An intelligent real time 3D vision system for robotic welding tasks

RODRIGUES, Marcos <http://orcid.org/0000-0002-6083-1303>, KORMANN, Mariza, SCHUHLER, C and TOMEK, P

Published version

RODRIGUES, Marcos, KORMANN, Mariza, SCHUHLER, C and TOMEK, P (2013). An intelligent real time 3D vision system for robotic welding tasks. In: Mechatronics and its applications. IEEE Xplore, 1-6.

MARWIN: An Intelligent Real Time 3D Vision System for Robotic Welding Tasks*

Marcos Rodrigues[1], Mariza Kormann[1], Clement Schuhler[2], Geoff Melton[2], Jakub Shejbal[3] and Peter Tomek[3]

***Abstract*— MARWIN is an intelligent system for automatic robotic welding tasks. It extracts welding parameters and calculates robot trajectories directly from CAD models which are then verified by real-time 3D scanning and registration. The focus of this paper is on describing a novel mathematical formulation for structured light scanning together with the design and testing of the 3D vision system and show how such technology can be exploited within an anthropomatic context. The expected end result is a 3D assisted user-centred robot environment in which a task is specified by the user by simply confirming (and/or adjusting) MARWIN's suggested parameters and welding sequences.**

I. INTRODUCTION

Welding by robots has experienced a vigorous upsurge in recent years with an estimated 25% of all industrial robots being used in connection to welding tasks [1]. The challenge is to develop flexible automation systems that can be set up quickly and can be switched over to another product line while maintaining quality and profitability. Small and medium enterprises (SMEs) normally do not have the resources to invest in technology requiring extensive human training. The MARWIN project offers a solution to human-robot interaction by developing a cognitive welding robot where welding tasks and parameters are intuitively selected by the end-user directly from a library of CAD models. Robot trajectories are then automatically calculated from the CAD models and validated through fast 3D scanning of the welding scene. The role of the user is limited to high level specification of the welding task and to the confirmation and/or changing of welding parameters and sequences as suggested by MARWIN.

This paper focuses on describing the MARWIN 3D vision system and a novel mathematical formulation for fast 3D scanning using structured light, and on methods to register scanned surfaces to CAD models and estimation of registration errors. The main idea behind the 3D vision system is that, if the scanned model matches the description of their CAD counterparts, then welding can proceed as calculated by MARWIN. Similarly, after the welding task is completed, a new scanning can be used for welding quality control.

*This work is supported by the European Commission, MARWIN Project Grant 286284 *Research for SMEs – SME-2011-1*, from Nov 2011 to Oct 2013.

[1]Marcos Rodrigues and Mariza Kormann are with the Geometric Modelling and Pattern Recognition Group, Communication and Computing Research Centre, Sheffield Hallam University, UK, {m.rodrigues, m.kormann} at shu.ac.uk

[2]Clement Schuhler and Geoff Melton are with TWI Cambridge, UK, {clement.schuhler, geoff.melton} at twi.co.uk

[3]Jakub Shejbal and Peter Tomek are with MFKK, Budapest, Hungary, {jakub.shejbal, peter.tomek} at mfkk.hu

