

Modeling of video projectors in OpenGL for implementing a spatial augmented reality teaching system for assembly operations

Carlos M. Costa¹, Germano Veiga¹, Armando Sousa¹, Luís Rocha²,
A. Augusto Sousa³, Rui Rodrigues³ and Ulrike Thomas⁴

Abstract—Teaching complex assembly and maintenance skills to human operators usually requires extensive reading and the help of tutors. In order to reduce the training period and avoid the need for human supervision, an immersive teaching system using spatial augmented reality was developed for guiding inexperienced operators. The system provides textual and video instructions for each task while also allowing the operator to navigate between the teaching steps and control the video playback using a bare hands natural interaction interface that is projected into the workspace. Moreover, for helping the operator during the final validation and inspection phase, the system projects the expected 3D outline of the final product. The proposed teaching system was tested with the assembly of a starter motor and proved to be more intuitive than reading the traditional user manuals. This proof of concept use case served to validate the fundamental technologies and approaches that were proposed to achieve an intuitive and accurate augmented reality teaching application. Among the main challenges were the proper modeling and calibration of the sensing and projection hardware along with the 6 DoF pose estimation of objects for achieving precise overlap between the 3D rendered content and the physical world. On the other hand, the conceptualization of the information flow and how it can be conveyed on-demand to the operator was also of critical importance for ensuring a smooth and intuitive experience for the operator.

I. INTRODUCTION

Teaching humans new manufacturing skills or advanced assembly / maintenance operations can be a long and error prone process that requires reading extensive manuals and a period of tutoring in which they are trained by field experts. This teaching period can be shortened and done without the need of other workers by relying on immersive Human Machine Interface (HMI) teaching systems that are able to transmit the knowledge more effectively using step by step instructions containing text and video along with visual cues showing the work areas and pick / place locations tagged with contextual help. Moreover, when coupled with active perception systems that can detect the assembly objects and what the operator is doing, the teaching systems can also act as a supervisor, alerting the operator when a mistake is made or when a damaged component needs to be replaced. This

approach to skill transfer using immersive HMIs along with dynamic feedback can speedup and improve the effectiveness of the training sessions while also giving continuous quality control, allowing to reduce the cost and time of product assembly, repair and maintenance.

With these goals in mind, a spatial augmented reality teaching system was developed for projecting into the operator workspace contextual assembly instructions that provide detailed information about the operations and tools that are required to assemble a given product. The main advantage of this approach is its ability to provide accurate 3D information directly into the environment and only when it is needed. For production lines that may receive custom products or have their employees rotating between workstations, a spatial augmented reality system can quickly and intuitively guide them throughout the assembly process. Moreover, the proposed system can also be used for coordinating tasks between human operators and robotic systems, because it is able to highlight 3D work areas or objects and provide visual cues for informing the operator what the robot will be doing next and where it will be working.

In the following section, a brief overview of the augmented reality systems that were developed over the years will be given. Then in Section III, the mathematical modeling and calibration of video projectors will be discussed. Later on, Section IV will describe the immersive HMI that was developed. Given the lack of Computer Aided Design (CAD) models of the starter motor used for testing our system, Section V will describe how the 3D model was retrieved using a structured light 3D scanner. Then, Section VI will present the 6 Degrees of Freedom (DoF) object pose estimation system. Finally, Section VII will discuss the results of a training session while Section VIII will summarize the conclusions and present possible future work.

II. RELATED WORK

Augmented Reality (AR) interfaces [1] have a wide range of applications within the manufacturing industry, from the design [2], simulation and planning phase for fast prototyping, to the training and guidance [3] of the operators that will be manufacturing, assembling and providing maintenance for the final product [4]. They offer an immersive way of exchanging information between a human operator and a robot / machine, allowing the development of cooperative assembly lines [5]. This immersive environment can be created with a wide range of devices, such as projectors, smart glasses / watches / phones, tablets, Virtual Reality (VR) headsets,

¹ Authors with the Centre for Robotics in Industry and Intelligent Systems (CRIIS) of INESC TEC and with the Faculty of Engineering of the University of Porto, Portugal (emails: {carlos.m.costa, germano.veiga}@inesctec.pt and asousa@fe.up.pt)

² Author with the Centre for Robotics in Industry and Intelligent Systems (CRIIS) of INESC TEC, Portugal (email: luis.f.rocha@inesctec.pt)

³ Authors with the Centre for Information Systems and Computer Graphics (CSIG) of INESC TEC and with the Faculty of Engineering of the University of Porto, Portugal (emails: {augusto.sousa, rui.rodrigues}@fe.up.pt)

⁴ Author with the Robotics and Human Machine Interaction Laboratory at the Technical University of Chemnitz, Germany (email: ulrike.thomas@eti.tu-chemnitz.de)

among others. Projectors allow to perform accurate overlay of virtual information into physical objects when properly calibrated [6] and modeled within a 3D rendering engine. They are typically installed on top of the workstations for spatially augmenting the environment with digital information without requiring the operator to wear any special hardware (that may cause discomfort or impact their productivity). This approach is very useful for assisting operators in new and complex operations while also helping them perform their tasks faster by having dynamic contextual information shown directly into the environment where it is needed (such as projection of geometric information for assembly / maintenance [7], cutting / welding [8] or even painting operations [9]).

