J Intell Robot Syst
DOI 10.1007/s10846-016-0334-3

Performance Evaluation of Human Detection Systems for Robot Safety

William Shackleford · Geraldine Cheok · Tsai Hong · Kamel Saidi · Michael Shneier

Received: 9 April 2015 / Accepted: 11 January 2016

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 for performing tasks. 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-robot collaboration · Human tracking · Performance evaluation · Performance metrics · Robot safety

M. Shneier (✉) · W. Shackleford · G. Cheok · T. Hong · K. Saidi
Intelligent Systems Division, National Institute of Standards and Technology, Gaithersburg, MD, USA
e-mail: michael.shneier@nist.gov

W. Shackleford
e-mail: william.shackleford@nist.gov

G. Cheok
e-mail: geraldine.cheok@nist.gov

T. Hong
e-mail: tsai.hong@nist.gov

K. Saidi
e-mail: kamel.saidi@nist.gov

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 robot safety standards [1–3] have been modified to specify requirements that 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 different from those of safety. Losing track of a person for a few frames is usually not critical for concern in surveillance applications, whereas losing track in a safety situation is not acceptable. Similarly, reporting

Published online: 22 January 2016

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 paper 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. An independent measurement system is 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 and the ground truth instrument.

Section 2 of this paper refers to previous work on human and object detection and tracking. Some of the performance measures used in this paper 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 the performance measures and their results. Most of the measures are designed specifically for safety applications. The paper ends with a discussion and conclusions. Detailed information about the experiments and results may be found in Shneier et al. [4].

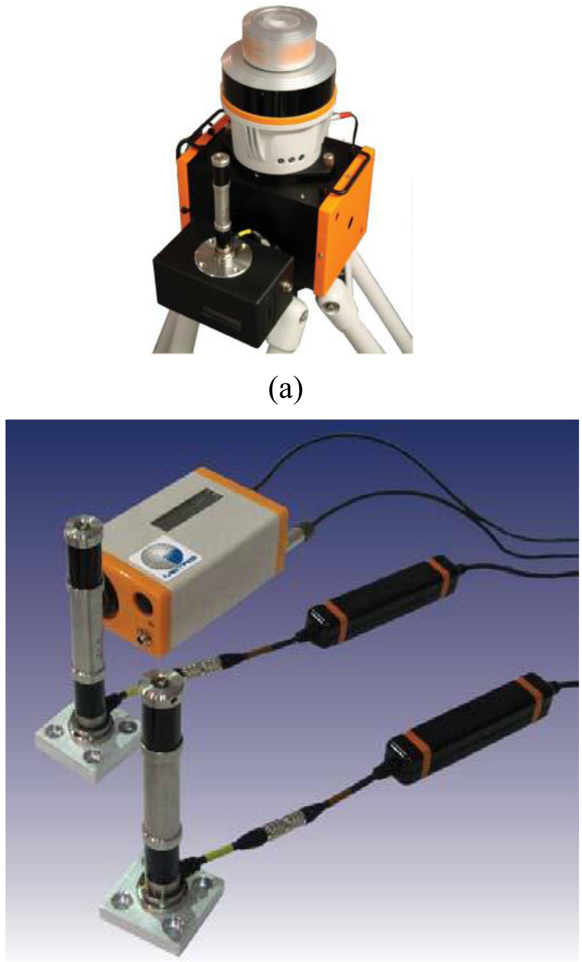
2 Related Work

Detecting and tracking people and objects has a long history, including a number of projects that focused on performance evaluation. Ogale [5] provides a survey of video-based human detection. The PETS (Performance Evaluation of Tracking and Surveillance) workshops, Ferryman, Crowley [6], focused on algorithm development and performance evaluation of tasks such as multiple object detection, event detection, and recognition. Nascimento, Marques [7], 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, Garofolo [8], provides performance evaluation of people, faces, cars, and object tracking and ETISEO, Nghiem et al. [9], 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) [10, 11] 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. [12], 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. [13] 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. [14], 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, Porikli [15], 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. [16], 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 [17–22].

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 Department of Homeland Security, with the British Home Office, and with the European CHIL program (Computers in the Human Interaction Loop), [14], and the CLEAR evaluations, [23]. 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. [24]).

The biggest difference between the current paper 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 here 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.



(a)

(b)

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

3 Measuring Ground Truth

Godil et al. [25], describe four types of ground truth data for object detection and tracking: annotation or 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. This paper adopts the last of these methods, making use of a sensor-based ground truth system, described in the following subsection.

3.1 Indoor Global Positioning System

The Indoor Global Positioning System (iGPS),[1] [26], is a high-resolution measurement system, shown in Fig. 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 (Fig. 1a) and receivers

[1] 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.

(Fig. 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. 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 m^2. Detailed analyses of the system are presented by Schmitt et al. [27], and Mosqueira et al. [28]. In [29], 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 [30], 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 paper, a pair of iGPS vector bars is attached to the top of a hard hat worn by each person, as shown in Fig. 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.

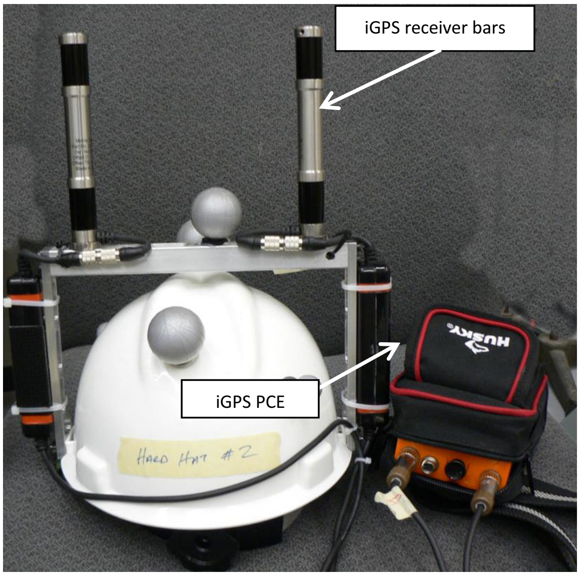


Fig. 2 iGPS vector bars attached to hardhat to be worn by a human. The position calculation engine (PCE) is worn around the waist

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.

