



# Highly accessible platform technologies for vision-guided, closed-loop robotic assembly of unitized enclosure systems

Daniel Tish<sup>1,2</sup> · Nathan King<sup>1,2,3</sup> · Nicholas Cote<sup>1</sup>

Received: 15 December 2019 / Accepted: 17 April 2020 / Published online: 7 May 2020  
© The Author(s) 2020

## Abstract

Equipping robotics with the capacity to make decisions based on real-time information about their physical environment is crucial to the success of in situ robotics and offers many process advantages in prefabrication scenarios as well. However, the perceived technical challenges of producing environmentally-aware closed-loop robotics have limited their use in construction and design applications. To address this challenge, a low-cost and largely open-source computer vision-guided closed-loop robotic control paradigm is developed. The system is used here to identify construction materials in the workspace and calculate their position in space and determine their place in the facade panel assembly. The industrial robot arm is equipped with an RGB-depth camera in an eye-in-hand configuration to give control over the positioning of the camera for greatest accuracy. The control system relies on a simple TCP client/server connection between the robot and a central control computer to pass information and instructions from the computer vision system to the robot and vice versa. This setup delivers process flexibility, enabling pick-and-place procedures of the material positioned randomly within the workspace. In this work, the technologies are deployed in a factory-type setting but would also be necessary for any on-site robotic construction system, building towards an on-site robotics future. The final product of this research is a unitized spandrel panel wherein the vision-guided robot finds and places the insulation, cement board, and masonry cladding materials.

**Keywords** Computer vision · Closed-loop control · Depth camera · Pick-and-place · Low cost · Unitized spandrel panel

## 1 Introduction

Robotic fabrication has been the subject of intense research in architecture and engineering for the last decade. For design, it has enabled new modes of expression and the making of many exciting projects, prototypes, and pavilions. Poorly, however, are the realities of the construction industry addressed in these new robotic processes, and building construction remains a slow, expensive, imprecise, and largely manual process. While highly controlled, offsite manufacturing of construction materials and assemblies (e.g. prefabrication) has shown to improve precision and reduce costs, they often resist robotic

automation due to the complex and heterogeneous material flows required for their assembly. The growing accessibility of closed-loop robotic fabrication promises the industry a way to automate aspects of the construction process, either on- or off-site, that require feedback from the physical environment for their success. This feedback may come as information about the state of a task, the progress of an assembly, or the location of items or obstacles in the workspace, but critically allows the robotic fabrication to progress upon certain physical events or with parameters from the physical environment. This is in contrast to open-loop robotic processes which can only be executed according to their pre-programmed logic.

Simultaneously, advances in computer vision and their incorporation into open-source libraries have placed powerful image processing techniques within reach of non-experts in the field. Computer vision techniques hold advantages over mechanical means of providing feedback within the workspace as the camera can be programmed to execute a variety of tasks with high precision without extra hardware costs. Mechanical methods, such as sliding tables or hardware stops, typically require unique hardware and fixturing

---

✉ Daniel Tish  
daniel.tish@autodesk.com

<sup>1</sup> Autodesk Technology Centers, Autodesk Inc., San Rafael, CA, USA

<sup>2</sup> Graduate School of Design (GSD), Harvard University, Cambridge, MA, USA

<sup>3</sup> School of Architecture + Design, Center for Design Research, Virginia Tech, Blacksburg, VA, USA

for each material or process, which would require significant investment in the case of complex material assemblies such as façade systems.

There exists, therefore, the latent potential to combine these increasingly accessible closed-loop robotic and computer vision technologies into a flexible and easy-to-develop platform to enable environmentally aware robotic fabrication processes. This platform would be highly useful to construction industry experts who are keenly aware of the challenges of applying robotic fabrication to complex assembly processes in a factory-type setting or to the unstructured and chaotic nature of construction sites. This platform lowers the burden of developing custom computer vision workflows for experimental and practical applications such as these and many others. The research presented here assembles low-cost hardware components with open-source libraries to offer researchers and practitioners in design and construction a simple method of implementing closed-loop, vision-guided robotic processes.

## 1.1 Precedents

While robots have proven to be adept at many things, the ease with which humans can identify and grasp various objects in our environment has belied the complexity of pick and place operations in terms of robotics applications. Much of the early research in the fields of computer vision and robotics was conducted in the 1980s as automation worldwide boomed. Key algorithms from this period are still in use today as a part of the OpenCV library, including hand-in-eye calibration methods and shape approximation algorithms (Tsai and Lenz 1989; Douglas and Peucker 1973; Feddema and Mitchell 1989; Bradski 2000). In particular, the Japanese construction industry made a major push towards automation in the 1980s, which featured advanced laser feedback systems to provide quality control information. However, the complexity and constantly changing nature of construction sites made these solutions extremely costly and time-consuming to implement especially in relation to their relative benefits (Bechthold 2010).

The pick and place procedures that are at the heart of autonomous construction assembly processes have a clear affinity to the bin picking research aimed at automated manufacturing and packaging workflows. Ikeuchi first demonstrated the utilization of a three-dimensional (3D) CAD model to generate identifying features of a non-continuous 3D object at any angle in a bin (Ikeuchi 1987). More recent work has developed a novel multi-directional flash camera which uses flashes from multiple angles to generate robust depth-edge images to identify objects in a cluttered environment (Liu et al. 2012). This research also used a second camera to determine the pose of the object once it is in the gripper and calculate the error in the picking procedure, which is

useful for the precise placement of the object. Oh et al. developed a generic method to estimate the pose of objects in a bin by using feature detection on a two-dimensional image and then querying the feature points in three-dimensional space through stereo vision (Oh et al. 2012). Finally, Mahler et al. have developed a convolutional neural network to estimate grasp quality trained on a dataset of millions of point clouds representing everyday objects (Mahler et al. 2017).

As construction environments are typically chaotic and unstructured, there is a similar need for computer vision and other sensing regimes in construction robotics applications to provide the robot with real-time information about its physical surroundings. Giffthaler et al. report the usage of laser range finders and camera systems paired with fiducial markers to solve the localization of their mobile in situ fabricator platform (Giffthaler et al. 2017). The authors ultimately suggest that camera systems offer greater flexibility due to their ability to be tuned to different usages. In their work, the stereo camera rig mounted on the robot's end effector was able to verify the progress of the assembly and catch any deformations that occurred during fabrication. Feng et al. also approached the issue of on-site construction robotics by developing a vision-guided autonomous assembly system (Feng et al. 2015). The computer vision system used fiducial markers on the assembly components, in this case wood blocks, to identify them and calculate their position in space using a 3D camera.

