Human Localization in Warehouse Environments based on a Wearable Camera Sensor Suite and Dynamic Ultra-Wide Band Nodes

Goran Popović, Igor Cvišić, Ivan Marković, Ivan Petrović

***Abstract*— Modern warehouses, equipped with an autonomous robots fleet, suffer from an efficiency deficit during human interventions in the shop floor area, since, during such interventions, all the robots must stop for safety reasons. This particularly affects large warehouses which could extremely benefit from a solution that would allow human-robot collaboration during such interventions. This solution should only stop the robots that are in the vicinity of the human and reroute the others, and it requires the following prerequisites: (i) reliable relative ranging of robots with respect to human workers that would ensure safety, and (ii) reliable localization of human workers in the warehouse. In this paper we propose a system for reliable human localization that is based on fusing cues from measured relative ranges with respect to moving robots and a camera system that is worn by the human worker. Since global locations of the robots in the warehouse are known, we use the relative ranges to robots in an indoor GPS-like fashion to provide global human location updates, while the wearable cameras provide visual odometry, i.e., the relative motion of the human, and another global cue by detecting ground fiducial markers used by robots. The fusion is performed within a graph-based optimization framework that also includes a visual odometry error model. We carried out an experimental analysis on three different datasets to validate the performance of the proposed approach and compared it to a state-of-the-art visual SLAM solution, namely ORB-SLAM2. The results demonstrate that our approach yields reliable human location within warehouse like environments.**

I. Introduction

Robotized warehouses, where the fleet of autonomous robots is coordinated by a warehouse management system distributing repetitive and physically exhausting tasks, are becoming a standard in modern logistics. Productivity and flexibility of such warehouses are shown to be superior to the ones where humans perform all the tasks [1]. However, such warehouses strictly separate the robot shopfloor from human workers as the robots are not equipped with safety lasers to ensure worker safety. Once the human worker enters the shopfloor, all the robots must stop. Due to these stops, the efficiency of the warehouse is decreased [2] and a model for better human-robot collaboration in warehouse environments is needed.

Human-robot collaboration in manufacturing environments has been in focus of researchers for some time now. An overview of existing human-robot collaboration methods for a variety of industrial applications and safety assurance mechanisms can be found in [3]. In manufacturing, human-robot collaboration is based on safe interaction between the human operator and an industrial manipulator [2], [4]. Authors in [2] use a high-speed vision chip that triggers the emergency stop once the robot and human are too close. In [4] safe collaboration is assured with a set of stationary cameras that use a difference image method to prevent collision. A different type of collaboration is presented in [5], where the human and a mobile manipulator perform a heavy load transportation task. Besides physical safety, authors in [6] take into account the stress of the worker during collaboration with a robot. However, such a tight human-robot collaboration is not necessarily required in automated warehouses, and the aim of the Horizont 2020 project SafeLog[1] was to develop a safe human-robot collaboration framework for such systems. The key idea of the project was to assure safety during collaboration by stopping only the robots in the vicinity of the humans and reroute the other which might come close to the worker. In this way, the safety of the human is assured and the productivity of the warehouse is maintained, since the continuous information about the worker pose enables early rerouting of robots whose path traverses the worker area.

Indoor localization is still in the focus of many researchers since it is crucial for many applications. In the overview [7] authors classify localization systems into active and passive systems. Active systems, such as radiofrequency identification (RFID), ultra-wideband (UWB), and Bluetooth, require a carry-on device on the object being tracked. UWB technology is nowadays a popular tool for indoor positioning achieving centimeter-level accuracy. In recent years, researchers started recognizing the advantages of UWB sensors for indoor localization and began developing robust algorithms for complex indoor environments [8], [9], [10]. The UWB sensors are also commonly used in combination with other sensors and authors in [11] integrated UWB and an inertial navigation system in a factor graph to reduce the positioning errors of unmanned aerial vehicles in indoor environments, whereas authors in [12] used filtering techniques to fuse the information from IMU and UWB cues. Authors in [13] used a fusion of laser range finders and a network of ranging sensors to localize the robot in a changing environment with very accurate positioning near the docking stations. Except for localization purposes, the accuracy of the UWB ranging makes the UWB technology suitable for collision avoidance tasks in multi-robot envi-

All authors are with the Laboratory for Autonomous Systems and Mobile Robotics, University of Zagreb Faculty of Electrical Engineering and Computing, Zagreb, Croatia. <first name>.<last name>@fer.hr

This research has been supported by the European Regional Development Fund under the grant KK.01.1.1.01.0009 (DATACROSS).

