2015 IEEE International Symposium on Mixed and Augmented Reality

[POSTER] Remote Welding Robot Manipulation using Multi-view Images

Yuichi Hiroi[1, †], Kei Obata[1, ‡], Katsuhiro Suzuki[1, †], Naoto Ienaga[1, ‡],
Maki Sugimoto[1,3, †], Hideo Saito[1,3, ‡], Tadashi Takamaru[2,*]
[1] Keio University, [2] Takamaru Engineers, [3] CREST, Japan Science and Technology Agency

Abstract

This paper proposes a remote welding robot manipulation system by using multi-view images. After an operator specifies two-dimensional path on images, the system transforms it into three-dimensional path and displays the movement of the robot by overlaying graphics with images. The accuracy of our system is sufficient to weld objects when combining with a sensor in the robot. The system allows the non-expert operator to weld objects remotely and intuitively, without the need to create a 3D model of a processed object beforehand.

Keywords: Remote robot control, multi-view images, industrial robot, augmented reality.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces − Interaction styles; I.2.9 [Artificial Intelligence]: Robotics — Operator interfaces

1 Introduction

A welding robot is generally programmed by two methods. One method is by using a teach pendant. When using a teach pendant, the operator controls the movement of the robot by having the robot in front and pressing the operational button. However the operation of the robot via the teach pendant requires technical abilities and the results may vary depending on the operator. Another method is an offline simulation by utilizing CG models. This method uses 3D models created in an offline environment to automatically generate the welding path by selecting any arbitrary ridge lines on the components. However, this requires a 3D model of all the components so this method is not suited for procedures that are only done a couple of times.

To overcome these problems, we propose a tele-operating welding robot manipulation system by using multi-view image (Figure 1). The operator first specifies the path of the robot by tracing the two dimensional path on to the captured image. By doing so for each captured image, our system is able to automatically transfer the two dimensional path into three dimensional one. Our system also generates CG animation that visualizes the posture of the robot when it traces the calculated path. The animation is overlaid with each captured image to check the behavior of robot the before actual manipulation.

Since our system is able to predict how the robot will move by using augmented reality, a non-expert operator can weld objects remotely along the path that he/she wants to weld. Moreover, since our system calculates high precision pathing without the need of creating an accurate 3D model of the processed object, this system contributes to higher efficiency when welding high-mix low-volume production. We implemented the prototype of the system and tested its accuracy. We also welded objects by using an actual welding robot.

† {y.hiroi, katsuhirosuzuki, sugimoto}@imlab.ics.keio.ac.jp
‡ {obata, ienaga, saito}@hvrl.ics.keio.ac.jp
* tadashi@takamaru.com

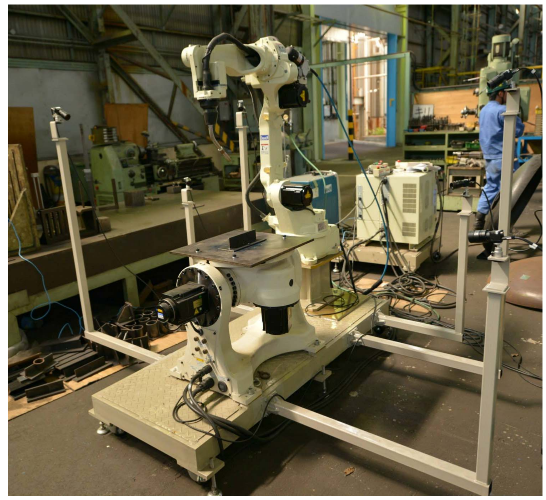

Figure 1: The multiple cameras are installed around working area of a welding robot.

2 Related Works

A lot of related methods have been proposed for estimating 3D position and orientation from images. One approach is using an interface that requires the operator to specify the motions of the robots onto 2D camera image. Hashimoto et al. proposed a method called TouchMe that allows the operator to manipulate each part of the robot by touching it on the camera image looking at the robot from a third-person view [1]. This interface overlays a virtual CG robot onto the actual robot in the camera image. The actual robot follows the virtual robot that is manipulated by the operator. Sakamoto et al. developed Sketch and Run, an interface for controlling the robot by sketching the robot's future trajectory with a touch display [2]. Interfaces for controlling via a first-person view (from the robot's eye-view) were also proposed. Sekimoto et al. proposed a simple driving interface for the mobile robot using a touch panel and first-person view images [3]. The system generates the path of the robot from the 2D point that the operator touches on the camera image. Okabe et al. expanded this interface to a multi-DOF (Degrees of Freedom) robot [4]. They proposed a tele-operation system for the rescue robot by using a touch panel and a camera that is mounted on the rescue robot arm. These systems focus on the operator's operability in moving the robot. Moreover, these systems have constraints for specifying 3D position and orientation of the robot because they specify the positions using only image from single camera.