Note that in the experiments, two ground truth systems were used; the gray spheres on the hard hat in 2 are markers for a motion capture system. The results presented in this paper 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.

The human being tracked by the iGPS was represented by a point - the center of the person's head. An experiment was conducted to determine the uncertainty of this point as measured by the iGPS. The average standard deviation of the x, y, z coordinates was $\pm$ 20.5 mm.

4 Methodology

A series of experiments was conducted with human subjects to collect data to evaluate the performance metrics. The focus of the metrics is on applications where safety in human-robot interaction is the main concern, although some of the measures have been used in previous, more general studies. 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 performance metrics are then described and their application to the human detection system is evaluated.

4.1 The Human Detection and Tracking Algorithm

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 they can be used in safety applications.

The algorithm that was selected is that of Chambers et al. [31], because it claims a very high accuracy: 99 % recall with less than 10^{-6} false positives per window on the INRIA person dataset (Dalal [32]). The algorithm operates on disparity data acquired from either a stereo camera system or an RGB-D camera (in this case we used an 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. [33]), classified first using adaBoost (Freund, Schapire [34]), then a support-vector machine (SVM) (Cortes, Vapnik [35]), 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, Triggs [36]), 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 sensors.[2] The people wore hard hats instrumented with 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 corresponding readings 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 3 shows a view of an experiment in progress. Note the hard hat with reflectors for ground truth measurements.

Due to the limited field of view of the sensors being tested, in particular the ASUS Xtion sensor, the active area used for testing was small, only 5 m by 5 m, with a 1 m buffer area surrounding it. One side of the floor was labeled with letters and the other with numbers. This meant that a person could be asked to stand in a

[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.

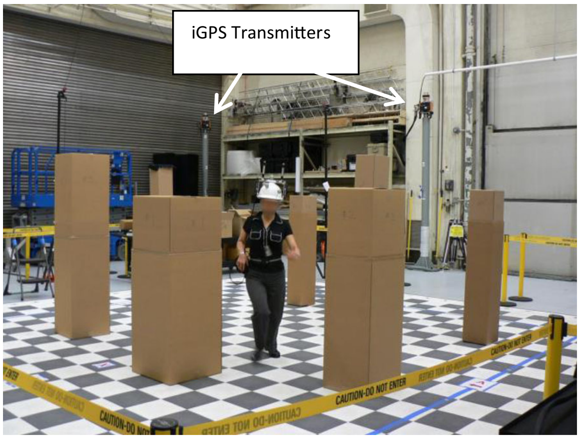


Fig. 3 A view of the test area showing an experiment in progress with obstacles in place

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, each of which was repeated five times with different participants. More details of the 34 different experiments may be found in Shneier et al. [4].[3] A total of fifteen subjects participated. The experiments were designed to include factors such as distance from the system under test, occlusion, clutter, speed of motion, different types of trajectories, and different human shapes and features. To give a flavor of the experiments, two are briefly described below. The data 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 the ground truth system and for the sensor being evaluated. Each sensor collected data at its own rate, but the time stamp made it possible to align the data for analysis.

4.3 Representative Experiments

In the experiments with moving people, the same trajectory was covered for four different cases: walking without obstacles, jogging without obstacles, walking with obstacles, and jogging with obstacles. Each person was given a trajectory at the start of the experiment and told to walk or jog at a comfortable speed (Fig. 4). Data were collected simultaneously from the ground truth and system under test sensors for the duration of the experiment.

To determine how well the algorithm can keep track of many people, a number of tests were conducted with up to five subjects (Fig. 5). In the limited space available because of the narrow field of view of the system being evaluated, there is a lot of crowding and, with obstacles in place, there is also 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.

5 Performance Measures

This section presents a new approach for measuring the performance of human detection systems developed specifically for applications in which safety is critical. We also describe some more traditional metrics for detection and tracking that are used to complement the new approach. Before describing the performance measures, we discuss the data obtained from the tests and registration of data from the iGPS ground truth (GT) sensors and the system under test (SUT).

5.1 Input Data

There are three types of input files:

GT Data Files There is a separate file for each sensor and for each test (i.e., two sensors * 100 tests = 200 files). Each row contains a timestamp, ID, X, Y, Z, where the timestamp represents the time in seconds since Jan 1, 1970 UTC, ID is a unique string corresponding to a person or physical target, and X, Y, Z are the Cartesian coordinates of the target. Each row may contain an optional radius which describes the distance from the center point of a detected person sufficient to contain the entire person, measured on the ground (XY plane).

[3] Available at http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8045.pdf.

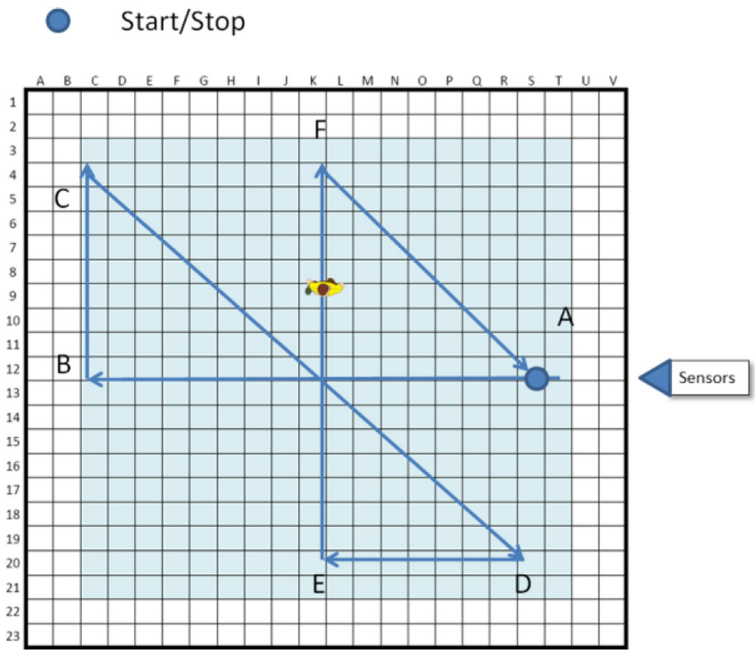


Fig. 4 A single person walks along a pre-defined path starting at A

1.) SUT Data Files. There are separate files for each test. Each row contains a timestamp, ID, X, Y, Z, as above. Each row may contain optional fields used by the analysis system if present, including velocities in X, Y, and Z directions, radius, and confidence.