Elashry and Glynn developed a Grasshopper plug-in which allowed for real-time feedback from a camera to be incorporated into a bricklaying procedure, taking on the challenge of automating the mortar-laying process using the camera feedback (Elashry and Glynn 2014). Vasey et al. developed a closed-loop feedback system to apply corrections to the process of robotic rod bending, and prototyped computer vision methods using stereo vision and laser range finders (Vasey et al. 2014). Jeffers assembled wood sticks of various lengths into an unplanned structure through the use of two sensory feedback systems, a mechanical limit switch and an IR depth camera (Jeffers 2016). Additionally, a closed-loop system was developed to feedback the dimensions of irregular materials from a 3D scanner into the fabrication of a timber envelope for a two-story structure (Eversmann 2018). Other research has developed techniques to 3D scan a collection of natural stone objects and autonomously determine their best stacking configuration and associated robotic pose planning (Furrer et al. 2017).

Computer vision techniques have also been used to identify the correct orientation of part geometries in irregular offcuts of stock material (Wu and Kilian 2016). Finally, recent work produced a workflow to compare point cloud information from an RGB-D camera to a Building Information Modeling (BIM) part database to correctly identify the assembly component and calculate the proper gripping pose for assembly (Dawod and Hanna 2019). This process

has the advantage of being able to identify complex three-dimensional objects positioned in any orientation, rather than modular parts, and can assemble them according to a pre-planned logic. However, in a number of the projects described here, the computer vision system was unable to deliver information with enough precision to ensure the success of the fabrication process, highlighting an area of necessary improvement moving forward.

With this context, the research presented here develops a computer vision aided robotic workflow for the assembly of prefabricated, materially heterogeneous construction panels. The goal of this research was to develop a computer vision system to identify and estimate the pose of construction materials in the workcell within the assembly tolerances of the construction industry without the use of fiducial markers or QR codes on the materials themselves. By developing a marker-less technology, the method saves time by eliminating the material labeling step and works with construction materials as they would be delivered to the site or factory. This is pursued by equipping the robot with an RGB-D camera and developing a control system for which the robot positions the camera strategically and information from the camera is processed and turned into pose targets for the robot. This process enables flexible robotic fabrication methods that are tested off-site but will aid the transition to on-site robotic assembly.

## 2 Methods

### 2.1 Computer vision

A computer vision system was developed to identify and estimate the pose of marker-less construction materials in the workcell. The end effector is fitted with an Intel RealSense D435 RGB-D camera. This depth camera uses active infrared stereo to construct three-dimensional point clouds of its field of view and includes a color sensor as well. Within 500 mm from the target, the accuracy of the depth sensor is less than 1 mm, making it precise enough for this task, although the error does increase exponentially to approximately 4 mm when the camera is placed 1 m away from the target. This accuracy combined with an affordable price under \$200 USD, small form factor and accessible SDK for the camera make it a logical choice for the development of the computer vision system.

#### 2.1.1 Eye-in-hand calibration and coordinate system transformation

When using computer vision with robots, one challenge rests on transforming the information from the camera into the robot's coordinate frame to use it to determine action targets. Using 3D information from an RGB-D camera,

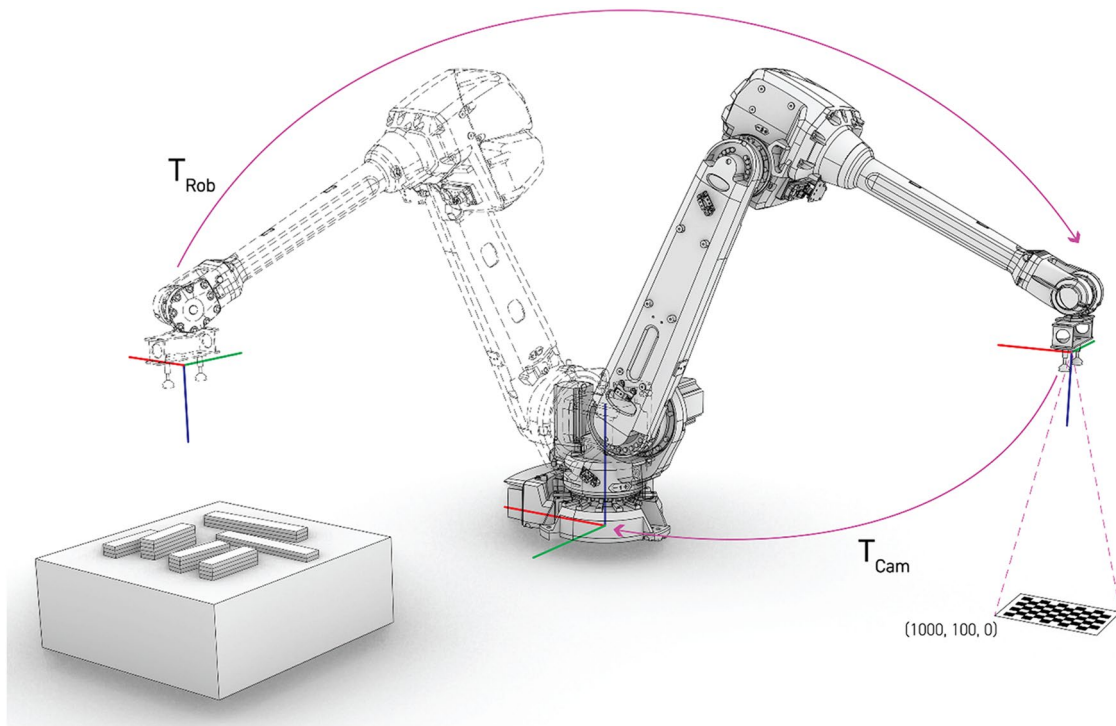
one can simply transform the camera's point cloud by the end-effector position and orientation to spatialize it in the robot's coordinate system. This method is error prone, however, as it requires the center point of the camera's imager to be known in relation to the tool center point with high precision, meaning that slight inaccuracies in the mounting position can compound into large errors. A more robust method is to calculate the transformation matrix based on a calibrated chessboard (Fig. 1). Using a set of vectors that describes the chessboard corners with respect to the camera frame and another with respect to the robot frame, the transformation can be calculated using the Kabsch algorithm, which finds the rotation and translation between two sets of vectors which minimizes the root mean squared deviation between the two (Kabsch 1976). The result may be applied to all 3D coordinates observed by the camera to spatialize them in the robot's coordinate system.

