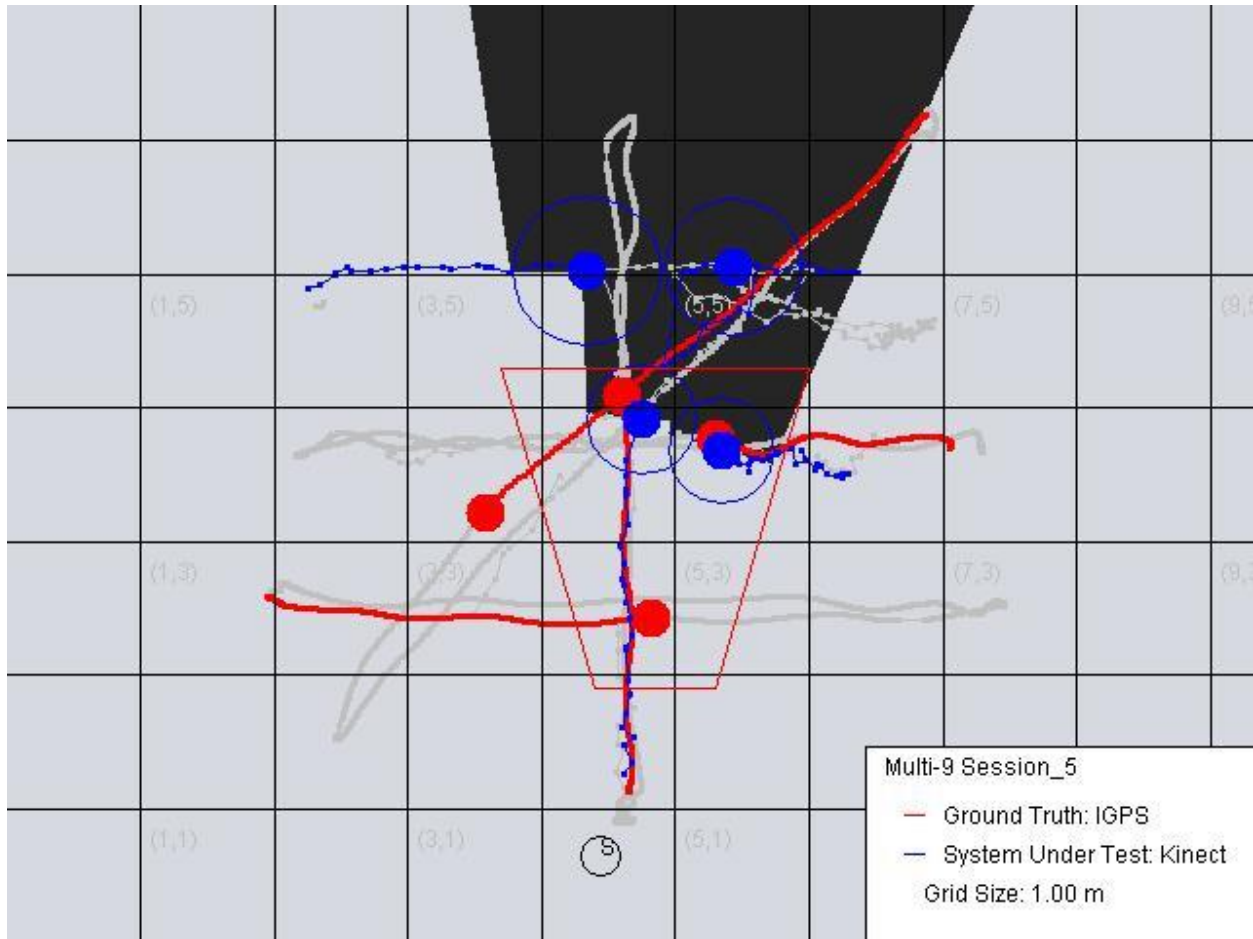


Figure 16. The tracks when the worst false clear area was detected without static obstacles. The sensor is located where the S in a circle is marked. None of the four ground-truth tracks (red) are matched by the tracks from the system under test (blue).

We provide some more detailed data below as an example of what one might obtain. It is important to avoid focusing on the many irrelevant details. To do this one should keep two questions in mind:

1. Would it be safe to deploy a robotic system that relied on this human detection system?
2. If it would be safe to deploy such a system, what insight can I get about how the human detection system would impact productivity?

If the answer for the first question for this system under test is no, there is no reason to even ask the second question. However it is hoped that there will be a different answer to the first question for future systems and the second question will become very relevant.

Table 7 and Table 8 report the data in both cluttered and uncluttered environments. Ideally we would have configured the system such that no frame of any of the tests showed false clear areas. If the system could be used safely, then the column labeled “Tests with Maximum False Clear

Area > 0 /Total Tests in category” would have 0 in the numerator of every cell. This column is designed to help answer the first question and in fact one can see that the answer is no without reading the entire table. If there is only a single non-zero value in the numerator, the system cannot be safely relied on. The column labeled “Average False Occupied Area / Safety Coverage Area” would then give a percentage measure of the area a robot would have to avoid unnecessarily which could be related to loss of productivity. This column is designed to help answer the second question which, although not relevant here given the answer to the first question, is the column where one would look if they were in the position of comparing two systems that were both safe to use and could then choose the one likely to provide the best productivity. A system that produced the smallest false occupied area would lead to the highest productivity.

Table 7. Performance statistics of detection in uncluttered environments using RGB-D sensor.

Run Type	Speed	Tests with Maximum False Clear Area > 0 /Total Tests in category	Average False Occupied Area / Safety Coverage Area
Stationary Humans	stationary	0/30 ¹	0.2184750
Single Human	walk	1/12	0.0655121
Single Human	jog	1/8	0.0876930
Multiple Humans	walk	11/15	0.1641399
Multiple Humans	jog	12/15	0.1975134
Sled & Humans	walk	2/6	0.1904649
Sled & Humans	jog	2/4	0.2127153

Table 8. Performance statistics of detection in cluttered environments using the using RGB-D sensor.

Run Type	Speed	Tests with Maximum False Clear Area > 0 /Total Tests in category	Average False Occupied Area / Safety Coverage Area
Stationary Humans	stationary	0/20 ¹	0.1225847
Single Human	walk	6/11	0.3059522
Single Human	jog	6/8	0.3686000
Multiple Humans	walk	17/19	0.2178781
Multiple Humans	jog	4/10	0.2240100
Sled & Humans	walk	5/7	0.1848025
Sled & Humans	jog	4/4	0.2420052

¹Many static tests passed for the trivial reason that there were never any people in the safety coverage area.

7. Discussion and Conclusions

The use of performance measures to evaluate human detection systems has previously been applied mainly to situations in which occasional false positive or false negative results have not been critically important. In the domain of workplace safety, however, missing a detection is entirely unacceptable, while incorrectly reporting a false detection is highly undesirable because of the resulting loss of productivity. Thus, the traditional performance metrics as described in Section 5.1, while useful for comparing different systems, are not by themselves sufficient to guarantee safety. The additional considerations discussed in Section 5.2 must also be addressed.

The scenarios used in the performance evaluation were selected to be representative of real-world situations in which one or more people can be expected to be present in the environment, there may be a lot of occlusion, other objects in the environment may have appearances that are difficult to distinguish from the people, and the lighting and people's clothing are not controlled. The gaits and speeds of the people were also not specified, except that jogging was expected to be faster than walking. The experiments also included a moving object that moved at a speed similar to a walking person and had the height and aspect ratio of a person. The scenarios were thus expected to challenge even the best current human detection and tracking systems that are not designed for safety or for the environments used in the experiments. The goal was not to show how well a particular system worked. Rather, it was to develop and test the performance metrics for this demanding type of application.

Using both traditional performance measures and measures designed specifically for safety may help to determine what modifications are needed to a system that performs well according to the traditional measures. This may make it easier to transfer the benefits of those high-performing systems to the safety domain. It must be emphasized, however, that the performance measures in this report operate in the three-dimensional world rather than the image domain because it is important to know where the people are in relation to objects that may pose hazards. Thus, systems developed for and evaluated only on detection and tracking in images may not be suitable for safety applications.

In many applications, it would be very useful to know not just where the people are and how they are moving, but also what their intentions are and, if they are working collaboratively with a robot, what their perceptions are about the joint task and the current goal. This would both improve task performance and let the robot make decisions about whether or not a person is planning to move into a dangerous region. Recognizing individual people would also be useful because the robot could know what role a person has in the task and how much freedom of motion they should be allowed. This is beyond the scope of the current work but will be the subject of future research.

More robots are being installed in factories as co-workers with people and the need for robust human detection and tracking will grow substantially to ensure the safety of the people in these environments. Performance measures aimed specifically at this domain will let developers

provide specifications for their products that are meaningful to buyers and will allow human detection systems to be incorporated into the overall safety system with a reasonable guarantee as to how well they will work. Because the problem of human detection is extremely complex when the environment is unconstrained, it is expected that detection and tracking systems will still need to be tested extensively in the target application domain before being accepted as safe for daily use.

Acknowledgements:

We are grateful to Afzal Godil for help with designing the experiments and participating in the IRB process. We also appreciate the assistance of Southwest Research Institute, and especially Dr. Chris Lewis, in developing the human detection and tracking algorithm.

References

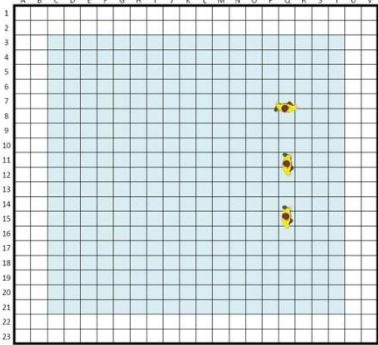
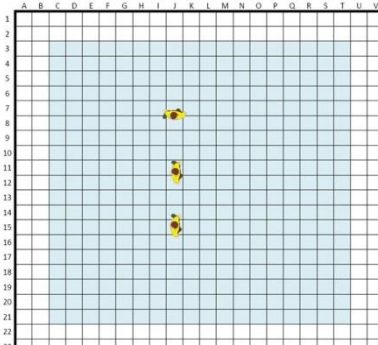
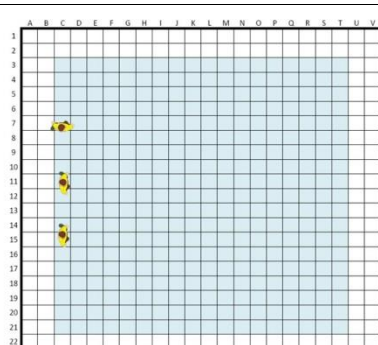
- [1] ISO TC 184, "ISO 10218-2:2011 Robots and robotic devices -- Safety requirements for industrial robots -- Part 2: Robot systems and integration.," ed, 2011.
- [2] ISO TC 184, "ISO 10218-1:2011 Robots and robotic devices -- Safety requirements for industrial robots -- Part 1: Robots.," ed, 2011.
- [3] American National Standards Institute and Robotics Industries Association, "ANSI/RIA R15.06-2012, Industrial Robots and Robot Systems - Safety Requirements," ed, 2012.
- [4] N. A. Ogale, "A survey of techniques for human detection from video," Masters Thesis, University of Maryland, 2006.
- [5] J. Ferryman and J. L. Crowley, "IEEE International Workshop on Performance Evaluation of Tracking and Surveillance," 2013.
- [6] J. C. Nascimento and J. S. Marques, "Performance evaluation of object detection algorithms for video surveillance," *Multimedia, IEEE Transactions on*, vol. 8, pp. 761-774, 2006.
- [7] R. Stiefelhagen and J. Garofolo, *Multimodal Technologies for Perception of Humans, First International Evaluation Workshop on Classification of Events, Activities and Relationships, CLEAR 2006* vol. 4122. Southampton, UK: Springer, 2007.
- [8] A. T. Nghiem, F. Bremond, M. Thonnat, and V. Valentin, "ETISEO, performance evaluation for video surveillance systems," in *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on*, 2007, pp. 476-481.
- [9] Home Office Centre for Applied Science and Technology, "Imagery Library for Intelligent Detection Systems (i-LIDS): The i-LIDS User Guide," 10/11, v4.9, 2011 2011.
- [10] British Home Office. (2013, 5/30/2014). *Imagery Library for Intelligent Detection Systems*. Available: <https://www.gov.uk/imagery-library-for-intelligent-detection-systems>
- [11] L. M. Brown, A. W. Senior, Y. Tian, J. Connell, A. Hampapur, C. Shu, *et al.*, "Performance Evaluation of Surveillance Systems Under Varying Conditions," presented at the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Colorado, 2005.
- [12] F. Yin, D. Makris, and S. Velastin, "Performance Evaluation of Object Tracking Algorithms," presented at the Tenth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS 2007), Rio de Janeiro, Brazil, 2007.
- [13] N. Lazarevic-McManus, J. R. Renno, D. Makris, and G. A. Jones, "An object-based comparative methodology for motion detection based on the F-Measure," *Comput. Vis. Image Underst.*, vol. 111, pp. 74-85, 2008.

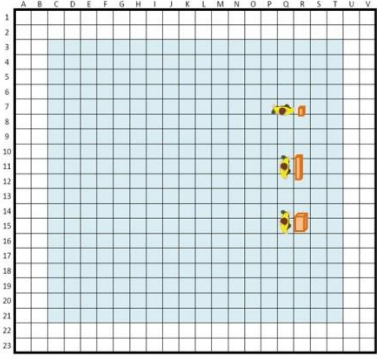
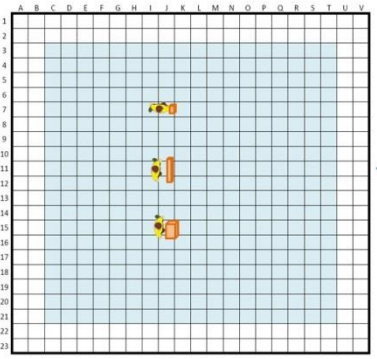
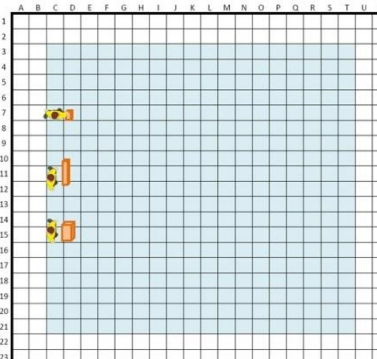
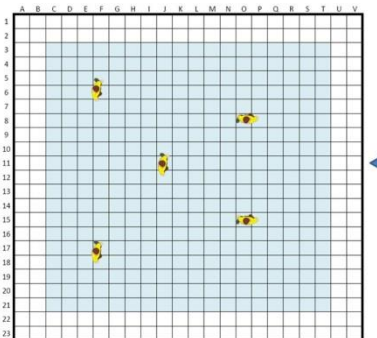
- [14] F. Bashir and F. Porikli, "Performance Evaluation of Object Detection and Tracking Systems," presented at the Ninth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS 2006), New York, NY, 2006.
- [15] J. Black, T. Ellis, and P. Rosin, "A Novel Method for Video Tracking Performance Evaluation," presented at the The Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, Nice, France, 2003.
- [16] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," presented at the Proceedings of the 23rd international conference on Machine learning, Pittsburgh, Pennsylvania, 2006.
- [17] Z. Kalal, J. Matas, and K. Mikolajczyk, "Online learning of robust object detectors during unstable tracking," in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, 2009, pp. 1417-1424.
- [18] J. Popoola and A. Amer, "Performance evaluation for tracking algorithms using object labels," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 2008, pp. 733-736.
- [19] K. Bernardin, A. Elbs, and R. Stiefelhausen, "Multiple Object Tracking Performance Metrics and Evaluation in a Smart Room Environment," presented at the The Sixth IEEE International Workshop on Visual Surveillance, VS 2006, Graz, Austria, 2006.
- [20] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Comput. Vis. Image Underst.*, vol. 104, pp. 90-126, 2006.
- [21] A. Baumann, M. Bolz, J. Ebling, M. Koenig, H. S. Loos, M. Merkel, *et al.*, "A Review and Comparison of Measures for Automatic Video Surveillance Systems," *EURASIP Journal on Image and Video Processing*, vol. 2008, 2008.
- [22] R. Stiefelhausen, R. Bowers, and J. G. Fiscus, *Multimodal Technologies for Perception of Humans, International Evaluation Workshops CLEAR 2007 and RT 2007, Baltimore, MD, USA, May 8-11, 2007, Revised Selected Papers*: Springer Berlin Heidelberg, 2008.
- [23] B. Bodt, R. Camden, H. Scott, A. Jacoff, T. Hong, T. Chang, *et al.*, "Performance measurements for evaluating static and dynamic multiple human detection and tracking systems in unstructured environments," presented at the Proceedings of the 9th Workshop on Performance Metrics for Intelligent Systems, Gaithersburg, Maryland, 2009.
- [24] A. Godil, R. Bostelman, K. Saidi, W. Shackelford, G. Cheok, M. Shneier, *et al.*, "3D Ground-Truth Systems for Object/Human Recognition and Tracking," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*, 2013, pp. 719-726.
- [25] Nikon Corporation. (6/2/2014). *iGPS/iSpace*. Available: http://www.nikon.com/products/instruments/lineup/industrial/metrology_assisted_production/large_volume/igps/index.htm
- [26] R. Schmitt, S. Nisch, Scho, x, A. nberg, F. Demeester, *et al.*, "Performance evaluation of iGPS for industrial applications," in *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, 2010, pp. 1-8.
- [27] G. Mosqueira, J. Apetz, K. M. Santos, E. Villani, R. Suterio, and L. G. Trabasso, "Analysis of the indoor GPS system as feedback for the robotic alignment of fuselages using laser radar measurements as comparison," *Robot. Comput.-Integr. Manuf.*, vol. 28, pp. 700-709, 2012.
- [28] Z. Wang, L. Mastrogiacomo, F. Franceschini, and P. G. Maropoulos, "Experimental comparison of dynamic tracking performance of iGPS and laser tracker," *The International Journal of Advanced Manufacturing Technology*, vol. 56, pp. 205-213, 2011.
- [29] C. Depenthal, "Path tracking with IGPS," in *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2010, pp. 1-6.
- [30] NaturalPoint Corporation. (6/2/2014). *OptiTrack*. Available: <https://www.naturalpoint.com/optitrack/products/flex-3/indepth.html>

