International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 14, Number 1 (2021), pp. 1-13


# A New Post-Processing Method for Stereo Matching Algorithm

**Maged Aboali[1], Nurulfajar Abd Manap, Rostam Affendi Hamzah, Abd Majid Darsono, Qazwan Abdullah**

*Machine Learning and Signal Processing (MLSP) Research Group, Center of Telecommunication Research & Innovation (CeTRI), Fakulti Kejuruteraan Elektronik Dan Kejuruteraan Komputer (FKEKK), Universiti Teknikal Malaysia Melaka (UTeM)*

*[1]E-mail: majedaboali199@gmail.com [1]ORCID: 0000-0003-2214-549X*

**Abstract**

Stereo matching is continuing to be a critical and challenging problem due to continuous progression in computer vision including the widely applied in various Artificial Intelligence (AI) applications to echo the human visual system. New developed algorithm approaches with more quality of information extraction and high frame rate of data processing are unique targets to control huge volume data and provide convenient solutions for the massive evolution in this field. Fundamentally, the process of stereo matching involved several stages to implement the disparity map which provides the depth information required in 3D reconstruction. Numerous studies and sophisticated algorithms have been developed in the stereo vision area with different concepts and properties to achieve disparity map implementation. However, their accuracy still low and algorithm structures are complicated which far from current requirements. As a popular breakthrough, post-processing algorithms have proved to achieve outstanding performances in stereo matching in terms of low error rate, less complex algorithm structures, and highly computational in their processing speed. In this paper, we present a novel post-processing framework known as Multistage Hybrid Median Filter (MHMF) with a new segment-based algorithm to surge up the accuracy and maintain low computational complexity. The developed framework consists of two main stages: in Stage 1, the Basic Block Matching (BBM) and Dynamic Programming (DP) are applied to obtain the initial disparity map. While, Stage 2 concerns on segment-based, hybrid median filtering, and merging process for the result of Stage 1 as the main contribution of MHMF. To prove the reliability with current available state-of-the-art algorithms, the experimental results and quantitative measurements on standard indoor and outdoor datasets have been compared. The results demonstrate that the proposed method outperforms many existing methods and works as a powerful approach especially in terms of low accuracy, computational cost, and handling of horizontal stripes.

**Keywords**: Stereo vision, stereo matching, Artificial Intelligence (AI), 3D vision, post-processing algorithms, disparity depth map.

## 1. INTRODUCTION

In recent years, there are a growing interest and attraction in Artificial Intelligence (AI) systems to echo the human visual concept which offers non-contact solutions for modern advancements. Computer vision and computer graphics are interdisciplinary fields that comprise methods and systems with ultimate goals of acquiring, processing, analyzing, and image understanding from videos or digital images. Stereo vision is the main part of them and the process that robustly and automatically convert videos or images data acquired from different approaches such as scene flow, recognition, object detection, semantic segmentation, depth acquisition, three dimensional (3D) reconstruction, mapping, and simultaneous localization into actionable information is called stereo matching [1]–[3].

Our surrounding environment is seen in three-dimensional (3D) scenes through stereo pair images captured by our visual system which genuinely in two-dimension (2D). Figure 1 presents the human vision basic principle of visualizing objects in our surrounding environment. Traditional stereo vision systems similarly capture image pairs in two-dimension (2D) and during the process, the third dimension of the scene is lost. Hence, in stereo vision and for three-dimensional (3D) reconstruction, all information related to depth is vital to be acquired through the raw data of image pairs. Similar to the human eye's concept, binocular stereo cameras can compute the disparity and determine the object's three-dimensional depth through triangulation theory. The analysis of stereo pairs has become a significant research area for three-dimensional (3D) scene model generation and its challenging tasks have motivated researchers in the past decades [4], [5].

Binocular stereo vision (SV) is a general term that addressed a critical research problem that is the three-dimensional (3D) reconstruction of points through images acquired from slightly different horizontal positions toward depth information estimation. The methods of stereo vision attempt to model our complex world based on the processing of multiple images and the reconstruction of their main properties including colors, dimensions, illumination, and shapes. In the image processing area and as a remarkable achievement, many concepts and mathematical methods of stereo vision have been applied to represent our visual environment in such different structures and unique forms. While, it has been exploited in a wide range of applications including 3D scene reconstruction, extraction of 3D information, image-based rendering, parts inspection, robot navigation, and virtual reality [6], [7].

The cameras in the binocular stereo vision (SV) system are coordinated in a parallel manner. For the matching point, the search space in the image can be minimized from a two-dimension (2D) plane to a line through the epipolar constraint since the input stereo pair can be converted to the parallel and horizontal location by applying the epipolar rectification. The disparity of a point in a scene is determined as the shift of pixel position of corresponding pixels from the reference (left image) to target (right image) covering the same scene in stereo vision pairs. In the real scenes, the prime input to define the depth location in three-dimensional (3D) is through identifying the disparity of a point in a scene. The depth map represents the depth at each pixel which is computed through the stereo matching process of image pairs [8]–[10].

Figure 2 shows the basic structure principle of binocular stereo vision (SV). Generally, the basic binocular vision principle consists of two stereo vision cameras 'left binocular camera and right binocular camera' correlated to each other by a horizontal distance known as a baseline. Each binocular camera captures the same object from a slightly different angle. The specific distance between the camera and object can be computed through $D = B \cdot f / d$, where $D$ refers to the distance of an object from the camera, $B$ specifies the baseline distance between the two stereo cameras, and $f$ indicates the focal length of the camera. The target point (real-world point) is perceived from the centers of the two cameras (the optical center $Ol$ and optical center $Or$). It produces two image planes, one left plane "optical plane of the left camera" and one corresponding right plane "optical plane of right camera" from each of the two cameras [11], [12]. Based on the disparity of each pixel, depth estimation is performed. The $d$ disparity is the apparent displacement 'difference in location' of an image of the object between the left (reference) and right (target) images.

