

# Human Intention Estimation based on Hidden Markov Model Motion Validation for Safe Flexible Robotized Warehouses

Tomislav Petković<sup>a</sup>, David Puljiz<sup>b</sup>, Ivan Marković<sup>a</sup>, Björn Hein<sup>b</sup>

<sup>a</sup>University of Zagreb Faculty of Electrical Engineering and Computing,  
Department of Control and Computer Engineering,  
Laboratory for Autonomous Systems and Mobile Robotics,  
Unska 3, HR-10000, Zagreb, Croatia,  
petkovic@fer.hr, ivan.markovic@fer.hr

<sup>b</sup>Karlsruhe Institute of Technology,  
Institute for Anthropomatics and Robotics,  
Intelligent Process Automation and Robotics Lab,  
Kaiserstraße 12, DE-76131, Deutschland, Karlsruhe  
david.puljiz@kit.edu, bjoern.hein@kit.edu

---

## Abstract

With the substantial growth of logistics businesses the need for larger warehouses and their automation arises, thus using robots as assistants to human workers is becoming a priority. In order to operate efficiently and safely, robot assistants or the supervising system should recognize human intentions in real-time. Theory of mind (ToM) is an intuitive human conception of other humans' mental state, i.e., beliefs and desires, and how they cause behavior. In this paper we propose a ToM based human intention estimation algorithm for flexible robotized warehouses. We observe human's, i.e., worker's motion and validate it with respect to the goal locations using generalized Voronoi diagram based path planning. These observations are then processed by the proposed hidden Markov model framework which estimates worker intentions in an online manner, capable of handling changing environments. To test the proposed intention estimation we ran experiments in a real-world laboratory warehouse with a worker wearing Microsoft HoloLens augmented reality glasses. Furthermore, in order to demonstrate the scalability of the approach to larger warehouses, we propose to use virtual reality digital warehouse twins in order to realistically simulate worker behavior. We conducted intention estimation experiments in the larger warehouse digital twin with up to 24 running robots. We demonstrate that the proposed framework estimates warehouse worker intentions precisely and in the end we discuss the experimental results.

## Keywords:

human intention estimation, hidden Markov model, virtual reality, Theory of Mind

---

## 1. Introduction

Substantial growth of logistics business in recent years has generated the need for larger and more efficient warehouse systems. State-of-the-art approaches in warehouse automation include solutions such as the Swisslog's CarryPick Mobile system and Amazon's Kiva system [1] which use movable racks that can be lifted by a fleet of small, autonomous robots. By bringing the product to the worker, productivity is increased by a factor of two or more, while simultaneously improving accountability and flexibility [2]. However, current automation solutions are based on strict separation of humans and robots; the worker is not allowed to enter the shop floor during operation due to safety reasons, since a robot with rack can weigh together up to a ton. When moving they are posing a significant risk to all humans nearby. Therefore, when human intervention is needed in the shop floor, the whole fleet of mobile robots has to be stopped and remain stopped until the worker has again left the shop floor. With the increasing size of warehouses, such events immensely impact operation efficiency. Therefore, a novel integrated paradigm arises where humans and robots

will work closely together and these integrated warehouse models will fundamentally change the way we use mobile robots in modern warehouses.

Besides safety as fundamental requirement in every human robot interaction (HRI) scenario, the proposed system has to take usability into account. It has to be ensured that the worker is assisted and not impeded during work. One way of achieving this is to on the one hand ensure that robots are avoiding the area near workers and on the other hand instruct the worker to reach its goal through robot free zones and corridors. Given that, we assert that a future warehouse system, which will have to orchestrate and coordinate human workers and robots, would significantly benefit if it were able to estimate worker's intentions correctly and control the robots accordingly, so that warehouse operation efficiency is ensured.

There exists a plethora of challenges in human intention estimation, because of the subtlety and diversity of human behaviors [3]. Contrary to some physical characteristics, such as the position and velocity, the human intention is not directly observable and usually needs to be estimated from human ac-

tions [4]. It is also imperative to put those actions in context because even basic behaviors, such as walking, running and jumping, are interpreted differently in, e.g., sports event, office and warehouse environment. Furthermore, because the human intention estimation should serve as input to decision making processes [5], the intentions should be estimated in real-time and overly complicated models should be avoided. Having that in mind, only the actions with the greatest influence on intention perception should be considered as inputs to human intention recognition model. For example, in the warehouse domain, worker’s orientation and motion as well as mobile robots’ movement have large effect on the goal intention recognition. On the contrary, observing, e.g., worker’s heart rate or perspiration could provide very little, if any, information on worker’s intentions. Therefore, such measurements should be avoided in order to reduce model complexity and ensure real-time operation [3]. In an integrated warehouse environment (Fig. 1), worker’s position and orientation, which are cues that will be used for intention estimation, need to be precisely estimated and in the present paper we assume that these quantities are readily available as well as preexisting and defined warehouse floorplan. Furthermore, given that conducting experiments in large scale operational commercial warehouses is impractical (which are usually 24h operational), methods for suitably testing the algorithms which require realistic human interaction need to be found. For these purposes, using augmented and virtual reality seems to be a proper and practical solution.



Figure 1: Illustration [6] of an integrated warehouse system where a fleet of mobile robots can move under the racks as well as carry them. Because of that the warehouse’s layout can change and opening or closing of existing passages can occur.

Augmented reality (AR) and virtual reality (VR) themselves have seen a big resurgence in robotics in the recent years [7]. VR refers to systems where the input from the outside world is totally blocked and replaced by a system-generated input. The first VR system was built in 1968 by Ivan Sutherland [8]. The device was extremely bulky and the screen resolution poor; however, it proved that VR was achievable. The first usable VR system came in 1992 with the CAVE system [9]. The CAVE is a special room where electromagnetic (now infrared) trackers track 3D glasses (now usually with active polarization) and

projectors display the appropriate images on the room’s surfaces. Although viable as a system, especially for prototyping [10], it is not flexible enough to be widely employed, requiring expensive sensors and a purposely designed room. Augmented reality, in contrast to VR, seeks to add information to the input from the real world. The first AR systems started to appear in the 1980s, mostly for military research. The most commonly quoted first functional AR system was the *Virtual Fixtures* system [11], developed in 1992 by Louis Rosenberg for the US Air Force. Quickly its usefulness became apparent in other fields as well, such as manufacturing, medicine, entertainment and robotics, where the first use cases focused on assistance in robotic teleoperation via a stereo camera pair [12]. These first AR systems in robotics added data to a camera stream and displayed the enhanced view on computer monitors [13]. Today’s AR can be mostly divided into projector based, tablet/smartphone based, head mounted and computer screen based. Given the described potential, in the present paper we will leverage these systems for providing human intention cues and construct realistic experimental scenarios. For example, AR can be used to track worker motion inside the warehouse as well as display valuable information, e.g, navigate the worker to a specific product that needs to be picked, or assist in repairing a broken robot, while VR can be used to construct elaborate virtual warehouses with realistic simulations of worker interaction [14].

In this paper, we propose an efficient warehouse worker intention estimation algorithm for safe flexible robotized warehouses motivated by the Bayesian Theory of Mind approach. The worker is placed inside of a warehouse with multiple potential goals. We assume that the position and orientation of the worker are measured and that the warehouse fleet management system knows the locations of all the robots. Based on this information, the proposed algorithm estimates the probability of the worker desiring each of the goals. Furthermore, in our model we have also included a specific state of an *irrational worker*, which could indicate to the supervisory system aberrant behavior and that intervention might be necessary. Intentions are estimated using an hidden Markov model (HMM), where the probability of desiring each goal is modeled as a hidden state within the framework. The HMM computes hidden state probabilities by leveraging observations generated as worker action validations based on comparing worker’s motion to possible generalized Voronoi diagram paths. Furthermore, in the present paper we also propose to use an interactive 3D simulation of a flexible warehouse in a VR environment allowing us to mimic worker’s interaction and to test the intention estimation exhaustively. Given that, we ran multiple experiments in virtual reality warehouses, as well as tested the proposed algorithm in a real-world laboratory warehouse with a worker wearing augmented reality glasses. The laboratory warehouse as well as data and use case for real-world experiments were provided by the SafeLog project’s [15] partner Swisslog.

The paper is organized as follows. In Section 2 we present related research and elaborate the contributions in details. In Section 3 we give preliminaries on HMMs and Voronoi graphs. Section 4 presents the proposed worker intention estimation

methodology. In the end, experimental results for the real-world and virtual reality warehouses are given in Section 5, while Section 6 concludes the paper.

## 2. Related Work

Our work follows up on the Bayesian Theory of Mind (BToM) framework described in [16], where authors introduced a model for estimating hungry student’s desires to eat at a particular food-truck by observing its movement. Therein, authors argue that machines lack the Theory of Mind – an intuitive concept humans have about other people’s mental state, and propose to emulate it by an intention recognition model based on Partially observable Markov decision processes (POMDPs). Though impressive, the BToM model does not predict the possibility of student’s change of mind and does not ensure real-time operation in changing environments which is crucial in an integrated warehouse problem. Many models addressing the problem of human intention recognition successfully emulate human social intelligence using Markov decision processes (MDPs). Examples of such models can be found in [3], where authors proposed a framework for estimating pedestrian intention to cross the road, and in [17], where authors proposed a framework for gesture recognition and robot assisted coffee serving.

There are also works from the gaming industry perspective, proposing methods for improving the non-playable character’s assisting efficiency [18, 19]. A driver intention recognition problem was approached in [20] and intention estimation based on gaze data was introduced in [21]. Both of those approaches use learning methods for training the models which has been criticized by [22] emphasizing the drawback of using motion pattern learning techniques for trajectory estimation or intention recognition. Authors assert that such techniques operate offline and imply that at least one example of every possible motion pattern is contained in the learning data set which does not hold in practice. They propose using growing hidden Markov models (GHMM) for predicting human motion, a problem which we consider dual to the human intention estimation in the warehouse domain. GHMMs can be described as a time-evolving HMMs with continuous observation variables where model structure and parameters are updated every time a new observation sequence is available. That kind of approach can be applied to human intention recognition problem, because it enables adding new goals during the experiment as well as an elegant framework for learning the model parameters online. Assistive technology such as smart homes [23], exoskeletons [24] and semi-autonomous wheelchairs [25] benefit also from precise human intention recognition. In [25], authors propose a POMDP driven algorithm for wheelchair control taking into account the uncertainty of user’s inputs because of, e.g., unsteady hands. The chair predicts user’s intention and autonomously enacts the intention with only minimal corrective input from the user. The authors also suggest that humans usually focus on moving from one spatial location to another, i.e., hallway to kitchen, without worrying about the optimality of exact steps that come in between. In the present paper, we build our model

using similar assumptions about human spatial understanding. In [5] Anh and Pereira offer thorough review of human intention recognition area emphasizing its potential applications in decision making theory.

