Stereo Vision Sensing: Review of existing systems

Conference Paper · December 2018

DOI: 10.1109/ICSensT.2018.8603605

4 authors:

Andrew O' Riordan
University of Limerick
7 PUBLICATIONS **108** CITATIONS

Thomas Newe
University of Limerick
197 PUBLICATIONS **2,676** CITATIONS

Daniel J F Toal
University of Limerick
178 PUBLICATIONS **2,542** CITATIONS

Gerard Dooly
University of Limerick
137 PUBLICATIONS **1,997** CITATIONS

Stereo Vision Sensing: Review of existing systems

Andrew O'Riordan,
Confirm Research Centre,
Centre of Robotics and Intelligent Systems,
Department of Electronic & Computer Engineering,
University of Limerick
Limerick, Ireland
Andrew.ORiordan@UL.ie

Thomas Newe,
Confirm Research Centre,
Centre of Robotics and Intelligent Systems,
Department of Electronic & Computer Engineering,
University of Limerick
Limerick, Ireland
Thomas.Newe@UL.ie

Gerard Dooly,
Confirm Research Centre,
Centre of Robotics and Intelligent Systems,
Department of Electronic & Computer Engineering,
University of Limerick
Limerick, Ireland
Gerard.Dooly@UL.ie

Daniel Toal,
Confirm Research Centre,
Centre of Robotics and Intelligent Systems,
Department of Electronic & Computer Engineering,
University of Limerick
Limerick, Ireland
Daniel.Toal@UL.ie

***Abstract*— Walking through any trade show or expo of 2018 it becomes obvious that 3D scanners have become the latest available in an arsenal of new technology being offered at an industrial level. Multiple technologies are implemented to achieve the task of 3D imaging from the real world. This paper focuses specifically on stereo vision sensing. A review of existing stereo vision systems is provided, their challenges and proposed solutions are outlined.**

Keywords—Vision Sensing, Stereo Vision, Image Matching, False Boundary, Boundary Overreach

I. Introduction

Stereo vision (SV) sensing has become one of the more popular methods of capturing 3D images. SV employs a number of cameras in order to capture multiple images of an object with the cameras at a known distance from each other. These images are combined using a triangulation of the distance between the cameras and the distance from the scanned object to form a 3d image, however matching these images to a high accuracy presents challenges and needs an implementation of specific procedures. These challenges and published solutions are outlined in this paper.

This article is divided into three primary sections; 1- SV, 2- the challenges of SV & 3- solutions to these challenges and advancement of SV as a sensing technology.

II. Stereo Vision

3D vision sensing can be found implemented in industry for navigation, collision detection systems and metrology, to name a few [1][2][3]. 3D vision can be carried out using a number of methods, depending on the intended use of the collected information. Some methods for accurate reconstruction include laser point or slit scanners, however, these do not allow the capture of dense, dynamic scenes achievable by SV[4]. SV is possible by measuring the disparity (d) (the distance between outer points on multiple images), which is then correlated to the distance (z) from the baseline of the object using the following equation:

Equation 1

$$d=(BF)/z$$

where B and F represent Baseline (the distance between camera centres at the moment the image is captured) and focal length respectively [5].

This can be advanced in a number of ways to allow more in-depth images to be obtained, such as passive stereo vision, active stereo vision or multiple-baseline stereo.

A. Passive Stereo Vision

Passive Stereo Vision (PSV) captures the boundary of a smooth surface. Where one camera is used, it is possible to derive the depth of the object using triangulation, however, this will not capture a boundary. In order to capture a boundary (i.e. a planar surface) and a depth, a minimum of two cameras or images must be used. The location of the camera for each image may be used to compute the 3D geometry of a scanned object as shown in Figure 1 below. A common approach to matching the two images together is to correlate a group of pixels (shown as the hatched area below), assuming that the light reflection off a point on the surface of the object will reflect the same light intensity to each camera [6].