2.) Position and Orientation Transform Files. A single transform file is used for each sensor for

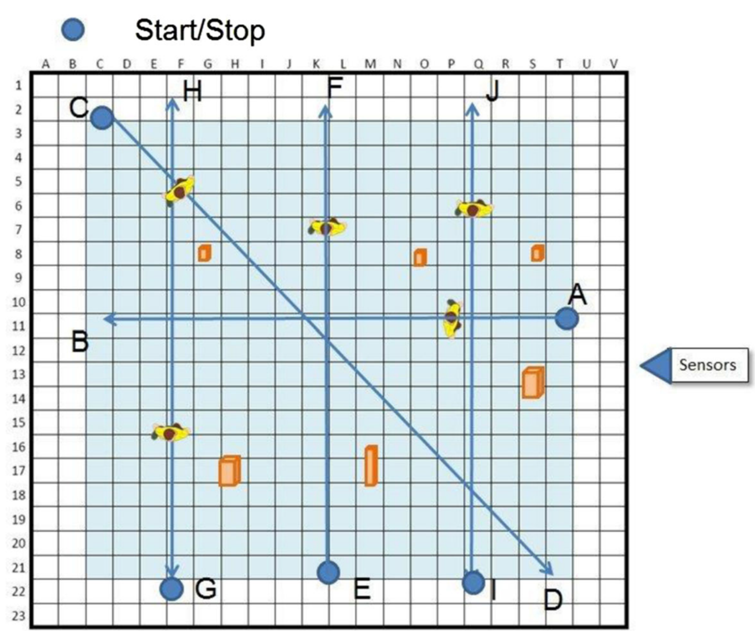


Fig. 5 Five people moving through the space simultaneously with obstacles. The *double arrows* indicate that the person will move along the path in both directions

all tests because the sensors were not moved between tests. The transform is a 4×4 homogenous matrix that encodes the registration between the SUT sensor and the GT obtained from a calibration and registration procedure.

5.2 Data Interpolation

The data were converted to units of seconds and meters and synchronized using the Network Time Protocol and a trigger signal. Data were interpolated so that the GT system provided the best estimate of the position of each person at each time given all data, including data collected after that point in time. For the SUT data, interpolation used only data produced prior to each sample time. The evaluation software limits the analysis to the common period of time when both GT and SUT systems were running.

5.3 Performance Measures for Safe Human-Robot Interaction

A new set of performance measures was defined with the following goals:

1) 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 5.6.
2) Be able to compare systems as configured by the vendor with safety guarantees met.
3) Provide fair comparisons between systems with different frame rates, detection mechanisms, and different numbers of outputs
4) Make it easy to combine multiple detection systems.
5) 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 is available from Shackleford [37]. Full source code is provided 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:

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

5.4 Detection Area vs Safety Coverage Area

The SUT is only expected to report the positions of people within its field of view and between its minimum and maximum operating 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 [38]; 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.

Ground-Truth Radius Estimate The GT system provides 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 can 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.

5.5 False Clear Area/ False Occupied Area

Given the above assumptions, two primary metrics, False Clear area and False Occupied area, 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 GT location of a person (see Fig. 6). The sizes of the circles are as described in Section 5.4.
4) Paint black circles surrounding each location where a person was found by the SUT.
5) Count the number of pixels that are white following steps 3 and 4 (see Fig. 7) to give the false clear area.

   To compute the false occupied area:
6) Create an image covering the area of interest at the desired resolution.

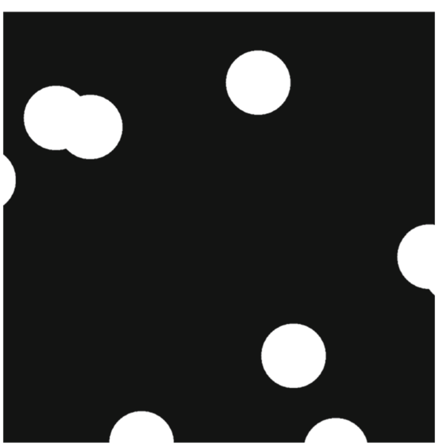


Fig. 6 Intermediate step in computing the False Clear Area. After Step 3, *white circles* indicate where GT located people

Fig. 7 The final false clear area, after Step 5

7) Paint the entire image black.
8) Paint white circles surrounding each location where the SUT detected a person.
9) Paint black circles surrounding the locations where the GT detected people.
10) The number of pixels that are white following this procedure gives the area that is falsely determined to be occupied (see Fig. 12).

Note: It would be incorrect to assume that false clear or false occupied regions smaller than the size that a person could occupy can or should 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.

5.6 Occlusions

Some regions of the coverage area will be occluded from the sensor by dynamic or static obstacles. A planner that produces a safe path 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 by the SUT. Painting these polygons requires that the evaluation system know the location of the SUT. 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

Fig. 8 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

as if it is known to contain a person. This results in an estimate of the occupied areas that appears something like Fig. 8.

5.7 Time Projection for Assumed Reaction Time

No system can react instantaneously. The planning and control system takes time to respond to new detections or changes in the estimated position of a detected person. What matters is a guarantee of where people cannot be over some time interval into the future. To account for this, false clear and false occupied regions are computed as in Section 5.6 except now from each point in time we integrate over a period into the future. This means that the SUT must not only estimate the current position of each person, but also predict potential future positions measured from that point until the reaction time has expired (Fig. 9).