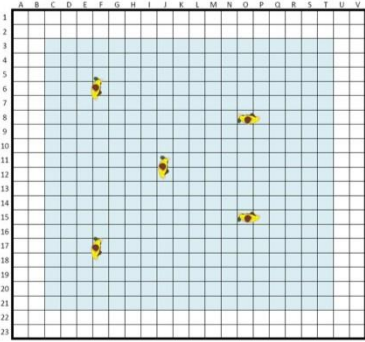
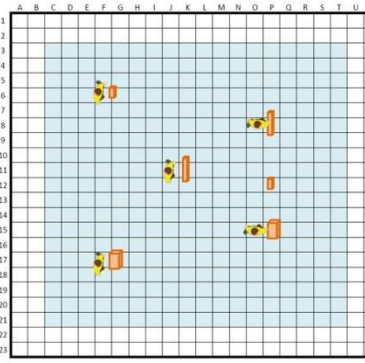
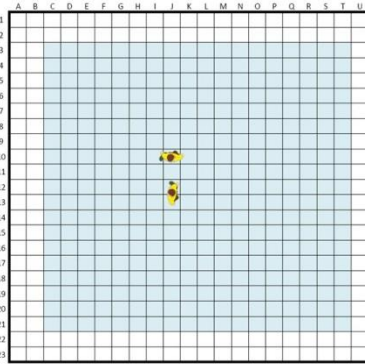
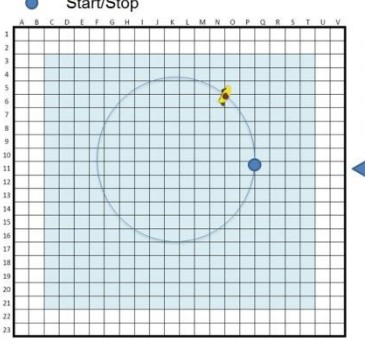
- [31] L. Sigal, A. Balan, and M. Black, "HumanEva: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated Human Motion," *International Journal of Computer Vision*, vol. 87, pp. 4-27, 3/2010 2010.
- [32] N. P. Van der Aa, X. Luo, G. J. Giezeman, R. T. Tan, and R. C. Veltkamp, "UMPM benchmark: A multi-person dataset with synchronized video and motion capture data for evaluation of articulated human motion and interaction," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, 2011, pp. 1264-1269.
- [33] A. O. Balan, L. Sigal, and M. J. Black, "A Quantitative Evaluation of Video-based 3D Person Tracking," presented at the Proceedings of the 14th International Conference on Computer Communications and Networks, 2005.
- [34] D. R. Chambers, C. Flannigan, and B. Wheeler, "High-accuracy real-time pedestrian detection system using 2D and 3D features," *SPIE Proceedings Three-Dimensional Imaging, Visualization, and Display*, vol. 83840G, pp. 83840G-1-83840G-11, 5/1/2012 2012.
- [35] N. Dalal. (2005). *The INRIA Person Dataset*. Available: <http://pascal.inrialpes.fr/data/human/>
- [36] C. P. Papageorgiou, M. Oren, and T. Poggio, "A general framework for object detection," in *Sixth International Conference on Computer Vision*, 1998, pp. 555-562.
- [37] Y. Freund and R. E. Schapire, "A short introduction to boosting," *Journal of Japanese Society for Artificial Intelligence*, vol. 14, pp. 771-780, 1999.
- [38] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273-297, 1995.
- [39] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR*, 2005, pp. 886-893.
- [40] B. Bodt and T. Hong, "UGV Safe Operations Capstone Experiment," presented at the Army Science Conference, Orlando, Florida, 2010.
- [41] W. Shackleford. (2014). *PTPM*. Available: <https://github.com/usnistgov/PTPM>
- [42] ISO TC 199, "ISO 13855:2010 Safety of machinery -- Positioning of safeguards with respect to the approach speeds of parts of the human body," International Organization for Standardization, Ed., ed, 2010.

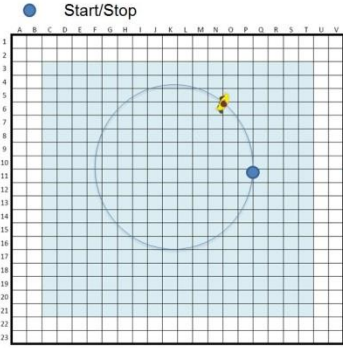
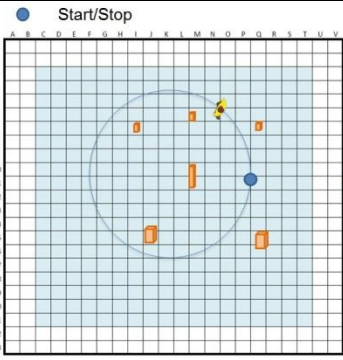
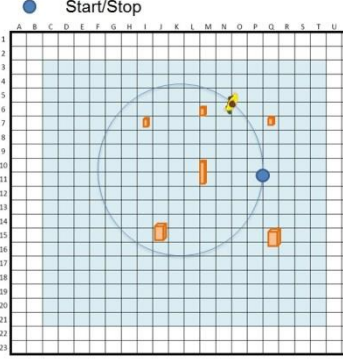
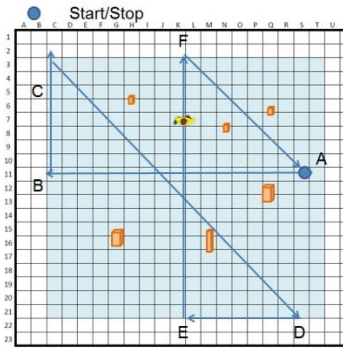
Appendix A – Description of the 34 experiments

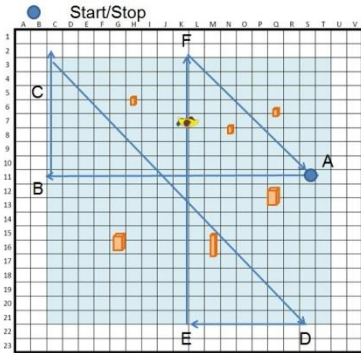
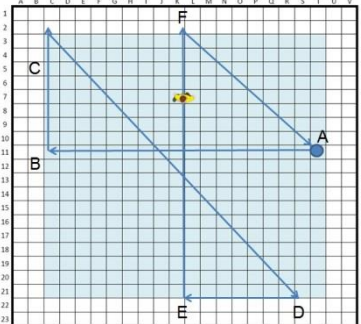
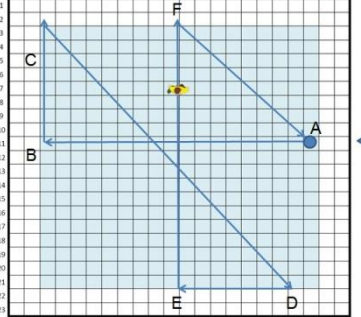
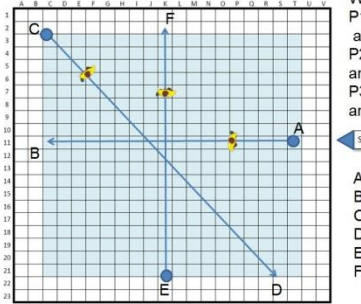
Key for obstacles: #1 – (521 x 521 x 1448) mm [(20.5 x 20.5 x 57) in] box
 #2 – (419 x 419 x 1524) mm [(16.5 x 16.5 x 60)] in box
 #3 – (419 x 419 x 1861) mm [(16.5 x 16.5 x 73.25) in] box
 #4 – (267 x 267 x 1905) mm [(10.5 x 10.5 x 75) in] box
 #5 – (267 x 267 x 1543) mm [(10.5 x 10.5 x 60.75) in] box
 #6 – (267 x 267 x 1454) mm [(10.5 x 10.5 x 57.25) in] box

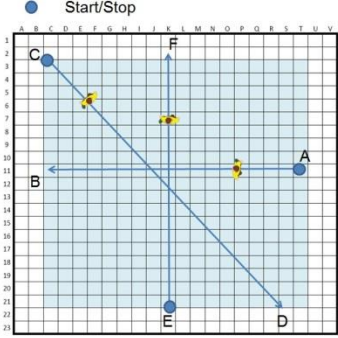
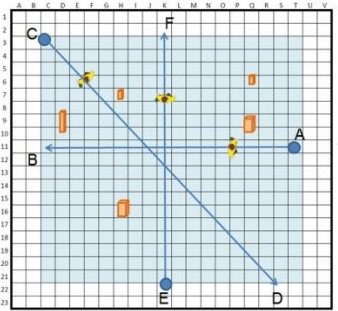
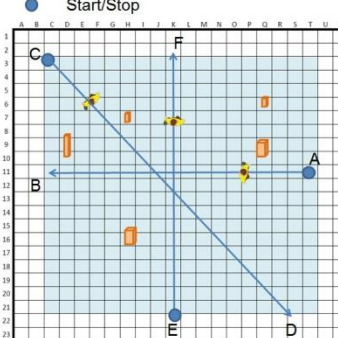
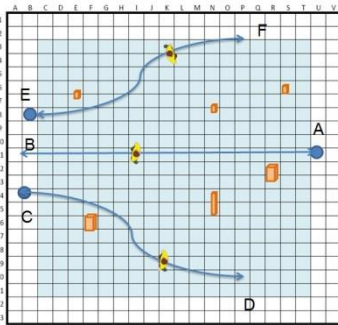
Test #	Description	Test Layout
1	Stat-1: Three stationary people on Col. Q.	 <p>People at Q7 Q11 Q15</p> <p>Sensors</p>
2	Stat-2: Three stationary people on Col. J.	 <p>People at J7 J11 J15</p> <p>Sensors</p>
3	Stat-3: Three stationary people on Col. C.	 <p>People at C7 C11 C15</p> <p>Sensors</p>

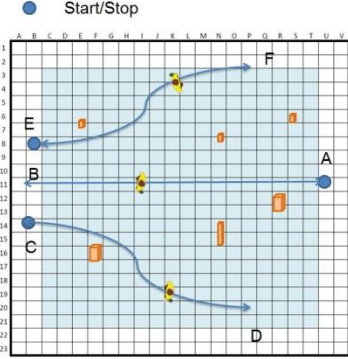
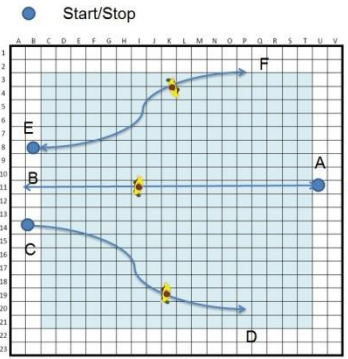
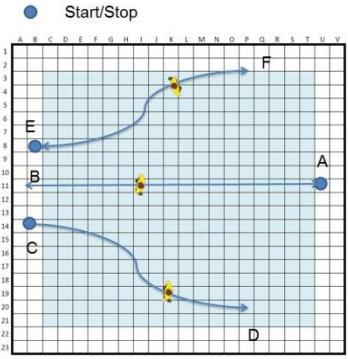
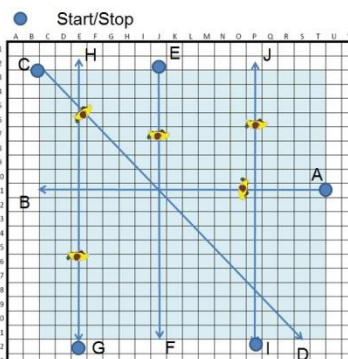
4	<p>Stat-4: Three stationary people on Col. Q with obstacles</p>	 <p>People at Q7 Q11 Q15</p> <p>Box 6 at R7 Box 1 at R11 Box 2 at R15</p>
5	<p>Stat-5: Three stationary people on Col. J with obstacles</p>	 <p>People at J7 J11 J15</p> <p>Box 6 at K7 Box 1 at K11 Box 2 at K15</p>
6	<p>Stat-6: Three stationary people on Col. C with obstacles</p>	 <p>People at C7 C11 C15</p> <p>Box 6 at D7 Box 1 at D11 Box 2 at D15</p>
7	<p>Stat-7: Five stationary people</p>	 <p>People at O8 O15 J11 F6 F17</p> <p>Note: We may have to move the people near the front closer together to fit in the Kinect FOV</p>

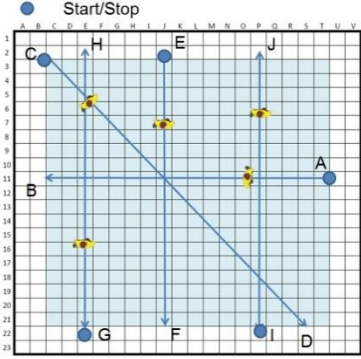
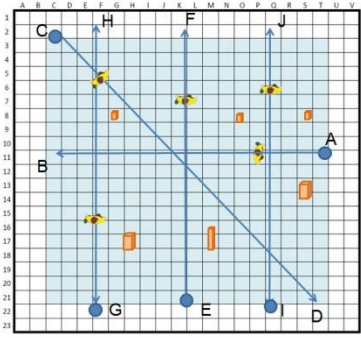
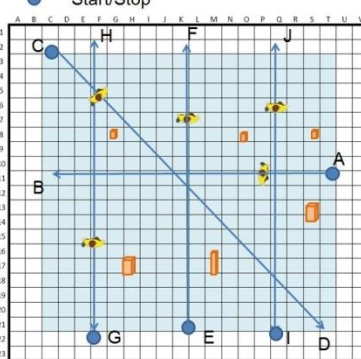
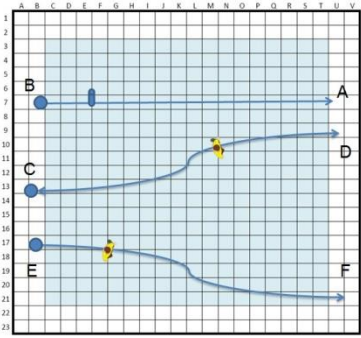
8	Stat-8: Five stationary people gesturing	 <p>People at O8 O15 J11 F6 F17</p> <p>Sensors</p> <p>Note: We may have to move the people near the front closer together to fit in the Kinect FOV</p>
9	Stat-9: Five stationary people with obstacles	 <p>People at O8 O15 J11 F6 F17</p> <p>Sensors</p> <p>Box 1 at P8 Box 2 at P15 Box 3 at G6 Box 4 at G17 Box 5 at K12.5 Box 6 at P12</p>
10	Stat-10: Two stationary people 0.5 m apart	 <p>People at J10 J12.5</p> <p>Sensors</p>
11	Single-1: Person walking in elliptical path	 <p>Start/Stop</p> <p>Person walks in circle, starting at P11, passing through K4, F11, K16.5</p> <p>Sensors</p>

12	<i>Single-2: Person jogging in elliptical path</i>	 <p>Person jogs in circle, starting at P11, passing through K4, F11, K16.5</p>
13	<i>Single-3: Person walking in elliptical path with obstacles</i>	 <p>Person walks in circle, starting at P11, passing through K4, F11, K16.5</p> <p>Box 1 at M10.5 Box 2 at Q15 Box 3 at I7 Box 4 at M6 Box 5 at Q7 Box 6 at J15</p>
14	<i>Single-4: Person jogging in elliptical path with obstacles</i>	 <p>Person jogs in circle, starting at P11, passing through K4, F11, K16.5</p> <p>Box 1 at M10.5 Box 2 at Q15 Box 3 at I7 Box 4 at M6 Box 5 at Q7 Box 6 at J15</p>
15	<i>Single-5: Single person walking a complex path with obstacles</i>	 <p>Person walks from</p> <p>A: T11 to B: C11 to C: C2.5 to D: S21.5 to E: K21.5 to F: K2 to A: T11</p> <p>Box 1 at M16 Box 2 at Q12.5 Box 3 at G15.5 Box 4 at N7.5 Box 5 at Q6.5 Box 6 at H5.5</p>

