A Review on Stereo Vision Algorithms: Challenges and Solutions

Kai Yit Kok [1] and **Parvathy Rajendran** [2]

ABSTRACT

This paper presents a survey on existing stereo vision algorithms. Generally, stereo vision algorithms play an important role in depth estimation. The quality of disparity map as well as computational load vary based on different approaches of stereo vision algorithms. The existing stereo vision algorithms are discussed in terms of concept, performance and related improvements. Also, a brief analysis of a performance comparison among existing stereo vision algorithms is presented. Moreover, available improvements and solutions for stereo vision challenges, such as computational complexity, occlusion, radiometric distortion, depth discontinuity and textureless region are reviewed.

Keywords: Stereo Vision Algorithm, Obstacle Detection, Object Tracking, Shape Reconstruction, Geometric Mapping

1. INTRODUCTION

In the last two decades, stereo vision development has received a lot of attention from researchers due to its huge potential in the fields of robotics, automobiles and aerospace. The applications of stereo vision include obstacle detection, 3D object recognition [1], object tracking [2], 3D shape reconstruction [3], and 3D geometric mapping [4].

Generally, stereo vision is a process whereby pixel information of two images is extracted to estimate a disparity map. A binocular camera (also known as a stereo camera), in which two cameras are well aligned horizontally, is required to operate stereo vision. Using this camera, we are able to capture two images (left image and right image) simultaneously and a disparity map is obtained based on the image sets [5].

The quality of generated disparity maps and computation load vary with different types of stereo vision algorithms. Furthermore, image noise and other uncertainties on stereo images will affect the accuracy of the disparity maps directly. Hence, new approaches and modified versions of stereo vision algorithms are developed and updated frequently to refine accuracy and reduce computational cost as much as possible.

Performance evaluation and comparison are essential to differentiate the pros and cons of various stereo vision algorithms. This information is crucial in order to allow developers to choose a suitable stereo vision algorithm according to application requirements. Scharstein, et al. [6] have presented an excellent taxonomy for dense stereo vision algorithms.

In order to evaluate performance of stereo vision algorithms impartially, they have developed an online platform for researchers to use to analyze performance of proposed algorithms using standard benchmark datasets. Researchers are able to compare results with others easily with a ranking table. There are other review papers listed in Table 1 which discuss different aspects of stereo vision including software based and hardware based techniques.

In this paper, a review of the state-of-the-art developments in stereo vision algorithms is presented. Also, the challenges and solutions of stereo vision algorithms, with available solutions, are discussed. In Section 2, a taxonomy for every processing stage of stereo vision disparity map estimation is presented. Then, existing stereo vision algorithms are reviewed in Section 3. Next, Section 4 explains major challenges of stereo vision algorithms, together with appropriate solutions. Lastly, the conclusion is presented in Section 5.

2. A TAXONOMY FOR THE PROCESS FLOW OF DISPARITY MAP ESTIMATION

A stereo vision algorithm mostly consists of four stages: 1) Matching cost computation, 2) Cost aggregation, 3) Disparity optimization and selection, and 4) Disparity refinement as described by Scharstein, et al. [6]. Every stage is necessary to ensure accurate disparity values can be obtained. This is done by utilizing all available information on the images and filtering unfavorable elements that can degrade the quality of the disparity map.

2.1 Matching Cost Computation

Using epipolar geometry, demonstrated in Fig. 1, we can determine the distance to a target scene from a binocular camera accurately based on the position of the target on the left and right images respectively, if the camera is well calibrated. Matching cost compu-

Manuscript received on June 12, 2019 ; revised on July 31, 2019.

Final manuscript received on August 14, 2019.

[1,2]The authors are with School of Aerospace Engineering, Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Malaysia., E-mail: kok901221kaiyit@student.usm.my and aeparvathy@usm.my

tation is the process used to measure the cost function between two pixels from the left and right images in order to estimate the distance of the target from the camera ultimately.

This is a fundamental process in all stereo vision algorithms used to obtain a disparity map from the two images. In general, a lower cost value means the two pixels are more likely to be correct match. In addition, the epipolar constraint makes the disparity estimation simpler and easier, since target P is located at P_l in the left image I_l and at P_r on the right image I_r and they individually are on the same horizontal line. This translates the searching space from two dimensional into a one dimensional problem. Normally, the cost function is obtained by comparing pixel intensities between two points from both images. Similar pixels with near zero cost are assumed to be identical regions with the same depth.

Table 1: *Demographic and Converted Data.*

Author	Description
Scharstein, et al. [6]	Presented a taxonomy for stereo vision algorithms in order to enable quality comparison. Also, they have proposed a test bed for performance evaluation of stereo vision algorithms in which the datasets and results are available online.
Brown, et al. [7]	Reviewed stereo vision algorithms with real time implementations and solutions for tackling occlusion.
Gong, et al. [8]	Investigated different cost aggregation methods on real time stereo platform, and evaluation of quality and computational cost are conducted with benchmark datasets available online.
Nalpantidis, et al. [9]	Presented a survey on stereo matching methods and usage of advanced intelligence techniques in development of stereo vision algorithms. Evaluation of various image sizes and datasets are demonstrated. Development of stereo vision algorithms via hardware is discussed.
Tombari, et al. [10]	Reviewed performance with a comparison of existing cost aggregation approaches in terms of accuracy and computational cost using standard datasets available online.
Lazaros, et al. [11]	Reviewed performance of up to date existing stereo vision algorithms including global and local approaches. Implementation of stereo vision algorithms in hardware are discussed.
Fang, et al. [12]	Analyzed various types of cost aggregators and performance of local stereo vision algorithms using OpenCL and executed on GPUs.
Kumari, et al. [13]	Presented a taxonomy of recent stereo vision algorithms and provided a parameterized comparison of these techniques.
Hamzah, et al. [14]	Presented a survey on general concepts of stereo vision and various types of existing stereo vision algorithms. They have also reviewed development and implementation of stereo vision using software and hardware with accuracy measurements.

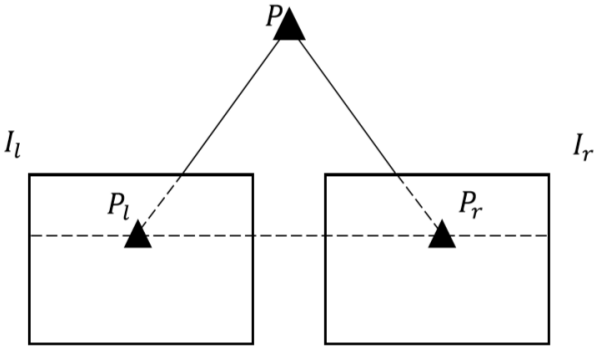

Fig.1: *Epipolar geometry of target P on left and right images.*

2.2 Cost Aggregation

Still, a cost function calculated from two pixels might not have enough information to differentiate the correct match. This in turn may lead to high rates of incorrect matching especially in a local approach. Therefore, more information should be considered around the interest point of the pixel to increase the uniqueness of the particular point through a cost aggregation approach. In this process, cost calculation is done over a predefined region around the interest point and the matching cost will be summed up as the final cost of the interest point. Through this, every pixel of the image becomes more distinct and the correct matching rate is improved significantly.

There are many ways of doing cost aggregation. The simplest way is to use a fixed size window with square shape. The main advantages of this method are easy implementation and low computational cost. The larger the size of predefined window, the more distinctive the calculated cost will be. Anyway, matching at homogenous or repetitive areas using a fixed size window is ambiguous [15].

Moreover, the number of matching errors will increase when the window size is over a certain threshold. This is due to the fact that correct matches of two points from both images mostly are not totally the same in terms of pixel intensity distribution, especially near the depth discontinuity, because of different angles of view. Increasing the window size will magnify the errors [16]. Also, large predefined windows tend to blur boundaries or discontinuity regions [17]. Furthermore, the performance of fixed-size windows varies when different sizes of images or different datasets are used.

In order to minimize the errors, various methods of cost aggregation have been developed, such as robustly adjusting the center pixel of windows in the shiftable window approach [18] or forming multiple smaller windows in the multiple windows approach [19] which intends to avoid occluded regions in the discontinuity area. Finding an appropriate shape and

size for the window is difficult. The rectangularly constrained shape of the support window makes it hard to estimate accurate disparity values for the pixels near the depth discontinuity. Veksler [20] has introduced a variable window approach to allow the size of the support window to change according to surrounding features of the interest point. This is because larger sizes of support windows can provide satisfactory results on textureless regions, while small support windows are good for dealing with discontinuity areas [21].

A better solution for this is using an adaptive weight window [22], whereby the weight parameter is assigned according to color similarity. The influence of depth discontinuity errors can be decreased near outliers of a window as the weight is very small when it is far from the central pixel. Analysis by Yoon, et al. [22] has shown that the algorithm has lower pixel errors in general compared to shiftable windows and variable windows, especially for the pixels near depth discontinuity. However, it has high complexity because of the non-linearity characteristic in computation of the weight parameter. In addition, image noise will severely affect the performance of this approach since support weight computation is measured based on individual pixel colour.