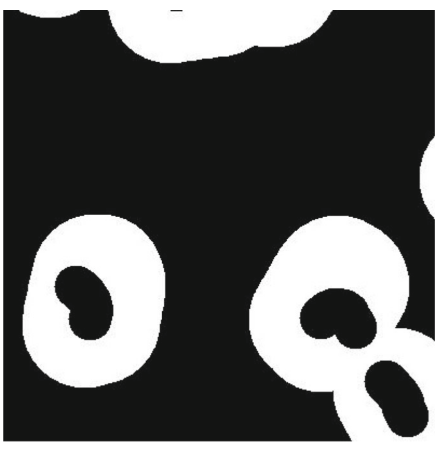


Fig. 9 Display of False Occupied regions as computed by integrating over the reaction time. SUT Occupied area is *white stretched circle*. GT Occupied area is *black region*

The simplest way of accomplishing this prediction is to increase the radius around each detected person until it contains the entire area through which the person can travel during the 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 small amount. The evaluation software is currently limited to projecting along a single vector at constant speed since this is the only information the SUT could provide. To test systems capable of more sophisticated predictions, the evaluation software would have to be updated.

5.8 Simulated Examples

In the first simulated example we take only a single frame of data and both the GT system and the SUT return a single position for the only person in the scene (Fig. 10). We assume that occlusions are not an issue for this system and either that the person is stationary or the robot reacts fast enough that time projection is not necessary. We compute the false clear area as in Section 5.5, giving the result in Fig. 11. In this example, the SUT reported a position that was somewhat to the left of the true position. This, combined with the fact that the system reported only one position and reported only the same minimum radius the GT was set to, means that the person is in danger of being hit by the robot on the right side. For the false clear image, the areas in black are areas where no robot-to-human

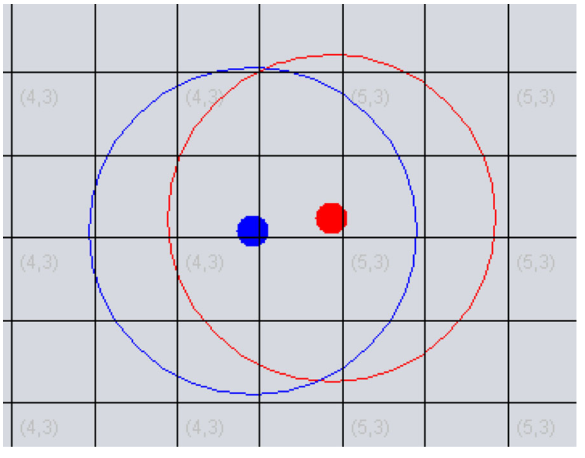


Fig. 10 Simulated example 1. Ground truth position in red surrounded by ground truth radius in red. System under test position in blue surrounded by system under test reported radius

Fig. 11 False clear area for simple example 1

collision can occur either because there is no human there or because the robot has been told not to enter that area. The areas in white are regions where a collision can occur because it is close enough to the GT position that a portion of the person occupies that area and the robot has no reason to avoid the area. Given that the false clear area is not zero, the conclusion is that this sensor is not safe to use and no further analysis is needed, although it may only require a simple fix. Later examples suggest some possible fixes, but one should be careful about allowing system developers to make a large number of changes to work around each test failure, since the test results may not generalize.

The false occupied area is shown in Fig. 12. Here, the SUT reported a position that was somewhat to the left of the true position. This means there is an area to the left of the person that the robot will be forced to avoid unnecessarily and this will reduce productivity. The black areas represent regions of maximum productivity that either the robot is avoiding because it truly needs to or it is able to traverse. The white areas represent areas that the robot cannot use but which would be usable if we had chosen a better sensor. There are 95 white pixels with an area of 4 square centimeters each. This means the false occupied area is 380 square centimeters and the ratio of "Average False

Fig. 12 False occupied region for simulated example 1

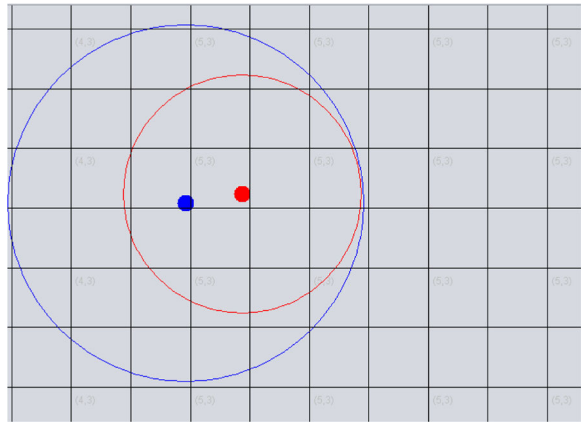


Fig. 13 Simulated example 2. Ground truth position in red surrounded by ground truth radius in red. System under test position in blue surrounded by system under test reported radius

Occupied Area to Safety Coverage Area" is 0.01 since there were 9500 pixels in the entire safety coverage area (Fig. 13).

The second example is the same as the first except the system reports a larger radius for the person.

In this example, computing the false clear area results in a completely black field, indicating that the SUT reported a position somewhat to the left of the true position but, since it increased the area around that position sufficiently, there is no position that the robot will go where it may collide with a person.

Figure 14 shows the false occupied area, which reflects that the SUT reported a position that was somewhat to the left of the true position. This means there is an area to the left of the person that the robot will be forced to avoid unnecessarily and this will reduce productivity. The increased radius which makes the system safe also increases this area. The false occupied area is 1560 square centimeters and the ratio of "Average False Occupied Area to Safety Coverage Area" is 0.041 since there were 9500 pixels in the entire safety coverage area (Figs. 15 and 16).