Wearable devices such as smart glasses / watches offer a flexible alternative [10] which is more suitable for providing guiding information when the operator is performing complex jobs in environments that are hard to reach by an overhead projector or when the operator needs to navigate in a large workspace. Screens with rear mounted sensors provide a quick and low cost approach for adding environment annotations which are useful for assembly / maintenance operations. They may use AR markers [11], markerless 2D / 3D perception [12] or a combination of both to analyze the environment and detect where are the target objects and what the operator is doing [13] in order to overlay virtual models on top of the real objects or provide contextual and corrective information for helping the operator work faster while also doing less mistakes. The exchange of information between a training system and an operator may take advantage of multimodal HMIs [14] for effectively guiding the operator during the assembly / maintenance tasks (using vibrotactile bracelets along with visual and audio cues). On the other hand, VR headsets provide an immersive virtual environment for teaching the robot / operator [15] without requiring access to the physical objects / robots / environment layout, allowing fast testing and prototyping of new products.

III. SPATIAL AUGMENTED REALITY

Projection mapping is a Spatial Augmented Reality (SAR) [1] approach in which video projectors are used to overlay virtual geometry on top of real objects or surfaces, allowing the creation of an immersive environment that together with 3D perception systems can be used to develop interactive interfaces that show contextual information for helping or teaching human operators performing complex tasks faster. The next sections describe the mathematical modeling of video projectors within the Open Graphics Library (OpenGL) and the associated calibration that is necessary for performing proper 3D rendering of the virtual world in order to achieve high accuracy projection.

A. Projector modeling

Over the years, several projection technologies were developed according to the requirements of color fidelity / saturation, image sharpness, brightness, contrast, refresh rate and price. Currently, the video projection market is split between

reflective Digital Light Processing (DLP) and transmissive Liquid Crystal Display (LCD) projection technology, with a small percentage consisting of a hybrid between the two technologies, such as the Liquid Crystal on a Silicon (LCoS) projectors. For video projection mapping purposes, reflective projectors are usually better suited than the remaining technologies given their ability to create images with smaller gaps between the projected pixels (smoother images) and they also have higher contrast, better color accuracy / uniformity, fewer dead pixels and the image quality does not degrade over time.

Despite the wide range of technologies and hardware configurations (example for DLP shown in Figure 1), the output of a video projector can be seen as an inverse pinhole camera (diagram shown in Figure 4) due to the grid disposition of the projected image and the very low distortion that modern projectors have. As such, rendering of 3D virtual environments for spatial augmented reality can be performed efficiently using an extended version of the OpenGL projection matrix¹ (presented in Equation (8)).

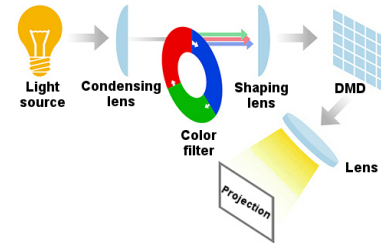


Figure 1: Diagram of a single chip DLP projector²

The extension of the default OpenGL projection matrix is of critical importance for being able to model video projectors as OpenGL cameras because unlike normal cameras that have a centered lens (which results in a principal point close to the image center), projectors normally have an off-centered lens (causing the principal point to be around the bottom of the image). This is due to the fact that projector lenses are assembled slightly tilted for redirecting the projection light upwards. This way the projector can be installed almost parallel to a table / ceiling and its projected image will not intersect the table / ceiling. In higher-end projectors, the lens shift³ can even be adjusted after the projector is manufactured, making its installation more flexible (diagram in Figure 2). The lower cost alternative to lens shift is keystone correction, but this approach is typically avoided in SAR systems since it distorts the image and wastes projection pixels.

