

## Article

# Objective Image Quality Measures for Disparity Maps Evaluation

Saad Merrouche , Milenko Andrić , Boban Bondžulić \*  and Dimitrije Bujaković 

Military Academy, University of Defence in Belgrade, Generala Pavla Jurišića Šturma 33, 11000 Belgrade, Serbia; saad.merrouche@va.mod.gov.rs (S.M.); milenko.andric@va.mod.gov.rs (M.A.); dimitrije.bujakovic@va.mod.gov.rs (D.B.)

\* Correspondence: boban.bondzulic@va.mod.gov.rs

Received: 8 August 2020; Accepted: 2 September 2020; Published: 2 October 2020



**Abstract:** The evaluation of disparity (range) maps includes the selection of an objective image quality (or error) measure. Among existing measures, the percentage of bad matched pixels is commonly used. However, it requires a disparity error tolerance and ignores the relationship between range and disparity. In this research, twelve error measures are characterized in order to provide the bases to select accurate stereo algorithms during the evaluation process. Adaptations of objective quality measures for disparity maps' accuracy evaluation are proposed. The adapted objective measures operate in a manner similar to the original objective measures, but allow special handling of missing data. Additionally, the adapted objective measures are sensitive to errors in range and surface structure, which cannot be measured using the bad matched pixels. Their utility was demonstrated by evaluating a set of 50 stereo disparity algorithms known in the literature. Consistency evaluation of the proposed measures was performed using the two conceptually different stereo algorithm evaluation methodologies—ordinary ranking and partition and grouping of the algorithms with comparable accuracy. The evaluation results showed that partition and grouping make a fair judgment about disparity algorithms' accuracy.

**Keywords:** disparity maps; error measures; image quality analysis; quantitative evaluation; stereo correspondence algorithms

## 1. Introduction

A stereo correspondence algorithm uses a stereo image pair as an input and produces an estimated disparity map (a new image) as an output [1,2]. The estimation of a disparity map is a fundamental problem in computer vision. This problem has been addressed in multiple domains such as outdoor mapping and navigation [3] and 3DTV [4].

With additional information about the stereo vision system, a disparity map can be transformed into a map of distances. Therefore, in addition to the term disparity map, the terms range map and depth map also appear in the literature.

The accuracy of stereo correspondence algorithms can be assessed by the evaluation of disparity maps, either qualitatively or quantitatively. A quantitative (or objective) approach is robust against several human-related biasing factors, offering advantages over a qualitative (subjective) approach. This assessment has practical applications such as component and procedure comparison, parameter tuning, supports decision-making by researchers and practitioners, and in general, to measure the progress in the field. It is useful to be able to quantify the quality of these disparity maps in order to benchmark range-finding devices and stereo correspondence algorithms [5].

Among the existing quantitative approaches used to estimate the accuracy/quality of disparity maps, Mean Squared Error (MSE), Root Mean Squared Error (RMSE) [6], Mean Relative Error (MRE) [7],

and the percentage of Bad Matched Pixels (*BMP*) [8] have been used extensively. A modification of the Multi-Scale Structural Similarity (*MS-SSIM*) index for range image quality assessment was proposed in [9], while the Sigma Z-Error (*SZE*) objective measure was introduced in [10]. The strengths of the *BMP* and *MRE* measures are combined in the Bad Matching Pixel Relative Error (*BMPRE*) measure [11].

The results of objective measures are used as inputs for quantitative evaluation methodologies, which aim to evaluate the performance of stereo algorithms. Among the existing quantitative evaluation methodologies, Middlebury's methodology is commonly used [12]. It uses the percentage of *BMP* as the error measure. As an alternative to this evaluation methodology, the  $A^*$  groups methodology has been proposed [13].

Among the current research in the field of stereo vision, disparity estimation based on deep learning methods [14] and real-time disparity estimation for high-resolution images [15] have been extensively studied, and significant progress has been made.

This paper proposes new disparity map evaluation measures, as modified versions of the original Image Quality Assessment (*IQA*) methods. The modification consists of the introduction of the capability to handle missing data in both the ground-truth disparity map (due to the lack of measures to determine ground-truth data) and in the estimated disparity map (due to the lack of an applied algorithm for image disparity estimation). The paper is focused also on consistency evaluation of the proposed measures using the two conceptually different state-of-the-art evaluation methodologies described in [12,13]. The obtained results show that the methodology extensions proposed in this research provide more accurate image disparity algorithm ranking.

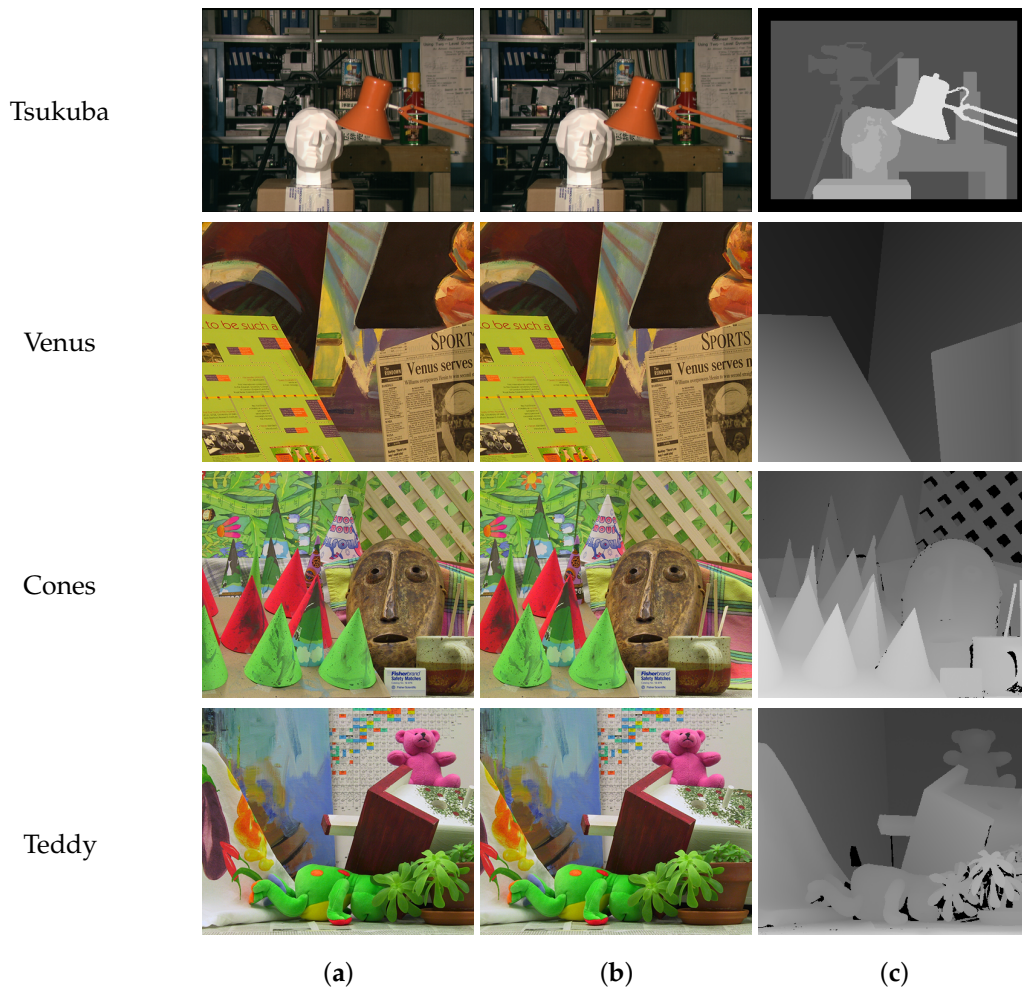
The rest of this paper is organized as follows. The next section will describe the methodologies and dataset used for disparity maps' evaluation. This is followed by the sections for experimental evaluation of known and the newly proposed objective error measures. The comparison of the results is shown in the next section. Finally, the conclusions are stated in the last section.

## 2. Methodologies and Dataset Used for Disparity Maps' Evaluation

Different methodologies have been used to compare the results of stereo correspondence algorithms. Publicly available datasets of stereo image pairs with ground-truth disparity maps are the basis for the application of the methodologies. Furthermore, different objective measures can be used when comparing stereo algorithms.

### 2.1. Dataset Description

Middlebury's stereo benchmark dataset [12] is used in this research. This dataset was selected because it is widely used and known by the stereo vision research community. It is composed of four indoor stereo image pairs (the Tsukuba, the Venus, the Teddy, and the Cones) captured under controlled conditions, which are illustrated, with their associated disparity ground-truth data, in Figure 1. The disparity ground-truth data of these images were generated using various methods. The Tsukuba stereo image pair is composed of front-parallel objects, while it has a disparity range of 16 pixels. The disparity ground-truth data of the Tsukuba stereo image pair was generated manually and is of integer precision. It excludes a border of 18 pixels, where no disparity value is provided. The Venus stereo image is composed of piecewise, planar slanted objects, while having a disparity range of 20 pixels. Each planar component was manually labeled, and a direct alignment technique was used on each planar region for estimating the affine motion of each patch. The horizontal component of these motions was used to compute the ground-truth disparity map. The Teddy and the Cones stereo images contain several objects with a different geometry, with the disparity range of 60 pixels. Their disparity ground-truth data were generated using a structured light technique. These techniques rely on projecting one or more special light patterns onto a scene, in order to directly acquire a scene range map.



