Depth-sensor–projector safety model for human-robot collaboration

Antti Hietanen[1], Roni-Jussi Halme[2], Jyrki Latokartano[2], Roel Pieters[3], Minna Lanz[2] and Joni-Kristian Kämäräinen[1]

Abstract— We propose a depth sensor and projector based safety model for human-robot collaboration in a compact shared workspace. The model consists of three spatial zones, *robot zone*, *danger zone* and *human zone*. The zones are online modelled, updated and monitored using a single depth sensor and user notification and interaction is provided by a projector-mirror display. Our model includes methods for detection of safety zone violations (an obstacle enters the danger zone) and prevents the robot to move to "unverified" (changed) workspace regions. Unverified regions are verified by user interaction. In the experimental part, we define an assembly task with the standard Cranfield benchmark parts where our proposed model reduces robot idle time by 43% and achieves 12.4% average reduction in task completion time as compared to a baseline.

I. INTRODUCTION

Human-robot interaction (HRI) for collaborative manufacturing requires special attention for HRI safety systems since heavy robots and payloads can lead to potentially dangerous situations. In addition to be effective and efficient, the HRI safety systems should be affordable and easy-to-install to be easily deployable. In this work, we focus on depth sensor and projector based safety in a shared workspace for collaborative manufacturing (HRI based assembly and disassembly) since this hardware is affordable, easy to install and re-configure.

Previous works on vision/depth-based human-robot safety have mainly focused on single safety components, such as collision avoidance [1], [2], [3] or collision injury minimization [4], [5], [6], but the role of each component is unclear until it is evaluated within a complete safety system. Recently, vision-only HRI safety systems have gained momentum in the industrial context [7], [8], [9]. In this work, we propose a HRI safety model for collaborative environments, specifically targeted for industrial manufacturing. For the model, we adopt the concept of workspace safety zones by Bdiwi et al. [7], [10] and we model them through captured point clouds (Figure 1). For interaction, we adopt the projector-mirror system in Vogel et al. [11], [12].

The novel contributions in this work are:

- A complete HRI safety model for collaborative manufacturing in a shared workspace S. The model is based on three zones (human, robot and danger) adapted from [7], [10] (Figure 1).
- Algorithms to detect danger zone violations and to detect and update changes in the shared workspace
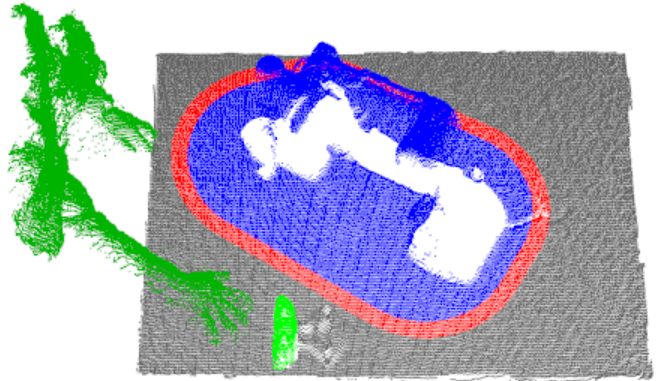
[1]Laboratory of Signal Processing, [2]Laboratory of Mechanical Engineering and Industrial Systems and [3]Laboratory of Automation and Hydraulic Engineering, Tampere University of Technology (TUT), Finland `First.Family@tut.fi`

Fig. 1. Illustration of our safety model. The shared human-robot collaboration workspace S is modelled as a point cloud captured by a depth sensor. There are two zones where robot and human can freely operate, the *robot zone* Z_r (blue) and the *human zone* Z_h, respectively. The two zones are separate by the third "border zone", *danger zone* Z_d (red), where any change causes immediate halt of the robot. During normal operation changes in Z_h are recorded as change regions R_i (green) and the robot cannot move to these regions before they are verified by a human co-worker.

automatically by the robot or manually through human verification.

- An experimental setup for HRI assembly task with Cranfield benchmark parts where the proposed and a baseline safety model are experimentally evaluated.

In the experimental part, we provide quantitative results for the proposed system and a baseline system.

II. PROPOSED FRAMEWORK AND METHODS

Our shared workspace model consists of three spatial zones: robot zone Z_r, danger zone Z_d and human zone Z_h, and methods to update and monitor the zones online.