[1] http://safelog-project.eu/

ronments, e.g., automated warehouses [14]. Passive systems for indoor localization, such as computer vision, do not require a stationary positioning system. Popular choice for passive localization is visual odometry and visual SLAM [15], [16]. The drawback of the odometry is the drift that accumulates over time and increases the uncertainty of the estimate. To take drift error into account, odometry uncertainty modeling has been studied in [17], [18], [19], where authors focused on finding analytical expressions for the Kalman filter covariance matrix representing the pose uncertainty. Analysis of visual odometry error propagation from the perspective of image features is given in [20], while in [21] authors used a convolutional network to learn the uncertainty of visual odometry. Although standard visual and visual-inertial SLAM solutions achieve impressive accuracy on public datasets like KITTI [22] and EuRoC [23], it can be challenging to directly apply them as such in our use case, since: (i) the static environment assumption of SLAM algorithms is not fulfilled in robotized warehouses where robots reorder racks continuously, (ii) visual aliasing, due to endless self-similar rows, increases the risk of wrong loop-closures, and (iii) all the processing power must be fitted in a wearable setup respecting ergonomic constraints.

In this paper we propose an approach for worker localization that fuses information from a wearable visual sensor suite and dynamic UWB nodes. Namely, the robots and the human worker are equipped with UWB nodes that perform relative ranging to ensure safety (when the robot gets too close it stops). Since the robots can localize themselves with absolute accuracy, we can use the relative ranges in an indoor GPS like manner to provide global cues about the human location. In our previous work [24], we presented a system based solely on visual odometry for relative motion estimation and a downward looking camera for absolute localization that detects ground fiducial markers used by the robots for localization. In this paper we complement this approach with UWB location cues and an error model for visual odometry. The error model makes the localization process aware of low-quality visual odometry input; thus, in conjuction with non-visual UWB cues we are making the localization framework more robust to visually challenging scenes. We carried out an experimental analysis on three different datasets to validate the performance of the proposed approach and compared it to a state-of-the-art visual SLAM solution, namely ORB-SLAM2. The results demonstrate that our approach yields reliable human location within warehouse like environments.

II. Proposed Human Localization Framework

The main idea of our approach is to avoid new infrastructure and reuse the existing one of the warehouse for human localization. The visual part of our method focuses on stereo visual odometry and detection of ground fiducial markers that robots use for localization. The camera setup has a high-resolution downward looking camera at the lower back of the human and the ground-marker detections are achieved using approach developed in [25]. Stereo visual odometry

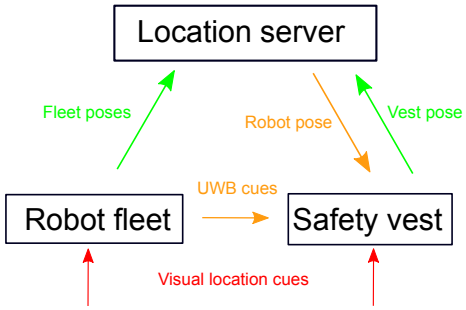

Fig. 1: Information flow in the proposed localization system. Red arrows show visual cues used for localization of the robots and the human worker. The *Location server* acts as an information broker and the worker localization algorithm can query the location of any robot at any time.

[15] estimates the relative poses between two ground-marker detections. However, during dynamic human motion or challenging scenes visual cues can get disrupted; hence, the localization could benefit from having another independent source of information – and this is where we leverage UWB. To guarantee safe collaboration, UWB nodes are installed on all the robots and the so-called safety vest that is worn by workers. The safety vest is also equipped with cameras and can communicate with the warehouse management system. The relative ranging of UWB nodes stops all the robots that come too close to the worker. The spatial information flow between the robot fleet, the safety vest, and the location server, which acts as an information broker, is shown in Fig. 1. Given that, through the Location server we can acquire all the necessary information to globally localize the worker by leveraging visual and UWB cues.

