A Stereo-Matching Technique for Recovering 3D Information from Underwater Inspection Imagery

Michael O'Byrne, Vikram Pakrashi*

Dynamical Systems and Risk Laboratory, School of Mechanical and Materials Engineering, University College Dublin, Ireland
Marine Renewable Energy Ireland (MaREI), University College Dublin, Ireland

Franck Schoefs

Université Bretagne-Loire, Université de Nantes, Research Institute of Civil Engineering and Mechanics (GeM)/Sea and Littoral Research Institute (IUML), CNRS UMR 6183/FR 3473, Nantes; IXEAD/CAPACITES Society, Nantes, France

&

Bidisha Ghosh

Department of Civil, Structural and Environmental Engineering, Trinity College Dublin, Ireland

Abstract: *Underwater inspections stand to gain from using stereo-imaging systems to collect 3D measurements. While many stereo matching algorithms have been devised to solve the correspondence problem, i.e. find the same points in multiple images, these algorithms often perform poorly when applied to images of underwater scenes due to the poor visibility and the complex underwater light field. This paper presents a new stereo matching algorithm, called PaLPaBEL (Pyramidal Loopy Propagated BELief), that is designed to operate on challenging imagery. At its core, PaLPaBEL is a semi-global method based on a Loopy Belief Propagation (LBP) message passing algorithm applied on a Markov Random Field (MRF). A pyramidal scheme is adopted which enables wide disparity ranges and high-resolution images to be handled efficiently. For performance evaluation, PaLPaBEL is applied to underwater stereo images captured under various visibility conditions in a laboratory setting, and to synthetic imagery created in a virtual underwater environment. The technique is also demonstrated on stereo images obtained from a real-world inspection. The successful results indicate that PaLPaBEL is well-suited for underwater application and has value as a tool for the cost-effective inspection of marine structures.*

1 INTRODUCTION

Marine structures are particularly susceptible to aesthetic, functional or structural degradation. Inspections are therefore crucial to ensure that these structures remain safe and fit for service. On top of this, inspections are invaluable as the findings guide owners/operators when deciding on repair or rehabilitation works. Most decision makers are acutely aware that the quality of inspection information has a significant bearing on the continued health of structural assets and this carries with it major financial implications. Visual inspections are the most common means of collecting data about the health of marine structures, however, they have some inherent limitations. They are affected by issues such as operator boredom, concentration lapses, subjectivity, and fatigue, all of which contribute to increased variability and reduced accuracy (Estes & Frangopol, 2003; Komorowski & Forsyth, 2000). These issues are exacerbated for underwater inspections as divers must endure cold and uncomfortable conditions (Busby, 1979). Gallwey and Drury (1986) found that incorporating image-processing methods into the inspection regime leads to increased quality of the inspection results and greater reproducibility compared with visual inspections. Adopting image-processing methods requires little additional effort as visual inspections almost always capture photographs despite rarely being exploited to their fullest potential in either a qualitative or quantitative sense. By using stereo camera systems, inspectors can obtain an even richer source of quantitative 3D data.

Capturing 3D shape of underwater structural components, instead of just a 2D projection as a standard camera does, naturally lends itself to many applications. For instance, it opens the possibility of performing realistic Computational Fluid Dynamics (CFD) simulations using the reconstructed

3D geometries. Moreover, once engineers have access to 3D reconstructions of underwater targets, physical properties can be easily extracted, such as the size of damage forms like cracks and corrosion. Detection of these damages may be automated using existing image-based damage detection techniques (Nishikawa et al., 2012; O'Byrne et al., 2014b).

Stereo systems have been used for numerous underwater applications such as robotic control (Sagara et al., 2013), archeology (Liarokapis et al., 2017) and surveying (Williams et al., 2016). These systems consist of two horizontally displaced cameras that simultaneously photograph a scene from slightly different perspectives. By taking photographs of a target from at least two different vantage points, 3D information is encoded in the imagery and can be extracted by applying stereo matching algorithms to the images with minimal human supervision.

The use of two cameras represents a good balance between reliability and operational complexity. 3D shape information can be obtained from a stereo pair once the stereo system has been calibrated, and unlike Structure-from-Motion (SfM) approaches which use just a single unconstrained camera, stereo systems can be used even if the scene does not remain static since both stereo images are captured simultaneously. This is especially relevant for underwater situations as floating particles and the continuously changing light field mean that underwater scenes are seldom static. While there are various other approaches for recovering 3D shape information, such as laser scanning, and coded structured light systems (Bruno et al., 2011), these approaches are often bulky, expensive, and not as easily adapted for underwater use compared with camera-only solutions.

While stereo imaging has the potential to be a convenient and useful underwater tool, the poor underwater visibility conditions diminish the ability of cameras, and subsequent stereo matching techniques, to effectively recover depth information. Underwater images are affected by light attenuation, scattering, colour absorption, suspended particles, and air bubbles, as identified by Massot-Campos & Oliver-Codina (2015) who carried out a survey on optical sensors and methods for 3D reconstruction in underwater environments. These issues lead to blurriness, reduced contrast, and an overall loss of image quality. These problems are even more profound in high turbidity conditions or when intense artificial light sources are employed. Artificial lighting often causes high specular reflections that mask detail and create bright spots that may mislead stereo-matching algorithms.

Many existing stereo algorithms perform poorly when applied to images of underwater scenes due to the poor visibility conditions and the complex underwater light field. Bianco et al. (2013) compared 3D imaging techniques using stereo imaging and structured lighting for underwater scenes with varying turbidity levels. They employed the popular Patch-based Multi-View Stereo (PMVS) algorithm (Furukawa & Ponce, 2010). The authors found that the performance of the PMVS-based stereo matching technique declines markedly in high turbidity conditions.

Liarokapis et al. (2017) used Agisoft's PhotoScan®, a state-of-the-art commercially available software package, for 3D modelling and mapping of underwater archaeology sites. Agisoft's PhotoScan® handles all stages of the 3D reconstruction pipeline, including the dense multi-view stereo matching stage. It generally produces good results, especially when there are a lot of overlapping input images which provides a high-degree of redundancy. However, since the software is geared towards general purpose applications, it cannot be expected to perform optimally when applied to underwater images, especially compared with techniques specifically designed for the task. Moreover, the commercial nature of the software means that users only have limited control over certain parameters and workflows. In this work, we applied Agisoft's PhotoScan®, with highest accuracy settings, to the real-world inspection photos presented in Section 5.3, however, the stereo cameras could not be aligned properly in the first instance and as a result, the rest of the 3D reconstruction workflow could not be implemented. This provides motivation for the development of a stereo matching algorithm that can deal with the poor underwater visibility and produce accurate results using high-resolution images in a reasonable timeframe.

To facilitate the 3D reconstruction of underwater inspection targets, this paper presents a new stereo matching algorithm, called PaLPaBEL. The paper focusses on the application of recovering the shape of structural members colonised by marine growth. The presence of marine growth increases the diameter and roughness of structural members. This introduces several problems; most notably, it increases drag forces and creates hydrodynamic instabilities. These factors lead to a loss in structural performance and reliability, and potentially, shortened lifespans for these structures. Owners/operators, therefore, have a keen interest in tracking the progression of marine growth so that they can optimise cleaning regimes, and so that they have more reliable estimates of the wave and current loading on their structures.

The following section outlines the related work in the field of stereo imaging, especially with regards to underwater application. Following that, the methodology of the proposed technique is described in Section 3. In Section 4, the performance of PaLPaBEL is evaluated on real images under varying visibility conditions and on synthetic imagery. PaLPaBEL is also demonstrated on a colonised mooring line from a real-world inspection. Section 5 concludes the paper.

2 LITERATURE REVIEW

A detailed taxonomy of stereo matching algorithms is provided by Olofsson (2010) and Scharstein and Szeliski (2002). Matching algorithms may be broadly classified into two categories, namely, local and global methods.

Local methods are generally based on computing the matching cost between patches, or small windows, in the left

image (reference image) with potential matching patches in the right image, followed by aggregation of the computed costs. Patches may be compared on the basis of pixel intensity values, texture patterns, or using census/rank transformed data. Common measures of similarity for computing the matching cost include the sum of squared distances, the sum of absolute distances, normalised-cross correlation, and the Hamming distance (Giachetti, 2000). The disparity value that is assigned to each point in the reference image is the disparity associated with the minimum matching cost. Local methods can often run in real-time, making them suitable for a range of on-line applications such as automated inventorying of road signs (Wang et al., 2010) and traffic surveillance (Sappa et al., 2008). However, their computational efficiency comes at the expense of reduced matching accuracy and increased sensitivity to noise.

For underwater inspections, there is often no requirement for real-time processing. Instead, the emphasis lies in obtaining more reliable disparity measurements, which calls for the use of global matching methods. Global methods demonstrate better handling of textureless regions and better tolerance to noise. Global methods usually entail computing the matching cost followed by optimising the disparity map. Disparity optimisation is usually done by minimising some energy function, usually consisting of a correspondence data term and a smoothness term, which penalises cases where adjacent pixels have different disparity values.

Popular types of global methods include dynamic programming (Bai et al., 2009) and Markov Random Field (MRF) based methods. The latter have attracted a lot of recent interest owing to their strong performances in the realm of stereo. Moreover, the emergence of fast algorithms for approximate inference – such as Graph Cuts (Papadakis & Caselles, 2010) and Belief Propagation (BP) (Sun et al., 2002) have enabled MRF-based methods to become computationally viable. These methods have been found to yield good results, largely thanks to how they explicitly model smoothness. However, Li and Zucker (2006) have shown that these methods have drawbacks if the smoothness assumption is violated or if it is inappropriately modelled.

