STEREO MATCHING ALGORITHM BASED ON ILLUMINATION CONTROL TO IMPROVE THE ACCURACY

ROSTAM AFFENDI HAMZAH,†1,2, HAIDI IBRAHIM1 AND ANWAR HASNI ABU HASSAN1

†School of Electrical & Electronic Engineering, Engineering Campus, Universiti Sains Malaysia, 14300 Nibong Tebal, Penang, Malaysia.; †Faculty of Engineering Technology, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Melaka, Malaysia
e-mail: rostamaffendi@utem.edu.my, haidi_ibrahim@ieee.org, eeanwar@usm.my
(Received July 14, 2015; revised December 11, 2015; accepted January 9, 2016)

ABSTRACT

This paper presents a new method of pixel based stereo matching algorithm using illumination control. The state of the art algorithm for absolute difference (AD) works fast, but only precise at low texture areas. Besides, it is sensitive to radiometric distortions (i.e., contrast or brightness) and discontinuity areas. To overcome the problem, this paper proposes an algorithm that utilizes an illumination control to enhance the image quality of absolute difference (AD) matching. Thus, pixel intensities at this step are more consistent, especially at the object boundaries. Then, the gradient difference value is added to empower the reduction of the radiometric errors. The gradient characteristics are known for its robustness with regard to the radiometric errors. The experimental results demonstrate that the proposed algorithm performs much better when using a standard benchmarking dataset from the Middlebury Stereo Vision dataset. The main contribution of this work is a reduction of discontinuity errors that leads to a significant enhancement on matching quality and accuracy of disparity maps.

Keywords: computer vision, digital image processing, disparity map, gradient matching, pixel-based matching, stereo matching algorithm.

INTRODUCTION

In recent years there have been various progresses in the field of vision and image processing. These researches are useful for commercial and scientific purposes in many fields. One of the great advances pertaining to these aspects is the extraction of three dimensional (3D) information from two dimensional (2D) data. Many research groups have studied this discipline in depth, gaining mechanism for 3D illustration by using more than one camera. Trying to resemble the vision of human beings, the researchers have investigated the camera-based system inspired by human vision (Dominguez-Morales et al., 2011). Following this, the stereo vision is established. It uses two parallel digital cameras to acquire the depth of a scene. The benefits of using the stereo camera are high resolution, low price and the images can be used for other application as well (Humenberger et al., 2010). For robot applications such as people or scene recognition and robot navigation, the stereo vision camera is used. Schmid et al. (2013) used a stereo camera as the main sensor for autonomous navigation and obstacle avoidance on indoor and outdoor flying robot. The task of the sensor is to continuously track the state of the visualized environment in order to provide visual information of the path planning and decision modules of the robot. This is to adapt the movement of the robotic system in tandem with to the state variations appearing in the imaged scene. This sensor provides 3D data perception of the surrounding environment for on-board planning, frontier-based exploration and wall following. Another interesting function of stereo vision application is the visual support system for blind navigation. Fernandes et al. (2010) developed a prototype system to acquire a 3D structure representing the environment in front of the user. Then, the object recognition is executed. The results are converted and provided to the user via a 3D virtual sound for navigation. The loudness of the sound is proportional to the intensity of the pixel. The usage of the stereo camera is relatively suitable because it is a purely passive technology and does not affect the surrounding environment.

In the field of stereo vision, stereo matching is a process of establishing the correspondence between one pair of the images. The matching process which is the search for corresponding predictions of the same scene point onto both camera planes has to be solved. The problems associated with developing image matching are the computational cost required to achieve the appropriate results (Humenberger et al., 2010). The result of the matching process is presented in a disparity map. This map provides the depth data which is important to 3D image reconstruction. The disparity map estimation comprises of finding the
correspondence for each pixel pair from two images at a designated coordinates (i.e., reference image coordinates). According to Scharstein and Szeliski (2002), most of the stereo matching algorithms rely on four steps. The four steps are explained below:

**Step 1:** Matching cost computation (i.e., matching process at each pixel from left and right images).

**Step 2:** Cost aggregation (i.e., aggregate initial costs over support region).

**Step 3:** Disparity optimization (i.e., select the disparity level that optimized the function).

**Step 4:** Disparity refinement (i.e., post-processing to refine the final disparity map). The algorithm of disparity map development can be classified as either local or global methods. This classification depends on the technique in which the disparity map is calculated (Scharstein and Szeliski, 2002).