#### 2.1.2 Image segmentation

With the camera properly calibrated, the initial step for the material identification process is the image or point cloud segmentation. This process filters the incoming information to separate objects of interest from the background. To provide the most accurate estimation of the material's centroid, it is required that only the top face of the materials be returned from the image segmentation. To achieve the separation of material top face from side faces, a process was developed to segment the point cloud with respect to the robot coordinate space. By working from the transformed point cloud rather than the depth frame alone, the method proved to be accurate from all camera angles and for all objects. After filtering null pixels, each point with a Z-component in the 90th percentile of the current scene could be separated from the rest of the scene, then used to produce a thresholded image for object identification (Fig. 2).

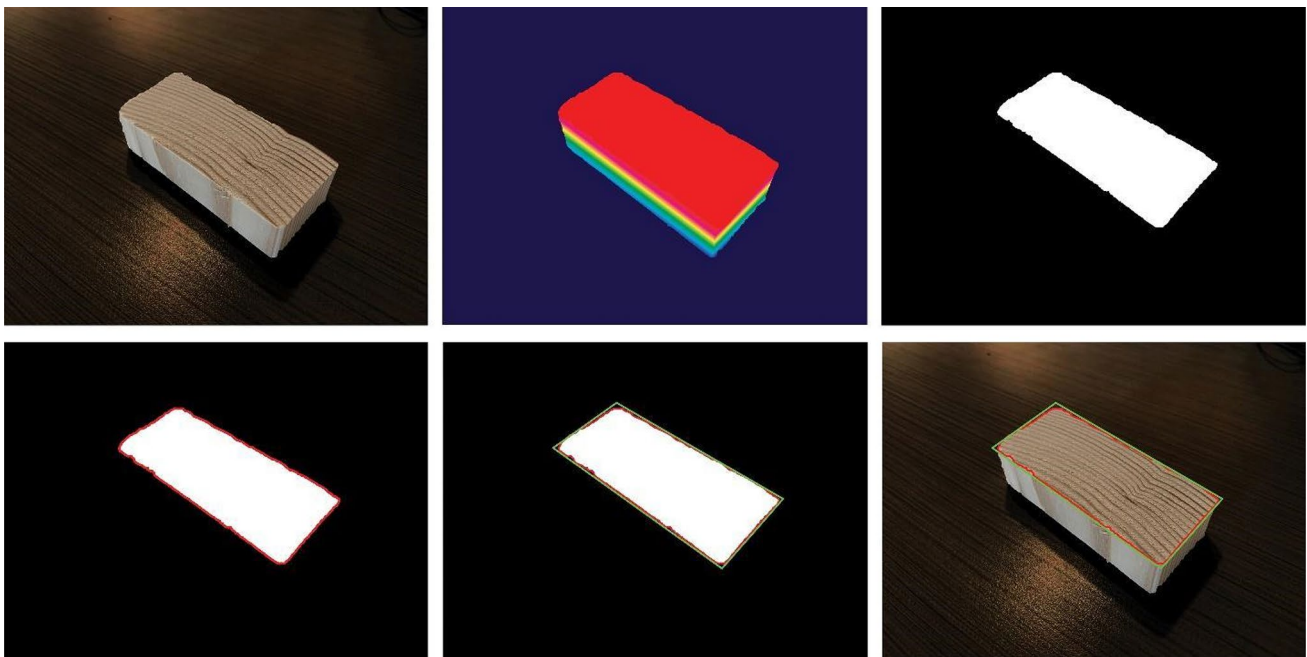
#### 2.1.3 Object identification

To detect the various material types in our assembly, a dictionary of the material types and their dimensions was constructed. As all of the construction materials in our assembly were standardized and are delivered in rectilinear forms, the dictionary simply holds the length and width dimensions of each object. Object identification starts by determining the image contours from the segmented image produced from the 3D point cloud thresholding operation. Using the OpenCV library, the shape of these contours is then approximated and checked to determine if the objects in view are mostly quadrilateral. If the object is quadrilateral, the corner points of the object are transformed from the camera frame to the robot frame.



**Fig. 1** Eye-in-hand calibration method with RGB-D camera assisted by chessboard grid of known points. Final transformation matrix is composed of the transformation between the camera and the robot

space from the calibration pose ( $T_{Cam}$ ) and the transformation from the current robot pose to the calibration pose ( $T_{Rob}$ )



**Fig. 2** Depiction of computer vision image segmentation and object detection workflow. Top, L to R: Base RGB image, transformed point cloud Z height map, thresholding by Z value. Bottom, L to R: Con-

tour of threshold image, object identification and rectangle reconstruction, final computer vision and RGB image composite



Two methods to serve as the final object detection layer were developed from this stage. The first compared the ratio of length to width of the observed object to the expected material dimensions. This method served as a quick and robust prototyping method but proved failure prone when partial objects appear in the camera's view that have the same ratio as the target objects. To fix this issue, the approximate width and length dimensions of the object as observed in the transformed robot coordinate space are compared directly to the expected dimensions of the current material type from the pre-stored dictionary. If the measured dimensions are within the allowable tolerance, the location of the object centroid and rotational deviation of the object from the robot's *XY* axes are recorded for pose estimation. As there may be many appropriate materials in the workspace, all the identified objects are collected from the scene, and the algorithm selects for assembly the object with the highest *Z* value in the space to avoid collisions with other materials.

### 3 Closed-loop robotic assembly

The research culminated in the robotic assembly of a unitized spandrel panel as a demonstrator. To fabricate the unitized spandrel panel system, the research developed many tools and workflows particular to this application. The main structural system of the unitized spandrel panel was a structural aluminum frame which was welded and bolted together manually; it also provided the base for subsequent fabrication processes. Additionally, a light gauge steel stud framing system was fabricated to specification by a Howick custom steel roll-forming machine and was sheathed in plywood. This assembly accepted the bolts from the structural frame to hold it in place and attach the robotically assembled elements to follow to the structural frame. While it was planned to coordinate the placement of this framing assembly with the robot, the tolerances on the bolt alignment proved to be too tight to accurately position with the robot and were done

manually. From this basis, rigid insulation panels, cement backer boards, and a bio-based finish cladding system were all positioned in place by the robot and affixed manually. The ABB IRB 4600 industrial robotic arm used here is mounted on a 9 m track, to move from the material loading area and the assembly area.