16	Single-6: Single person jogging a complex path with obstacles	 <p>Person jogs from</p> <ul style="list-style-type: none"> A: T11 to B: C11 to C: C2.5 to D: S21.5 to E: K21.5 to F: K2 to A: T11 <p>Box 1 at M16 Box 2 at Q12.5 Box 3 at G15.5 Box 4 at N7.5 Box 5 at Q6.5 Box 6 at H5.5</p>
17	Single-7: Single person walking a complex path	 <p>Person walks from</p> <ul style="list-style-type: none"> A: T11 to B: C11 to C: C2.5 to D: S21.5 to E: K21.5 to F: K2 to A: T11
18	Single-8: Single person jogging a complex path	 <p>Person jogs from</p> <ul style="list-style-type: none"> A: T11 to B: C11 to C: C2.5 to D: S21.5 to E: K21.5 to F: K2 to A: T11
19	Multiple-1: Three people walking on straight paths	 <p>Walking People:</p> <ul style="list-style-type: none"> P1 walks from A to B and returns to A P2 walks from C to D and returns to C P3 walks from E to F and returns to E <ul style="list-style-type: none"> A: T11 B: C11 C: C2.5 D: S21.5 E: K21.5 F: K2

20	Multiple-2: Three people jogging on straight paths	 <p>Jogging People: P1 jogs from A to B and returns to A P2 jogs from C to D and returns to C P3 jogs from E to F and returns to E</p> <p>Sensors</p> <p>A: T11 B: C11 C: C2.5 D: S21.5 E: K21.5 F: K2</p>
21	Multiple-3: Three people walking on straight paths with obstacles	 <p>Walking People: P1 walks from A to B and returns to A P2 walks from C to D and returns to C P3 walks from E to F and returns to E</p> <p>Sensors</p> <p>Box 1 at D9 Box 2 at Q9 Box 3 at H16 Box 4 at H7 Box 5 at Q6</p>
22	Multiple-4: Three people jogging on straight paths with obstacles	 <p>Jogging People: P1 jogs from A to B and returns to A P2 jogs from C to D and returns to C P3 jogs from E to F and returns to E</p> <p>Sensors</p> <p>Box 1 at D9 Box 2 at Q9 Box 3 at H16 Box 4 at H7 Box 5 at Q6</p>
23	Multiple-5: Three people walking curved paths with obstacles	 <p>P1 walks from A to B and returns to A P2 walks from C to D and returns to C P3 walks from E to F and returns to E</p> <p>A = U11, B = B11 C = B14, D = O19</p> <p>Sensors</p> <p>E = B8, F = P2</p> <p>Box 1 at N14.5 Box 2 at R12 Box 3 at F16 Box 4 at S6 Box 5 at N7.5 Box 6 at E6.5</p>

24	Multiple-6: Three people jogging curved paths with obstacles	 <p>Start/Stop</p> <p>P1 jogs from A to B and returns to A P2 jogs from C to D and returns to C P3 jogs from E to F and returns to E A = U11, B = B11 C = B14, D = O19</p> <p>Sensors</p> <p>E = B8, F = P2</p> <p>Box 1 at N14.5 Box 2 at R12 Box 3 at F16 Box 4 at S6 Box 5 at N7.5 Box 6 at E6.5</p>
25	Multiple-7: Three people walking curved paths	 <p>Start/Stop</p> <p>Walking People: P1 walks from A to B and returns to A P2 walks from C to D and returns to C P3 walks from E to F and returns to E</p> <p>Sensors</p> <p>A = U11, B = B11 C = B14, D = O19 E = B8, F = P2</p>
26	Multiple-8: Three people jogging curved paths	 <p>Start/Stop</p> <p>Jogging People: P1 jogs from A to B and returns to A P2 jogs from C to D and returns to C P3 jogs from E to F and returns to E</p> <p>Sensors</p> <p>A = U11, B = B11 C = B14, D = O19 E = B8, F = P2</p>
27	Multiple-9: Five people walking on straight paths	 <p>Start/Stop</p> <p>P1 walks from U11 to B11 and back P2 walks from B2 to S22 and back P3 walks from J2 to J2 and back P4 walks from E22 to E2 and back P5 walks from P22 to P2 and back</p> <p>Sensors</p> <p>A = U11, B = B11 C = B2, D = S22 E = J2, F = J22 G = E22, H = E2 I = P22, J = P2</p>

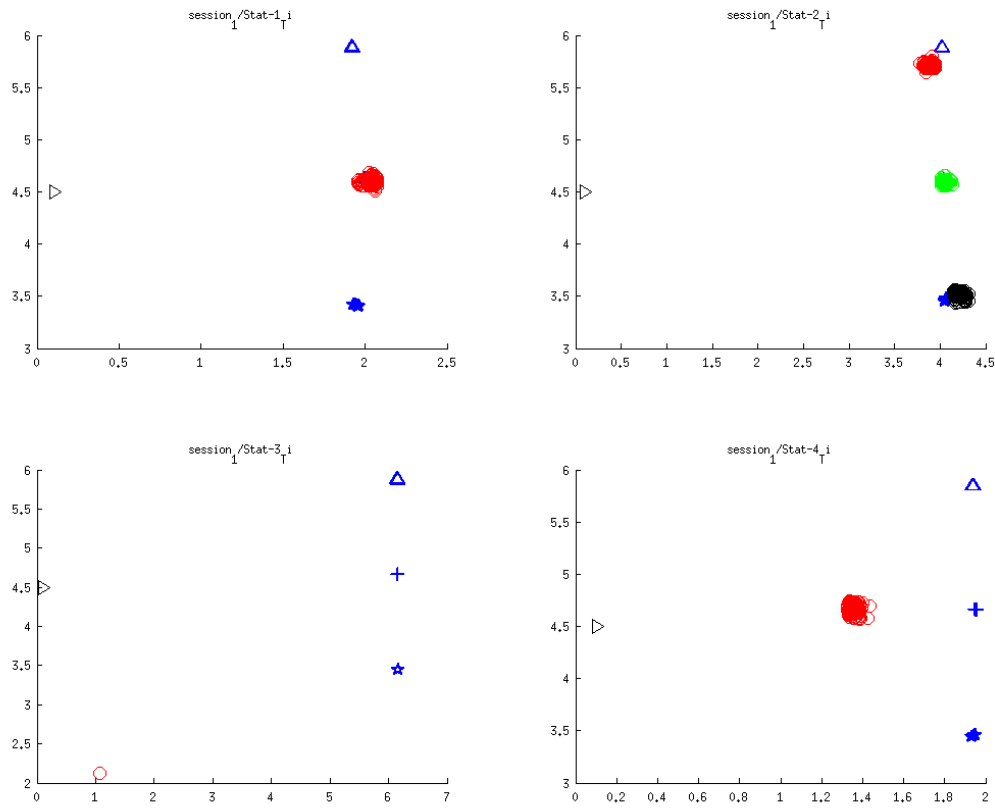
28	Multiple-10: Five people jogging on straight paths	 <p>Start/Stop</p> <p>P1 jogs from U11 to B11 and back P2 jogs from B2 to S22 and back P3 jogs from J2 to J2 and back P4 jogs from E22 to E2 and back P5 jogs from P22 to P2 and back</p> <p>Sensors</p> <p>A = U11, B = B11 C = B2, D = S22 E = J2, F = J22 G = E22, H = E2 I = P22, J = P2</p>
29	Multiple-11: Five people walking on straight paths with obstacles	 <p>Start/Stop</p> <p>P1 walks from U11 to B11 and back P2 walks from B2 to S22 and back P3 walks from J2 to J2 and back P4 walks from E22 to E2 and back P5 walks from P22 to P2 and back</p> <p>Sensors</p> <p>Box 1 at M17 Box 2 at S13 Box 3 at H17 Box 4 at S8 Box 5 at O8 Box 6 at G8</p>
30	Multiple-12: Five people jogging on straight paths with obstacles	 <p>Start/Stop</p> <p>P1 jogs from U11 to B11 and back P2 jogs from B2 to S22 and back P3 jogs from J2 to J2 and back P4 jogs from E22 to E2 and back P5 jogs from P22 to P2 and back</p> <p>Sensors</p> <p>Box 1 at M17 Box 2 at S13 Box 3 at H17 Box 4 at S8 Box 5 at O8 Box 6 at G8</p>
31	Sled-I: Sled with walking people	 <p>Start</p> <p>P1 walks from B13 to U9 and back P2 walks from B17 to U21 and back Sled moves from B7 to U7 and back</p> <p>Sensors</p>

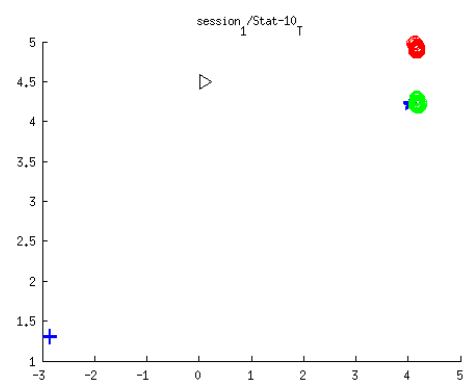
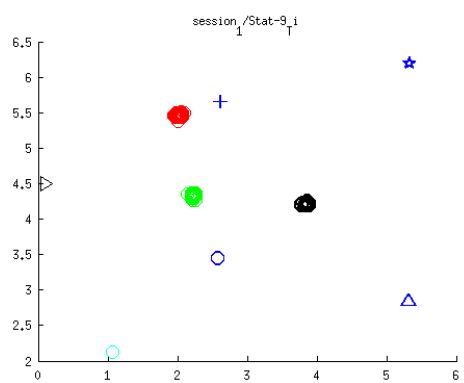
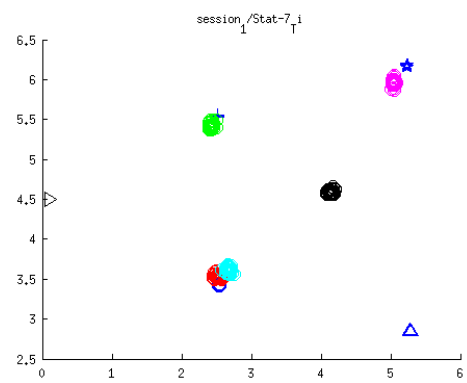
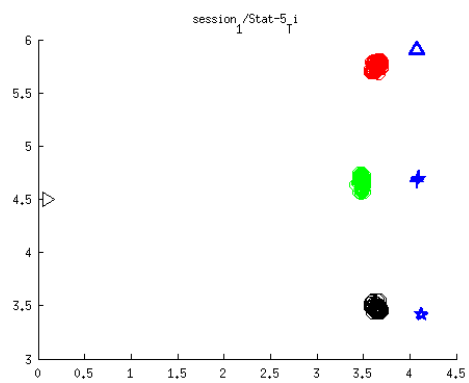
<p>32</p>	<p><i>Sled-2: Sled with jogging people</i></p>	<div data-bbox="743 195 1255 562"> <div> <div>● Start</div> <div> P1 jogs from B13 to U9 and back P2 jogs from B17 to U21 and back Sled moves from B7 to U7 and back </div> <div>Sensors</div> </div> </div>
<p>33</p>	<p><i>Sled-3: Sled with walking people and obstacles</i></p>	<div data-bbox="743 575 1255 940"> <div> <div>● Start</div> <div> P1 walks from B13 to U9 and back P2 walks from B17 to U21 and back Sled moves from B7 to U7 and back </div> <div>Sensors</div> <div> Box 1 at M15 Box 2 at R14 Box 3 at G15 Box 4 at R8 Box 5 at J9 Box 6 at F5 </div> </div> </div>
<p>34</p>	<p><i>Sled-4: Sled with jogging people and obstacles</i></p>	<div data-bbox="743 957 1255 1325"> <div> <div>● Start</div> <div> P1 jogs from B13 to U9 and back P2 jogs from B17 to U21 and back Sled moves from B7 to U7 and back </div> <div>Sensors</div> <div> Box 1 at M15 Box 2 at R14 Box 3 at G15 Box 4 at R8 Box 5 at J9 Box 6 at F5 </div> </div> </div>

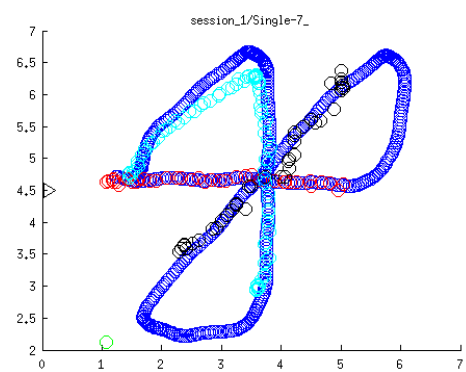
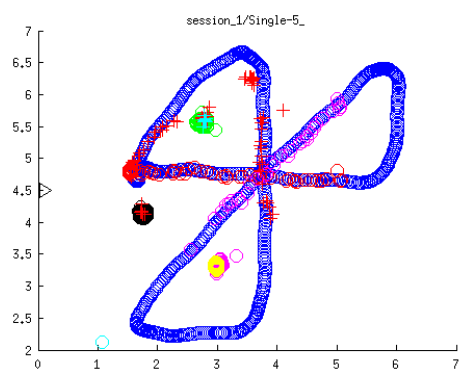
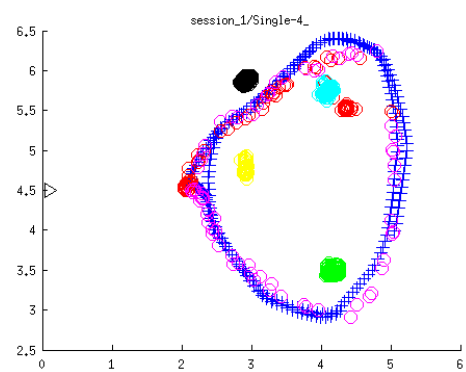
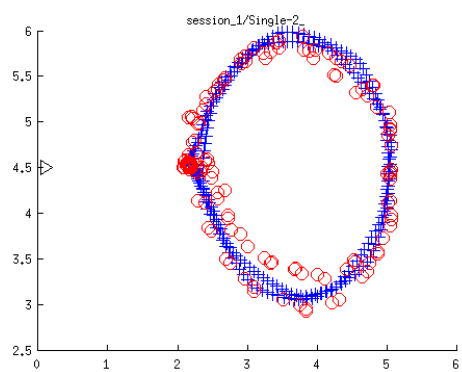
Appendix B – Data Plots for Detection and Tracking Measurements

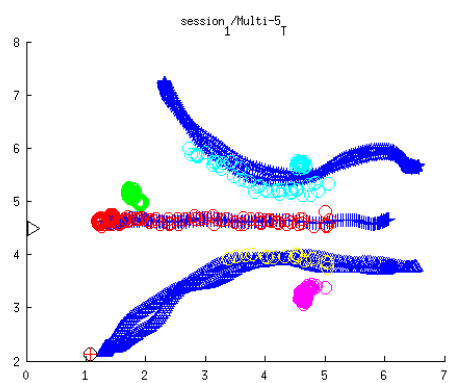
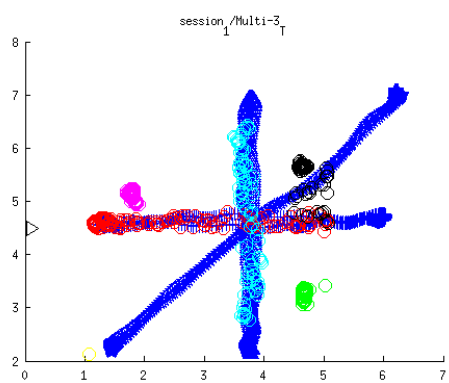
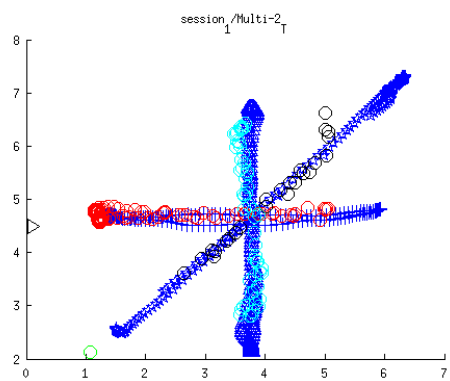
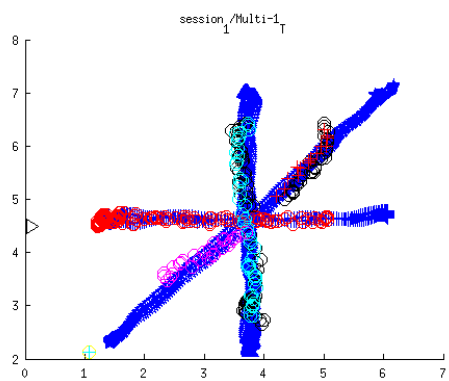
The experiments were carried out over five sessions. This Appendix provides plots of the detections for each experiment in each session. The individual experiments were the same for each session (as shown in Appendix A) but the human subjects were different. A total of thirteen people participated.