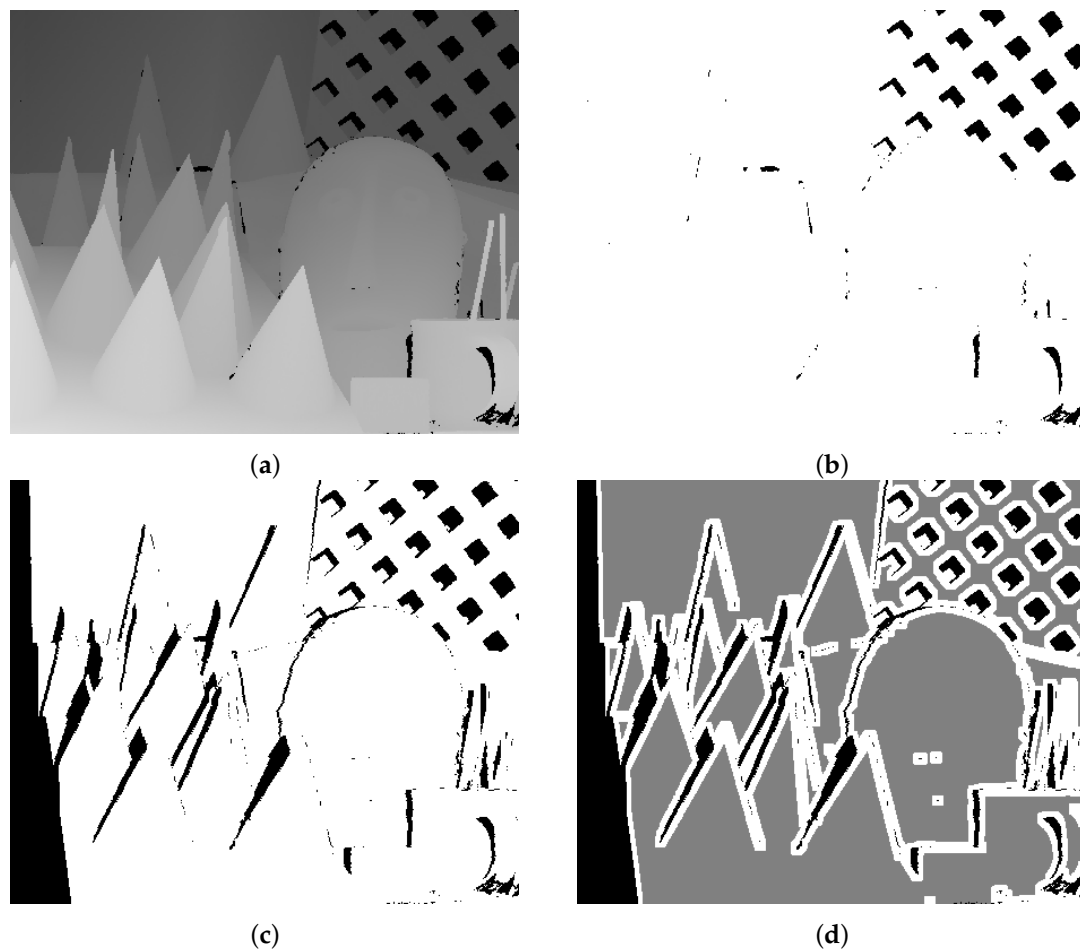
**Figure 1.** Middlebury Stereo Vision dataset: (a) left images, (b) right images, and (c) ground-truth disparity maps.

By observing stereo image pairs and the corresponding disparity maps from Figure 1, it can be noticed that objects that are closer to the observer on the disparity maps have higher gray level values (brighter regions). Therefore, disparity maps can be used to determine the depth of a scene and to determine the distance to objects.

## 2.2. Middlebury's Methodology

Middlebury's methodology was introduced in [6] and implemented for online stereo evaluation at [12]. The methodology is based on a testbed of four stereo image pairs [16], and the *BMP* measure is used as an error measure. It considers three error criteria related to different image regions: the entire image (*all*), the visible area in both images of the stereo pair (not occluded (*nonocc*)), and regions near depth discontinuities (*disc*), as seen in Figure 2.

The *all* criteria cover all regions known in the ground-truth (white pixels on Figure 2b); the *nonocc* criteria include all regions that are not occluded (white pixels on Figure 2c); while the *disc* criteria address all the areas that are near depth discontinuities and near occlusions (white pixels on Figure 2d).



**Figure 2.** Cones ground-truth image from the Middlebury stereo dataset with three error criteria considered: (a) ground-truth, (b) all known regions (*all*), (c) non-occluded regions (*nonocc*), and (d) regions near depth discontinuities (*disc*).

The evaluation model is based on comparing, sorting, and ranking the scores produced by the *BMP* measure, using the three error criteria. An overall ranking, which acts as an accuracy indicator, is computed as the average of the ranks obtained in each error criterion. The interpretation of the results is based on the final ranking. It is assumed that the algorithm with lower ranking is more accurate. Based on this methodology, it is possible to determine a set of top-ranking algorithms. However, the cardinality of this set is a free parameter, and this may lead to discrepancies among researchers about algorithm accuracy [13].

### 2.3. The $A^*$ Groups Methodology

The evaluation of disparity maps was addressed as a multi-objective optimization problem in [17], and the whole evaluation process is termed the  $A^*$  methodology. The introduced methodology computes the set  $A^*$ , which is composed by the stereo correspondence methods, with comparable accuracy among them (i.e., not better, nor worst) and, at the same time, superior to the rest of the algorithms under evaluation, according to the scores of selected error functions and considered error criteria. In this way, it avoids a subjective interpretation of evaluation results. This research is focused on presenting the methodology rather than on discussing the results achieved by introducing the evaluation model. Additionally, the *BMP* measure is used as the error measure for evaluating disparity maps' accuracy. Moreover, the  $A^*$  methodology is only capable of determining which algorithms are the most accurate overall, ignoring the rest of the algorithms.

The  $A^*$  methodology was extended in [13] by introducing the  $A^*$  groups methodology. It incorporates partitioning and grouping algorithms that allow a complete evaluation of the entire set of algorithms. In this way, multiple sets  $A^*$  are computed iteratively and associated with a label of accuracy that identifies the particular set as a group. The group  $A_1^*$  is of special interest since it is composed of the top performing stereo algorithms under evaluation.

### 3. Disparity Evaluation Using Objective Measures

Although there are a number of techniques for assessing the similarity between two images [18], simple pixel-based objective measures are generally used to assess the similarity between the ground-truth and the estimated disparity maps.

In most stereo applications, range images are obtained through range-finding devices or from structured light techniques. These methods often produce unknown regions (regions where the stereo algorithm fails to compute a depth estimate or regions where the range finder is shadowed by an obstacle). These unknown regions must be handled properly in order to obtain an accurate score from the quality metric. The local-based quality metric, termed  $R$ -SSIM [9], is capable of handling missing data in both the disparity map under evaluation and ground-truth data. Missing data in the disparity map may be associated with a fault of the stereo correspondence algorithm under evaluation.

In this section, pixel-based objective measures will be described first, and then, the local objective measures based on the Structural SIMilarity Index (SSIM) [19] will be introduced.

#### 3.1. Disparity Evaluation Using Pixel-Based Objective Measures

The five commonly used pixel-based error measures are considered for the disparity map evaluation—MSE, MRE, SZE, BMP, and BMPRE. MSE is formulated as:

$$MSE = \frac{1}{N} \sum_{(x,y)} (D_t(x,y) - D_e(x,y))^2, \quad (1)$$

where  $D_t$  and  $D_e$  are the ground-truth and estimated disparity maps, respectively, and  $N$  is the image size in pixels. However, the MSE measure does not distinguish well between disparity estimates with a large number of small errors and disparity estimates with a small number of large errors.

The MRE measure is formulated as:

$$MRE = \frac{1}{N} \sum_{(x,y)} \frac{|D_t(x,y) - D_e(x,y)|}{D_t(x,y)}, \quad (2)$$

while the SZE measure is defined as [10]:

$$SZE = \sum_{(x,y)} \left| \frac{f \cdot B}{D_t(x,y) + \mu} - \frac{f \cdot B}{D_e(x,y) + \mu} \right|, \quad (3)$$

where  $f$  is the focal length,  $B$  is the stereo system baseline (i.e., the distance between camera optical centers), and  $\mu$  is a constant added to avoid the instability due to missing estimations. In our research, it is assumed that  $f \cdot B = 1$ , while the value of the constant is  $\mu = 1$ . This measure is based on the inverse relation between disparity and depth using the magnitude of the disparity estimation error. It aims to measure the impact of a disparity estimation error on the 3D reconstruction. This impact depends on the true distance along the Z optical axis between the stereo camera system and the captured point and the point position according to the estimated disparity map.