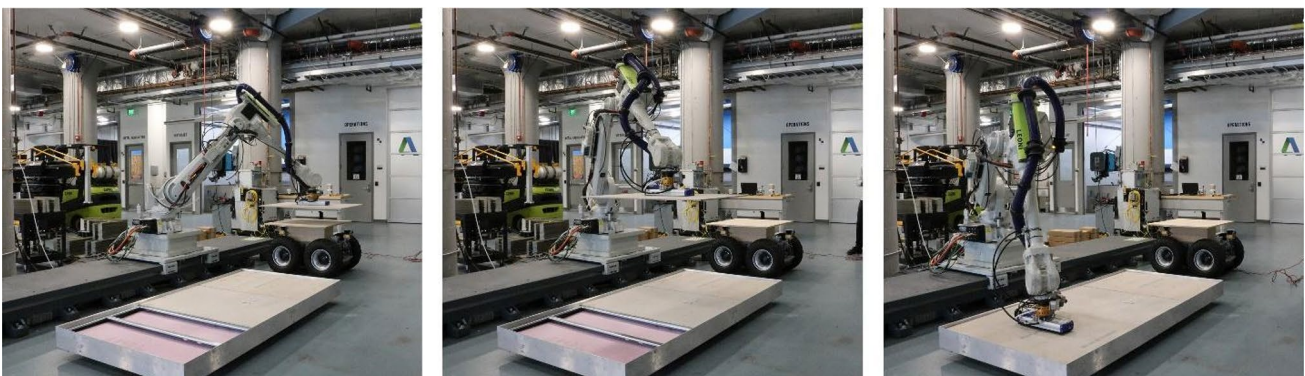
#### 3.1 End effectors

To facilitate the assembly of the panel and handle varied materials during that process, two different robot end effectors are used. The first uses a Schmalz vacuum area gripping system to handle the larger construction panels. This gripping system features internal vacuum generation and therefore only needs to be supplied pressurized air to function. This end effector is capable of lifting and positioning full 1.2 m × 2.4 m plywood sheets when mounted on the ABB IRB 4600 (Fig. 3).

The second tool is custom-built and sized for cladding elements that are placed in the assembly process and features two 40 mm suction cups fed by an in-line venturi vacuum generator each. These venturi generators are capable of producing enough vacuum pressure to overcome the pressure loss from the porosity of the cladding elements. This tool includes a mount for the Intel RealSense camera, positioned so that the depth camera imager is placed directly above the Tool Center Point, which is the midpoint between the two suction cups.

#### 3.2 Control flow

To drive the robot, we use a simple, off-the-shelf TCP client/server implementation (Dawson-Haggerty 2017). In practice, it allows for minimal and periodic communication between the two endpoints, does not require a continuous or high-frequency connection, and is easy to configure. The server executes on the robot and is written in RAPID. It listens for incoming instructions, executes a routine using



**Fig. 3** One sheet of cement backer board being placed in the spandrel panel assembly

the parameters included with those instructions, then replies with information about the success or failure of the routine. The client runs on a central PC and is written in Python. It sends outgoing instructions then listens for a reply.

The control system for the assembly process utilizes a finite state machine, which steps through a sequence of high-level routines: *Home*, *Calibrate*, *Pick*, *Place*, and, as needed, *Change Tool*. This state machine also coordinates the storage and passage of information between the computer vision and robotic processes (Fig. 4).

### 3.2.1 Home and calibrate routines

During the home routine, the robot moves to a known location in the material pickup zone and unwinds any rotations from previous sequences. During the calibrate routine, the robot maneuvers the camera over the calibrated chessboard pattern and runs the computer vision calibration process described above. The camera is calibrated before each pickup operation to ensure precise pose estimations, especially when the seventh axis track is in use.

### 3.2.2 Pick routine

The pick routine begins by positioning the camera about 1000 mm above the material pickup area to fit the entire platform in the field of view of the camera. The algorithm then performs the image segmentation based on the projected 3D point cloud of the scene in view. With the active material information taken from the global parameters of the state

machine, the object identification function then checks all of the objects that are present in the image segmentation to determine if they match the expected dimensions. From all of the positive matches in the scene, the algorithm selects the object with the highest Z coordinate as the target material (Fig. 5).

As the cladding material comes in two sizes, the object identification function looks for a match to either module size and then stores which material has been selected. Knowing the limitations of the camera’s accuracy at that distance, the robot is then positioned 450 mm directly above the target object to perform a more detailed target estimation using the same methods as above. During this verification step, the center point of the object and rotation of the object in the robot’s coordinate frame are calculated three times and averaged together for a more accurate result.

The robot then uses these target parameters to execute a pickup routine which slows the approach to the material and activates the vacuum generation. Each of the end effectors feature some form of dampening to allow for tolerance in the Z direction so that the robot is sure to make firm contact with the material without crushing it.

### 3.2.3 Place routine

With the proper assembly material now in hand, the place routine determines where to put the material in the assembly. As the rigid insulation panels and the cement backer boards are all the same size, this is handled by simply stepping through a list of pre-calculated positions for each