Another issue with MRF-based approaches is that they do not scale well for large disparity ranges and high-resolution images. For instance, an image of size 3000 x 2000 pixels and a disparity range of [0, 299] would require an excessively large array of 3000 x 2000 x 300 x 300 using a conventional BP algorithm applied on an MRF. In contrast, the pyramidal scheme employed by PaLPaBEL means that a more manageable array of size 3000 x 2000 x 5 x 5 is needed.

3 METHODOLOGY

This section provides an overview of stereo imaging. The methodology of PaLPaBEL is then outlined, which includes a description of i) the pyramiding scheme, ii) occlusion handling, iii) matching cost computation, iv) anisotropic diffusion, and v) MRFs and LBP message passing.

3.1 Stereo imaging

To understand how depth is recoverable from a stereo image pair, it is helpful to look at the geometry of a stereo rig. With reference to Figure 1, a point X in the 3D world space is projected onto the left image plane at point x^l and onto the right image plane at point x^r. Working backwards, it is possible to compute the location of this real-world point X by finding the intersection of the rays passing through the left camera's centre of projection and x^l, and passing through the right camera's centre of projection and x^r. This process is known as triangulation and depends on successful matching of corresponding points in the left and right images.

The epipolar geometry provides a helpful constraint that reduces the complexity of this matching problem. For a given x^l in the left image, the epipolar geometry asserts that the corresponding x^r must lie somewhere on the right image plane as well as on the same epipolar plane. The only points that satisfy both constraints are along the intersection of both planes, which is a line known as epipolar line. Hence the search-space in the right image for potential correspondences is drastically reduced from a whole image search to just a single line search. Epipolar geometry is extensively described in (Hartley & Zisserman, 2003).

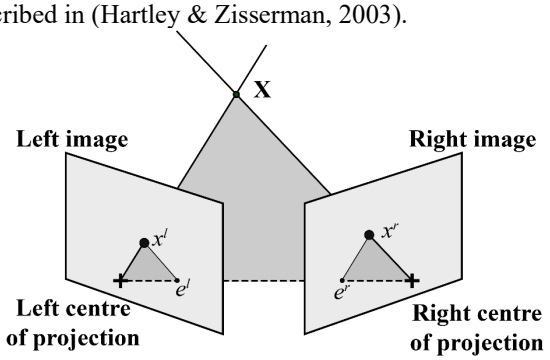


Figure 1 Projection of 3D point X onto left and right images

The process of transforming two images such that the epipolar lines are horizontal, and therefore corresponding points in the left and right images are separated only by a horizontal offset and not by a vertical offset, is known as rectification and is described in (Hartley, 1999). It can be performed either with knowledge of the intrinsic and extrinsic parameters of the stereo system or by computing the fundamental matrix from at least 8 preliminary point correspondences. Once a stereo pair is rectified, a pair of point correspondences can be expressed using the pixel location in the left image $x^l = (u, v)$ and a corresponding horizontal offset, the disparity d, to represent the location of the corresponding point in the right image, $x^r = (u + d, v)$, where u and v indicate the horizontal and vertical spatial coordinates of a pixel, respectively. The distance or depth, z, of the original point in 3D space from the stereo system can be recovered using:

$$z = \frac{f \bullet B}{d} \tag{1}$$

where f denotes the focal length (measured in pixels), B is the baseline (measured in real world units), which is the distance between the two camera centres, and d is the disparity (measured in pixels). Thus, it is sufficient to find the disparity between corresponding points in the left and right images to recover depth information. However, identifying corresponding points is non-trivial, and a wide range of techniques have been devised to solve this problem.

3.2 Overview of procedure

The methodology of PaLPaBEL is illustrated in the flowchart in Figure 2.

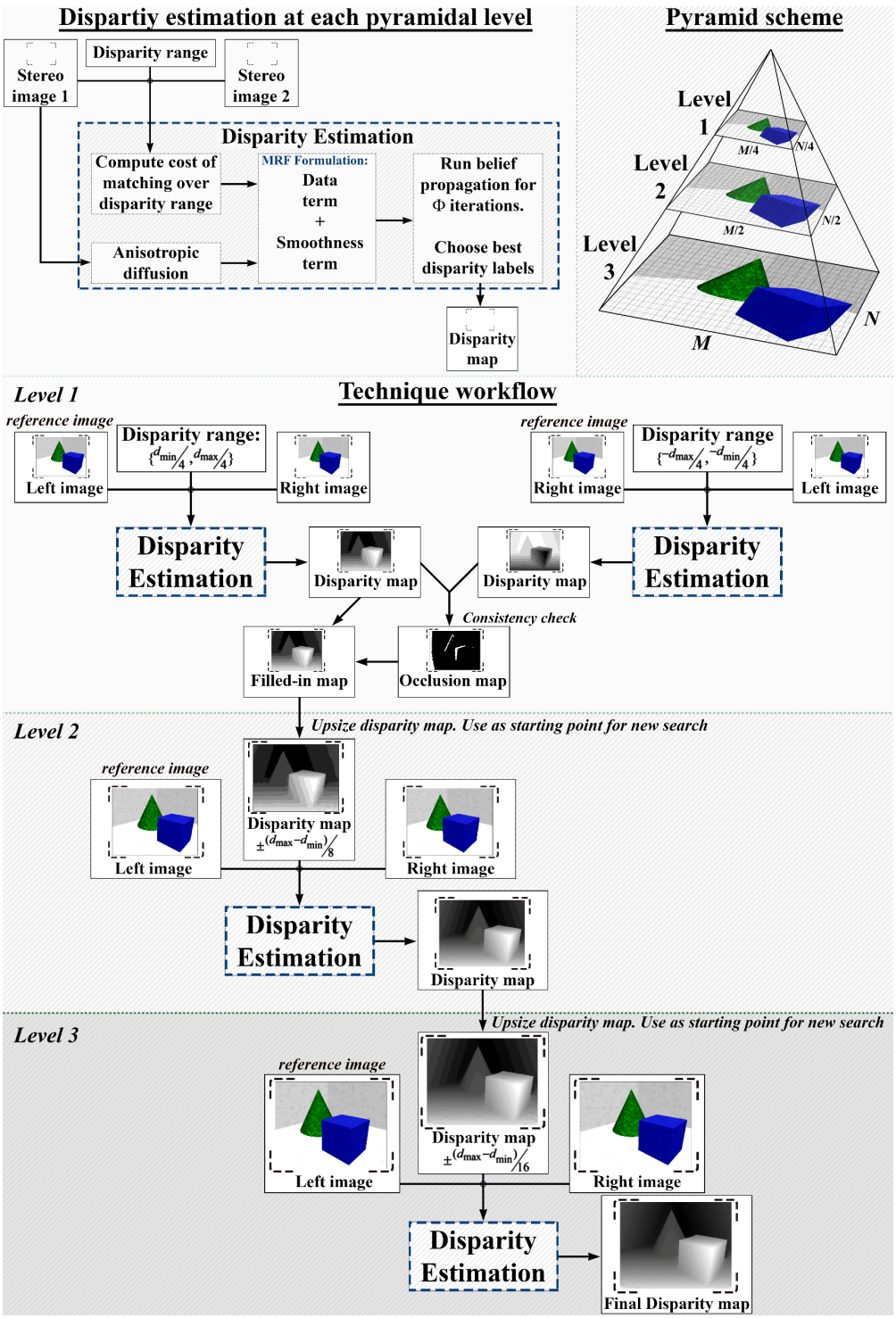
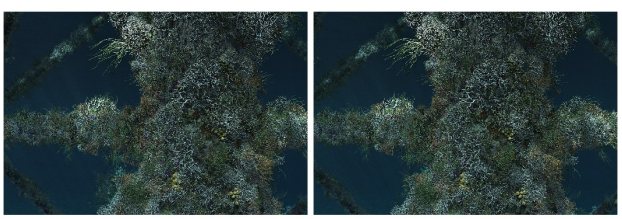


Figure 2 PaLPaBEL flowchart

Given two images of the same scene, PaLPaBEL seeks to match pixels in one image with the corresponding pixels in the other image. The input for PaLPaBEL is a rectified stereo image pair. For illustrative purposes, PaLPaBEL is demonstrated on the stereo pair shown in Figure 3. These synthetic images are renders of a virtual underwater scene and they depict marine growth on a jacket-type structure. To start, the left and right input images are down-sampled several times. The extent of down-sampling depends on the number of pyramid levels, K, and the scaling factor, s. Using a larger scaling factor and fewer pyramid levels produces a more aggressive convergence to the final solution while a smaller scaling factor with more pyramid levels results in a steadier but more time-consuming progression to the final solution.



(a) **(b)**

Figure 3 (a) Left, and (b) Right synthetic images of a virtual marine growth colonised structure

For simplicity, three pyramid levels ($K = 3$) and a scaling factor of $s = 2$ are used for all images shown in this paper. This means that the image resolution at the first pyramid level is a quarter of the full resolution and the resolution at the second pyramid level is a half of the full resolution. The image resolution at the third and final pyramid level uses the full image resolution. While these values typically work well, the number of pyramid levels, K, can be generalised to any natural number, $K = \{1, 2, 3, ...\}$, and the scaling factor, s, can by any real number satisfying $s \geq 1$. Finding the optimal values for K and s calls for a detailed exploration of scale-space analysis that is beyond the scope of this paper. However, a scaling factor of $s = 2$ is frequently used elsewhere in the literature, as is $s = \sqrt{2}$ which achieves greater interscale continuity at the expense of slower convergence (Cyganek & Siebert, 2009).

