

Structured light techniques for 3D surface reconstruction in robotic tasks

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

Available from Sheffield Hallam University Research Archive (SHURA) at:

<https://shura.shu.ac.uk/7280/>

This document is the Accepted Version [AM]

Citation:

RODRIGUES, Marcos, KORMANN, Mariza, SCHUHLER, C and TOMEK, P (2013). Structured light techniques for 3D surface reconstruction in robotic tasks. In: KACPRZYK, J, (ed.) Advances in Intelligent Systems and Computing. Heidelberg, Springer, 805-814. [Book Section]

Copyright and re-use policy

See <http://shura.shu.ac.uk/information.html>

Structured Light Techniques for 3D Surface Reconstruction in Robotic Tasks

M.Rodrigues¹, M.Kormann¹, C.Schuhler², P.Tomek³

Abstract Robotic tasks such as navigation and path planning can be greatly enhanced by a vision system capable of providing depth perception from fast and accurate 3D surface reconstruction. Focused on robotic welding tasks we present a comparative analysis of a novel mathematical formulation for 3D surface reconstruction and discuss image processing requirements for reliable detection of patterns in the image. Models are presented for a parallel and angled configurations of light source and image sensor. It is shown that the parallel arrangement requires 35% fewer arithmetic operations to compute a point cloud in 3D being thus more appropriate for real-time applications. Experiments show that the technique is appropriate to scan a variety of surfaces and, in particular, the intended metallic parts for robotic welding tasks.

1 Introduction

One the main challenges in automated robotic tasks is the development of flexible systems that can be set up quickly and easily with minimum user intervention and can be switched over to another product line while maintaining quality and productivity. Small and medium enterprises normally find it difficult to invest in new technologies requiring expert knowledge and extensive human training. Addressing those issues, the MARWIN Project [8] (Decision Making and Augmented Reality Support for Automatic Welding Installations) funded by the EU offers a solution to human-robot interaction by developing a cognitive welding robot where welding tasks and parame-

¹GMPR – Geometric Modelling and Pattern Recognition Group, Sheffield Hallam University, UK, {m.rodriques, m.kormann}@shu.ac.uk · ²TWI – The Welding Institute, Cambridge, UK, clement.schuhler@twi.co.uk · ³MFKK – Invention and Research Center Services, Budapest, Hungary, peter.tomek@mfkk.hu

ters 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 or changing of welding parameters and sequences as suggested by MARWIN.

This paper focuses on describing optical techniques for fast 3D reconstruction including image processing methods and a comparative analysis of sensor configurations concerning mathematical formulation and computational requirements. Optical methods for 3D surface reconstruction can be divided into two categories: passive methods that only require images from fairly uncontrolled illumination, and active methods that require controlled patterns of light to be projected onto the scene [10]. Passive methods include stereo vision, Shape-from-X (SfX) and Simultaneous Localization and Mapping (SLAM). Active methods include Time-of-Flight (ToF) and Structured Light where surface patches are reconstructed from the geometric relationships existing between the source of light and sensor devices. Our research is focused on structured light techniques using coded and uncoded light patterns (e.g. [2, 15]). The main advantages of our techniques are speed and accuracy as a surface is scanned and reconstructed from a single 2D image in 40ms.

Structured light techniques can take many forms; a review of current techniques is provided in [11]. Salvi *et al.* [16] have proposed a classification of structured patterns based on their coding strategy. For instance, we can identify recent techniques such as the Kinect Box based on projecting dot patterns [17], projection of a single pattern of uncoded stripes [13], coded stripes [15], colour patterns such as in [4], and alternating stripe patterns or phase shifting [18]. In general, structured light scanners in the literature assume an angled configuration in which 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 (e.g. [1, 3, 5, 7, 12]).

In this paper we further our previous work on fast 3D reconstruction using structured light [14] by comparing a novel configuration and corresponding mathematical formulation where the light source and camera sensor are parallel to each other against a standard angled configuration. Section 2 presents the mathematical formulation of both sensor configurations, Section 3 discusses image processing functions, Section 4 presents experimental results and a conclusion is presented in Section 5.