Local methods determine the disparity using the correspondence between the gray colors, values or texture patterns within a given local support window. The local method applies a small number of pixels around the pixel of interest. These methods are also referred as window-based or area-based methods (Hamzah and Ibrahim, 2016). There are several approaches related to window-based such as fixed window (Yang et al., 2014), multiple window (Hirschmuller et al., 2002) and adaptive window (Lu et al., 2008). Such methods use only local information. Therefore, they have a low computational requirement and a short runtime. The disparity map is determined by selecting the smallest matching cost value from disparity candidates. This selective approach is commonly known as the winner take all (WTA) approach. Although the local method can produce disparity maps quickly, but the precision is low, especially in depth discontinuity regions.

On the other hand, global optimization methods perceive the disparity assignment problem as a problem of minimizing a predefined global energy function. This method applies energy minimization in lines with the entire image itself. They are usually less sensitive to local individualities and tend to be computationally expensive. The measurement is taken from the global data and an additional smooth constraint for neighboring pixels (Wang et al., 2013a). The smooth constraint is utilized in order to preserve the disparity smoothness of the pixels of the same region while simultaneously refines the object boundaries. Numerous methods for solving the global energy minimization problem by using a graph from Markov random field (MRF), have been proposed. These methods can be categorized as either graph cut (GC) method (Wang et al., 2013b) or belief propagation method (Xiang et al., 2012). The GC method obtains the minimal energy solution by applying a minimum cut and max-flow algorithm to the energy flow structure extracted from the MRF graph. In contrast, the belief propagation method minimizes the energy function by iteratively passing the messages from the current node to neighboring nodes in the MRF graph. Global methods can obtain high accuracy, but unfortunately they have a high computational requirement.

**CONTRIBUTION**

In this work, the survey of the main challenge of existing methods is conducted based on the steps of algorithm development. By means of this survey (Hamzah and Ibrahim, 2016), the crucial part is to develop the preliminary of disparity map which contributes to the overall performance of disparity map accuracy. The preliminary of disparity map is produced at the matching cost computation process (i.e., Step 1). Based on this finding, this work proposes a new approach of matching cost function to increase the overall accuracy. The new proposed local method algorithm uses a combination of two different similarity measurements which consist of gradient-based and absolute difference (AD) approaches. The AD algorithm is imposed with an illumination control to enhance the quality of the preliminary disparity map. The aggregation step is implemented by using a guided filter as proposed by He et al. (2013). The implementation of post-processing steps use left-right cross consistency checking, fill-in the invalid pixels and a weighted bilateral filter. Based on the proposed algorithm, the accuracy of the disparity maps at low texture and discontinuity areas are improved with the implementation of illumination control. The computational complexity can also be reduced if compared to window-based approaches at the matching costs step (i.e., Step 1).

The rest of this paper is organized as follows. The next section will provide a detailed discussion of previous works on local stereo matching algorithms. Then, the subsequent section will explain the structure of the proposed stereo matching algorithm. This is followed by the section for experimental arrangements and the results. The conclusion and future works from the proposed stereo matching algorithm are provided in the last section.

**STEREO VISION GEOMETRY**

Basically, stereo vision can be illustrated by using triangulation analysis. This analysis computes the 3D position of points in the images. It gives the disparity
selection and depth estimation from the geometry of stereo setting as given by Eqs. 1, 2.

\[ d = x_l - x_r, \quad (1) \]
\[ z = \frac{bf}{d}, \quad (2) \]

where \( d \) represents the disparity value while \( x_l \) and \( x_r \) represent the coordinates of left and right image planes respectively. The \( z \) is the depth of a target point, \( b \) is the baseline and \( f \) denotes the focal length of the stereo camera. The triangulation model from the representation of \( P \) on the image plane is shown by Fig. 1.