The BMP measure is formulated as [12]:

$$BMP = \frac{1}{N} \sum_{(x,y)} \epsilon(x,y), \quad (4)$$

where  $\varepsilon(x, y)$  is a binary function defined as:

$$\varepsilon(x, y) = \begin{cases} 1, & |D_t(x, y) - D_e(x, y)| > \delta \\ 0, & |D_t(x, y) - D_e(x, y)| \leq \delta \end{cases} \quad (5)$$

and  $\delta$  is the error threshold, which was fixed to one pixel in our research (commonly used error threshold).

The *BMPRE* measure is formulated as [11]:

$$BMPRE = \sum_{(x,y)} \begin{cases} \tau(x, y), & \Delta(x, y) > \delta \\ 0, & \Delta(x, y) \leq \delta \end{cases} \quad (6)$$

where  $\tau$  is a function defined in order to avoid divisions by zero:

$$\tau(x, y) = \begin{cases} \rho(x, y), & D_t(x, y) > 0 \\ 0, & otherwise \end{cases} \quad (7)$$

$\Delta$  is the disparity error magnitude, computed as the absolute difference between the estimated disparity and the true disparity value:

$$\Delta(x, y) = |D_t(x, y) - D_e(x, y)|, \quad (8)$$

and  $\rho$  is the ratio between  $\Delta$  and the true disparity value—the relative error:

$$\rho(x, y) = \frac{\Delta(x, y)}{D_t(x, y)}. \quad (9)$$

The *BMPRE* measure considers, in a simple way, both the error magnitude and the inverse relation between depth and disparity. Moreover, the impacts of these considerations are not simple at all, since they allow a proper and deeper quantitative evaluation of disparity maps than the evaluation obtained using the *BMP*. In contrast to error measures that consider the inverse relation between depth and disparity (*SZE*), the *BMPRE* does not require additional information about the stereo camera system.

The pixel-based objective scores for the disparity maps of Tsukuba, Venus, Teddy, and Cones, estimated by the *PMF* and the *ADCensus* stereo correspondence algorithms [12], are shown in Table 1 (better objective scores of stereo algorithms for all stereo image pairs, criteria, and the complete dataset are highlighted in grey). The averaging of the 12 objective scores (4 stereo pairs  $\times$  3 criteria) points out the success of the algorithm on the global (dataset) level. These objective values can also be viewed as errors introduced by stereo algorithms (the lower, the better), i.e., the accuracy of the algorithms can be observed through them.



**Table 1.** Pixel-based objective scores for disparity maps computed by the *PMF* and *ADCensus* stereo correspondence algorithms.

		<i>PMF</i>					<i>ADCensus</i>				
		<i>BMP</i>	<i>SZE</i>	<i>MSE</i>	<i>MRE</i>	<i>BMPRE</i>	<i>BMP</i>	<i>SZE</i>	<i>MSE</i>	<i>MRE</i>	<i>BMPRE</i>
Tsukuba	<i>nonocc</i>	1.74	417.3	0.53	0.026	791.3	1.07	928.7	0.57	0.054	524.5
	<i>all</i>	2.04	449.9	0.57	0.028	983.1	1.48	970.0	0.68	0.057	832.9
	<i>disc</i>	8.07	131.8	2.33	0.056	669.2	5.73	156.1	1.85	0.061	516.73
Venus	<i>nonocc</i>	0.33	466.4	0.06	0.017	260.1	0.09	784.4	0.1	0.032	72.8
	<i>all</i>	0.49	512.4	0.09	0.018	414.7	0.25	819.5	0.12	0.033	193.6
	<i>disc</i>	4.16	79.4	0.62	0.044	247.1	1.15	76.7	0.39	0.044	72.2
Cones	<i>nonocc</i>	2.13	91.2	1.23	0.01	586.5	2.42	75.4	1.67	0.013	712.9
	<i>all</i>	6.80	844.5	5.11	0.02	1943.5	7.25	157.2	4.26	0.021	2144.8
	<i>disc</i>	6.32	76	3.69	0.02	582.8	6.95	45.5	4.89	0.023	705.2
Teddy	<i>nonocc</i>	2.52	76.9	0.56	0.01	512.6	4.1	295.1	6.23	0.017	990.4
	<i>all</i>	5.87	108.3	1.54	0.01	1213.5	6.22	370.6	7.91	0.022	1856.2
	<i>disc</i>	8.30	24.1	1.60	0.02	427.6	10.89	128.3	11.71	0.025	644.8
Average Value		<b>4.06</b>	<b>273.2</b>	<b>1.50</b>	<b>0.023</b>	<b>719.3</b>	<b>3.97</b>	<b>400.6</b>	<b>3.37</b>	<b>0.034</b>	<b>772.3</b>

The obtained *PMF* and *ADCensus* scores using the *BMP* measure are identical to the Middlebury stereo evaluation results [12]. Furthermore, the *BMP* and *BMPRE* measures, for the Cones and the Teddy images, indicate a superior accuracy of the *PMF* algorithm. On the contrary, they indicate a superior accuracy of the *ADCensus* algorithm for the Tsukuba and the Venus images. On the other side, the *MRE* scores indicate a superior accuracy of the *PMF* algorithm. Although the *SZE* and *MSE* scores depend on the stereo image pair selection and criteria used, it can be concluded that the *PMF* algorithm provides better disparity accuracy than the *ADCensus* algorithm. If the accuracy of the disparity map is observed through the average value, based on the *BMP* measure, the advantage is on the side of the *ADCensus* algorithm, while for the remaining four objective measures, the *PMF* algorithm provides a disparity map with better accuracy. Consequently, the selection of the error measure, in the evaluation process, has a great impact on the algorithm selection.

### 3.2. Disparity Evaluation Using SSIM-Based Objective Measures

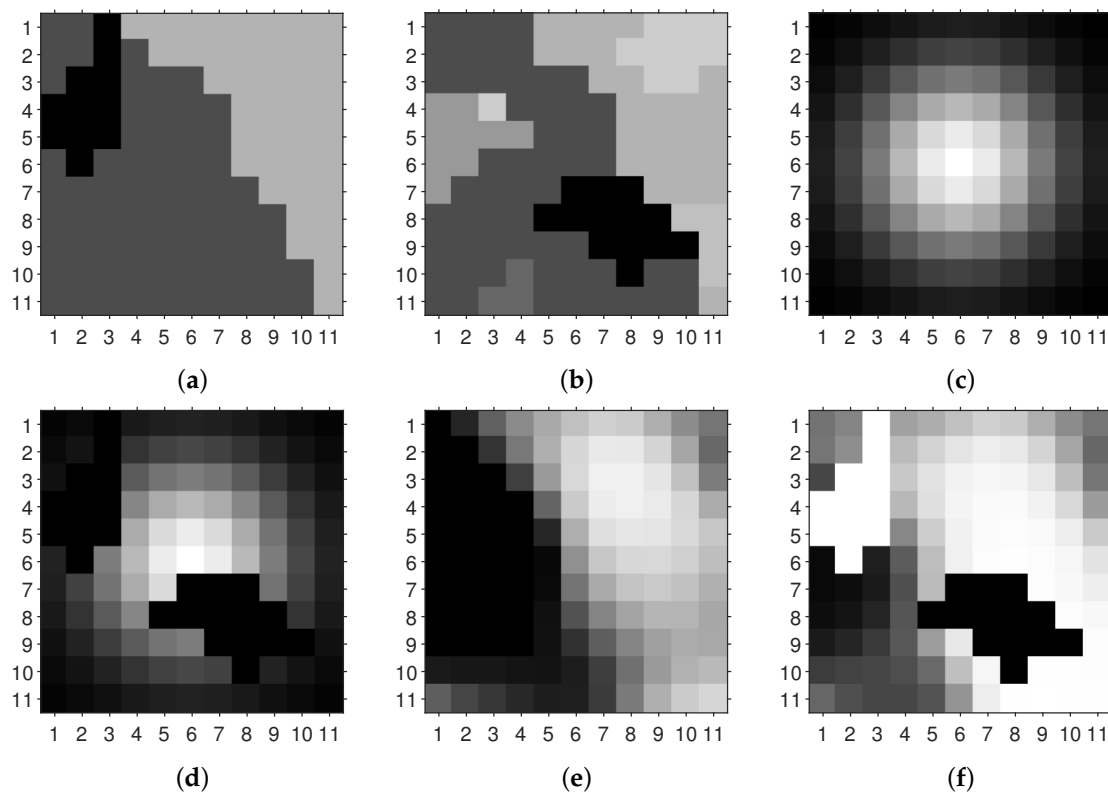
The *SSIM* index is a very popular algorithm for image quality evaluation. The basic idea behind the *SSIM* technique is that the images of natural scenes are rich in structures and that the human eye is sensitive to structural distortions. The index describes the quality by comparing the local luminance (mean value), contrast (local variance), and structure (local correlation) between the reference and the test image using an  $11 \times 11$  window [19]. The dynamic range of this measure is  $[-1, 1]$ . The Universal Image Quality Index (*UIQI*) [20] is a special case of the *SSIM* index. It is used to quantitatively assess a structural distortion between two images, and the dynamic range of this measure is also  $[-1, 1]$ . Both measures are calculated in the local regions of the image, using a moving window. The final quality indexes are obtained as the mean of all local quality values.