2 The Geometry of Structured Light

The principle of operation of the MARWIN structured light scanner 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.

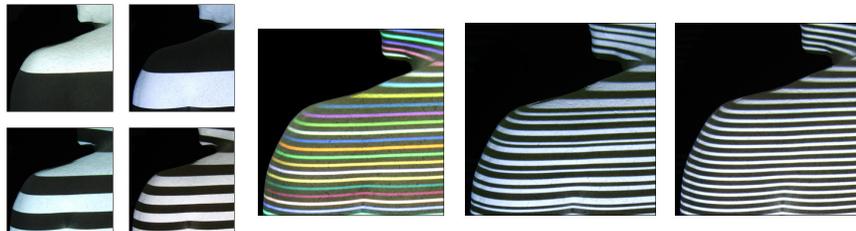


Fig. 1: Examples of coding schemes. From left to right: time-multiplexing, colour coding, variable width, and uncoded stripes.

The mapping from a captured image to a 3D surface requires knowledge of correspondences between captured patterns and projected elements. Many coding schemes have been proposed, a review is given in [16] and a few examples are depicted in Figure 1. In order to avoid taking multiple shots as in time-multiplexing, several coding schemes can be used based on colour or variable width stripes. In this research we use grey-level stripes as in uncoded stripes of Figure 1 in which every stripe has the same width allowing a maximum resolution with a sufficiently dense pattern of stripes.

In order to proceed to the mathematical formulation, first it is necessary to define a coordinate system; all image and surface parameters will be related to this reference frame. We choose the coordinate system to be in relation to the light source as shown in Figure 2: the origin is set at a known distance from the centre of the projection. Concerning the physical arrangement, the camera and projector can be either in parallel or angled configuration as illustrated in Figure 3.

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 each pixel belongs to, and then through trigonometric relationships find the coordinates of the surface point imaged by the pixel.

Assuming that image processing functions are available to detect stripes and their respective indices in the image (see Section 3), the mapping from the image sensor to a surface point is defined by the pixel location and its

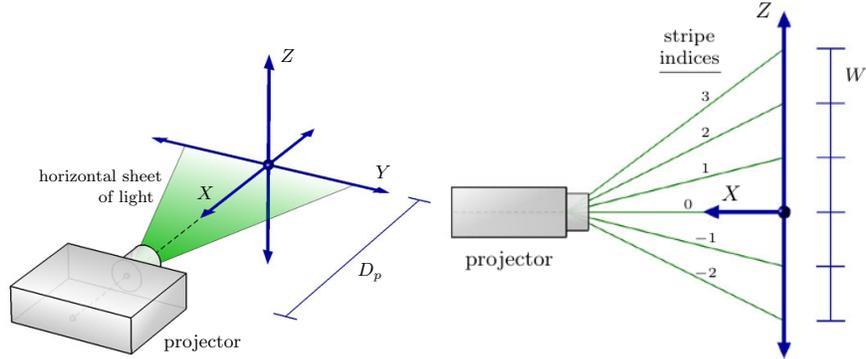


Fig. 2: A coordinate system is defined in relation to the light source which also defines the indices of the light planes.

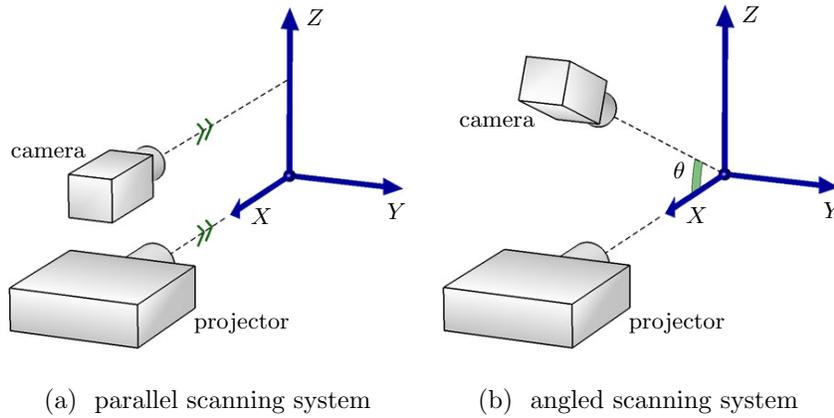


Fig. 3: Left, parallel configuration; right, standard angled configuration.