**Fig. 1: Two pinhole camera models (i.e., left (L) and right (R)) with parallel optical axes.**

**RELATED WORKS ON STEREO MATCHING ALGORITHMS**

Recently, many local methods have been developed to obtain a disparity map. There are some similarities in cost matching function through the use of window-based techniques. They are the sum of absolute differences (SAD) (Tippett et al., 2011), the sum of squared differences (SSD; Yang and Pollefeys, 2003) and the normalized cross correlation (NCC; Satoh, 2011). Observed from all of these works, the main problem of window based technique is that it normally assumes that the pixels within the support region have the same disparity. This is not certainly valid for pixels near depth discontinuities or edges. Hence, incorrect selection of the size and shape of the matching window leads to poor disparity value estimations. In the work by Gupta and Cho (2010), the adaptive binary window was used to build the matching cost. They formed a variable shape adaptive binary window by marking all positions into the support window as a one region. Their approach increases the mismatch pixels at the boundaries and textureless areas. This happens due to wrong pixel assumptions in window-based matching technique if there are boundaries or untextured areas detected. In the work by Hu et al. (2011), the virtual support window was introduced, which developed a more complex solution to resize the windows. Another complex solution is proposed by Einecke and Eggert (2013), who implemented a modification of normalized cross correlation (NCC) algorithm at the matching stage. This method produced the same problem as the implementation by Gupta and Cho (2010). The window-based matching approach tends to produce an incorrect matching approximations of the pixel value at every support regions calculated especially at the boundaries.

Sharma et al. (2011) and Wang et al. (2014) introduced a method which used a feature based matching technique. Sharma et al. (2011) developed a matching technique using the scale invariant feature transform (SIFT) algorithm. They adapted SIFT algorithm by using self-organizing mapping to perform a more efficient feature matching. Wang et al. (2014) implemented feature-based matching through two different expansion phases. The first phase used image segmentation process to obtain the sparse points of matching. Then, the second phase used a regular seed-growing algorithm was used to produce a quasi-dense disparity maps. From their results, the feature-based matching accuracy was low because this approach only targeted the object features and less sensitive to occlusion and textureless areas. Another approaches for pixel-based matching technique were shown by the works of Samadi and Othman (2013), Jung et al. (2014) and Ma et al. (2013a). Samadi and Othman (2013) stated that the census based pixel matching was implemented with the reduction of bits on the comparison numbers. Their modifications increase the speed and accuracy which have been applied to the mobile robot navigation. Jung et al. (2014) used a census filter based on higher order matching cost with an assumption that the patch of all disparities is at the same values. Their work produce a better quality disparity map. Ma et al. (2013a) implemented census transform based on neighbourhood information. They have utilised more bits to represent the difference between the pixel and its neighbour. However, their results showed that the discontinuity regions still produce high errors.

The essential property of the disparity maps that support high quality processing is the accuracy of the maps at the object boundaries and depth discontinuities. Generally, local methods do not achieve high accuracy in those regions due to inappropriate selection of window sizes. However, in previous years, at the cost aggregation stage, the adaptive support weight (ASW) approaches have gained consideration due to high quality disparity map approximation as shown by Yoon and Kweon.
This approach delivers weighting between the support window-based on the intensity similarity. Therefore, during the aggregation process of each pixel, this approach marks the neighbouring pixels which have closer intensities as it will exert a greater weight (i.e., higher support). This method produces a constant weight sharing for smooth regions, thus increases robustness against noise which helps to preserve the edge features. Therefore, some approximations over ASW have been proposed in order to increase their effectiveness. The most well-organized approach that use ASW was proposed in (Yang et al., 2009). They applied a constant time complexity of bilateral filter in cost aggregation stage. Through their approach, a piecewise linear discretization of intensity levels was implemented to reach the weighted average within a support window. They produced good results but with high computational complexity.

However, the support weight methods require independent calculations for each pixel in a neighbourhood that increases computational complexity and a huge amount of memory. Moreover, in the work of Kowalczyk et al. (2013), the two-stage adaptive support weights (ASW) algorithm was used, with a combination of window-based cost aggregation and iterative refinement technique. Their work produced high accuracy. However, the complexity of their work was high as it used ASW and iterative approaches in the same algorithm. ASW approach requires high computational complexity to adaptively determine the shapes or textures. Furthermore, the iterative refinement technique comprises multiple loops which involve complex programming technique to refine the mismatch pixels. In the work by Liu et al. (2012), the algorithm was implemented using a multi-scale Weber (MSW) and weighted linear regression bilateral filter. The MSW descriptor was utilized to combine raw matching cost in two layers of structures. The matching cost was aggregated by spatial moving discrete sampling and weighted least squares. Through this technique, the implementation of bilateral in aggregation stage managed to produce high complexity and resulted in slow processing time.

