DISPARITY REFINEMENT BASED ON DEPTH IMAGE LAYERS SEPARATION FOR STEREO MATCHING ALGORITHMS

Nurulfajar Abd Manap[1],John J. Soraghan[2]

[1]Faculty of Electronic & Computer Engineering,
Universiti Teknikal Malaysia Melaka, Malaysia

[2]Centre for Excellence in Signal and Image Processing (CeSIP),
Electronic and Electrical Engineering,
University of Strathclyde, Glasgow, UK

Email: nurulfajar@utem.edu.m, j.soraghan@eee.strath.ac.uk

Abstract

This paper presents a method to improve the raw disparity maps in the disparity refinement stage for stereo matching algorithm. The proposed algorithm will use the disparity depth map from the stereo matching algorithm as initial disparity depth output with a basic similarity metric of SAD. The similarity metric finds the pixel points between the left and right under the fixed window (FW) searching process. With this approach, the raw disparity depth map obtained is not smooth and contained errors particularly with the depth discontinuities and unable to detect the uniform areas and repetitive patterns. The initial disparity depth will be used to identify the layers of disparity depth map by adapting the Depth Image Layers Separation (DILS) algorithm that separate the layers of depth based on disparity range. Each particular disparity depth map distributed along the disparity range and can be divided into several layers. The layer will be mapped to segmented reference image to refine the disparity depth map. This method will be known as the Depth Layer Refinement (DLR) that using the disparity depth layers to refine the disparity map.

Keywords: *Depth map, stereo matching algorithm, disparity refinement stage, similarity metric, layered depth map.*

I. INTRODUCTION

Binocular stereo is one of the most significant and active areas in the field of computer vision. Recently, the number of publications on stereo is increasing due to the Middlebury Stereo Vision Page by Scharstein and Szelinski [1] with their taxonomy of stereo matching algorithms development. The Middlebury page provides some common benchmark datasets and evaluation systems that all researchers can utilize to examine their proposed methods objectively and universally. Based on the rank given by the website, the common techniques can be found and adopted in many sophisticated algorithms. According to Scharstein [1] that build the foundation of the page, stereo algorithms generally consist of four steps including matching cost computation, cost aggregation, disparity computation optimization and disparity refinement. However, not all stereo algorithms take all the four steps depending on the individual implementation.

The post-processing step for the stereo matching algorithm is the disparity refinement has received a lot of attention in recent years. Most pixel-based matching algorithms compute disparities as integer values and need to be refined. In this step, raw disparity maps computed by correspondence algorithms contain outliers that must be identified and corrected. Several approaches aimed at improving the raw disparity maps computed by stereo correspondence algorithms such as sub-pixel interpolation [2], image filtering techniques, Bidirectional Matching [3] and Single Matching Phase [4]. Even though the proposed algorithm provides

exceptional accurate disparity depth map, it suffered with complexity for the implementation particularly for real-time application.

In this paper, the main aim of this research is to improve the raw disparity maps in the disparity refinement stage. The algorithm will use a simple stereo matching correspondence with a basic similarity metric of SAD. The similarity metric finds the pixel points between the left and right under the fixed window (FW) searching process. With this approach, the raw disparity depth map obtained is not smooth and contained errors particularly with the depth discontinuities and unable to detect the uniform areas and repetitive patterns. The proposed algorithm will use the disparity depth map from the stereo matching algorithm as initial disparity depth output. The initial disparity depth will be used to identify the layers of disparity depth map since the depth consists of range of disparity. This approach is adapted from the Depth Image Layers Separation (DILS) algorithm that separate the layers of depth based on disparity range. In general, each particular disparity depth map distributed along the disparity range and can be divided into several segments, which is known as layers. Instead of using each layer to synthesize inter-view images in the DILS, the layer will be mapped to segmented reference image to refine the disparity depth map. This method will be known as the Depth Layer Refinement (DLR) that using the disparity depth layers to refine the disparity map.

This paper is organized in six sections. Section 2 provides an overview of the system design and also outlines the main features of the model that consist two main modules: stereo matching algorithm and disparity refinement module. Section 3 covers the proposed algorithm for the disparity refinement by adapting the Depth Image Layers Separation (DILS) algorithm. In section 4, performance evaluation used for the disparity depth map is presented. The results and performance are discussed in Section 5, which comparing the proposed algorithm with the state-of-the-art stereo matching algorithm in the Middlebury Ranking Stereo Page. And finally in Section 6 concluding remarks are provided.

II. SYSTEM DESIGN

The proposed system design of DLR is shown in Fig. 1 that consists of two stages: stereo matching engine and disparity refinement module. The first stage of DLR system is basically adapted from the stereo matching algorithm according to Scharstein [1] that contained three main components: matching cost computation, cost aggregation and disparity computation/optimization. In the matching cost computation step, it can be divided into two main categories that are pixel-based matching costs and area-based matching costs. Some similarity metric used in the matching are the Sum of Absolute Differences (SAD), Sum of Squared Differences (SSD) and Normalized Cross Correlation (NCC). The classification and evaluation of cost aggregation strategies for stereo correspondences [5] depends on the position, shape, position and weights.