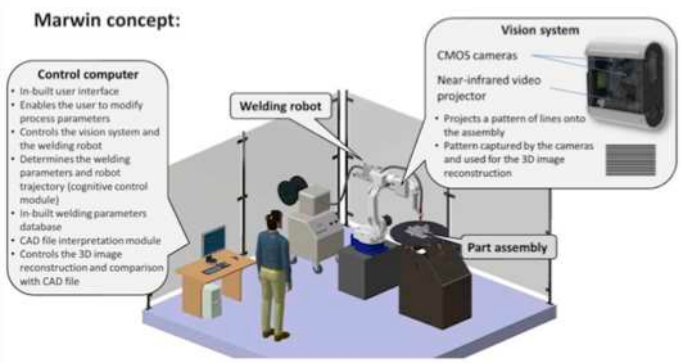

Fig. 1. The MARWIN concept for intelligent robotic welding

II. THE MARWIN 3D VISION SYSTEM

The principle of operation of the MARWIN vision system is to project multiple planes of light onto the target surface whose image is recorded by a camera. The shape of the captured pattern is combined with the spatial relationship between the light source and the camera, to determine the 3D position of the surface along the pattern. This is known as structured light scanning and is illustrated in Figure 2. Structured light scanning is an active method of surface reconstruction that relies on the projection of a known pattern of light [2], [3], [4], [5], [6], [7]. The main advantages of the method described here are speed and accuracy; a surface is scanned from a single 2D image and processed into 3D in 40ms. Structured light scanners in the literature are configured such that the normals from the centre of the projector and the centre of the camera sensor meet at the calibration plane with a number of methods being proposed for reliable pattern detection [8], [9], [10], [11], [12]. In this paper we further our previous work on fast 3D reconstruction using structured light [5], [6], [13], [14], [15], [16], [17], [18], [19] by presenting a novel configuration and corresponding mathematical formulation where the light source and camera sensor are parallel to each other [20] as shown in Figure 3.

The advantages of such system are that both camera and projector can be placed as close together as practically possible which may offer advantages to design miniaturisation.

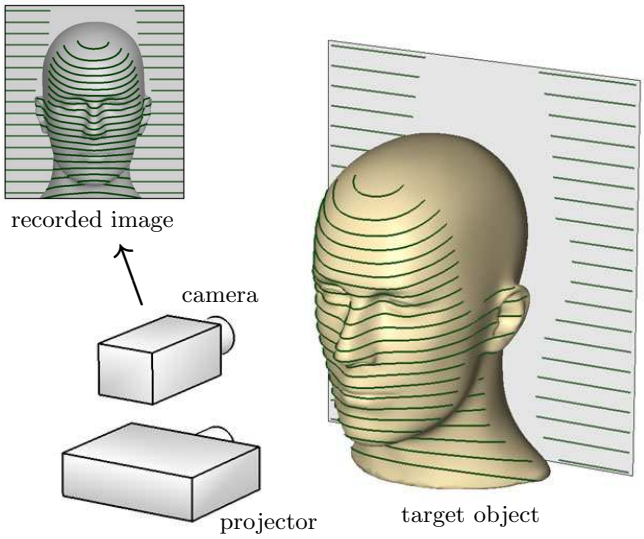

Fig. 2. The MARWIN vision system projects multiple, simultaneous light planes onto the target surface

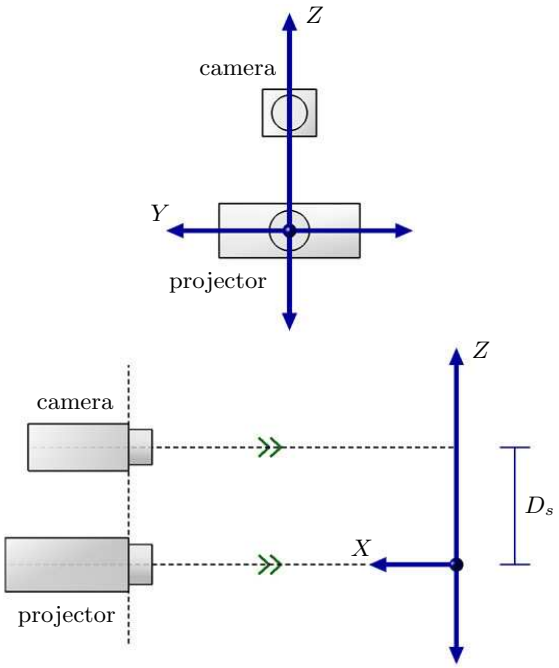

Fig. 3. A novel parallel arrangement of camera and projector

Moreover, the mathematical formulation of such arrangement is simpler than of those of standard scanners which results in less computing cycles thus, making the parallel design more appropriate for 3D real-time processing.

In order to proceed to the mathematical formulation, first it is necessary to define the coordinate system. We choose this to be in relation to the light source as shown in Figure 4. The problem we are trying to solve is defined as follows. Every pixel in the image as captured by the camera needs to be mapped into 3D to the chosen world coordinate system, also referred to as the system space. This is solved by determining to which light plane or stripe that pixel belongs to, and then through trigonometric relationships find the coordinates of the surface point imaged by that pixel. Such mapping can be determined with the help of Figure 5 where the parameters are defined as follows.