The Multi-Scale *SSIM* or *MS-SSIM* index [21] is the most popular variation of the *SSIM* index. It utilizes the *SSIM* algorithm over several scales. The reference and distorted images are iteratively driven through a low-pass filter and down-sampled by a factor of two. The resulting image pairs are processed with the *SSIM* algorithm. The *MS-SSIM* algorithm compares details across resolutions in a multiplicative manner, providing the overall image quality score [21].

The *R-SSIM* algorithm is a variation of the *MS-SSIM* algorithm with the ability to handle unknown regions or missing data in disparity maps' evaluation [9]. Treating unknown regions depends on their appearance—on the reference (ground-truth) image or on the distorted (estimated) image. Pixels in the unknown regions of the reference image should be ignored in objective measure calculations. Pixels in the unknown regions of the distorted image should be ignored when they fall inside the

sliding window used to calculate the local objective value, but the objective value of the unknown pixels themselves should be set to zero.

Figure 3 shows and explains how the  $R$ -SSIM algorithm was implemented on one scale. The same idea in this research is applied for the  $UIQI$  and  $SSIM$  objective measures, and these modified measures we named as  $UIQI_m$  and  $SSIM_m$ .



**Figure 3.** Explanation of  $R$ -SSIM (on one scale): (a) ground-truth disparity image patch, (b) estimated disparity image patch, (c) Gaussian weighting function, (d) renormalized weighting function with ignored unknown (black) pixels, (e) local structural similarity scores, and (f) local similarity scores with ignored unknown pixels.

Figure 3a depicts an  $11 \times 11$  patch in the ground-truth (reference) disparity image where unknown pixels are shown in black. Figure 3b shows the same patch in the estimated (test) disparity image, which also contains unknown (black) pixels. Figure 3c shows the Gaussian weighting function, which is used in local  $SSIM$ / $MS$ - $SSIM$  calculations, and Figure 3d shows it in the renormalized variant with ignored unknown (black) pixels. Finally, Figure 3e shows a map of the  $SSIM$  values of that patch, and Figure 3f shows its modification, where the pixel in the middle was calculated from Figure 3a,b,d. In Figure 3f, the unknown region from Figure 3b is indicated in black with a  $SSIM$  score of zero, while the unknown region from Figure 3a (shown in white) will be ignored in the final  $R$ - $SSIM$  score.

The  $SSIM$ -based objective evaluations for the disparity maps of the Tsukuba, Venus, Teddy, and Cones images, estimated using the *AdaptWeight* and the *TreeDP* stereo correspondence algorithms [12], are shown in Table 2 (better objective scores of stereo algorithms for all stereo image pairs and complete dataset are highlighted in grey).  $UIQI$  local scores are calculated using a sliding window of  $8 \times 8$  pixels and a uniform weighting function [20]. For these window-based measures, it is not reasonable to consider three criteria (*all*, *nonocc*, and *disc*) used for pixel-based error measures [9]. The averaging of the four stereo pairs' obtained error scores points out the success of the algorithm on the global (dataset) level.

Generally, we can notice that the scores after the modification ( $UIQI_m$ ,  $SSIM_m$ , and  $R$ - $SSIM$ ) are higher than the scores provided by the original objective measures ( $UIQI$ ,  $SSIM$ , and  $MS$ - $SSIM$ ).

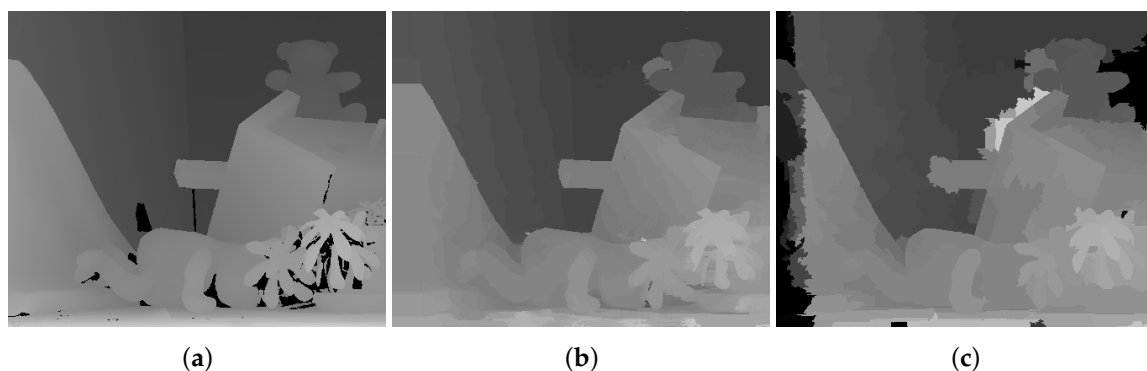


Objective image quality measures without and with modification provide equal scores in situations where both disparity maps (ground-truth and algorithm computed) do not contain unknown regions (Venus image and *AdaptWeight* algorithm). Nevertheless, modified objective measures yield more real and precise results. Furthermore, it can be noticed that the objective measures give results that differ on the used algorithm and the selected image pair. Based on the objective measures, it can be concluded that the *AdaptWeight* algorithm provides disparity estimations that are closer to the ground-truth data than the *TreeDP* disparity map calculations.

**Table 2.** SSIM-based evaluation for disparity maps computed by the *AdaptWeight* and *TreeDP* algorithms.

	<i>AdaptWeight</i>					<i>TreeDP</i>				
	Tsukuba	Venus	Cones	Teddy	Average Value	Tsukuba	Venus	Cones	Teddy	Average Value
<i>UIQI</i>	0.540	0.166	0.232	0.234	<b>0.293</b>	0.535	0.167	0.170	0.161	<b>0.258</b>
<i>UIQI<sub>m</sub></i>	0.687	0.166	0.250	0.255	<b>0.340</b>	0.684	0.168	0.186	0.177	<b>0.304</b>
<i>SSIM</i>	0.692	0.960	0.870	0.885	<b>0.852</b>	0.670	0.934	0.778	0.780	<b>0.791</b>
<i>SSIM<sub>m</sub></i>	0.862	0.960	0.936	0.941	<b>0.925</b>	0.834	0.943	0.845	0.840	<b>0.866</b>
<i>MS-SSIM</i>	0.806	0.963	0.868	0.854	<b>0.873</b>	0.762	0.936	0.756	0.702	<b>0.789</b>
<i>R-SSIM</i>	0.865	0.963	0.882	0.866	<b>0.894</b>	0.819	0.937	0.771	0.708	<b>0.809</b>

Confirmation of the precedence of the *AdaptWeight* stereo algorithm over the *TreeDP* algorithm (measured by objective scores; Table 2) can be seen in Figure 4.



**Figure 4.** (a) Ground-truth disparity map of the Teddy stereo image pair, (b) *AdaptWeight* obtained disparity map, and (c) *TreeDP* obtained disparity map.

By visual inspection of the disparity maps from Figure 4, it can be concluded that the disparity map obtained with the *AdaptWeight* algorithm is closer to the ground-truth disparity map than the map obtained with the *TreeDP* algorithm. Therefore, the objective similarity scores of the *AdaptWeight* obtained map are higher than the *TreeDP* generated one (Table 2).

#### 4. A New Local-Based Objective Measures for Disparity Evaluation

In this research, new objective measures are proposed for disparity image evaluation. Despite the SSIM-based objective measures, we used four state-of-the-art local-based objective image quality measures, which we adapted for range image quality assessment.

The gradient-based objective image quality assessment measure,  $Q^{AB}$ , is based on the preservation of gradient magnitudes and orientations [22]. A comparison of gradient information is carried out at the local level in a  $3 \times 3$  window, after which the local values of information preservation are used to determine the final quality score. As a result, a numerical  $Q^{AB}$  value is obtained that reflects the

quality of the test image—the accuracy with which the original gradient information is presented in the test image. The  $Q^{AB}$  values are in the range [0, 1].

Gradient Magnitude Similarity Deviation (*GMSD*) is also a gradient-based image quality assessment measure, which calculates a local similarity between the gradient magnitude maps of reference and test images in order to create a Local Quality Map (LQM) of the image degradation/similarity, in a  $3 \times 3$  window. After this step, the standard deviation of the LQM is calculated in order to achieve a final image quality estimate: *GMSD* [23]. In this research, we used the mean value of the LQM (Gradient Magnitude Similarity Mean (*GMSM*)) as a final quality score.