corresponding stripe index. From the point of view of the image sensor as illustrated in Figure 4, the mapping of a pixel to its coordinates \mathbf{p} is defined as $\mathbf{p} = \mathbf{c} + (0, -hPF, vPF)$ [14]. Here \mathbf{c} is the centre of the image, F is the focal length of the camera, P is the pixel size and each pixel in the sensor plane is a square of size $PF \times PF$.

The mathematical formulation for the parallel configuration as depicted by the geometry of Figure 5 is derived entirely from trigonometric relationships between the various measured distances [14]. The following parameters are defined:

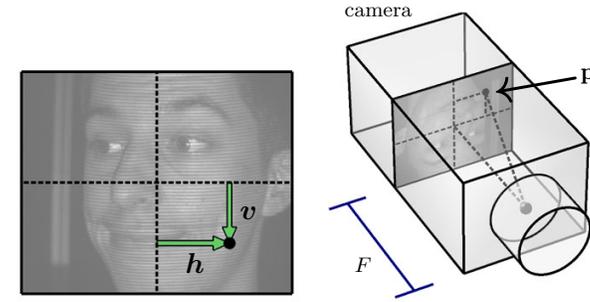


Fig. 4: The position of a pixel lying on a stripe of index n in the image plane is transformed to coordinates (h, v) then to a point \mathbf{p} .

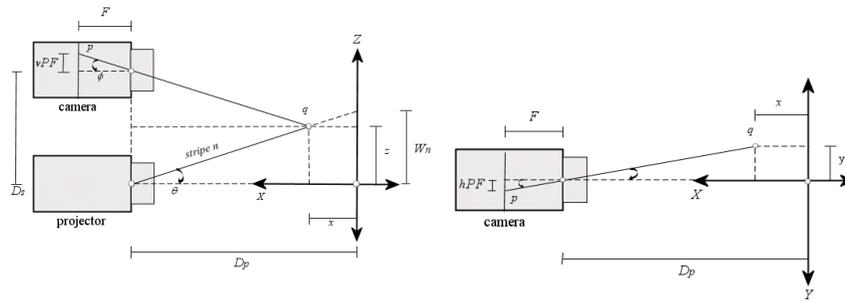


Fig. 5: Mapping of a surface point q from (h, v, n) in camera space to (x, y, z) in system space. Left, side view of scanner; right: view from the top.

- 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

The expressions to compute the coordinates (x, y, z) in system space from a pixel location (v, h) on stripe n (mapping to a point on the surface of the scanned object) is defined by the geometry of Figure 5 as:

$$x = D_p - \frac{D_p D_s}{v P D_p + W_n} \quad (1)$$

$$y = \frac{h P D_p D_s}{v P D_p + W_n} \quad (2)$$

$$z = \frac{W_n D_s}{v P D_p + W_n} \quad (3)$$

The mathematical formulation for the angled configuration has originally been described in [13]. In this case, the mapping of a pixel (v, h) on stripe n to its corresponding surface point (x, y, z) can be written as:

$$x = \frac{D_c v P + W_n (\cos \theta - v P \sin \theta)}{v P \cos \theta + \sin \theta + \frac{W_n}{D_p} (\cos \theta - v P \sin \theta)} \quad (4)$$

$$y = h P (D_c - x \cos \theta - z \sin \theta) \quad (5)$$

$$z = W_n \left(1 - \frac{x}{D_p}\right) \quad (6)$$

where the expressions for x in equation (4) can be substituted into equation (6) for z , and then into equation (5) for y , so that each is formulated as a function of v , h and n .