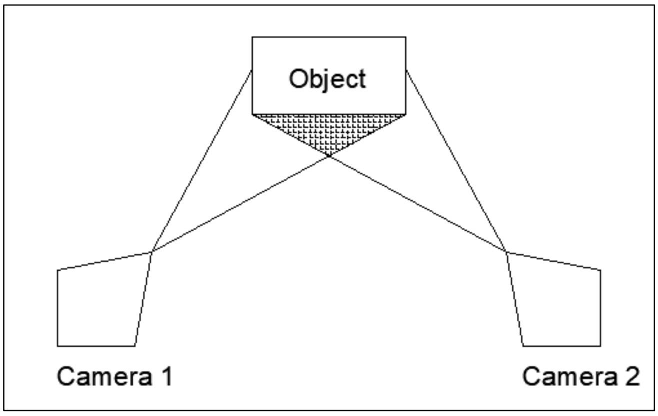

Figure 1 Passive Stereo Vision

978-1-5386-5147-6/18/$31.00 ©2018 IEEE

PSV requires the distance from the cameras to the front of the object to be known. This results in a restrictive 3D vision system as it is static, it does not allow for dynamic movement with accuracy as it is difficult to establish multiple baselines without assistance, such as structured light (active stereo vision) [7]. Motion of cameras is possible but image matching and construction are then based on motion compensation [4]. Often the employment of multiple cameras, where the images will have pixel correlations on multiple base points on the object (multiple baselines), will yield a much higher accuracy of 3D modelling than dual cameras [6].

Some of the main limitations of PSV are the existence of the false boundary issue (III.B) and reflectivity issues (III.C). These problems disallow the capture of the true boundary or surface topography due to image mismatching or lack of diversity of light intensity captured by pixels on the object surface due to high levels of reflectivity. These challenges limit the application of PSV within industry as it is not consistently reliable as a 3D image capturing technology [7].

B. Active Stereo Vision

Active Stereo Vision (ASV) shares the same basic principles as PSV with regards to capturing a frame, however, it allows for dynamic motion of the cameras and object. As a result, multiple frames per second can be processed and merged to form a much more detailed 3D image. This is achieved by the use of a structured projected light source (often grids of white light or infrared light). The light projected onto the object allows the multiple images captured to be matched together by correlating the pixels which are picking up the gridlines. Motion through this method is permitted as the light source will allow for dynamic image matching as the cameras detect the approaching grid lines as movement occurs. Major benefits of ASV include the possibility of mounting cameras on manipulators such as robotic arms etc. allowing for automation of the 3D vision technique [4][7]. This is shown in Figure 2A & 2B.

While ASV allows for dynamic systems to capture 3D images, it is not without its limitations. Where a complex surface is being captured, simpler shapes of light grids such as circles will not provide enough detail for accurate image matching as they provide fewer matching points than more complex features such as pentagons or hexagons. Although this has been addressd and attempts to ease the problem have been made, the limitation exists and has restricted the useability of ASV for practical applications within industry. This is due to its lack of versatility when scanning surfaces, caused by simple grid-shapes lacking the detail necessary to match points [7].