A. Localization with a dynamic UWB network

Unlike standard applications of localization with UWB nodes that use installed static anchors, in our case we have dynamic UWB anchors on robots whose positions serve as inputs for localization. Nevertheless, since the Location server keeps track of robots' poses, by querying the Location server we can obtain the poses of UWB anchors. In this paper, we compute the location of the human worker UWB node only when there is a sufficient number of UWB anchors. The fusion algorithm incorporates the location of the UWB node along with the visual location cues. The fusion algorithm avoids the UWB anchors' scalability problem by not depending only on the UWB anchors but also on the visual cues such as ground-marker localization and visual odometry.

A prerequisite for our localization, which is based on trilateration, is to have the coordinates of the anchors (x_k, y_k, z_k) and the range measurements from the anchors to the worker r_k. Generally, in 3D space, three anchors are sufficient to constrain the sought pose to 2 points, one of which is the true pose. We compute the human's pose (x_v, y_v, z_v) by solving a linear system of equations, but first we need to convert the following nonlinear system of equations

$$
\begin{aligned}
(x_1 - x_v)^2 + (y_1 - y_v)^2 + (z_1 - z_v)^2 &= r_1^2 \\
(x_2 - x_v)^2 + (y_2 - y_v)^2 + (z_2 - z_v)^2 &= r_2^2 \\
&\vdots \\
(x_n - x_v)^2 + (y_n - y_v)^2 + (z_n - z_v)^2 &= r_n^2
\end{aligned} \tag{1}
$$

to a linear one

$$
2 \begin{bmatrix} \Delta x_{1,n} & \Delta y_{1,n} & \Delta z_{1,n} \\ \Delta x_{2,n} & \Delta y_{2,n} & \Delta z_{2,n} \\ & \vdots & \\ \Delta x_{n-1,n} & \Delta y_{n-1,n} & \Delta z_{n-1,n} \end{bmatrix} \begin{bmatrix} x_v \\ y_v \\ z_v \end{bmatrix} = \begin{bmatrix} \Delta r_{1,n}^2 \\ \Delta r_{2,n}^2 \\ \vdots \\ \Delta r_{n-1,n}^2 \end{bmatrix}, \tag{2}
$$

where $\Delta A_{k,n} = A_k - A_n$ and $\Delta A_{k,n}^2 = A_k^2 - A_n^2$. To achieve this, we require one extra equation and, without loss of generality, we take the last equation from (1) and subtract it from all other equations, thereby removing the nonlinear members x_v^2, y_v^2, z_v^2. This gives us a linear system that can be solved with least squares as in [26].

B. Stereo odometry error modeling

To fuse data from all the sensors, we use a graph optimization framework as depicted in Fig. 2. Given that, in this section we aim to provide a measure of odometry uncertainty that will further increase reliability of the localization framework, so that there is relative distinction of odometry edges in our optimization graph. Green nodes and variables X_k represent relative localization estimates, e.g., coming from the stereo odometry. The orange nodes and variables G_k represent global localization estimates, e.g., coming from the UWB network trilateration or ground fiducial marker detection as in our previous work [24]. If odometry uncertainty is not modelled, then weights o_{xy} coming from the visual odometry would have the same value, meaning that the optimization process will equally believe the odometry pose estimate no matter the potential error. Improved performance is expected from an optimization process that would account for odometry uncertainty, translating to each edge in the graph in Fig. 2 weighted with the value inversely proportional to the odometry uncertainty.

In this paper we seek a simple solution that will not degrade the run-time performance of previously developed localization algorithm running on limited computational resources. Given that, our uncertainty model is based on two simple cues: (i) the magnitude of traversed distance and (ii) the number of features used for pose transformation. The first cue represents simply the increase of uncertainty as the human worker moves through the environment, while the second cue accounts for errors due to low-textured and badly-illuminated scenes where the odometry becomes error-prone.