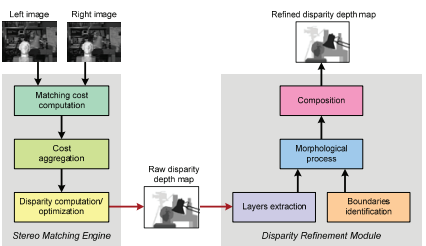


Fig. 1. Overview of DLR system

In this stage, raw disparity depth map obtained from the stereo matching based on left-to-right matching by using block-based fixed window similarity metric. In this case we are using the SAD that has been proven to be trade-off between reliability and computational cost [6]. However, other similarity metric can be used as well. Window-based methods

implicitly make the assumption of continuity by assuming constant disparity for all pixels inside the matching window. This assumption is broken at depth boundaries where occluded regions lead to erroneous matches, resulting in the familiar foreground flattening effect. Generally, the choice of an appropriate window size is a crucial decision. Small windows do not capture enough intensity variation to give correct results in less-textured regions. On the other hand, large windows tend to blur the depth boundaries and do not capture well small details and thin objects. This motivates the use of adaptive windows [7], shiftable window [8], multiple window [9], variable windows [10], bilateral filtering [11] and adaptive weights [12]. The newly algorithms adopted some of these approaches to improve the disparity depth map. In spite of its limitation, SAD with FW is the most frequently used algorithm for real time applications due to easy implementation, fast and has limited memory requirements. Therefore, the fixed window similarity stereo matching technique is adequate to obtain the estimated depth map. This configuration can be adapted for computation optimization in real-time hardware implementation [4].

The disparity computation or optimization step aims at finding the best disparity assignment that minimizes a cost function over the whole stereo pair. The relevant approaches are with the Graph Cuts [13-14], Belief Propagation [15] and Dynamic Programming [16-18]. The most common and effective method is a simple winner-takes-all (WTA) minimum or maximum search over all possible disparity levels. The matching can be done from right to left and vice versa (bidirectional matching), so occlusions and uncertain matches can be filtered with a left right consistency check (LRCC). This means only disparities with the same value (within a certain range) for both directions are accepted. In this case, only a single matching is needed for the DLR algorithm. The main reason of this is to use the depth layer and edge maps to remove the uncertain matches.

In the second stage, the disparity depth map will be separated into a number of layers based on the disparity range of the stereo pair. The disparity depth map can be improved with the same techniques such as sub-pixel interpolation [2], image filtering techniques, Bidirectional Matching [3] and Single Matching Phase [4]. Even though these algorithms provide exceptional accurate disparity depth map, it also required extra iterations to compute the mismatch between the uncertain pixels in the Bidirectional Matching and computational complexity within some of the proposed technique for real-time and practical implementation. The proposed disparity refinement developed through the layer extraction and separation process implemented using DILS algorithm. A new approach to refine the disparity image map is presented in this stage with boundaries identification, morphological and composition process, which are the DLR components. The layers mapped and adaptively fused with a reference image to identify the edge, border, depth discontinuities, uniform areas and repetitive patterns. The description on the disparity refinement module will be described in the next section.

III. DISPARITY LAYER REFINEMENT ALGORITHM

This section described the proposed algorithm of disparity layer refinement module. The overall algorithm for disparity refinement of DLR based on DILS algorithm can be divided into several major steps that are summarized in Fig. 2, which are the stereo matching and layers extraction (Part 1), boundaries and edges identification (Part 2), morphological process (Part 3) and lastly the layer composition stage (Part 4). The inputs to the matching engine are two stereo images in epipolar geometry.

The first processing step is the stereo matching and layer extraction as described in Section III.A. We calculate an initial disparity map using a fixed window-based correlation technique. The DILS algorithms will separate the disparity depth map into several numbers of layers depending on the complexity of the image pairs. The disparity levels and layers can be determined with the histogram distribution, which has been described in the DILS algorithm. The number of layers symbolized with i, from 1 to maximum *D*.

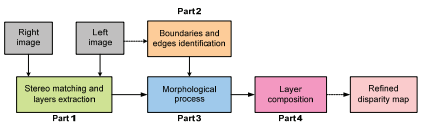


Fig. 2: Block diagram of the proposed algorithm on disparity refinement based on DIL

Since the discontinuities in the disparity map are usually reflected by discontinuities in the edge and color information, the borders of the segmented regions can be considered as a set of candidates for the boundaries of the disparity layers that we aim to compute and refine. With the layers has been identified, the left image is selected as the reference image. This image will undergo the edge detection and color segmentation process to obtain the edge and borders in the reference image. From this stage, the new edge map obtained as the reference image mask to create the edge boundaries of the layers. This process is explained in more detail in Section III.B through boundaries and edge detection stage.

In the next step (Section III.C), we create an initial representation for each extracted layer by separating the disparity depth obtained from the DILS in Part 1. The computed layer in the Part 3 obtained by fusing the disparity layer and edge map from Part 1 and 2 respectively. The mapping process of the layer and edge map created a new binary image mask layer that will be processed with the morphological operation. Each disparity depth map refined individually through layers separation and mapping process. As described in the previous section, the disparity depth map obtained based on left-to-right matching. The raw disparity depth map consists with the false matches and is not address any occlusion presence. Through individual layer refinement process, the noise, false matches can be removed without degrading the discontinuities in the edge map.

The last block in the DLR module is the layer composition, which explained in Section III.D. During this stage, the final disparity map composed by the extracted refined layers from the Part 2 and 3. The top layer is the object that closest to the camera view. It next layer under is the layer identified by the disparity range in the DILS algorithm. All the layers combined as a single disparity depth map. The cracks and holes corrected with the hole filling techniques.