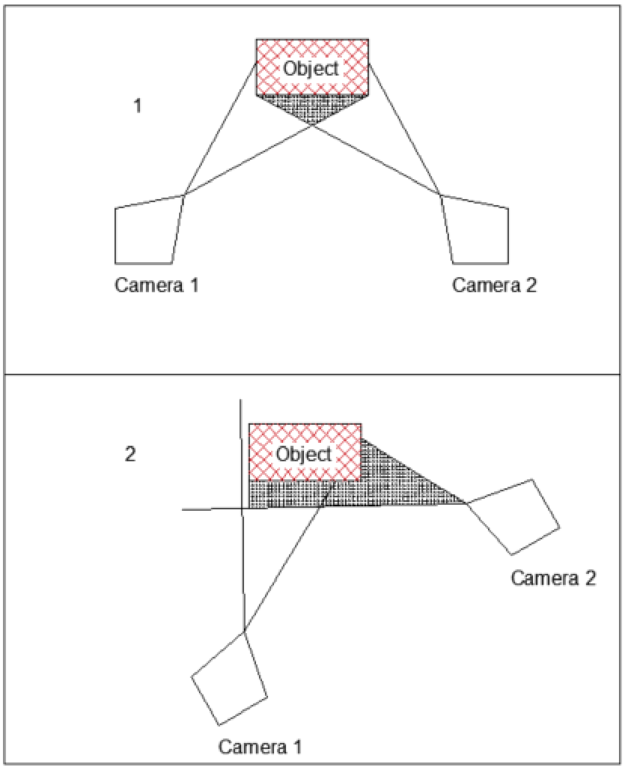

Figure 2-A Active Stereo Vision

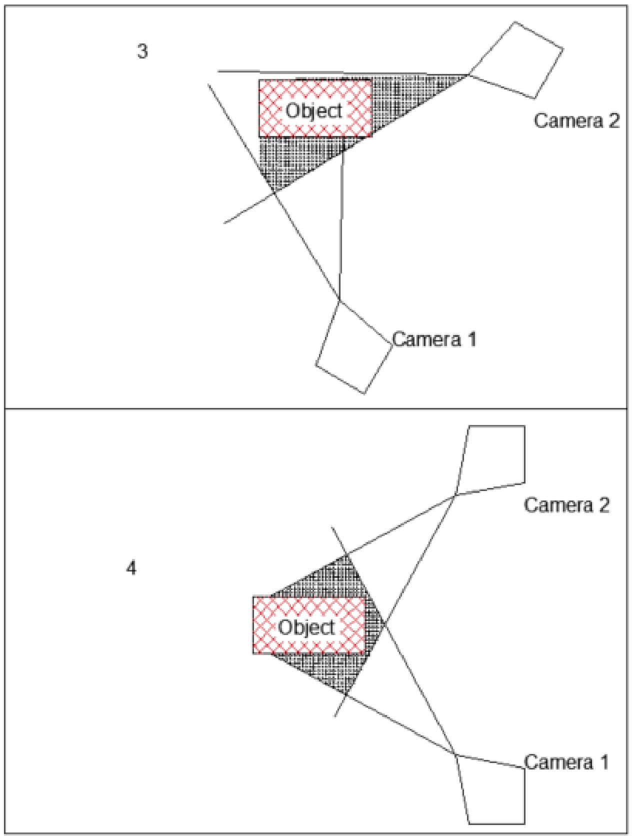

Figure 2-B Active Stereo Vision

C. Multiple-Baseline Stereo

Multiple-Baseline Stereo (MBS) utilizes many cameras to capture many images to match together to improve the accuracy of the 3D image captured and to allow a wider angle of capture of the object. MBS is usually a form of ASV, as the grid will allow the multiple cameras to synchronise easily. A major benefit of MBS used as a form of SV is that it will minimise any errors caused by discontinuities [4]. Trinocular stereo approaches have been used and proven successful for robotic navigation in indoor applications such as autonomous intelligent vehicles (AIV) [2].

Although MBS allows for a more accurate 3D image to be captured, by establishing multiple baselines between camera pairs, there occurs a trade-off between accuracy of pixel correlation (and thus image matching) and precision of the 3D image created. Equation 1 shows that the baseline and disparity are proportional for a constant distance. This means that if the baseline is extended (i.e. if the cameras are moved further from each other) a more accurate distance can be obtained, this, however, increases the disparity. An increased disparity results in a larger area to be searched for matching pixels and gives a larger possibility for false matching to occur as shown in Figure 3 [5].