A. Depth-based workspace model

We model the workspace S as a single $W \times H$ depth map image I_S. For convenience, the depth map coordinate frame is aligned with the robot coordinate frame. Therefore the depth map is an image that directly represents the 3D structure of a monitored workspace captured from the top. In our setting, the three zones (Figure 1) represent volume and volume changes (depth) are detected.

Changes in the workspace are detected by fast and simple element-wise subtraction:

$$I_\Delta = ||I_S - I|| \tag{1}$$

where I is the most recent depth data from the depth sensor re-projected to the robot frame and sampled in a regular grid

of the size $W \times H$ that matches the current workspace model I_S. The bins (pixels) in the difference image are thresholded by a depth threshold τ ($\tau = 10mm$ used in our experiments) that detect bins where substantial changes have occurred. Operation on the detected bins depends on each zone:

$$\forall \boldsymbol{x} \mid I_\Delta(\boldsymbol{x}) \geq \tau \begin{cases} \text{if } \boldsymbol{x} \in Z_d & HALT \\ \text{if } \boldsymbol{x} \in Z_r & I_S(\boldsymbol{x}) = I(\boldsymbol{x}) \\ \text{if } \boldsymbol{x} \in Z_h & M_h = \boldsymbol{0}, \ M_h(\boldsymbol{x}) = 1 \end{cases}. \quad (2)$$

a) Case 1: the change has occurred in the danger zone Z_d and therefore the robot must be immediately halted to avoid collision. For maximum safety this processing stage must be executed first and must test all pixels $\boldsymbol{x}$ before the next stages.

b) Case 2: the change has occurred in the robot working zone Z_r and is therefore caused by the robot itself by moving and/or manipulating objects and therefore the workspace model I_S can be safely and automatically updated.

c) Case 3: the change has occurred in the human safety zone Z_h and we create the mask M_h that represents the changed bins. The mask is re-computed for every depth frame to allow temporal changes. Robot can continue operation normally, but if the danger zone intersects with any changed bin in M_h, then the robot is halted and changed regions must be verified by a human. Again, no automation is adopted for maximum safety. The verified regions are added to the workspace model and operation continues.

In the shared workspace, the danger zone Z_d isolates a human co-worker working in the human zone Z_h and the robot operating in the robot zone Z_r by a spatial margin ω ($\omega = 20mm$ in the experiments) which size depends on the system reaction speed or the safety regulations in the industrial standard (ISO/TS 15066).

B. Coordinate transforms

The depth map model I_S of the workspace S in Section II-A is defined in a Cartesian coordinate frame that is aligned with the robot coordinate frame and is thus our *world frame*. A 3D point $\boldsymbol{p}$ in the world frame can be transformed to the depth sensor frame by 3D rotation and translation ($\boldsymbol{p}$ in homogeneous coordinates):

$$\boldsymbol{p}' = \mathcal{T}_s^w \boldsymbol{p} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{bmatrix} \boldsymbol{p}. \quad (3)$$

In our experimental platform $\boldsymbol{R}$ and $\boldsymbol{t}$ were solved by standard calibration procedure using a checkerboard pattern. The calibration procedure also provided the intrinsic camera matrix $\boldsymbol{K}$ for the depth sensor. For simplified computation we omitted the small skew value s and the lens correction step and adopted the inverse mapping of the standard pinhole camera model and adapted it for the depth measurements of current capture $I = \{\boldsymbol{p}'\}_i = \{x, y, d\}_i$ (a depth image). Now, the point $\boldsymbol{p}$ in the world frame can be computed from (homogeneous coordinates):

$$\boldsymbol{p} = \mathcal{N}^{-1} \left(\boldsymbol{R}^T \boldsymbol{K}^{-1} \boldsymbol{p}' + \boldsymbol{t} \right) \quad (4)$$

where $\boldsymbol{K}$ is the depth camera intrinsics matrix

$$\boldsymbol{K} = \begin{bmatrix} f_x & 0 & 0 & o_x \\ 0 & f_y & 0 & o_y \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (5)$$

and $\mathcal{N}^{-1}$ is the inverse coordinate normalization function

$$\boldsymbol{p} = \mathcal{N}^{-1}(\boldsymbol{p}) = \begin{bmatrix} p_x p_z \\ p_y p_z \\ p_z \end{bmatrix} \quad (6)$$

where p_z is the depth value measured by the depth sensor.