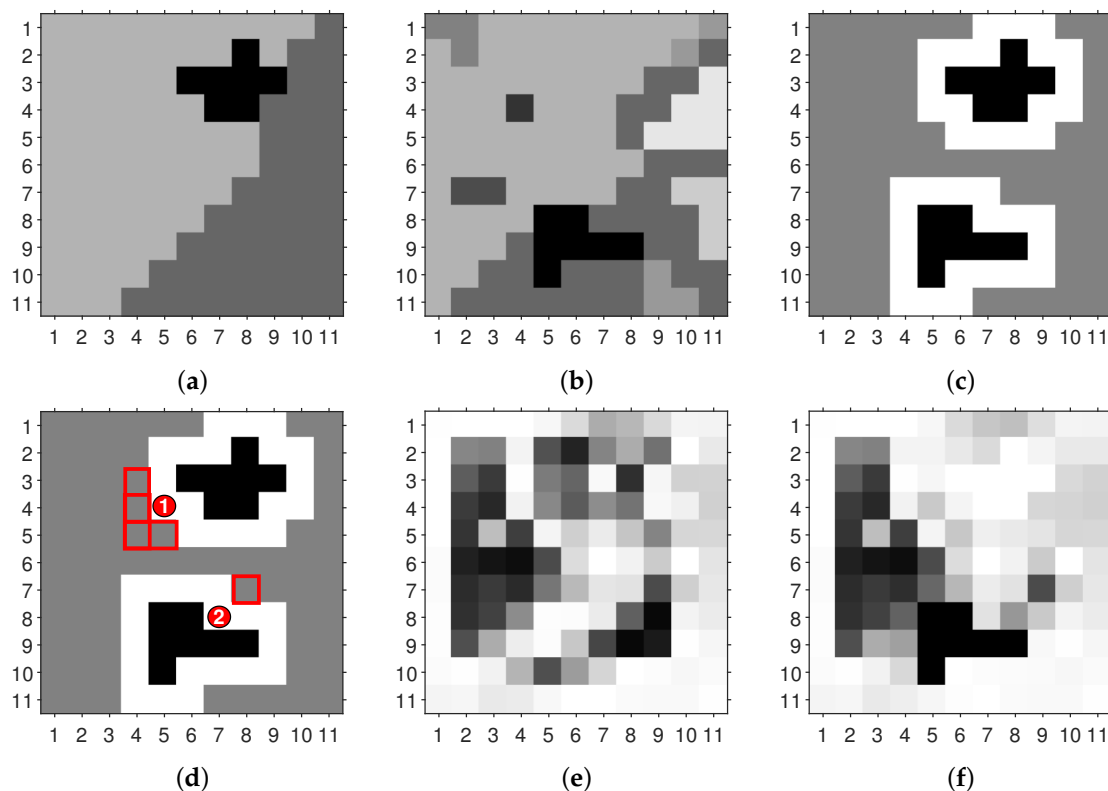
The Riesz transform and Visual contrast sensitivity-based feature SIMilarity index (*RVSIM*) is a full reference IQA method, which combines Riesz transform and visual contrast sensitivity [24]. *RVSIM* takes full advantage of the monogenic signal theory and log-Gabor filter by exploiting the contrast sensitivity function to allocate the weights of different frequency bands. At the same time, gradient magnitude similarity is introduced to obtain the gradient similarity matrix using a  $3 \times 3$  window. Then, the monogenic phase congruency matrix is used to construct the pooling function and obtain the *RVSIM* index.

Perceptual fidelity Aware MSE (*PAMSE*) is a variant of *MSE*, produced by introducing an l2-norm structural error term to it, using a  $7 \times 7$  Gaussian filter [25]. Lower *PAMSE* values indicate a higher similarity between the compared images, contrary to the other three local-based measures ( $Q^{AB}$ , *GMSM* and *RVSIM*) for which higher values indicate higher similarity.

For local-based objective measures  $Q^{AB}$ , *GMSM*, *RVSIM*, and *PAMSE*, in this research, a new filtering method has been proposed to treat the unknown regions from disparity maps. Unlike the *R-SSIM* in which, for each pixel influenced by the unknown regions, a new computed value is calculated from the sliding window, in this approach, the new value for pixels affected by unknown regions is determined as the average value of valid (unaffected) local objective scores.

The range adapted objective measures, based on the four previously mentioned measures, are named in this research as  $Q_m^{AB}$ ,  $GMSM_m$ ,  $RVSIM_m$ , and  $PAMSE_m$ . Figure 5 shows and explains how this adaptation was implemented for *GMSM*, which is also valid for  $Q^{AB}$  and *RVSIM*.

Figure 5a depicts a patch in the ground-truth (reference) image where there are some unknown pixels (shown in black), while Figure 5b shows the same patch in the computed range image, which also contains unknown (black) pixels. Figure 5c shows these unknown pixels and pixels affected by them (shown in white), in a  $3 \times 3$  local-based objective calculations. Figure 5d illustrates the proposed filtering method, where the objective score for the pixel numbered with one should be calculated from four local scores of unaffected pixels (shown as red squares), while the objective score for the pixel numbered with two should be the same as a local score of the only one unaffected pixel in a  $3 \times 3$  neighborhood. Finally, Figure 5e shows a map of the *GMSM* values of that patch, and Figure 5f shows its modification, where affected pixels were calculated using the proposed filtering method. In Figure 5f, the same unknown region from Figure 5b is indicated in black, with a *GMSM* score of zero, while the unknown region in Figure 5a (shown in white) will be ignored in the final  $GMSM_m$  score. Since *PAMSE* has an inverted scale (lower is better), the unknown pixels from the computed disparity image are penalized using the maximum disparity value, which is 255 for 8 bit/pixel grayscale images. Furthermore, for the *PAMSE* objective measure, it should be taken into account that the window size is  $7 \times 7$  pixels, which will give a larger number of affected pixels.



**Figure 5.** Explanation of the Gradient Magnitude Similarity Mean ( $GMSM_m$ ) index: (a) ground-truth disparity image patch, (b) estimated disparity image patch, (c) unknown pixels (in black) and pixels affected by them (in white), (d) local calculations based on unaffected pixels, (e)  $GMSM$  similarity scores, and (f) final similarity scores.

The local-based objective evaluations for the disparity maps of the Tsukuba, Venus, Teddy, and Cones images, estimated using the *AdaptWeight* and the *TreeDP* stereo algorithms [12], are shown in Table 3 (better objective scores of stereo algorithms for all stereo image pairs and the complete dataset are highlighted in grey).

**Table 3.** Local-based evaluation for disparity maps computed by the *AdaptWeight* and *TreeDP* algorithms.

	<i>AdaptWeight</i>					<i>TreeDP</i>				
	Tsukuba	Venus	Cones	Teddy	Average Value	Tsukuba	Venus	Cones	Teddy	Average Value
$GMSM$	0.887	0.983	0.947	0.949	<b>0.941</b>	0.869	0.973	0.910	0.912	<b>0.916</b>
$GMSM_m$	0.920	0.983	0.964	0.964	<b>0.957</b>	0.907	0.958	0.901	0.890	<b>0.914</b>
$Q^{AB}$	0.695	0.575	0.459	0.565	<b>0.573</b>	0.720	0.573	0.425	0.517	<b>0.559</b>
$Q_m^{AB}$	0.846	0.575	0.476	0.580	<b>0.619</b>	0.861	0.574	0.437	0.525	<b>0.599</b>
$RVSIM$	0.530	0.757	0.655	0.620	<b>0.640</b>	0.492	0.696	0.501	0.484	<b>0.543</b>
$RVSIM_m$	0.471	0.757	0.618	0.601	<b>0.612</b>	0.429	0.686	0.468	0.442	<b>0.506</b>
$PAMSE$	1754.6	14.3	274.4	313.5	<b>589.2</b>	1773.3	305.4	1551.5	1314.3	<b>1236.1</b>
$PAMSE_m$	84.1	14.3	48.4	51.7	<b>49.6</b>	232.4	1546.6	2629.4	3691.2	<b>2024.9</b>

Image quality scores after the modification ( $Q_m^{AB}$ ,  $GMSM_m$ ,  $RVSIM_m$ , and  $PAMSE_m$ ) differ from image quality scores provided using original implementations that do not handle unknown regions ( $Q^{AB}$ ,  $GMSM$ ,  $RVSIM$ , and  $PAMSE$ ). Objective image quality measures without and with modification provide equal scores for the Venus image and the *AdaptWeight* algorithm, for the same reason mentioned for the *SSIM*-based measures—neither disparity map contains unknown regions.

The results vary according to the choice of the objective algorithm, where the *AdaptWeight* algorithm provides disparity estimations that are closer to the ground-truth data than the *TreeDP* disparity map calculations (see Figure 4), which is also valid for the previously described *SSIM*-based objective measures (see Table 2).