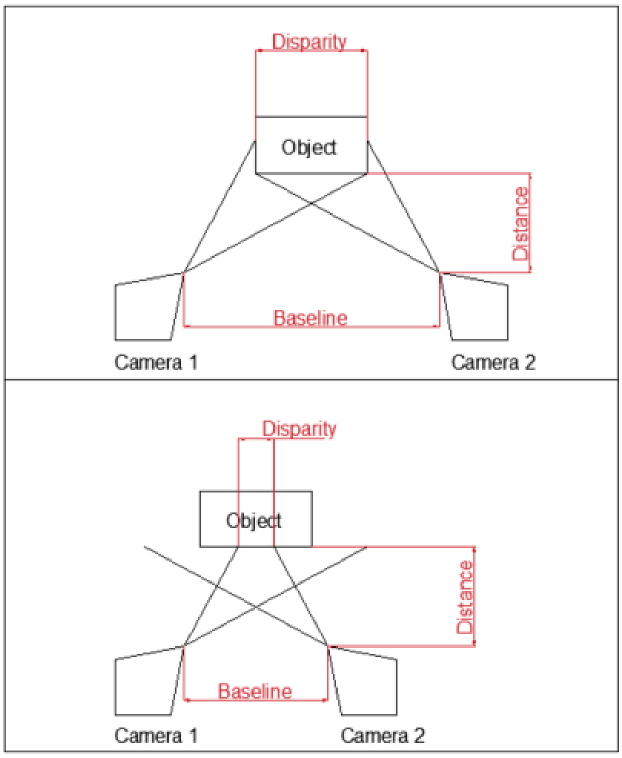

Figure 3 Disparity vs. Baseline

A commonly used solution for this is a combination of a coarse-fine image capture. A coarse image is taken with a short baseline to determine which data points collected provide a sufficient correlation to act as a base-point. This disparity section can then be matched, the baseline increased but with the matching pixels remaining within the disparity region established, allowing for a higher resolution result to be obtained. This allows for precision of results while significantly reducing the possibility of false matches [5].

If this method is repeated amongst a group of camera pairs, it will allow for multiple-baseline 3D image computation with minimised error and high precision in results [8].

III. Challenges of Stereo Vision

SV is not without its challenges; the advancement of SV has given way to solutions for some challenges, but new challenges have also appeared. Challenges faced include image matching, false boundary problems, reflection issues due to surface texture (which can also lead to false boundary issues) and identification of obstacles.

A. Image Matching

Image Matching (IM) can be one of the most problematic procedures to carry out using SV for 3D imaging. While it is often explained quite simply as matching pixels using correlation, establishing this correlation can prove a challenge.

Some methods for this include line matching on boundaries or edges of the object being scanned, others include matching all common features between images (i.e. 100% of pixels that have captured the same features are matched) known as global optimization. Some methods include area-based measures which find a correlation between image patches. One of the most effective methods, however, is to analyse each pixel as a group of collected data points which can be identified by an algorithm. This allows an analysis of each pixel from all cameras. This requires a large amount of processing power. This has led to many different techniques being employed to optimise this method as mentioned in section IV.A, including the coarse-fine technique mentioned in section II.C [8].

B. False Boundary Problem

The false boundary problem, also known as boundary overreach occurs when the edge of an object is out of view of the cameras, and a parallax occurs as shown in Figure 4, other instances include boundaries appearing outside of the geometry of the object due to false-matching. This problem is exacerbated if the surface texture is not suitable for light reflection. Although this can often be solved easily by using dynamic or exploratory sensing, involving the movement of cameras to allow the outer vision points to grasp something in the background of the object, this is not always satisfactory if the background is a continuation of the same object with the same surface texture [7]. Algorithms have been employed to overcome boundary overreach and false boundary issues as caused by mismatching, as addressed in sections IV.A. & IV.B.

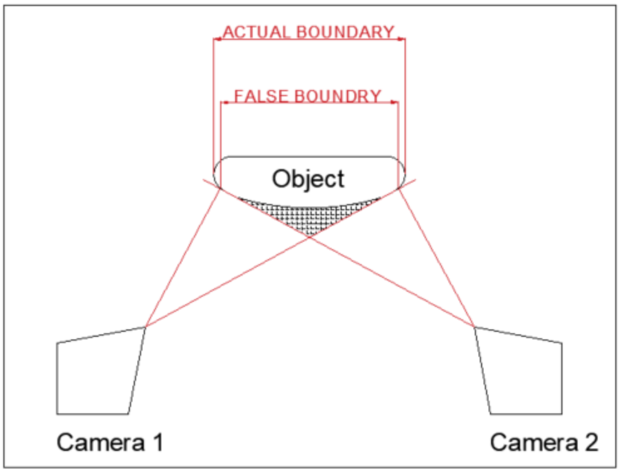

Figure 4 False Boundary Problem

C. Reflection issues due to Surface Texture

SV relies on the change of reflection of light from the surface of an object, which can lead to difficulty when trying to obtain a 3D image of a smooth, reflective metal surface, as where the surface geometry changes, the light reflected may not [7]. ASV is employed to assist this, however, due to the uniformity of a surface, structured white light may still lead to false matching or false surface imaging. When images are matched it is possible for false spikes of light to give way to a false geometry or for the geometry to be unmapped due to a lack of variance in light reflection across the surface of an object, this is demonstrated in Figure 5. In Figure 5-A, the object can be seen with a non-uniform light reflection, showing some of the geometry but also showing false lines where no geometry changes exist. Figure 5-B shows a side profile of the object, showing there should only be two boundary changes across the top of the object. Figure 5-C shows in black the true geometry changes (only portional visible in Figure 5-A), the red lines show the false geometry caused by the surface texture of the material [9].