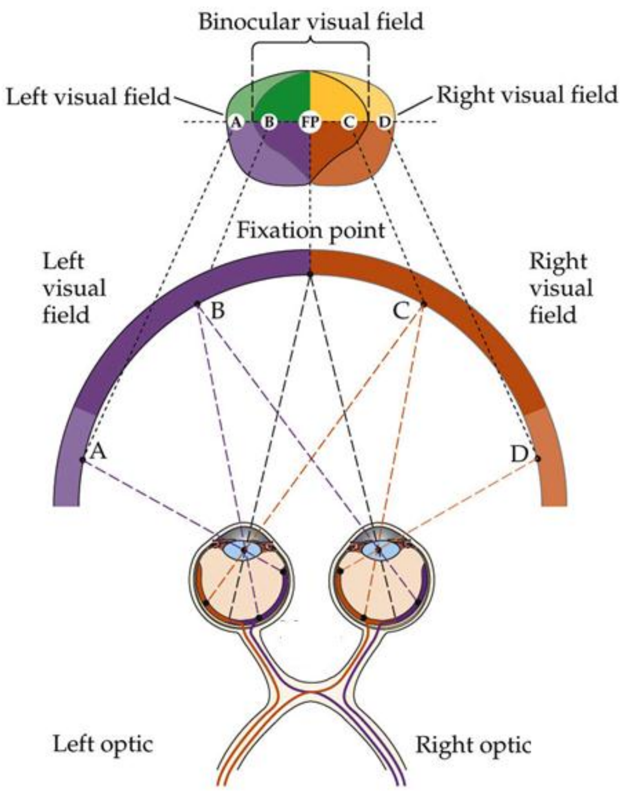



**Figure 1**. Human binocular vision processing principle [13]

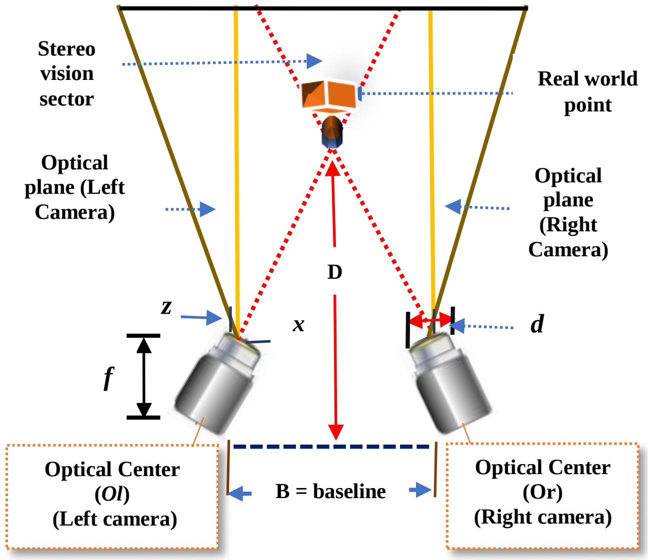



**Figure 2:** Binocular stereo vision (SV) principle basic structure

## 2. BACKGROUND

In general, depth reconstruction is fundamentally classified into passive and active methods. Active three-dimensional (3D) technologies with different performance scenarios of depth reconstruction ranging from the active sensor (e.g. RGB-D camera, three-dimensional LiDAR scanners) to sensing stereo cameras 'depth camera' such as time of flight camera (ToF), stereo camera, and structured-light depth sensors. However, despite enormous progress made in these technologies, most of them still far from performing well in all practical scenarios. Passive methods of depth reconstruction such as stereo matching recover the depth information of two pair images through the disparity of corresponding pixels [14]–[16].

Stereo matching has always been an extensively researched area which plays a crucial role in producing a highly accurate depth image. The process of stereo matching is a type of depth estimation that mainly focuses on the establishment of correspondence between two pixels in stereo image pairs from the same scene captured from slightly different viewpoints to produce a disparity depth map. The depth map is applied for depth estimation based on the triangulation principle which can be embedded as a core part of multiple applications including three-dimensional (3D) tracking, monitoring systems, 3D reconstruction, automotive vehicles, free-view video, Artificial Intelligent (AI**)** applications, robot vision, mobile and navigation system, photogrammetry, virtual reality, [17], [18].

The classification of stereo matching algorithms can be mainly categorized into three main approaches: local methods, global methods, and semi-global methods (SGM) [19]. The categorization is mostly based on the way of disparity map computation. In local methods, the computation of disparity map depends on color/intensity values at a specified given pixel of the support window. These approaches employ the information located near the neighborhood of pixels which are compared with each other. The significance and benefit of local methods rely on their simple structure and fast execution which make them suitable and attractive in practical applications such as real-time implantations. However, drawbacks are mismatches, low accuracy due to narrow limitations, and high noise sensitivity [20].

Global methods focus on the determination of disparities for all pixels of reference images at once. This is carried out by minimizing a predefined energy function by applying an optimization technique. In stereo matching, the energy function has two terms, the correspondence data, and smoothness terms. The correspondence data part is commonly a sum of costs for a given disparity map, depend on some matching cost. The smoothness part is normally a function that relies on the differences in disparity and intensity values of neighboring pixels. In the global methods, the cost aggregation step is generally skipped. Several optimization techniques are used with these methods including Belief Propagation (BP) method [21], [22], Graph Cut (GC) algorithm [23], scanline optimization or Dynamic Programming (DP) [24], [25].These approaches tend to be computationally expensive, but they achieve high accuracy.

Semi-global matching (SGM) algorithm and data-driven (deep learning) are other methods applied to handle stereo-matching issues. Semi-global matching (SGM) uses the local approximation to form matching costs and aggregates steps applying the global cost function. Nevertheless, the data-driven (deep learning) technique and the semi-global matching (SGM) method have been investigated and applied to solve stereo-matching algorithms issues, in recent years. The semi-global method (SGM) is generally to define the correspondence pixels between stereo images. This method applies local approximation to perform matching costs and aggregates using a global cost function across the entire MRF image along linear 1D pixel paths [26].

## 3. STEREO VISION AND STEREO MATCHING ALGORITHM

The stereo vision algorithm is generally built based on three fundamental and main stages, the pre-processing stage, the stereo matching stage, and the post-processing algorithm stage. The pre-processing stage is one of the main critical stages within computer vision. It is used within stereo vision input images to make more reliability of the images for stereo matching. The preprocessing approach includes multiple processes such as noise elimination, deblurring, contrast enhancement, edge detection, and sharpening [26]. The raw disparity map is generated within the stereo matching stage as a key step after the preprocessing. The stereo matching approach is used to compute matching cost at each pixel for all possible disparity levels, where all costs are generally aggregated with the minimum error. The disparity map enhancement comes in the post-processing stage as a significant and highlighted process after the stereo matching algorithm step to improve the quality of the produced disparity map [27].

Based on Scharstein and Szelinski and their comprehensive taxonomy and review stereo matching algorithm involves four highlighted steps: The matching cost step, which generally to match process for every pixel between the left image and the right stereo image. Then, the cost aggregation step, which to aggregate the initial matching costs over the support region. The disparity optimization /computation step (i.e., to select the disparity level that optimized the function). The disparity refinement (i.e., a critical step to precisely refine the final disparity map) [28].

Scharstein and Szelinski revealed in their comprehensive review and taxonomy that, stereo matching algorithms are performed in four steps in general: Matching cost (i.e., to match process for each pixel between left and right stereo images). Cost (support) aggregation (i.e., to aggregate the initial matching costs over the support region). Disparity optimization /computation (i.e., selecting the disparity level that optimized the function). Disparity refinement (i.e., a critical step to precisely refine the final disparity map). Recently, remarkable advances have been recognized in computer vision. The stereo matching algorithm domain is among the fields with high investigation and focus. Numerous studies and researches within stereo matching have been investigated and reported in order to increase the accuracy and decrease the computational process of the produced disparity depth map [29].

An exhaustive evaluation for the methods of stereo matching algorithms has been presented on scheme and taxonomy by Scharstein and Szelinski. In recent years, multiple sophisticated approaches have been implemented and produced on GPU to overcome the computational process drawbacks. These approaches showed remarkable results with high accuracy. However, their processing time for the targeted high-resolution dataset is generally far from the real-time requirement. The stereo matching field continues to attract numerous researchers, developers, and scholars since it particularly poses problems and challenges that are not resolved yet, and due to the continuous progression in computer vision including the wide advances in artificial intelligence applications which echo the human visual system. The field also receives high demands for approaches that perform accurately under multiple operating conditions with less complex algorithm structures.

A general and highly critical stage of disparity map enhancement is the post-processing [30]–[32]. This targeted step of the matching algorithm is required toward smoothing the raw disparity map produced from the optimization/ computation process. Meanwhile, huge fundamental research and studies have been accomplished within the post-processing approach in order to overcome the drawbacks and errors of the produced raw disparity map.

However, most of the proposed and accomplished approaches still suffer from undesirable portions that come with the implemented raw disparity map including occlusions problems, noise drawbacks, textured regions, non-edge preserving, and horizontal streaks. Therefore, in this paper, a new method of stereo matching and post-processing algorithm as a hybrid filtering approach is presented. The proposed Multistage Hybrid Median Filter (MHMF) algorithm is a new post-processing hybrid filtering approach with particular robust algorithms that perform efficiently and accurately in producing the disparity depth map with convenient computation time.