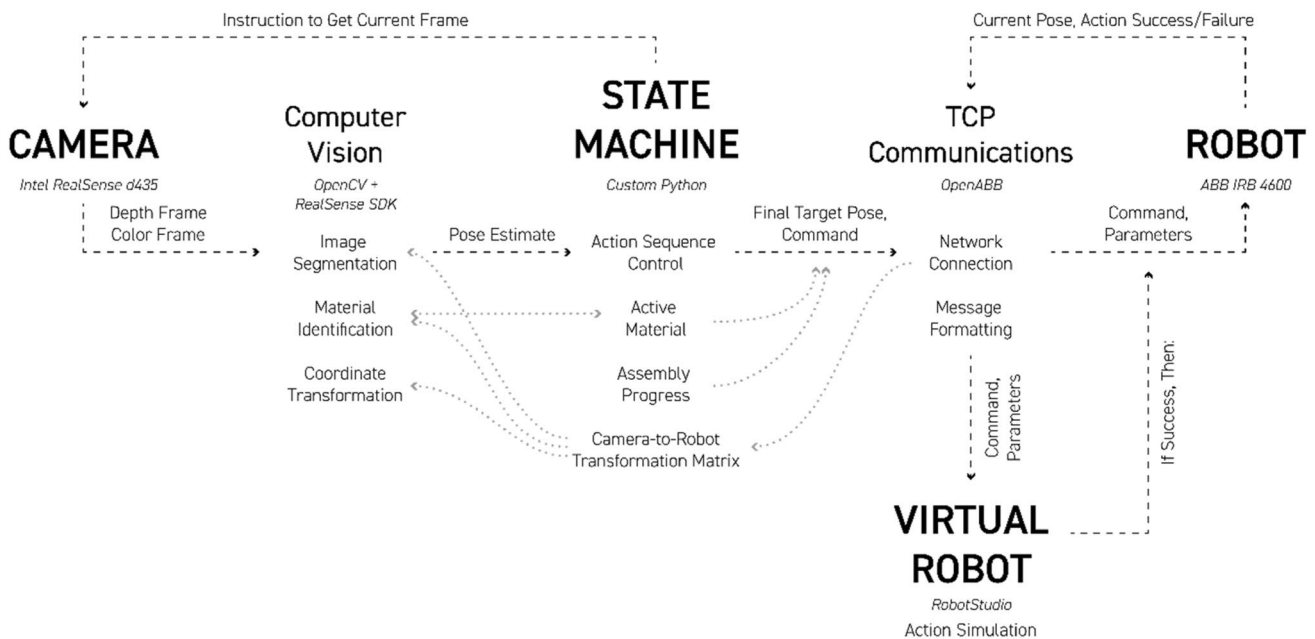
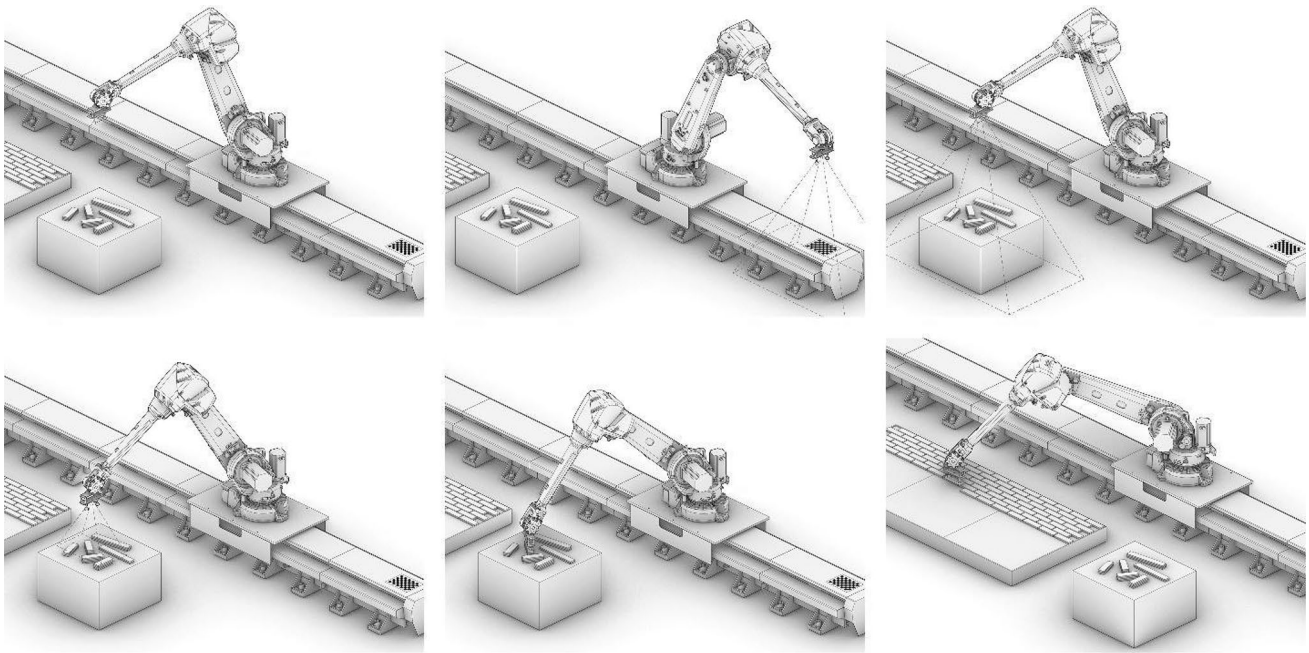


Fig. 4 Information flow diagram of the closed-loop control system



**Fig. 5** Vision-guided pick and place workflow. From L to R, top to bottom: 1. Home, 2. Calibrate camera and calculate transform matrix 3. Full workspace CV object identification 4. Hi-Res object identification verification and pose estimate 5. Pick with vacuum 6. Place in the assembly

type of panel. The robot takes this pose and runs a routine to slowly place the material on the assembly and deactivate the vacuum.

However, as the cladding material features two different module sizes, a more flexible system had to be developed. Since the pickup procedure does not give preference to one module over the other, the position that the object will be placed in needs to be calculated once the dimensions of the material in hand are known. The progress of the assembly is tracked through two global parameters, used as counters, to which the  $X$  and  $Y$  dimensions of the cladding modules are added cumulatively. The target point for the placement of the module in hand is then half of the dimension of that module plus the progress of the row so far. Nearing the end of a row, a check is made to determine if the longer module would not fit but the shorter module would. In that case, the picking procedure is constrained to only look for the shorter modules to complete the row.

This system thus generates quasi-random cladding patterns based on the vision system's choice of the material available (Fig. 6). It, therefore, suggests a shared design agency between the autonomous assembly system and the designer. After the placing procedure is finished, the control algorithm loops back to the Home state to start the process of picking and placing the next material in the assembly.