1) Uncertainty due to traveled distance: We express this uncertainty source based on the fact that translational and rotational errors increase given their proportion in the motion and that initial rotational error affects the final translational error. Given that, our total error is a superposition of those three error sources. The translational error source e_{trans} rises

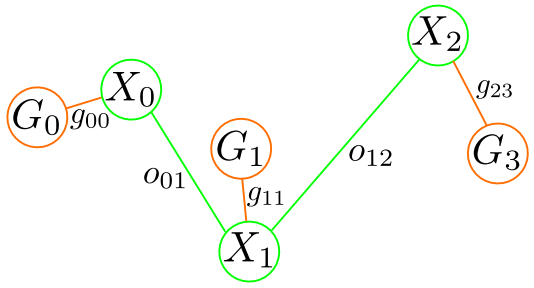

Fig. 2: Pose graph created with location cues from visual odometry (green nodes) and marker-detection or UWB ranging (orange nodes).

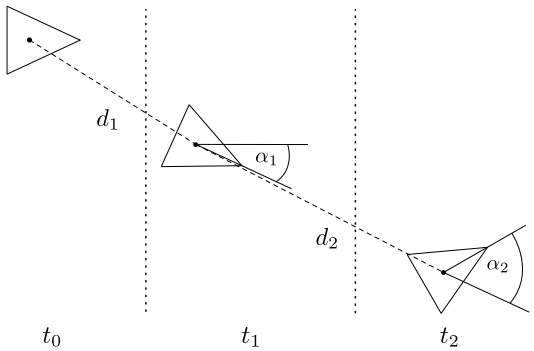

Fig. 3: A simplified 2D motion with marked differences in rotation and translation.

proportionally with the magnitude which we measure using the Euclidean distance. The rotational error source e_{rot} is proportional to the amount of rotation α expressed with

$$
\alpha_i = 2\cos^{-1}(\mathrm{Re}\{q_{i+1}q_i^{-1}\}) \tag{3}
$$

where quaternions q_i and q_{i+1} describe the orientation in two consecutive steps in the trajectory. The last error source e_{rt} comes from the initial rotation error affecting the succeeding translation error.

Now we formulate the visual odometry error model based on the aforementioned error types. In the first step t_1, the error is described with

$$
e_1 = \overbrace{k_1 d_1}^{e_{trans}} + \overbrace{k_2 \alpha_1 d_1}^{e_{rt}} + \overbrace{k_3 \alpha_1}^{e_{rot}} \tag{4}
$$

where d_1 is the amount of traveled distance and α_1 amount of rotation as depicted in Fig. 3. Factors k_1, k_2, k_3 map the rotation and translation magnitude to the odometry error. The second term $k_2\alpha_1 d_1$ corresponds to the error caused by the initial rotation error, which propagates to translation error as the human moves. In the next step, the error members are increased with the new distance d_2 and angle α_2:

$$
e_2 = k_1(d_1 + d_2) + k_2\left(\alpha_1(d_1 + d_2) + \alpha_2 d_2\right) + k_3(\alpha_1 + \alpha_2) \tag{5}
$$

In step t_n, the amount of error is given with:

$$
e_n = k_1 \sum_{i=1}^{n} d_i + k_2 \sum_{i=1}^{n} \alpha_i \sum_{j=i}^{n} d_j + k_3 \sum_{i=1}^{n} \alpha_i \tag{6}
$$

If we compute the error in a recursive manner the equation becomes:

$$e_n = k_1 e_n^1 + k_2 e_n^2 + k_3 e_n^3 \tag{7}$$

where $e_n^1 = e_{n-1}^1 + d_n$, $e_n^2 = e_{n-1}^2 + d_n e_n^3$, and $e_n^3 = e_{n-1}^3 + \alpha_n$. With the computed error e_n we compute the weight of the odometry edge $o_1^{n,n-1}$

$$o_1^{n,n-1} = max(O_{max} - e_n, O_{min}).$$

The weight is bounded by O_{max}, weight previously used for all odometry edges and O_{min}, a small positive value.