Usually, disparity map refinement methods reach an effective refinement by manipulating the values near the neighbouring pixels of interest (Gu et al., 2008; Min et al., 2012). Two established local disparity refinement techniques are median filtering and Gaussian convolution. Median filter is able to get rid of small, isolated mismatch disparities through its edge preserving characteristic. This filter picks the middle value within the window pixels to obtain the final result of the center pixel. In the work by Michael et al. (2013), the median filter was used in refinement stage for real time applications. Ma et al. (2013b) used a constant time weighted technique at refinement stage. Their results demonstrated high accuracy in removing noise and errors with respect to the edge of depth map produced. Another approach uses a Gaussian convolution, which combines disparity estimation with those of its neighbours, according to weights defined by a Gaussian distribution. This method decreases the noise in the disparity map. Unfortunately, it decreases the amount of fine detail existing in the final disparity maps too. An effort was made by Vijayanagar et al. (2013) who imposed weights on the Gaussian filter, approximating the lost depth values by using the nearby good depth pixels as a guide to prevent the filtering across the object boundaries.

MATERIAL AND METHODS

THE PROPOSED STEREO MATCHING ALGORITHM

Similar to standard local stereo matching approaches, the proposed development of an algorithm involves the steps as shown in Fig. 2. The matching cost computation uses the gradient and absolute difference values with an illumination control. The cost aggregation is implemented by using a guided filter. Then, the disparity optimization using winner take all (WTA) strategy. The WTA strategy absorbs the minimal aggregated corresponding values for each valid pixel. At this stage, the invalid pixels still occurred especially at the occlusion and untextured areas. The left-right (L-R) cross consistency checking process is applied to determine those invalid pixel areas. Then, the process of the invalid pixels fill-in takes place. In this work, the invalid pixels are replaced with a valid minimum pixel value detected by the previous process. The final stage consists of implementing the weighted bilateral filter to remove the remaining noise which usually occurs in the fill-in pixel processes.

MATCHING COST COMPUTATION

In this sub-section, the new proposed pixel-based matching stage uses a combination of gradient-based similarity measurement and absolute differences (AD) with an illumination control. To calculate the components of the gradient for each image, the gradient value of $G_x$ and $G_y$ masks are used as implemented in De-Maeztu et al. (2011) and von Gioi et al. (2012). Fundamentally, the gradient process
Using both gradient components, the magnitude $I_m$ and phase $\phi$ are computed using Eqs. 5,6 respectively.

$$m = \sqrt{G^2_x + G^2_y},$$

$$\phi = \arctan \left( \frac{G_y}{G_x} \right).$$

In the $x$-direction of gradient displacement, the gradient matching cost $CG'(p, d)$ is given by Eq. 7.

$$CG'(p, d) = |\nabla_x (m_l(p)) - \nabla_x (m_r(p-d))|,$$

where the coordinates pixel of interest $(x, y)$ are represented by $p$, and $d$ is the disparity value, while $m_l$ and $m_r$ are the magnitude of grayscale image gradient operator that applied to the left and right respectively. The final value of the gradient difference $CG(p, d)$ depends on the value which do not exceed the truncated value at $\tau_{CG}$. This is given by Eq. 8;

$$CG(p, d) = \begin{cases} \tau_{CG}, & \text{if } |CG'(p, d)| > \tau_{CG} , \\ CG'(p, d), & \text{otherwise} . \end{cases}$$

Meanwhile, the absolute difference algorithm $A_F$ is given by the Eq. 9 which relies on the intensity difference between two pixels at the left image $I_l$ and right image $I_r$.

$$A_F(p, d) = |\beta(I_l(p) - I_r(p-d))|. $$

In this work, the illumination control $\beta$ is imposed on the $A_F$ algorithm. This new control parameter will enhance the image features on the preliminary of the disparity map. The given Eq. 10 is the condition when the truncated value of $\tau_{AD}$ is applied as implemented by Yoon and Kweon (2006) in order to increase the robustness against the outliers.

$$AD(p, d) = \begin{cases} \tau_{AD}, & \text{if } |A_F(p, d)| > \tau_{AD} , \\ |A_F(p, d)|, & \text{otherwise} . \end{cases}$$

The final cost function $M(p, d)$ at this stage is the combination of $AD(p, d)$ and $CG(p, d)$ as given by Eq. 11.

$$M(p, d) = \alpha AD(p, d) + (1 - \alpha) CG(p, d),$$

where the $\alpha$ is added to balance the color and gradient terms as implemented by Yang et al. (2014). The value of $\alpha$ controls the sensitivity of radiometric differences. However, in this work, the weighted of $\alpha$ value is added to $AD(p, d)$ instead of $CG(p, d)$.

