Human/Robot Multi-initiative Setups for Assembly Cells

Dan Gadensgaard
Department of Mechanical and Manufacturing Engineering
Aalborg University
Aalborg, Denmark
Email: dan.gadensgaard@gmail.com

David Alan Bourne
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA, USA
Email: db@ri.cmu.edu

***Abstract*—New products and small batch production entail a disproportionate amount of time is spent setting up automated assembly tasks. We have designed and implemented an automation tool to radically reduce this time. The setup for assembly automation involves: assembly planning, fixture-tool selection and positioning as well as part loading. This automation tool provides the robot with the ability to provide a human operator with precise and convenient instructions to follow through augmented reality, while at the same time allowing the robot to read information supplied by the human operator's actions. In this way, a complex setup task can be collaboratively executed, while allowing both the robot and the human to do what each does best.**

***Keywords*-Augmented Reality (AR), setup for assembly automation, Human-Robot Interaction (HRI), Robot-Human Interaction (RHI), active vision, multi-initiative tasks.**

I. INTRODUCTION

In today's modern manufacturing, there is an increasing demand for flexible, adaptable, yet efficient manufacturing systems. One of the key goals for achieving this is to ensure a rapid changeover between products, thus a reduction of the setup time is a necessity. Even in highly automated manufacturing systems, the setup of these processes are, for the most part, an expensive, time consuming and manual task performed by skilled operators.

The demand for costumer-product diversity along with the short product life-cycle has led to an increased focus on bringing the skills of human operators (flexibility, adaptability, etc.) together with the skills of robots (efficiency, repeatability, etc.). This emerging area of hybrid manufacturing systems has especially focused on Human-Robot Interaction (HRI) and Collaboration [1][2]. The work carried out in this area, covers a wide range of issues like effective teaching methods [3], ensuring human safety [4][5][6], communication [7][8], etc. However, there has been little focus on the initial setup of robotic work cells. We argue that this is an area that could also benefit from the principles of HRI, by using them as a tool for semi-automating parts of the setup process. Related to this we distinguish between systems where either the human has the initiative in the task (HRI) or the robot has the task initiative (Robot-Human Interaction (RHI)). This concept is foreshadowed in previous work [9][10].

In this paper, a concept for a new tool allowing for semi-automated multi-initiative setups is presented. This concept should provide a basis for improving the existing setup-processes by combining the precision of the robot with the judgment of a human operator.

The remainder of this paper is structured as follows: Section II discusses our approach to augmented reality applied to setups for robotic assembly. Section III illustrates the design and implementation of a new augmented reality tool, while section IV describes its general applications. Section V details two applications including guiding setups for robotic assemblies and part loading. Finally, sections VI and VII discuss future work and offer our conclusions.

II. BACKGROUND

In a multi-agent collaborative system, one agent typically has the initiative. In other words, that agent is leading the task step-by-step by marking the beginning and end of a step with spoken or gestural commands. In a multi-initiative system, it is expected that the agents can trade the lead role based on which agent has superior knowledge or which agent has a superior vantage point for the task at hand. In this case, we are describing a system with one human and one robot agent, while both agents share a common goal of building a setup for the robotic assembly cell.

A. Human-Robot communication and collaboration

In general, the human agent has more complete knowledge about the general environment (e.g., lighting conditions, objects not modeled by the robotic agent or the moment-by-moment positions of human agents and managing their safety). In addition, there may be system goals that are not explicitly known by the robotic system (e.g., optimizing a machine component that is not modeled, giving priority to human safety or knowledge about delicate equipment and how it can be best safeguarded). For all of these reasons and others, the human agent may decide to take the task initiative. The robot on the other hand has it's own clear-cut position of superiority. For example, the robot can accurately model its own motions and do collision checking with other modeled elements. The robot can map its positions easily into world coordinates so that precise

ISBN: 978-1-61208-134-2

positions can be displayed for the benefit of a cooperative human agent and in many cases the robot simply can ”see” a developing situation better from its unique vantage point. Based on the give and take of the specific task elements, it is appropriate to allow the initiative to easily change hands between the agents.