The resolution in 3D space (in millimetres) for both parallel and angled configurations are mathematically equivalent: it 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. The vertical resolution is affected by the spacing between the projected stripes which occupies several pixels in the image plane (the distance between two stripes). The horizontal resolution measured along each stripe is dependent on the horizontal dimension of the pixel space. We can formalize the resolution (δ) along the (x, y, z) dimensions by [14]:

$$\delta x = \frac{P D_p^2 D_s}{[v P D_p + W_n][(v + 1) P D_p + W_n]} \quad (7)$$

$$\delta y = P (D_p - x) \quad (8)$$

$$\delta z = \frac{W}{D_p} (D_p - x) \quad (9)$$

In practice, it is possible to process one vertex in 3D per image pixel along each stripe (in the horizontal direction assuming that stripes are horizontal). To illustrate the magnitude of resolution, for an IDS camera with 1280×1024 pixels calibrated at 300mm from the object's surface, the measured pixel dimension at the calibration plane is 240 microns. Across stripes, the peak-and-through distance of a stripe is 5 pixels. Therefore this means that along stripes the horizontal resolution is 240 microns (0.24mm) and across stripes the vertical resolution is $240 \times 5 = 1200$ microns (1.2mm).

It can be clearly seen from equations (1)–(6) that the mathematical formulation for the parallel configuration is simpler requiring less computing cycles than those of standard angled scanners: the parallel configuration requires 24 arithmetic operations per vertex in 3D against 37 operations for the angled (assuming that $\cos \theta$ and $\sin \theta$ are pre-computed). Thus, the parallel configuration design is more appropriate for 3D real-time processing, as it requires 35% fewer operations than the angled configuration. In terms of mechanical construction and calibration, both designs are equivalent; the significant difference is how the surface points (x, y, z) are calculated as described above from the known calibrated (i.e. measured) values of extrinsic parameters D_s , D_p , W and θ .

3 Image Processing

In order to reconstruct a surface in 3D according to the mathematical definitions in the previous section, the following steps are necessary: (1) image noise filtering; (2) detect stripe patterns in the image; (3) index the stripes in relation to the centre of the system; then (4) map to 3D space. The mapping to 3D is accomplished using equations (1), (2), (3) for a parallel configuration and equations (4), (5), (6) for an angled configuration.

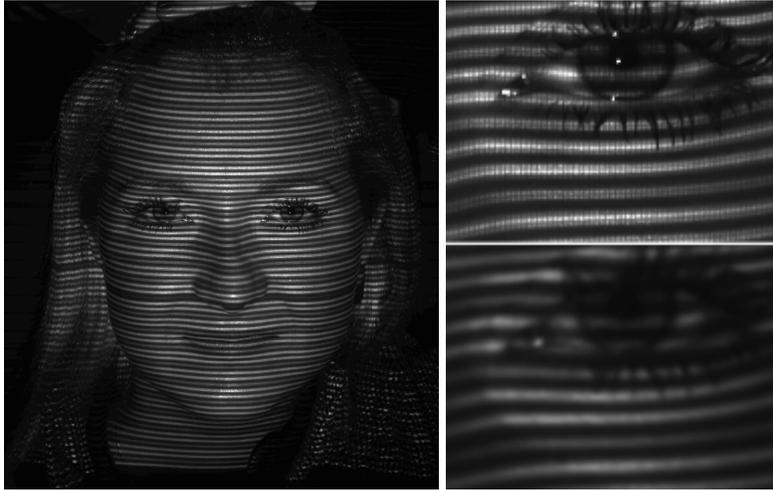


Fig. 6: Left, raw image; top right, raw detail; bottom right, filtered detail.