There has been many other proposed methods in order to convey the 3D position and orientation. One approach is

978-1-4673-7660-0/15 $31.00 © 2015 IEEE
DOI 10.1109/ISMAR.2015.38

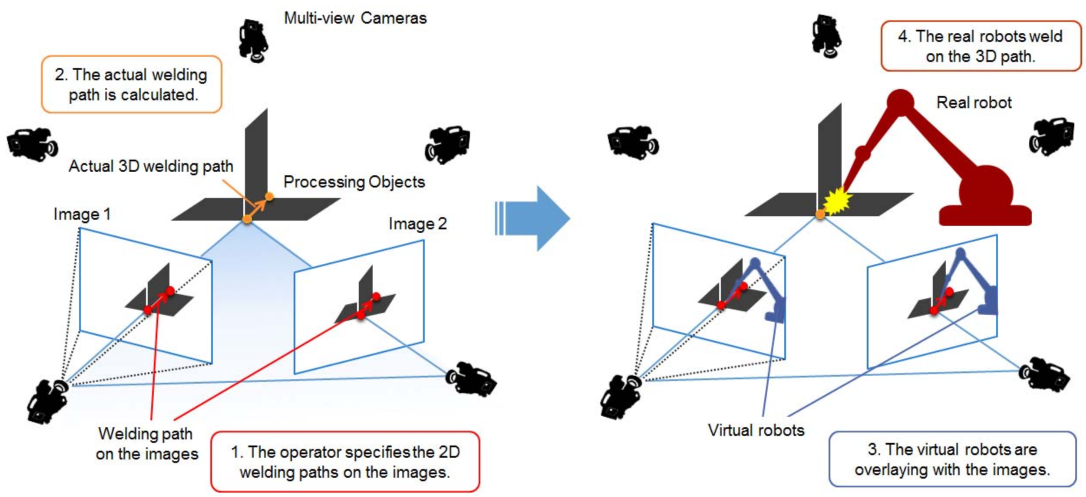

Figure 2: System overview.

reconstructing the model of the robot and the processing objects from camera images and sensor data. Liu et al. reconstructed the condition surrounding a vehicle via bird's-eye view images that are integrated by deforming images from multiple fish-eye cameras mounted on the vehicle [5]. Sato et al. reconstructed a wider range of bird's-eye view images by storing the images from fish-eye cameras spatio-temporally [6]. However, the presentation of the models of the processing objects via bird's-eye view images is not suitable for welding which requires 3D operation of the robot.

On the other hand, several interfaces for controlling robots by reconstructing a 3D model of the robot and processed objects are proposed. Maeda et al. proposed an interface that reconstructs the objects around the robot by using RGB-D camera and overlays the future positions of the robot onto the camera image [7]. This interface predicts whether the robot collides with the objects or not in the future when it moves along the current trajectory. Okura et al. proposed a virtual free-view point robot control system by unifying multiple depth images [8]. However, it takes time to reconstruct the 3D model of the objects and the accuracy of processing depends on the accuracy of the reconstruction. Moreover, this approach has to reconstruct the whole environment again when the processing object has changed.

3 System Overview

We propose a system for welding robots that allows the operators to specify the 3D position of the welding tip of the robot intuitively with high precision by using multiple cameras. Our system focuses on objects that will be processed. Our approach is able to check the operation of the robot immediately for any kind of processing object by specifying the welding path on multi-view images and by checking the movement of the robot by using the virtual robot.

Figure 2 shows an overview of our system. In our system, multiple cameras are installed around the working area of the welding robot. The system receives videos from multi-viewpoints. After the operator specifies the 2D path to weld on the captured videos by clicking on the image, the system then transforms the 2D path into a 3D path. The system imports the prepared 3D model of the robot and generates the CG animation that simulates the robot welding along the calculated path. The animation generated is overlaid with the current image that the operator selects. After the operator checks the movement of the real robot by using the virtual robot, the real robot welds the 3D path.