To allow for mixed-initiative collaboration between a human and a robot agent respectively, it is necessary to ensure that they reach a common understanding (i.e., ”grounding”), easily and confidently. In previous work this area has been investigated based on human-human interaction, where both audio, visual, and environmental cues, respectively, are used [1]. This indicates that a single channel of communication (e.g., speech) is insufficient, and a multi-channel approach (as seen in [4]) to human-robot communication and collaboration is necessary. Ideally, this means that the robot must have the capability to both understand and respond across the available communication channels. This work focuses on providing the robot with the ability to understand and respond to environmental cues in the workspace. The robot communicates by reading the environment and human intent by means of active vision, while the robot can send information to the human through Augmented Reality (AR).

B. Augmented Reality as a tool for communication

In an assembly task (or setting up for an assembly task) we must choose a communication mode most suitable to both the robot and human agent. This communication mode could be speech [10] as it often is between human agents, but the bandwidth for this modality is very low and noisy, along with the ”vocabulary problem” [8]. Instead we have conceived a system that can both easily display and ”see” visual information. To accomplish this, we have designed a tool, seen on Figure 1, that can augment reality with complex laser displays and at the same time capture these visual images. With this tool, laser-displays can be registered to the real world so that the projective displays can provide precise ”pointing data” as well as embedded information.

With augmented reality we are able to ”enhance the real world” by merging virtual objects with real objects. An example of this is seen in [11], where a projector-camera system is used to display information directly onto objects in an unconstrained environment. Implementing this kind of functionality into a robotic work cell has the advantage that the operator is able to maintain focus on the working environment while receiving information and instructions visually. Previous work has also shown that assembly instructions given to an untrained operator through AR result in a faster execution with fewer errors [12].

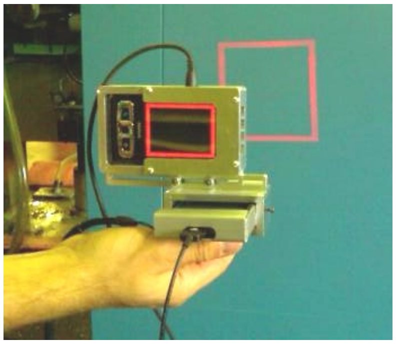


Figure 1. The augmented reality tool composed of a small laser-projector and a smartphone with a camera in a special fixture.

C. Active vision

To simplify the task of computer vision, the projected data can provide a way to see the structure of the environment via active vision [13]. The main idea behind active vision is that a simple, very bright pattern can be projected and then captured in an image. With simple image processing (e.g., thresholding followed by finding contours in the resulting binary image) it is possible to recover the distorted projections. Further, by analyzing the distortions of each projection, it is possible to directly infer the shapes of the 3D objects. In many ways, this is the holy grail of computer vision, but in the past the projection-setup has been bulky and awkward to use except for in the most constrained applications (e.g., single part inspection).

III. HARDWARE AND SETUP

The robotic hardware is as follows: an ABB IRB-140 6 axis robotic arm equipped with an Applied Robotics Smartgripper^{TM} electric gripper, a steel table as a foundation for varied configurations of fixtures (e.g., vise. clamps and compliant platforms), sensors (e.g., loadcells) and tools, which can be attached to the table with magnetic feet.

The augmented reality tool is constructed from: a Nokia N95 8G smartphone, a Microvision ShowWX Pico-P projector and a special fixture mount. The specific hardware used in this setup is just one of several possibilities, as other choices of hardware may be used. The smartphone requires several special features (i) programmable graphics (e.g., OpenGL used in this case) with, (ii) video output, (iii) wireless connectivity (e.g., Bluetooth and WiFi) and (iv) a camera with wireless video streaming. The projector needs to be laser based so that it is focus free and only the graphics are directly illuminated. The smartphone in turn streams data to a high-end PC suitable for doing real time computer vision and robot control.