The structure of the proposed Multistage Hybrid Median Filter (MHMF) algorithm mainly consists of two highlighted stages:

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 14, Number 1 (2021), pp. 1-13


In stage 1, both Basic Block Matching (BBM) and the Dynamic Programming (DP) algorithm are applied. The Basic Block Matching (BBM) is particularly used to achieve the matching process and finding the corresponding pixels with Sum Absolute Difference (SAD). Then, the Dynamic Programming (DP) algorithm is applied to achieve the minimum matching cost and operates as an optimization/computation part. However, the raw disparity map from the optimization/computation process came with several drawbacks including horizontal streaks and noises from the inter-scanline process of Dynamic Programming (DP) which certainly affect the quality of the produced disparity depth map [33], [34], [35], Therefore, to minimize and overcome these drawbacks of the implemented raw disparity map, a new proposed and developed post-processing approach is presented and known as Stage 2 of our proposed approach. Stage 2 is the main core of our presented algorithm, it involves in particular a Hybrid Median Filtering (MHF), a segmentation approach, and merging processes. In this post-processing stage, the proposed segmentation part is used to segment the implemented raw disparity depth from stage 1 to multiple segments based on pixel color values for every object. Then, the extracted segments will proceed to the hybrid median filter to remove horizontal streaks and noises. Eventullay, the merged process is followed to generate a new efficient disparity map.

This paper is mainly organized and proceeded as follows: In Section 1, a general introduction and specific review of computer vision, stereo vision, and the stereo matching algorithms are presented. Section 2 focused on background studies and previous research works related to our topic. In Section (3 and 4), the detail and the outline of our proposed approach Multistage Hybrid Median Filter (MHMF) algorithms of the stereo matching are presented including the overall structure and the fundamental stages. Section 5 is particularly discussing with more details, the results of the obtained disparity depth map from our proposed approach. In Section 6, the conclusion of our paper's contents is conducted.

## 4. PROPOSED ALGORITHM

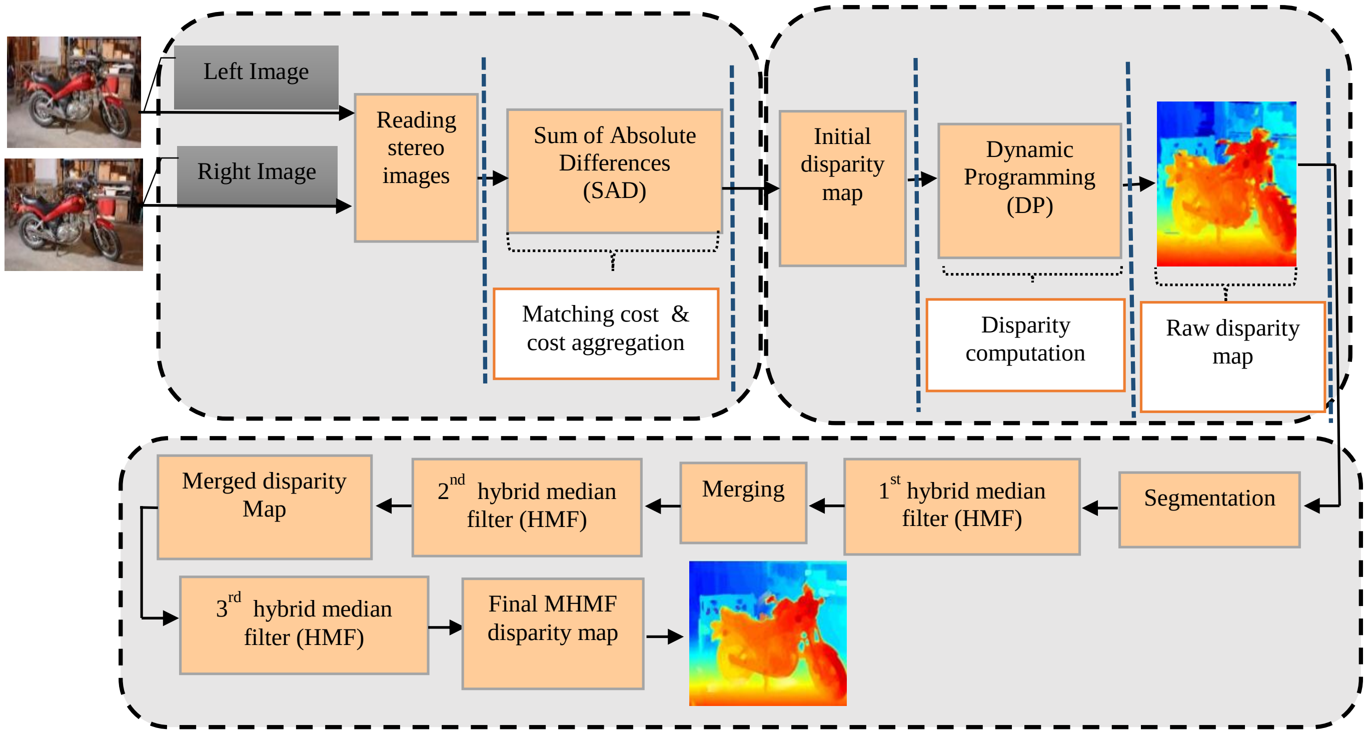



**Figure 3:** Overall structure of the proposed Multistage Hybrid Median Filter (MHMF) algorithm

In section 4, the framework and design of our proposed approach are presented with a complete flowchart description. The fundamental and general algorithm's structure of our proposed (Multistage Hybrid Median Filter) are shown within several main steps based on the taxonomy and fundamental approaches reported by Scharstein and Szeliski [36]. The significant steps and outlines of the proposed algorithms can be discussed within two major stages as follows:

### 4.1 STAGE 1 (Raw Disparity Map Implementation)

In this stage, the raw depth map is computed and implemented within particular and crucial steps as shown in Figure 3 which can be further explained as follows:

#### a) Matching Cost and Cost Aggregation

All disparity map of stereo matching algorithms requires a similarity measure "cost criterion" which compares pixel values in both stereo pair images to find how similar they are to be in correspondence. The step considers a cost metrics "a matching cost computation" to measure the similarity between pixels, where the closeness between reference pixel to a candidate pixel is defined "e.g., how much each pixel under consideration is either similar or dissimilar to the candidate matching pixel". Specifically, the matching cost computation is the step in which the values of two corresponding pixels of the same point in a view are determined.

In this proposed method, in order to define the corresponding pixels between the reference image and the target image, the template is run to produce the disparity map, and we compute the correspondences using the Basic Block Matching (BBM) algorithm based on Sum of Absolute Differences (SAD) function. The search for corresponding pixels is 1-dimensional since a pixel from the reference image and its equivalent in the target image are located on the same row as presented in Figure 4. Thus, the search for all candidate pixels in the left and right images is performed, and the values of two corresponding pixels of the same point in a view are computed.

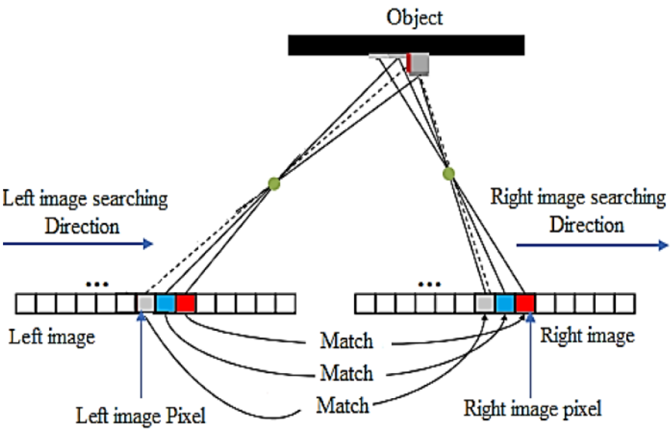



**Figure 4:** Example of the search of correspondence pixels between left and right stereo images

The Sum of Absolute Differences (SAD) is generally used to minimize the matching errors between all the points of the block in our proposed method. The coordinate ($x$, $y$) of the (target image) $I_t$ and the coordinate of the (reference image) $I_{t-1}$ addressed as ($x$+$u$, $y$+$v$), where the $u$ and $v$ are representing the motion vector.

$$\mathrm{SAD}_{(x,y)}(u,v)$$

$$= \sum_{j=0}^{p-1}\sum_{i=0}^{p-1} |I_t(x+i, y+j) - I_{t-1}(x+u+i, y+v+j)| \quad (1)$$

The definition of the (SAD) function is presented in equation (1) [37]. While the (a, b) the equation is defined as follows:

$$(a, b) = \arg\min_{(u,v)\in z} \mathrm{SAD}_{(x,y)}(u, v) \quad (2)$$

Z = {($u$, $v$) | - B ≤ $u$, $v$ ≤ B and ($x$+$u$, $y$+$v$) refer particularly to the valid or preferable position of pixel within the stereo reference image $I_{t-1}$, while B is mainly an integer to find or search for a particular range.