Fig. 14 False occupied region for example 2

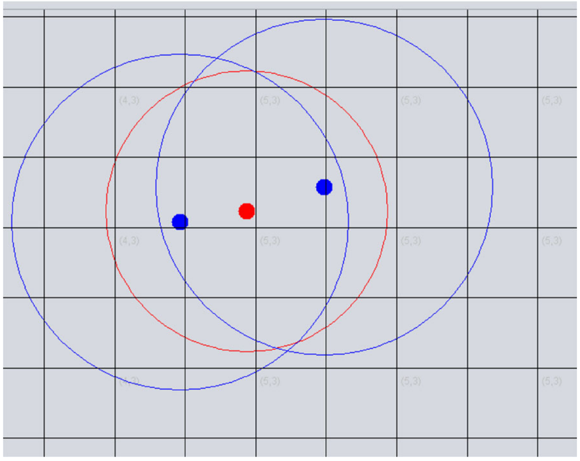


Fig. 15 Simulated example 3. Ground truth position in red surrounded by ground truth radius in red. System under test positions in blue surrounded by system under test reported radius

In the third example, the GT reported a single person at a single position while the SUT reported two positions. The radii reported by the SUT are larger than in example 1 but smaller than in example 2. One was somewhat to the left of the true position and one to the right. This means there is an area on both sides of the person that the robot will be forced to avoid unnecessarily and this will reduce productivity. The increased radius which makes the system safe also increases this area. The false occupied area is 1444 square centimeters and the ratio of "Average False Occupied Area to Safety Coverage Area" is 0.038.

The conclusion from this set of examples (Table 1) is that the system from example 1 is not suitable by itself as a safety sensor. If it were the only choice available, conventional fences or manual machines should be used instead. No more tests are needed to determine whether example 1 is safe to use since it has already failed. Examples 2 and 3 appear to be safe in this limited test, although more testing should be done in the

Fig. 16 False Occupied region for example 3

Table 1 Safety of system under test in the three example scenarios

System	Safe ?	Average False Occupied Area / Safety Coverage Area
Example 1	NO	0.010
Example 2	Yes (but further testing is required)	0.041
Example 3	Yes (but further testing is required)	0.038

target environment. Between example 2 and example 3, example 3 allows the robot to traverse more space which should lead to higher productivity.

Note that the metrics are not simply related to a minimum safety radius. The minimum safety radius cannot be computed from the localization error. If the person is entirely within the safety coverage area and there is only one SUT point, then the minimum radius equals the distance between the two points plus the distance of the furthest point from the detection to any portion of the person. However, with more than one point the distance need not reach the furthest point on the person, but only the furthest point not covered by another detection point. Some portion of the person could be within a known area of occlusion that the robot would have to avoid anyway. This would change the position on the person that need be reached by the minimum safety radius. Some portion of the person might also be outside the coverage area and there would be no need for the minimum safety radius to extend to cover this portion of the person.

With increasing distance it becomes less likely that the closest point reported by the SUT corresponds to the same person as that from the GT. In this case, it is better to mark the area as false clear. The exact distance at which this should be set is sensor dependent. The sensor vendor should commit to a radius before tests are performed and the tests should determine if the results are valid.

The false occupied area also cannot be computed from localization errors with or without a known minimum safe radius. The SUT could report a position during a period of time when no GT positions reported. There would therefore be no distances to measure during this period of time and therefore no effect on the localization error although there would be a large increase in the false occupied area. Even when there are GT positions in the scene at the same time, the correspondence function does not guarantee that all SUT positions will have a corresponding point with which to compute a distance. The correspondence function also does not guarantee that when a distance is computed that there is no closer GT position that could have produced a smaller distance. Even in the absence of GT positions, two sets of SUT positions of the same size can produce different amounts of false occupied area. In one set the points could largely overlap producing a false occupied area only slightly larger than the area of a circle for one person while in the other set they could be disjoint, producing an area larger by a factor of the number of points reported. Reported positions could also be close to the edge of the safety coverage area, thus reducing the increase in the false occupied area. Sensors that are susceptible to occlusion will necessarily have to mark large areas of the area as occupied and thus have a larger false occupied area than sensors not susceptible to occlusion even when reporting the same set of positions with the same minimum safe radius.

5.9 Conventional Performance Measures

In addition to the new metrics, several others that have been used for human detection were also applied. Also, since the new approach does not include tracking, some previously-used tracking metrics were used. These measures allow us to compare human detection algorithms to others in the literature and to provide an indication of how well the algorithms maintain the identities of people as they move about in the environment. They do not, however, provide a measure of the safety of the systems. The human detection metrics that we used include:

1) False Positive (FP): A human is present in the SUT data but not in the GT data.
2) False Negative (FN): A human is present in the GT data, but not in the SUT data.
3) True Positive (TP): A human is present in the GT data and the SUT data.
4) 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.
5) 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.
6) 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.
7) 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
8) 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.

Unlike the metrics for safety described above, conventional metrics provide no clear indication that a system is or is not safe. They also require a correlation between GT data and SUT data. For example it may be necessary to know that the third position reported by the SUT system corresponds to the second position reported by the GT system. While correspondences may seem obvious in some cases, there is no clear general rule for establishing them. The safety metrics proposed above require no assumptions about correspondences. It should also be noted that the conventional metrics focus on average or typical behavior while the safety of the system primarily depends on atypical or extreme cases. Furthermore since they deal only with discrete people they cannot include an effect for a person only partially within the coverage area or only partially within an area of known occlusion.

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. 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. Our matching method uses centroids of regions labeled as people and 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 also used (Bodt, Hong [39]).

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. The human tracking metrics are:

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

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

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).

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

$$HTA = 1 - \frac{\sum_t (\mathrm{FN}_t + \mathrm{FP}_t)}{\sum_t GT_t}$$

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.

6 Performance Analysis

For reasons of space, we present a performance analysis only for the new metrics introduced in this paper. The results for the conventional measures can be found in Shneier et al. [4].