Another emerging assistive technology is Mixed Reality (MR), a term usually including Augmented Reality, Virtual Reality, or a mix of thereof. As stated before one can leverage localization and sensor systems required for immersive AR for intuitive information display and the tracking of human motions, while VR can be used to realistically simulate environments for training and testing purposes. Here we give a brief overview of current MR systems and their applications.

Advancements in VR technology have seen the CAVE systems of the early 90s being replaced by cheaper and more flexible headsets such as the *Oculus Rift* or the *HTC Vive*. This has sparked a boom in the field of robotics where VR has recently seen the most use as a more intuitive method for teleoperation of stationary [26], mobile [27] and humanoid robots [28]. A natural extension of such a teleoperation systems is a more immersive tool for telepresence robots [29]. It has also emerged as a method to teach virtual robots how to preform tasks [30], where the knowledge is then successfully transferred to a real robot. In manufacturing [31] VR has seen use as a virtual prototyping [32] and training tool [33]. As a training tool, VR has been shown to increase performance of trainees in other areas as well, such as medicine [34], safety training in construction [35], and mining [36].

On the AR side of things, projector based AR has seen use for visualizing robot’s intentions and intuitive programming of robots in robot work cells [37], displaying intentions of mobile robots operating in human environments [38], as well as for debugging and rapid prototyping of robotic systems through visualization of sensor data [39]. Tablets have seen use in AR-assisted robot programming [40] as well. Devices like the *Google Tango* tablet, in addition, with inbuilt SLAM [41] can be used for markerless AR applications. The main challenge with tablets is that it occupies the hands, preventing any work while the AR information is visualized. With the recent releases of the ARCore and ARKit toolkits, for Android and iOS devices respectively, tablet and smartphone based AR applications are becoming an economical and straightforward interaction modality for home robots [42]. Head mounted systems can be further divided into Heads-up Display (HUD) systems and “full-AR” systems, with the most famous member of the former being the *Google Glass* and of the later *Microsoft Hololens*. As the name implies, HUD-based systems do not have any advanced localization or computing systems and therefore are only able to display interfaces, while full AR systems are able to perform localization using SLAM and display persistent, full 3D holograms in space. HUD systems have found applications in logistics, where they have been used for *pick-by-vision* systems to quicken and ease the picking of items in warehouses [43]. Since full AR systems, starting with the *Hololens*, have not been on the market for long, research is just starting [7], with the most prominent field perhaps being AR assisted robot programming [44].

The present paper draws upon our earlier work [45], where

we have presented the preliminary version of the human intention estimation algorithm. Therein, the algorithm was also motivated by the Bayesian Theory of Mind [16] and tested in a simulated  $20 \times 20$  cells large environment. Worker actions were validated by comparing worker actions with the optimal policy within an MDP framework, where the value iteration algorithm needed to be executed beforehand. In case of any changes in the environment, such as robots blocking a path that was previously free, value iteration would need to be rerun in order to find the optimal policy. Such an approach proved to be infeasible for large flexible warehouses where multiple robots could constantly block worker’s path, thus changing often the configuration of the environment. Given that, in the present paper we propose a worker intention estimation algorithm for safe flexible robotized algorithm that solves the aforementioned problem by creating first a generalized Voronoi diagram of the warehouse and running the D\* algorithm in order to find the optimal path between each two nodes. In case of a robot blocking the path, we can simply cut the edge between the nodes, thus ensuring efficient worker motion validation that is fed to the HMM as observations. Hidden states of the HMM each model worker intentions towards the goals, additionally including a specific state which models an irrational worker. This approach also enables us to efficiently add goals during the experiment, which was previously not possible. In order to test the intention estimation algorithm we ran experiments in a real-world laboratory warehouse with worker wearing Microsoft HoloLens augmented reality glasses. The HoloLens localization algorithm was used to generate worker location and orientation estimates. Furthermore, in order to demonstrate the scalability of the approach to larger warehouses, we propose to use VR digital warehouse twins in order to simulate realistically worker behavior. First, the real-world experiment warehouse was modeled in VR and the pertinent experiment was recreated in order to showcase the realistic nature of the VR experiments and compare the outputs of the intention estimation algorithm. Then, further intention estimation experiments were ran in a larger warehouse digital twin with up to 24 running robots. The results corroborate that the proposed framework estimates warehouse worker’s desires precisely and in an intuitive manner.

### 3. Preliminaries on HMMs and Voronoi graphs

Markov decision processes (MDPs) constitute a mathematical framework which models a system taking a sequence of actions under uncertainty to maximize its total rewards [3]. More precisely, an MDP is a discrete time stochastic process represented as tuple  $(\mathcal{S}, \mathcal{A}, T, \mathcal{R}, \gamma)$ , where  $\mathcal{S}$  is set of states and  $\mathcal{A}$  is set of actions. After an action  $a \in \mathcal{A}$  is taken, system moves from the current state  $s \in \mathcal{S}$  to a new state  $s' \in \mathcal{S}$ . We define the conditional probability function  $T(s, a, s') = p(s'|s, a)$  which gives the probability that the system lies in  $s'$  after taking the action  $a$  in state  $s$ , thus capturing system’s uncertainty. Taking an action also yields an immediate reward  $R(s, a)$  and the goal of the system is to choose the sequence of actions that maximizes the expected total reward  $E(\sum_{t=0}^{\infty} \gamma^t R(s_t, a_T))$ . To prevent an infinite-horizon case where all positive rewards sum

to infinity [46], one uses a discount factor  $\gamma \in (0, 1)$  which reflects system preference of immediate rewards over future ones. The MDP model can be too restrictive to be applicable to many problems of interest [47] because it assumes that all states are fully observable.

The hidden Markov model (HMM) is an MDP extension including the case where the observation is a probabilistic function of the state, i.e., the resulting model is a doubly embedded stochastic process with an underlying process that is not observable (it is hidden), but can only be observed through another set of stochastic processes that produce the sequence of observation. In general, when using HMMs we are interested in solving one of the following three problems. First, given an existing HMM and an observed sequence, we want to know the probability that the HMM could generate such a sequence (the scoring problem). Second, we want to know the optimal state sequence that the HMM would use to generate the sequence of such observations (the alignment problem). Third, given a large amount of data, we want to find the structure and parameters of the HMM which best account for the data (the training problem). In this paper we focus on optimal state sequence problem, i.e., the alignment problem, for warehouse worker actions with the aim of estimating worker intentions.

The HMM gives us a procedure for handling observations and turning them into intention estimations. In order to produce observations we need to also find a way to efficiently encode the environment and validate worker’s actions within. For this task we selected the Voronoi diagram. Voronoi diagram partitions a plane into regions based on distance to predefined points. The idea is that for each predefined point a corresponding region consisting of all the points closer to that point than to any other predefined point is found. More formally, the definition of Voronoi region for point  $p_i \in P$  is given by:

$$V(p_i) = \{\vec{x} \mid \|\vec{x} - \vec{p}_i\| \leq \|\vec{x} - \vec{p}_j\|, \forall_{j \in \{1, \dots, |P|\}} i \neq j\}, \quad (1)$$

where  $\|\cdot\|$  is usually the Euclidean distance [48]. We call the set given by  $V = \{V(p_1), V(p_2), \dots, V(p_n)\}$  the Voronoi diagram of  $P$ . The question is, how to construct a Voronoi diagram in practice? The generalized Voronoi diagram (GVD) is a discrete form of the Voronoi diagram defined as the set of points in free space to which the two closest obstacles have the same distance [49]. In mobile robotics applications, the GVD can be constructed from an occupancy grid map of the environment which is usually obtained by mapping the environment with a mobile robot [50] or by parsing the existing floorplans. In the present paper, we will use the latter approach, since we have warehouse plans at our disposal.

### 4. Worker intention estimation method

Theory of Mind (ToM) is the intuitive grasp that humans have of their own and other people’s mental states, how they are structured, how they relate to the world, and how they cause behavior [16]. Human beings understand that others have certain desires and that those desires guide them to act using means

most likely to achieve them. However, *explanation by rationalization* reasoning, which ToM assumes, is highly contextually dependent and translating such causal behavior model to machines is not an easy task. Having that in mind, we limit the proposed model to a problem of estimating intention of a human worker in a highly flexible robotized warehouse. The worker can perform tasks such as maintaining the robots or picking the items from the racks containing goods. We assume that there is finite number of possible goal locations which are usually in front of the *interesting* racks and that at least one goal is known before the start of the experiment. Furthermore, we also assume that the position and orientation of the worker are measured. This can be achieved by, e.g., augmented reality glasses, other types of wearable sensors, or special vests equipped with vision sensors as developed in the scope of the project SafeLog [15]. Dennett [51] has proposed that social reasoning abilities rely on *intentional stance*, i.e., the assumption that agents have beliefs about the world and their situation in it and will behave rationally to achieve their desires. We argue that, in the warehouse domain, rationally behaving with respect to the desired goal manifests as moving towards that goal’s location, and that since our agents are workers, trained professionals, they are highly likely to behave rationally within this context.

In order to determine the most likely goal the worker is moving to, we need to apply a complete and globally optimal path planning algorithms and compare the worker’s motion on-line with the algorithm output (details are discussed later). We assert that worker following approximately a globally optimal path is a reasonable assumption, since the worker is acquainted with the warehouse layout and will plan its motion in accordance to it. Given that, we find that for the problem at hand, where action uncertainties are reduced, frameworks such as POMDPs used in [16] are not necessary. Moreover, the planning algorithm must be able to quickly replan the path with the appearance of moving obstacles such as mobile robots carrying the racks. Planning using MDP solvers [46] offers a well grounded tool for human intention recognition by allowing elegant comparison of agent actions with respect to the optimal policy [45]; however, it lacks the ability to quickly replan, since the optimal policy for efficient intention recognition must be computed offline for realistic warehouses. Given that, in the present paper we choose to use the D\* algorithm [52] for finding the globally optimal path to the goals. However, having in mind that the modern warehouses are growing in size we aim to reduce the complexity of mentioned search problem. One approach to alleviating the complexity is to reduce the precision of the warehouse occupancy grid representation, thus reducing the D\* algorithm search space; however, we assert that it could jeopardize the proposed human intention recognition performance and cannot be applied to arbitrary large warehouse, thus directly impacting the proposed algorithm’s ability to generalize.