Pham, et al. [23] have made an improvement in terms of computational cost reduction by implementing domain transformation to change the 2D operation into a 1D operation. The integration of the domain transformation technique into cost aggregation to compute cost using 1D operations not only reduces memory cost, but also eliminates the impact of input parameters on computational cost. The computational load of this method is at least 20 times faster than fast cost volume filtering [24], and is 40 times faster than adaptive support weight [25] and geodesic diffusion [26]. These methods are inspired by well-known edge-aware filters such as bilateral filter and guided filter.

Einecke, et al. [27] introduced a different multi-window scheme to take into consideration four types of window: a horizontal window, a vertical window, a small square window and a large square window. Every window plays a different role in defining the disparity map and combining them can generate a disparity map with satisfactory accuracy while having low computational cost using a local approach. For instance, the horizontal window is used to estimate horizontal structures accurately, such as ground surface. The vertical window is used for capturing vertical structures such as sign poles. The large square window can minimize the streaking artifacts caused by horizontal and vertical windows. The small square window is used to lower the fattening effect on depth discontinuity regions. Under the same cost computation, this approach can produce a disparity map with lower cost compared to the standard square window as indicated in their analysis using KITTI benchmark datasets [27].

These approaches have a common limitation that the support window is constrained as a square or rectangular shape and it is hard to remove matching errors at discontinuity regions thoroughly. Cross-based windows are another variant of cost aggregation presented by Zhang, et al. [28] whereby the support window is a cross shape and the four arms lengths vary based on color similarity and connectivity constraints to yield an appropriate window shape close to the object structures in the image. Through this, disparity estimation at object boundaries can be done more accurately. Based on Zhang, et al. [28], this method can achieve lower error than variable windows and shiftable windows using the Middlebury benchmark datasets but, unfortunately, it is unable to compete with adaptive weight windows as indicated in the analysis.

In addition, a segmentation based window [29] is similar to a cross-based window in some ways. The disparity value in the same segment will be considered as a uniform disparity value, assuming they are lying at same object structure. It has similar accuracy to a variable window when analyzed using Middlebury the datasets. The concerns of using this method is that errors might arise at regions with slight differences such as slanted surfaces in the same segment.

Wang, et al. [16] focus on cost aggregation analysis. They found that an adaptive weight window is able to produce a disparity map with a smaller error than a fixed window or shiftable window, but it has larger computational cost. Moreover, Tombari, et al. [10] have done performance evaluation of existing cost aggregation methods using Middlebury datasets. They found that segment based windows and adaptive weight windows have the highest accuracy, but they take more than 10 minutes to compute. Computation with a fixed window takes less than 1 second using same platform. Besides, a variable window has moderate cost aggregation among existing cost aggregation approaches with considerable accuracy and needs only 25 seconds of computation.

2.3 Disparity Optimization and Selection

The strategy of choosing a final disparity value at a certain pixel is different when using the local or global approaches. Generally, the local approach is using a Winner Takes All (WTA) strategy (see Eq. (1)) in selecting the final disparity value for each of the pixels, whereby γ is the disparity range and $C(p,d)$ is the cost function of pixel p.

$$d_p = \arg\min_{d \in \gamma}(C(p,d)) \tag{1}$$

This process is easy and simple. For instance, after obtaining all cost values within the disparity range, the one with minimum value will be chosen as the fi-

nal disparity value of the particular pixel. Nevertheless, since only a local window with limited information is considered to make the final decision, a local approach is very sensitive to image noise, occlusion and blur areas. These 'bad' regions will obviously affect the cost value and cause the correct match not to have the correct minimum cost value within the disparity range. Thus, a denoising filter is often used to improve image quality in order to minimize the errors.

On the other hand, the global approach is more intelligent in making decisions on the final disparity value by having the assumption that similar regions except object boundaries should have a uniform disparity distribution. Hence, a smoothness parameter is included in the disparity estimation as part of the basic matching cost. This assumption makes the global approach have fewer errors caused by discontinuity, occlusion and textureless regions. It normally generates disparity maps with higher accuracy and precision than local approaches at the expense of computational cost. More details of this will be discussed in the global approach section.

2.4 Disparity Refinement

Disparity refinement is a 'post-processing' stage to remove uncertainties and noise from a disparity map as well as optimizing the disparity map to have a higher accuracy level. For instance, occluded regions exist even in the image set that has the best quality without any noise. This region can be detected using algorithms such as left-right consistency check [30], bimodality [31], match goodness jumps [32], and occlusion constraint [33].

We are unable to get correct matches in this region as the occluded region only can be found in one side of the image sets. This is done through marking particular pixels of left images as irrelevant pixels when the intensity value differs from the corresponding pixel of the right image by a certain threshold value [23]. Therefore, occlusion filling is essential to approximate the disparity level in this area using adjacent areas as reference.

Regularization is another common refinement step to smoothen the disparity map by eliminating and filtering image noise in the map. The Median filter is one of the common refinement methods in stereo vision development to remove image noise [23]. It is a well-known filter that is able to eliminate salt-and-pepper noise from images. In stereo vision, there are always textureless regions and false matches will occur easily due to uniqueness of pixels within the area which are not significant.

The Median filter plays the role of identifying these false matches and adjusts the disparity value according to the majority of neighboring pixels with a low computation cost [34]. Besides, sometimes desirable accuracy disparity estimation cannot be determined for certain pixels. Interpolation can be used to measure the disparity value based on local information in the adjacent neighborhood in order to obtain a disparity value with higher accuracy.

3. STEREO VISION ALGORITHM

Various stereo vision algorithms have been developed. The vast majority of them have three stages. They can be divided into two categories: 1) local approach, and 2) global approach. Generally, the local approach is efficient in terms of computational cost while the global approach has higher accuracy. The following section will discuss several common algorithms using the local and global approaches.

3.1 Local Approach

In this approach, for each pixel a disparity estimation is calculated by comparing the cost function within a predefined local window or area. This reduces the computational complexity drastically, compared to considering the whole image, when calculating the disparity for every pixel. Thus, it is also known as a window-based or area-based approach.

Basically, the computational cost is low and it is easier to implement in real time applications than the global approach. However, the errors of disparity maps generated using a local approach are significant, especially in occlusion areas, textureless regions, and at any discontinuity boundary, since it only considers limited information for measuring disparity depth. Normally, local approaches can be categorized as methods with parametric or non-parametric based costs.

3.1.1 Parametric Based Costs

Absolute Intensity Differences (AD) (see Eq. (2)) and Squared Intensity Differences (SD) (see Eq. (3)) [35] are the simplest cost matching functions. They compare the intensity difference of left and right images directly without any complicated calculations.

$$C_{AD}(x,y,d) = \frac{1}{n}\sum_{i=1,n} |I_l(x,y)_i - I_r(x-d,y)_i| \quad (2)$$

$$C_{SD}(x,y,d) = \frac{1}{n}\sum_{i=1,n} (I_l(x,y)_i - I_r(x-d,y)_i)^2 \quad (3)$$

Assuming the left image I_l to be the reference image, the right image I_r is the target image and d is disparity value, pixels with coordinates (x, y) from the reference are subtracted from pixels with coordinates $(x - d, y)$ from the target to obtain matching cost for disparity estimation.

The AD algorithm has been used in real time stereo vision due to its simplicity [36]. Still, the performance of the AD algorithm is notable only when

image resolution is small enough or in regions with simple textures. This is also true for the SD algorithm. The SD algorithm will produce larger errors since the intensity difference is magnified, especially in regions with light and noise.

The Sum of Absolute differences (SAD) (see Eq. (4)) and the Sum of Squared Differences (SSD) (see Eq. (5)) [37] can be considered as enhanced versions of the AD algorithm and the SD algorithm respectively by involving surrounding pixels around each interest pixel in cost calculation.

$$C_{SAD}(x,y,d) = \frac{1}{n} \sum_{(a,b)\in N(x,y)} C_{AD}(a,b,d) \tag{4}$$

$$C_{SSD}(x,y,d) = \frac{1}{n} \sum_{(a,b)\in N(x,y)} C_{SD}(a,b,d) \tag{5}$$

Generally, they total up the intensity differences over a predefined support window $N(x,y)$ and compare the cost value with another image. The SAD algorithm is one of the common methods in stereo vision in this decade as it is easy implement for real time applications. However, both the SAD and SSD algorithms have poor performance in the presence of radiometric distortion.

In 2010, Georgoulas, et al. [2] implemented SAD in their proposed algorithm for real time stereo vision applications because of the low computational cost and simplicity characteristics. In 2013, it was improved to develop a modified version with better performance [38]. Posugade, et al. [39] have tested the performance of the SAD method using a FPGA for real time application. In 2018, it was still being studied by researchers to utilize RGB differences to increase the accuracy of the disparity map [40].