4 Implementation

Our system is implemented by the following steps: (1) Multi-view camera calibration, (2) 3D path calculation, (3) Generating the animation of a virtual robot, (4) Overlaying the animation with the images.

4.1 Multi-view camera calibration

First, the internal parameters of each camera are calculated by using a specified checker pattern while changing its angles and distances from the camera.

Then, an operator moves the welding tip wearing a marker to the known spatial coordinates. External parameters (position and rotation) of each camera are calculated by detecting the marker position on each image. We are able to calibrate it by using the parameters for the transformation between the two-dimensional coordinates of multi-view images and the three-dimensional coordinates that robot is placed on. In the current setup, the operator specifies the marker of the tip by clicking on each image. An interface for specifying the marker is the same that is referred in Section 4.2. As a result, the system gets a pair between the 3D coordinates of the welding tip and the 2D position in the image. The projection matrix is able to calculate this by using at least 6 pairs. This calibration process is carried out only once after the cameras are fixed around the robot.

4.2 3D path calculation

To receive the 3D coordinates of the path, the system calculates the 3D position of the robot's coordinate from the clicked 2D image coordinates. Figure 3 is an interface for specifying and calculating the 3D path on images. After the calibration, the system captures multi-view images of the objects. By clicking on the welding position on the objects in each image, the 2D coordinates are acquired. The operator can select the viewpoint by switching the tabs on the left side of the window. If the 2D coordinates are obtained in at least two images, it is possible to acquire the 3D position of the welding position in the robot coordinate system by using the calibrated projection matrix. To reduce the operator's burden and the probability of error, the interface has functions such as changing the brightness of the images and zooming on a specific part of the image.

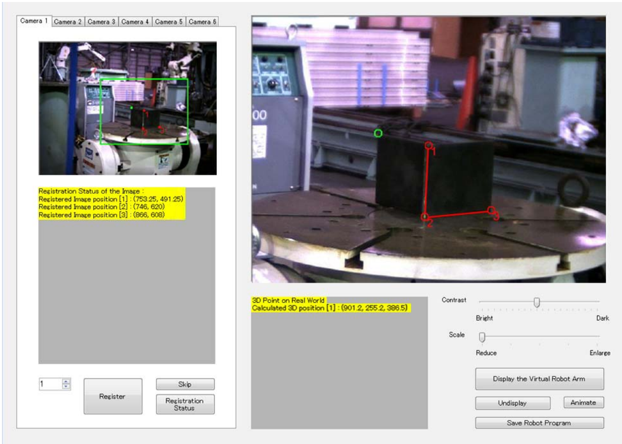

Figure 3: An interface for specifying and calculating the welding path.

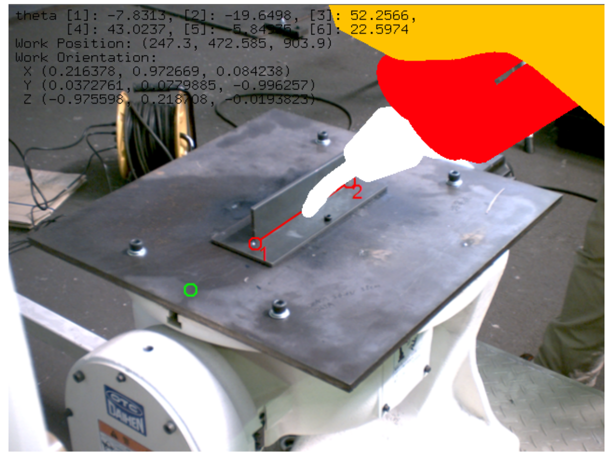

Figure 4: The virtual robot is overlaid with the captured image. The welding tip of the virtual robot moves on the path.

4.3 Generating the animation of a virtual robot

Six joint angles of the robot arm are calculated by inverse kinematics (IK) using the coordinate of the welding tip on the path. Continuous Path (CP) control method is selected for the interpolation method of IK because it considers the trajectory with which the robot draws from the current position to the final position.

4.4 Overlaying the animation with the images

To overlay the robot CG animation to the captured image, it is necessary to arrange the virtual camera, the captured image, and the robot CG at the correct position and orientation on the 3D space. The position and the orientation of the virtual camera are specified by the camera parameters obtained in section 4.1. The position of the captured image is calculated by the optical axis direction vector of the camera. The image is arranged at a position that is parallel to this vector and far enough in front of the camera. The normal vector of the image is set to be parallel to the optical axial vector. It is possible to display the image that fits the size of the viewport by equalizing the camera angle of the virtual camera and the real camera. The virtual robot is placed in accordance with the data obtained in section 4.3. Figure 4 shows a state when the virtual robot is overlaid with the camera image.