A. Stereo Matching and Layers Extraction

The stereo matching and layers extraction are based on the left-to-right image matching. For this implementation, the matching cost computation use the basic similarity metric, SAD as conventional approach for many stereo matching algorithms. As described earlier, any similarity metric and approaches can be selected to enhance the accuracy and reliability of the disparity map obtained from this process. The main idea for this implementation is to show that by using a basic similarity metric, the disparity map can be improved significantly by using DLR algorithm. The matching process calculated by the Eq. (1) in the following,

$$SAD(x,y,d) = \sum_{i,j=-n}^{n} \left| I_R(x+i,y+j) - I_T(x+d+i,y+j) \right| \quad (1)$$

where I_R(x,y) and IT(x,y) are the gray-level intensities of the reference (left) and target (right) image respectively, window size of n*x*n, and d is the disparity. The

disparity value at (x,y), d(x,y) is the point where the correlation value of SAD is the smallest. Therefore, the disparity value is as follows,

$$d(x,y) = \underset{d_{\min} \le d \le d_{\max}}{\arg\min} \; SAD(x,y,d) \tag{2}$$

The cost aggregation is done by summing matching cost over fixed square windows searching with constant disparity. The accuracy of the depth map can be increased with bigger window size. However, there is the trade-off between the accuracy and the depth discontinuities of the objects. Many methods have been proposed to improve the disparity map with efficient and robust approach within the cost aggregation. As observed by Kanade and Okutomi [7], the correlation window covers a region with non-constant disparity is not performed well and the error in the depth discontinuities grows with the window size. Reducing the window size makes the computed disparity more noise-sensitive. To overcome this problem, Kanade proposed an adaptive window, which can statistically select at each pixel that minimizes the uncertainty in the disparity estimation. This approach has been improved by Fusiello [19] with the symmetric multi-window to provide efficient and robust disparity estimation in the present of occlusions. Although the presented cost aggregation by [7, 19-20] performed very well by improving the disparity map, the fixed square window sufficient for basic area-based stereo matching. This will provide faster implementation and low complexity. Furthermore, the configuration of the fixed square window can be adapted for computation optimization in the hardware parallel implementation that has been proposed by Stefano [4].

The raw disparity map can be visualized by selecting the minimal aggregated value at each pixel. For applications such as robotic navigation or people tracking, the disparity map obtained from this stage may be perfectly adequate. However for image-based rendering, the raw disparity maps lead to errors and unappealing view synthesis results. To enhance the performance for DILS algorithm, the raw disparity map filtered with a median filter, which can clean up mismatches, holes and noises. In our implementation, we are not performing bidirectional matching to calculate the occlusion since we want to measure the performance of the DILS and DLR algorithm components. Within this stage, we obtained two main results that are the raw disparity depth map and the layers of the disparity depth (from the DILS algorithm). The layer of depth can be easily identified with the number of matched pixels, p quantized as the following equation,

$$\begin{aligned} p'(d_k) &= 1, if\ p(d_k) > Tk \in 0,1,\ldots,d_{\max} \\ p'(d_k) &= 0, \quad elsewhere \end{aligned} \tag{3}$$

where T is the threshold to set the minimum number of pixels to be selected as the matched corresponding points for the stereo pair.

B. Boundaries and Edge Detection

In the second part, the boundaries and edges of the reference (left) image will be identified. By assuming that for regions of homogeneous colour, the disparity varies smoothly and the depth discontinuities coincide with the boundaries of those regions, which hold true for most natural scene as described by Bleyer [21]. This assumption is incorporated by applying colour segmentation to the reference image and by using a disparity layer to represent the disparity inside the new layer segments. In addition to the colour segmentation, the reference image derived the edge boundaries based on the edge detection algorithms. In theory, any algorithm that able to identify sharp edges and discontinuities in the edge detection can be used for the proposed boundaries and edge identification stage. Also, any algorithm that divides the reference image into regions of homogeneous colour can be used for this stage. In our implementation, we used mean-

shift segmentation algorithm proposed by Comaniciu [22] and incorporates edge information by using Canny edge detection algorithm.

Edge detection refers to the process of identifying and locating sharp discontinuities in an image. The discontinuities are abrupt changes in pixel intensity, which characterize boundaries of objects in a scene. Classical methods of edge detection involve convolving the image with an operator of 2-D filter, which is constructed, to be sensitive to large gradients in the image while returning values of zero in uniform regions. There are an extremely large number of edge detection operators available, each designed to be sensitive to certain types of edges. Variables involved in the selection of an edge detection operator include orientation, noise environment and structure. In edge orientation, the geometry of the operator determines a characteristic direction in which it is most sensitive to edges. Operators can be optimized to look for horizontal, vertical, or diagonal edges.

Edge detection is difficult in noisy images, since both the noise and the edges contain high-frequency content. Attempts to reduce the noise result in blurred and distorted edges. Operators used on noisy images are typically larger in scope, so they can average enough data to discount localized noisy pixels. This resulted less accurate localization of the detected edges. In the edge structure, not all edges involve a step change in intensity. Effects such as refraction or poor focus can result in objects with boundaries defined by a gradual change in intensity. The operator needs to be chosen to be responsive to such a gradual change in those cases.