There are two identical augmented reality tools situated at two distinct stations. The first station, seen in Figure 2, is attached to the robot gripper so that the robot can (i) use it as an intelligent pointer composed of a graphics window and (ii) use it for varied active vision tasks by pointing the camera at points of interest. The second station, seen in Figure 3, is fixed and can be used to analyze the robot, it’s gripper and the parts being held. In addition, there are tasks

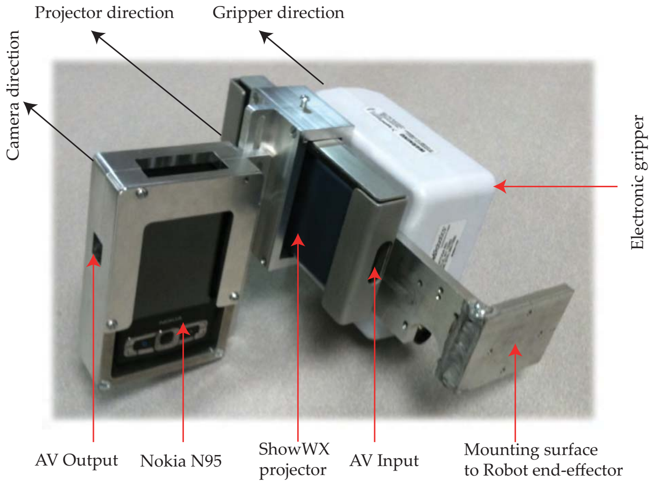


Figure 2. The AR tool assembled with the electric gripper for mounting on the robot end-effector.

that depend on projecting from one vantage point, while imaging from another (e.g., tasks involving triangulation) so for these tasks both stations work together.

There is a data command loop that operates between the workstation controller and the augmented reality tool. The typical steps for this command cycle are as follows:

1) *Workstation: Send* a high-level graphics command to the phone based on the application.
2) *Phone: Receive* draw command from the workstation.
3) *Phone: Draw* relevant pattern on phone's screen.
4) *Phone: Project* pattern on world.
5) *Phone: Snap* a picture of the resulting projection.
6) *Phone: Send* picture back to workstation.
7) *Workstation: Analyze* image and extract information.
8) *Workstation: Send* (optional) key information to phone - projector to be seen by human agent. This information can be registered on objects in the real world based on image analysis.

IV. APPLICATION AREAS FOR HRI IN SEMI-AUTOMATED SETUPS

There are many application areas for automated assembly where both human and robot agents can effectively cooperate. These include (i) cooperative setups where many elements are not suited to robot handling (ii) human training of assembly methods and sequences (iii) robotic programming by demonstrations (iv) and success or failure determination in a given assembly step [14]. In an attempt to capture all of these possible applications, we have composed a general framework, illustrated in Figure 4. If needed, the human agent is able to perform some offline planning of robot tasks, environment, etc. These plans are then ”mapped” to the robotic work cell where they are put into reality through the two AR stations described. Reversely, the robot may send information read from the environment through active vision back to the offline plan.

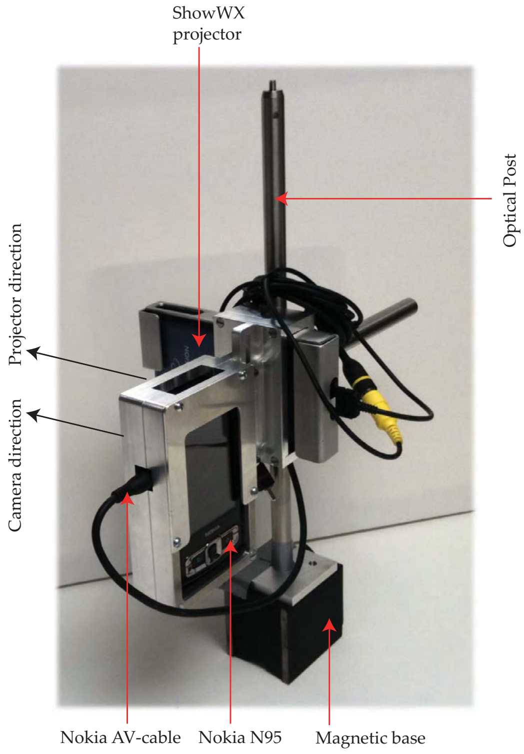


Figure 3. The AR tool mounted on a pair of optical posts for easy adjustments. The station can be conveniently placed by utilizing a magnetic base.

V. DETAILED APPLICATIONS

To demonstrate the general framework, we will present two applications that use multi-initiative execution of setups.