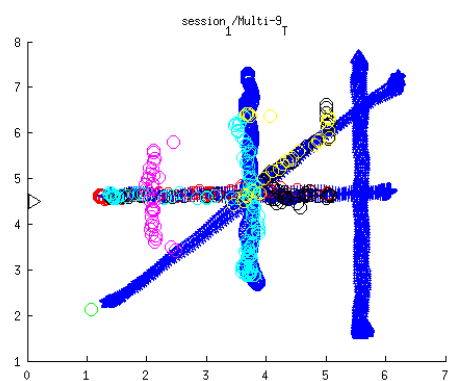
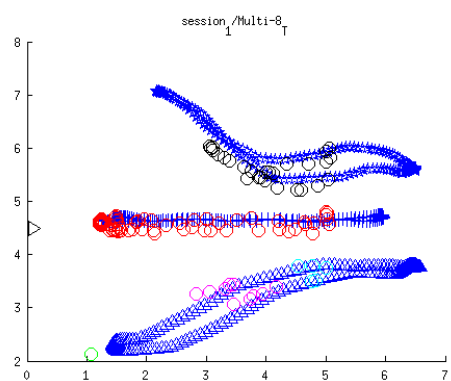
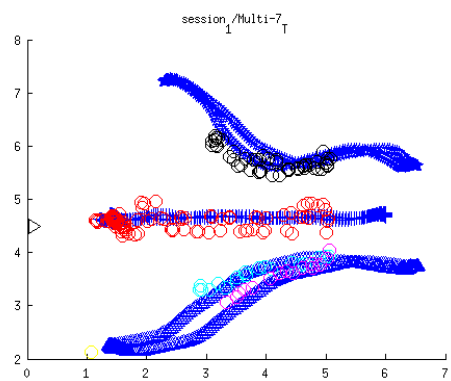
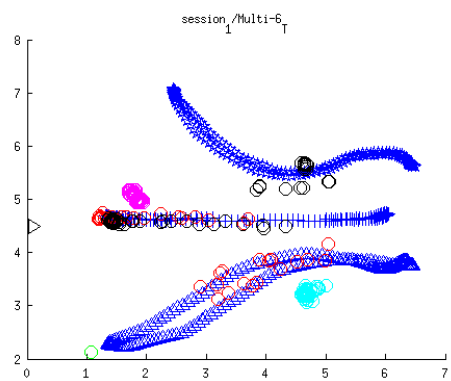
B.1: Shows the results between the GT data and the SUT data for the session 1 experiments. The blue color is the GT data. Other colors are the SUT data (different colors indicate different humans). The axes indicate the spatial positions or paths of the people in meters.

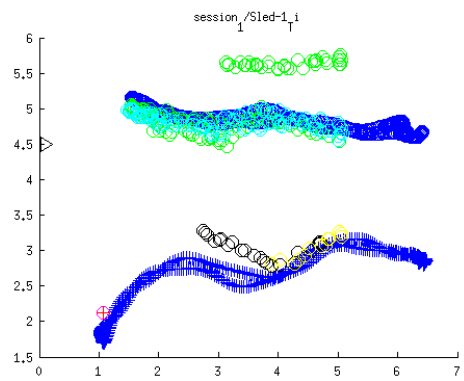
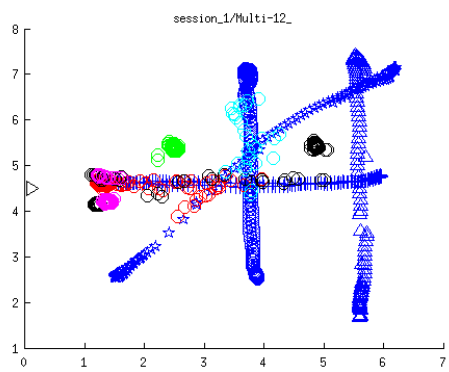
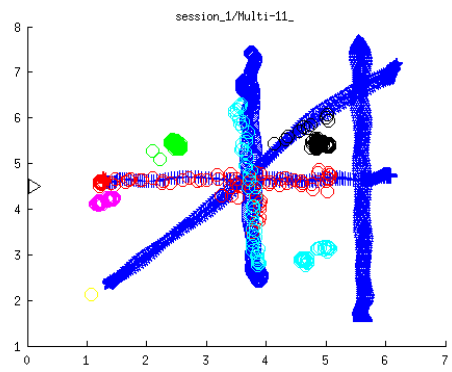
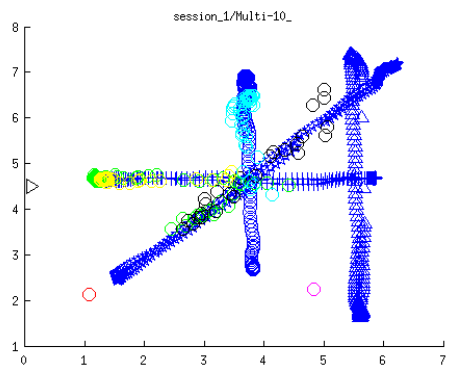




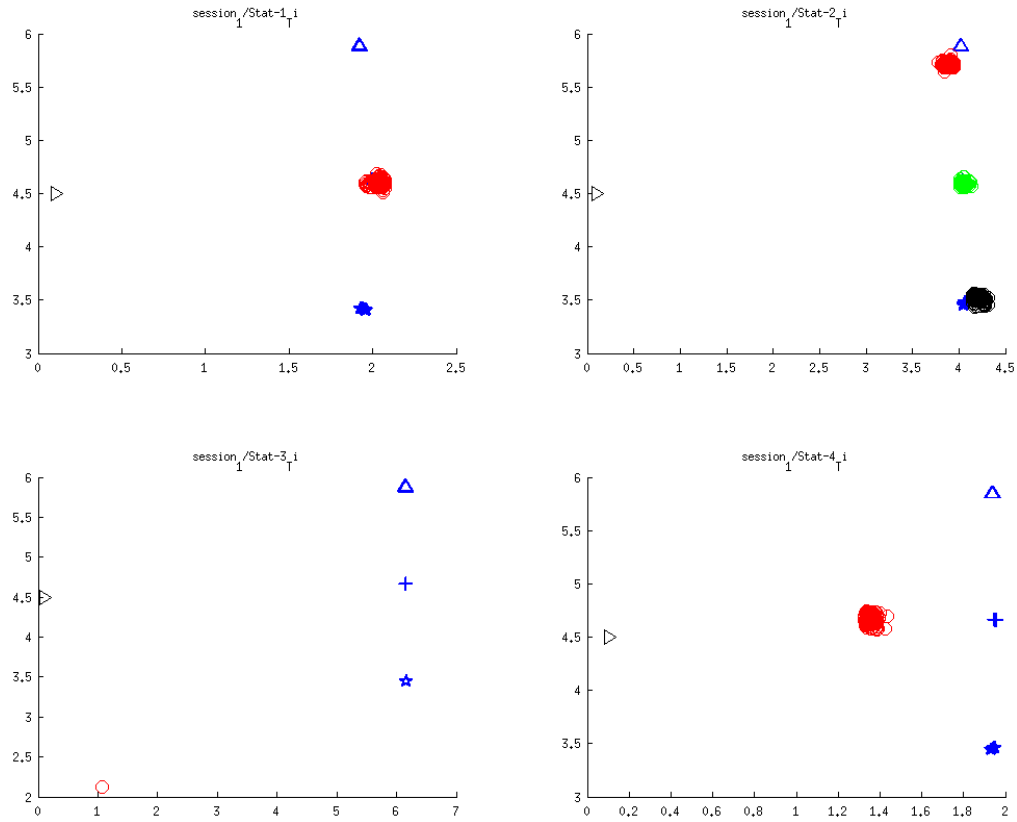


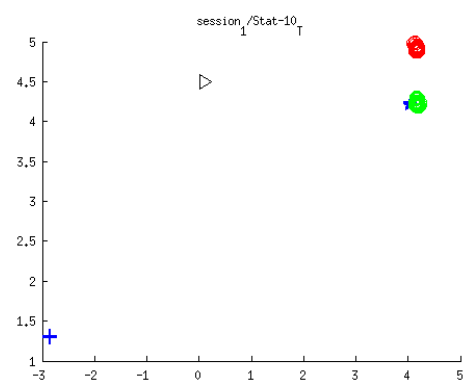
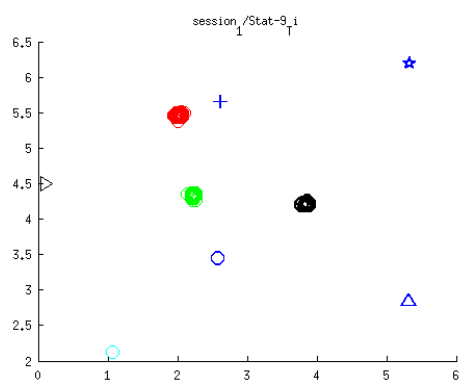
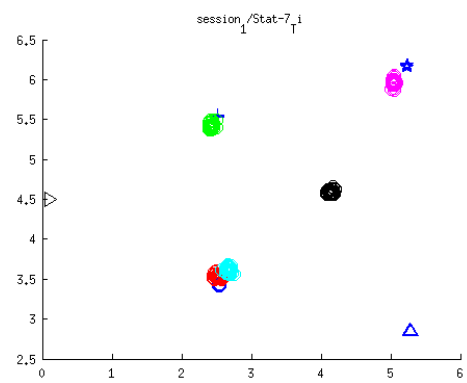
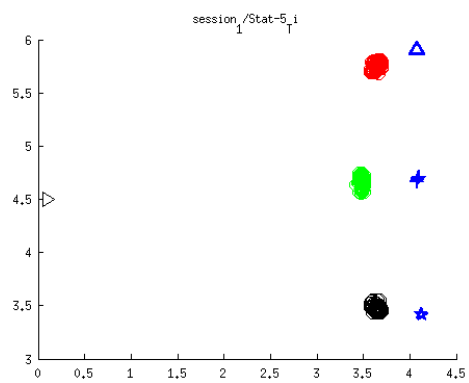


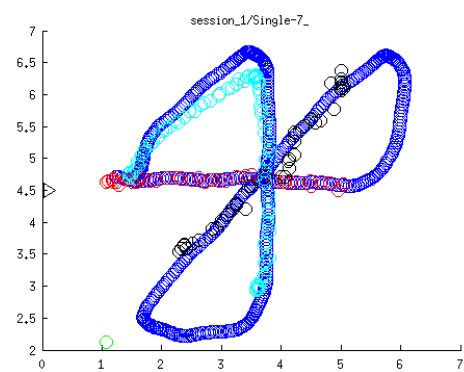
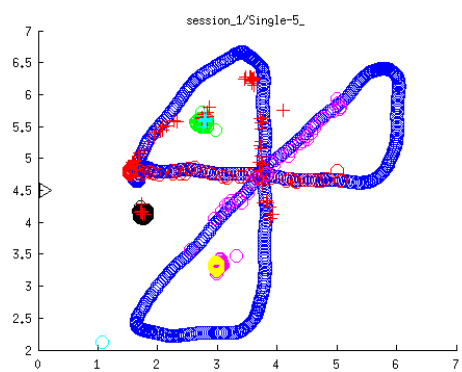
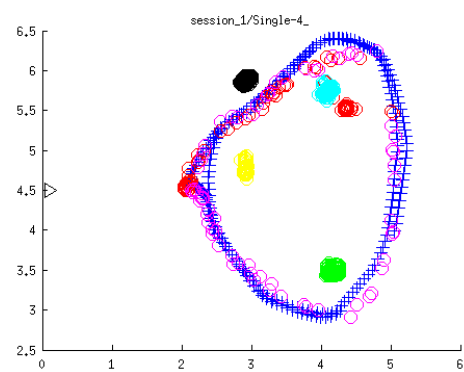
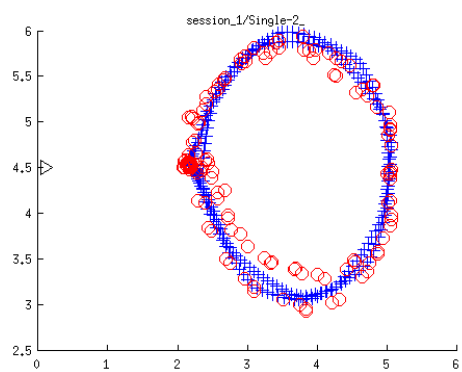


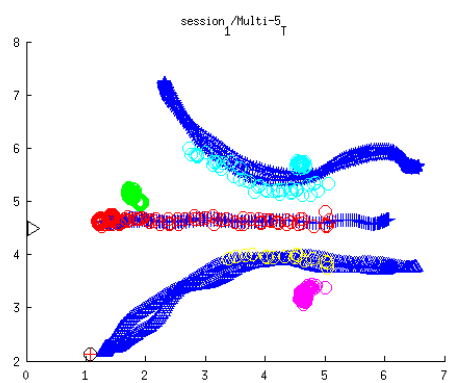
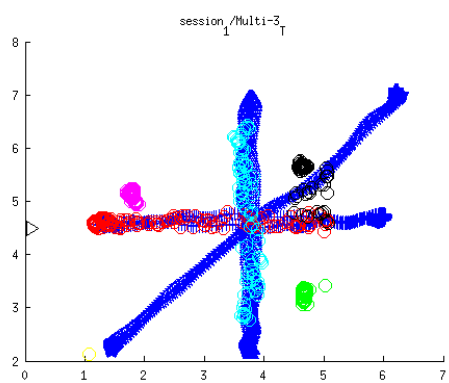
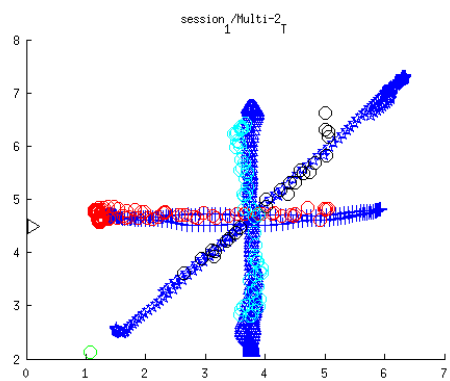
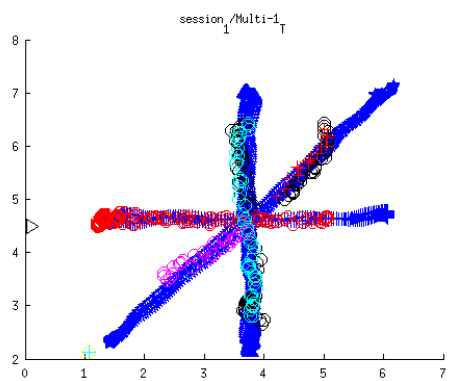


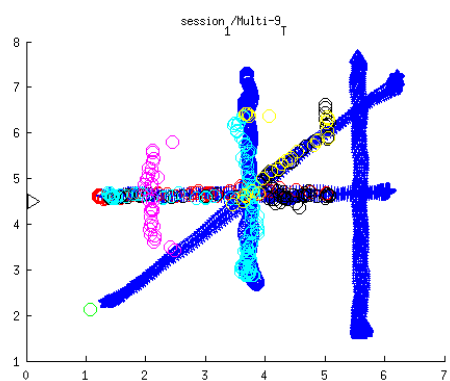
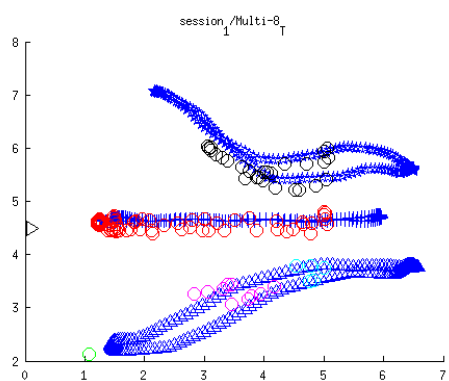
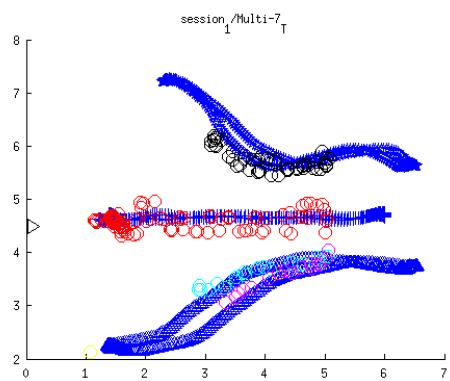
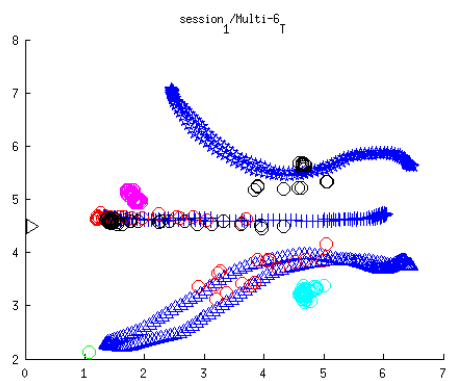
B.2: Shows the results between the GT data and the SUT data for the session 2 experiments. The blue color is the GT data. Other colors are the SUT data (different colors indicate different humans). The axes indicate the spatial positions or paths of the people in meters.