Basic Block Matching (BBM) algorithm is applied in this research as an initial step due to its less computation cost and less complex structure in producing disparity depth map, which made it widely targeted by numerous developers and researchers

to obtain the disparity map [38]. The cost aggregation step is followed to minimize all matching uncertainties of the matching cost. This step is often depend on the window-based which chooses the average above the cost values of disparity space image [39]. The fixed square window is selected among multiple existed methods through our developed algorithms to perform the cost aggregation part since it has less calculation complexity [40]–[42].

The main purpose of this step (cost aggregation) is to reduce the matching uncertainties. It is typically required since the information obtained for a single pixel upon computing the matching cost is not sufficient for precise matching. The step is relying on the window-based that selection of the average above the cost values of Disparity Space Image (DSI) [39]. Many methods have been introduced to perform cost aggregation by using fixed window size. Thus, through this algorithm, the fixed square window of cost aggregation is chosen due to its less calculation complexity and uncomplicated strategy [40]–[42].

### b) Disparity Optimization

Stereo matching algorithms area highlighted some major classification of disparity map optimization as discussed in section 3. Meanwhile, one of the most well-known approaches is the Dynamic Programming (DP) algorithm as an energy minimization framework. This approach is applied by numerous researchers and scholars for disparity depth map enhancement of stereo matching algorithm. The Dynamic Programming (DP) is generally processed and executed for every scan line (row) independently and effectively. Through our developed method, the DP algorithm has a certain role which performs as an optimizer approach to remove all noises on the produced disparity map. The Dynamic Programming (DP) mainly optimizes the energy function for the non-deterministic polynomial-time hard (NP-hard) toward disparity enhancement. The definition of the optimizing equation is represented as shown in Equation (3) [43].

$$E(d) = E_{data}(d) + E_{smooth}(d) \quad (3)$$

generally the term $E_{data}$ *(d)* symbolizes particularly the disparity map function through the correspondence pixels for the disparity depth map *(d)*. While the term $E_{smooth}$ *(d)* refers to the conjecture for smoothness generated from the method that settles the disparity between the pixels on the pixel grid [44]–[47]. One-dimension optimization of Dynamic Programming (DP) algorithm is selected as an optimizer approach of our developed method in order to the energy minimization problems.

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 14, Number 1 (2021), pp. 1-13


### 4.2 STAGE 2 (Disparity Map Enhancement)

This stage represents the main core of this research study. In this vital step (disparity map refinement) there are three main steps involved as the most significant steps which include a new segment-based, a Hybrid Median Filtering (MHF), and the merging process approach. The segmentation approach is particularly used directly to the result of Stage 1 (the raw disparity map) of our proposed Multistage Hybrid Median Filter (MHMF). The raw disparity map of Stage 1 will be segmented into several parts based on the pixel color values. Then, Hybrid Median Filtering (MHF) step will follow, where each extracted segment will be filtered up and remove out the noises. After the hybrid median filtering is done, the merging step will follow to combine all the extracted segments which have been smoothed out through the hybrid median filter process to generate a new disparity map. The focused details and description for each step of disparity depth map refinement on MHMF can be further discussed as follows:

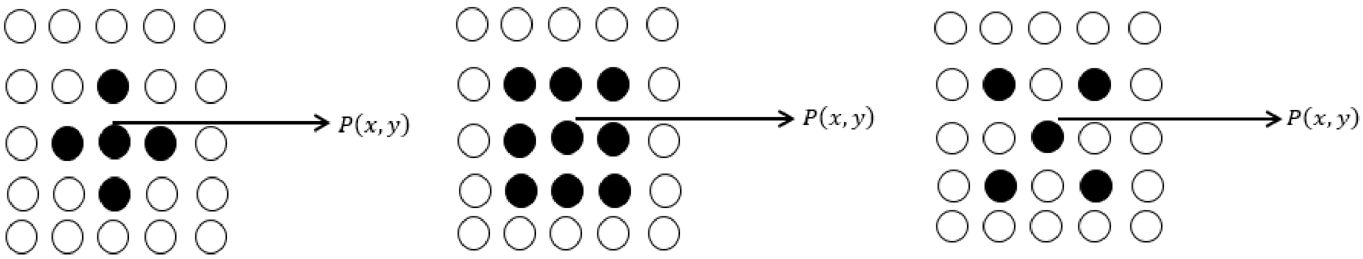
#### (a) Segmentation Approach

The segmentation process is the first step of Stage 2 after the completion of all steps in Stage 1. The raw disparity map obtained from the Dynamic Programming (DP) contains some noises, outliers, errors, and horizontal streaks on the raw disparity map. Thus, in order to solve and overcome these problems, a new segmentation step is developed as a significant part of this research study to extract the raw disparity map obtained from Dynamic Programming. The segmentation process will extract the contents of the raw disparity map into several segments based on the pixel colors of each individual object in the map. Then, each extracted segment from the raw disparity map will proceed with the first Hybrid Median Filter (HMF) to be filtered up.

#### (b) Hybrid Median Filter

In our developed approach, there are three major operations of the hybrid median filtering process is used on each stereo pair to enhance and smooth the extracted segments separately. These operations are applied and represented as shown in Figure 5 (a, b, and c) as follows:

The 4-neighbors of a point $P(x, y)$ are significantly its four horizontal and vertical neighbors $(x \mp 1,\ y)$ and $(x,\ y \mp 1)$. where the point $'p'$ and its 4-neighbors is denoted by $N_4(p)$ as shown in Figure 5 (a). Besides, the 8-neighbors of a point $P(x, y)$ consist of its 4-neighbors together with its four diagonal neighbors $(x + 1,\ y \mp 1)$ and $(x - 1,\ y \mp 1)$ The point $'p'$ and its 8-neighbors is denoted by $N_8(p)$ as shown in Figure 5 (b). While the cross neighbors of a point $P(x, y)$ is significantly consist of its 4-neighbors together with its four diagonal neighbors $(x + 1,\ y \mp 1)$ and $(x - 1,\ y \mp 1)$. The point $'p'$ and its 8-neighbors is denoted by $C_4(p)$ as shown in Figure 5 (c).




(a) (b) (c)

**Figure 5:** The neighbors of a point P

#### (c) Merging Algorithms

Merging algorithm is the process of combining all segmented parts by normal addition to produce a whole merged disparity image. The merging process is applied after the first time of hybrid median filtering on the extracted segments. All extracted segments will go through the hybrid median filter in order to remove out the streaks and noises. Then, the segments will be merged and go through the second hybrid median filter to remove the rest of the errors and noises found on the segment and merged up become a merged disparity map which will go through the third hybrid median filter to further remove the outliers and noises. Eventually, all horizontal streaks, less-textured areas, occlusion, depth discontinuities, and noises are removed and the final generated disparity depth map is smoothed. Figure 6 shows the final result of the disparity depth map by our proposed method (MHMF algorithm) and the merging process. Figure 6 (a) shows the original Motorcycle image from Middlebury Webpage, Figure 6 (b) presents the merged disparity map result, and Figure 6 (c) shows the final disparity map from our proposed MHMF method.