Unlike the SAD algorithm, the SSD algorithm was more popular in previous decades [41, 42]. There is little research on SSD algorithm development in stereo vision recently. However, performance comparison of the SAD and SSD algorithms has been done [37] and the results have shown that the SAD algorithm with a large support window will produce better quality disparity maps than the SSD algorithm.

Similar to the SAD and SSD algorithms, Normalized Cross Correlation (NCC) (see Eq. (6) & (7)) uses predefined windows around interest points when determining matching cost. It is favorable because of robustness in the face of contrast changes and intensity offsets. The NCC algorithm compensates differences of gain and bias and is better at dealing with Gaussian noise statistically. Still, the calculation in Eq. (7) tends to increase blurring in regions of discontinuity more than other existing algorithms [43]. Also, the computational cost of the NCC algorithm is higher than other window-based algorithms, especially in matching large scale images. That makes it unsuitable to be a practical solution [44].

$$C_{NCC}(x,y,d) = 1 - NCC(x,y,d) \tag{6}$$

$$NCC(x,y,d) = \frac{\sum_{a,b\in N_{(x,y)}} I_l(a,b) I_r(a-d,b)}{\sqrt{\sum_{a,b\in N_{(x,y)}} I_l^2(a,b) \sum_{a,b\in N_{(x,y)}} I_r^2(a-d,b)}} \tag{7}$$

However, the SAD, SSD and NCC algorithms depend on similarity of pixel values in pairs of left and corresponding right images to produce high quality disparity maps. Hence, they are sensitive to radiometric distortion, whereby overall pixel values of the left and right images differ by a certain offset or gain factor [45]. In order to solve this issue, zero mean based algorithms were developed, including Zero-mean Sum of Absolute Differences (ZSAD) (see Eq. (8) & (9)), Zero-mean Sum of Squared Differences (ZSSD) (see Eq. (10) & (11)), and Zero-mean Normalized Cross Correlation (ZNCC) (see Eq. (12) & (13)). The main drawback of zero-mean based algorithms is high computational cost due to additional calculations (see Eq. (14)), in which the pixel values are subtracted with the mean value of the support window $N(x,y)$ [46].

$$C_{ZSAD}(x,y,d) = \sum_{(a,b)\in N(x,y)} ZSAD(a,b,d) \tag{8}$$

$$ZSAD(a,b,d) = \frac{1}{n} \sum_{\substack{i=1,n \\ (a,b)\in N(x,y)}} |ZV(I_l,a,b)_i - ZV(I_r,a,b)_i| \tag{9}$$

$$C_{ZSSD}(x,y,d) = \sum_{(a,b)\in N(x,y)} ZSSD(a,b,d) \tag{10}$$

$$ZSSD(a,b,d) = \frac{1}{n} \sum_{\substack{i=1,n \\ (a,b)\in N(x,y)}} (ZV(I_l,a,b)_i - ZV(I_r,a,b)_i)^2 \tag{11}$$

$$C_{ZNCC}(x,y,d) = 1 - ZNCC(x,y,d) \tag{12}$$

$$ZNCC(x,y,d) = \frac{\sum_{(a,b)\in N_{(x,y)}} ZV(I_l,a,b) ZV(I_r,a-d,b)}{\sqrt{\sum_{(a,b)\in N_{(x,y)}} (ZV(I_l,a,b))^2 \sum_{(a,b)\in N_{(x,y)}} (ZV(I_r,a-d,b))^2}} \tag{13}$$

$$ZV(I,x,y) = I(x,y) - \bar{I}_{N(x,y)}(x,y) \quad (14)$$

3.1.2 Non-Parametric Based Costs

The fundamental idea of non-parametric based costs is similar to zero-mean based algorithms, both of which are designed to deal with changes in image gain and bias. The typical algorithms are Rank transform (RT) (see Eq. (15)-(17)) and Census Transform (CT) (see Eq. (18) & (19)) developed by Zabih, et al. [47]. The main difference of non-parametric based methods compared with parametric based methods is that the matching cost is not related to the intensity value at all, but is an integer in the range$(0,\ldots,x*y-1)$.

$$C_{RT}(x,y,d) = \sum_{(a,b)\in N(x,y)} |Rank_l(a,b) - Rank_r(a-b,d)| \quad (15)$$

$$Rank(a,b) = \sum_{(a,b)\in N(x,y)} L(a,b) \quad (16)$$

$$L(a,b) = \begin{cases} 1: & I(a,b) < I(x_c,y_c) \\ 0: & otherwise \end{cases} \quad (17)$$

$$C_{CT}(x,y,d) = \sum_{(a,b)\in N(x,y)} Hamming(CT_l(a,b) - CT_r(a-d,b)) \quad (18)$$

$$CT(a,b)) = Bitstring_{(a,b)\in N(x,y)}(I(a,b) < I(x_C,y_C)) \quad (19)$$

In terms of the RT algorithm, the cost function is defined as number of pixels in a predefined support window where the intensity value is less than the center pixel (x_C, y_C) [48]. The performance evaluation [47] has shown that the RT algorithm is better than parametric based algorithms such as SAD and SSD. It has been used in real time stereo vision with implementation in a FPGA to generate disparity maps with an image size of 640×480 pixels under 30 frames per second [48]. Also, with the help of a GPU, RT can achieve up to 36 frames per second with the same image size [49].

The CT algorithm, on the other hand, is identical to the RT algorithm except for only two differences. First, the structure of calculation in the CT algorithm is in the form of a bit string. Next, the integers of neighborhood pixels after comparing with the center pixel are stored, and the similarity with the next image is identified using Hamming distance [50]. That is unlike the RT algorithm where the integers of neighborhood pixels are summed before computing matching cost between two images. This makes the CT algorithm more accurate in matching cost compared to the RT algorithm [47], as well as other algorithms [51, 52]. However, the computational cost of the CT algorithm cannot be ignored when implemented in general CPUs. It is more suitable for use on FPGAs or ASICs which have faster performance [53, 54].

Moreover, accessed size and memory of FPGAs can be further decreased to allow performance of the CT algorithm to achieve up to 130 frames per second with an image size of 640×480 pixels [55]. Nevertheless, there is a study where the CT algorithm performance was able to reach up to 42 frames per second using CPUs [56]. In Hernandez-Juarez, et al. [57], an extension of the CT algorithm is selected and implemented in an embedded GPU system with up to 42 frames per second. Besides, the modified CT algorithm is developed to minimize noise. Experimental results have shown the robustness of the proposed algorithm against noise using Middlebury datasets [58].

3.2 Global Approach

Unlike the local approach, the global approach treats disparity estimation as a labeling problem with an energy minimization framework [59]. It considers the influence of all pixels to eliminate the impact of local areas which can cause errors because of homogenous regions or occlusion. Thus, the computational cost using global methods is usually larger than that of local methods, but they yield higher quality disparity maps. The common methods of the global approach are Dynamic Programming, Graph Cuts, and Belief Propagation.

Generally, this energy function consists of two terms, a data term and a smoothness term (see Eq. (20)). The data term estimates the consistency between pixels of both images. The smoothness term determines the similarity of pixels in the neighborhood. This function can be solved using either horizontal optimization (1D) or both sides optimization (2D). The typical 1D optimization is using classic DP optimization [18]. It is simple and it requires less time to complete the optimization. Still, the main drawback is that it will lead to a well-known streaking effect due to insufficient coherence with vertical pixels [60].

$$E(f) = E_{data}(f) + E_{smooth}(f) \quad (20)$$

A 2D optimization, on the other hand, can produce smoother and better quality disparity maps since it considers both horizontal and vertical direction scanlines. In this decade, most of the global approach studies are using 2D optimization to ensure the disparity estimation is as close as possible to the ground truth image. The main challenge of this implementation is minimizing computational cost to make it suitable for real time applications.

3.2.1 Dynamic Programming (DP)

DP is a mathematical approach that divides an optimization problem into smaller subproblems to decrease complexity [61]. In this method, each corresponding horizontal scanline from both the left and right images will be computed to form a cost matrix, which is also known as the disparity space image (DSI). The local cost of each pixel in a DSI can be obtained using one of the local approaches. Then, minimum cost accumulation is done using the DSI. This can be achieved by comparing left bottom pixels in the local region and searching for the lowest value of accumulated cost. An optimal path is then estimated with a backward direction from the upper right corner, which is demonstrated in Fig. 2.

A lot of research has been done to improve DP performance, since the inability of computing vertical consistency severely affects the quality of disparity estimation while the computational complexity is also significant. Fan, et al. [63] have simplified DP stereo vision by limiting disparity range and down sampling to reduce computational complexity. Kim, et al. [64] presented a Two-Pass DP which examines horizontal and vertical scanlines for estimating disparity depth.