The Canny edge detection algorithm is known as the optimal edge detector. It is important that edges occurring in images should not be missed and that there be no responses to non-edges. The second criterion is that the edge points be well localized. In other words, the distance between the edge pixels as found by the detector and the actual edge is to be at a minimum. A third criterion is to have only one response to a single edge. This was implemented because the first 2 were not substantial enough to completely eliminate the possibility of multiple responses to an edge.

Based on these criteria, the canny edge detector first smoothes the image to eliminate and noise. It then finds the image gradient to highlight regions with high spatial derivatives. The algorithm then tracks along these regions and suppresses any pixel that is not at the maximum (non-maximum suppression). The gradient array is now further reduced by hysteresis. Hysteresis is used to track along the remaining pixels that have not been suppressed. Hysteresis uses two thresholds and if the magnitude is below the first threshold, it is set to zero (made a non-edge). If the magnitude is above the high threshold, it is made an edge. And if the magnitude is between the 2 thresholds, then it is set to zero unless there is a path from this pixel to a pixel with a gradient above threshold. Therefore, the Canny edge detection is used along with the colour mean-shift segmentation.

The algorithm outline for Part 1 and 2 can be summarized in Fig. 3, where the results of stereo matching and layer extraction in Part 1 and the edge map image obtained in Part 2 used in the next in Part 3, which is the morphological process. The new segmented and edge map image defined as IS. The segmented image IS will be mapped and fused together with the layer i. The fusion process of the disparity depth layer and edge map image is described in the next section.

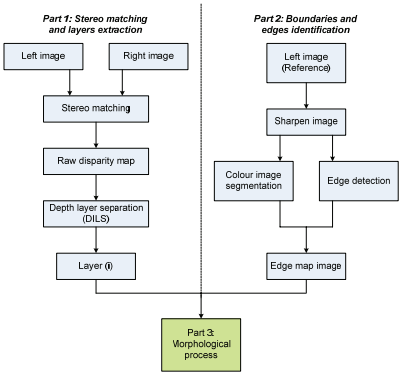


Fig. 3: Part 1 and 2 blocks for the DLR that consist: a) stereo matching and layers extraction, and b) boundaries and edges identification.

C. Morphological Process

The disparity depth map can be refined by median filter approach, where the outliers and noise can be removed. However, some of the noise unable to be removed automatically without affecting the whole portion of the disparity depth map obtained from the stereo matching algorithms. With disparity layer separation, particular noise can be easily removed while maintaining the quality some of the disparity layers. The accuracy of the disparity depth map can be enhanced with each layers processed with morphological process.

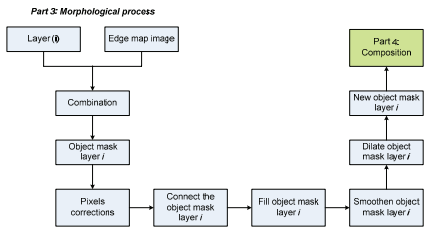
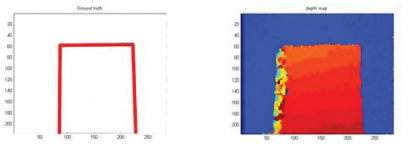


Fig. 4: Part 3 block of the DLR algorithm, morphological process

Fig. 4 shows the block diagram of the DLR algorithm in the Part 3 which taking the input of edge map image and layer i (separated by DILS algorithm). The combination of the input created the binary object map that holds the boundary of the edge layers. Each layers will be mapped on the same segmented edge map image I_S. Any edges and borders of the objects mapped and crossed with the same region on the layer i remained in the image, while the remaining will be removed. The new-segmented image now fused with the same region of layer i. The edge on the segmented image will now create a cross path along the layer i. The cross path is defined the new boundary notated as br and illustrated in Fig. 5(a), with the disparity depth map in Fig. 5(b).



(a) Cross path defined by the boundary (b) Sample of disparity depth map

Fig. 5: Sample of boundary path for layer I and the disparity depth map

The pixels in the object map will be corrected by removing unwanted pixels. The technique used in this block is based on erosion process combined with the algorithm proposed by Fergusson [23]. After that, the object map pixel of the layer connected with convex hull, which creates the closed-loop boundary region. The boundary region will be filled to produce the binary object map image.

During this stage, two regions of the disparity layer i can be distinguished based on the boundary created, which are the inner region and outer region. The inner region is the disparity depth map that contained inside the boundary. Any zero pixels on this region will be filled with the same value of layer i. The inner region is dilated till the boundary that sets as the threshold. Meanwhile, the outer region is for the disparity depth map that beyond the boundary edge of the segmented image. Any outer region of the disparity map will be eliminated. With this, the new disparity layer i created adaptively based on the boundary of object from the segmented reference image. This approach addresses the disparity depth discontinuities problems and able to detects the uniform areas and repetitive patterns on the stereo pairs.

The process can be illustrated in Fig. 6. The sample of output from the layer i and edge map image combination is shown in Fig. 7(a). The object mask layer i processed in the morphological stage that finally produced the new binary object mask layer I (Fig. 7(b)). This process iterated for all the layers of the disparity depth map before the layers can be composed as a single refined disparity depth map.