Table 1: The mean and standard deviation of the re-projection error in each camera.

Camera No.	Mean (px)	Standard Deviation (px)
1	0.413	0.202
2	0.533	0.429
3	0.435	0.268
4	0.661	0.475
5	0.461	0.134
6	0.584	0.325
Ave.	0.392	0.209

5 HARDWARE

We used 6 cameras (Flea 3 FL-U3-13E4C-C) for obtaining the multi-view images. The cameras are able to control the capture timing of the videos for synchronizing the videos between multiple cameras. Also, the cameras are equipped with a USB 3.0 interface that enables high-speed image transmission. For these reasons, these cameras are suitable for high-speed capturing for overlaying the virtual robot and checking the state of the robot. The resolutions of the cameras are all 1280×960px.

We used Daihen RT3500H as the welding robot. Accuracy required for the direct welding was within 1mm. However, if the accuracy is within 3mm, we can use the touch sensor that is mounted on the welding robot to detect and correct the deviation of the welding position. We used OpenGL as the CG library.

6 EXPERIMENT

We conducted two experiments to evaluate our system.

6.1 Evaluation of the camera calibration

First, we calculated the re-projection error using the projection matrix to evaluate the accuracy of the camera calibration. The 3D points which the coordinates are known were re-projected to the camera coordinate and we compared those points with the 2D points which were set as ground truth. We used 12 sets of 2D and 3D points of the robot arm for the ground truth: the sets are created based on the images that are used for the calculation of the extrinsic parameter of the camera.

From the result shown in Table 1, the re-projection was done in a mean of 0.392 pixels and the standard deviation of 0.209 pixels. Even when the error is at its maximum, it is within 1.2 pixels. The error was small enough in contrast to the camera resolution so it can be seen that the camera calibration was done precisely.

6.2 Evaluation of the calculated 3D position

Next, we conducted an experiment to evaluate the error between the actual 3D position and the 3D position calculated from the multi-view images. This error indicates how accurate the operator specifies the points using our system. In order to eliminate the error of the camera calibration and only evaluate the error of the points specified, we used the same 12 sets of points that were re-projected in the previous experiment for the ground truth.

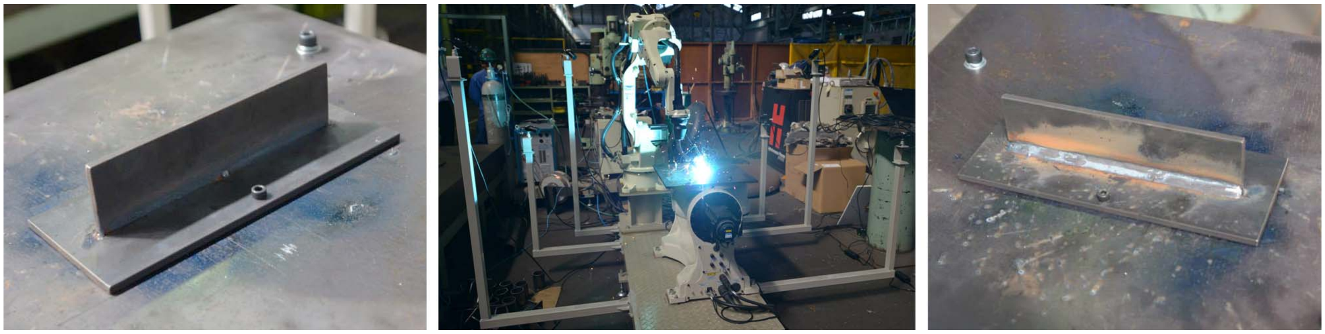

Figure 5: Left: Objects before welding. Two pieces are perpendicular to each other. These are tack welded at the start, the center and the end of welding path; Center: The robot is welding objects with our calculated path; Right: The welded object.

For each set, all points from the 6 images are shifted in 4 directions by a specific distance from 1 pixel to 6 pixel at an interval of 1px, and calculated the 3D point. In other words, we calculated 49,152 sets of 3D point for each deviation (4 directions to the 6^{th} power of the camera images $\times$ 12 points). Using the set of points chosen, we calculated the 3D points and compared the Euclidean distance with the ground truth.