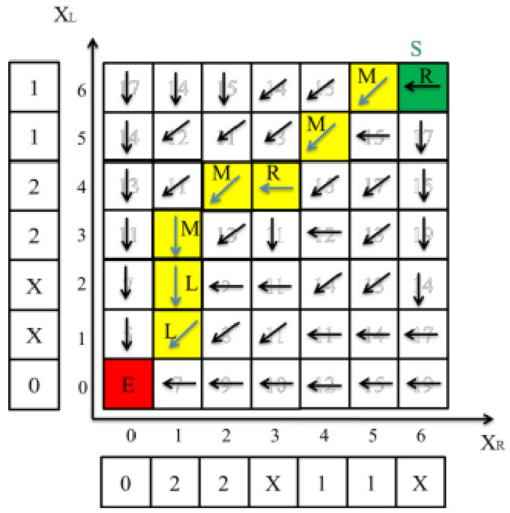

Fig.2: *Back tracking to search for optimal solution [62].*

Moreover, considering both horizontal and vertical consistency using sub-optimal cost has been proposed by Sawires, et al. [65] in order to smoothen and improve the accuracy of disparity maps. Witt, et al. [66] combined edge based matching with DP optimization. In this proposed algorithm, DP is used for optimization along edges instead of scanlines and the results have shown that most of the horizontal edges can be recovered but not the slanted plane textures.

Besides, Baek, et al. [59] have proposed a DP based stereo vision with the ability to detect occlusion and error regions. This is done by modelling an energy function with constraints of uniqueness, ordering, and color similarity. After obtaining occlusion or error regions, these areas are refined, and the disparity is estimated using DP and an edge preserving noise filter. Apart from this, Suhr, et al. [67] proposed a robust stereo vision with a combination of DP and Hough Transform to measure the vertical road profile. The results indicate that the proposed algorithm can estimate the road profile, even if the road is distant from the camera.

3.2.2 Graph Cuts (GC)

A weighted directed graph (See Eq. (21)) consists of two main parameters, a set of vertices V and weighted directed edges E (See Eq. (24)). There are two particular vertices, known as the source s and the sink t (See Eq. (22) & (23)), where the graph axes is according to the image xy axes and disparity range d. In this method, the nodes relate to edges constraints, following by calculating cost using a local method. A cut is a partition of vertices(V^s, V^t), in which the cost is the sum of the weights of the edges from the source to the sink (See Eq. (25)). The cut with minimum cost will be considered as the optimal solution for disparity depth estimation.

$$G = (V, E) \tag{21}$$

$$V = V^* \cup (s, t) \tag{22}$$

$$V^* = [(x, y, d), x \in (0, x_{\max}), y \in (0, y_{\max}), d \in (0, d_{\max})] \tag{23}$$

$$E = \left\{ \begin{array}{c} (u, v) \in V^* x V^* : \|u - v\| = 1 \\ (s, (x, y, 0)) : x \in [0, x_{\max}] \\ ((x, y, d_{\max}), t : y \in [0, y_{\max}] \end{array} \right\} \tag{24}$$

$$C_{GC}(V^s, V^t) = \sum_{\substack{(u,v)\in\varepsilon \\ u\in V^s, v\in V^t}} C_{GC}(u, v) \tag{25}$$

Similar to DP, much research has been carried out to enhance GC performance since minimizing the energy function is expensive in terms of computational cost while the disparity depth calculation is far from mature. For GC, Boykov, et al. [68] presented two techniques, namely expansion moves and swap moves, to search for the local minimum with higher efficiency. Instead of changing the labels of pixels one by one traditionally, these new approaches allow huge numbers of pixel label changes simultaneously. They have tested with simulated annealing and the results demonstrated that the new techniques are able to converge faster with lower values of energy.

Hong, et al. [69] used a segment domain in the energy minimizing problem which is different from the conventional pixel domain. It helps to increase the

accuracy disparity estimation and minimize the errors due to occlusion, textureless regions, and discontinuity. Furthermore, Altantawy, et al. [70] boosted the estimation by introducing the Non-Local-Mean method in the segment domain to reduce image noise while preserving the edge property.

Moreover, the Patchmatch algorithm was integrated with GC by Feng, et al. [71] to reduce computational complexity. Using the Patchmatch method, the label of pixels is randomly selected and propagated to adjacent pixels. Also, a convolutional neural network is applied to GC for similarity measures on image patches in order to refine the precision of the result. Compared to existing BP based algorithms and other GC based approaches, the proposed GC based method outperforms them on the Middlebury datasets.

3.2.3 Belief Propagation (BP)

In this method, the pixels of an image are connected to each other in order to exchange information through message transferring so that the probability distribution of the disparity of the pixels can be determined [72]. The message here is in the form of a probability value and the value is updated every iteration and replaces the previous value. The word 'Belief' in the name of this approach means that the probability value has been replaced. Message transferring from pixel p to pixel q in iteration n with minimum energy can be formulated as in Eq. (26).

$$m^n_{p\to q}(d_q) = \min_{d_p\in\gamma}(V(d_p,d_q) + D_p(d_p) + \sum_{r\in N(p)/q} m^{n-1}_{r\to p}(d_p)) \quad (26)$$

γ is the disparity range, $V(d_p, d_q)$ is a parameter that measures similarity or smoothness of two neighbor pixels, $D_p(d_p)$ is the matching cost associated with the central pixel and $m_(r \to p)(n-1)(d_p)$ is the latest updated message from adjacent pixels [73]. Fig. 3 demonstrates the concept of message passing in BP, in which pixel x_1 obtains a message from adjacent pixels and passes the message to pixel x_2. This process continues to spread to the whole image with iterations until the energy is decreased and converged to an acceptable level to obtain the optimal solution of the disparity map.

In terms of enhancement or modification, Sarkis, et al. [74] proposed measuring sparse depth map by matching a sparse image and a dense image for preserving the quality of depth estimation. Recovering the dense disparity map can be done using the sparse depth map using an approximation method. Liang, et al. [75] improved performance of BP by introducing a stability factor in the energy function, and include other occlusion handling such as a left-right consistency check, as well as a median filter to eliminate noise and refine the disparity map.

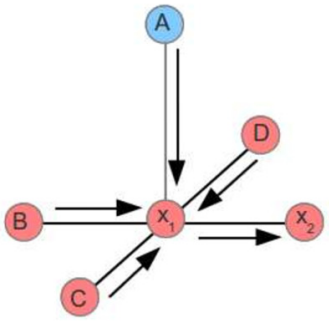

Fig.3: *Illustration of message passing in Belief Propagation [4].*

Semi-limited BP is presented by Luo, et al. [76] whereby passing messages is only allowed within regions of the same object to minimize errors caused by passing irrelevant messages to other regions of an image. Recently, there is study on developing BP to have 1D optimization similar to a classic DP approach [77]. By implementing it on a multicore embedded system, it can run with 1080p resolution images at 24.5 frames per second with only 10-Watt power consumption. Still,
the streaking effect remains even though a median filter is used to reduce the noise.

3.3 Overall Performance Comparison of Stereo Vision Algorithms

Table 2 illustrates a summary of recent stereo vision approaches. Miron, et al. [78] have done performance evaluation of local approaches for algorithms both with and without cost aggregation using datasets from Middlebury and KITTI. Results have shown the ZNCC algorithm without cost aggregation outperforms the others with around 40% errors. The rest have more than 50% errors. This indicates the zero mean feature can estimate more desirable disparity values without the use of cost aggregation. When fixed window cost aggregation is included, the result is further improved until there are only 29% errors on average. The proposed algorithm presented by Miron, et al. [78] combining CT with mean sum of relative differences of intensities inside a support window has a better disparity map with only 22% errors.