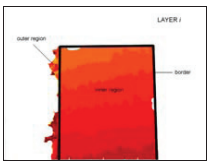
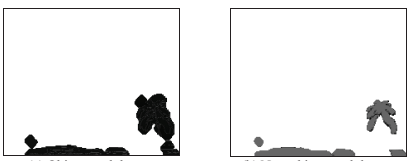


Fig. 6: Mapping and diffusing for layer i with the border set by the segmented reference image.



(a) Object mask layer *i* (b) New object mask layer *i*

Fig. 7: Layer extraction with edge map image. (a) Raw object mask layer i; (b) Refined layer i through morphological process

D. Layers Composition

Each layers of disparity depth map undergo the same process of mapping and fusing (diffusion) with the same reference segmented image. After all of the layers have been processed, the collection of layered disparity depth images merges into a new single refined disparity depth image. The process of the layer composition of the DLR algorithm can be summarized in Fig. 8.

IV. PERFORMANCE EVALUATION

In order to evaluate the performance of a stereo algorithm, a quantitative way is needed to estimate the quality of the computed correspondences. Two general approaches to this are to compute error statistics with respect to some ground truth data and to evaluate the synthetic images obtained by the disparity depth map. Two quality measures based on known ground truth data provided by the Middlebury Vision Page are RMS (root-mean-squared) error and percentage of bad matching pixels. RMS error measured in disparity units between the computed disparity map $d_c(x,y)$ and the ground truth map $d_T(x,y)$,

$$R = \left(\frac{1}{N} \sum_{(x,y)} \left| d_C(x,y) - d_T(x,y) \right|^2 \right)^{\frac{1}{2}} \quad (4)$$

where N is the total number of pixels. The percentage of bad matching pixels is given by,

$$B = \frac{1}{N} \sum_{(x,y)} \left(\left| d_C(x,y) - d_T(x,y) \right| > \delta_d \right) \quad (5)$$

where δ_d is a disparity error tolerance. For the experiments and evaluations, the disparity error tolerance, is set 1.0. In addition to compute these statistics over the whole image, two different kinds of regions are evaluated which are the non-occluded and depth discontinuities regions.

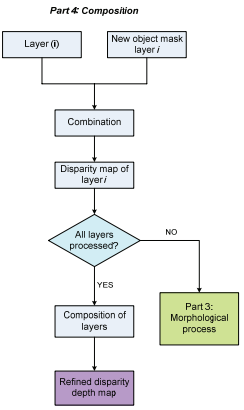


Fig. 8: Part 4 block of DLR algorithm, layers composition

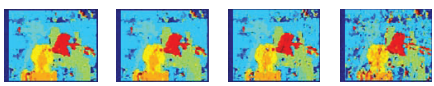
V. RESULTS AND DISCUSSION

The results for the depth refinement algorithms evaluated based on the performance evaluation with two approaches. The first performance is tested based on different similarity metric for the cost aggregation. Although the selected similarity metric is SAD, the comparison with different approaches will be shown. This includes the selection of window size for the correspondence matching. The second performance is based on the Middlebury Stereo Evaluation. The evaluation platform provides stereo image datasets consisting of the stereo image pair and the ground truth image. The proposed algorithm evaluated by using the Middlebury datasets and compared with results with many others through online. The performance evaluation based on this platform considered state-of-art for the reliability and constantly updated.

A. Performance Evaluation Based on Different Similarity Metric

This section gives a detailed evaluation of the proposed algorithm in term of results, quality and processing time. The DLR algorithm can be used with any stereo matching algorithm since it was developed to refine the raw disparity map images (in the post-processing stage). For this case, the evaluation has been made with Map (284x216 pixels) and Tsukuba (384x288 pixels) image with different similarity metric including SAD, SSD, SHD and NCC. The parameter of the stereo pair images set to 9x9 window size with maximum disparity 30(Map) and 16 (Tsukuba).

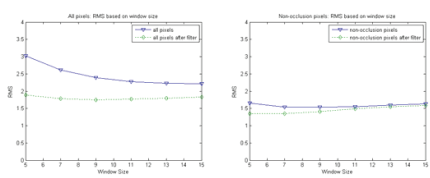
The results of stereo matching for Tsukuba image based on different similarity metric are shown in Fig. 9. The raw disparities based on the block-based window searching contained errors with unmatched pixels especially with the similarity metric in SHD. For this sample, the disparity depth map has been mapped with colour to show the hotter the colour, the closer of the object to the camera. In this case, the red colour (the lamp) is the closest object. The output of the disparity depth maps can be improved with the post-processing stage by using a median filter to smooth the result. The bidirectional matching can be used to eliminate the unmatched pixels, which can produce accurate disparity depth map.



(a) SAD (b) SSD (c) NCC (d) SHD

Fig. 9: Results of stereo matching based on different similarity metric

The size of window for the block-based matching affecting the performance as indicated in Fig. 10(a), where the RMS errors reduced accordingly when the window size increased for the all-pixels evaluation. The non-occlusion pixels errors are not affected with different window size as shown in Fig. 10(b). The errors reduced significantly when the disparity depth map filtered (in this case by using 11x11 median filter).



(a) (b)

Fig. 10: RMS error based on window size for all pixels and non-occlusion pixels.