To solve the aforementioned challenges, we propose to use generalized Voronoi diagrams (GVDs) for reducing the search space without losing valuable precision for intention recognition. Motivation for using GVDs in our work is manifold. First, partitioning the plane using GVDs allows us to limit the search space on the Voronoi graph nodes, thus greatly reducing the

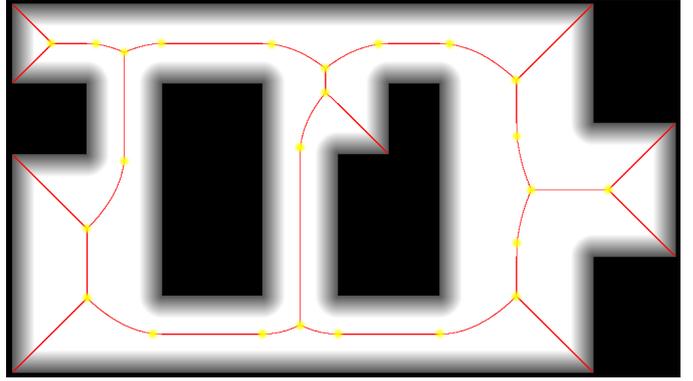


Figure 2: Generalized Voronoi Diagram (red) of Swisslog’s warehouse in Ettlingen with highlighted graph’s nodes (yellow). Untraversable parts of the warehouse such as walls or racks are denoted with black color. Shades of grey denote the distance of the traversable part of the warehouse from the obstacles. The distance is used for GVD generation.

search time. Second, moving along the edges of a Voronoi graph ensures the greatest possible clearance when passing between obstacles. This property resembles assumed human path planning in a warehouse application, because human beings are generally not prone to walking in the proximity of warehouse racks (note that in our example robots can also pass under the racks). Finally, moving obstacles, such as mobile robots, in flexible warehouse systems can obstruct worker’s path necessitating replanning of the optimal path towards that goal. The replanning can easily be achieved using graph search algorithms by discarding the edge of the Voronoi graph the robot is currently occupying. On the other hand, we can imagine a scenario where new possible passages could appear. This could happen when a mobile robot takes a rack and frees up space in the middle of the rack block. In order to handle that event, a new GVD would have to be generated from the occupancy grid map of the warehouse. This could impede the online application of the proposed approach; however, in order for that to happen, since usually one rack block has two columns of racks, multiple racks close to each other would have to be carried away at similar time intervals. Furthermore, we assert that workers in robotized warehouses would generally not be allowed to use such passages for safety reasons. Having that in mind, from the methodological perspective for the approach, we propose to use the D\* algorithm on GVD nodes for worker path planning. In the sequel we describe the proposed human intention recognition algorithm in details and divide it in three parts: creation of the warehouse GVD (offline), worker action validation, and worker intention estimation (both online).

#### 4.1. Creating warehouse generalized Voronoi diagram

We have already emphasized the necessity of performing human intention recognition online. In order to ensure online operation, we need to perform time consuming parts of the algorithm before the start of the experiment. First, we generate the GVD [53] of the warehouse using its floorplan. Example of such a graph generated on Swisslog’s warehouse using an occupancy grid with cell size of 5 mm can be seen in Fig. 2. After

---

**Algorithm 1** Human intention recognition
 

---

```

1: while True do
2:   if Worker moved or turned significantly then
3:      $d \leftarrow$  Modulated distance to every goal
4:     for Proximate positions and orientations do
5:        $D_i \leftarrow$  Modulated distance to every goal
6:     Update intention estimation( $d, D_i$ )
  
```

---

generating GVD, we select all graph nodes for further processing. It is worth noting that it is possible to insert additional nodes at arbitrary locations, but erasing generated ones except dead ends, is not allowed because it would impede graph's connectivity. Also, at least one goal node must be added before the experiment starts, since it is not possible to emulate ToM inference without having any hypotheses about possible goal locations. In this work we add only the goal nodes and discard the dead ends creating a node set which we denote  $\mathcal{N}$ . With the obtained node set, we run the D\* algorithm to find the optimal path between each two nodes and save all the relative distances in a matrix  $\mathbf{F}$ , where element  $\mathbf{F}_{i,j}$  denotes distance in pixels between nodes  $i$  and  $j$ . This might not be the optimal approach to finding relative distances between the nodes, but this part is performed offline and done only once before the start of the experiment.

#### 4.2. Worker action validation

During the online phase we monitor worker's position and orientation provided by the Microsoft HoloLens augmented reality device as well as positions of mobile robots (provided by the warehouse fleet management system). Mobile robots are treated as moving obstacles with radius  $r = 1$  m and the worker's wearable device shows positions of nearby robots. If the robot is located on a GVD edge, we cut that edge from the graph and update the relative distance matrix  $\mathbf{F}$  using the D\* algorithm.

The final objective of the human intention algorithm is to estimate towards which goal the worker is currently going to by observing worker motion and comparing it to the optimal path to each of the goals. With each worker's position and orientation information update, we check if (i) worker's position mapped to the occupancy grid floor plan has changed or if (ii) worker's orientation has changed more than  $\frac{\pi}{8}$  from the last intention estimation update. If any of these conditions is met, we perform an intention estimation update by associating observed worker's position and orientation with each of the nodes in set  $\mathcal{N}$  using vector  $\mathbf{c}$  as follows:

$$c_i = \begin{cases} 0, & \text{an unobstructed straight line exists} \\ & \text{between the worker and the } i\text{-th node} \\ \mathcal{G}(d_i, \sigma^2) \cdot \Phi(\theta_i), & \text{otherwise.} \end{cases} \quad (2)$$

In (2)  $\mathcal{G}$  is a Gaussian function with variance  $\sigma^2 = 0.005$ , which we obtained experimentally, and  $d_i$  is distance between worker's position and  $i$ -th node's location. Bell shaped functions such as Gaussians have been used for navigating in contin-

uous spaces [54], which motivated us to choose them as a proximity measure. They are smooth and monotonically decreasing functions of distance and have a non-zero value on the entire domain, allowing an intuitive association of worker's position with every visible node. We also assert that every worker's position will always be associated with at least one node because of GVD's space covering properties. The only exception to fore-mentioned claim is if the worker is trapped by mobile robots but we argue that warehouse management system must never allow such event to occur for obvious safety concerns. Furthermore, we also modulate the Gaussian with the following triangular function:

$$\Phi(\theta_i) = \frac{\pi - |\theta_i|}{\pi^2}, \quad (3)$$

where  $\theta_i \in [-\pi, \pi]$  is the difference between worker orientation and the angle at which the worker sees  $i$ -th node. It amplifies the association with those nodes the worker is oriented at, since we assume that the worker will look at the path it is planning to take [45]. We also need to define the isolation matrix  $\mathbf{I}^{n \times g}$ , where  $n$  is the total number of nodes and  $g$  is number of goal nodes:

$$\mathbf{I}_{i,j} = \begin{cases} 1, & i = n - g + j \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

We normalize the association vector  $\mathbf{c}$  and obtain modulated approximate distance vector  $\mathbf{d}$  by multiplying it further with the distance matrix  $\mathbf{F}$  and isolation matrix  $\mathbf{I}$ :

$$\mathbf{d} = \mathbf{cFI}. \quad (5)$$

Each element of vector  $\mathbf{d}$  represents modulated measure of distance to the respective goal. In order to find out if the worker is moving towards or away from the goal, we compare this distance to alternative worker positions and orientations. We take the location  $l'$  at which the worker was prior to the last intention estimation update, and calculate the difference  $r$  between that position and current worker's position  $l$ . Then, we generate set of  $m = 16$  equidistant points  $p$  on a circle around  $l'$  with the radius  $r$  which is shown in Fig. 3. For each point  $p_i \in p$  and potential worker orientation  $\theta \in \{-\pi, -\frac{3\pi}{4}, \dots, \pi\}$  we repeat the calculation of vectors  $\mathbf{c}$  and  $\mathbf{d}$  and append the result to potential modulated distances matrix  $\mathbf{D}$ . Computing the matrix  $\mathbf{D}$  enables us to validate worker's motion with respect to states it *could be in*, rather than to the state it *had been to*. The proposed algorithm pseudocode can be seen in Algorithm 1.

We use the distance vector  $\mathbf{d}$  and the distance matrix  $\mathbf{D}$  to validate worker's actions, and we introduce the motion validation vector  $\mathbf{v}$ , which is computed as follows:

$$\mathbf{v} = \frac{\max_{1 \leq i \leq n} \mathbf{D}_{ij} - \mathbf{d}}{\max_{1 \leq i \leq n} \mathbf{D}_{ij} - \min_{1 \leq i \leq n} \mathbf{D}_{ij}}. \quad (6)$$

If the worker is moving towards the goal, the corresponding value of  $\mathbf{v}$  will be close to unity, and if it is moving away from that goal, the corresponding value will gravitate to zero. One could argue that  $\mathbf{v}$  may be interpreted as the estimate of worker intention, because it expresses a measure of approaching the goal. However, elements of  $\mathbf{v}$  are very sensitive to sensor noise

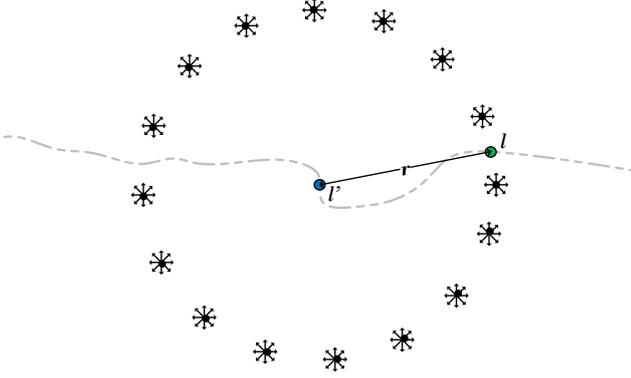


Figure 3: Illustration of generating additional points  $p$  with orientations  $\theta$ . We validate the worker’s motion by comparing the vector  $\mathbf{c}$  of the worker’s position and orientation at location  $l$  with vectors  $\mathbf{c}$  of the newly generated points.