In terms of the global approach, Tappen, et al. [79] have compared the performance of GC and BP algorithms under the same conditions and parameters. Overall, both the GC and BP algorithms have nearly the same error percentage regardless of non-occluded error, textureless error, or discontinuity error. The GC algorithm generates a smoother disparity map than the BP algorithm. But this advantage does not help the GC algorithm in obtaining lower error when compared to ground truth data because the latter has higher energy than both algorithms. This

Table 2: *Summary of major stereo vision methods.*

Stereo approach	Modification	Category	Device	Advantages	Limitations & Concerns
SAD [34]	Uses uniqueness of minimum and median filter	Local	CPU	Good accuracy with low computational cost	Not applicable for high texture images
SAD [2]	Uses pyramid reduction, ZNCC similarity measures and vergence angle control	Local	FPGA	Extremely low and consistent computational cost and low power consumption	Poor performance in homogenous regions
SAD [21]	Uses various sizes of correlation window	Local	CPU	Gives good results in non-texture regions as well as depth discontinuity	Performance evaluation under radiometric distortion is not justified
SAD [38]	Only considers object edge pixels in depth estimation calculation	Local	CPU	Faster than conventional SAD algorithm	Performance evaluation under radiometric distortion is not justified
SAD/SSD [37]	Uses nuclear norm minimization	Local	CPU	Able to reduce occlusion error and performs better in homogenous regions	Performance in terms of computational cost is unclear
SAD [40]	Uses bilateral filter	Local	CPU	Good accuracy and is able to reduce noise	Performance in terms of computational cost is unclear
NCC [83]	Modified to be more like template approximation	Local	CPU	Faster than conventional NCC	Quality of disparity map is not verified
NCC [84]	Uses shape adaptive window and orthogonal integral image technique	Local	CPU	Produces accurate disparity map quickly	Not suitable for real time application yet
NCC [44]	Uses dirty filtering	Local	CPU	High precision	High computational cost
ZNCC [85]	Integrated within neural network	Local	CPU	Good in dealing with textureless areas	Ineffective in dealing occlusion and discontinuity
CT [54]	None	Local	ASIC	Low computational cost	Disparity map quality is not justified
CT [53]	None	Local	FPGA	Low computational cost	Post-processing is not included, so disparity quality can be further improved improved
CT [55]	Optimizes size and memory access	Local	FPGA	High speed with reduced memory size	Poor performance when image has noise
CT [86]	Combines with SAD and uses permeability filtering	Local	CPU	Fast operation with promising accuracy	Computational cost is not low enough for real time application
CT [50]	Uses coherency sensitive hashing	Local	CPU	High accuracy with fewer bad pixels	High computational cost
CT [58]	Uses random walk and wavelet edge joint bilateral	Local	CPU	Minimizes noise significantly, even in the toughest situation with addictive Gaussian noise	Parameter selection is crucial to ensure optimum performance
DP [64]	Uses generalized ground control points scheme	Global	CPU	Scanline inconsistency problem is removed	Not applicable for real time application
DP [87]	Uses adaptive aggregation	Global	CPU	Faster and better than conventional DP based approach	Implementation on CPU has high computational cost
DP [88]	Uses reliability search for matches	Global	CPU	Processing speed is faster than conventional DP based approaches	The improvement is mainly on computational cost
DP [60]	Uses cross-based adaptive window aggregation	Global	CPU	More accurate and faster than most DP based approaches	The accuracy of the disparity map decreases as the range increases
DP [62]	Implements memory-based architectures	Global	FPGA	Requires less area cost than array-based architectures	Evaluation in terms of disparity quality is not justified
DP [67]	Combines with Hough transform	Global	CPU	Generates reliable vertical road profile distant from camera	Unable to work properly for severe roll angle and complete occlusion

Table 2: *Summary of major stereo vision methods (continued).*

Stereo approach	Modification	Category	Device	Advantages	Limitations & Concerns
GC [69]	Uses segment domain in minimizing energy	Global	CPU	Good in dealing with textureless areas, occluded regions and disparity discontinuities	Unable to handle high texture image where object boundaries are inside initial color segments
GC [89]	Uses SVD in plane fitting with 3 steps outliers filter and improved hierarchical clustering algorithm	Global	CPU	High accuracy	Significant computational cost
GC [70]	Uses Non-Local Mean technique for cost volume filtering and SVD in plane fitting	Global	CPU	Produces disparity map with satisfactory accuracy and deals with textureless regions well	Performance is affected when there are many or complicated objects in the image
GC [90]	Uses locally shared labels scheme	Global	CPU	High accuracy with great performance in tackling homogenous regions, disparity discontinuity and occluded areas	Computational cost is not stated. High cost is suspected
GC [91]	Uses previous sequence image as reference with noise filter and cross-checking to fill occlusion	Global	CPU	Outperforms conventional method in terms of accuracy	Computational cost analysis is not evaluated
GC [71]	Integrates patchmatch approach and neural network for similarity measure	Global	CPU	Higher accuracy than other patchmatch based or graph cuts-based methods	Computational cost is not analyzed
BP [74]	Uses sparse disparity map to recover dense disparity map	Global	CPU	Faster than conventional BP method and lower MSE than NCC method	Computational cost still much higher than local approaches (e.g. NCC)
BP [72]	Uses differential geometric constraint of disparity based on image segmentation	Global	CPU	Works well for slanted planes and curved surfaces	AdaptingBP is clearly better than this approach in terms of accuracy and processing time is not studied
BP [92]	Uses RGB vector measure into smooth function	Global	CPU	High accuracy and good in dealing with disparity discontinuity	Performance in dealing with homogenous regions is not satisfactory
BP [76]	Uses local method to estimate disparity range and regulates data term of energy function	Global	CPU	Accuracy is better than the original BP method	Computational cost is not analyzed in this study

phenomenon is caused by pixels with correct disparity within an occluded region in ground truth data which tend to have various intensity values. These increase the matching cost as well as the overall energy value.

Despite that, GC algorithm has lower computational cost than the BP algorithm, but the BP algorithm can be enhanced to be faster than the GC algorithm. Anyway, there are faster versions of the GC algorithm [80], thus the GC algorithm has more advantage in terms of computation. On the other hand, the conventional DP algorithm has lower computational load since it does not involve vertical scanlines with expense of higher error than the BP and GC algorithms [81]. Thus, most of the DP algorithm improvements include consideration of vertical consistency to ensure desirable accuracy can be reached.

4. VARIOUS CHALLENGES AND SOLUTIONS OF STEREO VISION ALGORITHM

This section discusses the major challenges of stereo vision algorithms and available solutions to tackle these challenges. A desirable solution with perfect output has not yet been found.

4.1 Radiometric Distortion

Radiometric distortion is a common defect of images captured by binocular cameras due to image noise, vignette effects, and slight setting changes [46]. We are often unable to avoid the presence of radiometric distortion, so stereo vision algorithms should take into consideration minimizing the influence of these defects. For this, Miron, et al. [78] have done a performance analysis of local approaches in the presence of radiometric distortion.

The results have indicated that parametric-based cost functions such as those in the SD algorithm and used by Klaus, et al. [82] (combining SAD with gradient based measure) algorithms have the highest number of errors, followed by the CT algorithm and the ZNCC algorithm with an almost similar error percentage. The proposed CT based algorithm has the lowest error percentage. Furthermore, Hirschmller, et al. [46] analysed performance of parametric

and non-parametric based cost functions with various gain changes, gamma changes, vignette effects, and signal-to-noise ratios (SNR).

In this evaluation, the CT algorithm has lower error and is more consistent than the ZNCC algorithm. Thus, the CT based algorithm is more suitable to deal with radiometric distortion among local approaches. A zero mean-based algorithm like the ZNCC algorithm is costly in terms of processing speed. In addition, a bilateral filter is used by Hamzah, et al. [40] with the SAD algorithm and the performance evaluation indicates that the bilateral filter is efficient in reducing the errors caused by radiometric distortion.

Nevertheless, it might not perform well when an image exists with many errors, but this situation can be enhanced by considering the original image information using a joint bilateral filter [93]. This will lead to better preserving effects. Still, this filter is prone to false matches at corners or edges on the disparity map. In order to maximize performance of the bilateral filter, Song, et al. [94] have proposed a combination of wavelet-based edge detection and joint bilateral filter (also known as wavelet edge joint bilateral filter) whereby information from the edge of the image is included for filtering image noise.

4.2 Occlusion, Depth Discontinuity and Homogenous Regions

Occlusion occurs in every stereo image pair because there must be some regions that only appear in one side of the stereo image. Depth discontinuity and homogenous regions are two other features of stereo images which we should be highly concerned about because false matching often occurs in this region. Hence, dealing with occlusion, depth discontinuity and homogenous areas is crucial to maximize the quality of the disparity map. Pandey, et al. [37] presented a low-rank sparse matrix completion algorithm to remove occluded areas as well as other falsely matched pixels by replacing the optimum disparity value in these areas using Nuclear norm minimization [95].

The advantage of this method is that it can be used in any approach. It works for both local and global types. Still, the performance of this implementation is not remarkable but can yield a 10% improvement in terms of accuracy. Zhang, et al. [84] used a cross-based window to improve the NCC algorithm by reducing the errors occurring at depth discontinuity areas, in which the shape of local window varies based on the discontinuity region with addition of computational load. Furthermore, Yang, et al. [60] apply a cross-based window in their DP algorithm and the results have shown that not only accuracy is improved, but the streaking effect is minimized.

Baek, et al. [59] focus on determining occlusion before eliminating it from a disparity map. They identified occluded regions through modelling an energy function using an ordering constraint, a uniqueness constraint, and a colour similarity constraint. The ordering constraint estimates candidates of occlusion, the uniqueness constraint justifies mutual consistency of both disparity map, while the colour similarity constraint further evaluates mutual consistency between the two stereo images using Euclidean distance. Then, these areas are removed using the DP algorithm with guided image filtering. Experiments have shown that the proposed algorithm is able to eliminate more errors of occlusion than conventional methods. Weight optimization of the three constraints is the main challenge to ensure the proposed algorithm can perform consistently.