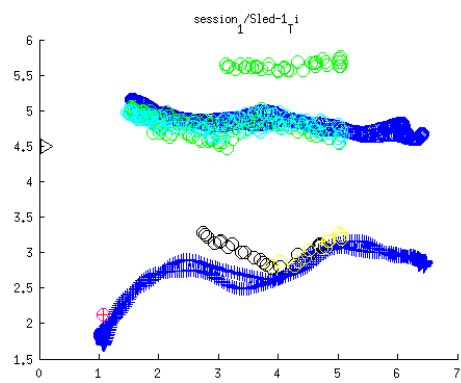
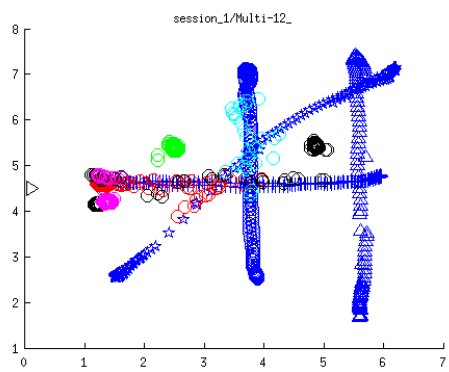
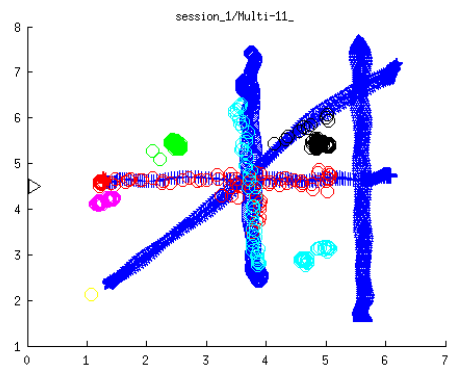
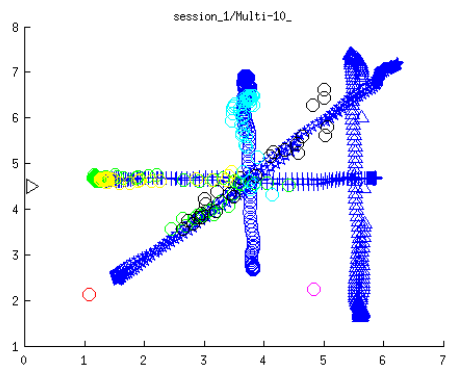




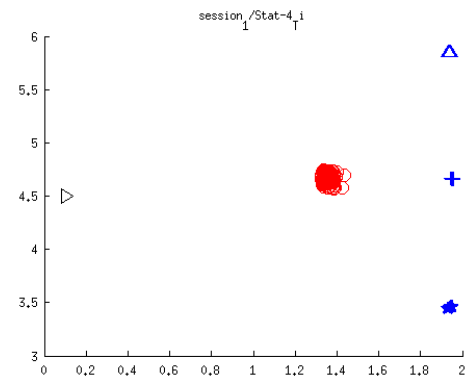
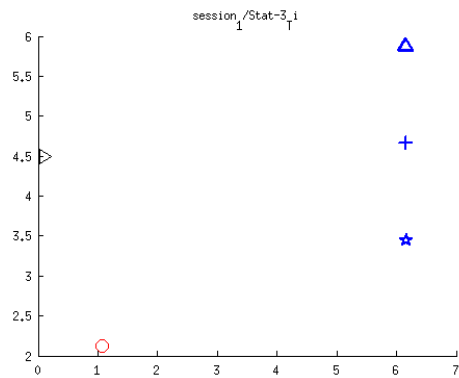
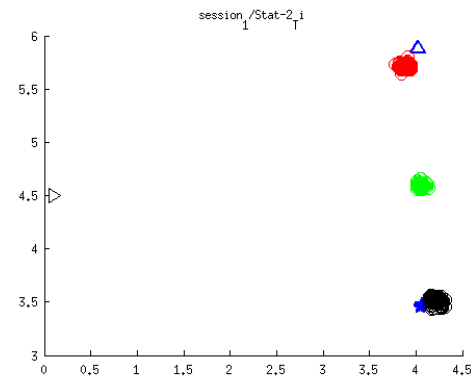
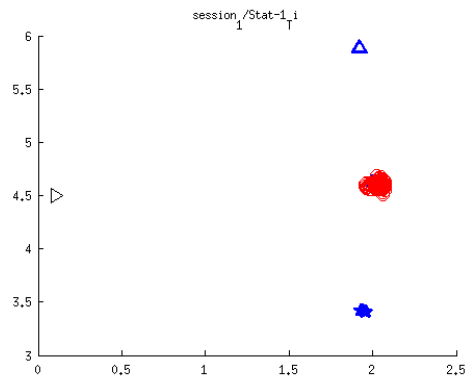


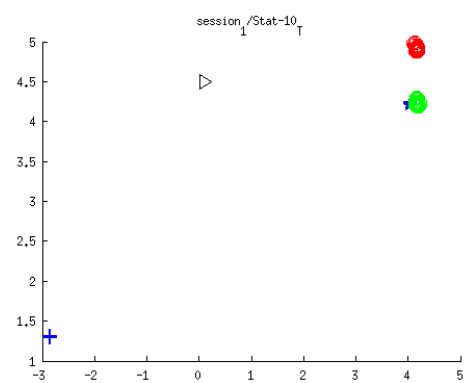
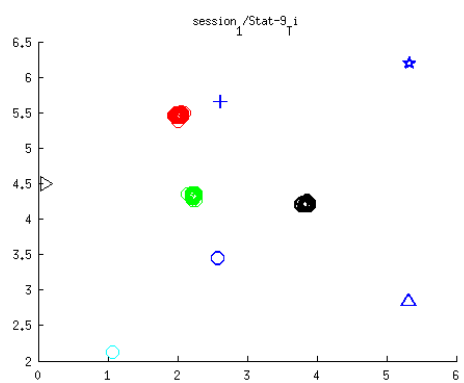
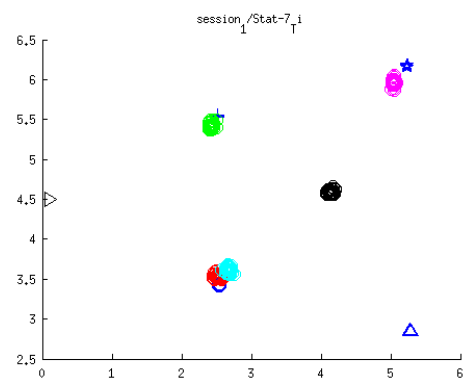
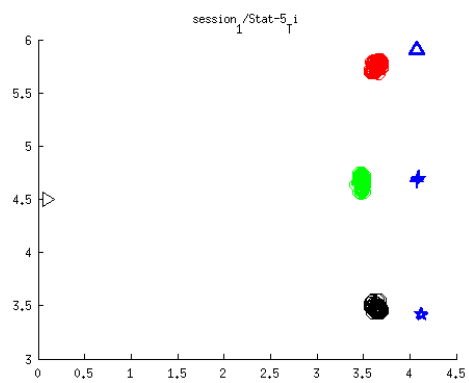


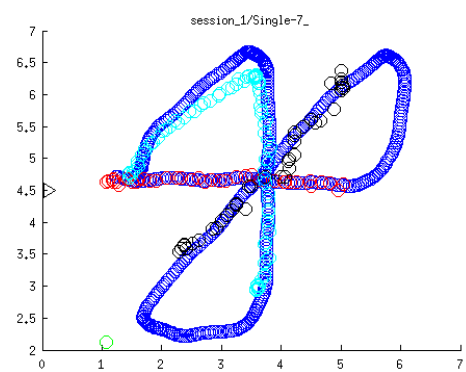
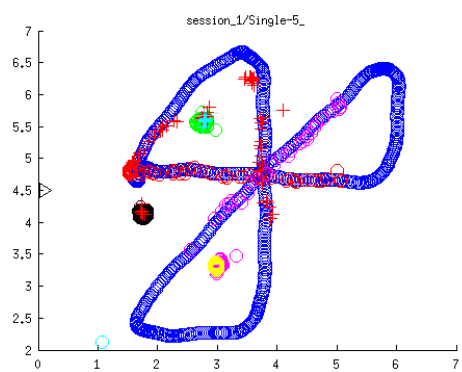
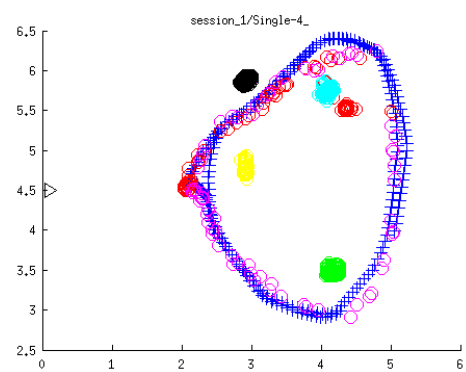
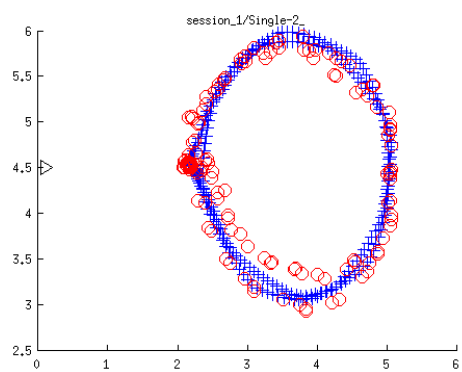


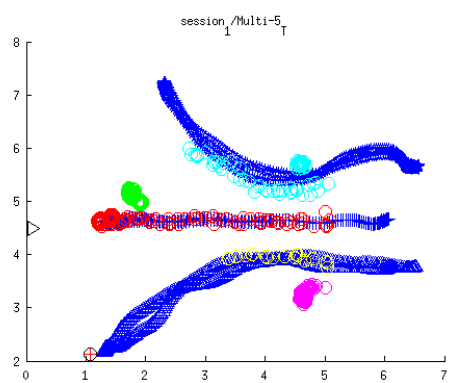
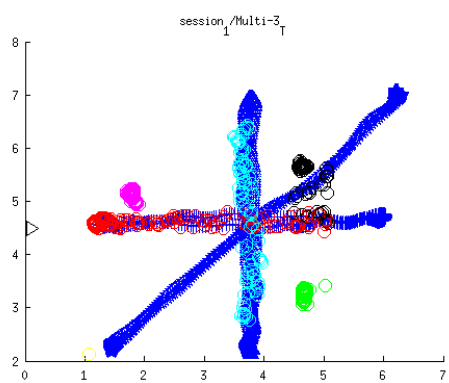
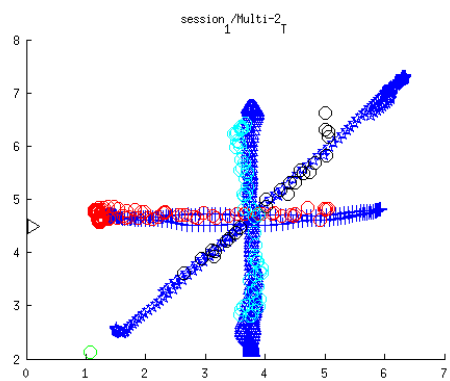
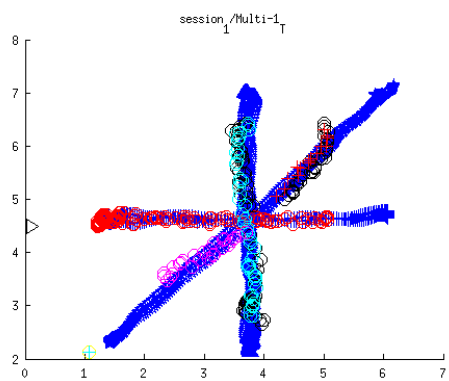


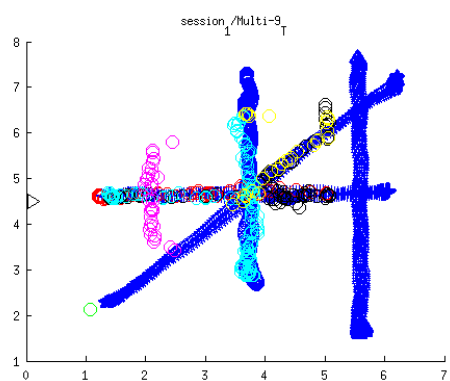
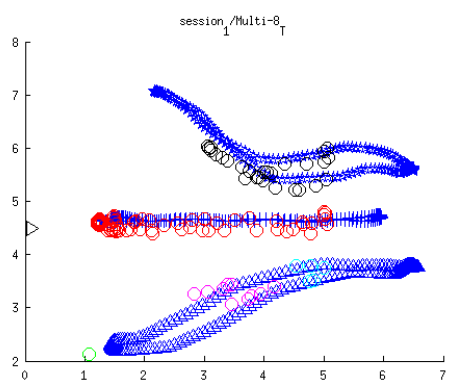
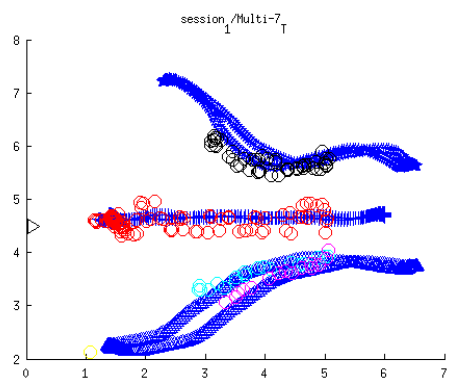
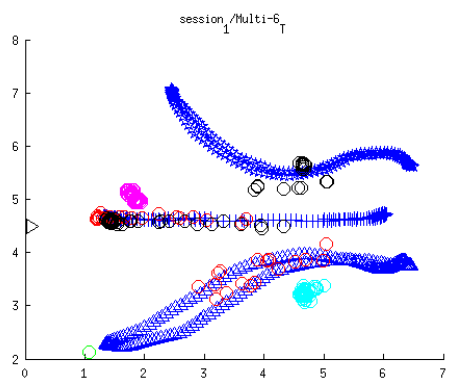
B.3: Shows the results between the GT data and the SUT data for the session 3 experiments. The blue color is the GT data. Other colors are the SUT data (different colors indicate different humans). The axes indicate the spatial positions or paths of the people in meters.