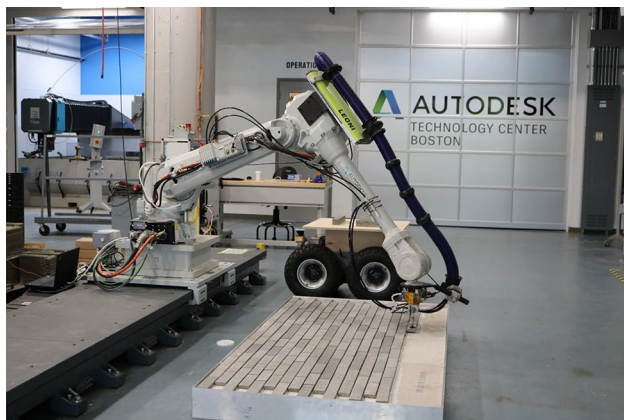


**Fig. 6** Placing a cladding module in the assembly. Placement targets are calculated on the fly based on the assembly progress thus far and the dimensions of the material in hand

## 4 Results

The low-cost and largely open-source computer vision system developed in this research proved to be accurate enough to assemble the panel within the given material tolerances, but certainly left room for improvement (Fig. 7). Previous studies have suggested that for assembly tasks in the building industry, the desired level of precision is between 2 and 3 mm positionally and  $0.1^\circ$  rotationally





**Fig. 7** Image of full panel being assembled during the cladding placement phase

(Shohet and Rosenfeld 1997). In the panel fabrication for this project, cladding materials are given a 5.25 mm gap between modules for the mortar between them. Therefore, the system could be lacking the desired precision but still avoid collisions between objects on the panel, just risking differences in mortar spacing.

During the cladding assembly process, no method of physically measuring the positional or rotational error was developed. However, the result of the computer vision object identification is saved for each object, which can be visually analyzed to determine the error in the computer vision decision-making. This was done by manually tracing the outline of the material in the color image taken by the camera and comparing the object centroid and rotation between the computer vision and manual processes. These results show that, during assembly, the computer vision system operated with an average positional error of 1.77 mm from the center point of the material, and an average rotational error of  $0.15^\circ$ . Due to inaccuracies in the image segmentation process and the presence of null pixels near the edges in the depth image, the computer vision system consistently misjudged the upper edge of the short modules, leading to a recurrent error along the long axis of the module of about 1.50 mm. However, since this error was repeated for nearly every module, it did not have a significant impact on the spacing between the modules. The rotational errors caused a larger visual impact in the panel assembly, with some nearing  $0.50^\circ$  within the computer vision system alone. Due mostly to failures in the image segmentation process, the computer vision system also failed to identify a material in the workspace with some regularity during assembly. The overall rate of identification failure was 30.4%, for which the control system would restart the pick state in these situations.

A series of tests were conducted post-assembly to more accurately measure the grasping error of the computer vision guided process. In these tests, the robot identifies

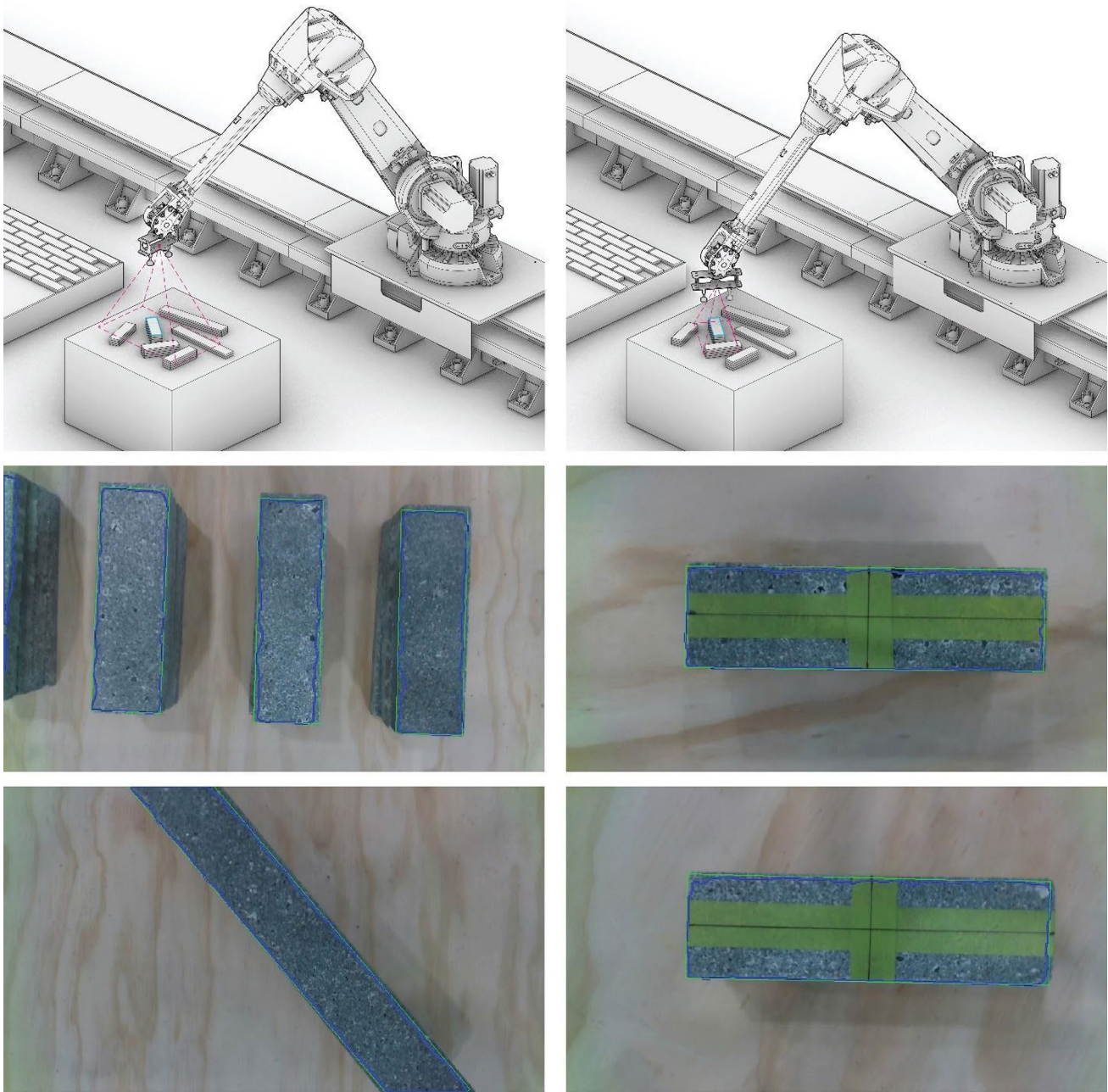
a material's pose in the workspace with the same methods used during the assembly process. Once the target is acquired, a clear acrylic piece is precisely aligned to the vacuum pads on the end effector and held in place using vacuum pressure. The acrylic sheet is etched to locate the tool center point and orientation of the end effector when it is properly aligned. Upon making contact with the selected material, rather than activating the vacuum to pick up the object, the robot turns off the vacuum to place the acrylic TCP marker on the object. With the physical center point of each object in the scene marked, the positional and rotational difference between the robot's TCP from the acrylic sheet and the actual object center point can be measured to represent the grasping error of the robotic pickup process.