In addition, colour segmentation is a good option for reducing computation complexity and errors due to occlusion, discontinuity, and textureless regions. Hong, et al. [69] used a mean-shift colour segmentation approach on a GC algorithm and changed from the pixel domain to the segment domain. There are several advantages when using a segment domain. First, the number of segments is far smaller than the number of pixels, hence computational cost will be decreased significantly.

Furthermore, uniform and smooth disparity distribution within a homogenous region can be guaranteed. With the use of smoothness terms in the energy function, the smoothness of disparity between segments can be preserved easily. However, the proposed algorithm still takes seconds to complete disparity estimation without considering colour segmentation computation. Further research needs to be carried out to reduce the cost in order to meet real time requirements. Recently, there are other improvements on the segment domain-based GC algorithm.

Wang, et al. [89] increased the accuracy of the algorithm by adding Singular Value Decomposition and setting three rules for disparity plane fitting: cross-checking, judging reliable region, and measuring the distance between previous disparity to the computed disparity plane. Also, Altantawy, et al. [70] introduced the Non-Local-Mean method into the segment based GC algorithm to eliminate image noise in a faster way with an edge preserving property.

4.3 Computational Cost

A lot of studies are done on reducing computational complexity of stereo vision algorithms. For software based improvement, Hoa, et al. [38] reduced the number of pixels being searched of the SAD algorithm to only 17% of the total pixels, which lowers computational cost approximately 89% compared to the conventional SAD algorithm. In this study, only edge or corner pixels are selected for matching cost and disparity estimation while the rest are excluded from estimation.

Still, this modification generates a sparse disparity map instead of a dense disparity map, since limited

pixels are chosen for disparity calculation without any filling process. Fućek, et al. [96] transformed the disparity map from previous frames for use with the current frame for live stream applications. This way is not only lowering computational cost, since less work is needed to update the disparity map, but also is increasing accuracy as the current disparity map is using the previous disparity map as reference.

Tippetts, et al. [97] have presented the Intensity Profile-Shape algorithm and the matching is based on image profile patterns using gradient measure with the purpose of increasing processing speed. Local maxima are identified as vertices for both images, and they are computed through comparing the gradient difference. The comparison will continue to spread away from vertices until the gradient difference is greater than a certain threshold and the region between the two vertices with uniform gradient will be labelled as the same shape.

Disparity estimation can be done at high speed once matching of vertices between left and right images is completed by filling the same disparity value inside the same region. It has some similarities with the classic DP algorithm, as it completes matching horizontally between image pairs. Although it performs a vertical smoothing pass at the end, the streaking effect is still significant, similar to the classic DP algorithm. Sarkis, et al. [74] decreased the computation complexity of the BP algorithm through the matching of sparse left stereo images with dense right stereo images. Although this matching will lead to a sparse disparity map, a dense disparity map can be recovered using a simple interpolation method. This modification saves 25% of the computational cost compared to the conventional BP algorithm.

Besides, for hardware based improvement, Georgoulas, et al. [2] increased stereo vision algorithm speed through vergence control of binocular cameras, in which shortening the input of disparity range means less work is needed to estimate the disparity value. This proposed algorithm can perform up to 320 frames per second on FPGA devices with an image size of 640x480, and is applicable for object tracking purposes. However, the quality of the generated disparity maps is a limitation of this algorithm. For applications with high requirements for accuracy and smoothness, it is not recommended to use this proposed algorithm.

Also, implementation in hardware, such as in a GPU, FPGA, or ASIC, can accelerate certain stereo vision algorithms drastically because of the hardwares parallel nature or optimality in terms of logic gates. Kuhn, et al. [98] have designed a hardware-efficient architecture which allows processing the SSD algorithm and the CT algorithm simultaneously with low computational load. Their work can achieve 50 frames per second, though the disparity range is only 25 and is using a small image size of 256×192 as input.

Banz, et al. [48] have implemented their proposed algorithm on a FPGA and they able to run 30 frames per second with a disparity range of 128 and 640x480 image input size settings. Hernandez-Juarez, et al. [57] presented a work applying the CT based semi-global matching algorithm on an embedded GPU system and the processing speed reached 42 frames per second with input settings similar to the work of Banz, et al. [48].

5. CONCLUSION

Development of stereo vision algorithms remains a challenge for researchers. More and more new approaches are developed to achieve a satisfactory disparity map or higher processing speed. It is a time-consuming task to become familiar with the state-of-the-art algorithms. This review paper provides a summary of recent stereo vision algorithms. Also, major challenges of stereo vision are discussed together with available solutions to tackle the issues. It will be useful as a reference when designing and developing new stereo vision algorithms with better performance in terms of disparity map quality and computation. Generally, research on local approaches mainly focuses on increasing quality of the disparity map through improving algorithms in terms of cost computation, cost aggregation, or refinement processes. For a global approach, on the other hand, the research focuses in two directions: increasing accuracy with more calculations and complexity, or decreasing computational load by simplifying algorithms without affecting disparity map quality or, if possible, improving it further. Appropriate implementation in hardware can greatly accelerate algorithm runtimes. Although the global approach has received more attention than the local approach recently due to higher accuracy, and a belief that the computational load issue will be minimized eventually with the help of technological advancement, current existing local approaches have yet to reach their potential limits. With appropriate occlusion, homogenous areas, and discontinuity handling, we believe that a local approach is able to estimate a comparable disparity map with limited information within constrained search space areas.

ACKNOWLEDGEMENT

This publication was supported by Malaysia Ministry of Education (MOE) - FRGS/1/2019/TK08/USM/02/1 Grant and a Universiti Sains Malaysia - USM Fellowship.

References

[1] Y. Sumi, Y. Kawai, T. Yoshimi, F. Tomita, “3D object recognition in cluttered environments by

segment-based stereo vision," *International Journal of Computer Vision*, vol.46, pp.5-23, 2002.

[2] C. Georgoulas, I. Andreadis, "FPGA based disparity map computation with vergence control," *Microprocessors and Microsystems*, vol.34, pp.259-273, 2010.

[3] S.N. Sinha, P. Mordohai, M. Pollefeys, "Multi-view stereo via graph cuts on the dual of an adaptive tetrahedral mesh," in: *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, IEEE*,pp. 1-8, 2007.

[4] S.G. Bahnemiri, A. Mousavinia, "Environment mapping with stereo vision and Belief Propagation algorithm," in: *Knowledge-Based Engineering and Innovation (KBEI), 2017 IEEE 4th International Conference on, IEEE*, pp.0101-0107, 2017.

[5] B.H. Bodkin, *Real-Time Mobile Stereo Vision*, (2012).

[6] D. Scharstein, R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International journal of computer vision*, vol.47, pp.7-42, 2002.

[7] M.Z. Brown, D. Burschka, G.D. Hager, "Advances in computational stereo," *IEEE transactions on pattern analysis and machine intelligence*, vol.25, pp.993-1008, 2003.

[8] M. Gong, R. Yang, L. Wang, M. Gong, "A performance study on different cost aggregation approaches used in real-time stereo matching," *International Journal of Computer Vision*, vol.75, pp.283-296, 2007.

[9] L. Nalpantidis, G.C. Sirakoulis, A. Gasteratos, "Review of stereo matching algorithms for 3D vision," in: *16th International Symposium on Measurement and Control in Robotics 21-23 June 2007-Warsaw*, POLAND, 2007.

[10] F. Tombari, S. Mattoccia, L. Di Stefano, E. Addimanda, "Classification and evaluation of cost aggregation methods for stereo correspondence," in: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, IEEE*, pp. 1-8, 2008.

[11] N. Lazaros, G.C. Sirakoulis, A. Gasteratos, "Review of stereo vision algorithms: from software to hardware," *International Journal of Optomechatronics*, vol.2, pp.435-462, 2008.

[12] J. Fang, A.L. Varbanescu, J. Shen, H. Sips, G. Saygili, L. Van Der Maaten, "Accelerating cost aggregation for real-time stereo matching," in: *2012 IEEE 18th International Conference on Parallel and Distributed Systems, IEEE*, pp.472-481, 2012.

[13] D. Kumari, K. Kaur, "A survey on stereo matching techniques for 3D vision in image processing," *Int. J. Eng. Manuf*, vol.4, pp.40-49, 2016.

[14] R.A. Hamzah, H. Ibrahim, "Literature survey on stereo vision disparity map algorithms," *Journal of Sensors*, 2016.

[15] M. Hariyama, T. Takeuchi, M. Kameyama, "VLSI processor for reliable stereo matching based on adaptive window-size selection," in:*Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No. 01CH37164), IEEE*, pp. 1168-1173, 2001.

[16] L. Wang, M. Gong, M. Gong, R. Yang, "How far can we go with local optimization in real-time stereo matching," in: *3D Data Processing, Visualization, and Transmission, Third International Symposium on, IEEE*, pp.129-136, 2006.

[17] Q. Yang, P. Ji, D. Li, S. Yao, M. Zhang, "Fast stereo matching using adaptive guided filtering," *Image and Vision Computing*, vol.32, pp.202-211, 2014.