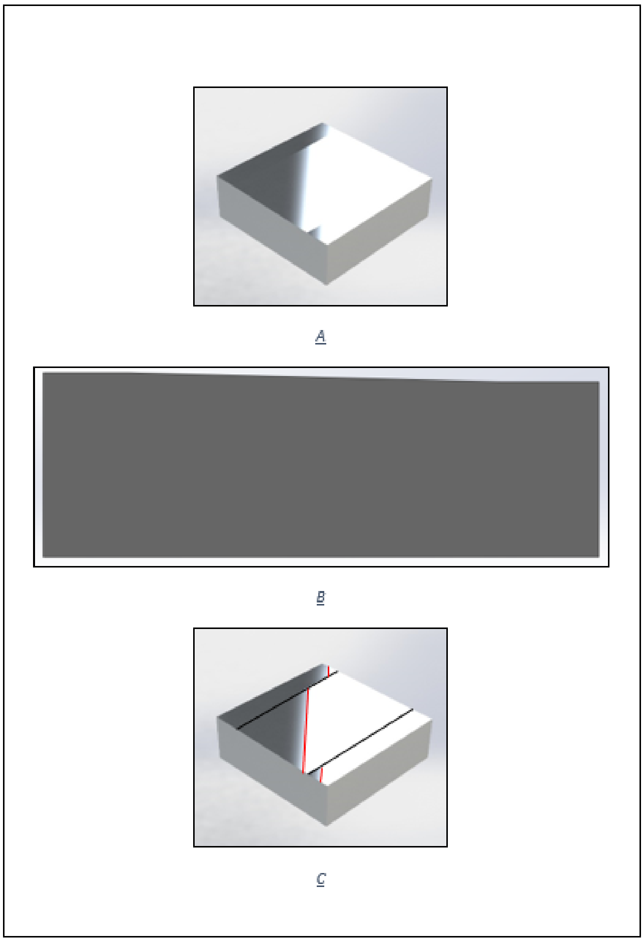

Figure 5 False Boundary due to Surface Reflection Problems

D. Identification of obstacles

Early attempts of robotic navigation employed a primary spatial sensor based on sonar transducers or laser range sensors. Experimentation of the use of SV for navigation proved successful when combined with occupancy grid mapping (OGM). OGM consists of dividing a plan of floor space into grids or zones, assigning each grid a probability of occupancy of 50%, where 100% represents a zone that is occupied and 0% represents an empty zone. As a zone comes into range of the spatial sensor, it is possible to determine if this zone is occupied, changing its probability of occupancy accordingly. The primary route-finding technique is based on choosing the path involving the lowest probability of occupancy, with a secondary objective of finding the shorted path between start and finish points. OCM proved very versatile a technique as it allows for utilization of an array of different sensor types. Where SV was employed it proved successful to a point, however, issues of identifying shadows or false obstacles arose. Examples of this include the SV sensing a table top and identifying it as a shadow as it seems to be a 2-D shape suspended in mid-air [2]. Algorithms used to identify obstacles more accurately were applied, making sure that if an imaged 2-D shape showed continuity toward another 3D surface, it was identified as an obstacle.

This method gave rise to the problem of moving obstacles, however, a combination of SV & ultrasonic sensors proved successful in resolving this issue [10]. The use of SV in these systems allowed continuous mapping and dynamic map updates of occupancy of each zone, allowing information to be communicated to multiple autonomous robots to path find based on occupancy of zones out of current range. Experimentation of exploratory, free-roams of floor plans have proven successful in applying these methods. SV has also been used for real-time navigation in outdoor environments when combined with inexpensive GPS. SV, in this case, was used not only for obstacle detection but also to estimate frame-to-frame motion [11].