At the first pyramid level, two disparity maps are computed. The first disparity map is created using the right image as the reference image and the search for corresponding points takes place in the left image. This situation is then reversed as the left image acts as the reference image and the search for corresponding points takes place in the right image to create a second disparity map. Two disparity maps are computed in order to check for occluded points. Occluded points refer to parts of the scene that are visible in one image but not in the other. Ideally, the two disparity maps should be consistent with one another; however, discrepancies often arise in textureless regions and at occluded points. Points, where such discrepancies occur, are considered to be occluded points, and the estimated disparity values at these points are regarded as being invalid, and consequently, they are discarded. They are replaced with the minimum valid disparity value from neighbouring regions. By identifying and filling in occluded regions, the overall quality of a disparity map is improved. Carrying out occlusion-handling at the first pyramidal level can be done quickly as the stereo images are significantly downsized at this stage. The disadvantage is that we obtain a cruder estimate of occluded regions, however, results are usually adequate.

The process of disparity estimation centres on minimising an energy cost that is composed of a correspondence data term and a smoothness term. The correspondence data term comes from computing a measure of similarity between points in both images of the stereo pair. In the case of PaLPaBEL, the similarity measures are the Zero-mean Normalized Cross-Correlation (ZNCC) score using colour information and the Hamming distance using census transformed data. These similarity measures were chosen as they can tolerate noise and can handle difficult lighting situations. The smoothness term penalises adjacent pixels that have different disparity values. Having a superior smoothness term has a significant bearing on the overall success of the stereo technique. For many stereo techniques, the weight of the smoothness term is controlled by the gradient magnitude in the original image. If the gradient magnitude is large, which often occurs at an object boundary, then the smoothness term does not overly penalise disparity variations. This reflects real-life situations as depth discontinuities in a scene tend to occur at object boundaries. Conversely, if the gradient magnitude is small, as would be expected for neighbouring pixels on the same object, then abrupt disparity changes are strongly penalised (but not prohibited). However, problems arise when the imagery is affected by noise as it becomes hard to distinguish between real and spurious edges. Consequently, the smoothness term is less effective. PaLPaBEL models the smoothness based on the gradient magnitude from an anisotropically diffused version of the original image. Anisotropic diffusion reduces noise without removing real edges. This leads to a more robust smoothness term, and ultimately, a better disparity map.

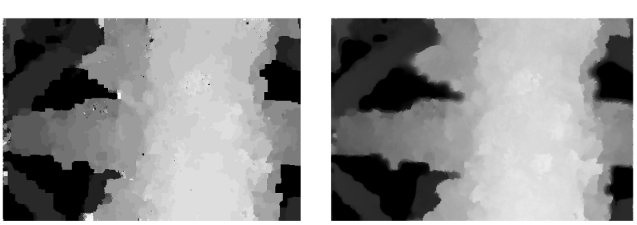
An optimal disparity map with minimum energy is approximated using Loopy Belief Propagation (LBP) where the message passing takes place between adjacent segments on a Markov Random Field (MRF), as discussed in greater detail in Section 3.7. The output disparity map from the first pyramid level is upsized by the scaling factor, s, and disparity value at each point in this upsized disparity map is used as an initial offset at the next pyramid level,

from which the new, and more precise, correspondence search will be centred on. The new search is carried out over a smaller disparity range. This process is continued until the highest level is reached, and a disparity map is obtained that is the same resolution as the original image.

The strength of PaLPaBEL lies in how takes advantage of the various components of the algorithm. The pyramidal scheme adds computational efficiency and it enables expensive tasks like determining occluded regions to be performed at the lowest pyramid scale where it requires much less time while the results remain acceptable. The choice of matching metrics (ZNCC and census transform-based matching) are carefully chosen as they perform well in harsh conditions typical of underwater scenes. Anisotropic diffusion is incorporated as it offers an effective way to model the smoothness term as part of the MRF formulation. Finally, the belief propagation applied on an MRF is chosen as it is amongst the leading strategies for stereo matching techniques. Considered collectively, these components tie together to create a technique that is well-suited to the demands of underwater inspections where traits such as resilience to poor visibility conditions, efficiency, the ability to scale well to high-resolution images, and strong performance are highly valued.

3.3 Pyramiding scheme

The primary challenge of using MRFs as part of a pyramidal scheme is accommodating high disparity values alongside small disparity values with access to only a limited number of labels. This proposed approach achieves this by disregarding the absolute disparity values and only concentrating on disparity values relative to the disparity computed at the preceding pyramid level. To begin with, the images are down-sampled a number of times, which in the case of images presented in this paper is 3 times. The disparity map computed at the first level is propagated up the pyramid to the next level. Here, the disparity map is upsized by the scaling factor, *s*, and it is used as an offset for the next higher level. The disparity values at each pixel act as the centre point on which the new search for a more precise disparity estimate is carried out. This is shown in Figure 4, where the disparity map at the first level in Figure 4(a) is notably cruder than the final map in Figure 4(b).



(a) (b)

Figure 4 (a) Disparity map at the first pyramid level, and (b) disparity map after the third pyramid level.

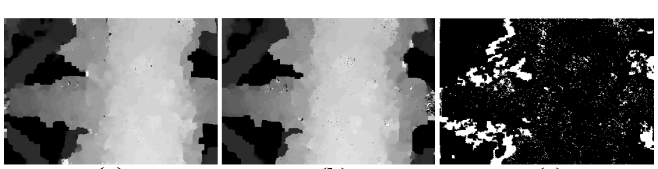
The new search takes place over a smaller range of disparities. If the original image size, M x N, is 3000 x 2000 pixels, the number of discrete integer disparity levels L is 300 (for a disparity search range of [0, 299]), the scaling factor s is 2, and the number of pyramid levels K is 3, then at the first pyramid level k =1, a search over a relatively small 75-pixel range is required to estimate disparities for an image of size (750 x 500) pixels. At the second level, when k=2, the image is upsized to 1500 x 1000 pixels, however, the disparity search range is reduced to 38 pixels. At the third and final pyramid level, when k=3, the image size returns to the full resolution. The disparity search range takes place over a 19-pixel range, and the output is the final disparity map. The results from this pyramidal approach are high quality and are achieved at a low computational cost.

3.4 Occlusion handling

Since stereo imaging involves capturing two images of a scene from different vantage points, there will be parts of the scene that will be visible in only one of the cameras. Pixels corresponding to these parts of the scene will therefore not have a match in the other image of the stereo pair. Occlusion handling is a way to detect these pixels and assign a reasonable disparity value to them. The simplest method for detecting occluded pixels is by mutual cross-checking (Ho & Jang, 2013). The left disparity map, D^l, and the right disparity map, D^r, that are computed after the first pyramid level act as the inputs for this check. The check tests whether disparity values in the left and right disparity maps are consistent with one another, i.e. that $D^l(u,v) \approx D^r(u + D^l(u,v), v)$. This is likely to hold true if pixels are not occluded and the estimated disparity values are reliable. Pixels that violate this expression are regarded as unreliable matches/occluded pixel. An occlusion map, O^l, is created according to Equation 2:

$$O^l(u,v) = \begin{cases} 0 & \text{if } \left| D^l(u,v) - D^r(u + D^l(u,v), v) \right| \le w \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

This sets points that do not map back and forth to each other, within a predefined threshold w, to 1. PaLPaBEL uses a value of w=4, meaning a disparity discrepancy of ± 4 pixels is tolerated. Otherwise the $O^l(u,v) = 0$. Left and right disparity maps are shown in Figure 5(a) and 5(b), along with the occlusion map in Figure 5(c).



(a) (b) (c)

Figure 5 (a) Left disparity map, (b) Right disparity map, and (c) Occlusion map (occluded pixels are white).

Occluded pixels are replaced with the minimum valid disparity value from neighbouring regions. This is achieved by extrapolating disparity values from valid, non-occluded regions, onto occluded regions in a process known as background filling (Patwardhan et al., 2005).

The occlusion handling process is confined to the first pyramid level. This is because it is quick to compute two disparity maps at this stage – one disparity map with the left image as the reference image and a second disparity map with the right image as the reference image, given that the input images are significantly downsized. Moreover, Occlusion map O^l acts as a visibility mask, in which the identified occluded pixels do not undergo any further refinement for successive pyramid levels.

3.5 Matching cost computation

PaLPaBEL uses a Zero-mean Normalized Cross-Correlation (ZNCC) function to compare pixel intensity values (RGB colour information) in the left and right images. On top of this, the Hamming distance metric is used to measure the quality of matches from the census transformed left and right images. The total cost, C, associated with matching a point (u,v) in the left image, I_{left}, to a point $(u+d,v)$ in the right image, I_{right}, is found by combining the negative ZNCC score, Z, with the Hamming distance, H, both of which are scaled to the range [0, 1] so that they have an equal contribution to the overall cost, C, as per Equation 3:

$$C(u,v,d) = Z(u,v,d) + H(u,v,d) \tag{3}$$

where, as before, d denotes the disparity value. Computing the ZNCC score and the Hamming distance are discussed in the following subsections.

3.5.1 Zero-mean Normalised Cross-Correlation (ZNCC)

The ZNCC metric is a widely-used measure of similarity. It is capable of tolerating variations in image brightness, which makes it a practical choice for situations where challenging lighting conditions exist, such as in an underwater environment, or when the left and right cameras in a stereo system have different exposure settings.