- D_s : the distance between the camera and the projector
- D_p : the distance between the projector and the system origin
- W : the width between successive light planes (or stripes) in the calibration plane
- P : the pixel size in the sensor plane of the camera
- F : the camera focal distance
- v, h : vertical and horizontal position of the pixel in camera space
- x, y, z ; 3D coordinates of a surface point
- n : stripe index from the light source
- θ, ϕ : angle between a light plane and its location in camera space

From trigonometric similarity ratios in Figure 5 we can derive the following expressions:

$$\frac{z}{D_p - x} = \frac{W_n}{D_p} \Rightarrow z = \frac{W_n}{D_p}(D_p - x) \qquad (1)$$

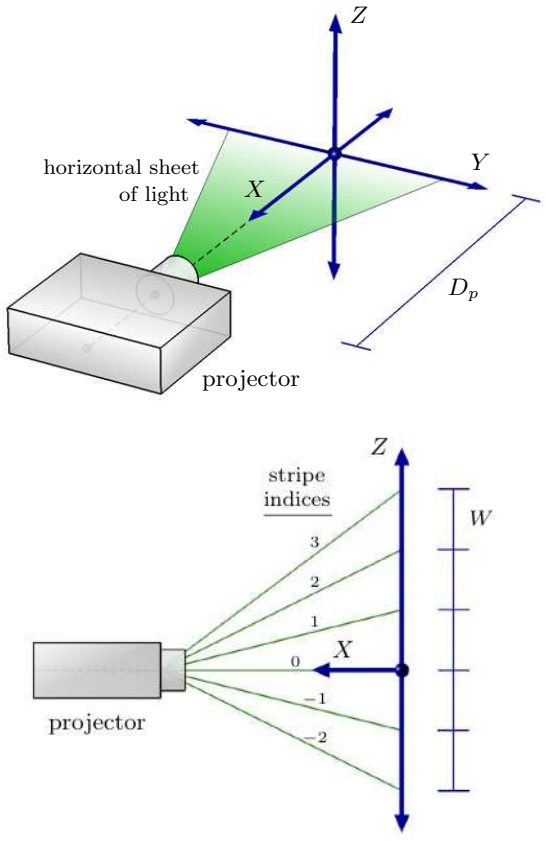

Fig. 4. A coordinate system is defined in relation to the light source which also defines the indices of the light planes

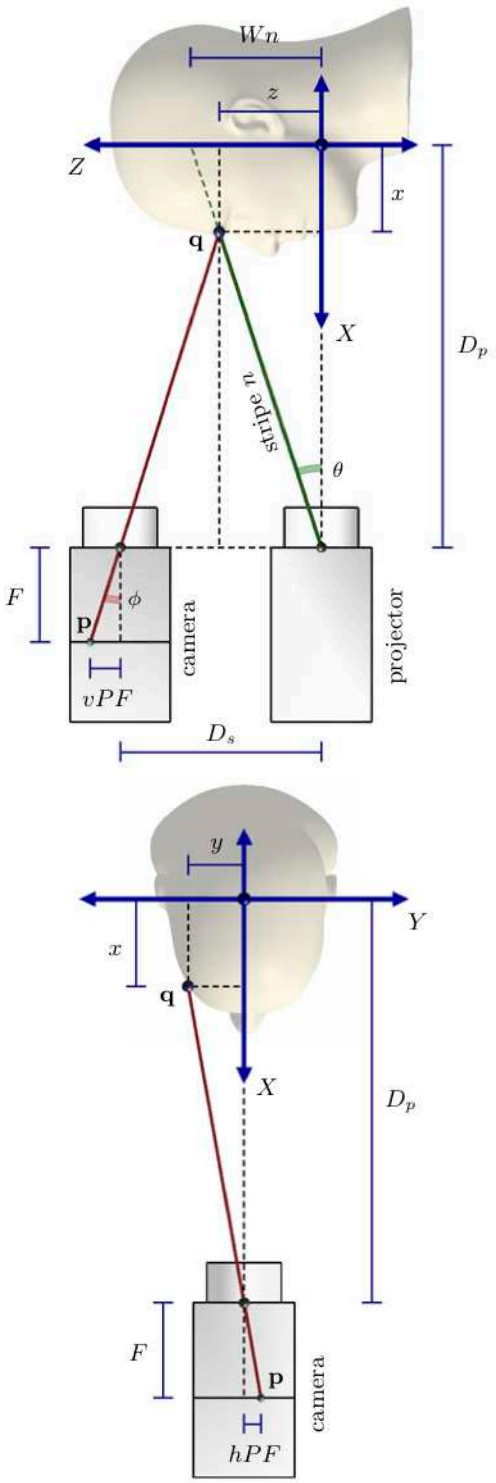

Fig. 5. Mapping (h, v, n) in camera space to (x, y, z) in system space

and

$$\frac{vPF}{F} = \frac{D_s - z}{D_p - x} \Rightarrow z = D_s - vP(D_p - x) \tag{2}$$