IV. Algorithmic Solutions and Applications

Multiple algorithmic solutions have been proposed to the challenges identified in section III. Although these have been successful, there have been many adapted applications and capabilities of SV sensing technology due to the successful implementations of additional algorithms, including colour analysis.

A. Image Matching

Many approaches have been applied to image matching to optimise the task, however, some have proven more successful than others. Some methods used include global-optimisation, algorithms based on dynamic programming or 2-D curve matching [8].

Area-based measures treat one image as a reference image. Within this image, there will be a window which will be analysed and statistically compared to other images captured simultaneously from other cameras, often referred to as target images. Image-hierarchy such as the coarse-fine approach may be utilised at this point, to obtain an accurate match and then allow for a precise image capture. Usually, a difference-based metric will be used such as a minimised RMS or maximised correlation (e.g. Mean and variance normalised cross-correlation) [8][12]. While more modern approaches provide many real-time stereo systems, most run on a correlation stereo engine [13]. Correlation-based methods allow for high confidence in the procedure but diminish in reliability where longer baselines are used, image texture is problematic or where there is high noise present. Global optimization techniques employ an image hierarchy. Global optimization will match all points within the two images that correspond [14].

As mentioned in section II.C, multiple baseline stereo presents its own image matching challenges. Where multiple baselines exist due to multiple cameras, a mathematical analysis must be carried out. This analysis is based on the idea that global mismatches may be reduced by adding the sum of squared difference (SSD) from multiple stereo pairs. Firstly, the SSD of each individual pair of stereo images is calculated, this can then be related to either disparity or inverse-distance (see Equation 1). A total SSD can be obtained by adding all SSD's from all pairs. This sum of SSD's will be at a minimum where the images are matched appropriately. This procedure will be repeated through all camera pairs [5]

Dynamic programming introduces a new primitive known as declivity, which is defined as "a cluster of contiguous pixels, limited by two endpoints which correspond to two local extrema of grey level intensity" [15]. Bensrhair et al. sort the declivities collected into two groups; those of low amplitude (which are then ascertained to be noise, and thus eliminated) and those of high-amplitude (which are deemed features and thus retained). Essentially, rather than matching individual points, curve geometry or statistically matching windows, this method allows for a region to be matched according to the amplitude of a cluster of points between two defined areas of grey intensity, limited by a defined threshold of standard deviation of the white Gaussian noise component in each image line [15].

Two-dimensional curve matching techniques for image matching makes use of algorithms which identify a group of individual points that collectively allow a continuous curve to be plotted. Feature points which fit into these curved geometries are extracted from image windows. Curves from the reference and target images are compared, a maximum variance is set as a matching parameter for these curves, with the angle formed by image pairs considered for geometric projection. This approach has proven successful with both trinocular and binocular cameras. Dynamic calibration is necessary for this method [16].

B. False Boundary, Boundary Overreach & Reflectivity Issues

While an adaptive window, a multiple window/Symmetric window and a shrinking window have been applied to solve false boundary and boundary overreach problems, they have proven to harm smooth object surfaces within the acquired 3D image, resulting in a trade-off between the smoothness of surfaces and the precision of a boundary [9].

Adaptive window techniques address the problem of having a window large enough to include intensity variation for matching purposes but also small enough to prevent projective distortion. Kanade et.al. [17] developed an adaptive window technique by performing an evaluation of local intensity variation and disparity variation. This allowed for a suitably sized window to minimise the uncertainty of disparity within a region.

Multiple window or symmetry window techniques perform multiple analyses on windows surrounding a pixel area, with the aim being to find the window with the lowest SSD. This window is more likely to cover a constant depth, matching is carried out at a base point within this window [18].

Shrinking window techniques use an image hierarchy system where the window shrinks around the area that provides a sharp depth, in which a base point will be found [19].

These techniques prevent boundary overreach and false boundary projection due to matching errors, however, do not prevent false boundaries due to object geometry. This can be addressed by use of dynamic exploratory sensing, making use of ASV, where the cameras can be centred facing the object in order for the centre line of the stereo cameras to coincide with the centre-most intersection of 2 tangents of the curves it is imaging. Another method is increasing the distance to capture a coarse shape upon which finer details may be later overlaid with fine image capture [7].