Figure 6 shows a raw image with detail on the right (raw and filtered). Normally images are noisy with high specularities especially if a laser-based projector is used. Image noise removal operations are required namely a median

filter followed by a weighted mean filter. Details of raw and filtered image are shown on the right – although it may seem less sharp, each stripe on the bottom right has a smooth Gaussian profile making it easier to detect the peak of each stripe with sub-pixel accuracy, which is determined as the local maxima in the greyscale intensity values of each stripe. Once peaks or local maxima are detected, algorithms are run (e.g. [2, 13] based on maximum spanning tree and flood filling) to detect each stripe correctly indexing them, where the stripe with index 0 (zero) corresponds to the light plane emanating from the centre of the projector (the reference projected light plane). Stripes with ascending positive indices are above the centre stripe, negative below. With the correct indices determined, the mapping given by equations (1)–(3) or (4)–(6) are then applied resulting in a point cloud of vertices in 3D space.

4 Experimental Results

We developed a flexible design of a compact scanner prototype with internal adjustable mechanisms allowing it to be configured and calibrated either as a parallel or angled configuration. To this end, we must choose a calibration plane at an arbitrary distance from the light source (300mm was used in this paper, albeit the calibration plane can be set at any distance between 200–800mm or until stripes are still detectable in the image) and adjust the camera until the line perpendicular to the image sensor is either parallel to the light source or until it intercepts the system origin at the calibration plane as shown in Figure 3(b) for an angled configuration. Note that the calibration plane does not set the maximum distance an object can be scanned. The limitation is due to the brightness of the projector and the reflective properties of the surface being scanned which, ultimately, will determine whether or not stripe patterns can be detected. For instance, it is known that metal surfaces are difficult to scan due to high levels of specular noise, requiring special techniques and filters to be used.

The prototype depicted in Figure 7 uses an IDS UI-1241LE camera board [6] with image size 1280×1024 and a Microvision PicoP laser projector [9] with resolution 848×480 pixels, equivalent to WVGA. The MARWIN project is concerned with 3D scanning of metallic components, and thus, tests were carried out on the actual metallic parts used in the robotic assembly. Furthermore, for a comparative analysis of scanner performance, wood and skin surfaces were also tested. The difference between the parallel and angled configurations is the number of arithmetic operations to evaluate a point cloud; the calibration procedures are equivalent, and the 2D image processing functions are the same and depend on the surface being scanned. Equally, if 3D post-processing is required such as hole filling and smoothing, these functions are also equivalent for both configurations.

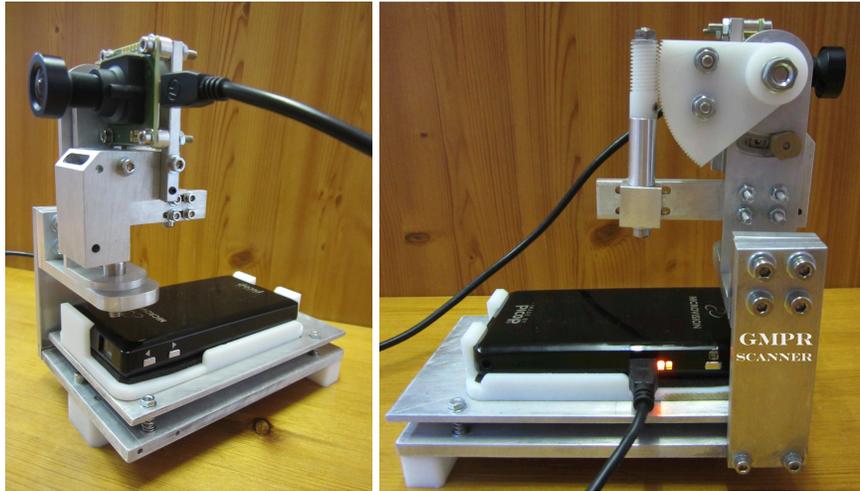


Fig. 7: The compact GMPR Scanner assembly with laser projector and CMOS camera. Dimension $W \times H \times L = 100 \times 150 \times 150\text{mm}$.