A. Positioning fixtures in robotic environment

The first application involves the precise positioning of magnetic feet on a steel table. These magnetic feet secure

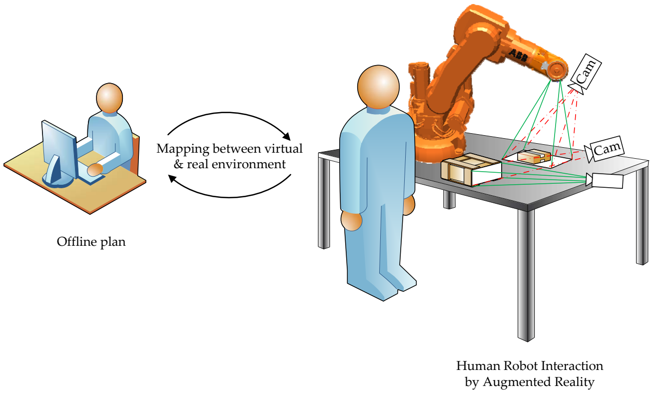


Figure 4. A general concept sketch for using augmented reality in partially automated assembly.

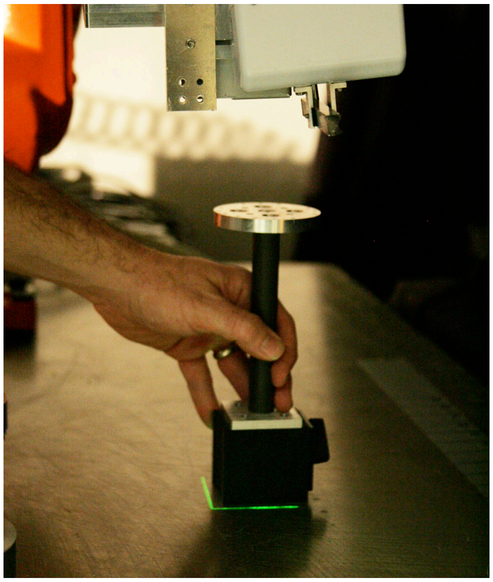


Figure 5. The AR tool marks a fixture's placement in green to aid the human operator.

a movable assembly platform on which assembled parts are held. In this application, planning software can locate an optimal location for the platform to avoid robot singularities and to provide collision free access to loading points. Since the robot controller knows the working coordinates and positions for all the fixturing elements, it can easily paint the environment with green laser marks to (i) request the human to load a fixture component and (ii) to provide a precise location. After the fixture elements have been positioned by the human-agent, the robot can then confirm the locations with the same system via active vision. The active vision algorithm projects light stripes at known locations and the fixture deforms the stripes into angles. The vertices of the angled laser stripes represent the transition between the surface of the table and the vertical wall of the magnetic foot. These positions can then be used to double check that the human operator has positioned the fixture correctly. Other researchers have shown [15] that more complex patterns (e.g., sinusoidal gradients) can be projected and in some cases it is possible to achieve better than sub-pixel resolutions, while reading object positions.

In this case, the robot has taken the initiative in the collaborative process of positioning fixtures. Alternatively, the human operator can ignore the robot's command and position the fixture in a fundamentally different place. In this case, the robot would have to discover the position and read the human's suggestion (again by active vision) of a better location. At this point, the planning system could change the planned fixture locations and run validity tests on that new set of locations. If the position is deemed feasible (e.g., no robot singularities prohibiting movement and collision free access to parts), then the plan could be altered in all of the specific details (i.e., part and fixture locations and robot motion plans). At this point, the robot would once again seize the task initiative and continue to give instructions of where additional fixtures need to be placed given the first position provided by the human. Thus, the time consuming process of relocating fixtures and other environmental components can be transformed into two phases: (i) initial conception of new approach/solution by human-agent and (ii) delegating the tedium of completing the task to the robotic agent.

B. *Loading parts in fixtures and robot gripper*

The second application helps the human-agent to precisely load a part into the robotic gripper during development to test, optimize, and/or train an assembly strategy. This entails that the operator is provided with information about (i) which part to feed to the robot and (ii) the orientation and location of the part in the robotic gripper. The second piece of information is especially important since the placement of the part might offer multiple solutions. For instance, the operator might choose to place the part upside-down in relation to the assembly strategy, causing the assembly operation to fail. In addition, the operator needs to place the parts at the same position in consecutive runs.