Reflectivity issues occur when the surface texture does not provide enough variation of light reflection due to the uniformity of the material (as shown in Figure 5). While ASV can often be the answer, it cannot be limited to grids of white light. Areas of continuous, reflectivity, while not flat, may show spikes in noise where there are no features. By moving the cameras but keeping the light-source stationary, there is an opportunity to layer a geometry repeatedly with frames from different angles with the motion of the sensing cameras compensated for. This allows for the elimination of spikes of points and depth which are outside of the standard deviation of the collected points within a defined proximity of the occurrence [7].

C. *Obstacle Identification*

SV has become more common as a solution for robotic navigation, however, as addressed in section III.D, obstacle identification has presented a challenge, as many obstacles such as table tops can be identified incorrectly as shadows, and shadows as obstacles. Murray et al. [2] identified this problem and addressed it using the following logic. If a spike is locally stable but not large and lacking in support from surrounding surfaces, it is treated as a spike error and eliminated from the analysis process. If the spike shows a lack of disparity discontinuities at all corners, is locally consistent and shows visual evidence of being globally part of a larger 3D surface, it is a solid obstacle. Segmentation of images allowed for a hypothesis to be formed on the size of a continuous 3D object (i.e. Connected pixels which show local consistency). This rejects false noise spikes that may pass noise filters, but also it allows for recognition of thin structures which belong to a part of a larger coherent structure [2].

D. *Colour Analysis*

Colour analysis involves the measurement of several metrics to establish the classification of colour in each pixel. The metrics used can involve the mean, variance and range of;

- Colour
- Hue
    - Hue is expressed as an angle where 0° = red, 60° = yellow, 120° = green, 240° = blue, 300° = magenta.
- Saturation
    - The saturation component signals how much the colour is polluted with white colour. The range of the saturation component is 0-1.
- Intensity
    - The intensity range is 0-1; 0 = black, 1 = white) [20].

These analyses have been used extensively for the identification of defects and damages within food production monitoring systems [20][21]. Colour analysis can also play a part in allowing for the capture of surface texture in machined parts or corrosion levels in metallic parts such as copper or steel [22].

E. *Metrology*

Metrology utilising SV requires calibration before a measurement can be taken. This can involve the 3D imaging of an object of know dimension, usually a calibration gauge. While the accuracy of metrology using 3D imaging is more difficult to assess than mechanical measurement (due to the variables that must be taken into consideration such as illumination of the object, surface structure and geometry of the object), results have shown to be promising. Aguilar et al [23] showed low positive results with 1mm uncertainties along each axis where the measuring field was 1500x200x4000mm. In more recent attempts, Weise et al [4] reported an accuracy to within 0.125mm when imaging at 1100mm distance.

V. CONCLUSION

SV sensing technology has played a big part in advancing the ability to construct 3D images of the real world. While it presents challenges for complete optimisation of the process purely for obtaining 3D images, these challenges are being met with many solutions developed in software and hardware. SV is not, however, being used to its full potential for industrial use, in the authors opinion, due to the limitations that still exist, as outlined in the respective sections of each technology.

Autonomous stereo scanning, carrying out of colour analysis simultaneously to 3D imaging an object will allow for the detection of defects in an object if colour-based thresholds are established (based on measured metrics). This would provide opportunity for the implementation of stereo scanning as a quality assurance measure within a production environment. Colour analysis would also allow for another metric to be included in the stereo matching image hierarchy techniques, and thus would be another step toward eliminating false matches, and by effect, boundary overreach.

Like colour analysis, implementation of metrology using SV sensing would allow for quality assurance to take place autonomously within production environments. Potential implementation of further visual analysis of acquired images would allow for SV sensing to become far more versatile as a sensing technology.

ACKNOWLEDGMENT

This work was supported by the Science Foundation Ireland under Grant No. 16/RC/3918.

REFERENCES

[1] Yuhara, H., Terauchi, A., Saka, M., Habaguch, M., & Kawai, M. (2000). Stereo vision sensor. JSAE Review, 529-534.

[2] Murray, D., & Little, J. J. (2000). Using Real-Time Stereo Vision for Mobile Robot Navigation. Autonomous Robots, 161-171.