The results from this test show that the computer vision error accounted for only part of the grasping error in the whole procedure during the assembly process. The results determined that, on average, the positional error was 2.33 mm and the rotational error was  $1.07^\circ$ . The remaining error in the pickup process can be assumed to come from inaccuracies in the calibration and coordinate system transformation processes. This may be attributed to the somewhat limited resolution of the depth camera and the distance at which the calibration operation was executed. This result still fits within the overall desired positional precision range of 2–3 mm, however, the rotational error is liable to cause issues and the precision would exclude some low-tolerance applications.

During the process of conducting these tests, however, it was discovered that the computer vision system was far more accurate when the long axis of the module was aligned to the X axis of the camera. In those situations, the precision was in the submillimeter range with errors as low as 0.53 mm. It appears that the alignment of the object to the natural pixel grid of the image helps improve the accuracy of both the image segmentation and the object identification processes.

As a result, a new control algorithm was designed to take advantage of this increased precision. For this new method, the robot executes the initial object identification from an elevated position as described earlier. However, the rotation of the selected object as well as its position is saved to be used for the verification step. In the new target verification process, the robot positions the end effector directly over the object and rotates the end effector to align the X-axis of the camera with the long axis of the target material. This alignment of the object to the camera's X-axis carries two crucial benefits. As previously mentioned, the alignment is beneficial for both of the key computer vision processes. Second, it also allows the camera to be positioned much closer to the target object, as the long edge of the material is aligned to the long axis of the image, which further increases the accuracy of the depth camera and provides more fidelity for the object identification process (Fig. 8). Taken together,





**Fig. 8** Diagrams of verification methods and their resulting computer vision object identification, blue lines represent threshold contour, green rectangle is the result of the object identification. Left: Results from panel assembly, note the top of each module is missed in the

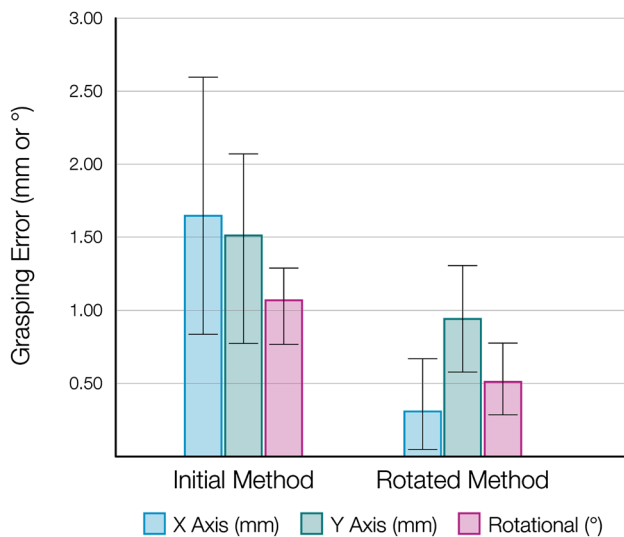
contouring and identification, producing an error in the pose estimation. Right: Results from rotated verification method with significantly less identification error

this new rotated verification method is significantly more accurate than the initial method implemented during the panel assembly.

The rotated verification method does introduce slightly more complexity in the coordinate system transformations, however. Since the camera is rotated in relation to the calibration pose, an additional rotational transformation about the TCP must be applied to the camera’s point cloud to

properly align it with the robot coordinate system to create an accurate target pose. These transformations can be easily managed through the use of the transforms 3d open-source library or a series of matrix algebra operations (Brett 2019).

With this new verification method, the average positional error is 1.07 mm and the rotational error is 0.51°. Interestingly, the error along the long axis of the modules, which was the most problematic during panel assembly is



**Fig. 9** Graph of the average grasping error results from the two verification methods. The initial method used during panel assembly shows significantly higher errors than the final rotated method across all measures. The black error bars represent the 25th and 75th percentile of results, demonstrating the increased reliability of the rotated method as well

reduced to 0.31 mm with the new method (Fig. 9, X-axis in the graph). The reliability of the process also improved significantly. With the initial method, the positional standard deviation in the data collected was about 1.03 mm. The new method reduced this standard deviation to 0.37 mm, expressing far more consistency in the process, an important characteristic in highly repetitive assembly sequences. Perhaps most significantly, the rotated verification method shows that through intelligent positioning and careful calibration, an inexpensive camera and open-source libraries can deliver robotic target generation nearing submillimeter levels of precision.

## 5 Discussion

The research here presents an attempt to establish a highly accessible suite of low-cost components and open-source libraries that can deliver high-precision vision-guided robotic processes for the construction industry. More work is required to automate more stages of the spandrel panel construction, including the placement and fastening of the steel-framed elements and the application of appropriate weather and vapor barriers (Fig. 10). This work should be undertaken with partners from the construction industry to guide the automation efforts towards the challenges that most require attention in the industry.

However, more work could be done on the computer vision system to improve the performance and avail

new paradigms for fabrication. More accurate targeting with the computer vision system could be accomplished by solving the object identification and pose estimation from multiple angles around the work area. Averaging the pose estimation result from multiple camera angles would counteract the tendency for the depth shadow to appear in one direction, skewing the results. This would require the registration and compilation of multiple point clouds into a single master 3D view but would result in a better representation of the work area. To increase the precision, another option is to introduce a second camera to the system to measure the grasping error of the object picked by the robot. A similar system was introduced by Liu et al. and could be used to measure the translation and rotation error of the material in hand, and calculate a new target pose for that object in the assembly to account for that error.