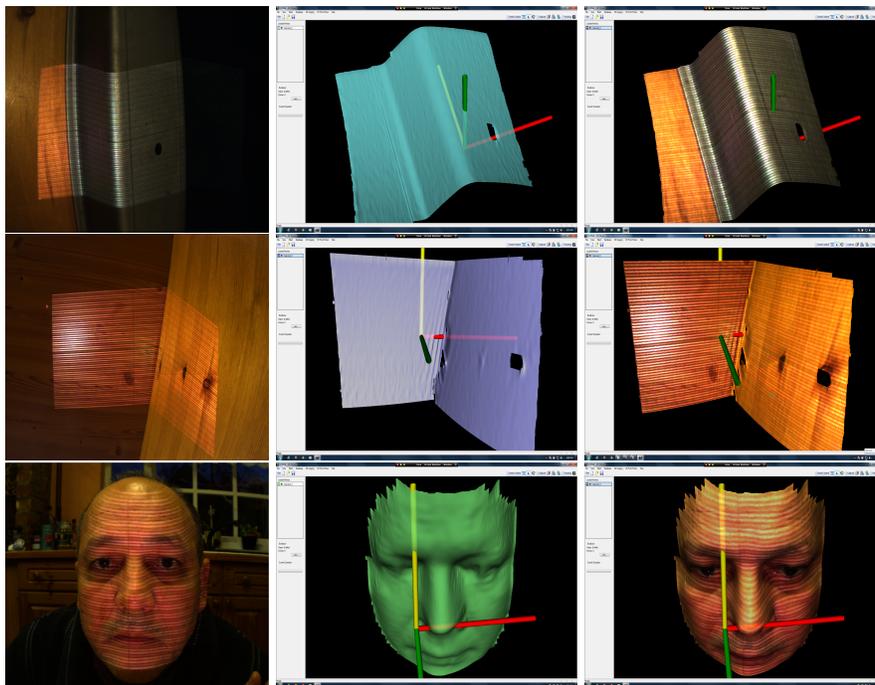


Fig. 8: From top to bottom: metal, wood, skin. Left column: 2D image; middle: 3D reconstructed surfaces; right: surface with texture mapping.

Figure 8 depicts several reconstructed surfaces. As 3D reconstruction maps a 2D pixel to its counterpart 3D vertex, texture mapping is a straightforward process and there is no need for a post-processing texture mapping step as in standard scanners. Due to high specularity, metallic parts require two distinct image filtering functions namely a median followed by a weighted mean filter to make stripes detectable. In contrast, the other surfaces only require either a weighted mean filter or a mean filter. In terms of overall processing time we can break down in several steps: 40ms for 2D image filtering (median and mean filters), 40ms for calculation of a point cloud in space for the angled configuration (26ms for the parallel configuration), and 120ms for 3D mesh post-processing (triangulation, small hole filling, mesh repairing and smoothing). With all the times computed, the current prototype in either parallel or angled configuration can operate in real-time at about 5 frames per second. If only the point cloud estimation is considered, the parallel configuration requires 35% fewer operations per vertex as compared to the angled one.

5 Conclusions

This paper describes research as part of the MARWIN project on 3D scanning of metallic components for robotic welding tasks. The focus is on the development of structured light scanners; a novel mathematical formulation for a parallel configuration is presented and compared to a standard angled arrangement of light source and image sensor. The built prototype consists of a camera board and a laser projector mounted on a mechanism designed for fine adjustments yielding either parallel or angled configurations. All interfaces to the control computer, camera boards and projector use standard USB connectors.

The main image processing requirement is identified as filtering images such that pixel intensities across stripes describe a smooth Gaussian profile for reliable detection of stripes peaks and troughs. Our experience has demonstrated that for reflective surfaces such as metals, employing a median filter followed by weighted mean filters can effectively remove extremes of noise resulting in the required Gaussian profile of pixel intensities. On less reflective surfaces such as wood and skin tissue, a combination of mean and weighted mean filters proved effective.