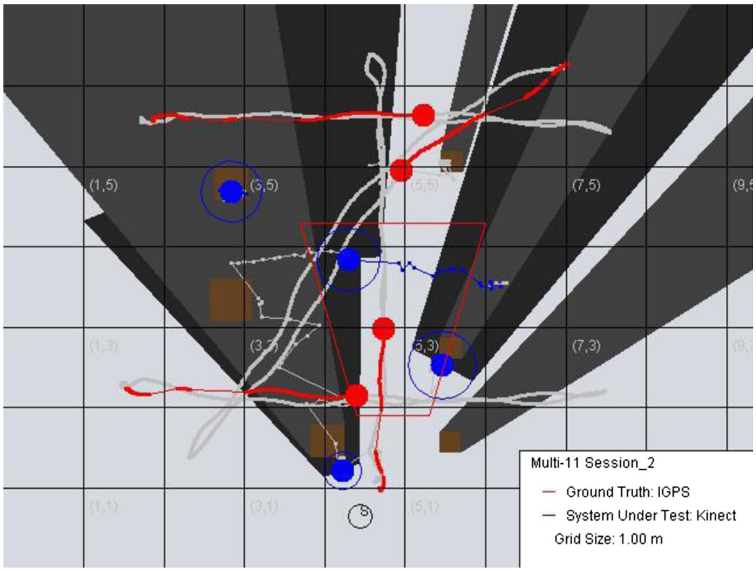


Fig. 17 The tracks when the worst false clear area was detected. Sensor is located where the S in a circle is marked. None of the four ground-truth tracks (*red*) are matched by the tracks from the system under test (*blue*). The plot shows a single snapshot in time until the end of the test

Table 2 Performance statistics of detection in uncluttered environments using 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[a]	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

[a]Many static tests passed for the trivial reason that there were never any people in the safety coverage area

The systems tested did not achieve 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 17 shows an example.

We provide more detailed data below as an example of what one might obtain, keeping 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 insights are available about how the human detection system would impact productivity?

If the answer for the first question for this SUT is no, there is no reason to even ask the second question. However when a system passes the first hurdle, the second question becomes relevant.

Tables 2 and 3 report the data in cluttered and uncluttered environments. If the system can be used safely, then the column labeled "Tests with Maximum False Clear Area > 0 / Total Tests in category" will have 0 in the numerator of every cell. This column is designed to help answer the first question. The column

Table 3 Performance statistics of detection in cluttered 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	0/20[a]	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

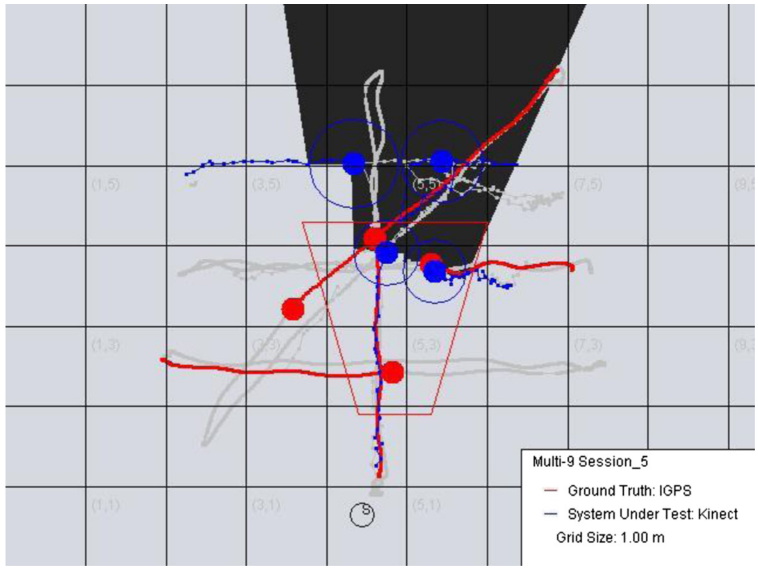


Fig. 18 The tracks when the worst false clear area was detected without static obstacles. The sensor is located where the S in a circle is marked. None of the four ground-truth tracks (*red*) are matched by the tracks from the system under test (*blue*)

labeled "Average False Occupied Area / Safety Coverage Area" then gives a measure of the area a robot would have to avoid unnecessarily whichis related to loss of productivity. This column is designed to help answer the second question and is the column where one would look to compare two systems that were both safe to use and choose the one likely to provide the best productivity. A system with the smallest false occupied area would give the highest productivity (Fig. 18).

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, while undesirable, 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, traditional performance metrics, while useful for comparing different systems, are not by themselves sufficient to guarantee safety. The additional considerations discussed in this paper 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 ground truth system used in this work is much less susceptible to occlusions than the system under test. This allows the system to evaluate the effect of varying amounts of occlusion, which would not be possible with some of the more commonly used ground-truth methods, especially those that rely on manually annotating the imagery.

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 such as those in this paper, 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.