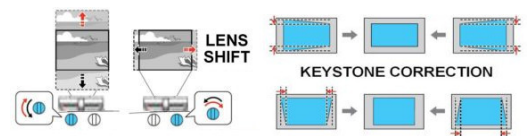


Figure 2: Diagram of lens shift (left) and keystone correction (right)⁴

The perspective projection of vertex data is a pivotal stage in any rendering engine. In OpenGL, the projection matrix is one

¹http://ksimek.github.io/2013/06/03/calibrated_cameras_in_opengl

²<https://vimeo.com/blog/post/display-tech-home-projectors>

³https://www.projectorpeople.com/resources/lens_shift.asp

⁴<http://www.theprojectorexpert.com/projector-lens-shift>

B. Projector calibration

High accuracy projection mapping requires proper hardware / software calibration of the camera / projector and also appropriate positioning within the intended workspace in order to avoid occlusions caused by the objects 3D shape or the human operators. The calibration process estimates the intrinsic parameters of the projector (that do not change when the projector is moved within the workspace) along with the extrinsic parameters that are needed to know where is the projector in the global reference frame in order to be able to do proper 3D rendering of the scene that will be projected.

The intrinsic parameters of a DLP projector can be computed using image analysis of complementary gray code patterns (example in Figure 7) projected into a chessboard. The calibration system proposed in [6] was used to retrieve the 5 intrinsic parameters (F_x , F_y , C_x , C_y , S) and 5 distortion coefficients of the projector along with the 3D position and rotation of the projector in relation to the camera (that remains firmly attached to the projector support for fast recalibration of the extrinsic parameters, as seen in Figure 18). It was used 5 sets of 42 gray code image patterns captured with the chessboard in different positions and orientations in relation to the projector, that was pointing to the table workspace at a distance of 0.81 meters.

For validating both the calibration and also the proposed camera modeling within OpenGL, the 6 DoF pose of a chessboard (in relation to the camera) was estimated using the OpenCV `calib3d` module, followed by the computation of the 6 DoF pose of the projector in relation to the chessboard using the camera-projector extrinsic parameters. Then, a Gazebo virtual world was created with a virtual camera positioned in the 6 DoF pose computed earlier (for simulating the video projector) and a 3D model was placed at the origin of the virtual world with the dimensions matching the physical chessboard. After rendering an image using the intrinsic projector parameters and projecting it on top of the physical chessboard (shown in Figure 8), it can be seen that the white squares were projected into the chessboard with sub-millimeter accuracy.

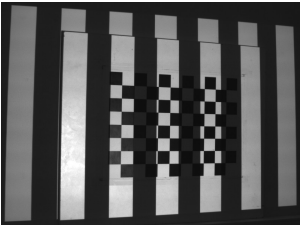


Figure 7: One of the projector gray code calibration patterns

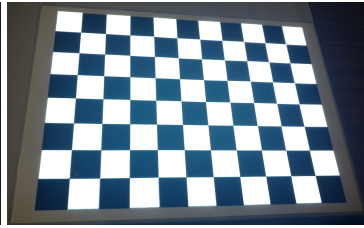


Figure 8: Projector validation pattern (projected white squares)

C. Scene rendering

For achieving accurate projection mapping, the Gazebo simulator⁹ camera implementation was improved for allowing the setting of a custom projection matrix in order to perform 3D rendering with the camera model proposed earlier, that

takes into account the projector intrinsic parameters. Moreover, it was added the possibility to dynamically change image, video and text during runtime for allowing the display of the relevant information for each assembly step. For efficient 3D scene rendering, the Gazebo simulator relies on the cross platform open source Ogre3D graphics engine¹⁰, that in turn uses the OpenGL Graphics Processing Unit (GPU) Application Programming Interface (API) to take advantage of the massively parallel graphics cards currently available to generate raster images for the DLP projector (example of a rendered scene for the last assembly step in Figure 9).

For user interaction the Gazebo simulator has a Qt¹¹ Graphical User Interface (GUI) that allows visual inspection of the scene while also giving the option to add new objects or move and rotate existing models. Moreover, for lightweight rendering it can also start in server mode without a GUI.

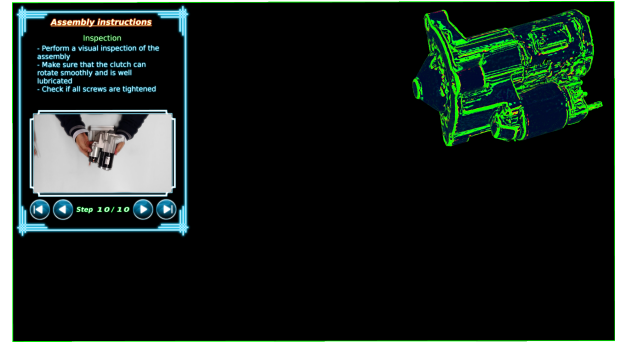


Figure 9: 3D scene rendering using the Gazebo simulator

IV. HUMAN MACHINE INTERFACE

The immersive HMI developed (shown in Figures 9 and 10) projects into the workspace detailed textual information of the current assembly task along with a video showing the operation being performed by an expert operator. Given the high variability of assembly / maintenance operations, the system was designed to decompose the assembly process into a set of small and concise operations. This allows to keep the operator focused on the current task and reduces the required projection area. Moreover, the operator can pause and move the video forwards and backwards, allowing him to inspect a given complex operation with more time.