The significance of the design shown here is that it provides a simple to build and calibrate scanner that can be used for a variety of tasks and surfaces. In particular, its small form factor and interfaces easily integrate within robotic tasks, such as welding and assembly of metallic components of the MARWIN project. Future work includes registration and fusion of scanned welding components and tests with LED light sources in the visible and near-infrared spectra. Furthermore, a detailed error analysis of the scanned

surfaces is required; the focus of this paper has been on initial development and testing of the designs for parallel and angled configurations and on the ability to scan a variety of surfaces.

Acknowledgements

This work is supported by the European Commission, MARWIN Project Grant 286284 Research for SMEs SME-2011-1, from Nov 2011 to Oct 2013. We also acknowledge Willie Brink at Stellenbosh University (South Africa) for his work on geometry.

References

1. Albitar, C., Graebbling, P., Doignon, C., 2007, "Robust structured light coding for 3D reconstruction", in: *Int Conf Comp Vis (ICCV)*, pp. 1–6.
2. Brink, W., A. Robinson, M. A. Rodrigues, "Indexing Uncoded Stripe Patterns in Structured Light Systems by Maximum Spanning Trees", *Proc BMVC British Machine Vision Conference 2008*, ISBN 978-1-901725-36-0.
3. Chen, S., Li, Y., 2008, "Vision processing for realtime 3D data acquisition based on coded structured light", *IEEE T Image Process* 17, 167–176.
4. 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, 3119–3128.
5. Gorthi, S., Rastogi, P., 2010, "Fringe projection techniques: Whither we are?" *Opt Laser Eng* 2, 133–140.
6. IDS Imaging GmbH, (2013) "USB2 uEye LE Camera Boards", www.ids-imaging.com. Last accessed 9 Jan 2013.
7. Kawasaki, H., Furukawa, R., Sagawa, R., Yasushi, Y., 2008, "Dynamic scene shape reconstruction using a single structured light pattern", in: *IEEE Int Conf Comp Vis and Pat Rec (CVPR)*, pp. 1–8.
8. MARWIN: *Decision making and augmented reality support for automatic welding installations*, EU-funded Project no. 286284, Research for the Benefit of SMEs, from Nov 2011 to Oct 2013. <http://www.marwin-welding.eu/>
9. Microvision Inc, (2013). "MicroVision SHOWWX+ Laser Pico Projector", microvision.com. Last accessed on 9 Jan 2013.
10. Mirota, D.J., Ishii, M., Hager, G.D., 2011. "Vision-based navigation in image-guided interventions", *Annu Rev Biomed Eng* 13, 297–319.
11. Mountney, P., Stoyanov, D., Yang, G.Z., 2010. "Three-dimensional tissue deformation recovery and tracking", *IEEE Signal Proc Mag* 27, 14–24.
12. Pavlidis, G., Koutsoudis, A., Arnaoutoglou, F., Tsioukas, V., Chamzas, C., 2007, "Methods for 3D digitization of cultural heritage", *Journal of Cultural Heritage* 8, 93–98.
13. Robinson, A., 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.
14. Rodrigues, M., M.Kormann, C.Schuhler, G.Melton, J.Shejbal, P.Tomek (2013). An Intelligent Real Time 3D Vision System for Robotic Welding Tasks, *IEEE ISMA '13 The 9th Int Sym on Mechatronics and Apps*, Amman, Jordan, April 9–13.

15. Rodrigues, M. 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.
16. Salvi, J., J. Pagès, J. Batlle, "Pattern codification strategies in structured light systems", *Pattern Recognition Volume 37*, Issue 4, April 2004, Pages 827–849.
17. Smisek, J. (2011). 3D with Kinect, *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 6-13 Nov. 2011, 1154–1160.
18. Wang Y., K. Liu, D.L. Lau, Q. Hao, L.G. Hassebrook (2010). Maximum SNR Pattern Strategy for Phase Shifting Methods in Structured Light Illumination, *J. Opt. Soc. Am. A*, 27(9), pp. 1962–1971.