We note that by construction, $D_p > x, D_p > 0$ and $F > 0$. Combining these expressions we obtain:

$$\frac{W_n}{D_p}(D_p - x) = D_s - vP(D_p - x) \tag{3}$$

thus,

$$x = D_p - \frac{D_p D_s}{vPD_p + W_n} \tag{4}$$

$$z = \frac{W_n D_s}{vPD_p + W_n} \tag{5}$$

$$y = hP(D_p - x) = \frac{hPD_p D_s}{vPD_p + W_n} \tag{6}$$

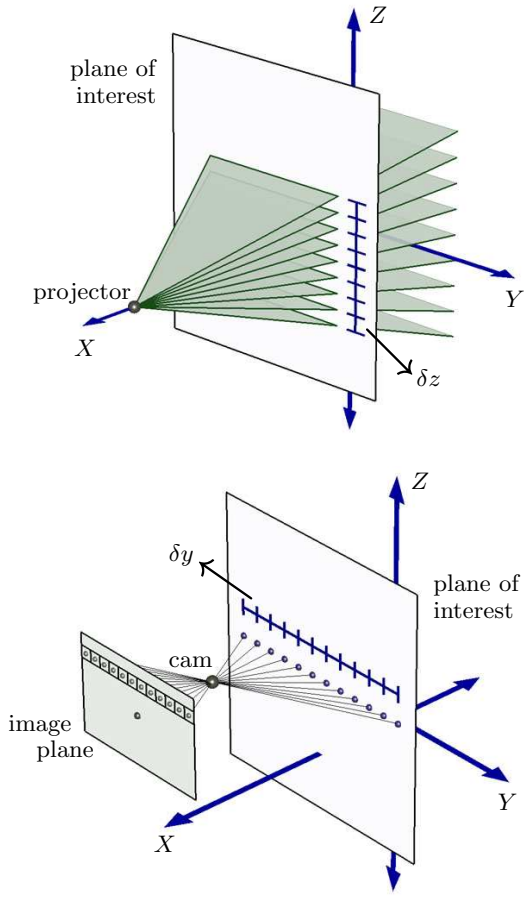

Fig. 6. Resolution depends on pixel size, stripe spacing and distance from the image plane

The resolution in 3D space (in millimetres) depends on the spacing between the projected planes, on the distance between the surface and the light source and the dimension of the pixel space as shown in Figure 6. The vertical resolution (Figure 6 top) is affected by the spacing of the projected stripes. The horizontal resolution (bottom) is dependent on

the horizontal dimension of the pixel space. The resolution (δ) along the (x, y, z) dimensions are given by:

$$\delta z = \frac{W}{D_p}(D_p - x) \tag{7}$$

$$\delta y = P(D_p - x) \tag{8}$$

$$\delta x = \frac{PD_p^2 D_s}{[vPD_p + W_n][(v+1)PD_p + Wn]} \tag{9}$$

In order to reconstruct a surface in 3D the following steps are required: locate stripes in the 2D image, index the stripes then map to 3D space. Stripe pixels are determined as local maxima in the greyscale values of every column in the image as depicted in Figure 7 (top). Once the stripes are detected, algorithms are run (e.g. [5], [6] based on maximum spanning tree and flood filling) to correctly index those, where the stripe with index 0 (zero) corresponds to the light plane emanating from the centre of the projector (the reference projected light plane). With the correct indices noted, the mapping given by (4), (5) and (6) is then applied resulting in a point cloud of vertices in 3D space. A wire mesh model can be easily defined as the vertices along each adjacent stripe define the intrinsic structure for either triangulated or quad meshes, so there is no need for demanding triangulated algorithms such as Delauney triangulation. A quad-edge mesh is represented at the bottom of Figure 7.