Acknowledgments 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. In. (2011)
2. ISO TC 184: ISO 10218-1:2011 Robots and robotic devices – Safety requirements for industrial robots – Part 1: Robots. In. (2011)
3. American National Standards Institute, Robotics Industries Association: ANSI/RIA R15.06-2012, Industrial Robots and Robot Systems - Safety Requirements. In. (2012)
4. Shneier, M., Hong, T., Cheok, G., Saidi, K., Shackleford, W.: Performance evaluation methods for human detection and tracking systems for robotic applications. In., vol. NISTIR 8045. National Institute of Standards and Technology, Gaithersburg (2015)
5. Ogale, N.A.: A survey of techniques for human detection from video. University of Maryland (2006)
6. Ferryman, J., Crowley, J.L.: IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. In: Ferryman, J., Crowley, J.L. (eds.) (2013)
7. Nascimento, J.C., Marques, J.S.: Performance evaluation of object detection algorithms for video surveillance. IEEE Trans. Multimedia **8**(4), 761–774 (2006). doi:10.1109/TMM.2006.876287
8. Stiefelhagen, R., Garofolo, J.: Multimodal Technologies for Perception of Humans, First International Evaluation Workshop on Classification of Events, Activities and Relationships, CLEAR 2006, Lecture Notes in Computer Science, vol. 4122. Springer, Southampton (2007)
9. Nghiem, A.T., Bremond, F., Thonnat, M., Valentin, V.: ETISEO, performance evaluation for video surveillance systems. In: IEEE Conference on Advanced Video and Signal Based Surveillance, 2007. AVSS 2007, pp. 476–481 (2007)
10. Home Office Centre for Applied Science and Technology: Imagery Library for Intelligent Detection Systems (i-LIDS): The i-LIDS User Guide. In. (2011)
11. British Home Office: Imagery Library for Intelligent Detection Systems. https://www.gov.uk/imagery-library-for-intelligent-detection-systems (2013). Accessed 30 May 2014
12. Brown, L.M., Senior, A.W., Tian, Y., Connell, J., Hampapur, A., Shu, C., Merkl, H., Lu, M.: Performance evaluation of surveillance systems under varying conditions. Paper presented at the IEEE international workshop on performance evaluation of tracking and surveillance, Colorado
13. Yin, F., Makris, D., Velastin, S.: Performance evaluation of object tracking algorithms. Paper presented at the 10th IEEE international workshop on performance evaluation of tracking and surveillance (PETS 2007), Rio de Janeiro
14. Lazarevic-McManus, N., Renno, J.R., Makris, D., Jones, G.A.: An object-based comparative methodology for motion detection based on the F-Measure. Comput. Vis. Image Underst. **111**(1), 74–85 (2008). doi:10.1016/j.cviu.2007.07.007
15. Bashir, F., Porikli, F.: Performance evaluation of object detection and tracking systems. Paper presented at the 9th

IEEE international workshop on performance evaluation of tracking and surveillance (PETS 2006), New York
16. Black, J., Ellis, T., Rosin, P.: A novel method for video tracking performance evaluation. Paper presented at the joint IEEE international workshop on visual surveillance and performance evaluation of tracking and surveillance, Nice, 2003
17. Davis, J., Goadrich, M.: The relationship between Precision-Recall and ROC curves. Paper presented at the Proceedings of the 23rd international conference on machine learning, Pittsburgh, 2006
18. Kalal, Z., Matas, J., Mikolajczyk, K.: Online learning of robust object detectors during unstable tracking. In: 2009 IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops), pp. 1417–1424 (2009)
19. Popoola, J., Amer, A.: Performance evaluation for tracking algorithms using object labels. In: IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008, pp. 733–736 (2008)
20. Bernardin, K., Elbs, A., Stiefelhagen, R.: Multiple object tracking performance metrics and evaluation in a smart room environment. Paper presented at the The 6th IEEE international workshop on visual surveillance, VS 2006, Graz
21. Moeslund, T.B., Hilton, A., Krüger, V.: A survey of advances in vision-based human motion capture and analysis. Comput. Vis. Image Underst. **104**(2), 90–126 (2006). doi:10.1016/j.cviu.2006.08.002
22. Baumann, A., Bolz, M., Ebling, J., Koenig, M., Loos, H.S., Merkel, M., Niem, W., KarlWarzelhan, J., Yu, J.: A review and comparison of measures for automatic video surveillance systems. EURASIP Journal on Image and Video Processing 2008 (2008). doi:10.1155/2008/824726
23. Stiefelhagen, R., Bowers, R., Fiscus, J.G.: Multimodal Technologies for Perception of Humans, International Evaluation Workshops CLEAR 2007 and RT 2007, Baltimore, MD, USA, May 8–11, 2007, Revised Selected Papers. Lecture Notes in Computer Science. Springer, Berlin Heidelberg (2008)
24. Bodt, B., Camden, R., Scott, H., Jacoff, A., Hong, T., Chang, T., Norcross, R., Downs, T., Virts, A.: Performance measurements for evaluating static and dynamic multiple human detection and tracking systems in unstructured environments. Paper presented at the Proceedings of the 9th workshop on performance metrics for intelligent systems, Gaithersburg, Maryland, 2009
25. Godil, A., Bostelman, R., Saidi, K., Shackleford, W., Cheok, G., Shneier, M., Hong, T.: 3D ground-truth systems for object/human recognition and tracking. In: 2013 IEEE Conference on 2013 Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 719–726. IEEE
26. Nikon Corporation: iGPS/iSpace. http://www.nikon.com/products/instruments/lineup/industrial/metrology_assisted_production/large_volume/igps/index.htm. Accessed 2 June 2014
27. Schmitt, R., Nisch, S., Schönberg, A., Demeester, F., Renders, S.: Performance evaluation of iGPS for industrial applications. In: 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–8 (2010)
28. Mosqueira, G., Apetz, J., Santos, K.M., Villani, E., Suterio, R., Trabasso, L.G.: Analysis of the indoor GPS system as feedback for the robotic alignment of fuselages using laser radar measurements as comparison. Robot. Comput.-Integr. Manuf. **28**(6), 700–709 (2012). doi:10.1016/j.rcim.2012.03.004
29. Wang, Z., Mastrogiacomo, L., Franceschini, F., Maropoulos, P.G.: Experimental comparison of dynamic tracking performance of iGPS and laser tracker. Int. J. Adv. Manuf. Technol. **56**(1–4), 205–213 (2011). doi:10.1007/s00170-011-3166-0
30. Depenthal, C.: Path tracking with IGPS. In: International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–6 (2010)
31. Chambers, D.R., Flannigan, C., Wheeler, B.: High-accuracy real-time pedestrian detection system using 2D and 3D features. SPIE Proceedings Three-Dimensional Imaging, Visualization, and Display, vol. 83840G, pp. 83840G-83841–83840G-83811 (2012)
32. Dalal, N.: The INRIA Person Dataset. http://pascal.inrialpes.fr/data/human/ (2005)
33. Papageorgiou, C.P., Oren, M., Poggio, T.: A general framework for object detection. In: 6th International Conference on Computer Vision, pp. 555–562 (1998)
34. Freund, Y., Schapire, R.E.: A short introduction to boosting. Journal of Japanese Society for Artificial Intelligence **14**(5), 771–780 (1999)
35. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995). doi:10.1007/BF00994018
36. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, pp. 886–893 (2005)
37. Shackleford, W.: Position tracking performance metrics. https://github.com/usnistgov/PTPM (2014)
38. ISO TC 199: ISO 13855:2010 Safety of machinery – Positioning of safeguards with respect to the approach speeds of parts of the human body. In: International Organization for Standardization (ed.) (2010)
39. Bodt, B., Hong, T.: UGV safe operations capstone experiment. Paper presented at the Army science conference, Orlando, 2010