Calculating the ZNCC score between a point in the left image and points in the right image involves the use of a sliding window. The window moves horizontally over the predefined disparity search range in the right image. At every disparity value, the window in the right image is compared with the reference window in the left image. The ZNCC score has a high value when there is a high degree of similarity between the two windows. However, to suit our purposes, we reverse the sign of the ZNCC score to express it as a cost, whereas now similar points have a low matching cost, while dissimilar points incur a large cost. The negative ZNCC score, Z, and its various parts can be expressed as follows:

$$Z(u,v,d) = \sum_{c\in\{R,G,B\}} \frac{1-\mathrm{cov}_{u,v,c,d}(l,r)}{2\times\sqrt{\mathrm{var}_{u,v,c}(l)\times\mathrm{var}_{u+d,v,c}(r)}} \tag{4}$$

$$\mathrm{cov}_{u,v,c,d}(l,r) = \sum_{i=-m}^{m}\sum_{j=-n}^{n} \frac{\left(l_{u+i,v+j,c}-\overline{l}_{u,v,c}\right)}{\left(r_{u+d+i,v+j,c}-\overline{r}_{u+d,v,c}\right)^{-1}} \tag{5}$$

$$\mathrm{var}_{u,v,c}(l) = \sum_{i=-m}^{m}\sum_{j=-n}^{n} \left(l_{u+i,v+j,c}-\overline{l}_{u,v,c}\right)^2 \tag{6}$$

$$\mathrm{var}_{u+d,v,c}(r) = \sum_{i=-m}^{m}\sum_{j=-n}^{n} \left(r_{u+d+i,v+j,c}-\overline{r}_{u+d,v,c}\right)^2 \tag{7}$$

where $l_{u+i,v+j,c}$ and $r_{u+i,v+d+j,c}$ represent the pixel intensity value in the c^{th} colour channel (red, green, or blue) at the point (u,v) in the left image and the point $(u+d,v)$ in the right image, respectively. $\overline{l}_{u,v,c}$ and $\overline{r}_{u+d,v,c}$ represent the average intensity of pixels in the c^{th} colour channel in a window of size $(2m+1)$ x $(2n+1)$ in the left image and the right image, respectively. A window size of 5x5 is used throughout this paper. It represents a reasonable balance between capturing the intensity variation within a window while keeping the window specific to the point on which it is centred on. Strategies for choosing an optimal window size are discussed by Kanade and Okutomi (1994).

3.5.2 Census Transform and Hamming Distance

The ZNCC function and the Hamming distance complement each other – while the ZNCC function accounts for the similarity in terms of colour information, the Hamming distance applied to census transformed data captures local structure/patterns in the image. Combining these two metrics creates a richer and more robust metric. Like the ZNCC metric, census transform-based matching is known to be robust to radiometric distortion since global lighting differences between two images will not affect the ordering of pixels at a local level. The census transform summarises local image structure. It is based on the relative ordering of local intensity values and not on the intensity values themselves. This aspect means that it can tolerate outliers and perform better near object boundaries compared with only using colour data (Tavera-Vaca et al., 2015). Also, in the context of stereo estimation, Lee et al. (2016) have demonstrated the usefulness of the census transform under harsh lighting conditions.

The census transform is applied on grayscale versions of the left and right images. A grayscale image, G, is obtained by averaging the three colour channels of an input colour image, I, at each pixel location. Census transform-based matching involves two steps: applying the transform and computing the Hamming distance. The census transform is realised by comparing the grayscale intensity value of a pixel at the centre of a window with other pixels in the window. Thus, the census transform captures how a pixel in an image relates to its neighbouring pixels. A window of size $(2m+1)$ x $(2n+1)$ is centred at a point (u,v) in the

grayscale image. If the pixel intensity value at the centre of the window, $G(u,v)$ is greater that the intensity value at some other point in the window, $G(u+i,v+j)$, then the census transform records this relationship as a 1, otherwise, the relationship is assigned a value of 0, as per Equation 8:

$$CT(u,v)_{ij} = \begin{cases} 0 & G(u,v) \le G(u+i,v+j) \\ 1 & G(u,v) > G(u+i,v+j) \end{cases} \quad (8)$$

where $i \in \{-m,...,m\}$ and $j \in \{-n,...,n\}$

where CT is the census-transformed image consisting of ones and zeros. As with computing the ZNCC score, a window size of 5x5 is employed throughout this paper. For this window size, the array that is required to store the census value at every point (u,v) has 5x5=25 elements, although strictly speaking, only 24 comparisons are necessary since the relationship between the centre pixel in the window with itself will always be the same.

The matching cost is computed using the Hamming distance as a measure of similarity. The Hamming distance is computed by performing an XOR (exclusive or) operation between a point in the census-transformed left image CT_{left} and a point in the census transformed right image CT_{right} that lies within the disparity search range. This entails counting the number of times that an element of $CT_{left}(u,v)_{ij}$ differs from the corresponding element of $CT_{right}(u+d,v)_{ij}$, as given by Equation 9:

$$H(u,v,d) = \sum_{i=-m}^{m} \sum_{j=-n}^{n} \frac{CT_{left}(u,v)_{ij} \oplus CT_{right}(u+d,v)_{ij}}{(2m+1)(2n+1)} \quad (9)$$

where $H(u,v,d)$ is the Hamming distance at the point (u,v) for a disparity value, d, and $\oplus$ denotes the XOR operation that counts the number of positions at which the corresponding values differ. Figure 6 shows an example of census transform-based matching for a 3x3 size window.

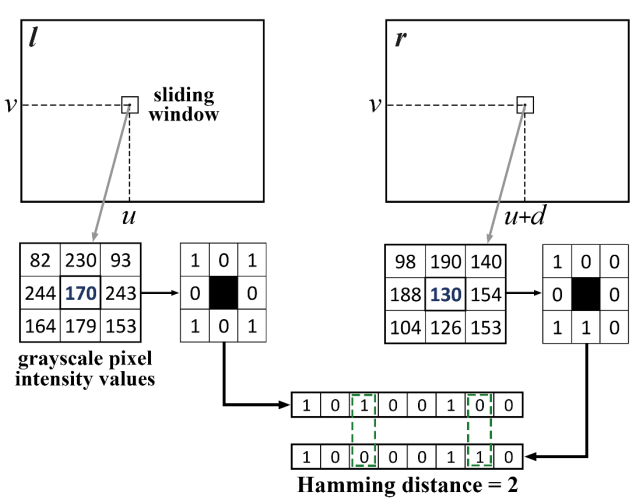


Figure 6 Census transform example

3.6 Anisotropic diffusion

Anisotropic diffusion is a smoothing method that encourages smoothing within a region as opposed to smoothing across region boundaries. It allows edges to remain sharp while smoothening out noise and other image artefacts. In particular, at the initial pyramid levels when the input images are down-sampled, the anisotropic diffusion softens light edges and surface texture while preserving strong edges that often coincide with the general outline of objects. Light edges, which may be from small protrusions on the surface of an object, can often be captured at the final pyramid level when imagery is at the full resolution and anisotropic diffusion is more sensitive to fine, local details. Anisotropic diffusion involves taking a grayscale image, G, as input and iteratively applying the following diffusion equation, as per Equation 10:

$$\begin{aligned} G^{t+1}(u,v) = G^t(u,v) + \lambda[&h^t_{North}(u,v) \cdot \nabla_{North} G^t(u,v)... \\ &+h^t_{South}(u,v) \cdot \nabla_{South} G^t(u,v)... \\ &+h^t_{East}(u,v) \cdot \nabla_{East} G^t(u,v)... \\ &+h^t_{West}(u,v) \cdot \nabla_{West} G^t(u,v)] \end{aligned} \quad (10)$$

where λ controls the speed of diffusion, which should be in the range $0 \le \lambda \le 1/4$ to maintain numerical stability. A value of $\lambda = 0.2$ is assumed in this paper. *North*, *South*, *East* and *West* indicate the direction in which the conduction coefficients h act. The symbol ∇ represents the difference between adjacent pixels, as per Equation 11:

$$\begin{aligned} \nabla_{North} G^t(u,v) &= G^t(u,v-1) - G^t(u,v) \\ \nabla_{South} G^t(u,v) &= G^t(u,v+1) - G^t(u,v) \\ \nabla_{East} G^t(u,v) &= G^t(u+1,v) - G^t(u,v) \\ \nabla_{West} G^t(u,v) &= G^t(u-1,v) - G^t(u,v) \end{aligned} \quad (11)$$

These conduction coefficients are updated at every iteration according to Equation 12:

$$\begin{aligned} h^t_{North}(u,v) &= e^{(-|\nabla_{North} G^t(u,v)|/\alpha)} \\ h^t_{South}(u,v) &= e^{(-|\nabla_{South} G^t(u,v)|/\alpha)} \\ h^t_{East}(u,v) &= e^{(-|\nabla_{East} G^t(u,v)|/\alpha)} \\ h^t_{West}(u,v) &= e^{(-|\nabla_{West} G^t(u,v)|/\alpha)} \end{aligned} \quad (12)$$

The constant α can be fixed at some value or it can be automatically based upon an estimate of the noise as described by Canny (1986). Images in this paper were obtained using a value of α = 0.1. The conduction coefficients are small when there is a large difference between adjacent pixels, e.g. at high-contrast boundaries.

As a result, pixels on either side of a boundary will only share minimal information with each other, and the boundary will remain largely intact after diffusion.

Anisotropic diffusion is quick to compute. For this reason, 20 iterations were carried out (t = 20). The results of anisotropic diffusion are shown in Figure 7. While the input 'noisy' image and the output smoothed image look perceptually similar, the effect of anisotropic diffusion becomes apparent when looking at a close-up section.