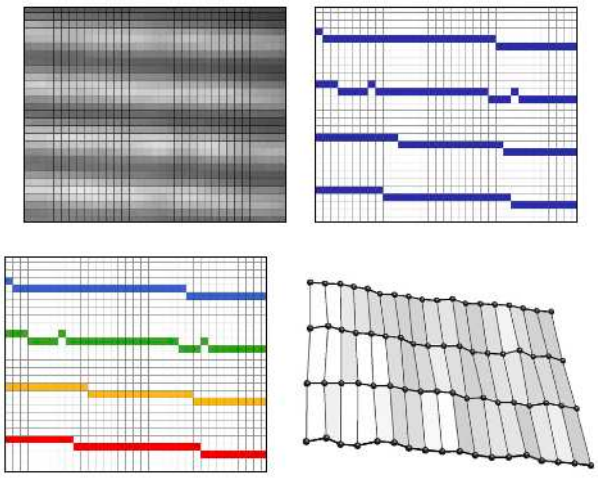

Fig. 7. Stripes are determined as local maxima, then indexed and converted into 3D point cloud and quad-edge mesh

III. Experimental Results

A. 3D Reconstruction

We developed a scanner prototype with an internal adjustable mechanism where the top camera is aligned parallel with the light source as depicted in Figure 8. The prototype uses IDS UI-1241LE camera boards [21] and a Microvision PicoP laser projector [22]. By construction, the light source at the centre of the projector is vertically aligned with the centre of the camera sensor. The parallel configuration means that the normals to the camera sensor and to the projector

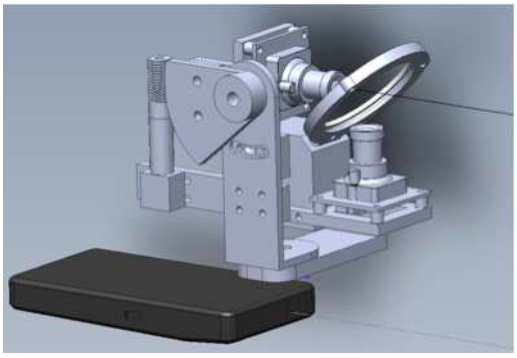

Fig. 8. The scanner assembly design with cameras and projector

lens are also normal to the calibration plane. This plane is set to 300mm from the centre of the projector in the current prototype, but this is obviously adjustable to any desired distance between 200–800mm given the properties of the projector. It is important to stress here that the calibration plane does not set the maximum distance an object can be imaged. The limitation is due to the brightness of the projector and the reflective properties of the surface being scanned; if the stripe pattern is detectable in the scene, then 3D reconstruction is possible, otherwise it will fail.

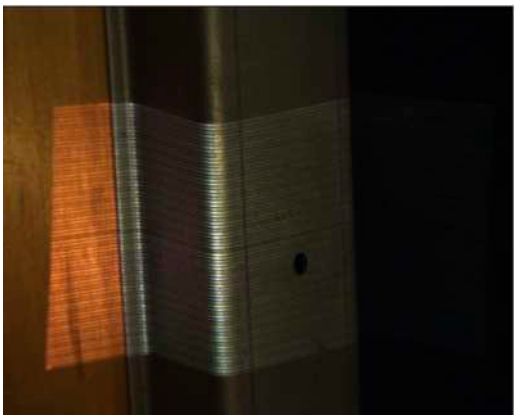

Fig. 9. 2D image of a welding part as seen by the camera

A welded part as seen by the image sensor is depicted in Figure 9. The light planes emanating from the projector appear as stripes in the scene. Note that to the naked eye stripes are barely visible; however, our stripe detection algorithms [5], [6] can detect each and every stripe reliably. After each stripe is assigned an index, each pixel lying on a stripe can be converted into a 3D surface point using (4), (5) and (6).

Several views of the reconstructed part are shown in Figure 10 (top). Given that the location in 2D of each reconstructed pixel is known, texture mapping is straightforward as shown on the top left and it can be turned on and off as required. A detailed view is shown at the bottom of Figure 10. Note that this is a wire mesh but the density of the 3D reconstructed points is such that it looks like a shaded or solid model.