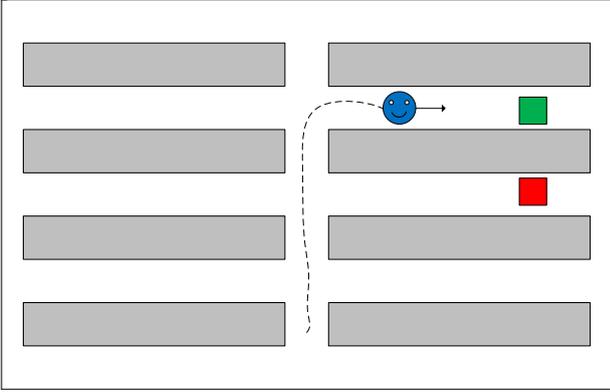


Figure 4: The warehouse worker (blue circle) has previously decided not to turn towards the goal labeled with the red box. It is intuitive that the worker desires the green goal more than the red goal because of its action history despite the fact it is now reducing its distance to both goals. The intention estimation algorithm has to take action history into account when estimating the worker’s intention.

and they would need to be filtered if one wanted to draw inferences about worker intentions directly from them. Additionally, even though observing the current value of  $\mathbf{v}$  does indeed carry crucial information for estimating worker intention, it is not sufficient since it lacks history of past values of  $\mathbf{v}$ . Consider the following simple example depicted in Fig. 4, where a warehouse worker moved past the goal labeled by red square, and further advanced towards the goal labeled with the green square. As soon as the worker turned right, the value of  $\mathbf{v}$  related to the green, as well as red, goal started to grow, since the distance between the worker and that goal started to drop. However, since the worker previously failed to turn towards the red goal, its intention estimation for that goal should have remained low. This example also demonstrates why there is a need to measure distance of traversable path as the approaching measure, instead of simply having, e.g., the Euclidean distance as elements of vector  $\mathbf{v}$ . Before further calculations we apply discrete first-order low-pass filter on  $\mathbf{v}$  in order to reduce the noise influence.

Table 1: HMM framework components

Symbol	Name	Description
$G_i$	Goal $i$	Worker wants to go to Goal $i$
$G_?$	Unknown goal	Worker’s intentions are not certain
$G_x$	Irrational worker	Worker behaves irrationally

### 4.3. HMM based intention estimation

Given that we know how to calculate the motion validation vector  $\mathbf{v}$ , we now introduce the model for solving the worker’s intention estimation problem. While worker’s actions, manifested as moving and turning, are fully observable, they depend on the worker’s inner states (desires), which cannot be observed and need to be estimated [3]. We propose a framework based on hidden Markov model for solving the worker’s intention estimation problem. HMMs are especially known for their application in temporal pattern recognition such as speech, handwriting, gesture recognition [47] and force analysis [55]. They are an MDP extension including the case where the observation (worker’s action) is a probabilistic function of the hidden state (worker’s desires) which cannot be directly observed. We propose a model with  $g + 2$  hidden states shown in Fig. 5 and listed in Table 1.

Hidden states  $G_i$  describe worker’s intention of going to  $i$ -th goal,  $G_?$  indicates that the worker prefers multiple goals and the model cannot decide on the exact desire with enough certainty. On the other hand, hidden state  $G_x$  indicates that the worker is moving away from all the goal locations. This hidden state models the case of the worker being irrational or worker desiring a goal we have not yet specified. The proposed model cannot distinguish between these two cases. Introduced hidden states enable the human intention recognition model to elegantly save the intention estimation history as probabilities  $P(G_i)$ . The first building block in this HMM architecture is the transition matrix  $\mathbf{T}^{g+2 \times g+2}$ :

$$\mathbf{T} = \begin{bmatrix} 1 - \alpha & 0 & \dots & \alpha & 0 \\ 0 & 1 - \alpha & \dots & \alpha & 0 \\ \vdots & & \ddots & & \vdots \\ \beta & \beta & \dots & 1 - g\beta - \gamma & \gamma \\ 0 & 0 & \dots & \delta & 1 - \delta \end{bmatrix}, \quad (7)$$

where the architecture and description of the matrix parameters can be seen in Fig. 5. We have obtained parameters experimentally as follows:  $\alpha = 0.5, \beta = 0.1, \gamma = 0.05, \delta = 0.1$ .

We use the calculated motion validation vector  $\mathbf{v}$  to generate the HMM’s emission matrix  $\mathbf{B}$ . Every time the worker moves or turns significantly, we estimate the worker intention using the Viterbi algorithm [56], which is often used for solving HMM human intention recognition models [57]. The inputs of the Viterbi algorithm are the hidden states set  $S = \{G_1, \dots, G_g, G_?, G_x\}$ , hidden state transition matrix  $\mathbf{T}$ , initial state  $\Pi$ , sequence of observations  $\mathbf{O}$ , and the emission matrix  $\mathbf{B}$ . The HMM framework generally assumes a discrete set of observations, but since our observation is the validation vector  $\mathbf{v}$  with

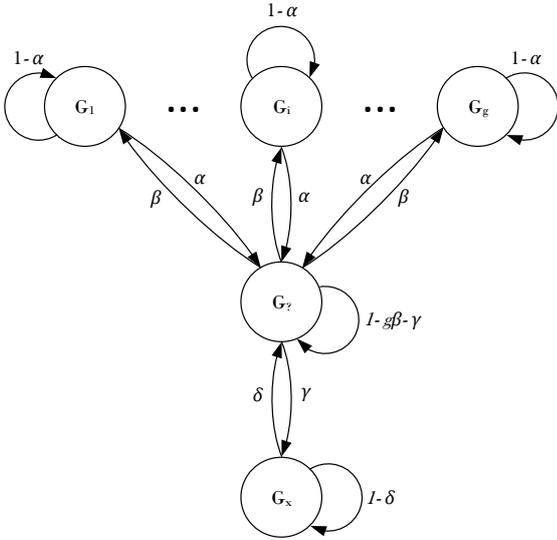


Figure 5: HMM architecture used for human intention recognition. Worker’s change of mind tendency is captured by the parameter  $\alpha$  and the parameter couple  $\beta$  and  $\gamma$  set the threshold for estimating intention for each goal location. Increasing  $\beta$  leads to quicker inference of worker’s intentions and increasing  $\gamma$  speeds up the decision making process. Parameter  $\delta$  captures model’s reluctance to return to estimating the other goal probabilities once it estimated that the worker is irrational.

continuous element values, we have decided to modify the input to the Viterbi algorithm by introducing an expandable emission matrix  $\mathbf{B}$ . While the classic HMM emission matrix  $\mathbf{B}^{n \times g}$  links hidden states with discrete observations via fixed conditional probability values, elements of the introduced expandable emission matrix  $\mathbf{B}^{k \times g}$ , where  $k$  is the recorded number of observations, are functions of the observation value. By using this modification of the emission matrix, we additionally simplify the Viterbi algorithm because there is no set of discrete observations it has to iterate through. Once a new validation vector  $\mathbf{v}$  is calculated, the emission matrix is expanded with the row  $\mathbf{B}'$ , where the element  $\mathbf{B}'_i$  stores the probability of observing  $\mathbf{v}$  from hidden state  $G_i$ . We also calculate the average of the last  $m$  vectors  $\mathbf{v}$  and the maximum average value  $\phi$  is selected. It is used as an indicator if the worker is behaving irrationally, i.e., is not moving towards any predefined goal. The value of the hyperparameter  $m$  decides how much evidence we want to collect before we allow the algorithm to declare the worker irrational. If the worker has been moving towards at least one goal in the last  $m$  iterations ( $\phi > 0.5$ ), we calculate  $\mathbf{B}'$  as:

$$\mathbf{B}' = \zeta \cdot \begin{bmatrix} \tanh(\mathbf{v}) & \tanh(1 - \Delta) & 0 \end{bmatrix}, \quad (8)$$

and otherwise as:

$$\mathbf{B}' = \zeta \cdot \begin{bmatrix} \mathbf{0}_{1 \times g} & \tanh(0.1) & \tanh(1 - \phi) \end{bmatrix}, \quad (9)$$

where  $\zeta$  is a normalizing constant and we calculate  $\Delta$  as difference of the largest and second largest element of  $\mathbf{v}$ . Using such way of calculating  $\Delta$  enables us to simply encode that, in

order for our model to decide in favor for any goal location, it has to significantly stand out from other goals. Humans often infer intentions of others by observing their actions [16], which are generally not optimal with respect to their goals. Nevertheless, we argue that the worker will globally move towards the goal it desires the most, but may locally take suboptimal actions such as looking around or swerving laterally. In order to encode such behavior, we use the hyperbolic tangent function in (8) and (9) to reduce the difference between the actions that indicate movement towards the goal whilst penalizing other actions approximately equal as linear function would. Finally, we set initial probabilities of worker’s intentions as:

$$\Pi = \begin{bmatrix} 0 & \dots & 0 & 1 & 0 \end{bmatrix}, \quad (10)$$

indicating that the initial state is  $G_7$  and the model does not know which goal the worker desires the most. The Viterbi algorithm outputs the most probable hidden state sequence and the probabilities  $P(G_i)$  of each hidden state in each step. These probabilities are the worker’s intention estimates.