**COST AGGREGATION**

Cost aggregation is the most critical stage, which minimize the matching uncertainties. It produces the overall performance of the disparity maps for local methods. From cost matching step, the raw disparity values are vast and too sensitive to noise. In this work, the guided filter is chosen since it is designed to reduce the noise and preserve the edges (Yang et al., 2014). The guided means that the filter is using the selected guided imaging, (i.e., left or right grayscale image as a guide for the filtering process). The left image is selected in this work as a reference and guidance for the process of filtering. The filter kernel of the weighted guided filter is given by Eq. 12.

$$G_{p,q}(l) = \frac{1}{|W|^2} \sum_{q \in W_k} \left( 1 + \frac{(I_p - \mu_k)(I_q - \mu_k)}{\sigma_k^2 + \varepsilon} \right).$$

$I$ is the guidance grayscale image and $p$ represents the $(x,y)$ coordinates pixel of interest within a support window on the same image. $q$ shows the neighbouring pixels in the support region with the size of $(r \times r)$. The $\sigma$ and $\mu$ are the intensity values of variance and mean in a squared window of $w_k$, which is centered at the pixel $k$ on the guidance image. $w$ is the number of pixels in a square window of $w_k$. The $\varepsilon$ value represents the control element for smoothness.
The aggregation cost volume $CA(p,d)$ at this stage is given by Eq. 13,

$$CA(p,d) = G_{p,q}(I)M(p,d), \quad (13)$$

where $G_{p,q}(I)$ is the weight of the guided filter and $M(p,d)$ represents the filtering input image. The weight of the edge preserving factor is determined by the sum of neighbouring pixels of $q$ in Eq. 12 within the support region of the guidance image.

**DISPARITY OPTIMIZATION**

To obtain the accurate disparity map, this work computes the final disparity by selecting the minimal aggregated corresponding value for each pixel using WTA strategy. The utilization of WTA strategy for local algorithms is able to reduce the computational complexity such as those implemented by Huang and Wang (2010); Zhang et al. (2011). However, through their findings, the disparity maps attained at this stage still consist of errors in the unmatched pixels or occluded regions. Given Eq. 14, the disparity associated with the minimum aggregated cost $d_p$ at each pixel is chosen. $CA(p,d)$ means the cost aggregation volume acquired after the process of cost aggregation and $D$ represents the set of all valid and allowed discrete disparity values,

$$d_p = \arg\min_{d \in D} CA(p,d). \quad (14)$$

**POST PROCESSING**

The final stage of the proposed algorithm consists of the post processing step. This step comprises three sequential processes which are occlusion handling or invalid pixels detection, fill-in the invalid pixels and filtering. The occlusion or invalid areas are detected by the left-right (L-R) cross consistency checking process. The task of this process is to find out the invalid pixels in the disparity map. This process performs from left reference disparity map image coincides with the right reference disparity map. The result is the invalid and rejected pixels due to some flat regions and occluded areas of the scene. Next, the fill-in invalid pixels process takes place. In this current work, since the left image is a reference, the filling process starts with the right to the left valid pixel replacement. The invalid pixel is replaced with the nearest valid pixel value. The valid pixel must be located on the same scan line or at the starting scan line as shown by Eq. 15.

$$d(x) = \begin{cases} d(x-1), & \text{if } d(x) = 0, \\ d(x), & \text{otherwise,} \end{cases} \quad (15)$$

where $d(x)$ is a pixel with an intensity value and $x$ represents the location of the pixel. However, this filling and replacing process will produce the unwanted streak artefacts in the disparity maps. To remove that noise, the weighted bilateral filter is utilized as given by Eq. 16. This filter is an edge-preserving filter and is able to improve the disparity map quality.

$$WM_{p,q}^{BF} = \exp \left( -\frac{|p-q|^2}{\sigma^2} \right) \exp \left( -\frac{|I_p-I_q|^2}{\sigma^2} \right), \quad (16)$$

where $p$ is a pixel needs to be denoised using the weight of the neighbouring pixel of $q$. $\sigma$ represents a spatial adjustment parameter and $\sigma_c$ corresponds to the color similarity parameter. The $p-q$ refer to spatial Euclidean distance and $|I_p-I_q|$ is the Euclidean distance in color space. This filter applies a higher weight to pixels that are spatially close and have a similar color according to the sigma adjustment (Tan and Monasse, 2014). The summation of histogram $h$ is calculated from Eq. 17 where each value is weighted from Eq. 16,

$$h(d) = \sum_{q \in w_p} WM_{p,q}^{BF}, \quad (17)$$

produce the results of Eq. 18;

$$d \in [d_{\min},d_{\max}], \quad (18)$$

where the $w_p$ is the window size with the radius of $(r \times r)$ at the centred pixel of the $p$. The $d_{\min}$ is the minimum disparity value and $d_{\max}$ denotes the maximum disparity value. The final disparity value is determined by the median value of $h(d)$ from Eq. 19.