In this application, the information is provided by digitally marking the robotic gripper through AR by projection. The placement of the part is found by planning software, and then fed to the robot and an AR tool. By placing the robotic gripper in the AR tool's field of view, the tool detects the location of the gripper fingers through active vision and then projects a digital stripe marking the proper location of the part.

To properly detect the gripper fingers, successfully calculate the stripe location, and finally project this line onto the finger, it is necessary to find a relation between three

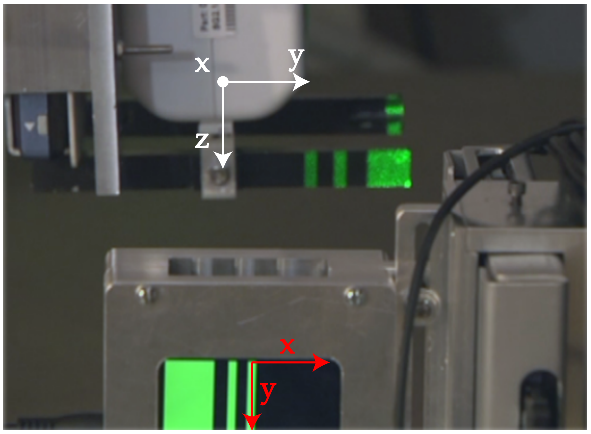
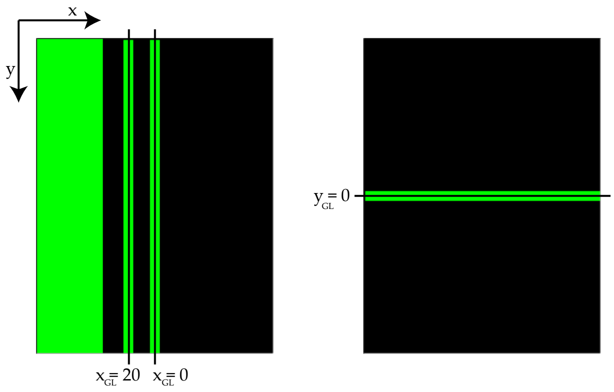


Figure 6. The robot and the projector coordinate frames are two out of the three frames that need to be correlated. The third coordinate frame is the incoming camera image seen in figure 8.



(a) (b)

Figure 7. These patterns are used to detect the gripper fingers and to correlate the three coordinate frames.

coordinate frames: (i) The robot frame and (ii) the projector image frame, as illustrated in Figure 6, and finally (iii) the camera image frame as seen Figure 8. To find this relation we utilize the patterns shown in Figure 7.

The relation between the camera image and projector image coordinate frames is found by relating the known pixel positions of the thin lines of the pattern in Figure 7(a) to the distance between the corresponding lines found in the camera image seen in Figure 8. Note that the projected pattern, seen in Figure 8, is rotated by 180 degrees as a result of the projector orientation. The relation to the robot coordinate frame is found by moving the gripper fingers into the path of the projected patterns in the horizontal (Figure 7(a)) and vertical (Figure 7(b)) directions respectively, until they are successfully detected in the camera image. Thus, the complete relation between the three coordinate frames are found in the following steps:

1) The pattern shown in Figure 7(a) is projected.
2) The robotic gripper is moved horizontally into the path of the projected pattern until the full pattern is recognized (as shown in Figure 8) in the resulting camera image.
3) The pattern is changed to the one shown in Figure 7(b), and step 2 is repeated in the vertical direction.
4) The resulting robot position is saved, and the stripe is calculated from two captured images showing the two projected patterns respectively, yielding the result shown in Figure 8.
5) A line is created based on the calculation and projected onto the gripper (left-most stripe in Figure 8).

Once the stripe has been successfully calculated and projected, the operator is able to load the part into the robotic gripper at the marked position.