Worker tasks are not always predefined in the beginning and can appear or cease during worker’s stay on the shop floor. We have taken such events into consideration and made it possible to add or remove goals during the experiment. If the goal has to be removed, e.g., because other worker took that job over or the task was canceled, we simply discard that goal from our calculations and add it to the unknown goal intention estimation. If the goal has to be added, we do this by setting its intention estimation to  $\max(\min(P(G_i), 0.1)$  with  $0 < i < g$  and by expanding the  $\mathbf{I}$  and  $\mathbf{T}$  matrices. We would like to emphasize that it is not possible to add an arbitrary number of goals, because of the fixed values of parameters in  $\mathbf{T}$  matrix. The maximum number of goals this model can handle is limited because elements of transition matrix  $\mathbf{T}$  must be greater than 0. The only value of matrix  $\mathbf{T}$  that depends on goal number is probability of staying in  $G_7$  state and equals to  $1 - g\beta - \gamma$ . We calculate the maximum number of goals by applying positivity condition to that expression:

$$g_{max} = \left\lfloor \frac{1 - \gamma}{\beta} \right\rfloor, \quad (11)$$

which for our current value of parameters amounts to 9 goals. However, after detailed testing, we have concluded that proposed model does not perform well if there are more than 5 goals. The main difficulty here is that if there are more potential goals, worker’s actions are not discriminatory enough and there is not enough evidence that the worker desires one goal more than the other goals. Because of that, the model estimates  $G_7$  as the most probable state throughout the experiment. While we argue that this still is accurate worker’s intention estimation, it is hardly useful to supervisory system that is taking our algorithm’s estimations as input.

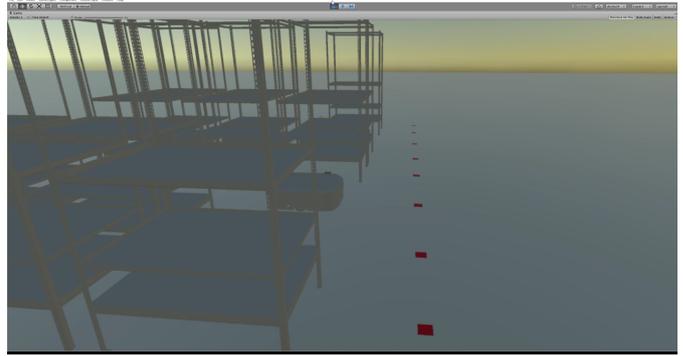
## 5. Experimental Results

In this section we demonstrate proposed human intention estimation algorithm performance and discuss the results. We have conducted experiments in both an industrial setup and in





(a) First-person view from the 2 Megapixel Hololens camera



(b) First-person view from HTC Vive

Figure 7: Comparison between the real-world experiment and the experiments in VR. Both experiments use exactly the same setup.

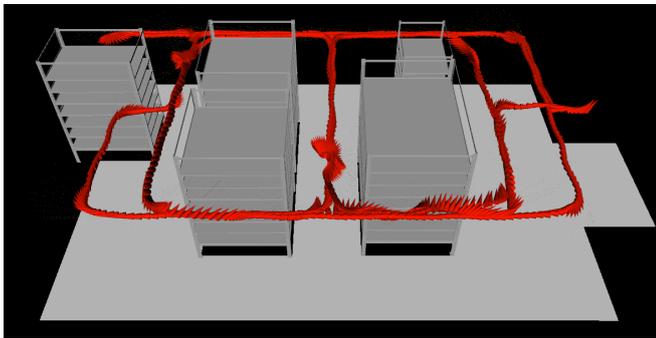


Figure 8: Conducted experiment showcasing the precision of the Hololens' localization. We have created a warehouse model in *rviz* - Robot Operating System's 3D visualization tool, which will be further used to demonstrate proposed algorithm's results.

straight line paths between nodes. A first-person view from the Microsoft Hololens is shown in Fig. 7a.

To test the proposed algorithm, two experiments were conducted (besides the localization experiment described previously). In both experiments the starting point was node R32 and featured two robots; one stationary and positioned at R15, while the other was mobile. In the first experiment shown in Fig. 9 the worker has three potential goals located in front of racks near nodes on R117, R108 and R17. The worker initially starts moving towards node R117, but the mobile robot moves forward from node R107 and turns right towards node R101, thus blocking the worker's initially intended path towards the R117. The worker then turns around and follows the only remaining path towards R117 which is also the best path for the other two goals. Once the worker reached R103 and continued to R104, the proposed algorithm detects that it has failed to turn towards the R17 goal and lowers its intention estimation value. The worker further continues to goal R109 but turns around before reaching it. After turning, the only goal the worker could be going to is R17, but the worker continues past it returning to node P. When worker passed the goal R17, the model recognized there are no goals it could be going to and declared the worker irrational. The estimated intentions by the proposed algorithm can be seen in Fig. 10a. The second experiment included a worker moving towards R117 and changing its mind

to go to R17 simultaneously with mobile robot blocking the best path towards that goal and eventually returning to R117. The main idea of the second experiment was to showcase algorithm's flexibility in scenarios where the worker changes its mind often. Given that, we only show the second experiment in the accompanying video<sup>1</sup>.

## 5.2. Virtual Reality Setup

We built a virtual reality framework for rapid prototyping of applications for flexible robotized warehouses, a tool that might also evolve into a training framework in the future. The main motivation behind a virtual framework, i.e., a virtual reality digital twin of a warehouse, is that commercial automated warehouses are often unavailable for experiments, but they can be simulated, thus avoiding the warehouse downtime and financial losses, yet enabling testing in full scale to identify potential problems and obtain realistic user experience. Furthermore, this approach also enables us to do tests with multiple users more freely and achieve the best possible interaction modalities. Note that besides testing user or worker behavior inside the warehouse, the virtual setup also serves for testing augmented reality applications for warehouse workers wearing glasses such as the *Microsoft Hololens*. Concretely, the application was developed in *Unity3D* and is used with the *HTC Vive* headset. An example of a user view within the virtual reality warehouse is shown in Fig. 7b.

The warehouses layout is planned using Swisslog proprietary network planner, from which an XML file is exported. We have developed an XML file parser which together with an available CAD model builds a VR warehouse from scratch. The fleet of robots is presently controlled by parsing a series of JSON messages, which are the output of a path planner [58].

Currently two AR interaction scenarios have been implemented for prototyping augmented reality applications: path visualization for worker navigation and rack object picking assistance. The virtual setup emulates the Microsoft Hololens: all of the *holographic objects* are only visible through a narrow field of view (FoV), which according to specifications ranges

<sup>1</sup><https://youtu.be/SDD-v-pH0v4>

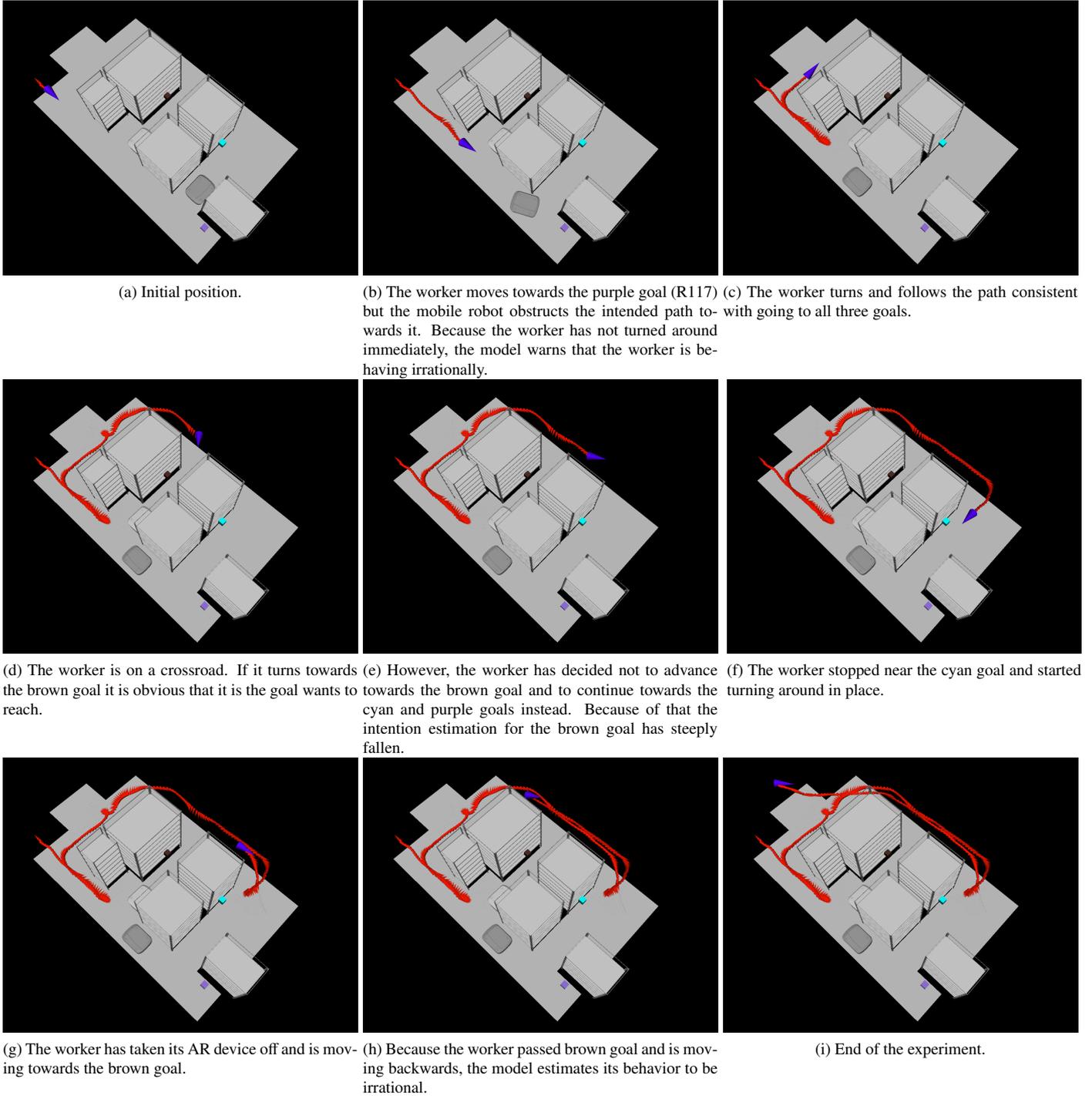


Figure 9: Key moments of the industrial experiment setup in the laboratory warehouse visualized in *rviz*. Goal nodes are labeled with cubes as follows: R117 purple, R17 brown and R109 cyan.

from  $30^\circ$  to  $35^\circ$  horizontally and  $17.5^\circ$  to  $17.82^\circ$  vertically (we selected the lower bounds). The interaction pointer uses ray-casting to position itself with respect to objects. Potentials of the proposed system are numerous. For example, such a system allows us to test if a certain interaction modality fails or is not as informative because of the low FoV in a realistic environment. It also allows us to inject localization errors to determine at which point various interaction modalities become unusable.