$$d(p) = \min \left\{ d \mid h(d) \geq \frac{1}{2}h(d_{\max}) \right\}. \quad (19)$$

**RESULTS**

In this work, the experimental images and the results use a standard benchmarking dataset that have been widely used by researchers from the Middlebury Stereo Vision database (Scharstein and Szeliski, 2015). Fig. 3 shows the input images which consist of Tsukuba, Venus, Teddy and Cones with the left scene reference, ground truth images and the frame size. According to (Scharstein and Szeliski, 2015), the accuracy levels of an image are measured from bad pixels percentage in all pixels in non-occluded regions (nonocc), all pixels detected with valid pixels (all) and pixels in regions near depth discontinuities and occluded regions (disc). The experiments in this work are carried out on the platform of Window 8.1 on a desktop PC with a 3.2GHz processor and 8GB
memory. The parameters in this work are explained as follows:

_Step 1:_ The value of $\beta$ is determined through the simulation of a test image (i.e., in this work, cones image is selected). As for the visualization of the illumination difference, Fig. 4 shows the slice of the cones image in two dimensional. In this work, the $\beta$ is selected at about 0.1. The image produced at this value is more reliable in terms of its brightness output compared to the other images. The $\tau_{CG}$ equals to 0.00855 and $\tau_{AD}$ is 0.02155. The constant value of $\alpha$ is 0.18.

_Step 2:_ The selection of a filter at this step is based on the experimental analysis of several established methods or filters. They are guided filter (GF) (He et al., 2013), non-local (NL)(Yang, 2012b), adaptive support weight (ASW) (Hosni et al., 2013), segment tree (ST)(Mei et al., 2013), recursive bilateral filter (RBF) (Yang, 2012a), median (MD) and box (BX) filters. The results are shown in Table 1. Based on this experiments, the GF is selected due to the lowest average error it produces. Since the cost aggregation step is important for the local method algorithms, this paper presents an analysis of the relationship of $\epsilon$ and filter window size ($r \times r$) on the GF. The range of $\epsilon$ is about 0 to 1 and the filter window size is from 7 to 11. Figs. 5-7 show the results on nonoc, all and disc. The average results of these analyses are shown in Fig. 8 with the lowest average error at $\epsilon$ equals to 0.0001 and window size size ($9 \times 9$).
HAMZAH RA ET AL: Stereo matching base on illumination control

Fig. 5: All pixels in non-occluded regions (nonocc) for different filter size.

Fig. 6: All pixels detected with the valid pixels (all) for different filter size.

Fig. 7: Pixels in regions near the depth discontinuities and occluded regions (disc) with different filter size.
**DISCUSSION**

The performance of the proposed algorithm is compared with some other established local algorithms in (Scharstein and Szeliski, 2015) through Table 3. This table summarizes the quantitative performance in descending order of the overall performances and the best results are in bold for every attribute. These algorithms are taken for comparison due to their different approaches to the development and their complexities compared to the proposed algorithm. The proposed method is highly ranked among the local approaches and more accurate than the window-based matching cost algorithms such implemented in RealTimeABW (Gupta and Cho, 2010), VSW (Hu et al., 2011) and SNCC+AM (Einecke and Eggert, 2013). Moreover, the proposed algorithm outperforms recently published well-known census-based algorithm (i.e., Differential, Samadi and Othman, 2013, RINCensus, Ma et al., 2013a, and feature-based, i.e., TwoStep Wang et al., 2014) in most parameters evaluated. The proposed algorithm also outperforms two complex algorithms with iterative method and two layer structures, (i.e., RTAdaptWgt, Kowalczyk et al., 2013, and MSWLinRegr, Liu et al., 2012) which produces the best result in nonocc, all and disc regions for high texture of the cones image. Additionally, Fig. 11 shows that the proposed algorithm has the lowest average value of bad pixels percentage on the discontinuity errors.

---

**Table 1:** Comparison results with different methods or filters at Step 2.

<table>
<thead>
<tr>
<th>Algorithm/Filter</th>
<th>GF</th>
<th>ST</th>
<th>NL</th>
<th>RBF</th>
<th>ASW</th>
<th>MD</th>
<th>BX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ave (%)</td>
<td>5.45</td>
<td>5.49</td>
<td>5.68</td>
<td>5.68</td>
<td>5.78</td>
<td>7.56</td>
<td>9.78</td>
</tr>
</tbody>
</table>

**Step 3:** The optimization stage in this work uses WTA strategy which is similar to other local method algorithms. The minimum value of disparity intensity is chosen within the support region from Step 2.