In this application, the robot also takes the initiative, telling the human agent which part to load, and how to place it. As previously discussed, it might be possible for the operator to load the part in several different orientations and locations. This raises the possibility that the operator misinterprets the information, or choose to ignore it, and as a result loads the part incorrectly. Thus, the robot needs to have the ability to check that the part has been loaded correctly, and perform a suitable correcting action in case it is not. This could involve requesting the operator to correct the mistake, or, if possible, change the motion plan according to the loaded configuration of the part.

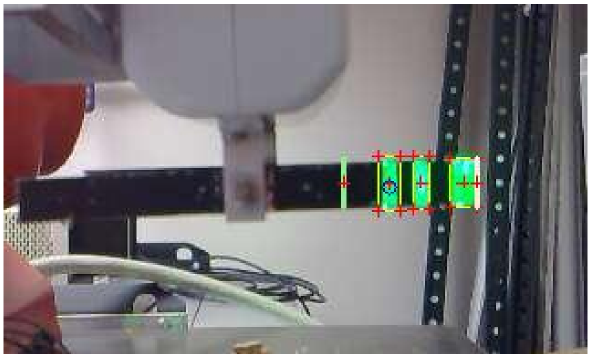


Figure 8. The result after image processing, which shows the calculated green stripe (left most) for part placement.

This application highlights a new design component for this approach; namely, we often must design different projected patterns for different applications. In this application, a 3 stripe pattern was chosen to simplify visual identification of the green stripes and its subsequent processing to determine positional information along the gripper fingers. We also have experimented with the mixing of primary colors to compose different information in the same space during projection. Once visual processing begins, the color channels can be separated and processed individually. The primary advantage of this approach is to allow one picture to encode several different messages simultaneously, while noting that sending pictures from the AR tool back to the workstation is a key processing bottleneck. Other well known patterns include a checkerboard pattern to calibrate the projection-camera system and sinusoidal wave patterns to achieve sub-pixel point cloud reconstructions. We fully expect that other new applications will demand new and creative projections to help manage and minimize the effect of characteristic limitations found in these applications.

VI. FUTURE WORK

We have discussed several applications with relatively simple steps performed by the human and robot agents. As the applications become more complex, it will be necessary to have dialog markers to determine when one step in an application ends and the next step begins. To seamlessly switch initiative between agents also requires clear dialog markers to avoid confusion or a time consuming end-of-step confirmation process. For example, in our first application we described a system where a person is directed to place a fixture at a given location, but they may choose to place it someplace entirely different. In that case, the robot would have to attempt to confirm the given placement and explicitly fail. This failure mode in itself is the dialog marker that allows the systems to swap initiative. However, this is not the most efficient method. One option might be to provide a simple visual marker (or coin) painted red on one side and green on the other. By flipping the coin, the human operator can either seize or relinquish the initiative. Of course, it is possible to do this process digitally (computer input), but that may require accessing a teach pendant or computer interface that would be relatively inconvenient. In addition, humans find speech to be the modality of choice especially when only dialog markers are required [7]. In the end, we need to find a method that does not impede progress in a given task and is more resistant to background noise.

The augmented reality tool we have described in this paper is ideal for tasks that require a narrow field of view. Unfortunately, that fails to address many applications that require a wide-field of view with less precision. For example, if we wished to monitor the movements of the human operators for safety analysis, then a different device such as the new Microsoft Kinect would be required.

Finally to achieve acceptance in the world of manufacturing, it is necessary to do a carefully controlled study comparing times and costs between these methods and the state-of-the-art.

VII. CONCLUSION

The rapid miniaturization of key components: full color laser projection, wireless computing and video streaming will make a wide array of augmented reality tasks feasible. By adding these functions to automated robotic systems, it also becomes apparent how humans and robots can collaborate on complex tasks without excessive programming overhead. This is particularly important in manufacturing where the time to setup machines is becoming a dominant bottleneck in production due to smaller and smaller batches and shorter product life-cycles.

ACKNOWLEDGMENT

We would like to thank Michael Dawson-Haggerty for his development work to guide the placement of fixtures. We would also like to thank Otto Moensteds Fond, Knud Hoejgaards Fond, Oticon Fonden, Vilhelm Pedersen & Hustrus Legat and Aalborg Stiftidenes Fond for the financial support for Dan Gadensgaard during the period of this project.