[18] A.F. Bobick, S.S. Intille, "Large occlusion stereo," *International Journal of Computer Vision*, vol.33, pp.181-200, 1999.

[19] H. Hirschmller, P.R. Innocent, J. Garibaldi, "Real-time correlation-based stereo vision with reduced border errors," *International Journal of Computer Vision*, vol.47, pp.229-246, 2002.

[20] O. Veksler, "Fast variable window for stereo correspondence using integral images," in: *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, IEEE*, pp. I-I, 2003.

[21] R.K. Gupta, S.-Y. Cho, "Window-based approach for fast stereo correspondence," *IET Computer Vision*, vol.7, pp.123-134, 2013.

[22] K.-J. Yoon, I.-S. Kweon, "Locally adaptive support-weight approach for visual correspondence search," in: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, IEEE*, pp. 924-931, 2005.

[23] C.C. Pham, J.W. Jeon, Domain transformation-based efficient cost aggregation for local stereo matching," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.23, pp.1119-1130, 2013.

[24] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.35, pp.504-511, 2012.

[25] K.-J. Yoon, I.S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE transactions on pattern analysis & machine intelligence*, pp.650-656, 2006.

[26] L. De-Maeztu, A. Villanueva, R. Cabeza, "Near real-time stereo matching using geodesic diffusion," *IEEE transactions on pattern analysis and machine intelligence*, vol.34, pp.410-416, 2011.

[27] N. Einecke, J. Eggert, "A multi-block-matching approach for stereo," in:*Intelligent Vehicles Sym-*

posium (IV), 2015 IEEE, IEEE, pp. 585-592, 2015.

[28] K. Zhang, J. Lu, G. Lafruit, "Cross-based local stereo matching using orthogonal integral images," *IEEE transactions on circuits and systems for video technology*, vol.19, pp.1073-1079, 2009

[29] F. Tombari, S. Mattoccia, L. Di Stefano, E. Addimanda, "Near real-time stereo based on effective cost aggregation," in: *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, IEEE*, pp. 1-4, 2008.

[30] R. Trapp, S. Drüe, G. Hartmann, "Stereo matching with implicit detection of occlusions," in: *European Conference on Computer Vision, Springer*, pp.17-33, 1998.

[31] R.P. Wildes, "Direct recovery of three-dimensional scene geometry from binocular stereo disparity," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, pp.761-774, 1991.

[32] B.L. Anderson, K. Nakayama, "Toward a general theory of stereopsis: binocular matching, occluding contours, and fusion," *Psychological review*, vol.101, pp.414, 1994.

[33] D. Geiger, B. Ladendorf, A. Yuille, "Occlusions and binocular stereo," *International Journal of Computer Vision*, vol.14, pp.211-226, 1995.

[34] K. Mühlmann, D. Maier, J. Hesser, R. Männer, "Calculating dense disparity maps from color stereo images, an efficient implementation," *International Journal of Computer Vision*, vol.47, pp.79-88, 2002.

[35] M.J. Hannah, "Computer matching of areas in stereo images," in *Stanford University, California, Department of Computer Science*, 1974.

[36] M. Gong, R. Yang, "Image-gradient-guided real-time stereo on graphics hardware," in: *3-D Digital Imaging and Modeling, 2005. 3DIM 2005. Fifth International Conference on, IEEE*, pp.548-555, 2005.

[37] P. Pandey, A. Goel, "Novel method for occlusion reduction and disparity refinement," in: *Image Information Processing (ICIIP), 2017 Fourth International Conference on, IEEE*, pp.1-4, 2017.

[38] D.K. Hoa, L. Dung, N.T. Dzung, "Efficient determination of disparity map from stereo images with modified sum of absolute differences (SAD) algorithm," in: *Advanced Technologies for Communications (ATC), 2013 International Conference on, IEEE*, pp.657-660, 2013.

[39] V.G. Posugade, R.P. Patil, "A novel framework for disparity estimation in FPGA," in: *Automatic Control and Dynamic Optimization Techniques (ICACDOT), International Conference on, IEEE*, pp.1102-1105, 2016.

[40] R.A. Hamzah, M.S. Hamid, A. Kadmin, S.F.A. Ghani, S.S. Fakulti, "Matching cost computation based on sum of absolute RGB differences," in: *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), IEEE*, pp.317-320, 2018.

[41] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," *International Journal of Computer Vision*, vol.2, pp.283-310, 1989.

[42] L. Matthies, T. Kanade, R. Szeliski, "Kalman filter-based algorithms for estimating depth from image sequences," *International Journal of Computer Vision*, vol.3, pp.209-238, 1989.

[43] H. Hirschmuller, D. Scharstein, "Evaluation of cost functions for stereo matching," in: *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, IEEE*, pp. 1-8, 2007.

[44] S.i. Satoh, "Simple low-dimensional features approximating NCC-based image matching," *Pattern Recognition Letters*, vol.32, pp.1902-1911, 2011.

[45] J. Banks, M. Bennamoun, P. Corke, "Non-parametric techniques for fast and robust stereo matching," in: *TENCON'97. IEEE Region 10 Annual Conference. Speech and Image Technologies for Computing and Telecommunications., Proceedings of IEEE, IEEE*, pp.365-368, 1997.

[46] H. Hirschmller, D. Scharstein, "Evaluation of stereo matching costs on images with radiometric differences," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, pp.1582-1599, 2008.

[47] R. Zabih, J. Woodfill, Non-parametric local transforms for computing visual correspondence," in: *European conference on computer vision, Springer*, pp.151-158, 1994.

[48] C. Banz, S. Hesselbarth, H. Flatt, H. Blume, P. Pirsch, "Real-time stereo vision system using semi-global matching disparity estimation: Architecture and FPGA-implementation," in: *Embedded Computer Systems (SAMOS), 2010 International Conference on, IEEE*, pp. 93-101, 2010.

[49] S.H. Lee, S. Sharma, "Real-time disparity estimation algorithm for stereo camera systems," *IEEE transactions on Consumer electronics*, vol.57, 2011.

[50] J. Lim, Y. Kim, S. Lee, "A census transform-based robust stereo matching under radiometric changes," in: *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2016 Asia-Pacific, IEEE*, pp. 1-4, 2016.

[51] S. Gautama, S. Lacroix, M. Devy, "Evaluation of stereo matching algorithms for occupant detection," in: *Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 1999. Proceedings. International Workshop on, IEEE*, pp. 177-184, 1999.

[52] H. Hirschmuller, "Improvements in real-time correlation-based stereo vision," in: *Stereo and Multi-Baseline Vision, 2001.(SMBV 2001). Proceedings. IEEE Workshop on, IEEE*, pp.141-148,

2001.

[53] C. Murphy, D. Lindquist, A.M. Rynning, T. Cecil, S. Leavitt, M.L. Chang, "Low-cost stereo vision on an FPGA," in: *Field-Programmable Custom Computing Machines, 2007. FCCM 2007. 15th Annual IEEE Symposium on, IEEE*, pp.333-334, 2007.

[54] J.I. Woodfill, G. Gordon, R. Buck, "Tyzx deepsea high speed stereo vision system," in: *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on, IEEE*, pp. 41-41, 2004.

[55] M.A. Ibarra-Manzano, D.-L. Almanza-Ojeda, M. Devy, J.-L. Boizard, J.-Y. Fourniols, "Stereo vision algorithm implementation in fpga using census transform for effective resource optimization," in: *Digital System Design, Architectures, Methods and Tools, 2009. DSD'09. 12th Euromicro Conference on, IEEE*, pp. 799-805, 2009.

[56] C. Zinner, M. Humenberger, K. Ambrosch, W. Kubinger, "An optimized software-based implementation of a census-based stereo matching algorithm," in: *International Symposium on Visual Computing, Springer*, pp. 216-227, 2008.

[57] D. Hernandez-Juarez, A. Chacón, A. Espinosa, D. Vázquez, J.C. Moure, A.M. López, "Embedded real-time stereo estimation via semi-global matching on the GPU," *Procedia Computer Science*, vol.80, pp.143-153, 2016.

[58] K. Song, X. Wen, Y. Zhao, Z. Dong, Y. Yan, "Noise robust image matching using adjacent evaluation census transform and wavelet edge joint bilateral filter in stereo vision," *Journal of Visual Communication and Image Representation*, vol.38, pp.487-503, 2016.

[59] E.-T. Baek, Y.-S. Ho, "Occlusion and error detection for stereo matching and hole-filling using dynamic programming," *Electronic Imaging, 2016* pp.1-6, 2016.

[60] Q.-Q. Yang, L.-H. Wang, D.-X. Li, M. Zhang, "Hybrid stereo matching by dynamic programming with enhanced cost entry for real-time depth generation," in: *Audio, Language and Image Processing (ICALIP), 2012 International Conference on, IEEE*, pp. 557-563, 2012.