2) Uncertainty due to lack of features: The second uncertainty source comes from badly-illuminated or low-textured scenes which are common in warehouse environments. To account for this effect, we track the number of features in frames as the human moves through the environment. The uncertainty is modeled as follows:

$$\phi = \frac{sum\_of\_features}{sum\_of\_frames} \tag{8}$$

$$o_2^{n,n-1} = \min\left(O_{max}, \max\left(k_f \phi - O_{min}, O_{min}\right)\right) \tag{9}$$

where ϕ is the average number of features per frame between two pose nodes and O_{max} and O_{min} are limits of the odometry edge value. The final edge weight accounting for total uncertainty is expressed as a geometric mean:

$$o^{n,n-1} = \sqrt{o_1^{n,n-1} o_2^{n,n-1}} \tag{10}$$

III. Experimental Results

We recorded several datasets, since evaluation of the proposed approach required a specific infrastructure. The *Fraunhofer IML* and *Swisslog* datasets were introduced in our previous work [24], which is based on fusion of odometry with constant uncertainty and ground fiducial markers, while in this paper we also introduce a new *UNIZG-FER Library* dataset that was recorded at our Faculty's library. All the datasets are publically available[2]. The first two have only visual location cues and on these two datasets we evaluate our odometry reweighting model. The new dataset has both visual and UWB cues and we use it to test both the odometry reweighting and fusion with UWB relative ranging.

On all datasets, sequences consist of three main steps: (i) the initialization procedure during which the ground-marker detection algorithm synchronizes the coordinate systems between the visual odometry and the global coordinate system, (ii) operation phase when the worker enters the testing area and performs usual tasks, and (iii) exit procedure when the worker returns to the initial position. In the evaluation, we compare the following trajectories: the *F-S* trajectory, which is obtained by fusing visual odometry with no reweighting and ground markers, the *F-SU* trajectory, which is computed with visual odometry with no reweighting and UWB location cues, the *F-S*$_r$ and *F-SU*$_r$ that additionally to their counterparts use the odometry edges reweighting presented in Section II-B. We compared our localization algorithms with the state-of-the-art SLAM algorithm ORB-SLAM2. For evaluation we used the *absolute trajectory error* (ATE) metric [27].

[2]https://zenodo.org/communities/safelog/

TABLE I: The results for the *Swisslog* dataset. All values are expressed in meters.

ATE	F-S	F-S$_r$	ORB-SLAM2	distance
AG01	0.328	0.270	**0.128**	170.6
AG02	0.191	**0.177**	0.514	140.9
AG03	0.303	**0.130**	0.661	83.9
AG04	0.719	0.712	**0.532**	117.0

TABLE II: The results for the *Fraunhofer IML* dataset. All values are expressed in meters.

ATE	F-S	F-S$_r$	ORB-SLAM2	distance
DM01	**0.044**	0.049	0.120	24.0
DM02	0.098	0.104	**0.057**	32.4
DM03	0.107	0.065	**0.057**	20.4
DM04	0.098	0.084	**0.029**	22.5
DM05	0.104	0.125	**0.022**	25.8
DM06	0.072	0.078	**0.038**	20.6
DM07	0.051	0.050	**0.032**	20.3
DM08	0.066	0.073	**0.025**	25.0
DM09	0.091	0.100	**0.020**	18.0
DM01-DM05	0.185	**0.175**	0.550	125.1

A. Swisslog and Fraunhofer IML datasets

Performance of the localization based on stereo visual odometry and ground markers was conducted on the *Fraunhofer IML* and *Swisslog* datasets in our previous work [24]. However, since none of them contains UWB cues, herein we use them just to test the odometry edges reweighting. Tables I and II show the results for the Swisslog and Fraunhofer IML datasets, respectively. In Table I *F-S*$_r$ trajectory shows improved performance on all sequences, indicating that edge reweighting with the proposed uncertainty model can yield more accurate trajectories. Although the ORB-SLAM2 approach still has better performance on 2 sequences, the proposed improvement brought the trajectory error closer to that of ORB-SLAM2. Table II shows that the *F-S*$_r$ trajectory had similar performance to the *F-S* trajectory, which is mainly due to the fact that for this experiment sequences where shorter and larger errors did not manage to manifest themselves. As warehouses are non-static environments, and the warehouse's safety system did not allow live recording, we stacked multiple sequences (DM01-DM05) with different rack positions to simulate redistribution of racks during the operation. On most sequences ORB-SLAM2 showed better performance; however, when five sequences were stacked together the proposed method produced the best result.