Finally, we would like to thank ABB Corporate Research for their partial support of the entire project. Thomas Fuhlbrigge (ABB) and Gregory Rossano (ABB) both provided significant feedback during the work.

REFERENCES

[1] S. A. Green, M. Billinghurst, X. Chen and J. Chase, ”Human-Robot Collaboration: A literature review and Augmented Reality approach in design,” *International Journal of Advanced Robotic Systems*, Volume 5, 2008, pp. 1-18.

[2] J. Krüger, T. K. Lien, and A. Verl, ”Cooperation of human and machines in assembly lines,” *CIRP Annuals - Manufacturing Technology*, Volume 58, 2009, pp. 628-646.

[3] F. Walhoff, J. Blume, A. Bannat, W. Rösel, C. Lenz, and A. Knoll, ”A skill based approach towards hybrid assembly,” *Advanced Engineering Informatics*, Volume 24, 2010, pp. 329-339.

[4] M. Zaeh and W. Roesel, ”Safety aspects in a Human-Robot Interaction scenario: A human worker is co-operating with an industrial robot,” in *Progress in Robotics*, Springer Berlin Heidelberg, 2009, pp. 53-62.

[5] J. A. Corrales, F. A. Candelas, and F. Torres, ”Safe Human-Robot Interaction based on dynamic sphere-swept line bounding volumes,” *Robotics and Computer-Integrated Manufacturing*, Volume 27, 2011, pp. 177-187.

[6] N. Lauzier and C. Gosselin, ”3-DOF cartesian force limiting device based on the delta architechture for safe physical Human-Robot Interaction,” in *IEEE International Conference on Robotics and Automation (ICRA 2010)*, May 3-8 2010, Anchorage, Alaska, USA, pp. 3420-3425.

[7] M. Lohse (2011, Jan. 6), ”The role of expectations in HRI,” in *AISB'08 Workshop on New Frontiers in Human-Robot Interaction*, 2009, Edinburgh, UK [Online]. Available: http://www.aisb.org.uk/convention/aisb09/Proceedings-/NEWFRONTIERS/FILES/LohseM.pdf

[8] G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais, ”The vocabulary problem in human-system communication,” *Commun. ACM*, Volume 30, 1987, pp. 964-971.

[9] J. Peltason, F. H. K. Siepmann, T. P. Spexard, B. Wrede, M. Hanheide, and E. A. Topp, ”Mixed-Initiative in human augmented mapping,” in *IEEE International Conference on Robotics and Automation (ICRA 09)*, May 12-17 2009, Kobe, Japan, pp. 2146 - 2153.

[10] I. Lütkebohle, J. Peltason, R. Haschke, B. Wrede, and S. Wachsmuth (2011, Jan. 8), ”The curious robot learns grasping in multi-modal interaction,” in *IEEE International Conference on Robotics and Automation (ICRA 2010)*, May 3-8, Anchorage, Alaska [Online]. Available: http://aiweb.techfak.uni-bielefeld.de/files/cr-video.pdf

[11] D. Molyneaux, H. Gellersen, G. Kortuem, and B. Schiele, ”Cooperative augmentation of smart objects with projector-camera systems,” *UbiComp 2007: Ubiquitous Computing*, Springer Berlin / Heidelberg, 2007, pp. 501-518.

[12] A. Tang, C. Owen, F. Biocca, and W. M. Mou, ”Performance evaluation of augmented reality for direct assembly,” in *Virtual and Augmented Reality Applications in Manufacturing*, Springer, 2004, pp. 311-331.

[13] B. Zhang, E. J. Gonzalez-Galvan, J. Batsche, and S. Skaar, ”Computer Vision” published by I-Tech, Vienna Austria, November 2008, pp. 111-123.

[14] A. Rodriguez, D. Bourne, M. Mason, G. Rossano, and J. Wang, ”Failure detection in assembly: Force signature analysis,” in *IEEE Conference on Automation Science and Engineering (CASE 2010)*, August 21-24, Toronto Canada, pp. 210-215.

[15] T. Peng and S. K. Gupta, ”Model and algorithms for point cloud construction using digital projection patterns”, *Transactions of ASME*, Volume 7, 2007, pp. 372-381.