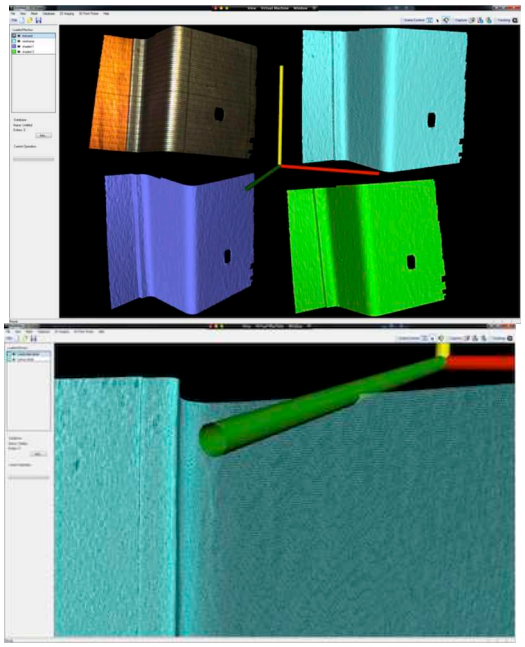

Fig. 10. Reconstructed part

In terms of overall processing time we can break down in several steps: 40ms for 2D image filtering (median and mean filters are used), 40ms for stripe detection and calculation of a point cloud in space, and 120ms for 3D mesh pos-processing (triangulation, small hole filling, mesh repairing and smoothing). Therefore, the current prototype can operate in real-time at about 5 frames per second. However, this is unlikely to be required for the task at hand given that the robot needs to position the scanner over the various surface patches in order to build a 3D model of the entire welding assembly, so an intermittent rather continuous 3D scanning is required.

The resolution in (x, y, z) given by (7), (8) and (9) normally refer to the detected peaks of each stripe as shown in Figure 7 meaning that the vertical resolution between stripes or the distance between vertices on consecutive stripes is 1.95mm for the current prototype. Along each stripe we get a higher horizontal resolution of 0.24mm as one vertex is processed for each pixel in the image. In order to increase the vertical resolution, we process the valleys (the darkest pixel of each dark stripe in Figure 7) effectively doubling the resolution and doubling it again by one step of mesh subdivision using a 4th order polynomial interpolation. The end effect is depicted by the high density mesh of Figure 10.

B. Registration with CAD Models

In order ensure correct calculation of robot trajectories for a given welding task, it is necessary to register the CAD model of the welding assembly to its scanned 3D model. This is to ensure that the real world scene matches its CAD model description. We use a variant of the ICP (Iterative Closest Point) estimation algorithm [23] with additional constraints such as checking the orientation and the surface normals of the data points in both frames. By iteratively finding the closest points in the second frame, a better transformation estimate is obtained. The closest points are found by calculating the Euclidean distances between a point $\mathbf{p}$ in the first frame (the CAD model) and a point $\mathbf{q}$ in the second frame (the scanned surface S) given by

$$d(\mathbf{p}, S_k) = \min_{j \in (1,\ldots,N_k)} d(\mathbf{p}, \mathbf{q}_{k,j}) \tag{10}$$

Equation (10) means that for every point in the CAD model we have to check every point in the scanned surface. Once the closest points are estimated, we should have two sets of points $\mathbf{p}_i$ and $\mathbf{q}_i$ where which are paired to each other. The following equation is used to minimise the distance between the two sets of points:

$$f(\mathbf{R}, \mathbf{t}) = \frac{1}{N} \sum_{i=1}^{N} ||\mathbf{R}\mathbf{p}_i + \mathbf{t}, \mathbf{q}_i||^2 \tag{11}$$

where $\mathbf{R}$ and $\mathbf{t}$ are the transformation rotation matrix and translation vector. A number of methods can be applied to perform this minimisation, here we used steepest descent. When the transformation model $(\mathbf{R}, \mathbf{t})$ has been estimated, transform every point in the CAD model. This iteration is repeated until convergence or when the transformation satisfies minimum set threshold. Here we have set a maximum of 25 iterations for the algorithm to converge.