However, in the present paper the virtual framework is used primarily as a controlled simulated warehouse environment for conducting experiments for worker intention estimation.

### 5.3. Virtual reality experiments

Due to unavailability of a full-scale commercial warehouse for testing, we conducted larger-scale tests in VR, using the system described in Section 5.2. The tracking method of the HTC

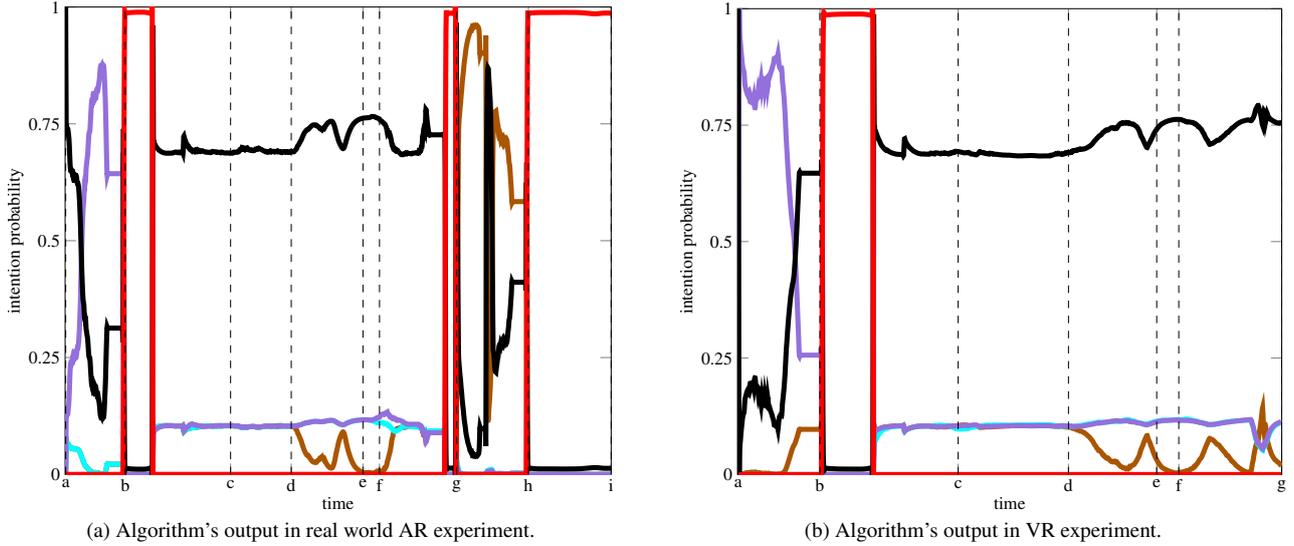


Figure 10: Comparison of the proposed algorithm's output between the real-world experiment and the experiment recreated in VR. Intention estimations for the three goal locations are labeled with respect to their color in Fig. 9 (brown, cyan and purple), the unknown goal state is labeled black and the irrational worker state is labeled red. One can see that the results in this scenario are similar for both AR and VR technology.

Vive has an RMS error of 1.9 mm, which offers very accurate tracking for a realistic VR experience for warehouse localization and worker behavior purposes. We tracked the position of the user, as well as of each robot. An example of a first-person view from the HoloLens used in the laboratory warehouse and HTC Vive in a VR digital twin of the same warehouse can be seen in Fig. 7.

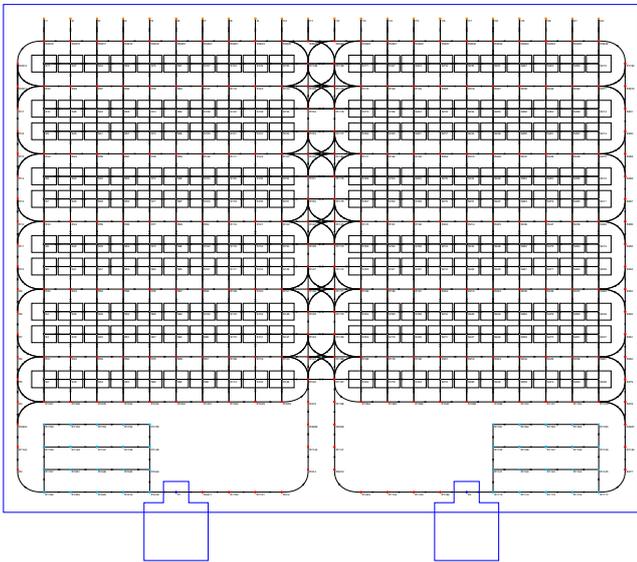


Figure 11: The larger warehouse layout used for VR experiments.

We repeated the first AR experiment in the constructed VR environment to demonstrate applicability of VR for human intention estimation. The only difference between scenarios was that we did not reproduce last part of the AR experiment when worker took off the AR device off to simulate a type of an irrational behavior. The results are shown in Fig. 10b, where we

can see that the biggest difference between real world and VR experiments is at the beginning of the experiment. Those differences are caused by the fact that VR experiment starts with the worker going directly towards the goal, whereas at the beginning of the real word experiment the worker is still putting on the AR device. Because they can be interpreted as semantically similar to AR results, we conclude that testing intention estimation in VR can produce credible results.

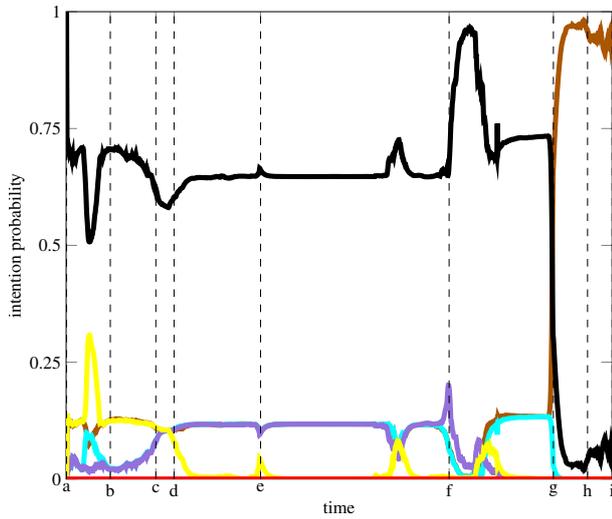
We then proceeded to conducting experiments in a larger VR warehouse (layout can be seen in Fig. 11) with four robots running on preprogrammed paths. These paths were generated by the path planner and read from a text file containing JSON messages. We considered two scenarios for this experiment: one with four initially known goals and one with three initially known goals with one goal being added during the experiment. Visualization of the key moments can be seen in Fig. 12 and the proposed algorithm output is shown in Fig. 13. The worker initially starts moving towards yellow and brown goals. If the brown goal is known, the model cannot decide which goal the worker desires more and stays in the unknown goal state (Fig. 13a). However, if the brown goal has not yet been added, worker moves only to the yellow goal and probability for that goal steeply rises (Fig. 13b). The worker then continues past the yellow goal which causes its intention estimation to fall. Shortly after this event, the brown goal is added in second scenario. Once the mobile robot blocks the shortest path towards the brown goal, the worker turns right. Because it is now also moving towards cyan and purple goal, the model cannot decide which goal worker desires the most. The worker continues moving towards the crossroad and hesitates with turning towards brown and cyan goals which manifests as a spike of estimation for purple goal. However, since worker eventually turns towards brown and cyan goals, purple goal's estimation steeply falls. The worker then continues moving towards the brown goal and the model recognizes its intention.



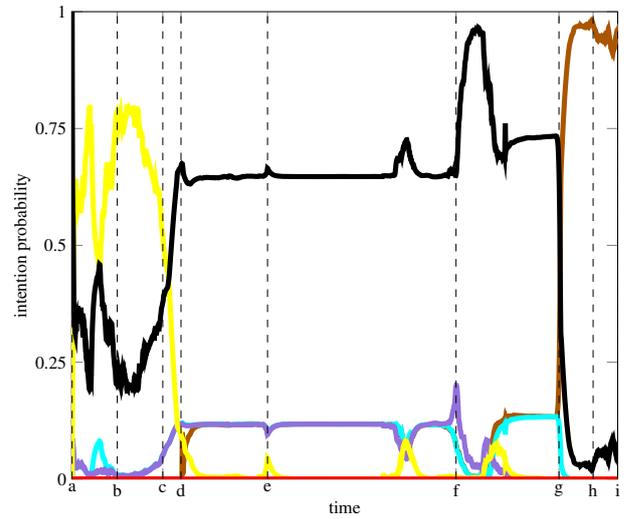
Figure 12: Key moments of the virtual reality generated scenario using larger warehouse visualized in *rviz*. Goal nodes are labeled with cubes and goal labeled with brown cube is being added during the experiment.

In the end, we have conducted another experiment on the large-scale VR test warehouse with 24 mobile robots. The robots were placed in groups of eight on the far left, middle and far right vertical corridors, equally spaced vertically. Each robot then selected one of the reachable adjacent nodes at random and continued selecting nodes in such a manner, except that it is not allowed to return to the node visited in the previous step. If another robot already selected the same node in the same time step, it stops and waits for it to pass. This works well at emulating a fleet of robots moving around the warehouse without collisions. The goal of this experiment was to show the

proposed algorithm's scalability with the respect to the number of mobile robots. We emphasize that mobile robot trajectories were not taking worker position into account. Given that, in some scenes the worker is moving in the proximity of a mobile robot carrying a warehouse rack, which would not meet safety requirements in a realistic flexible robotized warehouses. Nevertheless, we use this scenario solely for illustration purposes in order to demonstrate the proposed algorithm's scalability. The key moments of the experiment are shown in Fig. 14 and the result of intention estimation is shown in Fig. 15.



(a) Algorithm's output with brown goal being known through the whole experiment.



(b) Algorithm's output with brown goal being added during the experiment.

Figure 13: Comparison of the proposed algorithm's output in the larger VR warehouse between two different scenarios. Intention estimations for three locations are labeled with respect to their color in Fig. 12 (brown, cyan and purple), the unknown goal state is labeled black and the irrational worker state is labeled red. One can notice that the algorithm's outputs are different only on the segment where the brown goal has not been introduced in 13b.