[3] Tsai, R. Y. (1987). A Versatile Camera Calibration Techniaue for High-Accuracy 3D Machine Vision Metrology Using Off-the-shelf TV Cameras and Lenses. IEEE JOURNAL OF ROBOTICS AND AUTOMATION, 323-344.

[4] Weise, T., Leibe, B., & Van Gool, L. (2007). Fast 3D Scanning with Automatic Motion Compensation. Conference on Computer Vision and Pattern Recognition,.

[5] Okutomi, M., & Kanade, T. (1993). A Multiple-Baseline Stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence, 353-363.

[6] Seitz, S. M. (2018, 6 26). Carnegie Mellon University. Retrieved from Index of /~seitz/course/SIGG99/papers: http://www.cs.cmu.edu/~seitz/course/SIGG99/papers/seitz-passive-abs.pdf

[7] Chen, C., & Zheng, Y. F. (1995). Passive and Active Stereo Vision for Smooth Surface Detection of Deformed Plates. IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, 300-306.

[8] Hannah, M. J. (1989). Digital Stereo Image Matching Techniques. Photogrammetry & Remote Sensing, 1765-1770.

[9] Okutomi, M., Katayama, Y., & Oka, S. (2002). A Simple Stereo Algorithm to Recover Precise Object Boundaries and Smooth Surfaces. International Journal of Computer Vision, 261-273.

[10] Ohya, A., Kosaka, A., & Kak, A. (1998). Vision-Based Navigation by a Mobile Robot with. IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION Obstacle Avoidance Using Single-Camera Vision and Ultrasonic Sensing, 969-978.

[11] Agrawal, M., & Konolige, K. (2006). The 18th International Conference on Pattern Recognition. Real-time Localization in Outdoor Environments using Stereo Vision and Inexpensive GPS (pp. 1063-1068). Hong Kong: IEEE Computer Society.

[12] Barnard, S. T., & Fischler, M. A. (1982). Computational Stereo. ACM Computing Surveys (CSUR), 553-572.

[13] Gehrig, S. K., Eberli, F., & Thomas, M. (2009). A Real-Time Low-Power Stereo Vision Engine Using Semi-Global Matching. In M. Frits, B. Schiele, & J. H. Piater, Computer Vision Systems (pp. 134-143). Liége: Springer.

[14] Barnard, S. T. (1987). Tenth International Joint Conference on Artificial Intelligence. STEREO MATCHING BY HIERARCHICAL, MICROCANONICAL ANNEALING (pp. 832-835). Milan: International Joint Conference on Artificial Intelligence.

[15] Bensrhair, A., Miché, P., & Debire, R. (1996). Fast and automatic stereo vision matching algorithm based on dynamic programming method. Pattern Recognition Letters, 457-466.

[16] Deriche, R., & Faugeras, O. (1990). 10th International Conference on Pattern Recognition. 2-D Curve Matching Using High Curvature Points: Application to Stereo Vision (pp. 240-242). Atlantic City, NJ: IEEE.

[17] Kanade, T., & Okutomi, M. (1991). IEEE. A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment (pp. 1088-1095). International Conference on Robotics and Automation: IEEE.

[18] Fusiello, A., Roberto, V., & Trucco, E. (1997). IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Efficient Stereo with Multiple Windowing (pp. 858-863). San Juan: IEEE.

[19] Park, J.-I., & Inoue, S. (1997). Acquisition of sharp depth map from multiple cameras. Signal Processing: Image Communication, 7-19.

[20] Luo, X., Jayas, D. S., & Symons, S. J. (1999). Identification of Damaged Kernels in Wheat usinf a Colour Machine Vision System. Journal of Cereal Science, 49-59.

[21] Leemans, V., Magein, H., & Destain, M. -F. (1998). Defects segmentation on 'Golden Delicious' apples by using colour machine vision. Computers and Electronics in Agriculture, 117-130.

[22] Stratmann, M., Leng, A., Furbeth, W., Streckel, H., Gehmecker, H., & Grose-Brinkhaus, K. H. (1996). The scanning Kelvin probe; a new technique for the in situ analysis of the delamination of organic coatings. Progress in Organic Coatings, 261-267.

[23] Aguilar, J. J., Torres, F., & Lope, M. A. (1996). Stereo vision for 3D measurement: accuracy analysis, calibration and industrial applications. Measurement, 193-200.