**Figure 6:** (a) Motorcycle Left image, The dis-Range (70)

The size (750×474)

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 14, Number 1 (2021), pp. 1-13

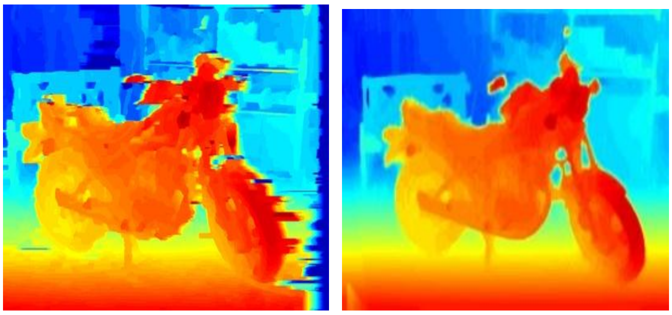



**Figure 6:** (b) Merge disparity map for Motorcycle image, (c) Final depth map for our proposed MHMF method.

## 5 RESULTS AND DISCUSSION

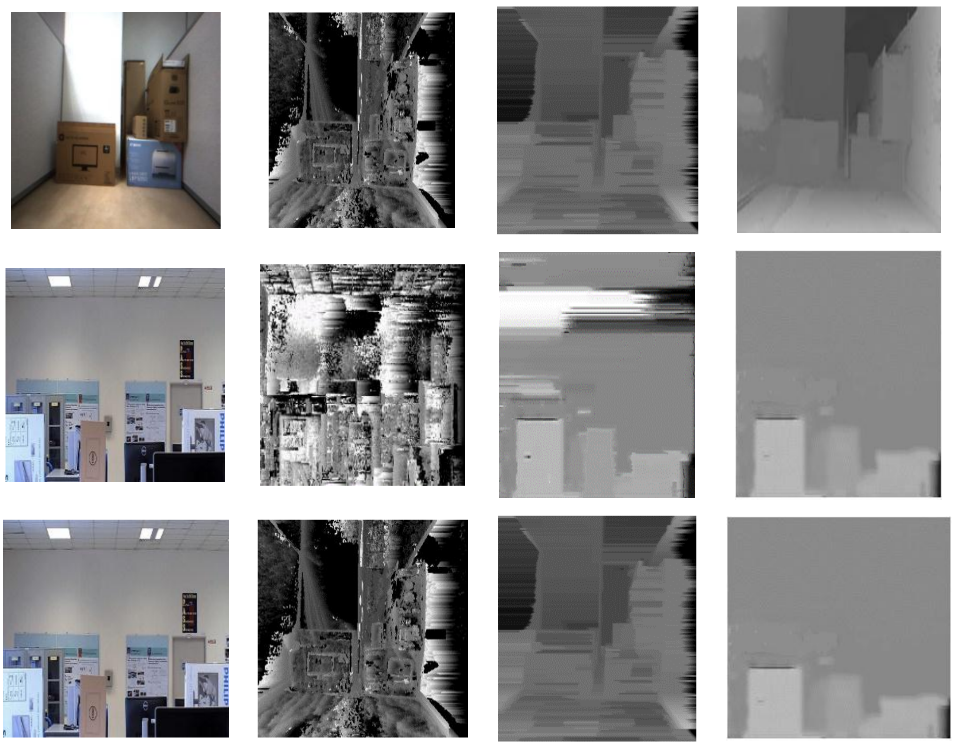
This section presents the results and analysis of the developed stereo matching algorithm: Multistage Hybrid Median Filter (MHMF). The experiments were carried out to evaluate the accuracy, precision, and high performance of our developed method. To verify and show the accuracy of the generated disparity map by our proposed Multistage Hybrid Median Filter (MHMF) and two other highlighting approaches: the Basic Block Matching (BBM) algorithm and the Dynamic Programming (DP) algorithm. All output results of the disparity maps were evaluated within two main directions which are: Subjective evaluation as the first direction and objective evaluation as the second direction to evaluate the performance, accuracy, and high precision.

In the first direction, subjective evaluation is mainly focused on the results of the datasets captured by MV BLUEFOX stereo camera. In this direction, the evaluation is based on the perception of the human's eyes. While, in the second direction evaluation, the objective evaluation is based on three significant functions: Peak Signal to Noise Ratio (PSNR), Mean Squared Error (MSE), and Structural Similarity Index Metric (SSIM) is conducted to investigate all the obtained results.

### 5.1 Subjective Evaluation Direction

Stereo pair images were captured within an indoor environment using MV BLUEFOX stereo cameras to justify and accurately test the high performance of the developed MHMF algorithms experimentally. The MV BLUEFOX stereo camera does not implement the ground truth image, therefore the evaluation of stereo pairs results is based on the perception of human eyes. Multiple characteristics are taken into consideration while captured stereo images including large and small objects, closer location, contrary brightness in the region, closer and further location in order to prove the robustness, precision, and high performance of our proposed method.

**Original Image** **BBM** **DP** **MHMF**



(a) (b) (c) (d)

**Figure 7:** Results for stereo images captured with MV BLUEFOX camera: (a) Original stereo images; (b) Result from BBM; (c) Result from DP; (d) Result from MHMF.

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 14, Number 1 (2021), pp. 1-13


From Figure 7, the first column (a) presents the original images which have been captured by MV BLUEFOX camera with a specified disparity range. The second column (b) presents the disparity map results obtained from (BBM) algorithm approach which contain many errors and unwanted aspects including occlusions, depth discontinuities, and random noises. Figure 6 (c) shows the disparity map results attaining from DP algorithm. Based on the results obtained, DP approach is capable of remove most of the noises in the results of BBM with scanline optimization. However, DP approach is unable to estimate good disparity for the image with color surface and textureless region. Furthermore, DP algorithm causes horizontal streaks on disparity map results. Figure 6 (d) shows the results of the proposed Multistage Hybrid Median Filter (MHMF) algorithm. Where all the noises and horizontal streaks are completely removed. The results show the proposed approach is able to estimate a high accurate disparity depth map on the color surface and textureless regions. The disparity maps obtained are also demonstrating clear detail of objects and edges. The presented results demonstrate clearly that, our proposed method has the capability to produce high-quality depth maps.

**5.2 Objective Evaluation Direction**

In this section, specified four stereo image datasets with provided ground truths are used as input which are: Teddy, Cones, Tsukuba, Venus. All disparity depth map results obtained from our proposed (MHMF) method and the other two highlighted stereo matching methods: BBM algorithm and DP algorithm are analyzed and investigated based on previous input datasets.

According to the results presented in Figure 8, column (a) indicates the ground truths of the image datasets. Column b) presents the results of BBM algorithm, the obtained results include error and undesired aspects including occlusions aspects, random noises, and depth discontinuities. Column (c) shows DP algorithm's result. Based on the gained results, the DP algorithm is able to eliminate the noises that appear in the results from the BBM algorithm with the approach of scanline optimization. However, the DP algorithm has poor disparity estimating for images with color surfaces and textureless areas. In addition, the DP algorithm produces 'streaks' on disparity map results. Column (d) refers to the results obtained by our developed (MHMF) method. Based on the results in Column (d) all the noises and streaks are completely removed. The result of our developed method has high accurate performance and precision in producing disparity depth maps.