## 5. Results

In this research, fifty disparity stereo algorithms used in Middlebury's dataset [12] were analyzed, where 34 algorithms were used in [9], and the 16 other algorithms were chosen among the state-of-the-art stereo algorithms. In order to perform image disparity algorithm evaluation, two methodologies were applied: Middlebury's methodology and the  $A^*$  groups methodology.

The ranking of these algorithms using pixel-based objective measures was done by sorting the average ranks obtained for different stereo image pairs (Tsukuba, Venus, Teddy and Cones) and criteria (*all*, *nonocc*, and *disc*); this means that for the 12 rank values' averaging while using *SSIM*-based and proposed local-based objective measures, the algorithms ranking was done by sorting the average ranks obtained for the four stereo image pairs (four ranking values averaging). The final algorithm rank was obtained by sorting the average ranks of all objective measures-based ranks.

A part of the ranked stereo correspondence algorithms according to Middlebury's evaluation methodology (*BMP*-based) and the rest of the considered objective measures are listed in Table 4.

**Table 4.** Part of the ranked stereo correspondence algorithms using the considered objective measures.

Algorithm	Error Measure												Rank	
	Pixel-Based				SSIM-Based				Proposed Local-Based				Average Rank	Final Rank
	<i>BMP</i>	<i>MSE</i>	<i>MRE</i>	<i>SZE</i>	<i>BMPRE</i>	<i>SSIM<sub>m</sub></i>	<i>R-SSIM</i>	<i>UIQI<sub>m</sub></i>	<i>GMSM<sub>m</sub></i>	<i>Q<sub>m</sub><sup>AB</sup></i>	<i>RVSIM<sub>B</sub></i>	<i>PAMSE<sub>m</sub></i>		
<i>TSGO</i>	3	1	5	5	1	1	4	5	2	2	1	1	2.58	1
<i>GC + LocalExp</i>	10	9	2	1	12	3	2	3	4	6	5	3	5.00	2
<i>AdaptingBP</i>	6	6	12	13	5	4	5	6	1	7	2	4	5.92	3
<i>PM-Forest</i>	7	7	4	2	4	9	6	11	6	3	10	7	6.33	4
<i>GC + LSL</i>	16	14	3	4	16	15	7	8	12	15	11	8	10.75	10
<i>SubPixDoubleBP</i>	11	5	11	7	13	13	13	13	11	26	13	14	12.50	11
<i>Segm + visib</i>	21	23	13	10	22	11	17	4	10	5	12	9	13.08	12
<i>IGSM</i>	1	8	14	20	2	17	21	25	17	13	14	21	14.42	13
<i>OverSegmBP</i>	26	19	19	14	20	23	16	30	19	35	24	20	22.08	23
<i>PlaneFitBP</i>	19	24	22	21	18	24	28	26	25	10	27	26	22.50	24
<i>AdaptWeight</i>	30	22	27	23	24	28	27	34	27	28	22	16	25.67	25
<i>RegionTreeDP</i>	27	27	29	28	23	27	25	29	26	31	18	23	26.08	26
<i>EnhancedBP</i>	28	36	35	38	32	32	34	40	39	32	37	40	35.25	37
<i>RealtimeBP</i>	34	35	39	36	37	37	38	38	37	30	36	32	35.75	38
<i>TensorVoting</i>	38	37	37	29	38	39	37	27	35	49	38	37	36.75	39
<i>RealTimeGPU</i>	39	33	40	34	36	40	39	43	40	38	40	30	37.67	40
<i>DP</i>	44	47	45	50	44	48	43	45	49	47	42	50	46.17	47
<i>SSD + MF</i>	46	46	48	48	48	46	47	47	47	41	49	48	46.75	48
<i>PhaseDiff</i>	49	49	49	42	50	49	49	49	48	48	48	46	48.00	49
<i>SO</i>	48	50	50	49	49	50	50	50	50	50	46	49	49.25	50

It can be observed that Table 4 lists 20 different stereo algorithms, and the best one of them is *TSGO* [12], which is reported among the top-five by all objective measures' rankings. Furthermore, it can be noticed that the algorithm rank in some cases differs significantly based on the objective measure used, i.e., the *IGSM* algorithm is the best ranked according to the *BMP* pixel-based measure, but it is ranked 25th using the *UIQI<sub>m</sub>* objective measure. This algorithm is positioned in 13th place on the global (final) level. For the bottom ranked algorithms (*DP*, *SSD+MF*, *PhaseDiff*, and *SO* [12]), there is a high level of agreement between objective-based rankings.

The affiliation of the above 20 selected algorithms with *A\** groups for the different objective error measures, according to the *A\** groups methodology [13], using 12 (pixel-based)/four (local-based) score errors, is shown in Table 5. The lower label values identify the algorithms with better accuracy. The last row of the table shows the number of *A\** subsets (labels) for each objective measure.

**Table 5.** Affiliation of the selected stereo correspondence algorithms to *A\** groups.

Algorithm	Algorithm $\in A^*$ Group											
	Pixel-Based				SSIM-Based				Proposed Local-Based			
	<i>BMP</i>	<i>MSE</i>	<i>MRE</i>	<i>SE</i>	<i>BMPRE</i>	<i>SSIM<sub>m</sub></i>	<i>R-SSIM</i>	<i>UIQI<sub>m</sub></i>	<i>GMSSM<sub>m</sub></i>	<i>Q<sub>m</sub><sup>AB</sup></i>	<i>RVSIM<sub>m</sub></i>	<i>PAMSE<sub>m</sub></i>
<i>TSGO</i>	1	1	1	1	1	1	1	1	1	1	1	1
<i>GC + LocalExp</i>	1	1	1	1	1	2	1	1	1	1	1	1
<i>AdaptingBP</i>	1	1	1	1	1	1	1	1	1	1	1	1
<i>PM-Forest</i>	1	1	1	1	1	1	1	2	1	1	1	1
<i>GC + LSL</i>	2	1	1	1	1	3	1	2	2	2	2	1
<i>SubPixDoubleBP</i>	2	1	2	1	1	1	1	3	1	3	2	1
<i>Segm + visib</i>	2	1	1	1	2	2	2	1	2	1	2	1
<i>IGSM</i>	1	1	1	2	1	3	3	3	3	2	2	3
<i>OverSegmBP</i>	3	2	2	2	2	4	2	4	3	4	2	2
<i>PlaneFitBP</i>	2	1	1	2	2	5	3	3	4	2	3	4
<i>AdaptWeight</i>	3	2	2	2	2	4	3	5	4	4	3	2
<i>RegionTreeDP</i>	2	2	3	2	2	4	3	4	4	4	2	3
<i>EnhancedBP</i>	2	2	3	3	3	7	4	5	7	4	7	5
<i>RealtimeBP</i>	3	3	4	3	3	7	6	5	6	4	5	5
<i>TensorVoting</i>	4	4	3	2	3	8	6	5	6	7	6	4
<i>RealTimeGPU</i>	4	2	4	3	3	7	6	6	7	6	5	4
<i>DP</i>	5	5	5	5	4	11	7	7	10	8	8	9
<i>SSD + MF</i>	6	5	5	5	5	10	8	7	9	6	8	9
<i>PhaseDiff</i>	6	5	5	4	5	11	8	7	9	8	9	8
<i>SO</i>	5	5	5	6	6	12	8	8	11	9	9	10
<i>A*</i> subsets	7	5	5	6	6	12	8	8	11	9	10	10

It can be seen that the first four average-based ranked stereo correspondence algorithms (Table 4) are members of almost all the instances of the *A<sub>1</sub>\** sets. There are two exceptions: the *GC+LocalExp* algorithm for the *SSIM<sub>m</sub>* measure and *PM-Forest* for the *UIQI<sub>m</sub>* error measure. The worst average-based ranked algorithms belong to the *A\** subsets with higher labels.

If we look at the individual ranks of the stereo correspondence algorithm *AdaptingBP* (Table 4), it can be concluded that this algorithm is positioned from first to 13th place. Based on all objective

measures, it is in third place. The accuracy of this algorithm was better evaluated using the  $A^*$  groups methodology, where Table 5 shows that it belongs to the  $A_1^*$  group by all objective measures.