## 6. Conclusion

In this paper we have proposed a real-time human intention estimation algorithm capable of a dynamic environment. Our goal was to estimate the intention of a human worker inside of a robotized warehouse whose layout can change due to robots blocking paths. We assumed that the worker position and orientation are measured and that the warehouse layout and robots positions are readily available. The worker has a set of potential goal locations that can be defined before or added during the experiment. The task of the proposed algorithm is to precisely estimate worker's desires for each goal. Given that, we evaluated worker actions with the respect to a modulated optimal path generated using the generalized Voronoi diagram and  $D^*$  algorithm. Then we used the resulting motion validation as observations of the hidden Markov model framework to estimate the final probabilities of worker intentions. We have carried out multiple experiments both in a real-world industrial setup using augmented reality glasses and in virtual reality generated warehouses in order to demonstrate the scalability of the algorithm. Results corroborate that the proposed framework estimates warehouse worker's desires precisely and within reasonable expectations.

For future work, the worker intention estimation is planned to be used for robot human-aware trajectories replanning to ensure more efficient operation of flexible robotized warehouses. Currently, the VR environment is used mostly for rapid prototyping and testing of AR interactions, while we are also looking into expanding it for training purposes and possibly system monitoring. Furthermore, the AR interactions were developed for the *Microsoft HoloLens* and built upon the standard use of AR in logistics as part of *pick-by-vision* systems and in the future it is envisaged to include navigational help, situational awareness and maintenance help for humans working in a mostly automated warehouse.

## Nomenclature

AR	Augmented reality
BToM	Bayesian Theory of Mind
FoV	Field of view
GHMM	Growing hidden Markov model
GVD	Generalized Voronoi diagram
HMM	Hidden Markov model
HRI	Human robot interaction
HUD	Heads-up display
MDP	Markov decision process
MR	Mixed reality
POMDP	Partially observable Markov decision process
SLAM	Simultaneous localization and mapping
VR	Virtual reality
<b>F</b>	Relative distances matrix
<b>I</b>	Isolation matrix
<b>c</b>	Association vector
<b>d</b>	Approximate distance vector
<b>D</b>	Approximate distance matrix
<b>v</b>	Motion validation vector
$G_i$	HMM hidden states
<b>T</b>	HMM transition matrix
$\alpha, \beta, \gamma, \delta$	Transition matrix parameters
<b>B</b>	HMM emission matrix
$\Pi$	HMM initial state

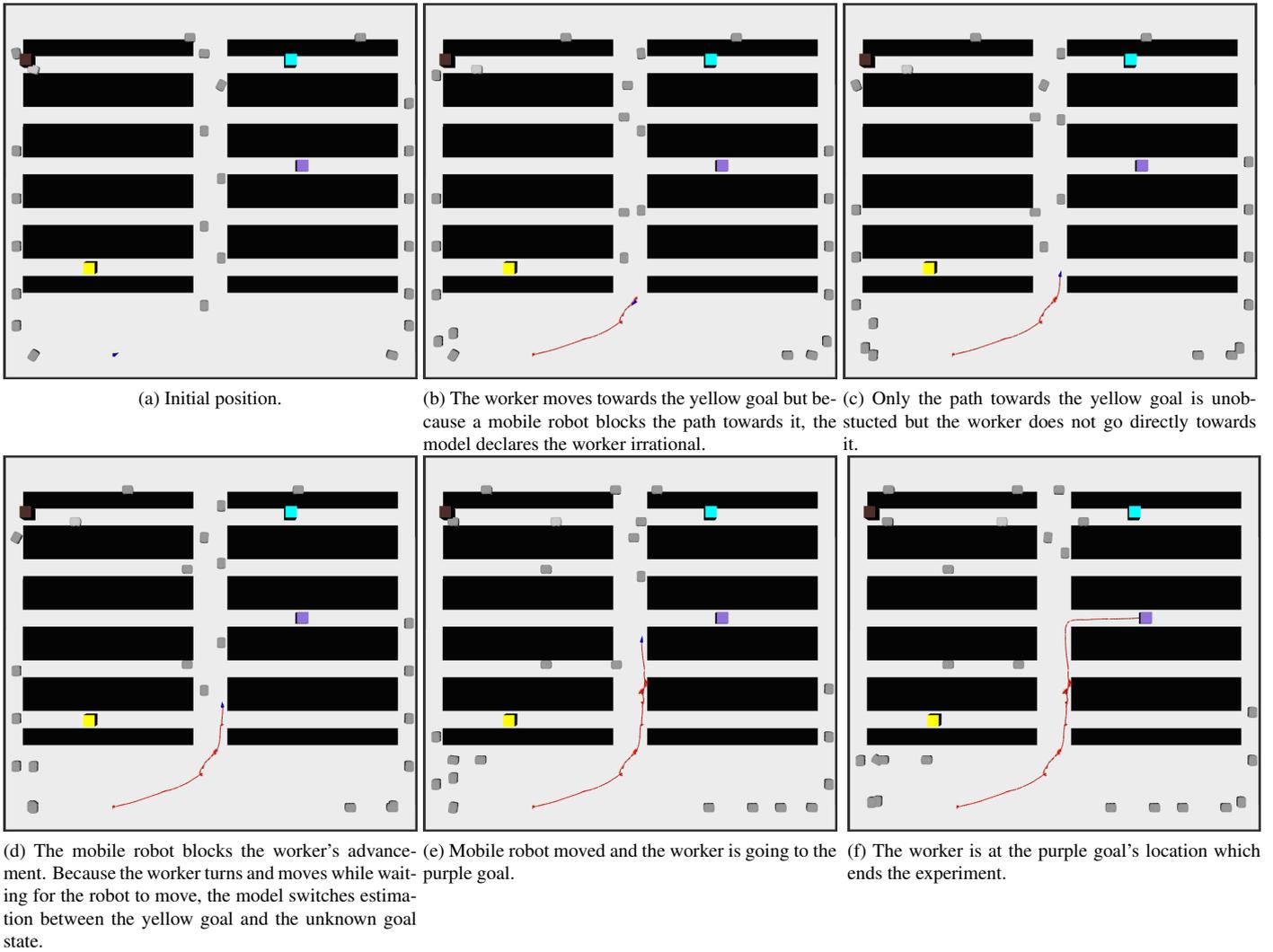


Figure 14: Key moments of the virtual reality generated scenario using a larger warehouse with 24 mobile robots visualized in *rviz*. Goal nodes are labeled with colored cubes.

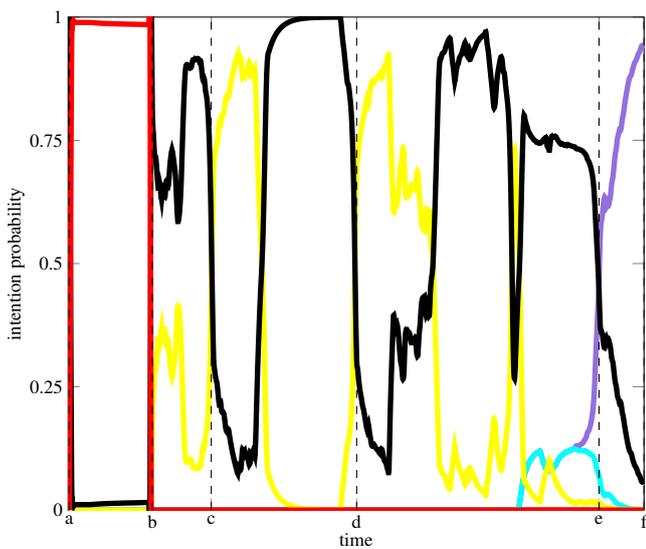


Figure 15: Proposed algorithm's output in VR scenario with 24 mobile robots. Intention estimations for goal locations are labeled with respect to their color in Fig. 14, the unknown goal state is labeled black and the irrational worker state is labeled red.

## Acknowledgment

This work has been supported from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688117 "Safe human-robot interaction in logistic applications for highly flexible warehouses (SafeLog)" and the DAAD/MZO grant "Human localization and intention recognition based on wearable sensors". The authors would like to thank Julian Schanowski of the Karlsruhe Institute of Technology for his help with the virtual reality experiments.

## References

- [1] R. D'Andrea, Guest editorial: A revolution in the warehouse: A retrospective on kiva systems and the grand challenges ahead, *IEEE Transactions on Automation Science and Engineering* 9 (4) (2012) 638–639.