The performance of different similarity metrics presented in Table 1 for Tsukuba and Map images. The table shows the statistics aimed at assessing the capability of the similarity metrics in term of processing time and RMS. The time is calculated in seconds for the processing time and the RMS in term of pixels. The stereo matching evaluated with Intel Quad CPU of 3.0 GHz, 3,25GB of RAM. The comparison of similarity metric with bidirectional matching (BM) also included. It is worth noticing that with the Map stereo pair, the similarity metrics of SAD, SSD and NCC perform similarly

and pretty well, with slightly better RMS yielded by BM. The BM shows the capability to deal with occlusions and not corrected disparities. The similarity metric performs better with Map image pairs compared to the Tsukuba due to the complex objects at different depths generating several occlusions, as well as poorly textured regions in the background. Moreover, this stereo pair contains some specular regions (such as the face of statue and some regions of the lamp) that quite difficult with stereo matching process.

Table 1: Processing time and RMS of Tsukuba and Map images

Algorithms	Map image		Tsukuba image	
	Time	RMS	Time	RMS
SAD	6.84	41.29	7.09	57.15
SSD	6.07	42.48	6.54	56.86
NCC	9.94	43.15	10.58	56.99
SHD	38.37	43.85	41.58	58.58
BM SAD	6.76	26.86	7.23	53.22
BM SSD	6.05	28.95	6.59	52.75
BM NCC	9.93	29.36	11.07	51.14
BM SHD	41.98	37.17	41.16	50.55

Based on this evaluation, it shows that the similarity metric by using SAD is satisfactory. Besides the simplicity, reliability and low computational cost, the SAD has been adapted for real-time implementation. Faster execution can be implemented by using the SAD through computational optimisation techniques which has been proposed by Stefano [3, 24].

B. Performance Based on Middlebury Stereo Evaluation

Scharstein and Szelinski [1] have developed an online evaluation platform, the Middlebury Stereo Evaluation [24], which provides a huge number of stereo image datasets consisting of the stereo image pair and the ground truth image. We evaluated our algorithm by using the Middlebury datasets and compared the results with many others online. The samples of these datasets are shown in the first row of Fig. 11, which consist the 'Tsukuba', 'Venus', 'Teddy' and 'Cones'. Since this evaluation is very well-known and state-of-the-art, the proposed algorithm in this work is also evaluated in this manner. In order to evaluate an algorithm on this website, the disparity maps of all four datasets have to be generated and uploaded to through online. The disparity maps have to correspond to the left stereo image and the disparities have to be scaled by a certain factor. The evaluation engine calculates the percentage of bad matched pixels within a certain error threshold by pixel-wise comparison with the ground truth image. This is done three times for each dataset. Firstly the disparity map image evaluated for all pixels where a ground truth value is available. Secondly, it will be evaluated for all non-occluded pixels. And lastly, the disparity map images compared for all pixels at disparity discontinuities. Many stereo algorithms researchers use this platform for evaluation and this gives a significant overview of how the developed algorithm performs in comparison to other algorithms. The platform is up-to-date and constantly mounting.

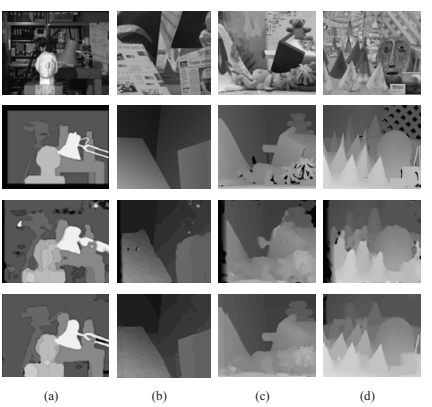


Fig. 11: Results of the proposed method by the Middlebury benchmark datasets: Tsukuba, Venus, Teddy and Cones. The first row images are the reference images of each set. The second row images are the ground truths. The third row images are the disparity maps by left-to-right SAD. The fourth row images are the disparity maps by DLR-SAD method.

Fig. 11 shows the Middlebury evaluation datasets, the ground truths of four datasets and the resulting disparity maps estimated by SAD and DLR-SAD

methods. The results based on third row in Figure 5.12 used the fixed window SAD of 21x21 ('Cones' and 'Teddy'), 11x11 ('Tsukuba') and 25x25 ('Venus'). Additionally, an 11x11 median filter is applied as a post-processing step for the SAD. The selected parameters are chosen to achieve the best possible result for the disparity maps. The results based on SAD has been further enhanced and refined by using the proposed algorithm, DLR where every disparity depth layers has been separated. Through the DLR, the new disparity maps have been formed. The fourth row of Fig. 11 indicates that disparity maps improved and removed the noise and errors in the basic SAD stereo matching. It is worth observing that several major occlusions and boundary discontinuities have been discarded, showing the ability of DLR to deal with this problem. The morphological operation processed in separated layers enables the unwanted regions, errors and noise can be removed efficiently. Due to the erosion and dilation process in the morphological operation, the final disparity map probably contained holes and cracks between the depth layers. Therefore, the missing values in the disparity maps have to extrapolate with adjacent pixel values by using hole-filing techniques.

Table 2 shows the performance of our method by the Middlebury ranking list with the error threshold of 1 pixel. The order of the algorithms is based on the average rank for bad pixels percentage. This value is very meaningful and shows how close together the algorithms in comparison. The basic fixed-window (FW) with SAD as cost aggregation methods placed in the last ranking. It shows that the stereo matching by using the basic approach is not accurate and contained with errors for the all regions, non-occluded and near depth discontinuities. However, after the FW-SAD refined by using the proposed method of DLR, the results significantly improved and the new results moved up to 13 places.