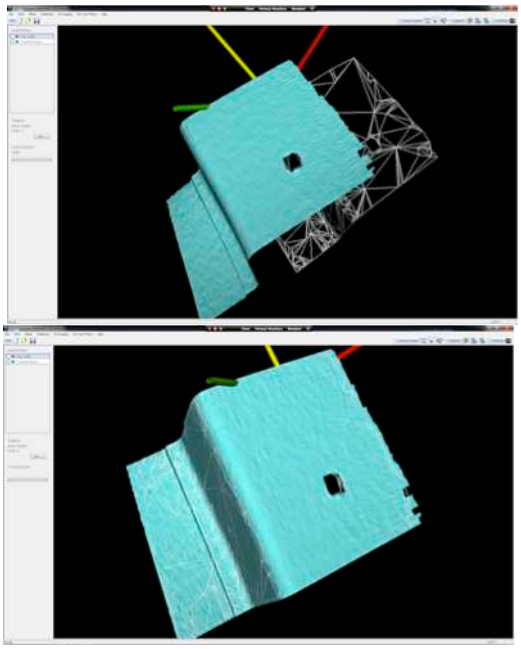

Fig. 11. Before and after registration

Results of the registration method as defined by (10) and (11) are depicted in Figure 11 where the top image shows the CAD model as a sparse wire mesh (white) and the scanned surface as a dense wire mesh (cyan) at their initial positions. The bottom image shows the end result where the CAD model closely agrees with their scanned counterpart with average mean square error of 0.3845. This means that, on average, each CAD model vertex is misplaced by just over the minimum mesh resolution which is a very close match.

The registration errors within prescribed thresholds mean that the various parts of the welding assembly (i.e. the scanned model) are correctly positioned and match their CAD model description. Furthermore, during the scanning process the scanner is attached to the robot wrist which means that we know exactly where the scanned model is in relation to the robot coordinate system. Given that the CAD model matches the scanned surfaces, we can proceed to the specification of the welding task. This allows the user to intuitively define the welding task negating the need for detailed programming as, after visual task specification (by selecting CAD models and parameters from a library) the user will take a more supervisory role. Thus, the ability to automatically calculate robot trajectories and welding sequences directly from CAD models satisfies the main aim of the MARWIN system: to provide a user-centred, cognitive system for robotic welding tasks.

IV. CONCLUSIONS

This paper describes research as part of the MARWIN project focusing on a novel mathematical formulation for structured light scanning and on demonstrating the effectiveness of the method both to fast 3D reconstruction and to the registration of acquired surfaces to their counterparts CAD models. By setting error thresholds for global registration, an automatic and objective measure is obtained of whether or not the scanned model (and thus the various assembly parts) are correctly positioned before welding starts. In the same way, a new scan and registration after the welding task is completed will also provide an objective measure of whether or not the welding assembly has been deformed or misplaced in any way during the welding operation and thus, can be used as post-welding quality control.

The significance of the MARWIN system is that it provides an intuitive, user-centred robotic assembly planning and control negating the need for detailed robot trajectory planning for welding sequences. The work described here represents a milestone in the project and the automatic task specification through CAD models and their integration into a full working system will be described in a follow on paper.

ACKNOWLEDGMENTS

We acknowledge the constructive comments and suggestions received from the MARWIN partners Rolan Robotics BV (The Netherlands), Recamlaser SL (Spain), Nemetschek ODD (Bulgaria), and G-Robots Szolgaltato Es Kereskedelmi KFT (Hungary). We also acknowledge Willie Brink (Stellenbosh University, South Africa) for his work on geometry.

REFERENCES

[1] European Commssion Report, 2003. “Smart, Sustainable Manufacturing: EU Research Sparks Revolution in Machine Tools, Robots and Automation”, Research and Innovation Report, Brussels, 2003.

[2] Mirota, D.J., Ishii, M., Hager, G.D., 2011. “Vision-based navigation in image-guided interventions”, Annu Rev Biomed Eng 13, 297319.

[3] Clancy, N.T., Stoyanov, D., Groch, A., Maier-Hein, L., Yang, G.Z., Elson, D.S., 2011. “Spectrally-encoded fibre-based structured lighting probe for intraoperative 3D imaging”, Biomedical Optics Express 2, 31193128.

[4] Mountney, P., Stoyanov, D., Yang, G.Z., 2010. “Three-dimensional tissue deformation recovery and tracking”, IEEE Signal Proc Mag 27, 1424.