Figure 6 shows the mean errors, the standard deviations and the max errors of the calculated 3D positions when the specified points on the images are shifted. From the result, it can be seen that the error will be always within 1mm when the operator specified the point with 1 pixel in accuracy. Moreover, the probabilities of the 3D points being within 3mm when the specified points of the camera image are shifted by 4~6 pixels were 99.54%, 94.97% and 84.21% each. From this result, it can be said that the 3D points calculated will be mostly within 3mm when the operator specified the point within 4 pixels of error.

When the position the operator clicked is displaced more than 1px, our result is not able to weld the objects directly. However, our system allows the operator 4px of range to specify a point on a camera image by using the touch sensor to correct the deviation. Also, our system can zoom the image up to 8 times, so it allows 32px of range for the operator to specify points. Therefore, by combining our system with the touch sensor we can weld objects at the correct position. By increasing the number of cameras and improving the method of camera calibration, the accuracy of the system will become higher and is able to weld without the touch sensor. Figure 5 shows the objects welded by using our system and touch sensor.

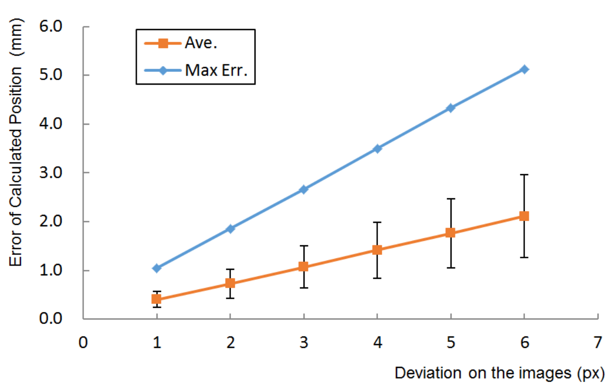

Figure 6: The mean errors, the standard deviations and the max errors of the calculated points when the specified points on the images are shifted from 1px to 6px.

7 Conclusion

This paper proposes a system which allows non-expert operators to weld objects remotely and intuitively by using multi-view images. After the operator specifies a two-dimensional path on the images, the system transforms the two-dimensional path into a three-dimensional path. The system also generates the animation of a virtual robot tracing along the path, and overlays it with captured images for predicting the movement of the real robot. The accuracy of our system is enough to weld objects by combining with the touch sensor.

For future works, we plan to use a tablet display and touch pens for more intuitive designation of the welding positions. Moreover, we plan to paste retro-reflective markers on objects and use IR cameras to recognize welding paths automatically.

References

[1] S. Hashimoto, A. Ishida, M. Inami, T. Igarashi. Touchme: An augmented reality based remote robot manipulation. In *Proceeding of the ICAT2011*, pp.28-30, 2011.

[2] D. Sakamoto, K. Honda. M. Inami, T. Igarashi. Sketch and run: a stroke-based interface for home robots. In *Proceeding of the CHI'09*, pp.197-200, 2009.

[3] T. Sekinoto, T. Tsubouchi and S. Yuta. A Simple Driving Device for a Vehicle Implementation and Evaluation. In *proceedings of the IROS'97*, pp. 147-154, 1997.

[4] D. Okabe, N. Sato, Y. Morita. Tele-operation system for rescue robot by inputting target position of end-effector. In *System Integration (SII) 2013 IEEE/SICE International Symposium on*, pp. 861-866. 2013.

[5] Y.C. Liu, K.Y. Lin and Y.S. Chen. Bird's-eye view vision system for vehicle surrounding monitoring. In *Robot Vision Lecture Notes in Computer Science*, Vol. 4931, pp. 207-218, 2008.

[6] T. Sato, A. Moro, A. Sugahara, T. Tasaki, A. Yamashita, and H. Asama. Spatio-temporal bird's-eye view images using multiple fish-eye cameras. In *System Integration (SII) 2013 IEEE/SICE International Symposium on*, pp. 753-758, 2013.

[7] N. Maeda and J. Morita and M. Sugimoto. Pathfinder Vision: Tele-operation Robot Interface in Consideration of Geometry for Supporting Future Prediction. In *Proceeding of the ICAT2014*, pp.29-35, 2014.

[8] F. Okura, Y. Ueda, T. Sato, N. Yokoya. Teleoperation of mobile robots by generating augmented free-viewpoint images. In *Intelligent Robots and System (IROS), 2013 IEEE/RSJ International Conference on (2013)*, pp. 665-671, 2013.