The user interaction with the projected HMI is done by analyzing the 3D point cloud sensor data that falls within a set of Regions of Interest (ROIs), that are shown in Figure 11 as 4 green cubes for navigating within the assembly steps (first, previous, next and last), 1 dark blue box for pausing / playing the video and 1 yellow box for the video seek functionality (examples of a user interacting with the HMI shown in Figures 12 to 14).

For ensuring robust detection of the user intentions, each ROI has a state machine that triggers its associated action only when a minimum number of points falls within the specified ROI boundaries (avoids sensor noise problems) and the user

⁹<http://gazebo.org>

¹⁰<http://www.ogre3d.org>

¹¹<http://www.qt.io>

holds his finger within it for at least 0.25 seconds. Moreover, to avoid unintentionally triggering the same action several times, the user needs to remove and insert the finger into the ROI to request the action again.

When a ROI state machine activates its action, the 3D sensor data centroid (shown as spheres in Figure 11) is computed for providing a visual debugging feedback of the HMI state and also for being used in higher level perception, namely in the seek bar ROI (the vertical yellow box in Figure 11), in which the seek time is computed by considering the relative position of the finger within the ROI (the bottom of the ROI is associated with the start of the current video while the top corresponds to the end of the current video).



Figure 10: Rendering of the human machine interface using Gazebo

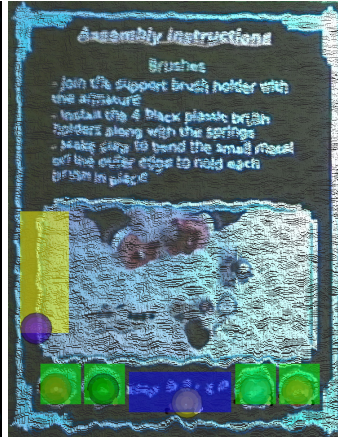


Figure 11: ROIs for the HMI overlaid on top of the Kinect 2 point cloud sensor data using Rviz)

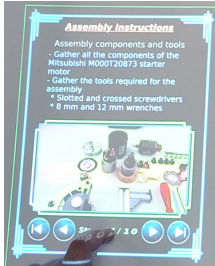


Figure 12: Example of video play / pause interaction

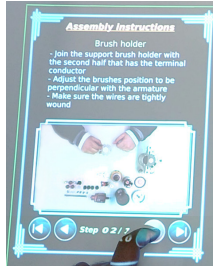


Figure 13: Visual highlight of the request to move the next assembly step

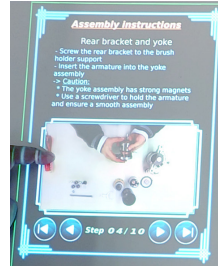


Figure 14: Example of video seek interaction

V. OBJECT 3D RECONSTRUCTION

For performing proper 3D rendering and also be able to estimate the 6 DoF pose of an object within the workspace, a 3D CAD or mesh model of the final product is required. Given the lack of public available CAD models for the Mitsubishi M000T20873 starter motor (shown in Figure 15), it was necessary to perform object 3D reconstruction. The 3D mesh model shown in Figure 16 was generated using the David Laser 3D structured light system¹², and was built by surface matching algorithms using sensor data retrieved from 38 scans in which the starter motor was moved and rotated several times

in order to capture enough sensor data for reconstructing the entire surface. This particular object created some challenges for the structured light scanner because it contains polished metallic sections and also black coated surfaces. As such, it was necessary to capture the same object sections several times with different projector brightness and camera exposure times (the dark regions required the maximum projector brightness and very high exposure times while the polished sections required dimmer projector brightness and very short exposure times for the surface to be fully reconstructed). Moreover, for ensuring that the point cloud matching algorithms worked as expected, all captured point clouds needed to be manually cleaned using the David Laser scanning software before performing the alignment and reconstruction of the mesh model.

The last step of the assembly process included a visual inspection stage performed by the operator in which he compared the final assembled product with a projected surface outline that was computed from the reconstructed mesh and then overlaid on top of the physical object after its 6 DoF pose was estimated. The outline was generated using a curvature estimator available in MeshLab¹³. Namely, the "Compute curvature principal directions" filter was used with the principal component analysis method and the mean curvature color mapping algorithm (result shown in Figure 16). For improving the visibility of the outline, the color palette was remapped to show the high curvature regions (associated with surfaces boundaries) using the green color and the remaining intermediate curvatures were changed to very dark blue, for giving high contrast between the outline and the other surfaces. This was achieved by exporting the mesh to the .dae format (required by Gazebo for displaying colored models) with the color information separated to a .png texture. This way, it was possible to use Gimp¹⁴ for remapping and enhancing the colors of the mesh outlines (as seen in Figures 9 and 29).