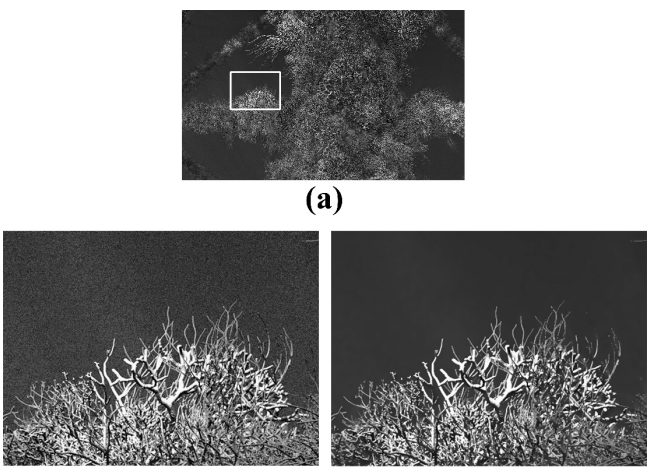


Figure 7 (a) Input grayscale image, close-up view of (b) original, and (c) anisotropic diffused image.

It may be observed from Figure 7 that real edges are preserved while background noise is suppressed. It is at the edge boundaries where disparity discontinuities are likely to occur. For this reason, the anisotropically diffused image acts as a good basis for the smoothness term for the next stage which involves MRF-based disparity optimisation.

3.7 MRF formulation and LBP message passing

Disparity map estimation is formulated on an MRF and solved using LBP. An MRF is a graphical model of a joint probability distribution consisting of visible and hidden nodes. Visible nodes are present at each pixel location and they represent information that we know or that we can compute (e.g. pixel colour). Hidden nodes represent unknown values (e.g. disparity). Statistical dependencies between hidden variables are expressed by explicitly associating hidden variables with one another. The value of MRFs stem from the fact that they enable local explicit relationships between relatively few pairs of pixels to give rise to implied longer-range relationships. MRFs have application in many image processing tasks, including segmentation (Nguyen & Wu, 2013), fusion (Sun et al., 2013), and stereo matching (Sun et al., 2003). For stereo matching purposes, MRFs enable spatial dependencies between nearby pixels to be encoded. MRFs are comprised of nodes and links and may have loops (i.e. they may be cyclic), which is why belief propagation algorithms applied on these networks are sometimes referred to as loopy belief propagation algorithms. The presence of loops means there is no guarantee of convergence to the optimal global solution regardless of the number of iterations of the belief propagation algorithm. Despite this, belief propagation approaches are known to yield good approximate solutions. An example of how MRFs can be used to model the stereo problem is presented in Figure 8. $B=\{b\}$ represents the observed nodes (the matching costs), and $A=\{a\}$ represents the hidden nodes (the unknown disparity labels).

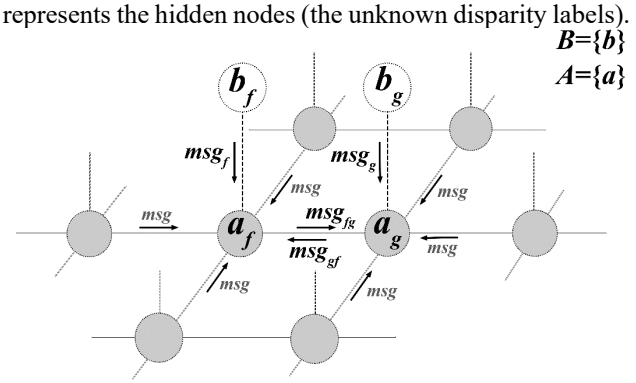


Figure 8 Message passing on a MRF. Grey nodes are hidden variables. White nodes are observable variables.

For simplicity, pixel locations in the image are denoted by a single index, *f*, as opposed to distinct horizontal and vertical spatial indices, (*u*,*v*), as before. It follows that *f* may be expressed by $f=u+(v-1)\times M$ where *M* is the image width. For every pixel in the image, there exists an observed node b_f that represents the matching cost (as computed in Section 3.2), and a corresponding hidden node, a_f, which represents the unknown disparity value. The observed nodes are white and the hidden nodes are grey in Figure 8. The links between each node indicate a dependency. It may be noted that while the hidden variables are connected to their neighbours, the observed data points are connected only to a single hidden variable. As such, the estimated disparity at a point depends on the estimated disparity values of the four immediately adjacent pixels as well as the observed matching cost.

An optimal disparity map with minimum energy is approximated using Loopy Belief Propagation where the message passing takes place between adjacent segments. There are various kinds of BP algorithms with different message update rules: max-product, sum-product, and min-sum (Szeliski et al., 2006). PaLPaBEL uses a min-sum form where the objective is to minimise the energy function consisting of a matching cost term and a smoothness term:

$$E(B,A)=\sum_{f} C_f(p_f)+\sum_{(f,g)\in L} S(p_f,p_g) \qquad (13)$$

where E is the energy associated with having a disparity label p_f at every node a_f. $C_f(p_f)$ gives the cost for assigning a disparity label p_f at the point f in the image, as found from Equation 2. The second term, S, is the pairwise term, or the smoothness term, which penalises instances of neighbouring pixels f and g having different disparity labels, where g is the spatial index of a point that is directly adjacent to f. The smoothness term is based on the anisotropically diffused image, as per Equation 14:

$$S(p_f, p_g) = e^{\left(-\mu\left|G_f^T - G_g^T\right|\right)} \times \min\left(\left|p_f - p_g\right|, \kappa\right) \quad (14)$$

where G_f^T and G_g^T are the intensity values of the anisotropically diffused image (after T iterations) at the points f and g, respectively. The smoothness cost promotes consistent and steady labelling across neighbouring hidden nodes. It increases as the difference between disparity labels p_f and p_g increases. The shape of the cost function is controlled be a constant μ and by the truncation threshold κ. The threshold κ prevents adjacent disparate labels from being overly penalised. Including this parameter often leads to a better and more robust solution. The results obtained in this paper use values of $\mu = 50$ and $\kappa = 3$. The smoothness cost is low when neighbouring pixels are assigned similar disparity values (i.e. p_f and p_g are similar) or if there is a large difference between G_f^T and G_g^T which may occur at an edge boundary in the image. On a related note, using a smoothness term like this one has the advantage of being eligible for expedited BP computation, as outlined by Coughlan and Shen (2010).

Let msg_{fg}^{ϕ} be the message passing vector that node a_f sends to a_g, on the ϕ iteration of the BP algorithm. It is defined over each label q by Equation 15:

$$msg_{fg}^{\phi}(q) = \min_{p}\left(E(B, A) + \sum_{r \in J(f) \setminus g} msg_{rf}^{\phi-1}(p)\right) \quad (15)$$

where r is the index of pixels that are members of $J(f)$ - the local neighbourhood of a_f, not including b_f. q is the disparity label for node a_q, which is a neighbour of a_p, $p, q \in \{d_{\min}, ..., d_{\max}\}$, where d_{min} is the minimum integer disparity label and d_{max} is the maximum integer disparity label. After Φ iterations, the best disparity label at every pixel can be found by calculating the belief vector bel_f, which is defined over each label p by Equation 16:

$$bel_f(p) = C_f(p) + \sum_{g \in J(f)} msg_{gf}^{\phi}(p) \quad (16)$$

The label p_f corresponding the smallest component of bel_f is taken as the MAP solution for the node a_f. As with most iterative algorithms, it can be run for a fixed number of iterations or terminated when the change in energy drops below a certain threshold. In the case of PaLPaBEL, the label p_f is selected after 20 iterations (Φ =20), and is assigned to the disparity map D at the point f, i.e., $D_f = p_f$.

4 RESULTS AND DISCUSSION

This section presents the results following application of PaLPaBEL on synthetic imagery, and on imagery captured in a laboratory setting. The performance of PaLPaBEL is compared with existing techniques. PaLPaBEL is also applied to stereo images from a real-world inspection.

4.1 Synthetic imagery

This paper demonstrates the PaLPaBEL technique on synthetic data as it provides a way of imitating the texture and irregular shape of natural marine growth in an underwater setting while retaining complete control of the scene and the stereo cameras. This cannot be achieved for laboratory based experiments, such as those carried out by Gaurier et al. 2014, where the shape of the artificial marine growth is known yet it does not appear natural, and errors from external sources, such as camera imperfections, make performance evaluation inexact as external problems may contribute towards incorrect matches. The synthetic data consists of a virtual 3D model of a jacket-type platform that was colonised by marine growth, as shown in Figure 3.

The 3D model was set in a physically accurate underwater environment, where the common issues affecting underwater optical imaging systems were present, such as scattering and colour absorption. Other factors that occur in real world scenes, such as luminous complexities (e.g. bright spots caused by artificial lighting, surface reflections etc.) and turbidity were also accounted for.

This section compares the performance of PaLPaBEL with two other matching methods algorithms. The first method is a simple block matching approach based on the well-known Birchfield and Tomasi matching cost that is insensitive to image sampling (Birchfield & Tomasi, 1998). This method compares each pixel in the reference image against a linearly interpolated function of the other image. It does not rely on any smoothness constraints, but rather, the disparity is computed by selecting the minimal (winning) aggregated value at each pixel.

The second method is a region based stereo matching algorithm by Alagoz (2008), which operates by semi-global error energy minimization by smoothing functions. This method chooses a root point in a region and then grows that point while the energy function remains equal or less than a certain value. Otherwise, a new root point is selected in another region. This algorithm employs the Sum of Absolute Differences (SAD) as a similarity measure.