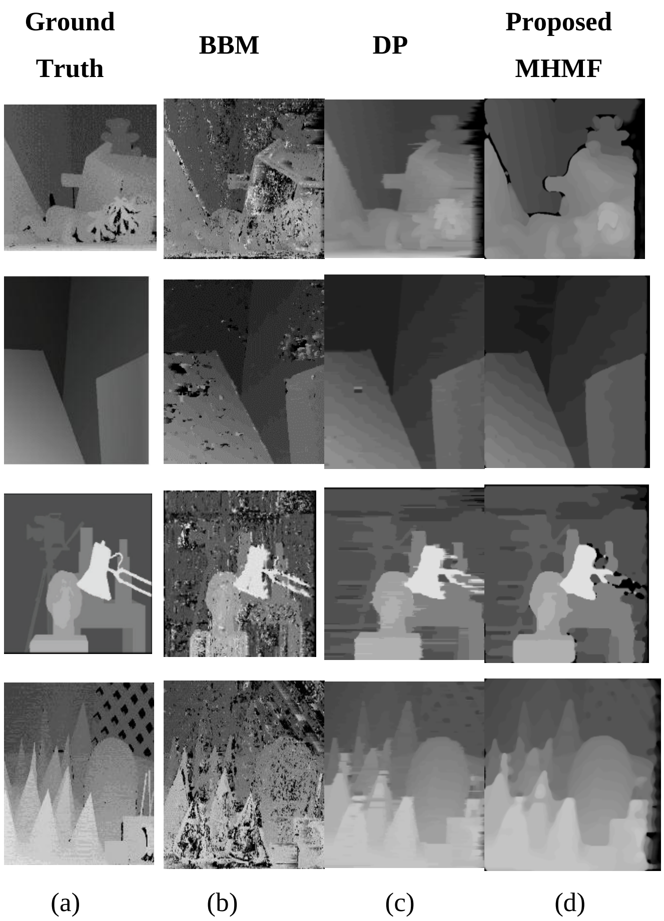



**Figure 8:** Results gained from the three stereo matching algorithms by applying the Middlebury benchmark datasets.

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 14, Number 1 (2021), pp. 1-13


**Table 1:** Results of MSE value for Tsukuba image

| Window size | BBM | DP | MHMF |
|---|---|---|---|
| 5×5 | 287.12 | 214.61 | 212.13 |
| 7×7 | 269.41 | 224.71 | 216.81 |
| 9×9 | 261.21 | 225.32 | 221.43 |
| 11×11 | 256.83 | 227.45 | 223.21 |
| 13×13 | 247.19 | 224.31 | 219.81 |
| 15×15 | 239.31 | 226.73 | 218.94 |

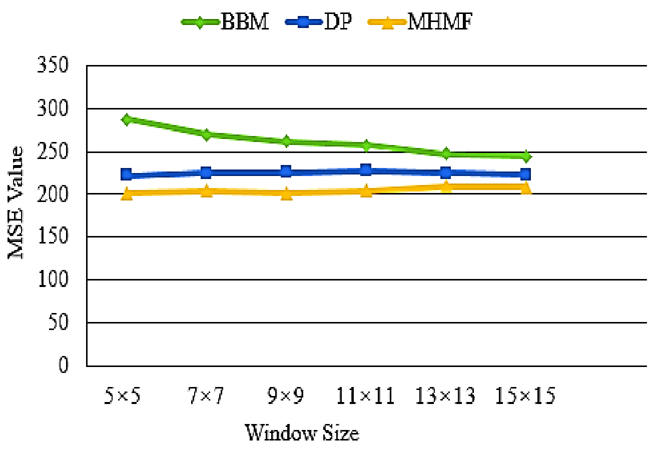



**Figure 9:** MSE graph result for Tsukuba

The disparity map output result based on the MSE function for Tsukuba stereo images is presented in Figure 9 and Table 1. The output results directly change in a proportional pattern based on the increase of window sizes. Based on Figure 9 and Table 1, the MSE values for Tsukuba datasets are gradually decreasing with the extending of the window sizes through Basic Block Matching (BBM) approach, and this indicates that as window size increases, more errors are reduced for Basic Block Matching algorithm because the bigger window sizes are able to provide a better prediction on the minimum value of the matching cost.

The MSE values for Tsukuba datasets through Dynamic programming (DP) is vice versa to the BBM. The MSE values for DP algorithm are increased with the extending of window sizes until a certain window size (window 11×11). Then, the MSE values of DP after window 11×11 is constant. The presented results by DP algorithm show that it performs well when using a smaller window size rather than the bigger window size due to its scanline optimization on every row of pixels. The MSE values for our developed MHMF algorithm specify a standard range with high performance than both basic block-matching and dynamic programming.

**TABLE 2:** Results of PSNR value for Tsukuba image.

| Window size | BBM | DP | MHMF |
|---|---|---|---|
| 5×5 | 12.97 | 14.11 | 14.95 |
| 7×7 | 13.52 | 13.98 | 14.93 |
| 9×9 | 13.56 | 14.43 | 15.30 |
| 11×11 | 13.62 | 14.69 | 15.50 |
| 13×13 | 13.75 | 14.95 | 15.56 |
| 15×15 | 13.85 | 14.82 | 15.00 |

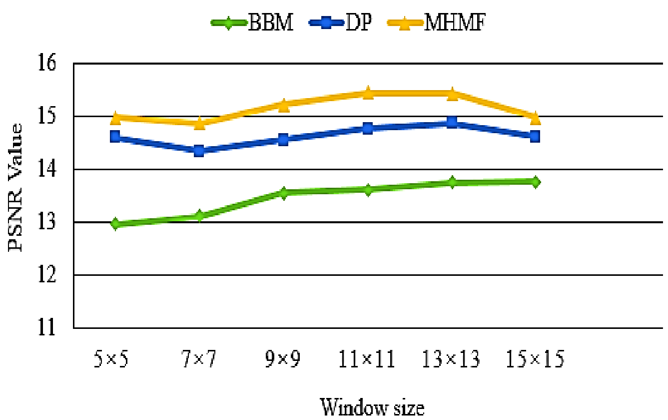



**Figure 10:** PSNR graph result for Tsukuba

Table 2 and Figure 10 present the results of the PSNR gained from BBM, DP, and MHMF algorithms by applying Tsukuba stereo image pairs. According to the obtained results, the PSNR values for BBM algorithm are gradually increasing with extending window sizes which indicates that more noises are minimized with the increase of window size through BBM approach. While the PSNR values through DP algorithm are gradually decreasing and then gradually increasing until a specific window size (window 13×13), where the values decreased again. This indicates that the values gained by DP are not stable, and noises are removed. For our developed MHMF algorithm, the window sizes are slightly affected on the performance of the whole algorithm, the results indicate that the increasing and decreasing is similar with a better performance of MHMF than both BBM and DP approaches. This also shows that the MHMF algorithm is able to operate efficiently with convenient window sizes.

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 14, Number 1 (2021), pp. 1-13


**TABLE 3:** Results of SSIM value for Tsukuba image.

| Window size | BBM | DP | MHMF |
|---|---|---|---|
| 5 × 5 | 0.41 | 0.51 | 0.59 |
| 7 × 7 | 0.44 | 0.51 | 0.58 |
| 9 × 9 | 0.45 | 0.52 | 0.57 |
| 11 × 11 | 0.47 | 0.52 | 0.59 |
| 13 × 13 | 0.46 | 0.52 | 0.59 |
| 15 × 15 | 0.48 | 0.53 | 0.58 |

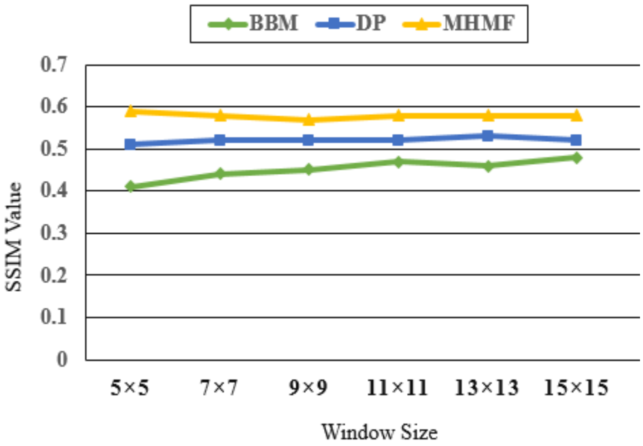



Figure 11: SSIM graph result for Tsukuba

Table 3 and Figure 11 present the SSIM index result of depth maps of Tsukuba datasets for each window size. According to SSIM index results, it is clearly indicated that SSIM values of the Basic Block Matching (BBM) algorithm have a large gap of improvement from smaller to larger window sizes. While, the Dynamic Programming (DP) algorithm, the SSIM values show that, there is a slight improvement on the disparity maps from the smaller to larger window sizes particularly on the background of the image. The SSIM index maps of the developed Multistage Hybrid Median Filter (MHMF) algorithm are the most accurate compared to BBM and DP algorithms, where the difference between the SSIM index maps are close to each other.