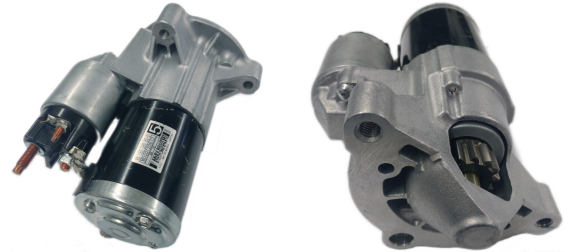


Figure 15: Mitsubishi M000T20873 starter motor

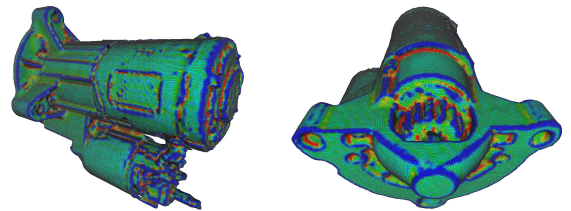


Figure 16: 3D model of the starter motor reconstructed using the David Laser structured light 3D scanner

¹²<http://www.david-3d.com>

¹³<http://www.meshlab.net>

¹⁴<https://www.gimp.org>

VI. 6 DOF POSE ESTIMATION OF OBJECTS

Robust and accurate estimation of the 6 DoF pose of objects is a requirement when virtual information must be overlaid on top of objects that may change their location over time. To achieve this goal, the 3D point cloud registration system (drl¹⁵) described in [17] was fine-tuned for our table top use case¹⁶. Namely, the reference point cloud preprocessing pipeline was configured to randomly select 3000 vertices of the starter motor reconstructed mesh (small green circles shown in Figure 17) and compute the Scale Invariant Feature Transform (SIFT) keypoints [18] (large yellow circles) and their associated Fast Point Feature Histogram (FPFH) descriptors [19]. Later on, the registration pipeline for the sensor point clouds was setup. Its filtering stage was configured to segment the target object (starter motor) by extracting the points that were within a ROI specified in a calibrated frame on top of the table. The ROI defined the plausible space in which the object could be placed and also allowed to remove the points associated with the table. Moreover, for ensuring that the initial pose estimation would take less than 3 seconds, a random sample filter was also configured to downsample the point cloud within the ROI to 2500 points.

Given that in this use case only a single object appears within the ROI, the clustering algorithms were not added to the drl runtime pipeline. But they will be very useful in the future when extending the proposed SAR system to provide contextual information and visual feedback for all the parts visible by the sensor (the clustering stage will then provide input for a object recognition module, which will estimate which CAD is the best fit for each group of 3D points).

After finishing the setup of the preprocessing stage for the sensor point clouds pipeline, its initial alignment algorithms were configured and fine-tuned. Namely, the SIFT keypoint detector along with the FPFH keypoint descriptor and the Random Sample Consensus (RANSAC) feature matcher [19]. Moreover, the Iterative Closest Point (ICP) [20] algorithm was also added for refining the point cloud registration in order to achieve a 6 DoF pose estimation with high accuracy (less than 2 mm of alignment error, as can be seen by the good overlap between the physical starter motor and its projected virtual outline displayed in Figure 29).

The drl modules discussed earlier were enough for estimating the 6 DoF pose of objects, but for improving the efficiency of the drl, the tracking and recovery matching pipelines were also configured. These two pipelines allow the definition of a set of cloud matchers that are fine-tuned to align the sensor data by successively updating the 6 DoF pose of the object as new sensor data arrives. This allows to run the robust, but computationally intensive feature matching algorithms once, and then rely on dense point cloud matching algorithms to quickly and accurately update the object pose at the frame rate of the 3D sensor (up to 30 Hz, depending on the CPU used). The tracking pipeline was configured to rely on the ICP

algorithm with the point-to-point metric with a search radius of 0.07 m for establishing correspondences during the iterative matching process and with the maximum number of iterations set to 300. On the other hand, the recovery pipeline was fine-tuned to use the ICP algorithm with the point-to-plane metric with the search radius and number of iterations increased to 0.2 m and 400 respectively. This approach of starting with robust feature matching algorithms and then rely on tracking algorithms with the sporadic usage of recovery methods was critical for being able to track the object with a reasonable frame rate, which will be useful in the future when extending the proposed SAR system to monitor the operator actions for ensuring that he is following the instructions and did not forget or misplaced assembly parts.