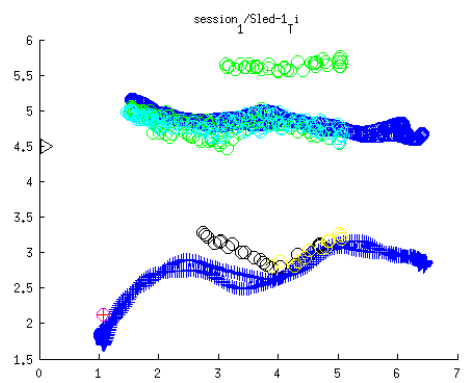
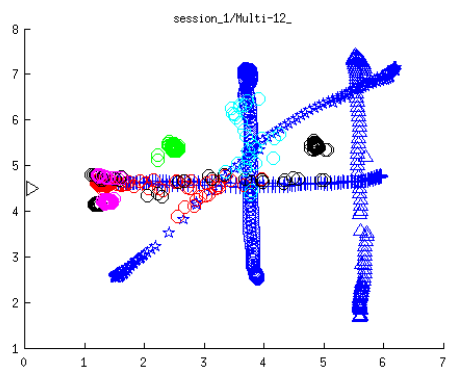
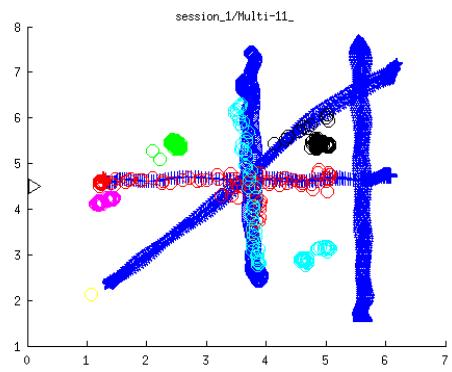
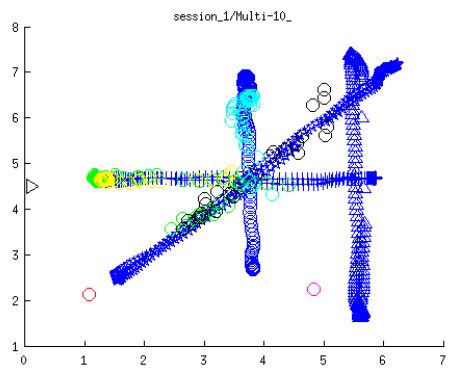




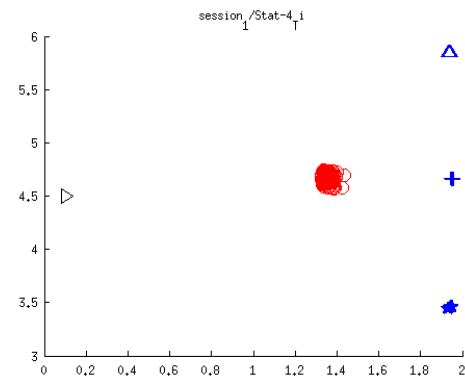
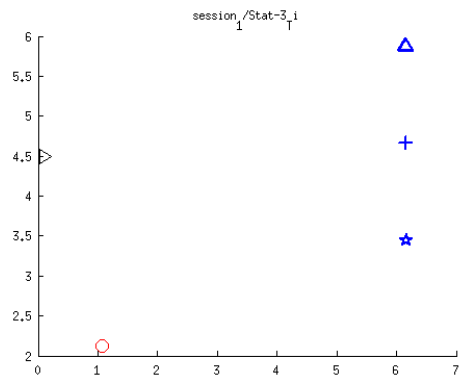
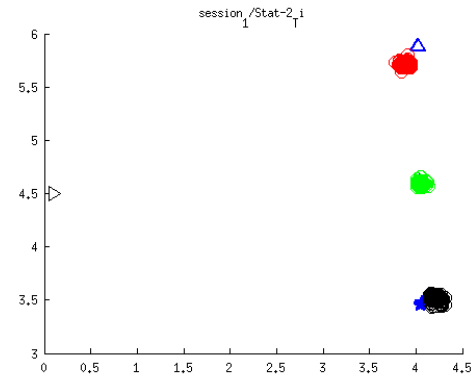
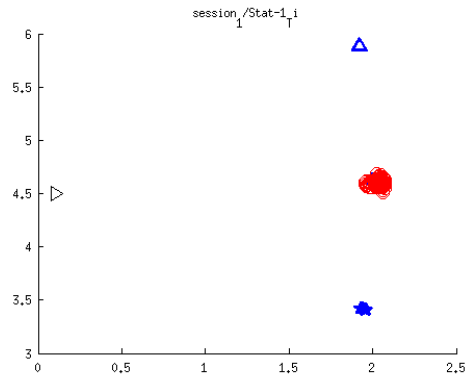


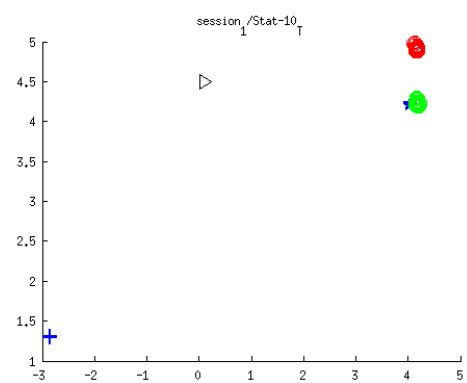
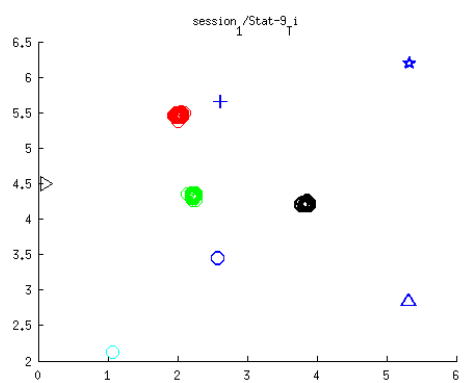
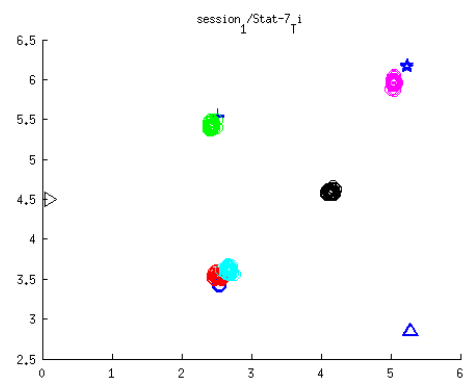
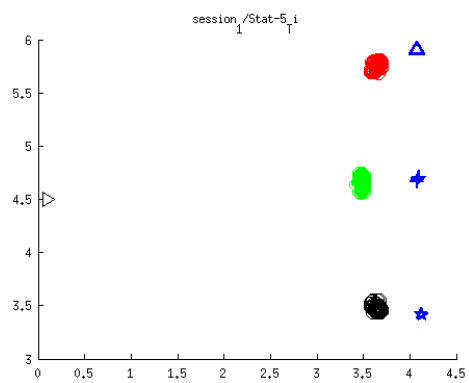


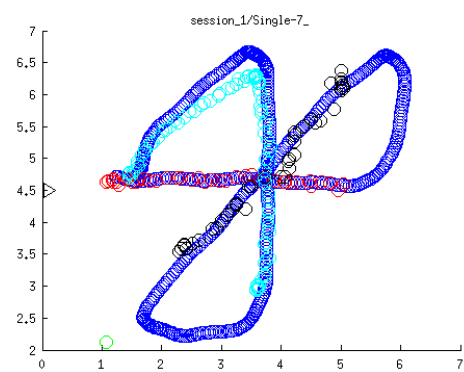
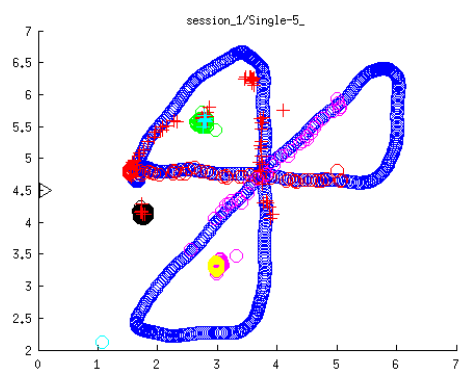
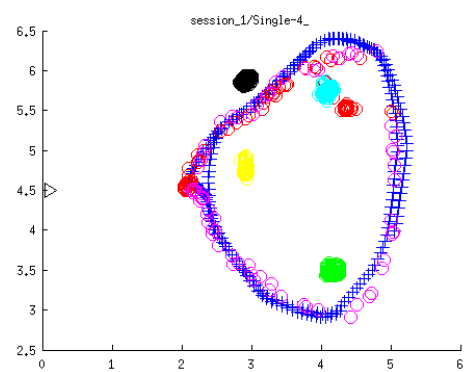
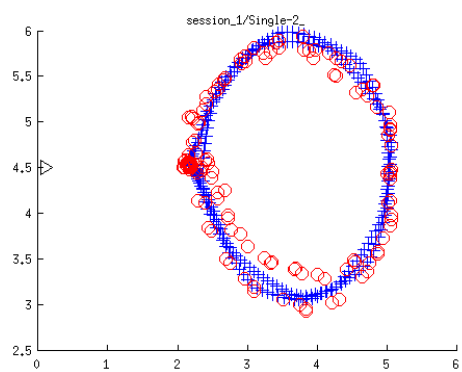


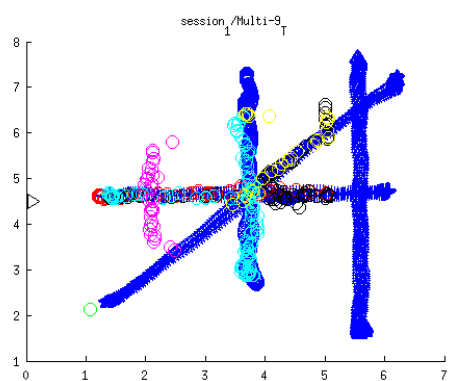
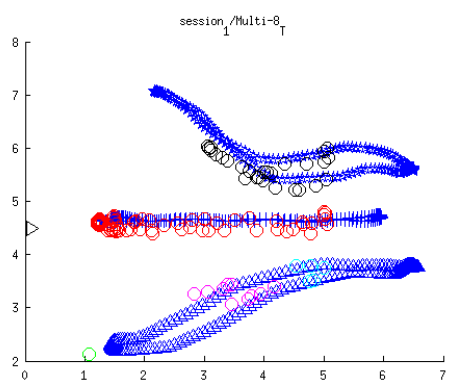
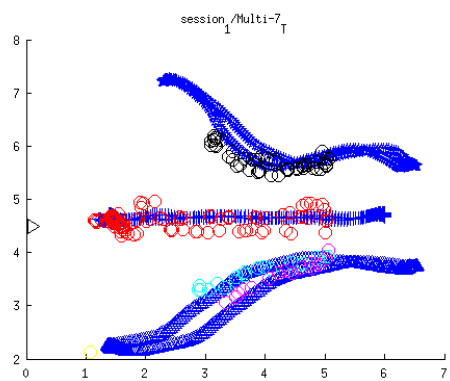
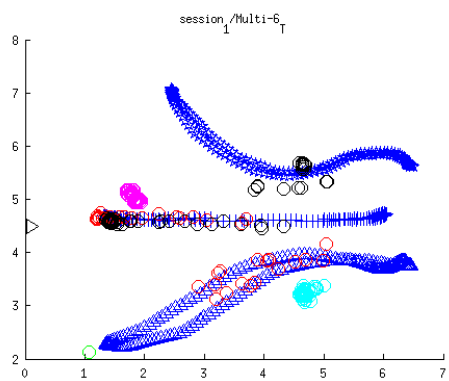


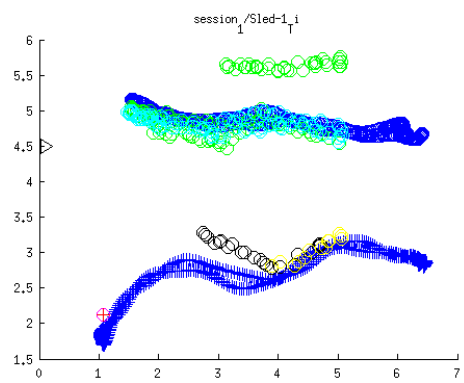
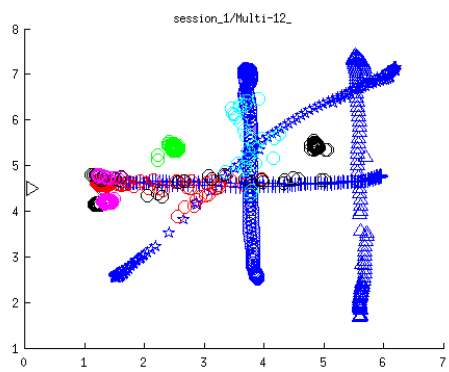
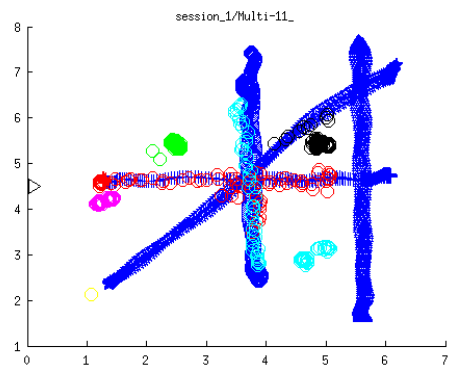
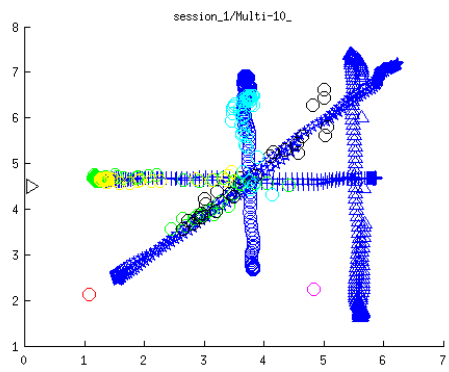
B.4: Shows the results between the GT data and the SUT data for the session 4 experiments. The blue color is the GT data. Other colors are the SUT data (different colors indicate different humans). The axes indicate the spatial positions or paths of the people in meters.



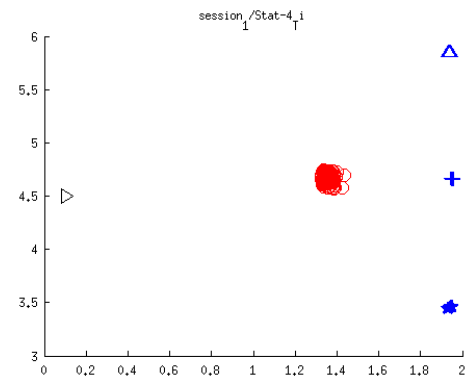
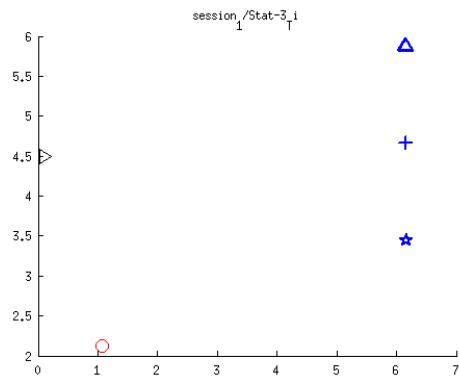
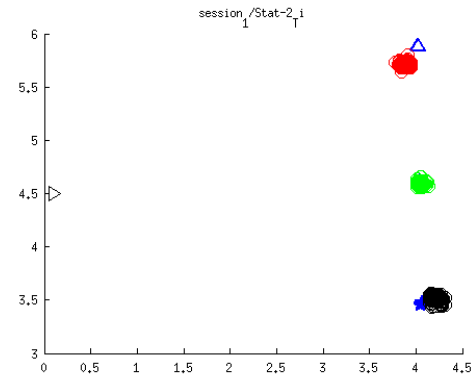
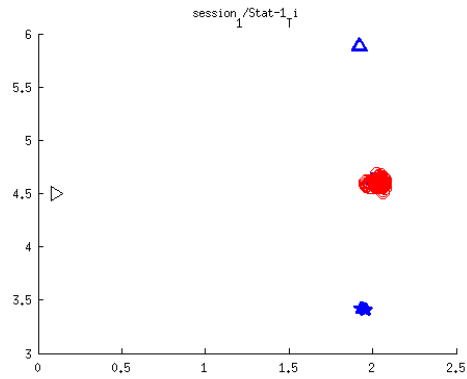


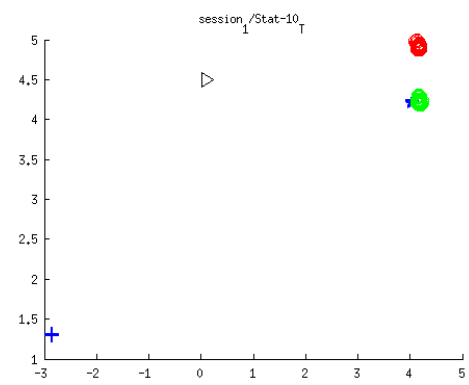
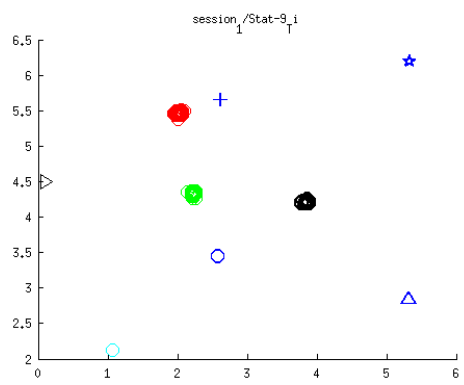
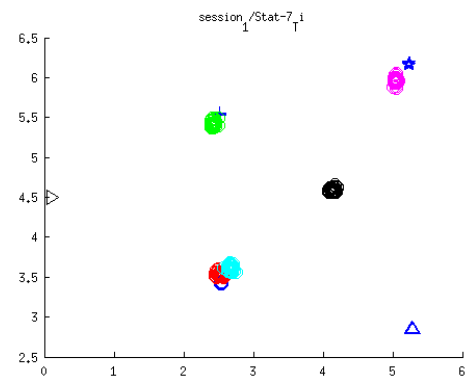
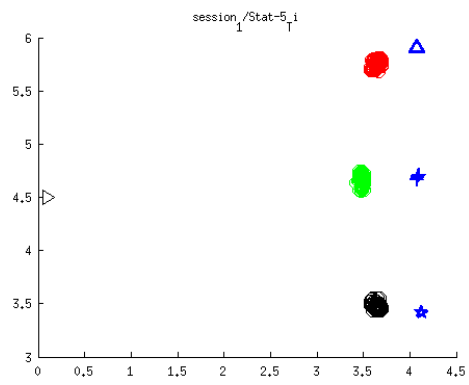