By similar procedure we also calibrated the projector to the world (robot) frame using inverse camera calibration [13]. With the calibrated projector we can render images based on the bin locations of the model I_S.
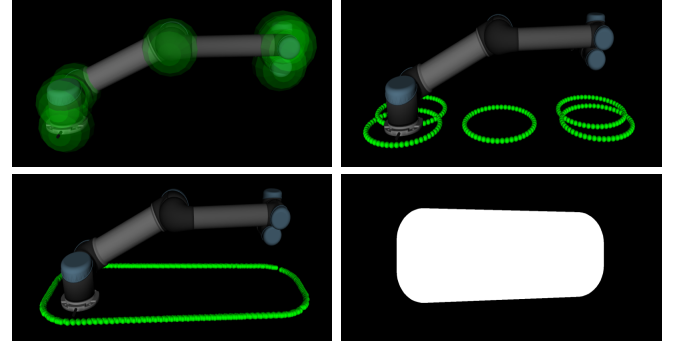
C. Robot zone construction

Fig. 2. Illustration of computation of the robot zone Z_r. Top-left: the selected 3D control points $\boldsymbol{\phi}_i$, i.e. robot joint locations, are provided by the robot controller (green spheres). Top-right: the control point x, y-coordinates map directly to the robot frame xy-plane - these are converted to regions by plotting circles of radius r. Bottom-left: [14] algorithm provides a convex hull around the circles. Bottom-right: convex hull is converted to a binary mask M_r that is mapped to the workspace image I_S dimensions.

Our zone modeling engine uses the 3D robot control points $\boldsymbol{\phi}_i$ for simulations (Figure 2). The control points can be read from the UR5 robot controller with the rate of 125 Hz. The robot joint locations provide sufficient information to cover all robot dimensions. Since the robot coordinate frame is also our world frame the points $\boldsymbol{\phi}_i$ can be directly mapped to the depth map I_S. In particular, we use only the x,y-coordinates that share the two axes of I_S, respectively. Around each coordinate pair we define a circle of radius r ($r = 250mm$ used in our experiments). Then we run the efficient convex hull algorithm of Graham et al. [14] that produces the actual robot zone Z_r in the same frame with I_S and can be efficiently rendered by a binary mask representation (Figure 2). All other zones can be generated from the known workspace S and the robot zone Z_r as illustrated in Figure 1.

III. EXPERIMENTS

A. A benchmarking task and evaluation

For comparing the proposed and the baseline model (Section III-C) we defined a suitable task to experiment safety in
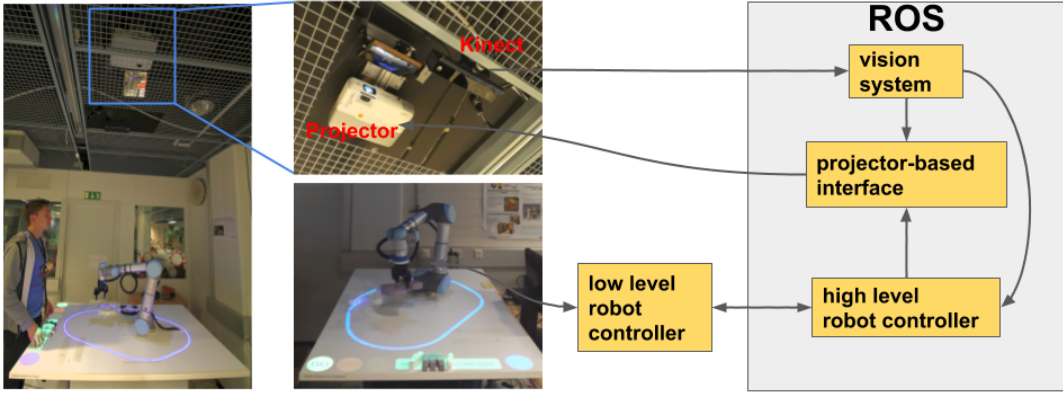
Fig. 3. Overall description of our task setup. A projector (Epson EB-905) with a mirror and a depth sensor (Kinect V2) are installed to the ceiling and pointing downwards toward the workspace. The workspace contains a flat surface (note that our model is not restricted to a flat surface) and a robotic arm (UR5) with Rototiq 85 gripper. Projector, robot and depth sensor are all connected to a single laptop computer that runs the Robot Operating System (ROS) on Ubuntu 16.04. and performs all computing. Video is available at https://youtu.be/YLF_oVbImPM.
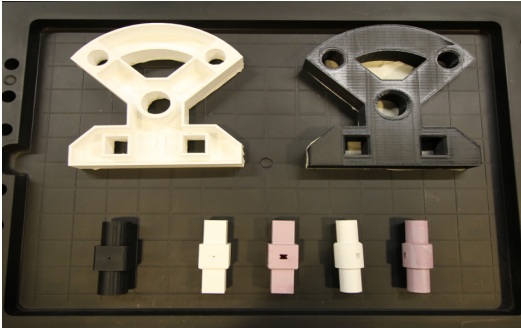