[5] W. Brink, A. Robinson, M. A. Rodrigues, “Indexing Uncoded Stripe Patterns in Structured Light Systems by Maximum Spanning Trees”, Proc BMVC British Machine Vision Conference 2008, www.bmva.org/bmvc/2008/papers/230.pdf, ISBN 978-1-901725-36-0

[6] A. Robinson, L. Alboul and M. Rodrigues, “Methods for Indexing Stripes in Uncoded Structured Light Scanning Systems”, Journal of WSCG, Vol.12, No.1-3, ISSN 1213-6972.

[7] J. Salvi, J. Pagès , J. Batlle, “Pattern codification strategies in structured light systems”, Pattern Recognition Volume 37, Issue 4, April 2004, Pages 827849.

[8] Gorthi, S., Rastogi, P., 2010, “Fringe projection techniques: Whither we are?” Opt Laser Eng 2, 133140.

[9] Chen, S., Li, Y., 2008, “Vision processing for realtime 3D data acquisition based on coded structured light”, IEEE T Image Process 17, 167176.

[10] Kawasaki, H., Furukawa, R., Sagawa, R., Yasushi, Y., 2008, “Dynamic scene shape reconstruction using a single structured light pattern”, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1 8.

[11] Albitar, C., Graebling, P., Doignon, C., 2007, “Robust structured light coding for 3D reconstruction”, in: International Conference on Computer Vision (ICCV), pp. 16.

[12] Pavlidis, G., Koutsoudis, A., Arnaoutoglou, F., Tsioukas, V., Chamzas, C., 2007, “Methods for 3D digitization of cultural heritage”, Journal of Cultural Heritage 8, 9398.

[13] L. Maier-Heina, P. Mountneyb, A. Bartolic, H. Elhawaryd, D. Elsone, A. Grocha, A. Kolbf, M. Rodriguesg, J. Sorgerh, S. Speideli, D. Stoyanovj, “Optical Techniques for 3D Surface Reconstruction in Computer-Assisted Laparoscopic Surgery” submitted to Medical Image Analys, 2012.

[14] M. Rodrigues, A. Osman and A. Robinson,“Partial Differential Equations for 3D Data Compression and Reconstrution”, Int Conf on Differential Equations, Difference Equations and Special Functions, Patras, Greece 3-7 Sep 2012, to be published in Advances in Dynamical Systems and Applications.

[15] M. Rodrigues and A. Robinson, 2011, “Real-time 3D Face Recognition using Line Projection and Mesh Sampling”, In: EG 3DOR 2011 - Eurographics 2011 Workshop on 3D Object Retrieval, Llandudno, UK, 10th April 2011. Eurographics Association.

[16] M. Rodrigues, A. Robinson and A. Osman, 2011, “Efficient 3D data compression through parameterization of free-form surface patches”, In: Signal Process and Multimedia Applications (SIGMAP), Proceedings of the 2010 International Conference on. IEEE, 130-135.

[17] M. Rodrigues and A. Robinson, 2011, “Fast 3D recognition for forensics and counter-terrorism applications”, In: AKHGAR, Babak and YATES, Simeon, (eds.) Intelligence management : knowledge driven frameworks for combating terrorism and organized crime. Advanced information and knowledge processing . London, Springer-Verlag, 95-109.

[18] M. Rodrigues and A. Robinson, 2010, “Novel methods for real-time 3D facial recognition”, In: SARRAFZADEH, Majid and PETRATOS, Panagiotis, (eds.) Strategic Advantage of Computing Information Systems in Enterprise Management. Athens, Greece, ATINER, 169-180.

[19] M. Rodrigues, A. Robinson and W. Brink, 2008, “Fast 3D reconstruction and recognition”, In: 8th WSEAS International Conference on Signal Processing, Computational Geometry and Artificial Vision, Rhodes, Greece, 20-22 August 2008. 15-21.

[20] W. Brink, “3D face scanning and alignment for biometric systems”, PhD dissertation, Sheffield Hallam University (UK), 2008.

[21] IDS Imaging GmbH, “USB2 uEye LE Camera Boards”, www.ids-imaging.com

[22] Microvision Inc, “MicroVision SHOWWX+ Laser Pico Projector”, microvision.com

[23] P. Besl and N. McKay, “A method for Registration of 3-D Shapes”, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 14(2):239 - 256, February 1992.