**Step 4:** To make the algorithm robust against the image size, different test set (i.e., Tsukuba image) is used as test image to determine the weighted bilateral filter parameters. Other parameters in the previous steps remain unchanged. The test results are shown in Fig. 9. The lowest average of bad pixel value from the attributes (i.e., nonocc, all and disc) is chosen at $\sigma_i$ equals to 17 and $\sigma_c$ is 0.3 respectively. The window size of $w_p$ is 19.

The quantitative results of every disparity map are submitted to the authoritative testing website (Scharstein and Szeliski, 2015). Table 2 shows the comparative results with and without $\beta$ implementations while other parameters are fixed at the same values. The results demonstrate that the proposed algorithm with the illumination control performs much better. The proposed algorithm with $\beta$ is able to reduce all of the attributes evaluated. The most significant reduction is at the discontinuity areas. This occurs due to the implementation of image enhancement especially at the depth of the object boundaries. The final results of this work are shown in Fig. 10. The figure includes the results (i.e., images) on the raw matching disparity at Step 1, before post processing at Step 3, the L-R cross checking process and the region of bad pixels images. The running time is estimated at 0.01 second for each image.
HAMZAH RA ET AL: Stereo matching base on illumination control

Fig. 9: The results for Tsukuba image with different values of sigma space $\sigma_s$, sigma colour $\sigma_c$ and window size $w_p$.

(a) $w_p = (17 \times 17)$  
(b) $w_p = (19 \times 19)$  
(c) $w_p = (21 \times 21)$

Fig. 10: The final results on the Middlebury dataset. The running time of the images is Tsukuba (0.005s), Venus (0.009s), Teddy (0.01s) and Cones (0.01s).

Table 2: Results with and without $\beta$ implementations.

<table>
<thead>
<tr>
<th>Algorithms (threshold=1)</th>
<th>Tsukuba</th>
<th>Venus</th>
<th>Teddy</th>
<th>Cones</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nonocc</td>
<td>all</td>
<td>nonocc</td>
<td>all</td>
</tr>
<tr>
<td>Proposed algorithm without $\beta$</td>
<td>1.81</td>
<td>2.06</td>
<td>7.96</td>
<td>0.38</td>
</tr>
<tr>
<td>Proposed algorithm with $\beta$</td>
<td>1.63</td>
<td>1.91</td>
<td>7.14</td>
<td>0.28</td>
</tr>
<tr>
<td>Total difference</td>
<td>0.18</td>
<td>0.15</td>
<td>0.82</td>
<td>0.10</td>
</tr>
</tbody>
</table>

To test the capability of the proposed algorithm, some other stereo images in Fig. 12 are tested with the proposed algorithm. In general, the results demonstrate that the proposed algorithm generates a good disparity mapping. It can be seen that for the Barn and Sawtooth images where the foreground objects are detached from the background with clear and precise contours and accurate disparity values in accordance with their depth order. For the complex scene objects (i.e., Reindeer, Books, Moebius, Art, Dolls stereo pairs), the disparity values of the layered objects are correctly reconstructed in accordance with their respective depths. The scene objects are situated at increasing depth and are assigned step by step according to the disparity values from near to far. The results testify that the proposed algorithm produces an accurate and smooth disparity maps with clear and detailed edge contours.
Table 3: Performance comparison with established methods.