Fig. 4. Cranfield benchmark [15] parts selected for our evaluation: the front and back plates (top row) and the five "screws" that connect the plates (bottom row).

a shared workspace and to provide quantitative measures for comparisons. For this purpose we adopted the well-known *Cranfield benchmark* [15] that was designed for an assembly task. We selected 7 parts (Figure 4) and defined which assembly stages are made by a robot and which by a human. The task was the same in all experiments and consisted of the following steps:

1) A human operator enters to the workspace and starts the experiment by pressing a "START" button.
2) The robot brings the back plate to the shared workspace and goes back to collect the front plate. The human co-worker starts inserting the screws into the back plate.
3) *Proposed* – Any new item (including the screws) entering the workspace creates an "unverified region" R_i which needs to be explicitly verified by pressing the "CONFIRM". If Z_d and R_i intersect, the robot stops. *Baseline* – The robot is halted using a "STOP" button or by violating the safety line before inserting the screws.
4) If necessary the robot is restarted and the robot inserts the front plate and the task is finished.

For task performance evaluation we selected two different metrics: the *total assembly completion time* and the *robot idle time*. The total assembly time measures how long it takes for

a co-worker to finish the experiment. The idle time measures how long during the experiment the robot was doing nothing.

B. Setup and configuration

The overall description of our benchmark system is shown in Fig. 3 which includes the workspace, hardware components and the main software interfaces. The workspace was captured by the depth sensor which was installed at the ceiling perpendicular to the workspace and overseeing both the robot and a human co-worker.

A standard 3LCD projector was installed to the ceiling and used to display the user interface and operational information. The projector outputs a 1920×1080 color projection image with 50 Hz frame rate. Due to the short distance from the ceiling to the workspace we increased the physical projection size by installing a mirror in $45°$ angle to re-project the image to the workspace. Interaction with the UI components was provided with the depth sensor (setting a hand on an UI component is detected as a depth change).

The *ur_modern_driver* ROS package [16] was used to establish a ROS interface between the high and low level robot controllers. The package provides official drivers for the Universal Robot family and two different controller modes: velocity and position (used in the experiments) based control.

C. Baseline method

For model comparison we implemented a baseline method inspired by [11], [12] using an RGB camera and a projector. The method does not specifically model or update the workspace but assumes it is a planar table. The danger zone boundaries are projected to the workspace and safety violation is detected by comparing the projected boundary to a simulated boundary. The method is based on two different masks: *current-state mask* (measured by the RGB camera) and *expected-state mask* (simulated using robot movement and known workspace structure). The boundary

was simplified by using a single line (see video[1]) without compromising the method performance.

D. Results

TABLE I

AVERAGE TASK COMPLETION TIMES AND AVERAGE ROBOT IDLE TIMES AND THEIR STANDARD DEVIATIONS FOR THE COLLABORATIVE CRANFIELD BENCHMARK ($N = 21$).

	Tot. time [s]	Robot idle [s]
Baseline	46.82 ± 6.22	13.45 ± 6.21
Our	$\mathbf{40.98 \pm 4.26}$	$\mathbf{7.66 \pm 4.28}$
Improv.	12.4%	43.5%

Human-robot interaction experiments were conducted using 21 undergraduate students of mechanical engineering and automation with various backgrounds, but nobody had prior knowledge about the system or the task. Before recorded experiments each student was introduced to the task and the user interface and they were able to test the system. The order of the experiments with our or the baseline model was chosen randomly and the robot motion (speed) was identical in both experiments. Each participant completed the task twice with both models and the runs with the smallest idle times were included to our comparison.