In Figure 17 is shown an example of the estimation of the 6 DoF pose of the starter motor. The left image displays the overlay of the reference point cloud on top of the 3D sensor data in the previously estimated pose (before the operator occluded the part with its hand and moved it to a new place) while the right image shows the updated pose after alignment, that correctly detected that the operator moved the part to the right and rotated it 90° clockwise.

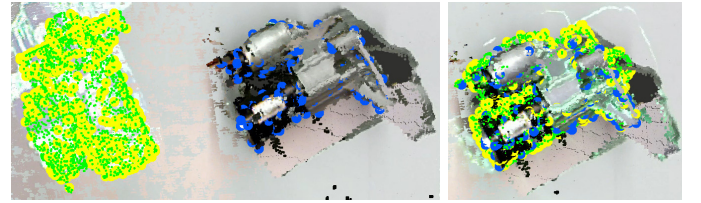


Figure 17: Example of 6 DoF pose estimation of the assembled starter motor before (left) and after (right) the alignment of the 3D reconstructed mesh with the 3D sensor data

VII. TRAINING OPERATORS TO ASSEMBLE A STARTER MOTOR USING AN IMMERSIVE HMI

Small parts assembly of flexible components is a very challenging task to automate given the advanced sensing and gripping technologies that it requires. As such, currently it is more cost effective to have cooperative assembly lines in which humans perform the tasks that require robust perception, adaptive grasping and high-level cognition while the robots do the remaining tasks. In the next sections, the application of our immersive training system for the assembly of a starter motor will be presented. This is a representative use case of small parts assembly given its diversity of operations and components. Moreover, since it has flexible parts (rubbers, wires, springs), it would be a prime candidate for a collaborative assembly line, in which besides teaching human operators, our immersive HMI system could also be used to assign and coordinate tasks in which human operators and robotic systems are involved.

A. Testing platform

Our immersive teaching system¹⁷ was developed as a Robot Operating System (ROS)¹⁸ package for fast integration into

¹⁵https://github.com/carlosmccosta/dynamic_robot_localization

¹⁶https://github.com/carlosmccosta/object_recognition

¹⁷https://github.com/carlosmccosta/assembly_projection_mapping_teaching

¹⁸<http://www.ros.org>

robotic workstations and relies on the Gazebo simulator for 3D rendering and the Point Cloud Library (PCL) for 3D perception. It was tested with a BenQ W1070 DLP projector for overlaying the teaching information into the environment, an Asus Xtion Pro Live structured light 3D sensor for 6 DoF pose estimation of objects and a Kinect 2 Time of Flight (ToF) 3D sensor for the user interaction analysis. In Figure 18 it can be seen the work area and the hardware disposition (in the right image the projector is on the top right, the Kinect 2 is on the left, the Asus Xtion is below the projector and the David Laser 3D structured light system camera is at the top).



Figure 18: Hardware setup

B. Training session

The training session started by gathering all the assembly parts and the required tools for performing the starter motor assembly (shown in Figure 19). Then using the proposed immersive teaching system, the operator read the instructions, watched the videos and navigated through the assembly steps using the projected interaction buttons (displayed in Figures 12 to 14) until he completed the assembly process. Namely, the operator started by assembling the brushes into the brush holder (seen in Figure 20) and then bended the braided cables for ensuring that the brushes were perpendicular to the armature, that was assembled later on (shown in Figure 21). These kind of operations that involve flexible parts with cables and rubbers are very hard to automate with robotic manipulators, and as such, are the ideal candidate for being assigned to operators. On the other hand, assembly steps that deal with large and rigid parts can be delegated to robots, which is the case of step 4 (displayed in Figure 22), in which the operator assembled the rear bracket and attached the large cylindrical yoke. This way, the operator could be performing steps 5 to 8 (which involved the lower section of the starter motor) while the robot would be finishing in parallel the assembly of the top sections. Later on, step 5 included the placement of the large bottom bracket on top of a fixture for holding it vertically, followed by the assembly of the clutch and shift lever (presented in Figure 23). Then, step 6 included the assembly of the 3 planetary gears (shown in Figure 24) while in step 7 the plunger and several isolation rubbers were installed (displayed in Figure 25). Later on, the plunger spring

and its magnetic switch were attached to the lower section of the starter motor (as seen in Figure 26). Finally, the lower and upper section of the starter motor were assembled together (presented in Figure 27), followed by a visual inspection of the assembled product using the projected 3D model outline (seen in Figures 28 to 30).