<table>
<thead>
<tr>
<th>Algorithms (threshold=1)</th>
<th>Tsukuba nonocc</th>
<th>Tsukuba all disc</th>
<th>Venus nonocc</th>
<th>Venus all disc</th>
<th>Teddy nonocc</th>
<th>Teddy all disc</th>
<th>Cones nonocc</th>
<th>Cones all disc</th>
<th>Ave (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed algorithm</td>
<td>1.63</td>
<td>1.91</td>
<td>7.14</td>
<td>0.28</td>
<td>0.49</td>
<td><strong>2.63</strong></td>
<td>6.10</td>
<td>11.4</td>
<td><strong>7.49</strong></td>
</tr>
<tr>
<td>HCFilter, 2013</td>
<td>1.56</td>
<td>1.78</td>
<td><strong>8.07</strong></td>
<td><strong>0.22</strong></td>
<td><strong>0.34</strong></td>
<td>2.96</td>
<td>6.18</td>
<td>11.5</td>
<td>16.1</td>
</tr>
<tr>
<td>MSWLinRegr, 2012</td>
<td>1.46</td>
<td>1.72</td>
<td>7.89</td>
<td>0.57</td>
<td>0.92</td>
<td>6.71</td>
<td>6.11</td>
<td>11.0</td>
<td><strong>15.6</strong></td>
</tr>
<tr>
<td>RTAdaptWgt, 2013</td>
<td>1.45</td>
<td>1.99</td>
<td>7.59</td>
<td>0.40</td>
<td>0.81</td>
<td>3.38</td>
<td>7.65</td>
<td>13.3</td>
<td>16.2</td>
</tr>
<tr>
<td>VSW, 2011</td>
<td>1.62</td>
<td>1.88</td>
<td>6.96</td>
<td>0.47</td>
<td>0.81</td>
<td>3.40</td>
<td>8.67</td>
<td>13.3</td>
<td>18.0</td>
</tr>
<tr>
<td>SNCC+AM, 2013</td>
<td>3.21</td>
<td>3.57</td>
<td>13.6</td>
<td>0.22</td>
<td>0.45</td>
<td>3.00</td>
<td>6.41</td>
<td>10.4</td>
<td>17.7</td>
</tr>
<tr>
<td>TwoStep, 2014</td>
<td>2.91</td>
<td>3.68</td>
<td>13.3</td>
<td>0.27</td>
<td>0.45</td>
<td><strong>2.63</strong></td>
<td>7.42</td>
<td>12.6</td>
<td>18.0</td>
</tr>
<tr>
<td>RealTimeABW, 2010</td>
<td><strong>1.26</strong></td>
<td><strong>1.67</strong></td>
<td><strong>6.83</strong></td>
<td>0.33</td>
<td>0.65</td>
<td>3.56</td>
<td>10.7</td>
<td>18.3</td>
<td>23.3</td>
</tr>
<tr>
<td>Differential, 2013</td>
<td>4.74</td>
<td>6.77</td>
<td>19.4</td>
<td>1.69</td>
<td>2.62</td>
<td>20.4</td>
<td>8.29</td>
<td>10.1</td>
<td>17.4</td>
</tr>
<tr>
<td>RINCencus, 2014</td>
<td>4.78</td>
<td>6.00</td>
<td>14.4</td>
<td>1.11</td>
<td>1.76</td>
<td>7.91</td>
<td>9.76</td>
<td>17.3</td>
<td>26.1</td>
</tr>
</tbody>
</table>

The performance of the proposed algorithm is also tested for adaptability to the real environment. The images from the KITTI dataset (Menze and Geiger, 2015) and the Universiti Sains Malaysia (USM) laboratory are taken into consideration. Fig. 13 shows the results of six continuous frames from the KITTI database. Every frame consists of three images which are the left image, ground truth and results of the proposed algorithm. It demonstrates that smooth disparity maps are generated from the proposed algorithm and is able to identify the vehicles and obstacles on each frame. The results on the USM images show smooth disparity maps are produced. Both of the above experiments show the good ability of the proposed method to reconstruct disparity maps from real environment.

CONCLUSION AND FUTURE WORKS

In this work, the accuracy of stereo matching algorithm is presented. With the standard qualitative benchmarking dataset tested, the proposed algorithm with the illumination control imposed at AD algorithm increases the robustness against the discontinuity regions and is able to reduce the errors as shown in Table 1. An average of error reduction with and without $\beta$ implementation is 1.15%. This proves that this method is able to increase the accuracy and most significantly reduces the discontinuity errors. The proposed algorithm is also tested with other dataset which contain low and high texture images. The good results are produced as shown in Figs. 12-14. The aggregation of the matching cost uses a guided filter. The advantage of applying this filter is its edge-preserving property. The optimization stage implements a WTA strategy which uses a proper minimum value. In the last stage, the weighted bilateral filter in post-processing step is able to reduce the existing noise that smoothers the disparity maps. For future works, this method will be tested in a...
standalone system (i.e., FPGA) and provide the energy consumption of the system to show the viability and its behaviors.

ACKNOWLEDGEMENT
This work was supported by Universiti Sains Malaysia's Research University Individual (RUI) with Account no. 1001/PELECT/814169 and Universiti Teknikal Malaysia Melaka.

REFERENCES


Gu Z, Su X, Liu Y, Zhang Q (2008). Local stereo matching with adaptive support-weight, rank transform...