These techniques are applied to the stereo pair shown in Figure 3 and the resulting disparity maps are shown in Figure 9. The performance is evaluated on the basis of the Normalised Root-Mean-Square-Error (NRMSE) score. Normalising the Root-Mean-Square Error facilitates the comparison between datasets or models with different

scales. This is achieved by dividing the Root-Mean-Square Error by the number of disparity levels that is being considered by a given stereo pair. A low NRMSE score is indicative of a strong performing stereo matching algorithm as it corresponds to cases where there is a smaller deviation between the computed disparity map and the ground truth disparity map. The NRMSE scores are summarised in Table 1. The performances are discussed in Section 4.4.

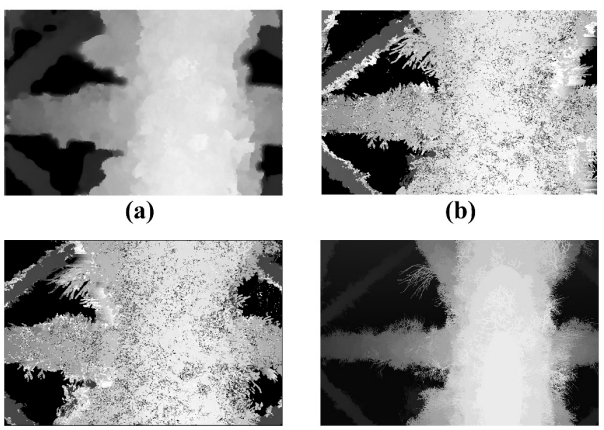


Figure 9 Disparity maps for (a) PaLPaBEL, (b) Block-matching using Birchfield and Tomasi, (c) Semi-global matching by Alagoz (2008), and (d) ground truth

Table 1
Performance of the methods for the synthetic imagery

Method	*NRMSE*
PaLPaBEL	0.05
Block Matching	0.20
Semi-Global Matching	0.14

4.2 Laboratory controlled imagery

These techniques are applied to stereo images that are taken from a large Underwater Lighting and Turbidity Image Repository (ULTIR) (O'Byrne et al., 2017). The images were captured using a stereo system comprised of two synchronised Canon EOS 600D DSLR cameras. Both cameras shared identical settings (e.g. aperture, ISO, shutter speed). The stereo images feature a range of submerged specimens that were photographed under varying controlled levels of turbidity and lighting. The quality of these photographs is assumed to be chiefly affected by luminosity, sharpness (focus accuracy), contrast and noise. These quality factors are directly related to the on-site operating conditions, for which lighting and turbidity are the most influential.

Turbidity is the cloudiness in a liquid caused by the presence of suspended solids. These suspended solids scatter and absorb light and therefore reduce visibility. Lighting is also crucial for achieving good visibility. Ambient light may be sufficient for near-surface inspections; however, it is unlikely to be sufficient at greater depths at which point artificial light sources become necessary. A side-effect of these artificial light sources is that they introduce luminous complexities such as 'bright-spots', which appear as over-exposed areas in the imagery. Such bright spots may fool the damage detection algorithms. The images in ULTIR were captured under three levels of turbidity were chosen (0 NTU (clear water), 6 NTU, and 12 NTU) and three lighting levels were used (100 Lux, 1000 Lux, 10000 Lux. An example of one of the specimens – a concrete cube - is shown in Figure 10. PalPaBEL, the block matching method and the semi-global method are applied to these stereo images and the estimated disparity maps are shown in Figure 11.

The stereo matching techniques are applied to all the stereo images contained in ULTIR. The performance is based on comparing the disparity maps with a ground truth, which is built by visually matching support points (e.g. corresponding vertices of the cube) and extrapolating between points to generate a full-field disparity map. The aggregated NRMSE scores for all specimens in the repository under each lighting and turbidity are shown in Figure 12. Specimen descriptions are provided in Table 2.

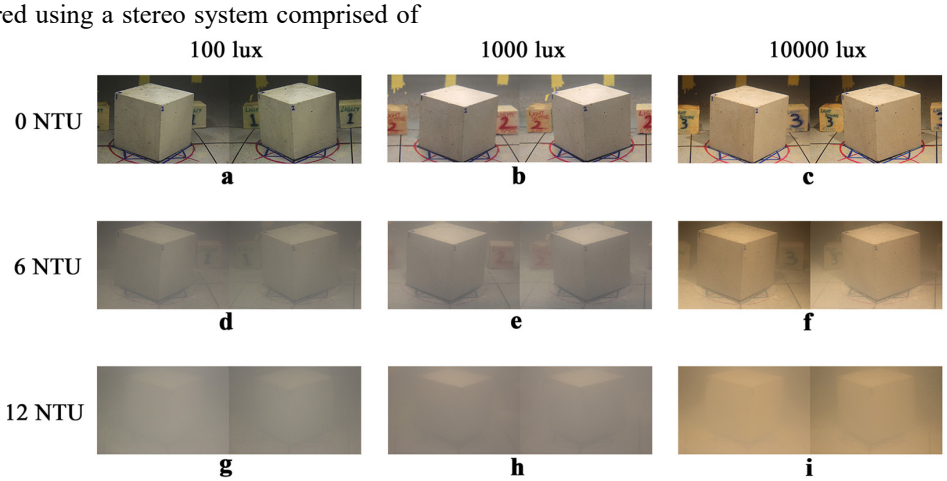


Figure 10 Stereo pairs featuring a concrete cube shown under varying lighting and turbidity levels. Columns: Low (100 lux), Medium (1000 lux), High (10000 lux) Light. Rows: Low (0 NTU), Medium (6 NTU), High (12 NTU) Turbidity.