[61] R.L. Rivest, C.E. Leiserson, *Introduction to algorithms*, McGraw-Hill, Inc., 1990.

[62] S.-F. Hsiao, W.-L. Wang, P.-S. Wu, "VLSI implementations of stereo matching using dynamic programming," in: *VLSI Design, Automation and Test (VLSI-DAT), 2014 International Symposium on, IEEE*, pp. 1-4, 2014.

[63] Y.-C. Fan, Y.-H. Jiang, C.-L. Chen, "Disparity measurement using dynamic programming," in: *Instrumentation and Measurement Technology Conference (I2MTC), 2012 IEEE International, IEEE*, pp. 2683-2686, 2012.

[64] J.C. Kim, K.M. Lee, B.T. Choi, S.U. Lee, "A dense stereo matching using two-pass dynamic programming with generalized ground control points," in: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, IEEE*, pp.1075-1082, 2005.

[65] E.F. Sawires, A.M. Hamdy, F.Z. Amer, E. Bakr, "Disparity map using suboptimal cost with dynamic programming," in: *Signal Processing and Information Technology (ISSPIT), 2010 IEEE International Symposium on, IEEE*, pp.209-214, 2010.

[66] J. Witt, U. Weltin, "Sparse stereo by edge-based search using dynamic programming," in: *ICPR*, pp.3631-3635, 2012.

[67] J.K. Suhr, H.G. Jung, "Dense stereo-based robust vertical road profile estimation using Hough transform and dynamic programming," *IEEE Transactions on Intelligent Transportation Systems*, vol.16, pp.1528-1536, 2015.

[68] Y. Boykov, O. Veksler, R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on pattern analysis and machine intelligence*, vol.23, pp.1222-1239, 2001.

[69] L. Hong, G. Chen, "Segment-based stereo matching using graph cuts," in: *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, IEEE*, pp.I-I, 2004.

[70] D.A. Altantawy, M. Obbaya, S. Kishk, "A fast non-local based stereo matching algorithm using graph cuts," in: *Computer Engineering & Systems (ICCES), 2014 9th International Conference on, IEEE*, pp.130-135, 2014.

[71] L. Feng, K. Qin, "Superpixel-based graph cuts for accurate stereo matching," in: *IOP Conference Series: Earth and Environmental Science, IOP Publishing*, pp. 012161, 2017.

[72] F.-z. Wang, D.-g. Huang, S. Ge, "Belief propagation stereo matching based on differential geometry constraint of disparity," in: *Digital Manufacturing and Automation (ICDMA), 2010 International Conference on, IEEE*, pp.324-327, 2010.

[73] S.-S. Wu, C.-H. Tsai, L.-G. Chen, "Efficient hardware architecture for large disparity range stereo matching based on belief propagation," in: *Signal Processing Systems (SiPS), 2016 IEEE International Workshop on, IEEE*, pp. 236-241, 2016.

[74] M. Sarkis, K. Diepold, "Sparse stereo matching using belief propagation," in: *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, IEEE*, pp.1780-1783, 2008.

[75] C. Liang, L. Wang, H. Liu, "Stereo matching with cross-based region, hierarchical belief propagation and occlusion handling," in: *Mechatronics and Automation (ICMA), 2011 International Conference on, IEEE*, pp. 1999-2003, 2011.

[76] C. Luo, J. Lei, G. Hu, K. Fan, S. Bu,

"Stereo Matching with Semi-limited Belief Propagation," in: *Genetic and Evolutionary Computing (ICGEC), 2012 Sixth International Conference on, IEEE*, pp. 1-4, 2012

[77] J.-F. Nezan, A. Mercat, P. Delmas, G. Gimelfarb, "Optimized belief propagation algorithm onto embedded multi and many-core systems for stereo matching," in: *Parallel, Distributed, and Network-Based Processing (PDP), 2016 24th Euromicro International Conference on, IEEE*, pp. 332-336, 2016.

[78] A. Miron, S. Ainouz, A. Rogozan, A. Bensrhair, "A robust cost function for stereo matching of road scenes," *Pattern Recognition Letters*, vol.38, pp.70-77, 2014.

[79] M.F. Tappen, W.T. Freeman, "Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters," in: *null, IEEE*, pp. 900, 2003.

[80] Y. Boykov, V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," in: *International workshop on energy minimization methods in computer vision and pattern recognition, Springer*, pp.359-374, 2001.

[81] J. Sun, N.-N. Zheng, H.-Y. Shum, "Stereo matching using belief propagation," *IEEE Transactions on pattern analysis and machine intelligence*, vol.25, pp.787-800, 2003.

[82] A. Klaus, M. Sormann, K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," in: *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, IEEE*, pp.15-18, 2006.

[83] K. Briechle, U.D. Hanebeck, "Template matching using fast normalized cross correlation," in: *Optical Pattern Recognition XII, International Society for Optics and Photonics*, pp. 95-103, 2001.

[84] K. Zhang, J. Lu, G. Lafruit, R. Lauwereins, L. Van Gool, "Robust stereo matching with fast normalized cross-correlation over shape-adaptive regions," in: *Image Processing (ICIP), 2009 16th IEEE International Conference on, IEEE*, pp.2357-2360, 2009.

[85] E. Binaghi, I. Gallo, G. Marino, M. Raspanti, "Neural adaptive stereo matching," *Pattern Recognition Letters*, vol.25, pp.1743-1758, 2004.

[86] C. Cigla, A.A. Alatan, "Information permeability for stereo matching," *Signal Processing: Image Communication*, vol.28, pp.1072-1088, 2013.

[87] L. Wang, M. Liao, M. Gong, R. Yang, D. Nister, "High-quality real-time stereo using adaptive cost aggregation and dynamic programming," in: *null, IEEE*, pp.798-805, 2006.

[88] M. Gong, Y.-H. Yang, "Real-time stereo matching using orthogonal reliability-based dynamic programming," *IEEE Transactions on Image Processing*, vol.16, pp.879-884, 2007.

[89] D. Wang, K.B. Lim, "A new segment-based stereo matching using graph cuts," in: *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on, IEEE*, pp. 410-416, 2010.

[90] T. Taniai, Y. Matsushita, T. Naemura, "Graph cut based continuous stereo matching using locally shared labels," in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1613-1620, 2014.

[91] E.-T. Baek, Y.-S. Ho, "Temporal stereo disparity estimation with graph cuts," in: *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2015 Asia-Pacific, IEEE*, pp. 184-187, 2015.

[92] Y. Geng, Y. Zhao, H. Chen, "Improved belief propagation based on RGB vector measure for stereo matching," in: *Wireless Communications and Signal Processing (WCSP), 2011 International Conference on, IEEE*, pp. 1-5, 2011.

[93] J. Kopf, M.F. Cohen, D. Lischinski, M. Uyttendaele, "Joint bilateral upsampling, ACM Transactions on Graphics (ToG), 26 (2007) 96.

[94] K. Song, Y. Yan, M. NIU, C. LIU, "Effective stereo matching method with equicrural triangle census transform," *J. Comput. Inform. Syst*, vol.8, pp.7769-7780, 2015.

[95] B. Recht, M. Fazel, P.A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM review*, vol.52, pp.471-501, 2010.

[96] L. Fućek, I. Marković, I. Cvišić, I. Petrović, "Dense Disparity Estimation in Ego-motion Reduced Search Space," *IFAC-PapersOnLine*, vol.50, pp.10122-10127, 2017.

[97] B.J. Tippetts, D.-J. Lee, J.K. Archibald, K.D. Lillywhite, "Dense disparity real-time stereo vision algorithm for resource-limited systems," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.21, pp.1547-1555, 2011.

[98] M. Kuhn, S. Moser, O. Isler, F.K. Gurkaynak, A. Burg, N. Felber, H. Kaeslin, W. Fichtner, "Efficient ASIC implementation of a real-time depth mapping stereo vision system," in: *Circuits and Systems, 2003 IEEE 46th Midwest Symposium on, IEEE*, pp.1478-1481, 2003.

Kai Yit Kok received the B.Eng. (Hons.) in aerospace engineering from Universiti Sains Malaysia in 2014 and the M.Sc. degree in aerospace engineering from Universiti Sains Malaysia in 2016, where he is currently pursuing the Ph.D. degree with the School of Aerospace Engineering. His research includes aircraft control, UAV path planning, image processing and stereo vision.

Parvathy Rajendran is currently an academic in the School of Aerospace Engineering at Universiti Sains Malaysia since 2013. She completed her Ph.D. in Aerospace Engineering from Cranfield University, United Kingdom in October 2012. There, her research includes UAV design, development and flight testing and UAVs systems development and testing. Rajendran has produced many high-impact publications and served as an editor-in chief, guest editor, international advisor and reviewer. She has been the chairman and member of the technical conference committee of various international conferences. In addition, she has maintained various grants worth more than RM 1 million.