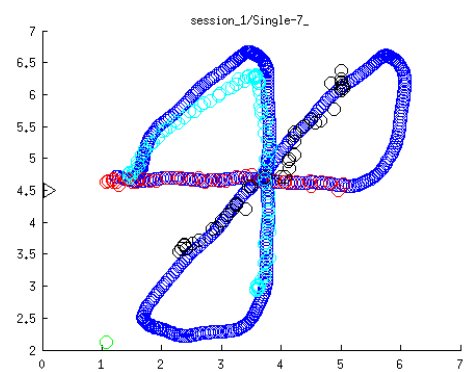
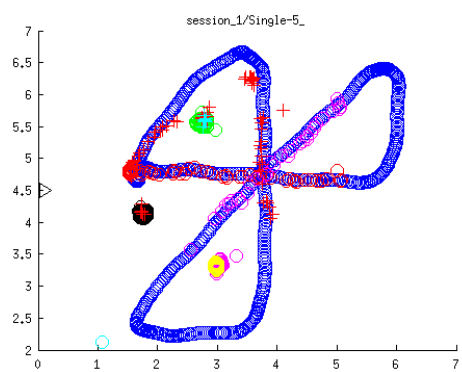
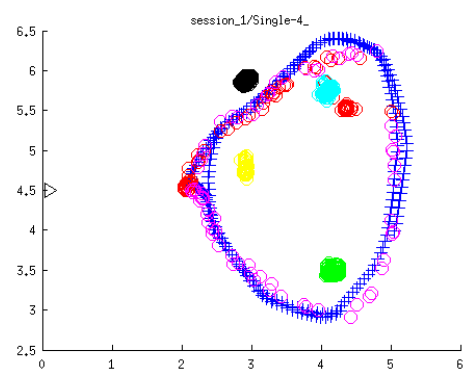
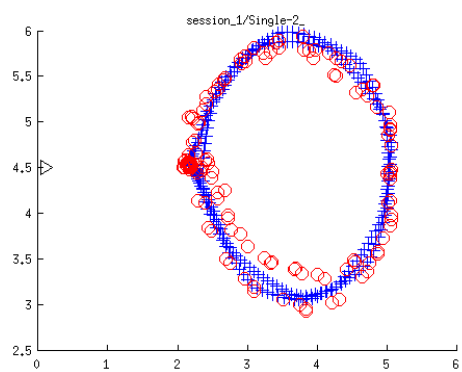


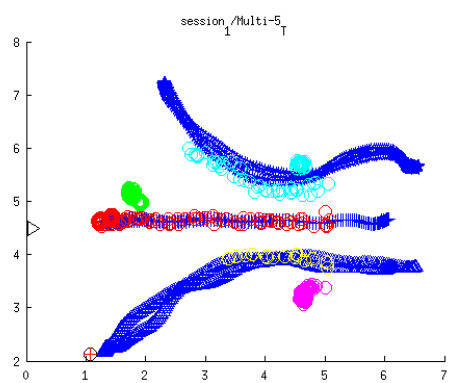
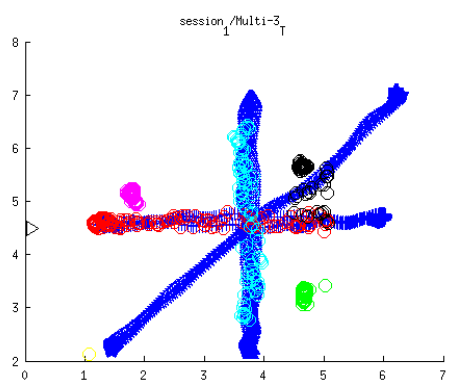
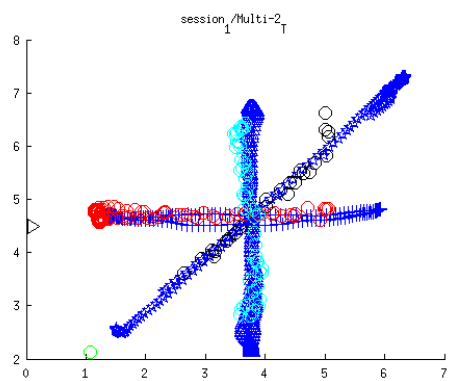
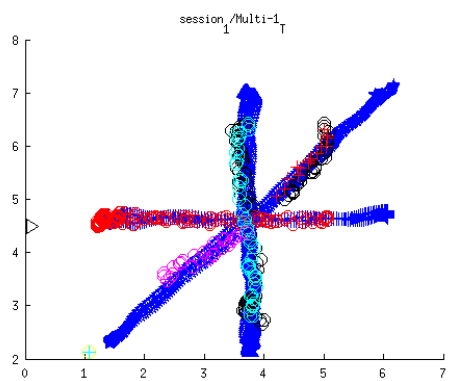


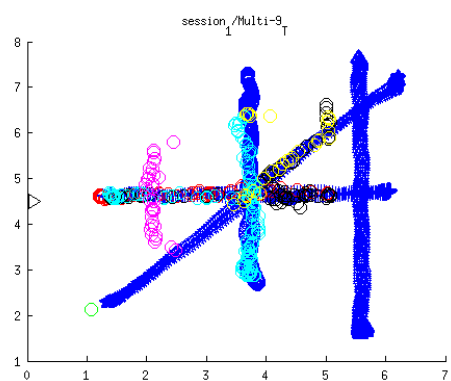
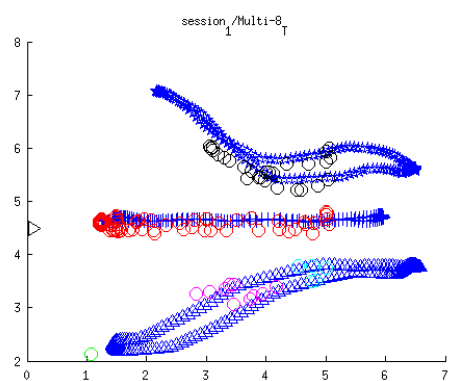
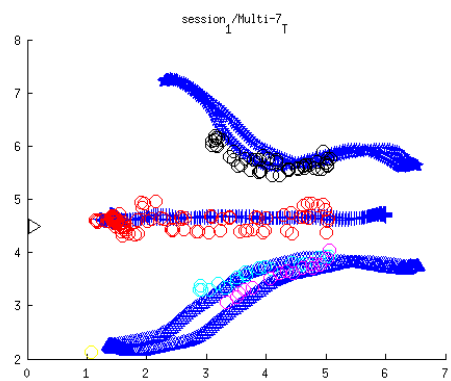
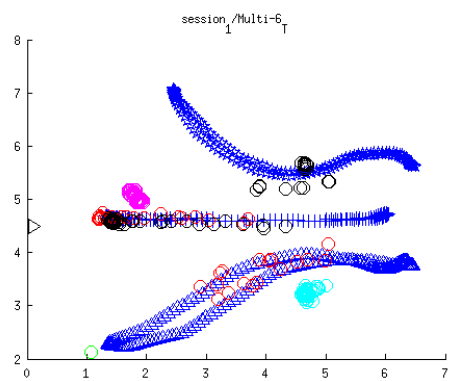
B.5: Shows the results between the GT data and the SUT data for the session 5 experiments. The blue color is the GT data. Other colors are the SUT data (different colors indicate different humans). The axes indicate the spatial positions or paths of the people in meters.

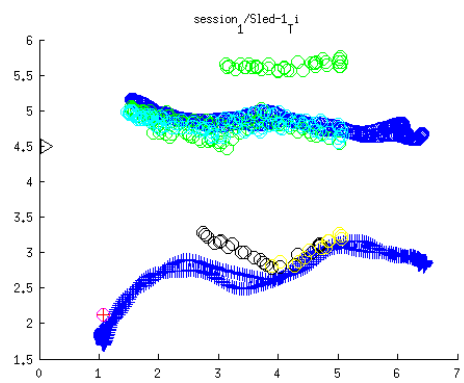
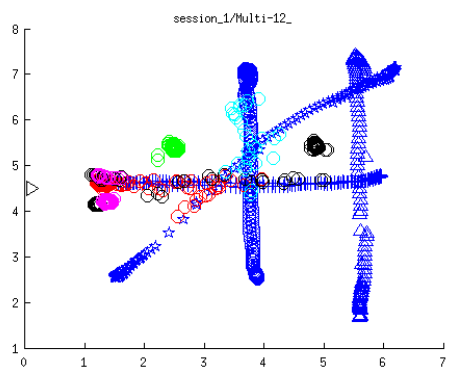
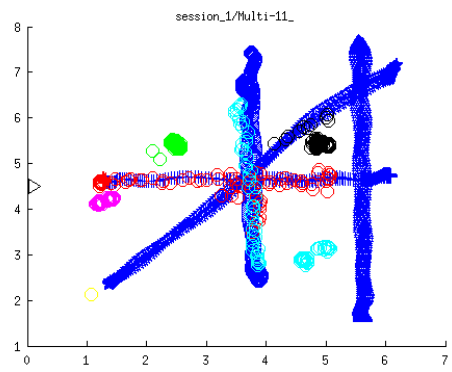
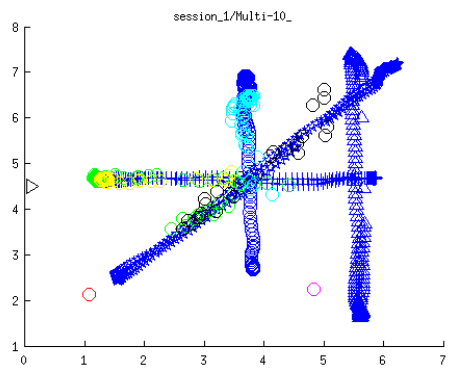












Appendix C – Data Plots for the grid based measurements

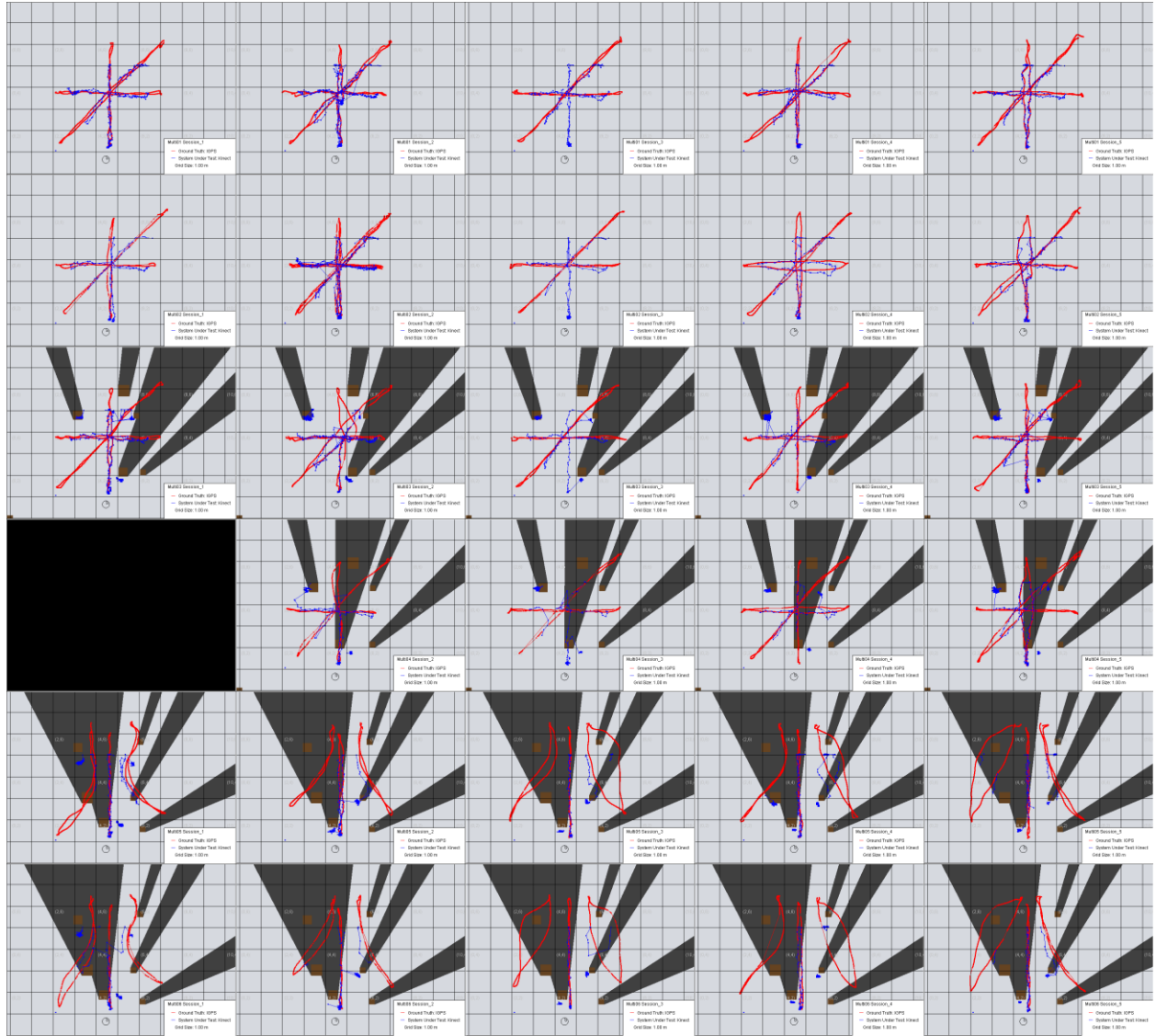


Figure 17. Plots of tests Multi01 to Multi06 (vertical) over Sessions 1 to 5 (Horizontal). Within each subplot tracks of XY positions recorded by Ground-Truth iGPS in red and the System-Under-Test using RGB-D in blue. Static obstacles are brown boxes.

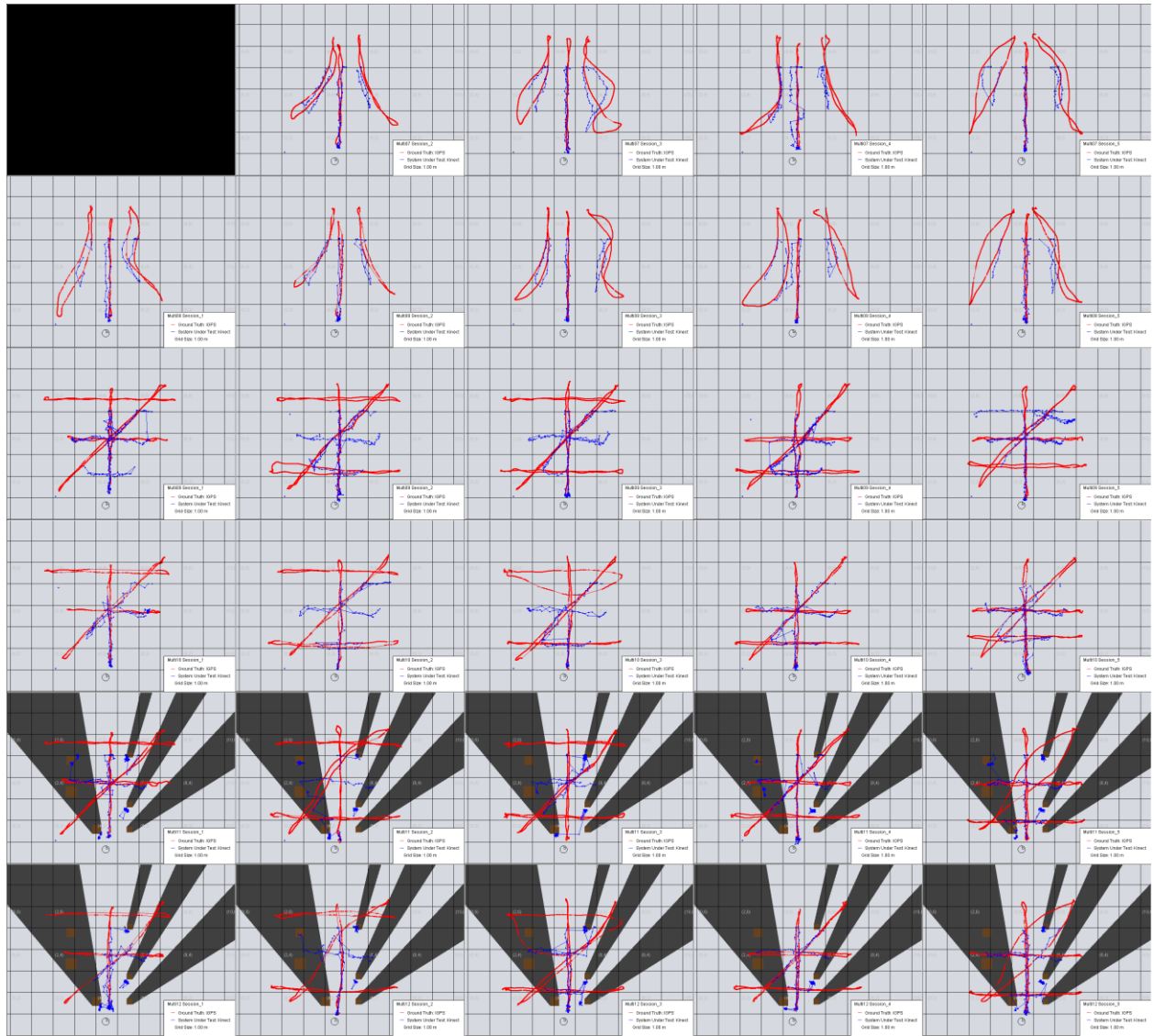


Figure 18. Plots of tests Multi07 to Multi12 (vertical) over Sessions 1 to 5 (Horizontal). Within each subplot tracks of XY positions recorded by Ground-Truth iGPS in red and the System-Under-Test using RGB-D in blue. Static obstacles are brown boxes.