B. UNIZG-FER Library dataset

Localization with UWB cues required recording of a new dataset that was additionally equipped with a network of UWB sensors. This dataset offers three location cues: visual odometry, ground-markers, and UWB nodes. It was recorded with a wearable sensor suite shown in Fig. 4. The sensor suite has an IMU-aided stereo camera PerceptIn Ironsides,

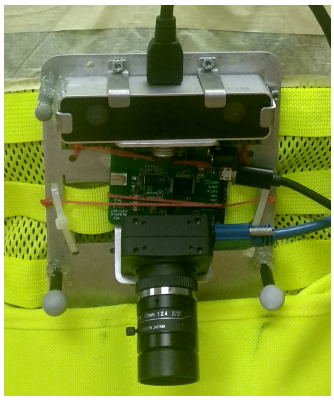

Fig. 4: The sensor suite mounted on the back of the human worker's vest. It is equipped with a stereo camera, monocular camera, UWB node, and motion capture markers.

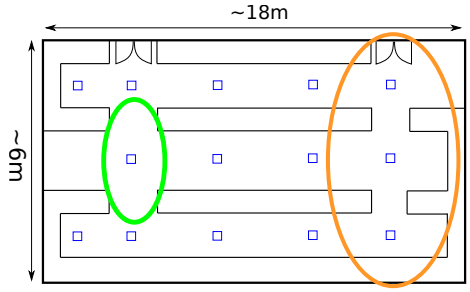

Fig. 5: The floorplan of a section of the UNIZG-FER library where experiments where recorded.

downward-looking monocular camera FLIR Chameleon3 CM3-U3-50S5M-CS with a Computar, 12mm, 2/3", 5 MP lens and Pozyx UWB node. The sensors are connected to an onboard computer Intel® Core™ i7-7700HQ CPU @ 2.80GHz × 8 which runs the localization algorithm. The intrinsic and extrinsic parameters of the camera setup were obtained through calibration with the Kalibr package [28], and the OpenCV library [29]. For the extrinsic parameters of the UWB node we used an approximation where we assumed that the position of the UWB node coincides with the position of the left camera in the stereo pair. The placement of the UWB node was near the stereo camera, as can be seen in Fig. 4, and the UWB ranging accuracy and the accuracy required for our use case justifies this approximation.

The library with its long corridors and metal shelves serves as a good environment for our use case since it provides similar conditions as expected in a warehouse. The floorplan is shown in Fig. 5, from which we can see that the area is approximately 6×18 m^2 and consists of three long corridors with two passages at the end of each corridor. In the testing area, we placed 14 ground markers, depicted as blue squares, equally distributed throughout the library. Five UWB anchors limited the area of UWB location cues to the part of the library marked with the orange ellipsoid and the anchors were placed to maximally cover the area, ensuring anchor's coordinate variety and line-of-sight between the human worker node and the anchors. The ground truth was acquired with the Optitrack motion capture system that,

TABLE III: The list of dominant movements per recording.

Recording	Movement type
FL01	recording of all markers
FL02	normal speed walk through the library
FL03	slow walk with occasional crunching
FL04	slow lateral walk
FL05	normal speed walk with loss of images in stereo camera
FL06	normal walk, dominantly around the UWB area
FL07	long normal speed walk
FL08	fast walk
FL09	fast walk and rotating during the walk
FL10	long fast walk
FL11	running through the library

TABLE IV: The results for the *UNIZG-FER Library* dataset. All values are expressed in meters.

ATE	F-S	F-SU	F-SU$_r$	ORB-SLAM2	distance
FL01	1.286	0.079	0.073	**0.052**	72.1
FL02	0.096	0.055	**0.048**	0.123	62.0
FL03	0.044	**0.036**	0.039	0.049	50.6
FL04	0.174	0.156	**0.109**	2.465	61.9
FL05	4.531	0.263	**0.193**	1.651	76.0
FL06	3.392	0.250	0.264	**0.065**	49.9
FL07	0.151	0.170	0.168	**0.041**	262.6
FL08	0.387	0.108	**0.096**	0.201	101.2
FL09	0.357	**0.200**	0.204	1.434	85.8
FL10	0.215	0.344	0.365	**0.138**	272.6
FL11	0.308	0.249	**0.188**	0.901	109.6

due to narrow corridors and library's ceiling-high shelves, constrained the availability of the ground truth to the area inside the green ellipsoid. We simulated a variety of human motions and the list per experiment is provided in Table III.