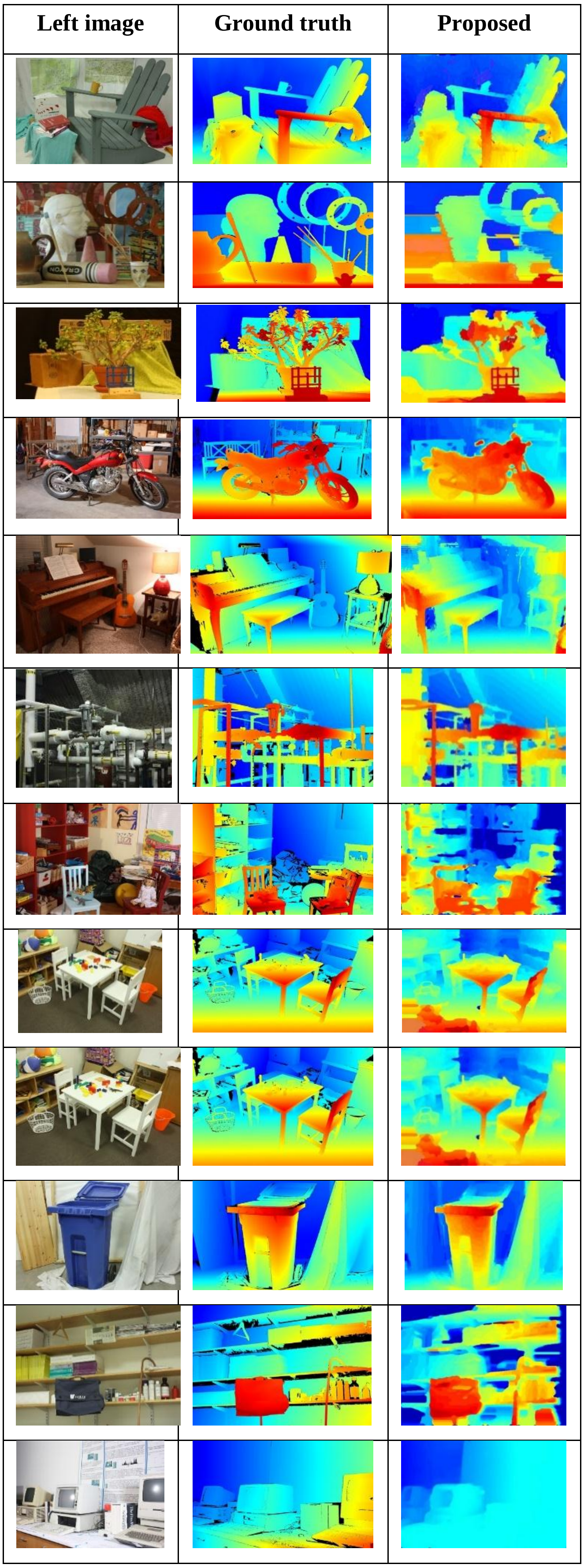



**Figure 12:** Disparity depth map results of the Middlebury training dataset for ground truths and our proposed method

Figure 12 shows the evaluation results from Middlebury Version 3 training datasets, which are based on bad pixels. The first column images are the Middlebury Version 3 training Left images. The second column is the ground truths of the disparity map. The third column shows the proposed Multistage Hybrid Median Filter (MHMF) results. From the observation, the MHMF algorithm is able to obtain accurate results with less errors for the images of Adirondack, Motorcycle, MotorcycleE, Piano, PianoL, Playtable, PlaytableP, and Recycle. There are many incorrect disparities on the datasets of ArtL, Jadeplant, Pipes, Playroom, Shelves, Vintage due to the texture regions and complexity of the content in these datasets. The results show that MHMF algorithm is able to smooth and remove noises and produce highly accurate disparity depth maps. On average, the proposed MHMF algorithm has the potential to overcome the problems and well-performed on each image as well as, the MHMF algorithm is a significant method to overcome the horizontal lines "streaks".

## 6 CONCLUSION

In this paper, we presented a new post-processing method for stereo matching algorithm known as Multistage Hybrid Median Filter (MHMF). The proposed and developed algorithm is a new highlighted hybrid filtering approach with high accuracy to implement a disparity depth map in comparison to the conventional algorithms. The proposed Multistage Hybrid Median Filter (MHMF) method can be applied to mainly improve the accuracy of the disparity depth map and embedded into numerous stereo matching applications. The structure of the developed Multistage Hybrid Median Filter (MHMF) consists of multiple approaches involving the Basic Block Matching (BBM), Dynamic Programming (DP) algorithms, segmentation process, Hybrid Median Filtering (MHF), and merging processes. The results of the produced disparity depth map by the proposed method (MHMF) and other (BBM), (DP) algorithms are evaluated using MSE, PSNR, SSIM functions, and the Middlebury webpage system. According to the overall results, the developed MHMF method is able to produce the disparity depth map with high accuracy, less computational efficiency, and less complexity algorithm structure compare to other existing algorithms. The results obtained demonstrate clearly that, the implemented and produced disparity map by the MHMF algorithm is closer to the ground truth rather than other algorithms. The MHMF method has the ability to eliminate noises, removing horizontal 'streaks' of DP algorithm, and minimizing the depth discontinuities of disparity depth maps. The introduced MHMF algorithm can be embedded as part of numerous stereo matching applications especially with its easy implementation and simple algorithm structure.

## ACKNOWLEDGEMENT

The authors gratefully acknowledge UTeM Zamalah Scheme, Universiti Teknikal Malaysia Melaka (UTeM) and Ministry of Higher Education for supporting this research work.

## REFERENCES

[1] S. Kim, D. Min, S. Kim, and K. Sohn, "Unified confidence estimation networks for robust stereo matching," IEEE Trans. Image Process., vol. 28, no. 3, pp. 1299–1313, 2019.

[2] C. Zhu and Y. Z. Chang, "Efficient Stereo Matching Based on Pervasive Guided Image Filtering," Math. Probl. Eng., 2019.

[3] R. A. Hamzah, M. Y. Wei, and N. S. Anwar, and Y. Nik, "'Development of stereo matching algorithm based on sum of absolute RGB color differences and gradient Matching.,'" Int. J. Electr. Comput. Eng., vol. 10, no. 3, p. 2375, 2020.

[4] C. Won, J. Ryu, and J. Lim, "SweepNet: Wide-baseline omnidirectional depth estimation," Proc. - IEEE Int. Conf. Robot. Autom., no. 10, pp. 6073–6079, 2019.

[5] S. Zhu, H. Xu, and L. Yan, "A Stereo Matching and Depth Map Acquisition Algorithm Based on Deep Learning and Improved Winner Takes All-Dynamic Programming," IEEE Access, vol. 7, pp. 74625–74639, 2019.

[6] C. Ttofis, S. Member, S. Hadjitheophanous, and S. Member, "Edge-Directed Hardware Architecture for Real-Time Disparity Map Computation," IEEE Trans. Comput., vol. 62.4, pp. 690–704, 2013.

[7] V. H. Borisagar and M. A. Zaveri, "Census and Segmentation-Based Disparity Estimation Algorithm Using Region Merging," J. Signal Inf. Process., vol. 6.3, p. 191, 2015.

[8] A. Blake and A. Zisserman, "Visual Reconstruction," Cambridge, MA, USA MIT Press, 1987.

[9] X. Ye, S. Member, Y. Gu, and L. Chen, "Order-Based Disparity Refinement Including Occlusion Handling for Stereo Matching," IEEE Signal Process. Lett., vol. 24, no. 10, pp. 1483–1487, 2017.

[10] T. Yan, Y. Gan, Z. Xia, and Q. Zhao, "Segment-Based Disparity Refinement with Occlusion Handling for Stereo Matching," IEEE Trans. Image Process., vol. 28, no. 8, pp. 3885–3897, 2019.