The results of the experiments are shown in Table I. From these results it is obvious that our model of a dynamically updated and shared workspace achieves substantial improvement even in the simple assembly task. The total improvement is 12.4% in overall performance, but the difference is particularly evident in robot idle time where improvement was more than 40%. The main reason for the improvements is the fact that the shared dynamic workspace in our model allows parallel working in the same shared workspace. It is important to notice that our system does not compromise anything in safety but provides more flexible working than the baseline. The baseline method does not allow simultaneous operation since it cannot update the workspace model and therefore any changes are detected as violations.

IV. CONCLUSIONS

We proposed a HRI safety system for collaborative manufacturing that requires only a single depth sensor and a single projector. The proposed system provides a flexible workspace modelling where human and robot can alter workspace contents (manipulate new or existing objects) and still avoid working too close to each other (defined by the danger zone width). We conducted real experiments on a simple assembly task where the proposed model provided clear improvements as compared to a baseline method without depth sensing or the dynamic workspace model with safety zones.

---

[1] https://youtu.be/YLF_oVbImPM

REFERENCES

[1] L. Wang, B. Schmidt, and A. Y. Nee, "Vision-guided active collision avoidance for human-robot collaborations," *Manufacturing Letters*, vol. 1, no. 1, pp. 5–8, 2013.

[2] M. Saveriano and D. Lee, "Distance based dynamical system modulation for reactive avoidance of moving obstacles," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5618–5623, IEEE, 2014.

[3] A. Mohammed, B. Schmidt, and L. Wang, "Active collision avoidance for human–robot collaboration driven by vision sensors," *International Journal of Computer Integrated Manufacturing*, vol. 30, no. 9, pp. 970–980, 2017.

[4] S. Haddadin, A. Albu-Schaffer, A. De Luca, and G. Hirzinger, "Collision detection and reaction: A contribution to safe physical human-robot interaction," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3356–3363, IEEE, 2008.

[5] A. De Luca and F. Flacco, "Integrated control for phri: Collision avoidance, detection, reaction and collaboration," in *IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, pp. 288–295, IEEE, 2012.

[6] A. Cirillo, F. Ficuciello, C. Natale, S. Pirozzi, and L. Villani, "A conformable force/tactile skin for physical human–robot interaction," *IEEE Robotics and Automation Letters*, vol. 1, no. 1, pp. 41–48, 2016.

[7] M. Bdiwi, "Integrated sensors system for human safety during cooperating with industrial robots for handing-over and assembling tasks," *Procedia CIRP*, vol. 23, pp. 65–70, 2014.

[8] C. Morato, K. N. Kaipa, B. Zhao, and S. K. Gupta, "Toward safe human robot collaboration by using multiple kinects based real-time human tracking," *Journal of Computing and Information Science in Engineering*, vol. 14, no. 1, p. 011006, 2014.

[9] C. Vogel, C. Walter, and N. Elkmann, "Safeguarding and supporting future human-robot cooperative manufacturing processes by a projection-and camera-based technology," *Procedia Manufacturing*, vol. 11, pp. 39–46, 2017.

[10] M. Bdiwi, M. Pfeifer, and A. Sterzing, "A new strategy for ensuring human safety during various levels of interaction with industrial robots," *CIRP Annals*, vol. 66, no. 1, pp. 453–456, 2017.

[11] C. Vogel, M. Poggendorf, C. Walter, and N. Elkmann, "Towards safe physical human-robot collaboration: A projection-based safety system," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3355–3360, IEEE, 2011.

[12] C. Vogel, C. Walter, and N. Elkmann, "A projection-based sensor system for safe physical human-robot collaboration," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5359–5364, IEEE, 2013.

[13] I. Martynov, J.-K. Kamarainen, and L. Lensu, "Projector calibration by "inverse camera calibration"," in *Scandinavian Conference on Image Analysis (SCIA)*, 2011.

[14] R. L. Graham and F. F. Yao, "Finding the convex hull of a simple polygon," *Journal of Algorithms*, vol. 4, no. 4, pp. 324–331, 1983.

[15] K. Collins, A. Palmer, and K. Rathmill, *Robot Technology and Applications*, ch. The Development of a European Benchmark for the Comparison of Assembly Robot Programming Systems. 1985.

[16] T. T. Andersen, "Optimizing the universal robots ros driver.," tech. rep., Technical University of Denmark, Department of Electrical Engineering, 2015.