For helping the operator during the final inspection phase, the proposed SAR system estimated the 6 DoF pose of the starter motor and then projected its expected outline on top of it. The main purpose of this final stage was to test the accuracy of the proposed SAR system. Namely, to evaluate if the proposed SAR system was able to achieve a good overlap between the physical and virtual objects. This would implicitly confirm that the approach proposed to model the 3D camera within Gazebo along with the subsystems that influence the rendering of the starter motor outline (calibration of the projector, cameras and 3D sensors along with 6 DoF pose estimation) were suitable for achieving a usable SAR system. Looking at Figures 28 to 30, the overall overlap error seems to be below 2 mm, making the proposed SAR system ready to be applied to other use cases and be used as a starting point for developers wanting to incorporate SAR into their products, such as cooperative workstations in which operators and robots work side by side.

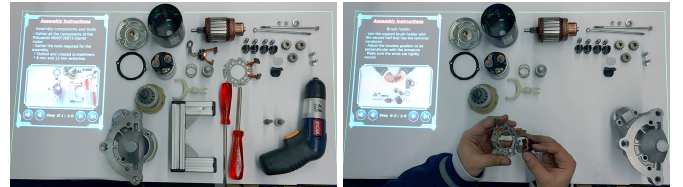


Figure 19: Step 1 - starter motor parts and assembly tools

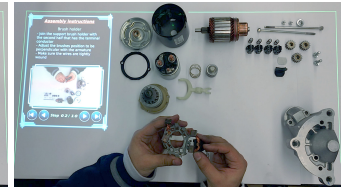


Figure 20: Step 2 - assembly of the brush holder

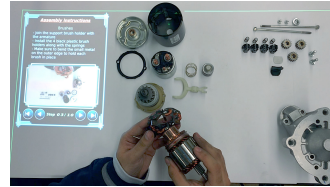


Figure 21: Step 3 - assembly of the brushes into the armature

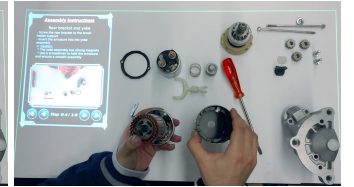


Figure 22: Step 4 - assembly of the rear bracket and yoke

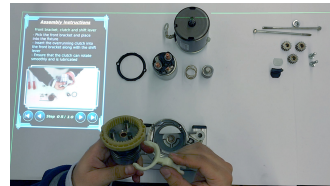


Figure 23: Step 5 - assembly of the front bracket, clutch and shift lever

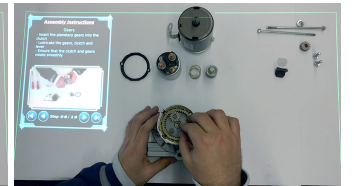


Figure 24: Step 6 - assembly of the planetary gears



Figure 25: Step 7 - assembly of the plunger and rubbers

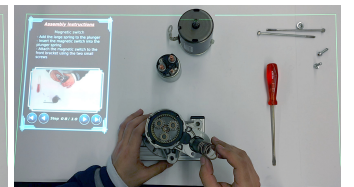


Figure 26: Step 8 - assembly of the plunger spring and magnetic switch

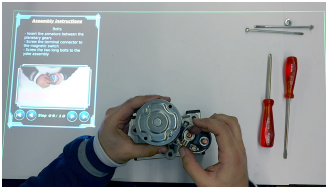


Figure 27: Step 9 - assembly of the armature and its bolts



Figure 28: Step 10 - visual inspection of the assembled product



Figure 29: Detailed view of the valve

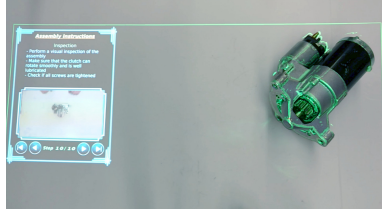


Figure 30: Projection of the outline of the reconstructed 3D model after 6 DoF pose estimation phase

VIII. CONCLUSIONS

This paper presented the fundamental technologies required to implement an accurate SAR system within the domain of small parts assembly. Namely, it proposed an efficient and accurate approach for modeling video projectors using the OpenGL projection matrix while also describing how to calibrate the intrinsic and extrinsic parameters of the sensing and projection hardware, which together with the 6 DoF pose estimation system proposed are critical for ensuring that the system as a whole is able to achieve an overlap error between the physical and projected virtual objects below 2 mm. These technologies were validated using our immersive teaching system, that is capable of guiding the operator throughout the assembly process using a projected HMI containing text and video content while also providing a visual inspection phase in which the expected product outline is overlaid on top of the assembled components. This proof of concept use case served to validate the approaches suggested and can be used as a starting point for other applications, namely, cooperative workstations in which a SAR system can be used to coordinate tasks between humans and robotic systems by showing directly on the environment the expected work areas and operations associated to each operator and robot.