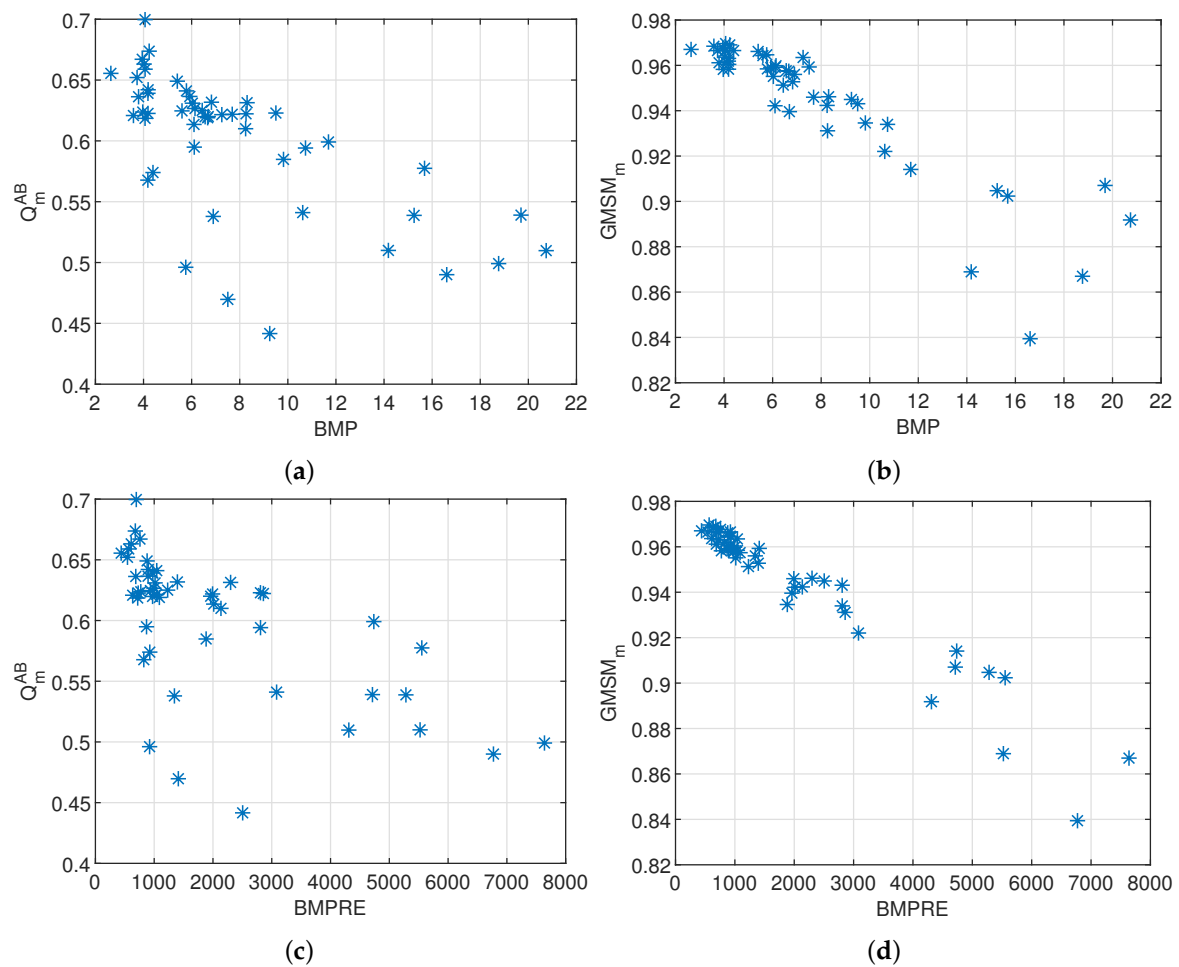
The Linear Correlation (LC) between average error values of objective measures applied for all the 50 stereo correspondence algorithms is shown in Table 6. From this table, it can be noticed that  $SZE$ ,  $UIQI_m$ ,  $Q_m^{AB}$ , and  $PAMSE_m$  have a weak mean correlation with the other objective measures (italicized scores in the last row). Moreover, we can see that proposed local-based measures  $GMSM_m$  and  $RVSIM_m$  have a high correlation with the other measures (bolded scores in the last row). Knowing that  $BMP$  and  $BMPRE$  are the most frequently used measures in the literature (anchor points), we can see that the proposed  $GMSM_m$  and  $RVSIM_m$  modifications can be good alternatives for them (bolded mutual scores), where  $GMSM_m$  is computationally more efficient.

**Table 6.** Linear correlation between objective evaluation measures for stereo correspondence algorithms.

	Error Measure											
	Pixel-Based					SSIM-Based			Proposed Local-Based			
	<i>BMP</i>	<i>MSE</i>	<i>MRE</i>	<i>SZE</i>	<i>BMPRE</i>	<i>SSIM<sub>m</sub></i>	<i>R-SSIM</i>	<i>UIQI<sub>m</sub></i>	<i>GMSM<sub>m</sub></i>	<i>Q<sub>m</sub><sup>AB</sup></i>	<i>RVSIM<sub>m</sub></i>	<i>PAMSE<sub>m</sub></i>
<i>BMP</i>	1	0.77	0.97	0.47	0.92	0.90	0.94	0.76	<b>0.90</b>	0.65	<b>0.95</b>	0.63
<i>MSE</i>	0.77	1	0.86	0.79	0.93	0.90	0.86	0.66	0.92	0.54	0.80	0.89
<i>MRE</i>	0.97	0.86	1	0.56	0.97	0.95	0.98	0.78	0.94	0.63	0.95	0.72
<i>SZE</i>	0.47	0.79	0.56	1	0.64	0.65	0.52	0.41	0.71	0.33	0.46	0.96
<i>BMPRE</i>	0.92	0.93	0.97	0.64	1	0.97	0.97	0.71	<b>0.96</b>	0.61	<b>0.90</b>	0.77
<i>SSIM<sub>m</sub></i>	0.90	0.90	0.95	0.65	0.97	1	0.96	0.73	0.99	0.63	0.89	0.76
<i>R-SSIM</i>	0.94	0.86	0.98	0.52	0.97	0.96	1	0.74	0.95	0.58	0.94	0.68
<i>UIQI<sub>m</sub></i>	0.76	0.66	0.78	0.41	0.71	0.73	0.74	1	0.73	0.75	0.84	0.54
<i>GMSM<sub>m</sub></i>	<b>0.90</b>	0.92	0.94	0.71	<b>0.96</b>	0.99	0.95	0.73	1	0.61	0.89	0.82
<i>Q<sub>m</sub><sup>AB</sup></i>	0.65	0.54	0.63	0.33	0.61	0.63	0.58	0.75	0.61	1	0.65	0.43
<i>RVSIM<sub>m</sub></i>	<b>0.95</b>	0.80	0.95	0.46	<b>0.90</b>	0.89	0.94	0.84	0.89	0.65	1	0.63
<i>PAMSE<sub>m</sub></i>	0.63	0.89	0.72	0.96	0.77	0.76	0.68	0.54	0.82	0.43	0.63	1
<i>minimum</i>	0.47	0.54	0.56	0.33	0.61	0.63	0.52	0.41	0.61	0.33	0.46	0.43
<i>maximum</i>	0.97	0.93	0.98	0.96	0.97	0.99	0.98	0.84	0.99	0.75	0.95	0.96
<i>mean</i>	0.81	0.81	0.85	0.59	0.85	0.85	0.83	0.70	<b>0.86</b>	0.58	<b>0.81</b>	0.71

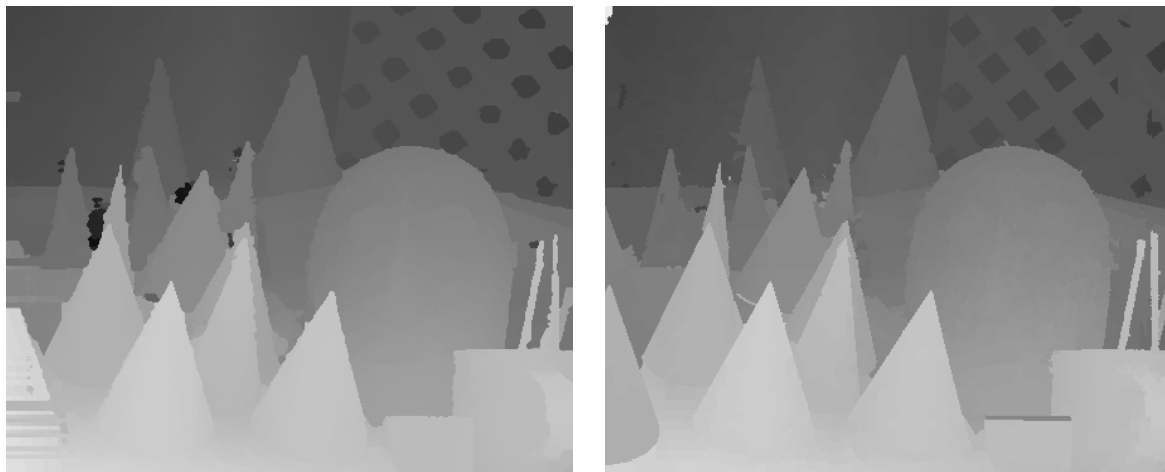
The correlation between objective measures can be illustrated through scatter plots. Figure 6 shows the scatter plots of  $GMSM_m$  and  $Q_m^{AB}$ , local-based objective error measures, versus  $BMP$  and  $BMPRE$ , pixel-based error measures, where each point (star) represents the average error of the utilized stereo algorithms (four values averaging for  $GMSM_m/Q_m^{AB}$  and 12 values averaging for  $BMP/BMPRE$ ). Scatter plots between these objective measures confirm the results given in Table 6. It can be observed that  $GMSM_m$  correlates better with  $BMPRE$  (Table 6, LC = 0.96) than  $BMP$  (Table 6, LC = 0.90). This can be explained by the fact that the  $GMSM_m$  objective measure is also sensitive to disparity error magnitudes as the  $BMPRE$  measure. For the disparity maps that are closer to the ground-truth (higher  $GMSM_m$  and lower  $BMP/BMPRE$  values), the correlation between  $GMSM_m$  and  $BMP/BMPRE$  objective measures is almost linear (in the upper left corner in the scatter plots). The degree of agreement between  $Q_m^{AB}$  and  $BMP/BMPRE$  is lower, which is visible through a big spreading of the objective scores on the scatter plots.