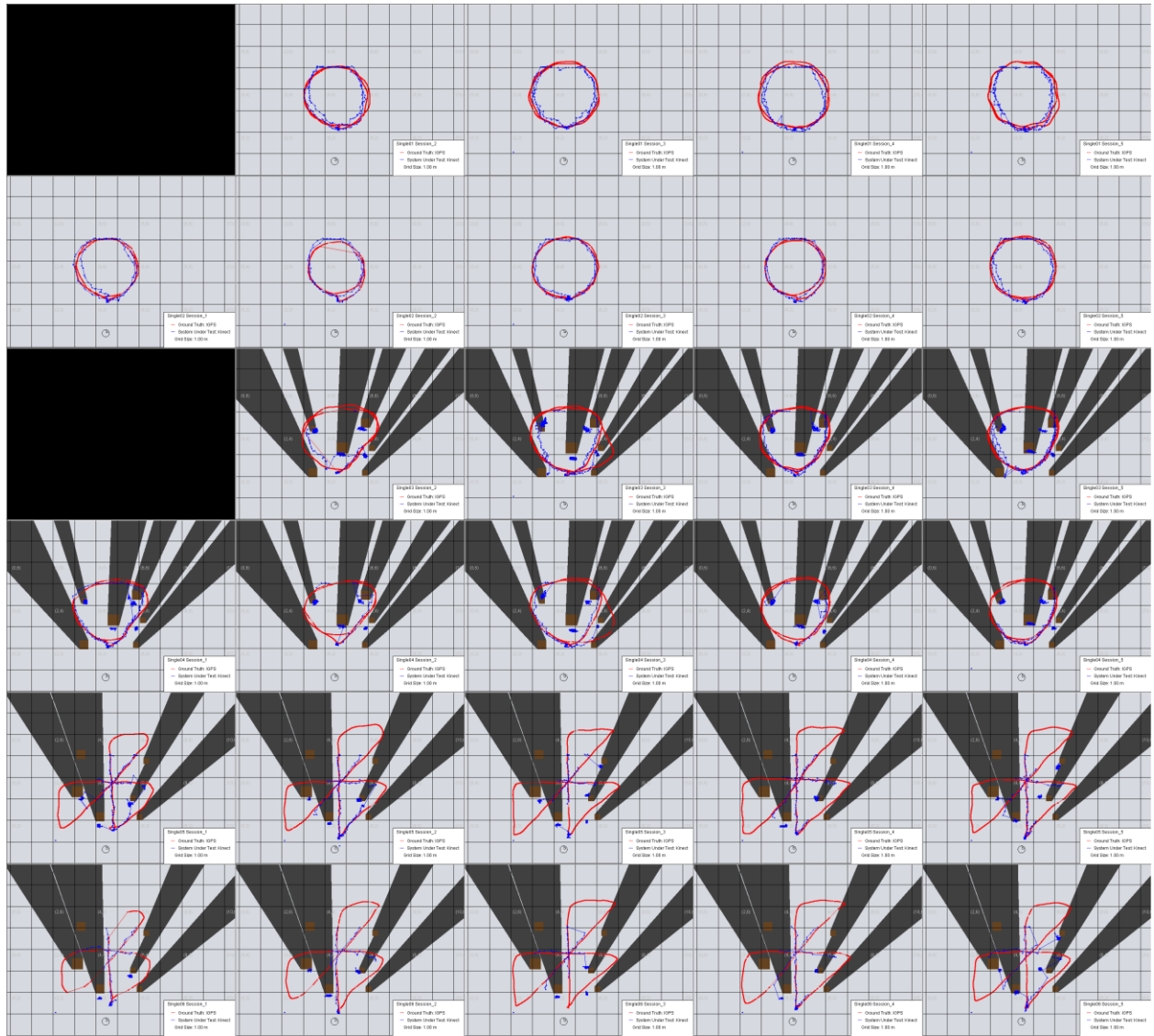


Figure 19. Plots of tests Single01 to Single06 (vertical) overs Sessions 1 to 5 (Horizontal). Within each subplot tracks of XY positions recorded by Ground-Truth iGPS in red and the System-Under-Test using RGB-D in blue. Static obstacles are brown boxes.

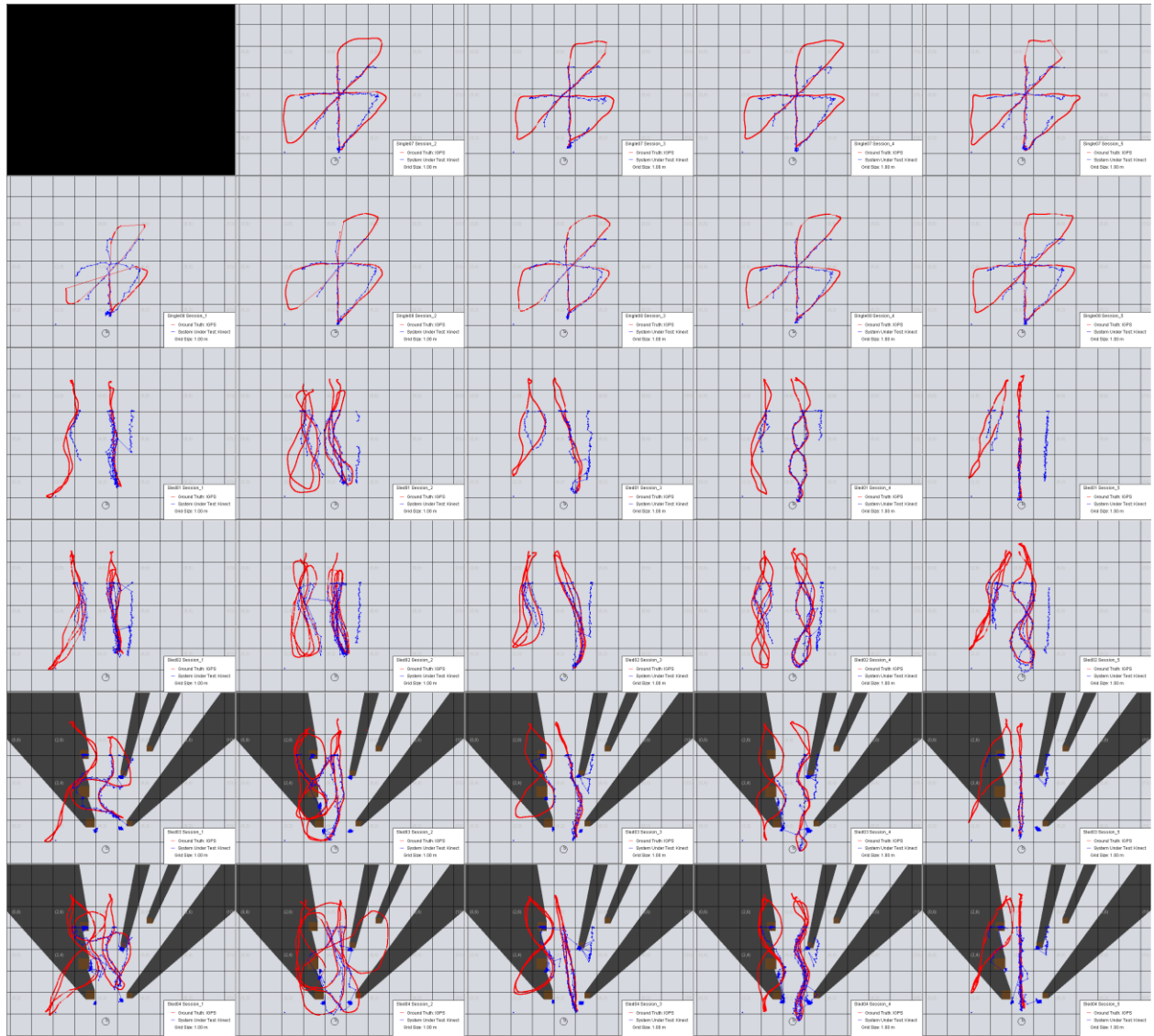


Figure 20. Plots of tests Single07,Single08 + Sled01 to Sled04 (vertical) over Sessions 1 to 5 (Horizontal). Within each subplot tracks of XY positions recorded by Ground-Truth iGPS in red and the System-Under-Test using RGB-D in blue.

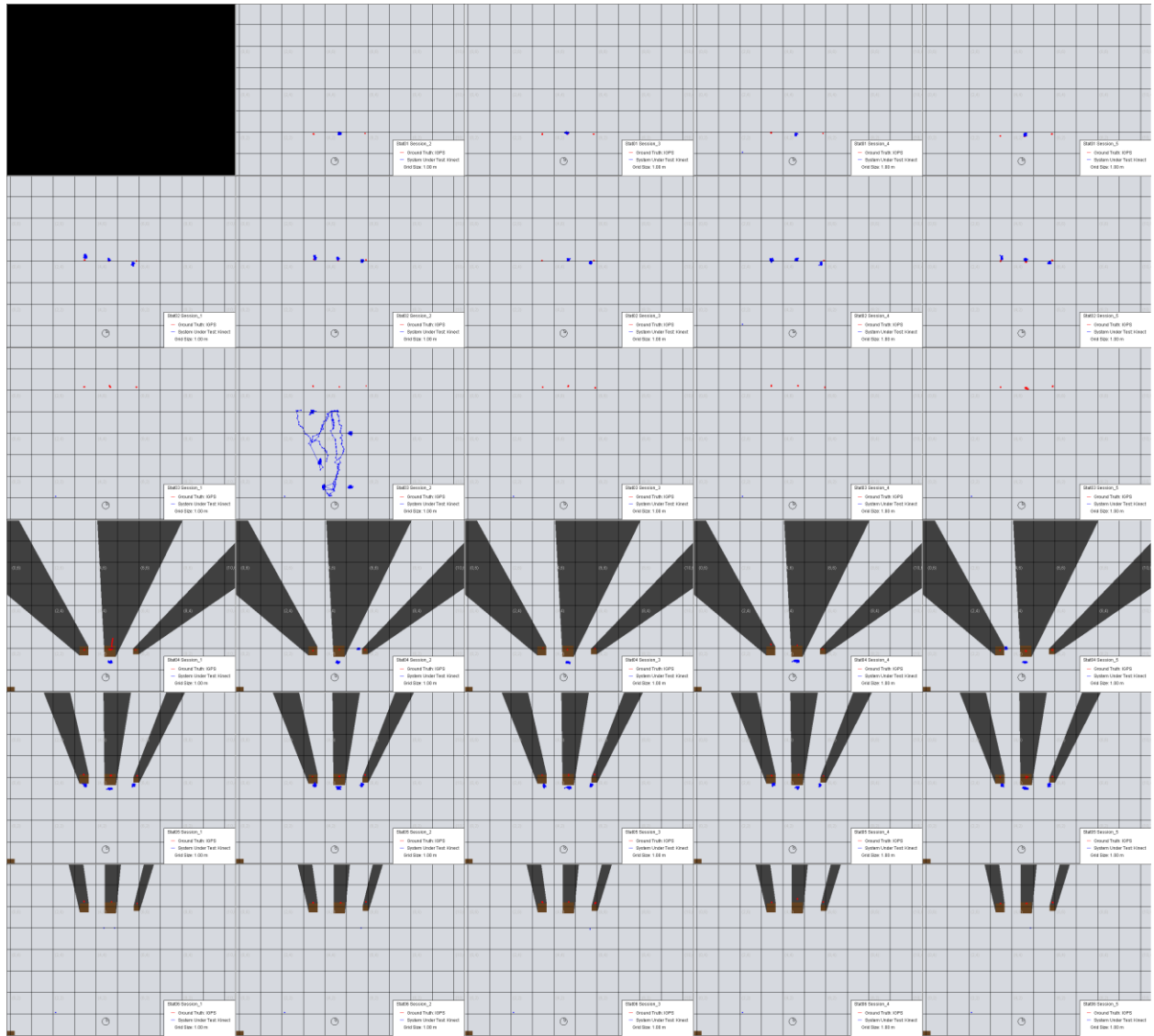


Figure 21. Plots of tests Stat01 to Stat06 (vertical) overs Sessions 1 to 5 (Horizontal). Within each subplot tracks of XY positions recorded by Ground-Truth iGPS in red and the System-Under-Test using RGB-D in blue. Static obstacles are brown boxes.

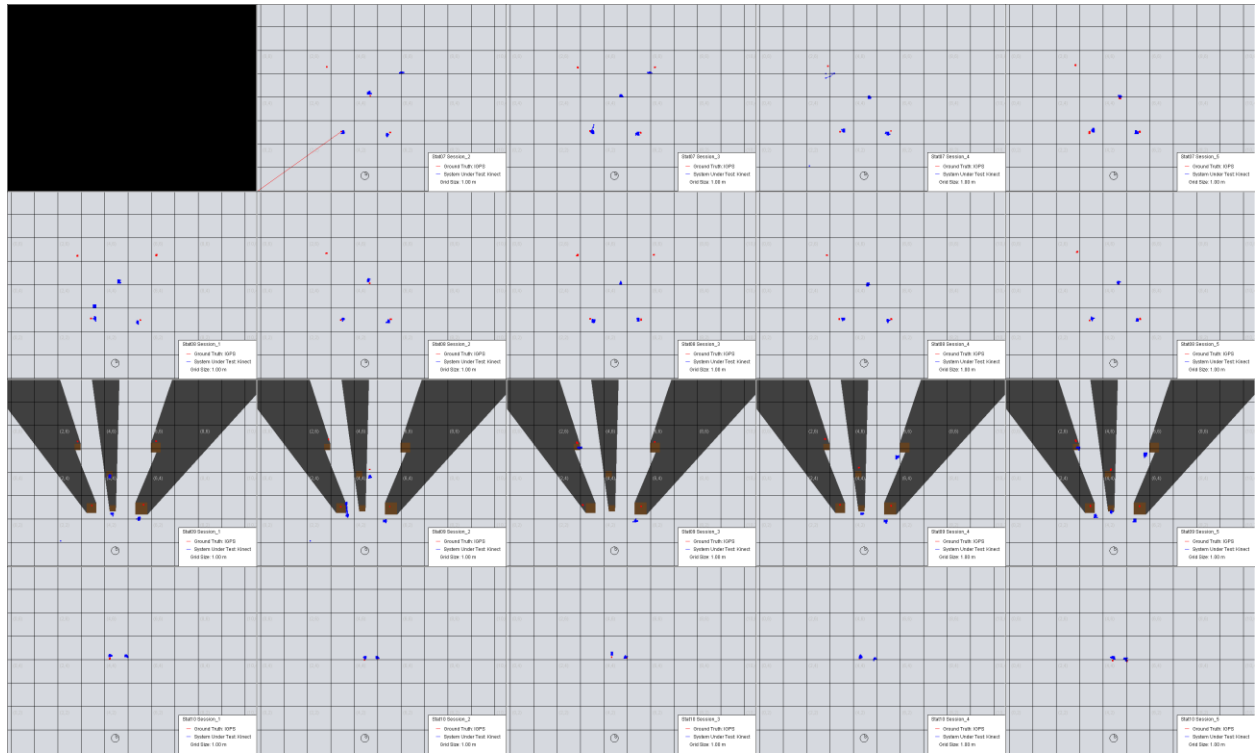


Figure 22. Plots of tests Stat07 to Stat10 (vertical) overs Sessions 1 to 5 (Horizontal). Within each subplot tracks of XY positions recorded by Ground-Truth iGPS in red and the System-Under-Test using RGB-D in blue. Static obstacles are brown boxes.