The presented immersive teaching system can be improved further by adding an assembly analysis module for monitoring what the operator is doing in order to provide contextual help (such as detecting that the current component was mounted correctly and projecting the next part that the operator needs to assemble) and also alert for possible mistakes done by the operator during assembly. This would allow continuous analysis and quality control of the assembly process, reducing the time required for the detection and correction of assembly problems. On the other hand, the expansion of the SAR system to other use cases and its evaluation with a large group of operators would provide useful feedback for its improvement and would allow to quantify how much effective it can be when compared with traditional methods.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union's Horizon 2020 - The EU Framework Programme for Research and Innovation 2014-2020, under grant agreement No. 723658.

This work is financed by the ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme, and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project SAICTPAC/0034/2015-POCI-01-0145-FEDER-016418.

REFERENCES

- [1] O. Bimber and R. Raskar, *Spatial Augmented Reality: Merging Real and Virtual Worlds*. A. K. Peters, 2005.
- [2] A. Nee, S. Ong, G. Chrysosouris, and D. Mourtzis, "Augmented reality applications in design and manufacturing," *CIRP Annals - Manufacturing Technology*, vol. 61, no. 2, pp. 657 – 679, 2012.
- [3] S. Webel, "Multimodal training of maintenance and assembly skills based on augmented reality," Ph.D. dissertation, Technische Universität Darmstadt, 2011.
- [4] Z. Zhu, V. Branzoi, M. Wolverton, G. Murray, N. Vitovitch, L. Yarnall, G. Acharya, S. Samarasekera, and R. Kumar, "Ar-mentor: Augmented reality based mentoring system," in *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, Sept 2014, pp. 17–22.
- [5] C. Lenz, "Context-aware human-robot collaboration as a basis for future cognitive factories," Ph.D. dissertation, Technische Universität München, 2011.
- [6] D. Moreno and G. Taubin, "Simple, accurate, and robust projector-camera calibration," in *Second International Conference on 3D Imaging, Modeling, Processing, Visualization Transmission*, Oct 2012, pp. 464–471.
- [7] A. E. Uva, M. Gattullo, V. M. Manghisi, D. Spagnolo, G. L. Cascella, and M. Fiorentino, "Evaluating the effectiveness of spatial augmented reality in smart manufacturing: a solution for manual working stations," *The International Journal of Advanced Manufacturing Technology*, vol. 94, no. 1, pp. 509–521, Jan 2018.
- [8] A. Doshi, R. T. Smith, B. H. Thomas, and C. Bouras, "Use of projector based augmented reality to improve manual spot-welding precision and accuracy for automotive manufacturing," *The International Journal of Advanced Manufacturing Technology*, vol. 89, no. 5, pp. 1279–1293, Mar 2017.
- [9] G. F. Barbosa, J. de Carvalho, and C. H. P. de Souza, "Deployment of a laser projection solution for stripes plotting based on six sigma dmaic methodology applied to aircraft painting shop," *Production & Manufacturing Research*, vol. 2, no. 1, pp. 697–711, 2014.
- [10] G. Michalos, N. Kousi, P. Karagiannis, C. Gkournelos, K. Dimoulas, S. Koukas, K. Mparis, A. Papavasileiou, and S. Makris, "Seamless human robot collaborative assembly – an automotive case study," *Mechatronics*, vol. 55, pp. 194 – 211, 2018.
- [11] S. Siltanen, *Theory and Applications of Marker-based Augmented Reality*. VVT, 2012.
- [12] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, and J. Wan, "3d object recognition in cluttered scenes with local surface features: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2270–2287, Nov 2014.
- [13] A. Bannat, J. Gast, G. Rigoll, and F. Wallhoff, "Event analysis and interpretation of human activity for augmented reality-based assistant systems," in *4th International Conference on Intelligent Computer Communication and Processing*, Aug 2008, pp. 1–8.
- [14] S. Webel, U. Bockholt, T. Engelke, N. Gavish, M. Olbrich, and C. Preusche, "An augmented reality training platform for assembly and maintenance skills," *Robotics and Autonomous Systems*, vol. 61, no. 4, pp. 398–403, Apr. 2013.
- [15] A. F. Abate, M. Guida, P. Leoncini, M. Nappi, and S. Ricciardi, "A haptic-based approach to virtual training for aerospace industry," *Journal of Visual Languages and Computing*, vol. 20, no. 5, pp. 318 – 325, 2009.
- [16] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2003.
- [17] C. M. Costa, H. M. Sobreira, A. J. Sousa, and G. M. Veiga, "Robust 3/6 dof self-localization system with selective map update for mobile robot platforms," *Robotics and Autonomous Systems*, vol. 76, pp. 113 – 140, 2016.
- [18] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [19] R. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (fpfh) for 3d registration," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2009, pp. 3212–3217.
- [20] P. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, Feb 1992.