- [2] P. R. Wurman, R. D'Andrea, M. Mountz, Coordinating Hundreds of Cooperative, Autonomous Vehicles in Warehouses, *AI Magazine* 29 (1) (2008) 9.
- [3] T. Bandyopadhyay, K. S. Won, E. Frazzoli, D. Hsu, W. S. Lee, D. Rus, Intention-aware Motion Planning, in: *Springer Tracts in Advanced Robotics*, Vol. 86, 2013, pp. 475–491.
- [4] C. Schlenoff, Z. Kootbally, A. Pietromartire, M. Franaszek, S. Foufou, Intention recognition in manufacturing applications, *Robotics and Computer-Integrated Manufacturing* 33 (2015) 29–41.
- [5] T. A. Han, L. M. Pereira, State-of-the-art of Intention Recognition and Its Use in Decision Making, *AI Communications* 26 (2) (2013) 237–246.
- [6] SwissLog, *Swisslog* (2018).  
URL <https://www.swisslog.com/>
- [7] T. Williams, D. Szafir, T. Chakraborti, H. Ben Amor, Virtual, Augmented, and Mixed Reality for Human-Robot Interaction, in: *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, HRI '18*, ACM, New York, NY, USA, 2018, pp. 403–404.
- [8] I. E. Sutherland, A Head-mounted Three Dimensional Display, in: *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I, AFIPS '68 (Fall, part I)*, ACM, New York, NY, USA, 1968, pp. 757–764.
- [9] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, J. C. Hart, The CAVE: Audio Visual Experience Automatic Virtual Environment, *Commun. ACM* 35 (6) (1992) 64–72.
- [10] A. Seth, J. M. Vance, J. H. Oliver, Virtual Reality for Assembly Methods Prototyping: a Review, *Virtual Reality* 15 (1) (2011) 5–20.
- [11] L. B. Rosenberg, The Use of Virtual Fixtures as Perceptual Overlays to Enhance Operator Performance in Remote Environments., *Tech. rep., Stanford Univ Ca Center for Design Research* (1992).
- [12] P. Milgram, S. Zhai, D. Drascic, J. Grodski, Applications of Augmented Reality for Human-robot Communication, in: *Intelligent Robots and Systems '93, IROS '93. Proceedings of the 1993 IEEE/RSJ International Conference on, Vol. 3, 1993*, pp. 1467–1472 vol.3.
- [13] R. Bischoff, A. Kazi, Perspectives on Augmented Reality Based Human-robot Interaction With Industrial Robots, in: *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on, Vol. 4, IEEE, 2004*, pp. 3226–3231.
- [14] E. Matsas, G.-C. Vosniakos, D. Batras, Prototyping proactive and adaptive techniques for human-robot collaboration in manufacturing using virtual reality, *Robotics and Computer-Integrated Manufacturing* 50 (2018) 168–180.
- [15] SafeLog, *Safelog-project* (2018).  
URL <http://safelog-project.eu/>
- [16] C. L. Baker, J. B. Tenenbaum, Modeling Human Plan Recognition Using Bayesian Theory of Mind, Plan, Activity, and Intent Recognition: Theory and Practice (2014) 177–204
- [17] H.-i. Lin, W.-k. Chen, Human Intention Recognition using Markov Decision Processes, *International Automatic Control Conference (CACSC)* (2014) 340–343
- [18] T.-H. D. Nguyen, D. Hsu, W.-S. Lee, T.-Y. Leong, L. P. Kaelbling, T. Lozano-Perez, A. H. Grant, CAPIR: Collaborative Action Planning with Intention Recognition, *Proceedings of the Seventh AAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-11)* (2011) 61–66.
- [19] A. Fern, P. Tadepalli, A Computational Decision Theory for Interactive Assistants, *Advances in Neural Information Processing Systems* 23 (NIPS) (2011) 577–585.
- [20] L. Jin, H. Hou, Y. Jiang, Driver Intention Recognition Based on Continuous Hidden Markov Model, *Transportation, Mechanical, and ...* 4 (3) (2011) 9–10.
- [21] C.-M. Huang, S. Andrist, A. Sauppé, B. Mutlu, Using Gaze Patterns to Predict Task Intent in Collaboration, *Frontiers in Psychology* 6 (2015) 1049.
- [22] D. Vasquez, T. Fraichard, C. Laugier, Growing Hidden Markov Models: An Incremental Tool for Learning and Predicting Human and Vehicle Motion, *The International Journal of Robotics Research* 28 (11-12) (2009) 1486–1506.
- [23] J. Rafferty, C. D. Nugent, J. Liu, L. Chen, From Activity Recognition to Intention Recognition for Assisted Living Within Smart Homes, *IEEE Transactions on Human-Machine Systems* 47 (3) (2017) 368–379.
- [24] Y. Long, Z.-j. Du, W.-d. Wang, W. Dong, Human motion intent learning based motion assistance control for a wearable exoskeleton, *Robotics and Computer-Integrated Manufacturing* 49 (2018) 317–327.
- [25] T. Taha, J. V. Miró, G. Dissanayake, POMDP-based Long-term User Intention Prediction for Wheelchair Navigation, *Proceedings - IEEE International Conference on Robotics and Automation (May)* (2008) 3920–3925.
- [26] J. I. Lipton, A. J. Fay, D. Rus, Baxter's Homunculus: Virtual Reality Spaces for Teleoperation in Manufacturing, *IEEE Robotics and Automation Letters* 3 (1) (2018) 179–186.
- [27] J. Jankowski, A. Grabowski, Usability Evaluation of VR Interface for Mobile Robot Teleoperation, *International Journal of Human-Computer Interaction* 31 (12) (2015) 882–889.
- [28] J. Allspaw, J. Roche, N. Lemiesz, M. Yannuzzi, H. A. Yanco, Teleoperating a Humanoid Robot With Virtual Reality, in: *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*, 2018.
- [29] J. Zhang, E. Langbehn, D. Krupke, N. Katzakis, F. Steinicke, Detection Thresholds for Rotation and Translation Gains in 360; Video-Based Telepresence Systems, *IEEE Transactions on Visualization and Computer Graphics* 24 (4) (2018) 1671–1680.
- [30] T. Zhang, Z. McCarthy, O. Jow, D. Lee, K. Goldberg, P. Abbeel, Deep Imitation Learning for Complex Manipulation Tasks from Virtual Reality Teleoperation, *CoRR* abs/1710.0.
- [31] T. S. Mujber, T. Szecsi, M. S. J. Hashmi, Virtual Reality Applications in Manufacturing Process Simulation, *Journal of Materials Processing Technology* 155-156 (2004) 1834–1838.
- [32] S. H. Choi, A. M. M. Chan, A Virtual Prototyping System for Rapid Product Development, *Computer-Aided Design* 36 (5) (2004) 401–412.
- [33] N. Gavish, T. Gutiérrez, S. Webel, J. Rodriguez, M. Peveri, U. Bockholt, F. Tecchia, Evaluating Virtual Reality and Augmented Reality Training for Industrial Maintenance and Assembly Tasks, *Interactive Learning Environments* 23 (6) (2015) 778–798.
- [34] N. E. Seymour, A. G. Gallagher, S. A. Roman, M. K. O'brien, V. K. Bansal, D. K. Andersen, R. M. Satava, Virtual Reality Training Improves Operating Room Performance: Results of a Randomized, Double-Blinded Study, *Annals of surgery* 236 (4) (2002) 458.
- [35] R. Sacks, A. Perlman, R. Barak, Construction Safety Training Using Immersive Virtual Reality, *Construction Management and Economics* 31 (9) (2013) 1005–1017.
- [36] A. P. Squelch, Virtual Reality for Mine Safety Training in South Africa, *Journal of the Southern African Institute of Mining and Metallurgy* 101 (4) (2001) 209–216.
- [37] A. Gaschler, M. Springer, M. Rickert, A. Knoll, Intuitive Robot Tasks With Augmented Reality and Virtual Obstacles, in: *Robotics and Automation (ICRA), 2014 IEEE International Conference on, IEEE, 2014*, pp. 6026–6031.
- [38] R. T. Chadalavada, H. Andreasson, R. Krug, A. J. Lilienthal, That's on My Mind! Robot to Human Intention Communication Through On-board Projection on Shared Floor Space, in: *2015 European Conference on Mobile Robots (ECMR)*, 2015, pp. 1–6.
- [39] S. Omidshafiei, A. A. Agha-Mohammadi, Y. F. Chen, N. K. Ure, S. Y. Liu, B. T. Lopez, R. Surati, J. P. How, J. Vian, Measurable Augmented Reality for Prototyping Cyberphysical Systems: A Robotics Platform to Aid the Hardware Prototyping and Performance Testing of Algorithms, *IEEE Control Systems* 36 (6) (2016) 65–87.
- [40] C. Mateo, A. Brunete, E. Gambao, M. Hernandez, Hammer: An Android Based Application For End-user Industrial Robot Programming, in: *Mechatronic and Embedded Systems and Applications (MESA), 2014 IEEE/ASME 10th International Conference on, IEEE, 2014*, pp. 1–6.
- [41] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, J. J. Leonard, Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age, *IEEE Transactions on Robotics* 32 (6) (2016) 1309–1332.
- [42] D. Sprute, K. Tönnies, M. König, Virtual borders: Accurate definition of a mobile robot's workspace using a rgb-d google tango tablet, *arXiv preprint arXiv:1709.00954*.
- [43] B. Schwerdtfeger, R. Reif, W. A. Günthner, G. Klinker, Pick-by-vision: There is Something to Pick at the End of the Augmented Tunnel, *Virtual Reality* 15 (2) (2011) 213–223.
- [44] J. Guhl, S. Tung, J. Kruger, Concept and Architecture for Programming Industrial Robots Using Augmented Reality With Mobile Devices Like Microsoft HoloLens, in: *2017 22nd IEEE International Conference on*

- Emerging Technologies and Factory Automation (ETFAs), 2017, pp. 1–4.
- [45] T. Petković, I. Marković, I. Petrović, Human Intention Recognition in Flexible Robotized Warehouses Based on Markov Decision Processes, in: ROBOT 2017: Third Iberian Robotics Conference, Vol. 694, 2018.
- [46] S. Thrun, Probabilistic Robotics, Communications of the ACM 45 (3) (2002) 480.
- [47] L. R. Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition (1989).
- [48] F. Aurenhammer, Voronoi Diagrams—a Survey of a Fundamental Geometric Data Structure, ACM Computing Surveys 23 (3) (1991) 345–405.
- [49] H. Choset, J. Burdick, Sensor-Based Exploration: The Hierarchical Generalized Voronoi Graph, The International Journal of Robotics Research 19 (2) (2000) 96–125.
- [50] S. Thrun, Learning Occupancy Grid Maps with Forward Sensor Models, Autonomous Robots 15 (2) (2003) 111–127.
- [51] D. C. Dennett, The intentional stance, MIT press, 1989.
- [52] A. Stentz, Optimal and Efficient Path Planning for Partially Known Environments, Intelligent Unmanned Ground Vehicles (May) (1997) 203–220.
- [53] B. Lau, C. Sprunk, W. Burgard, Efficient grid-based spatial representations for robot navigation in dynamic environments, Robotics and Autonomous Systems 61 (10) (2013) 1116–1130.
- [54] K. Hengster-Movrić, S. Bogdan, I. Draganjac, Multi-agent Formation Control Based on Bell-shaped Potential Functions, Journal of Intelligent and Robotic Systems: Theory and Applications 58 (2) (2010) 165–189.
- [55] C. Chen, Y. Xu, J. Yang, Human Action Learning via Hidden Markov Model, IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 27 (1) (1997) 34–44.
- [56] G. D. Forney, The Viterbi Algorithm, Proceedings of the IEEE 61 (3) (1973) 268–278.
- [57] C. Zhu, Q. Cheng, W. Sheng, Human Intention Recognition in Smart Assisted Living Systems Using a Hierarchical Hidden Markov Model, Automation Science and Engineering, 2008. CASE 2008. IEEE International Conference on (2008) 253–258.
- [58] M. Kulich, R. Sushkov, Speed-up of Self-Organizing Networks for Routing Problems in a Polygonal Domain (688117) 5–6. arXiv:1707.09809v1.