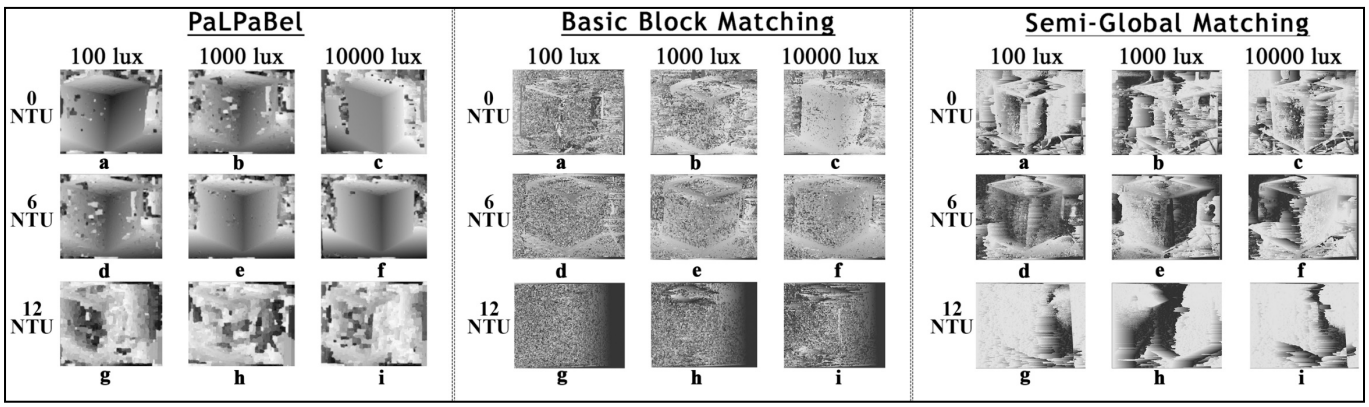


Figure 11 Disparity maps for the stereo imagery in Figure 7

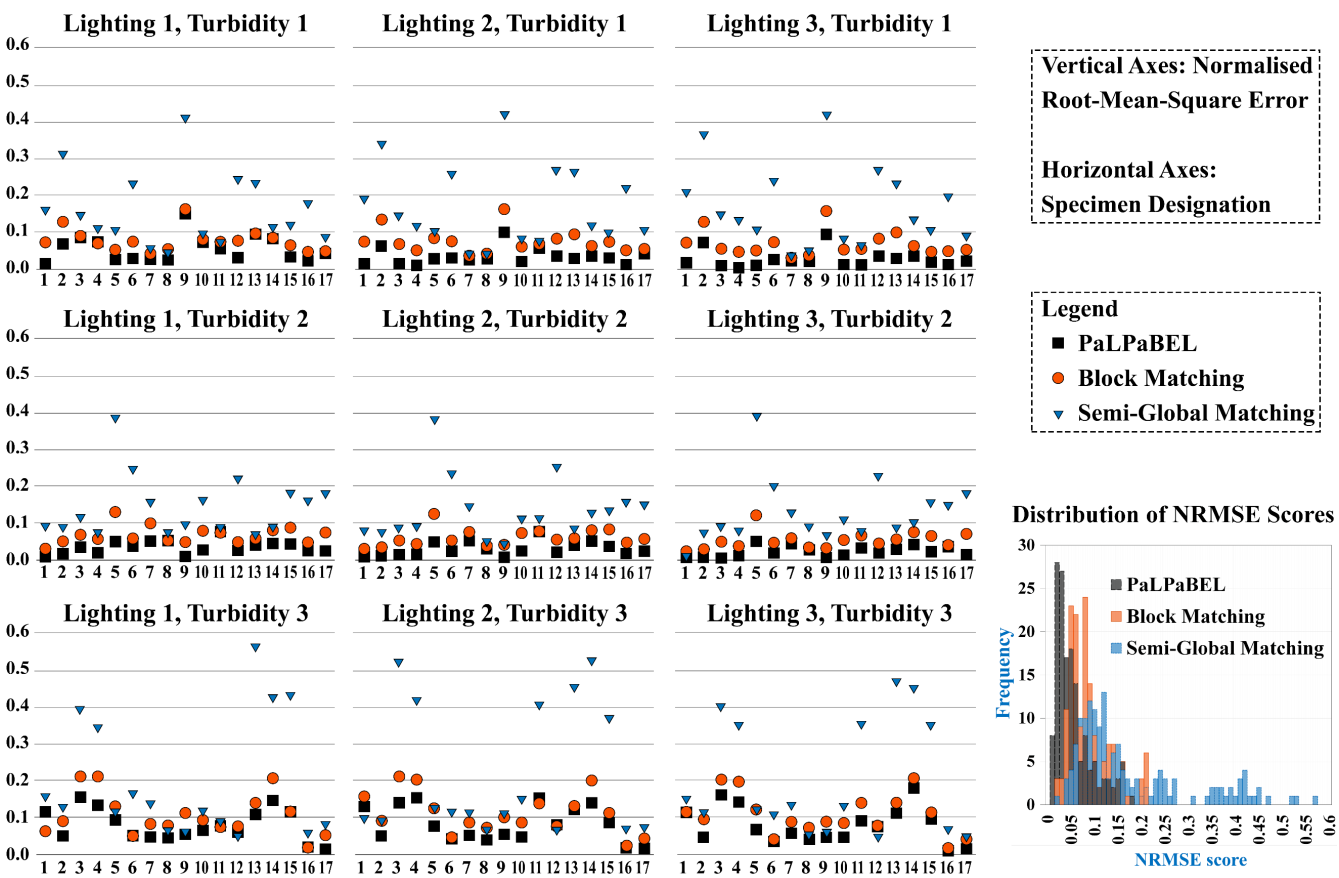


Figure 12 The aggregated NRMSE scores for all the specimens in the repository under each lighting and turbidity

Table 2
Description of specimens

Specimen No.	*Material*	*Surface Texture*	*Surface Finish*	*Orientation to Camera*	*Curvature*
1	Concrete	Rough	Matte	Normal	Planar
2	Concrete	Rough	Matte	Angled	Planar
3	Concrete	Smooth	Matte	Normal	Planar
4	Concrete	Smooth	Matte	Angled	Planar
5	Concrete	Smooth	Matte	Normal	Cylindrical
6	Concrete	Smooth	Matte	Normal	Spherical
7	Metal	Rough	Satin	Angled	Planar
8	Metal	Rough	Satin	Normal	Planar
9	Metal	Smooth	Glossy	Normal	Cylindrical
10	Metal	Smooth	Glossy	Angled	Planar
11	Metal	Smooth	Glossy	Normal	Cylindrical
12	Metal	Smooth	Glossy	Normal	Spherical
13	Plastic	Smooth	Satin	Normal	Planar
14	Plastic	Smooth	Satin	Angled	Planar
15	Plastic	Smooth	Satin	Normal	Cylindrical
16	Plastic	Smooth	Satin	Normal	Spherical
17	Rubber	Rough	Matte	Normal	Cylindrical

The specimens described in Table 2 pose various challenges for the stereo algorithms. Uniformly textured surfaces, such as the smooth plastic surfaces (e.g. specimen no.13), test the ability of stereo algorithms to handle ambiguous matches. Shiny surfaces, such as the glossy metallic surfaces (e.g. specimen no. 9), test how well the algorithms perform in the presence of lighting complexities and specular reflections. The specimen shape and orientation affect the projectivity between corresponding patches on the specimen in the left and right images. For instance, a patch on a flat surface that is normal to the cameras should appear similar in both the left and right views, whereas, a patch on a curved body, or on a surface that is non-normal to the cameras, should appear more distorted in the other image.

4.3 Real-world example

This section demonstrates the real-world utility of PaLPaBEL by applying it to a stereo image pair featuring a colonised mooring line. The images were obtained from an inspection off the south coast of France. The objective is to reconstruct the 3D shape of the mooring line so that it can be used for tasks such as establishing better estimates of the hydrodynamic loads acting on the mooring line, as well as improving understanding of processes like seasonal build-up/decline, spatial variability and rate of advance of marine growth. With reference to Figure 13(a), the first step is to acquire the stereo images. For the sake of simplicity, this demonstration only considers one stereo pair, as shown in Figure 13(b & c). If the goal is to create a full 3D model, multiple stereo pairs from all sides of the mooring line must be captured and integrated into an SFM framework.

Camera calibration is needed to ensure a Euclidean reconstruction, i.e. a reconstruction where the 3D model is appropriately scaled and angles remain true to reality. In this example, we use self-calibration, or auto-calibration, which is the process of obtaining intrinsic camera parameters (such as the focal length) and extrinsic parameters (the rotation and translation transformations between stereo cameras) using only the constraints in the scene (Zhang et al., 1994). Rectification follows on from calibration, whereby the left and right images are transformed so that corresponding points in both images are separated only by a horizontal offset and not by a vertical offset. Next, the proposed stereo matching algorithm, PaLPaBEL, is applied to the rectified stereo pair. The output from this algorithm is a disparity map and an occlusion map. The occlusion map identifies points where no reliable matches exist, which occurs mostly in background regions or for points on the mooring line that are visible in only one of the cameras. The disparity map for non-occluded points is presented in Figure 13(d). With knowledge of the disparity values and the camera parameters, triangulation can be performed which produces a 3D point cloud, as shown in Figure 13 (e). A common subsequent stage, although not discussed here, involves surface reconstruction which seeks to upgrade the 3D point cloud to a watertight mesh. This is the expected input for

many CAD or Computational Fluid Dynamic (CDF) packages where further analysis tasks can be performed.

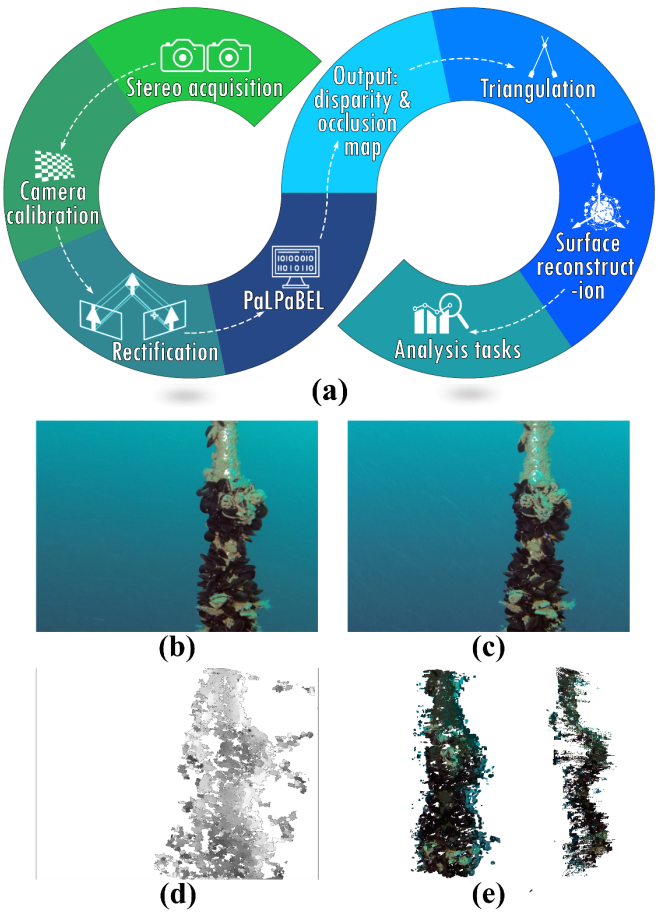


Figure 13 (a) Pipeline for reconstructing 3D shape from a stereo image pair using PaLPaBEL, (b) Left stereo image, (c) Right stereo image, (d) Disparity map for non-occluded points, and (e) Front view and side view of 3D point cloud.

4.4 Discussion

In the case of the synthetic stereo images, it is apparent from the disparity maps in Figure 9 that PaLPaBEL produces a smoother disparity map compared to the basic block matching and semi-global matching techniques, both of which were heavily affected by noisy estimates. This smoothness, however, comes at the expense of being able to resolve very fine details. Nevertheless, analysis of the NRMSE scores in Table 1 reveals that PalPaBEL is superior.

In the case of the laboratory specimens, the performances of the techniques vary markedly under different visibility conditions, but also, there is a marked variation in terms of the performance when considering different specimens under the same visibility conditions. For instance, specimens such as the rough concrete cube (specimen no. 1) and the aged metallic case (specimen no. 8), facilitate more reliable stereo matching due to the richly textured nature of their surfaces This is in contrast with specimens that have more uniform and homogeneous surfaces, where the lack of unique features hinder performance of stereo algorithms.

It may be observed from the disparity maps in Figure 11, and from the NRMSE scores in Figure 12, that PaLPaBEL produces the best results in most cases. Given that a key feature of this technique is that it explicitly accounts for smoothness between neighbouring pixels, a reasonable interpretation would be that algorithms which also model smoothness fare better when applied to underwater images. The usefulness of a smoothness term appears to be even more profound when operating at higher turbidity levels at which point some of the surface details become washed out. However, in the worst conditions (i.e. low light and high turbidity), the performances from all techniques are poor. This indicates that the operating limit has been reached.

Further analysis of the results indicate that the choice of similarity measure is a key factor. The SAD similarity measure, which is used for the semi-global technique, is not as effective as the ZNCC score in conjunction with census transform-based matching, as employed by PaLPaBEL. Both the ZNCC score and census transform based matching are widely regarded as being robust similarity measures, especially in harsh lighting conditions (Lin et al., 2017).