Table 2: Middlebury dataset ranking with the 1 pixel threshold. These values indicate the percentage of bad pixels whose errors are more than 1 pixel. In Venus and Tsukuba image, our method shows excellent results in non-occluded (N-o) region.

Algorithms	Tsukuba			Venus			Teddy			Cones			Avg (%)
	N-o	All	Disc	N-o	All	Disc	N-o	All	Disc	N-o	All	Disc	
2DPOC	2.88	4.8	10.5	6.55	7.8	17.4	14.4	22	27.8	15.2	23	24.5	14.7
Bipartite	2.54	4.4	13.6	6.62	7.5	18.6	16.9	24	30.2	15.1	22	23	15.4
SAD-DLR	4.22	5.1	19.5	2.50	3.2	18.3	18.2	19	37.2	18.0	21	32.9	16.5
Phase-based	4.26	6.5	15.4	6.71	8.2	26.4	14.5	23	25.5	10.8	21	21.2	15.3
RegionalSup	3.99	6.1	14.2	8.14	9.7	36.8	18.3	27	32.1	9.16	19	19.9	17.0
BioDEM	6.57	8.4	28.1	3.61	4.8	33.7	13.2	21	34.5	6.84	16	19.8	16.4
IMCT	4.54	5.9	19.8	3.16	3.8	23.2	18.0	23	35.3	12.7	19	19.8	16.3
SSD+MF [13]	5.23	7.1	24.1	3.74	5.2	11.9	16.5	25	32.9	10.6	20	26.3	15.7
SO [1]	5.08	7.2	12.2	9.44	11	21.9	19.9	28	26.3	13.0	23	22.3	16.6
MI-nonpara	5.59	7.5	18.8	7.50	9	35	17.4	26	36.9	10.2	20	22.6	18.0
PhaseDiff	4.89	7.1	16.3	8.34	9.8	26	20.0	28	29	19.8	29	27.5	18.8
STICA	7.70	9.6	27.8	8.19	9.6	40.3	15.8	23	37.7	9.8	18	28.7	19.7
Rank+ASW	6.51	8.4	19.7	10.5	12	32.7	15.7	24	32.8	14.1	23	21.7	18.4
LCDM+AdaptWgt	5.98	7.8	22.2	14.5	15	35.9	20.8	27	38.3	8.9	17	20	19.5
Infection	7.95	9.5	28.9	4.41	5.5	31.7	17.7	25	44.4	14.3	21	38	20.7
FW-SAD	7.51	9.5	30.0	9.15	11	48.7	22.0	30	47.3	15.7	25	36.3	24.3

As can be seen, the results in the Table 2 indicate our algorithm is competitive with other existing algorithms. In contrast to the others, the presented algorithms of DLR obtained by using a basic similarity metric. Therefore, the complexity of the algorithm is low and can be easily adapted with any stereo matching system. Our result is the best among all nominated algorithms for the non-occluded region in the Venus dataset, and the second for the Teddy dataset. The scenes of the Venus dataset consist of many textured surfaces, such as the background and printed document. With respect to the evaluations in 'all' sections, our results are moderate since the 'all' region includes occluded regions and the occluded regions mainly consist of planes of background.

Fig. 12 shows the analysis and error evaluation for the non-occluded regions based on bad pixel with (absolute disparity error > 1). The first row of Fig. 12 shown the samples images for evaluation provided by Middlebury Stereo Page. The non-occluded regions visualized by the white areas while the occluded and border regions shown by black. The second row shows the errors for non-occluded regions based on FW-SAD. By comparing the non-occluded regions for the disparity depth map of the proposed algorithm, Fig. 12 (in the third row) visually points where incorrect measurements are produced by the SAD-DLR. We can notice that the number of errors low for the Tsukuba and Venus datasets. The incorrect disparities are higher for the Teddy and Cones datasets due to the complexity and texture regions.

In general, the SAD-DLR improved the disparity maps obtained from the FW-SAD where most of the sparse small black regions (in the second row of Fig. 12) have been removed. One of the disadvantages by using the SAD is the incompetency of the similarity metric to calculate the discontinuity regions. This can be improved by selecting the different cost aggregation method.

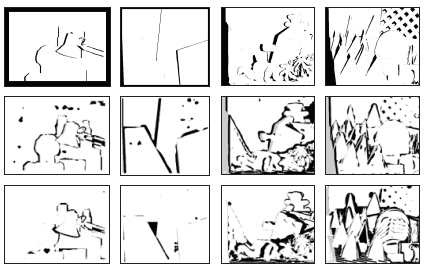


Fig. 12: Analysis for non-occluded region based on bad pixel (absolute disparity error > 1). Non-occluded regions (white) with occluded and border regions (black).

The results obtained have proven to be adequate for the DLR to improve the disparity depth map. Though the DLR does not deal with the cracks and hole due to the layer separations, the merging of disparity and edge boundaries regions change the new disparity maps significantly. The performance of the DLR can be improved by using advanced matching techniques such as graph cut, segmented-matching and dynamic programming, which can produce more accurate disparity depth map. Furthermore, a more sophisticated cost aggregation strategy could lead to better results. However, based on the performance evaluation of DLR with SAD, the results are satisfactory in term of accuracy and quality of the disparity depth maps.