The results for the *UNIZG-FER Library* dataset are shown in Table IV. On 4 out of 11 sequences the ORB-SLAM2 trajectory had the lowest error. In these sequences, the dominant movement was relatively simple, but on the more demanding sequences the proposed solution showed to be more robust. The UWB enhanced trajectory optimization showed improved performance over the original solution based on ground markers and odometry. The biggest differences were in sequence FL05 where the loss of stereo images appeared and FL06 in which the worker spent most of the time walking in the area with UWB localizations cues. Reweighting odometry edges showed further performance improvement, scoring the least error in most of the sequences and being in the centimeters range from the best result in other sequences.

1) Dynamic UWB nodes scenario: The *UNIZG-FER Library* dataset has UWB anchors with fixed locations, which are used to determine the location of the UWB node on worker's vest. In the warehouse environment, UWB anchors will be placed on autonomous robots whose locations change with time. Given that, we performed experiments in our laboratory where we aimed to imitate the warehouse scenario with moving robots. In these experiments, we introduced 2 moving UWB anchors each attached to a moving platform, while additional 5 UWB anchors had fixed positions simulating static robots. With these experiments, we wanted to demonstrate that our localization fusion with UWBs works

TABLE V: Dynamic UWBs localization scenario. All values are expressed in meters.

ATE	F-S	F-SU	F-SU$_r$	F-SU$_{rs}$	ORB-SLAM2	distance
AC01	0.329	0.268	0.225	0.299	**0.076**	56.9
AC02	0.291	0.262	0.192	0.201	**0.148**	62.2

even when the positions of UWB anchors change with time. This scenario has two sequences, dubbed AC01 and AC02, where in the first one human wears the vest in open space, while in the second one the environment contains an obstacle representing a rack. To simulate robot localization we used the motion capture system to get the poses of the robots.

The experimental results are shown in Table V. On both sequences, ORB-SLAM2 scored the lowest error, and this result was expected due to small size of the environment and frequent loop closures. Furthermore, the trajectory error of the proposed solution decreased as we added more location cues to optimization. Following the ORB-SLAM2 trajectory, the *F-SU$_r$* trajectory had the second-lowest error. Trajectory computed with only 5 static UWB anchors, i.e., F-SU$_{rs}$ had higher error than the one which also included the dynamic nodes, i.e., *F-SU$_r$* and we can see the positive effect of adding even dynamic node information.

IV. Conclusion

In this paper, we have proposed a system for human localization in warehouse environments. The system is based on the fusion of location cues from a wearable camera setup and relative ranging between the robots and human worker. The relative distances along with the known robot poses are used to compute the absolute location of the worker. The camera setup uses visual odometry from the stereo camera to provide relative pose estimates, and ground-marker detections from the downward looking camera to compute absolute pose estimate. All the estimates are then fused within a graph-based optimization framework that also includes a straightforward model of the visual odometry uncertainty. The proposed system was evaluated on three different datasets (all made publicly available). On all datasets, the localization system showed comparable performance to ORB-SLAM2, and superior performance in challenging scenarios that are expected in warehouse environments.

References

[1] P. R. Wurman, R. D'Andrea, and M. Mountz, "Coordinating hundreds of cooperative, autonomous vehicles in warehouses," 2008.

[2] D. Ebert, T. Komuro, A. Namiki, and M. Ishikawa, "Safe human-robot-coexistence: emergency-stop using a high-speed vision-chip," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Aug 2005, pp. 2923–2928.

[3] Z. Bi, C. Luo, Z. Miao, B. Zhang, W. Zhang, and L. Wang, "Safety assurance mechanisms of collaborative robotic systems in manufacturing," *Robotics and Computer-Integrated Manufacturing*, vol. 67, p. 102022, 2021.

[4] T. Gecks and D. Henrich, "Human-robot cooperation: safe pick-and-place operations," in *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, 2005, pp. 549–554.

[5] A. K. Ramasubramanian and N. Papakostas, "Operator - mobile robot collaboration for synchronized part movement," *Procedia CIRP*, vol. 97, pp. 217–223, 2021, 8th CIRP Conference of Assembly Technology and Systems.