Building off the assembly workflow introduced in this work, the computer vision system could be further developed to help track the progress of the assembly. Rather than relying on algorithmic counters, the camera could be used to identify where material had or had not been placed. This would help not only to calculate the position of the next object in the assembly but also verify the accuracy of the assembly as it progresses and introduce compensation for any errors that have accrued if necessary. Enabling this type of system could also lead to the ability to assemble non-standard parts, such as waste stream materials with unique shapes. The camera could identify the shape of all of the materials in view and decide which piece would fit best in the assembly. This type of process requires high-fidelity information of both the stock of available materials and the voids left in the assembly.



**Fig. 10** Image of the finished spandrel panel assembly in the robotic workcell

## 6 Conclusion

As industrial robot arms are generalized positioning devices, they require extra layers of intelligence and sensing to be aware of the environment around them. The computer vision system developed in this work allows for assembly processes to take place with greater flexibility to adapt to changes in the work environment without an appreciable loss of precision. This is critical to the development of low-cost automation strategies for prefabricated elements but will also be required of in situ robotic construction technologies, as construction sites are chaotic, unstructured, and in a constant state of change.

The research develops the image segmentation, object identification, and camera-to-robot coordinate space transformations required for pick and place assembly workflows using a RGB-D camera, and pairs these capabilities with a closed-loop control system for real-time control of the robotic arm. This system is built with inexpensive components and largely open-source libraries, allowing for easy deployment of the technology in a number of applications. This modular approach allows for many more possibilities than the ones presented here. This technology enables new modes of robotic fabrication and the ability to fabricate unplanned assemblies through a shared agency between designer and robot.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Bechthold M (2010) The return of the future: a second go at robotic construction. *Archit Des* 80:116–121. <https://doi.org/10.1002/ad.1115>
- Bradski G (2000) The openCV library. *Dr Dobb's J Softw Tools* 25:120–125
- Brett M (2019) Transforms 3D. <https://matthew-brett.github.io/transforms3d/>. Accessed 6 Mar 2020
- Dawod M, Hanna S (2019) BIM-assisted object recognition for the on-site autonomous robotic assembly of discrete structures. *Constr Robot*. <https://doi.org/10.1007/s41693-019-00021-9>
- Dawson-Haggerty M (2017) open\_abb: Control ABB robots remotely. [https://github.com/robotics/open\\_abb](https://github.com/robotics/open_abb). Accessed 6 Mar 2020
- Douglas DH, Peucker TK (1973) Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartogr Int J Geogr Inf Geovis* 10:112–122. <https://doi.org/10.3138/FM57-6770-U75U-7727>
- Elashry K, Glynn R (2014) An approach to automated construction using adaptive programming. In: McGee W, de Ponce Leon M (eds) *Robotic fabrication in architecture, art and design*. Springer, Cham, pp 51–66
- Eversmann P (2018) Robotic fabrication techniques for material of unknown geometry. In: De Rycke K et al (eds) *Humanizing digital reality: design modelling symposium Paris*. Springer, Singapore, pp 311–322
- Feddema JT, Mitchell OR (1989) Vision-guided servoing with feature-based trajectory generation (for robots). *IEEE Trans Robot Autom* 5:691–700. <https://doi.org/10.1109/70.88086>
- Feng C, Xiao Y, Willette A, McGee W, Kamat VR (2015) Vision guided autonomous robotic assembly and as-built scanning on unstructured construction sites. *Automat Constr* 59:128–138. <https://doi.org/10.1016/j.autcon.2015.06.002>
- Furrer F, Wermelinger M, Yoshida H, Gramazio F, Kohler M, Siegwart R, Hutter M (2017) Autonomous robotic scene stacking with online next best object target pose planning. *IEEE Int Conf Robot Autom*. <https://doi.org/10.1109/ICRA.2017.7989272>
- Gifftthaler M, Sandy T, Dörfler K, Brooks I, Buckingham M, Rey G, Kohler M, Gramazio F, Buchli J (2017) Mobile robotic fabrication at 1:1 scale: the in situ fabricator. *Constr Robot* 1:3–14. <https://doi.org/10.1007/s41693-017-0003-5>
- Ikeuchi K (1987) Generating an interpretation tree from a CAD model for 3D-object recognition in bin-picking tasks. *Int J Comput Vis* 1:145–165. <https://doi.org/10.1007/BF00123163>
- Jeffers M (2016) Autonomous robotic assembly with variable material properties. In: Reinhardt D, Saunders R, Burry J (eds) *Robotic fabrication in architecture, art and design*. Springer International Publishing, Cham, pp 240–249
- Kabsch W (1976) A solution for the best rotation to relate two sets of vectors. *Acta Cryst A* 32:922–923. <https://doi.org/10.1107/S0567739476001873>
- Liu M-Y, Tuzel O, Veeraraghavan A, Taguchi Y, Marks TK, Chellappa R (2012) Fast object localization and pose estimation in heavy clutter for robotic bin picking. *Int J Robot Res* 31:951–973. <https://doi.org/10.1177/0278364911436018>
- Mahler J, Liang J, Niyaz S, Laskey M, Doan R, Liu X, Ojea JA, Goldberg KY (2017) *Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics*. Robotics: science and systems. Cambridge University Press, Cambridge
- Oh J-K, Lee S, Lee C-H (2012) Stereo vision based automation for a bin-picking solution. *Int J Control Autom Syst* 10:362–373. <https://doi.org/10.1007/s12555-012-0216-9>
- Shohet IM, Rosenfeld Y (1997) Robotic mapping of building interior—precision analysis. *Automat Constr* 7:1–12. [https://doi.org/10.1016/S0926-5805\(97\)00030-7](https://doi.org/10.1016/S0926-5805(97)00030-7)
- Tsai RY, Lenz RK (1989) A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE Trans Robot Autom* 5:345–358. <https://doi.org/10.1109/70.34770>
- Vasey L, Maxwell I, Pigram D (2014) Adaptive part variation: a near real-time approach to construction tolerances. In: McGee W, de Ponce Leon M (eds) *Robotic fabrication in architecture, art and design*. Springer International Publishing, Cham, pp 291–304
- Wu K, Kilian A (2016) Developing architectural geometry through robotic assembly and material sensing. In: Reinhardt D, Saunders R, Burry J (eds) *Robotic fabrication in architecture, art and design*. Springer International Publishing, Cham, pp 240–249

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.