For the real-world example of the colonised mooring line, it is not possible to verify the accuracy of the results due to the uncontrolled nature of the exercise; however, the general shape that emerges from the 3D point cloud appears fit reasonably and to be in line with our expectations. Overall, the results underline the value of PaLPaBEL as an effective and technique that is well-suited for underwater application.

5 CONCLUSION

Recovering quantitative shape information is a challenging but useful task, and has wide applicability in many areas of Structural Health Monitoring (SHM). While stereo systems have the potential to be a convenient underwater tool capable of obtaining in-situ measurements, they must operate in challenging visibility conditions characterised by non-uniform lighting, limited range visibility, low contrast, blurring, light attenuation and colour absorption. The development of a new stereo matching algorithm is motivated by the need for a technique that can effectively deal with these issues and meet the needs of modern inspection services, where there is a demand for techniques capable of extracting as much information as possible from high-resolution images in a reasonable timeframe. In this vein, a new stereo correspondence algorithm, called PaLPaBEL, has been described in this paper. PaLPaBEL is based on a pyramidal loopy BP message passing algorithm applied on an MRF. Adopting a coarse-to-fine pyramidal scheme enables MRFs to be used for high-resolution images in a computationally efficient manner. By using the original image resolution, more accurate and precise depth information can be extracted. Pyramiding also has an inherent smoothing, which is useful in the noisy

underwater conditions and when specular reflections are encountered. Furthermore, PaLPaBEL utilises an anisotropic diffusion technique for sensible smoothing that minimises noise while preserving real features such as edges.

PaLPaBEL has been designed to work well in challenging lighting conditions. Central to this is the choice of similarity measures: a ZNCC score and the Hamming distance using census transformed data. These similarity measures are highly resilient to illumination variations.

PaLPaBEL is validated on a host of real-world underwater stereo images that were captured under various visibility conditions, as well as on virtual images of underwater scenes where the ground-truth depth information is precisely known. The results show that PaLPaBEL can handle reduced visibility conditions quite well as it demonstrates improved performance over other considered methods.

ACKNOWLEDGEMENTS

The authors acknowledge CAPACITES/IXEAD society and Marine Research Energy Ireland (MaREI), grant no. 12/RC/2302, a Science Foundation Ireland (SFI) project.

REFERENCES

Alagoz, B. B. (2008). Obtaining depth maps from color images by region based stereo matching algorithms. *OncuBilim Algorithm and Systems Labs,* **8**, 1-13.

Bai, M., Zhuang, Y., & Wang, W. (2009). Hierarchical adaptive stereo matching algorithm for obstacle detection with dynamic programming. *Journal of Control Theory and Applications,* **7**(1), 41-47.

Bianco, G., Gallo, A., Bruno, F., & Muzzupappa, M. (2013). A comparative analysis between active and passive techniques for underwater 3D reconstruction of close-range objects. *Sensors*, **13**(8), 11007–11031.

Birchfield, S., & Tomasi, C. (1998). Pixel dissimilarity measure that is insensitive to image sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **20**(4), 401-406.

Bruno, F., Bianco, G., Muzzupappa, M., Barone, S., & Razionale, A. (2011). Experimentation of structured light and stereo vision for underwater 3D reconstruction. *Journal of Photogrammetry and Remote Sensing,* **66**(4), 508-518.

Busby, F.R. (1979). Underwater inspection/testing/ monitoring of offshore structures. *Ocean Engineering*, **6**, 355-491.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **6**, 679-698.

Coughlan, J. M., & Shen, H. (2010). An embarrassingly simple speed-up of belief propagation with robust potentials. *arXiv preprint arXiv:1010.0012*.

Cyganek, B., & Siebert, J.P. (2009). *An Introduction to 3D Computer Vision Techniques and Algorithms*. John Wiley & Sons, Ltd.

Estes, A., & Frangopol, D. (2003). Updating Bridge Reliability Based on Bridge Management Systems Visual Inspection Results. *Bridge Engineering,* **8**(6), 374-382.

Furukawa, Y., & Ponce, J. (2010). Accurate, Dense, and Robust Multiview Stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(8), 1362-1376.

Gallwey, T. J., and Drury, C. G. (1986). Task Complexity in Visual Inspection. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, **28**(5), 595-606.

Gaurier B., Germain G., Facq J.V., Baudet L., Birades M., & Schoefs F. (2014). *Effect of marine concretions on the hydrodynamic behavior of circular structures*. 14th Hydrodynamics Days, Val de Reuil, France, 18-20 November 2014.

Giachetti, A. (2000). Matching techniques to compute image motion. *Image and Vision Computing,* **18**(3), 247-260.

Hartley, R. (1999). Theory and Practice of Projective Rectification. *International Journal of Computer Vision,* **35**(2), 115-127.

Hartley, R., & Zisserman, A. (2003). *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Ho, Y.-S., & Jang, W.-S. (2013). Occlusion Detection Using Warping and Cross-Checking Constraints for Stereo Matching, *Springer*, (pp. 363-372).

Kanade, T., & Okutomi, M. (1994). A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **16**(9), 920-932.

Komorowski, J. P., & Forsyth, D. S. (2000). Role of enhanced visual inspections in the new strategy for corrosion management. *Aircraft Engineering and Aerospace Technology,* **72**(1), 5-13.

Lee, J., Jun, D., Eem, C., & Hong, H. (2016). Improved census transform for noise robust stereo matching. *Optical Engineering,* **55**(6), 1-10.

Li, G., & Zucker, S. W. (2006). *Surface geometric constraints for stereo in belief propagation.* IEEE Computer Society Conference Computer Vision and Pattern Recognition, 17-22 June, 2006, New York, United States.

Liarokapis, F., Kouřil, P., Agrafiotis, P., Demesticha, S., Chmelík, J., & Skarlatos, D. (2017). 3D Modelling and Mapping for Virtual Exploration of Underwater Archaeology Assets, *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **42**(2), 2017, 425-431.

Lin, C., Li, Y., Xu, G., & Cao, Y. (2017). Optimizing ZNCC calculation in binocular stereo matching. *Signal Processing: Image Communication,* **52**, 64-73.

Massot-Campos, M., & Oliver-Codina, G. (2015). Optical Sensors and Methods for Underwater 3D Reconstruction. *Sensors, 15*(12), 29864.

Nguyen, T. M., & Wu, Q. M. J. (2013). A fuzzy logic model based Markov random field for medical image segmentation. *Evolving Systems,* **4**(3), 171-181.

Nishikawa, T., Yoshida, J., Sugiyama, T., & Fujino, Y. (2012). Concrete crack detection by multiple sequential image filtering. *Computer-Aided Civil and Infrastructure Engineering,* **27**(1), 29-47.

O'Byrne, M., Schoefs, F., Pakrashi, V., & Ghosh, B. (2014b). Regionally enhanced multi-phase segmentation technique for damaged surfaces. *Computer-Aided Civil and Infrastructure Engineering*, **29**(9), 644-658.

O'Byrne, M., Schoefs, F., Pakrashi, V., & Ghosh, B. (2017). An underwater lighting and turbidity image repository for analysing the performance of image based non-destructive techniques. *Structure and Infrastructure Engineering*. 1-20

Olofsson, A. (2010). *Modern stereo correspondence algorithms: investigation and evaluation.* Linköping.

Papadakis, N., & Caselles, V. (2010). Multi-label depth estimation for graph cuts stereo problems. *Journal of Mathematical Imaging and Vision,* **38**(1), 70-82.

Patwardhan, K. A., Sapiro, G., & Bertalmio, M. (2005). *Video inpainting of occluding and occluded objects.* IEEE International Conference on Image Processing. ICIP 2005.

Sagara, S., Ambar, R. B., & Takemura, F. (2013). A stereo vision system for underwater vehicle-manipulator systems - Proposal of a novel concept using Pan-Tilt-Slide cameras. *Journal of Robotics and Mechatronics,* **25**(5), 785-794.

Sappa, A. D., Dornaika, F., Ponsa, D., Geronimo, D., & Lopez, A. (2008). An efficient approach to onboard stereo vision system pose estimation. *IEEE Transactions on Intelligent Transportation Systems,* **9**(3), 476-490.

Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision,* **47**(1-3), 7-42.

Sun, J., Shum, H.-Y., & Zheng, N.-N. (2002). Stereo matching using belief propagation *Computer Vision—ECCV 2002* (pp. 510-524): Springer Berlin Heidelberg.

Sun, J., Zheng, N.-N., & Shum, H.-Y. (2003). Stereo matching using belief propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **25**(7), 787-800.

Sun, J., Zhu, H., Xu, Z., & Han, C. (2013). Poisson image fusion based on Markov random field fusion model. *Information Fusion,* **14**(3), 241-254.

Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., et al. (2006). A Comparative Study of Energy Minimization Methods for Markov Random Fields. In A. Leonardis, H. Bischof, & A. Pinz (Eds.), *Computer Vision – ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part II* (pp. 16-29). Berlin, Heidelberg.

Tavera-Vaca, C. A., Almanza-Ojeda, D. L., & Ibarra-Manzano, M. A. (2015). Analysis of the efficiency of the census transform algorithm implemented on FPGA. *Microprocessors and Microsystems,* **39**(7), 494-503.

Wang, K. C. P., Hou, Z., & Gong, W. (2010). Automated road sign inventory system based on stereo vision and tracking. *Computer-Aided Civil and Infrastructure Engineering,* **25**(6), 468-477.

Williams, S. B., Pizarro, O., Steinberg, D. M., Friedman, A., & Bryson, M. (2016). Reflections on a decade of autonomous underwater vehicles operations for marine survey at the Australian Centre for Field Robotics. *Annual Reviews in Control,* **42**, 158-165.