William Shackleford is an electronics engineer in the Sensing and Perception Systems group at the U.S. National Institute of Standards and Technology in Gaithersburg, Maryland. He received a B.S. in electrical engineering from the University of Maryland in 1993, and an M.S. in computer science from the Johns Hopkins University in 2013. His research interests include mobile robotics, autonomous guided vehicles, communications software and sensing. Past projects include the Enhanced Machine Controller (EMC), and open source machine-tool controller, an entry in the DARPA Learning Applied to Ground Robots (LAGR) challenge, and the Aerial Multi-axis Platform (AMP) a cable-driven system used for airplane de-painting. He is the developer of the Neutral Message Language (NML) and Real-Time Control System (RCS) library. He was a co-author of the book “The RCS Handbook: Tools for Real Time Control Systems Software Development” published in 2001 by John Wiley and Sons, Inc.

Geraldine Cheok is a research structural engineer in the Sensing & Perception Systems Group of the Intelligent Systems Division (ISD) of the Engineering Laboratory (EL) at the National Institute of Standards and Technology (NIST). Ms. Cheok graduated cum laude from the University of Maryland with a Bachelor of Science in Civil Engineering in 1983. She received her Master in Science in Civil Engineering from the University of Maryland in 1986. While working on her graduate degree, Ms. Cheok joined the National Institute of Standards and Technology (NIST) Structures Division in 1984, and has been with NIST since. Upon joining NIST, she worked for 14 years in the earthquake engineering area, and was mainly involved in the experimental testing and data analysis of model and prototype bridge columns, steel braced frame connections, and moment resisting precast beam-column connections for seismic regions. She also worked on projects involving the seismic response of lightly-reinforced masonry walls and strengthening of lightly reinforced concrete building frames. Between 1998 and present, her work involved the use of 3D imaging systems (e.g., laser scanners, LADAR, laser radars, range cameras) for construction applications. Her efforts involved research that will be used to support the development of standard test methods for 3D imaging systems. She has also been involved in the performance evaluation of perception systems for manufacturing applications. Ms. Cheok was awarded the Building and Fire Research Laboratory's Communicator Award in 1988 and 1997. In 1996, she was named Marylands Outstanding Young Engineer. In 1997, she received the American Concrete Institute Structural Research Award. That same year, Ms. Cheok accepted the Department of Commerce Bronze Medal. In 2001, she earned the Department of Commerce Silver Medal. She again received the Bronze Medal in 2004.

Tsai Hong received her PhD in Computer Science from the University of Maryland at College Park in 1982. Dr. Hong plays a major leadership role in research activities of perception systems and perception performance evaluation for manufacturing and human robot interaction perception applications. She was responsible for algorithm design and software development of several major modules for areas of perception. Dr. Hong is the chair for the ASTM perception standard group and is responsible for development and technical supervision of research projects in developing advanced measurements for perception manufacturing applications. Her research activities include realtime vision, world modeling, sensor fusion and calibration. She has published over 100 articles on the above research areas.

Kamel Saidi is a Mechanical Engineer in the Manipulation & Mobility Systems Group of the Intelligent Systems Division (ISD) of the Engineering Laboratory (EL) at the National Institute of Standards and Technology (NIST). His current research includes the development of metrics, test methods and ground-truth systems for evaluating the performance of robot perception, safety, dexterity, mapping and autonomy for manufacturing as well as emergency response robots. He is also involved in research aimed at developing performance evaluation methods, metrics, and facilities for 3D Imaging Systems (through ASTM E57). Dr. Saidi conducted research in construction automation for 10 years prior to shifting his focus to manufacturing and emergency response robotics. Dr. Saidi holds a Ph.D. in Civil Engineering from The University of Texas at Austin (UT), where he focused his research on information technology at the construction site and construction automation systems. His graduate research included the design and implementation of a tele-operated, hydraulic robot for the maintenance of electric power plant boilers (for which he is a co-inventor on a US patent). Dr. Saidi's graduate research also involved the design and implementation of a road crack-sealing robot and a largescale manipulator (robot arm) for construction operations. Dr. Saidi also holds a Master of Science and a Bachelor of Science in Mechanical Engineering with an emphasis on Mechanical Systems and Design (and a minor emphasis on Biomedical Engineering) from UT.

Michael Shneier is Group Leader of the Sensing and Perception Systems Group in the Intelligent Systems Division of the National Institute for Standards and Technology (NIST). He received his PhD in Artificial Intelligence from the University of Edinburgh, Scotland. After a postdoctoral appointment at the Computer Vision Laboratory at the University of Maryland, College Park, he moved to NIST (then the National Bureau of Standards), where he worked in the Automated Manufacturing Research Facility. He then worked at Philips Laboratories in Briarcliff Manor, NY, where he initiated the medical imaging program and managed the Intelligent Systems department. He spent five years managing the Artificial Intelligence research program at the Office of Naval Research before returning to NIST. Dr Shneier has conducted research in the areas of robotics, autonomous navigation, object recognition, diagnostic imaging, and manufacturing. He participates in standards activities related to manufacturing robot safety.