VI. CONCLUSION

The Depth Layer Refinement (DLR) module has been presented aimed to improve the raw disparity maps in the post-processing stage. The proposed system takes advantage of the Depth Image Layers Separation (DILS) algorithm that separate the layers of depth based on disparity range. The resulting disparity maps are evaluated on the Middlebury Stereo Vision website and perform well in comparison to other algorithms although it only using a basic similarity metric of SAD. Qualitative and quantitative evaluation proved the satisfactory quality of the achieved matching results. The proposed method improved up to 13 places from the last place after the basic FW-SAD refined by using DLR in the online evaluation on the Middlebury Stereo Vision website. We found that the proposed technique removes the noise and unmatched pixels on the fixed window searching SAD. It also improved the depth discontinuities of the disparity depth maps.

The limitation of the presented approach lies in the assumption that the scene can be well approximated by a set of rectified images. In the future development, the system can be incorporate with real-time implementation, which can be used with the novel inter-view synthesis algorithm for 3D video and free-viewpoint applications. The proposed algorithm is quite practical for robot navigation and autonomous operations.

ACKNOWLEDGMENT

The authors would like to acknowledge the funding from Ministry of Higher Education of Malaysia and Universiti Teknikal Malaysia Melaka.

REFERENCES

[1] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," International Journal of Computer Vision, vol. 47, pp. 7-42, 2002.

[2] L. D. Stefano and S. Mattoccia, "Real-Time Stereo within the VIDET Project," Elsevier Science Ltd: Real-Time Imaging, vol. 8, pp. 439-453, 2002.

[3] P. Fua, "Combining Stereo and Monocular Information to Compute Dense Depth Maps that Preserve Depth Discontinuities," International Joint Conference on Artificial Intelligence, pp. 1292-1298, August 1991 1991.

[4] L. D. Stefano and S. Mattoccia, "Fast Stereo Matching for the VIDET System using a General Purpose Processor with Multimedia Extensions," presented at the CAMP '00: Proceedings of the Fifth IEEE International Workshop on Computer Architectures for Machine Perception (CAMP'00), Washington, DC, USA, 2000.

[5] F. Tombari, *et.al.*, "Classification and evaluation of cost aggregation methods for stereo correspondence," CVPR, 2008.

[6] L. D. Stefano, *et.al.*, "A PC-Based Real-Time Stereo Vision System," Machine Graphics & Vision, vol. 13, pp. 197-220, 2004.

[7] T. Kanade and M. Okutomi, "A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment," IEEE International Conference on Robotics and Automation, vol. 2, pp. 1088-1095, April 1991 1991.

[8] D. Scharstein, "Stereo Vision for View Synthesis," Proc. of Conference on Computer Vision and Pattern Recognition, pp. 852-858, 1996.

[9] H. Hirschmuller and S. Gehrig, "Stereo Matching in the Presence of Sub-Pixel Calibration Errors," IEEE Conference on Computer Vision and Pattern Recognition, June 2009.

[10] O. Veksler, "Fast Variable Window for Stereo Correspondence using Integral Images," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 556-561, 2003.

[11] S. Birchfield and C. Tomasi, "Depth Discontinuities by Pixel-to-Pixel Stereo," International Journal of Computer Vision, vol. 35, pp. 269-293, 1999.

[12] S. Yoon, *et.al.*, "Fast correlation-based stereo matching with the reduction of systematic errors," Pattern Recognition Letters, vol. 26, pp. 2221 - 2231, 2005.

[13] Y. Boykov and O. Veksler, "Graph Cuts in Vision and Graphics: Theories and Applications.," Handbook of Mathematical Models in Computer Vision, 2006.

[14] V. Kolmogorov, "Graph Based Algorithms for Scene Reconstruction from Two or More Views," PhD thesis, Cornell University, 2003.

[15] J. Sun, *et.al.*, "Image Completion with Structure Propagation," SIGGRAPH, vol. 24, pp. 861-868, 2005.

[16] D. Tzovaras, *et.al.*, "Disparity Field And Depth Map Coding For Multiview 3D Image Generation," Signal Processing: Image Communication, vol. 11, pp. 205-230, Jan 1998 1998.

[17] M. C. Sung, *et.al.*, "Stereo Matching Using Multi-directional Dynamic Programming," presented at the Intelligent Signal Processing and Communications, 2006. ISPACS '06. International Symposium on, 2006.

[18] N. Grammalidis and M. G. Strintzis, "Disparity and occlusion estimation in multiocular systems and their coding for the communication of multiview image sequences," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 8, pp. 328-344, 1998.

[19] A. Fusiello, *et.al.*, "Expreriments with a new Area-Based Stereo Algorithm," ICIAP '97 Proceedings of the 9th International Conference on Image Analysis and Processing, vol. 1, 1997.

[20] A. Fusiello, *et.al.*, "Efficient Stereo with Multiple Windowing," Conference on Computer Vision and Pattern Recognition, pp. 858-863, 1997.

[21] M. Bleyer and M. Gelautz, "A layered stereo matching algorithm using image segmentation and global visibility constraints," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 59, pp. 128-150, 2005.

[22] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp. 603-619, May 2002.

[23] R. J. Fergusson, "Human Visual System Based Object Extraction for Video Coding," PhD, Electronic and Electrical Engineering, University of Strathclyde, Glasgow, 1999.

[24] M. C. Vision. Stereo Evaluation [Online]. Available: http://vision.middlebury.edu/stereo/