[6] Y. Niu, F. Schulte, and R. R. Negenborn, "Human aspects in collaborative order picking – letting robotic agents learn about human discomfort," *Procedia Computer Science*, vol. 180, pp. 877–886, 2021.

[7] G. Deak, K. Curran, and J. Condell, "A survey of active and passive indoor localisation systems," *Computer Communications*, vol. 35, no. 16, pp. 1939–1954, 2012.

[8] S. Djosic, I. Stojanovic, M. Jovanovic, T. Nikolic, and G. L. Djordjevic, "Fingerprinting-assisted uwb-based localization technique for complex indoor environments," *Expert Systems with Applications*, vol. 167, p. 114188, 2021.

[9] J. González, J. Blanco, C. Galindo, A. O. de Galisteo, J. Fernández-Madrigal, F. Moreno, and J. Martínez, "Mobile robot localization based on ultra-wide-band ranging: A particle filter approach," *Robotics and Autonomous Systems*, vol. 57, no. 5, pp. 496–507, 2009.

[10] W. Shule, C. M. Almansa, J. P. Queralta, Z. Zou, and T. Westerlund, "Uwb-based localization for multi-uav systems and collaborative heterogeneous multi-robot systems," *Procedia Computer Science*, vol. 175, pp. 357–364, 2020.

[11] Y. Song and L.-T. Hsu, "Tightly coupled integrated navigation system via factor graph for uav indoor localization," *Aerospace Science and Technology*, vol. 108, p. 106370, 2021.

[12] Y. Xu, C. K. Ahn, Y. S. Shmaliy, X. Chen, and Y. Li, "Adaptive robust ins/uwb-integrated human tracking using ufir filter bank," *Measurement*, vol. 123, pp. 1–7, 2018.

[13] C. Röhrig, C. Kirsch, J. Lategahn, M. Müller, and L. Telle, "Localization of autonomous mobile robots in a cellular transport system," 2012.

[14] S. Monica and G. Ferrari, "Low-complexity uwb-based collision avoidance system for automated guided vehicles," *ICT Express*, vol. 2, no. 2, pp. 53–56, 2016.

[15] I. Cvišić, J. Ćesić, I. Marković, and I. Petrović, "Soft-slam : Computationally efficient stereo visual slam for autonomous uavs," *Journal of Field Robotics*, 2017.

[16] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[17] L. Matthies and S. Shafer, "Error modeling in stereo navigation," *IEEE Journal on Robotics and Automation*, vol. 3, no. 3, pp. 239–248, 1987.

[18] A. Martinelli, "Modeling and estimating the odometry error of a mobile robot," *IFAC Proceedings Volumes*, 2001, 5th IFAC Symposium on Nonlinear Control Systems 2001, St Petersburg, Russia, 2001.

[19] C. Golban, P. Cobarzan, and S. Nedevschi, "Direct formulas for stereo-based visual odometry error modeling," in *2015 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, 2015, pp. 197–202.

[20] J. Fabian and G. M. Clayton, "Error analysis for visual odometry on indoor, wheeled mobile robots with 3-d sensors," *IEEE/ASME Transactions on Mechatronics*, vol. 19, no. 6, pp. 1896–1906, 2014.

[21] A. De Maio and S. Lacroix, "On learning visual odometry errors."

[22] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite."

[23] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The euroc micro aerial vehicle datasets," *The International Journal of Robotics Research*, 2016.

[24] G. Popović, I. Cvišić, G. Écorchard, I. Marković, L. Přeučil, and I. Petrović, "Human localization in robotized warehouses based on stereo odometry and ground-marker fusion," *Robotics and Computer-Integrated Manufacturing*, vol. 73, p. 102241, 2022.

[25] G. Écorchard, K. Kosnar, and L. Preucil, "Wearable camera-based human absolute localization in large warehouses," in *Twelfth International Conference on Machine Vision, ICMV 2019*.

[26] W. Hereman and S. Murphy Jr, "Determination of a position in three dimensions using trilateration and approximate distances," *Department of Mathematical and Computer Science (MCS), Colorado School of Mines, USA*, 1995.

[27] Z. Zhang and D. Scaramuzza, "A tutorial on quantitative trajectory evaluation for visual(-inertial) odometry," in *IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, 2018.

[28] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 1280–1286.

[29] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.