[11] R. A. Hamzah and H. Ibrahim, "Literature Survey on Stereo Vision Disparity Map Algorithms," J. Sensors, 2016.

[12] G. S. Hong and B. G. Kim, "A local stereo matching algorithm based on weighted guided image filtering for improving the generation of depth range images," Displays, vol. 49, pp. 80–87, 2017.

[13] T. Jessell, S. Siegelbaum, A. J. Hudspeth, and Thomas M. Jessell, "Principles of neural science," Dep. Biochem. Mol. Biophys. New York McGraw-hill, vol. 4, 2000.

[14] A. Bogomjakov, C. Gotsman, and M. Magnor, "Free-viewpoint video from depth cameras," Vision, Model. Vis. Proceedings, Novemb. 22-24 Aachen, Ger., p. 89, 2006.

[15] G. Marin, G. Agresti, L. Minto, and P. Zanuttigh, "A multi-camera dataset for depth estimation in an indoor

scenario," Data Br., vol. 27, p. 104619, 2019.

[16] M. Aboali, N. A. Manap, A. M. Darsono, and Z. M. Yusof, "Review on three dimensional (3-D) acquisition and range imaging techniques," Int. J. Appl. Eng. Res., vol. 12, no. 10, pp. 2409–2421, 2017.

[17] O. Zeglazi, M. Rziza, A. Amine, and C. Demonceaux, "A hierarchical stereo matching algorithm based on adaptive support region aggregation method," Pattern Recognit. Lett., vol. 112, pp. 205–211, 2018.

[18] I. Lee and B. Moon, "An Improved Stereo Matching Algorithm with Robustness to Noise Based on Adaptive Support Weight," J. Inf. Process. Syst., vol. 13.2, pp. 256–267, 2017.

[19] H. Li, Y. Gao, Z. Huang, and Y. Zhang, "Stereo matching based on multi-scale fusion and multi-type support regions," J. Opt. Soc. Am. A, vol. 36, no. 9, p. 1523, 2019.

[20] D. Chen, M. Ardabilian, L. Chen, and S. Member, "A Fast Trilateral Filter-Based Adaptive Support Weight Method for Stereo Matching," IEEE Trans. Circuits Syst. Video Technol., vol. 25.5, pp. 730–743, 2015.

[21] J. Sun, N. Zheng, and S. Member, "Stereo Matching Using Belief Propagation," IEEE Trans. pattern Anal. Mach. Intell. 25.7 787-800. 2003, 2003.

[22] B. Potetz, "Efficient Belief Propagation for Vision Using Linear Constraint Nodes," Comput. Vis. Pattern Recognition, 2007. CVPR'07. IEEE Conf. on. IEEE, 2007.

[23] et al Wang, Hao-qian, "Effective stereo matching using reliable points based graph cut," Vis. Commun. Image Process. (VCIP), 2013, IEEE, 2013.

[24] A. Qayyum, A. S. Malik, M. Naufal, B. Muhammad, F. Abdullah, and M. Iqbal, "Disparity Map Estimation Based on Optimization Algorithms using Satellite Stereo Imagery," Signal Image Process. Appl. (ICSIPA), Int. Conf. on. IEEE, pp. 127–132, 2015.

[25] M. Hallek, H. Boukamcha, F. Smach, and M. Atri, "Real time stereo matching using two step zero-mean SAD and dynamic programing," 2018 15th Int. Multi-Conference Syst. Signals Devices, SSD 2018, pp. 1234–1240, 2018.

[26] Deepa and K. Jyothi, "A robust and efficient pre processing techniques for stereo images," Int. Conf. Electr. Electron. Commun. Comput. Technol. Optim. Tech. ICEECCOT, pp. 89–92, 2018.

[27] K. S. Raviya and D. V. Vyas, "An Evaluation and Improved Matching Cost of Stereo Matching Method," Int. J. Image, Graph. Signal Process., vol. 8.10, pp. 42–52, 2016.

[28] R. Scharstein, D. and Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," nternational J. Comput. Vis., vol. 47.1–3, pp. 7–42, 2001.

[29] L. Nalpantidis, G. C. Sirakoulis, and A. Gasteratos, "A Dense Stereo Correspondence Algorithm for Hardware," Artif. Intell. Theor. Model. Appl., pp. 365–366, 2008.

[30] N. Einecke and J. Eggert, "Anisotropic median filtering for stereo disparity map refinement," VISAPP (2), 2013.

[31] S. Choi, J. Jeong, J. Chang, and H. Shin, "Implementation of Real-Time Post-Processing for High-Quality Stereo Vision," ETRI J., vol. 37.4, pp. 752–765, 2015.

[32] X. Wang, Y. Tian, H. Wang, and Y. Zhang, "Stereo Matching by Filtering-Based Disparity Propagation," PLoS One, vol. 11.9, p. e0162939, 2016.

[33] A. F. Bobick and S. S. Intille, "Large Occlusion Stereo," Int. J. Comput. Vis., vol. 33.3, pp. 181–200, 1999.

[34] N. Cross-correlation, "Robust Stereo Matching Using Adaptive," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, pp. 807–822, 2011.

[35] Y. Wang, C. Tung, and P. Chung, "Efficient Disparity Estimation Using Hierarchical Bilateral Disparity Structure Based Graph Cut Algorithm With a Foreground Boundary Refinement Mechanism," IEEE Trans. Circuits Syst. Video Technol., vol. 23, pp. 784–801, 2013.

[36] D. Scharstein and R. Szeliski, "'A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,'" Int. J. Comput. Vis., vol. 47, pp. 7–42, 2002.

[37] Y. S. Chen, Y. P. Hung, and C. S. Fuh, "Fast block matching algorithm based on the winner-update strategy," IEEE Trans. Image Process., vol. 10, no. 8, pp. 1212-1222.IEEE., 2001.

[38] L. Di Stefano, M. Marchionni, S. Mattoccia, G. Neri, and B. Sede, "A Fast Area-Based Stereo Matching Algorithm," Image Vis. Comput., vol. 22.12, pp. 983–1005, 2004.

[39] F. Tombari, L. Di Stefano, S. Mattoccia, A. Mainetti, and D. Arces, "A 3D Reconstruction System Based on Improved Spacetime Stereo," Control Autom. Robot. Vis. (ICARCV),11th Int. Conf. on. IEEE, 2010.

[40] M. Debella-gilo and A. Kääb, "Locally adaptive template sizes for matching repeat images of Earth surface mass movements," ISPRS J. Photogramm. Remote Sens., vol. 69, pp. 10–28, 2012.

[41] S. Lee, J. Sim, C. Kim, S. Member, and S. Lee, "Correspondence Matching of Multi-View Video Sequences Using Mutual Information Based Similarity Measure," IEEE Trans. Multimed., vol. 15.8, pp. 1719–1731, 2013.

[42] X. Mi, "Stereo Matching based on Global Edge Constraint and Variable Window Propagation," Image Signal Process. (CISP), 5th Int. Congr. on. IEEE, 2012.

[43] L. L. Zhu, Y. Chen, Y. Lin, C. Lin, and A. Yuille, "Recursive Segmentation and Recognition Templates for 2D Parsing," Adv. Neural Inf. Process. Syst., 2009.

[44] M. Bleyer and M. Gelautz, "A layered stereo matching algorithm using image segmentation and global visibility constraints," ISPRS J. Photogramm. Remote Sens., vol. 59, no. 3, pp. 128–150, 2005.

[45] R. Gouiaa and J. Meunier, "3D reconstruction by

fusioning shadow and silhouette information," Proc. - Conf. Comput. Robot Vision, CRV 2014, pp. 378–384, 2014.

[46] D. Neilson and Y. H. Yang, "A component-wise analysis of constructible match cost functions for global stereopsis," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 11, pp. 2147–2159, 2011.

[47] F. Tombari, S. Mattoccia, L. Di Stefano, and E. Addimanda, "Classification and evaluation of cost aggregation methods for stereo correspondence," 26th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR, 2008.