**Figure 6.** Scatter plots of: (a)  $Q_m^{AB}$  vs. BMP, (b)  $GSM_m$  vs. BMP, (c)  $Q_m^{AB}$  vs. BMPRE and (d)  $GSM_m$  vs. BMPRE on Middlebury dataset.

Figure 7 shows that the two objective methods can provide very different rankings to the same disparity maps. The two algorithms being assessed are graded in nearly reverse rank-order by the two assessment methods, BMP and  $GSM_m$ . Therefore, in the performance analysis of stereo correspondence algorithms, it is desirable to use more objective quality measures. Visual inspection of the two disparity maps suggests that in this instance, the  $GSM_m$  algorithm delivers a more meaningful assessment of the quality of the estimated disparity map. Nevertheless, it can be noticed that the  $GSM_m$  objective values of quality are very close (0.958 vs. 0.968) and that the ranks are significantly different (21 vs. eight). Therefore, it is reasonable to use the  $A^*$  groups methodology in the performance analysis of stereo correspondence algorithms, because algorithms whose accuracy is close will be classified in the same group.



*SubPixSearch* disparity map  
 BMP rank/score (all criteria): 6/6.87  
 GMSM<sub>m</sub> rank/score: 21/0.958

*OverSegmBP* disparity map  
 BMP rank/score (all criteria): 21/8.81  
 GMSM<sub>m</sub> rank/score: 8/0.968

**Figure 7.** Comparison of the results (disparity maps) of two computational stereo algorithms and their BMP and GMSM<sub>m</sub> ranks and scores.

We make the results for all 50 stereo algorithms publicly available to the research community as three Excel files in Supplementary Materials. These files contain extended results that are not presented in this paper: error scores for all analyzed stereo algorithms by all objective measures for each stereo pair and criteria, their rankings, and  $A^*$  groups subsets. Furthermore, in these files, researchers can find error scores before and after the proposed modifications for local-based objective measures.

## 6. Conclusions

In this paper, a characterization of error measures for evaluating disparity maps is presented. The impact on the results caused by the selection of the error measure was analyzed using Middlebury's and the  $A^*$  groups methodologies. This is the first attempt to analyze disparity maps using a plethora of pixel-based and local-based objective measures, using both methodologies.

We proposed modified objective measures as a new and needed quality metric for range images and demonstrated their utility by the evaluation of 50 stereo algorithms in the Middlebury Stereo Vision page, using four stereo image pairs with a rigid baseline. We confirmed that the new objective measures are effective for range image quality assessment; they complement the evaluation of stereo algorithms using conventional bad matched pixels approaches, but they can be stand-alone error measures. These modifications measure more than a loss of depth values; they are sensitive to errors in depth and surface structure, which cannot be measured using bad matched pixels approaches.

Experimental evaluations showed that the objective error measures may lead to different and contradictory ordinary rankings of stereo algorithms. These differences are less using the  $A^*$  groups methodology, where this methodology makes a fair judgment about the algorithms' accuracy. Consequently, the fair comparison of stereo correspondence algorithms may be a more difficult task than it has been considered so far.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2079-9292/9/10/1625/s1>, Excel S1: Error scores, Excel S2: Ranks, Excel S3: A groups.

**Author Contributions:** Conceptualization, B.B.; methodology, S.M.; software, S.M.; validation, B.B. and D.B.; formal analysis, D.B.; data curation, S.M.; writing, original draft preparation, S.M.; writing, review and editing, B.B., D.B., and M.A.; visualization, M.A.; supervision, M.A. All authors read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Chang, Y.J.; Ho, Y.S. Enhanced guided image filter using trilateral kernel for disparity error correction. In Proceedings of the IS&T International Symposium on Electronic Imaging, Burlingame, CA, USA, 13–17 January 2019; pp. 257–1–257-7.
2. Hamzah, R.A.; Ibrahim, H. Literature survey on stereo vision disparity map algorithms. *J. Sens.* **2016**, *2016*, 1–23. [\[CrossRef\]](#)
3. Menze, M.; Heipke, C.; Geiger, A. Object scene flow. *ISPRS J. Photogramm. Remote Sens.* **2018**, *140*, 60–76. [\[CrossRef\]](#)
4. Dufaux, F.; Pesquet-Popescu, B.; Cagnazzo, M. *Emerging Technologies for 3D Video: Creation, Coding, Transmission and Rendering*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2013.
5. Vargas, C.; Cabezas, I.; Branch, J.W. Stereo correspondence evaluation methods: A systematic review. In Proceedings of the International Symposium on Visual Computing, Las Vegas, NV, USA, 14–16 December 2015; pp. 102–111.
6. Scharstein, D.; Szeliski, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vis.* **2002**, *47*, 7–42. [\[CrossRef\]](#)
7. Van Der Mark, W.; Gavrilu, D.M. Real-time dense stereo for intelligent vehicles. *IEEE Trans. Intell. Transp. Syst.* **2006**, *7*, 38–50. [\[CrossRef\]](#)
8. Zitnick, C.L.; Kanade, T. A cooperative algorithm for stereo matching and occlusion detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 675–684. [\[CrossRef\]](#)
9. Malpica, W.S.; Bovik, A.C. Range image quality assessment by structural similarity. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, 19–24 April 2009; pp. 1149–1152.
10. Cabezas, I.; Padilla, V.; Trujillo, M. A measure for accuracy disparity maps evaluation. In Proceedings of the Iberoamerican Congress on Pattern Recognition, Pucón, Chile, 15–18 November 2011; pp. 223–231.
11. Cabezas, I.; Padilla, V.; Trujillo, M. BMPRE: An error measure for evaluating disparity maps. In Proceedings of the 11th IEEE International Conference on Signal Processing, Beijing, China, 21–25 October 2012; pp. 1051–1055.
12. Middlebury Stereo Evaluation—Version 2. Available online: <http://vision.middlebury.edu/stereo/eva> (accessed on 5 August 2020).
13. Cabezas, I.; Trujillo, M.P.; Florian, M. An evaluation methodology for stereo correspondence algorithms. In Proceedings of the International Conference on Computer Vision, Theory and Applications, Rome, Italy, 24–26 February 2012; pp. 154–163.
14. Pei, Z.; Wen, D.; Zhang, Y.; Ma, M.; Guo, M.; Zhang, X.; Yang, Y.H. MDEAN: Multi-view disparity estimation with an asymmetric network. *Electronics* **2020**, *9*, 924. [\[CrossRef\]](#)
15. Chang, Q.; Maruyama, T. Real-time stereo vision system: A multi-block matching on GPU. *IEEE Access* **2018**, *6*, 42030–42046. [\[CrossRef\]](#)
16. Scharstein, D.; Szeliski, R. High-accuracy stereo depth maps using structured light. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, WI, USA, 18–20 June 2003; pp. 195–202.
17. Cabezas, I.; Trujillo, M.P.; Florian, M. A non-linear quantitative evaluation approach for disparity estimation. In Proceedings of the International Conference on Computer Vision, Theory and Applications, Vilamoura, Portugal, 5–7 March 2011; pp. 704–709.
18. Athar, S.; Wang, Z. A Comprehensive performance evaluation of image quality assessment algorithms. *IEEE Access* **2019**, *7*, 140030–140070. [\[CrossRef\]](#)
19. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Wang, Z.; Bovik, A.C. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [\[CrossRef\]](#)
21. Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multi-scale structural similarity for image quality assessment. In Proceedings of the 37th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 9–12 November 2003; pp. 1398–1402.
22. Bondzulich, B.P.; Petrovic, V.S. Edge-based objective evaluation of image quality. In Proceedings of the 18th IEEE International Conference on Image Processing, Brussels, Belgium, 11–14 September 2011; pp. 3305–3308.

23. Xue, W.; Zhang, L.; Mou, X.; Bovik, A.C. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Trans. Image Process.* **2014**, *23*, 684–695. [[CrossRef](#)] [[PubMed](#)]
24. Yang, G.; Li, D.; Lu, F.; Liao, Y.; Yang, W. RVSIM: A feature similarity method for full-reference image quality assessment. *EURASIP J. Image Video Process.* **2018**, *2018*, 1–15. [[CrossRef](#)]
25. Xue, W.; Mou, X.; Zhang, L.; Feng, X. Perceptual fidelity aware mean squared error. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 705–712.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).