



**HAL**  
open science

## Structural Similarity Measurement Based Cost Function for Stereo Matching of Automotive Applications

Oussama Zeglazi, Mohammed Rziza, Aouatif Amine, Cédric Demonceaux

### ► To cite this version:

Oussama Zeglazi, Mohammed Rziza, Aouatif Amine, Cédric Demonceaux. Structural Similarity Measurement Based Cost Function for Stereo Matching of Automotive Applications. *Journal of Imaging*, 2020, 6 (8), pp.77. 10.3390/jimaging6080077 . hal-02916268

**HAL Id: hal-02916268**

**<https://hal.science/hal-02916268>**

Submitted on 17 Aug 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Article

# Structural Similarity Measurement Based Cost Function for Stereo Matching of Automotive Applications

Oussama Zeglazi <sup>1,\*</sup>, Mohammed Rziza <sup>1</sup>, Aouatif Amine <sup>2</sup> and Cédric Demonceaux <sup>3</sup>

<sup>1</sup> LRIT, Rabat IT Center, Faculty of Sciences, Mohammed V University, Rabat B.P. 1014, Morocco; rziza@fsr.ac.ma

<sup>2</sup> LGS, National School of Applied Sciences, Ibn Tofail University, Kenitra B.P. 241, Morocco; amine\_aouatif@yahoo.fr

<sup>3</sup> ERL VIBOT CNRS 6000, ImViA, Université Bourgogne Franche-Comté, 71200 Le Creusot, France; cedric.demonceaux@u-bourgogne.fr

\* Correspondence: zeglazi.oussama@gmail.com; Tel.: +212-670-281-916

Received: 22 June 2020; Accepted: 28 July 2020; Published: 3 August 2020



**Abstract:** The human visual perception uses structural information to recognize stereo correspondences in natural scenes. Therefore, structural information is important to build an efficient stereo matching algorithm. In this paper, we demonstrate that incorporating the structural information similarity, extracted either from image intensity (*SSIM*) directly or from image gradients (*GSSIM*), between two patches can accurately describe the patch structures and, thus, provides more reliable initial cost values. We also address one of the major phenomenons faced in stereo matching for real world scenes, radiometric changes. The performance of the proposed cost functions was evaluated within two stages: the first one considers these costs without aggregation process while the second stage uses the fast adaptive aggregation technique. The experiments were conducted on the real road traffic scenes KITTI 2012 and KITTI 2015 benchmarks. The obtained results demonstrate the potential merits of the proposed stereo similarity measurements under radiometric changes.

**Keywords:** stereo matching; structure similarity measurement; cross-based aggregation method; KITTI 2012; KITTI 2015

## 1. Introduction

Intelligent vehicles rely on active sensors (e.g., time-of-flight-camera [1], LiDAR [2]) in order to represent the cloud points of the surrounding environment. However, low cost passive computer vision offers the potential to produce richer geometric representations. In particular, our intention was paid to the stereo matching task, as it is vital for applications that are linked to intelligent vehicles.

The aim of stereo matching process is to estimate the depth of a scene viewed from two stereo images. Stereo matching algorithms can be roughly split into two categories. Sparse algorithms that rely on feature-based matching methods, generally used in camera calibration or orientation tasks [3,4], and dense algorithms, estimate depth values at every pixel value in the image.

Dense algorithms can be classified to global or local approaches. The global approaches formulate the stereo correspondence problem as an energy function over all image pixels with some smoothness constraints. This function is then minimized by global methods, such as the commonly used dynamic programming [5], belief propagation [6], and graph-cuts [7]. Generally, these approaches can effectively alleviate the matching ambiguities and, therefore, provide quite accurate depth results. However, they are inappropriate for real-time applications due to their slow convergence to optimal values. By contrast, local approaches consider for each individual pixel in the image a local smoothness

assumption to estimate its depth values [8–10]. This makes them computationally inexpensive but produce a lower disparity results, especially in textureless areas. A stereo matching algorithm can be performed in four steps [11]: cost computation, cost aggregation, disparity selection, and disparity refinement. The first step consists of matching pixels of the two stereo pairs. Several cost functions can be adopted in this step. Each of these have different characteristics that enable dealing with specific image regions. The second one, cost aggregation, is performed in order to filter out noisy matches that could have been occurred during the first stage. In the third step, disparity values are selected. The Winner-Take-All (WTA) strategy is often performed. It considers the disparity with the lowest or higher matching cost from the previous aggregation step. The last step, disparity refinement, is optional and it aims to refine erroneous disparity values by filtering out wrong matches using global smoothness assumptions.

Although all of these steps are required for accurate disparity results, the cost computation is the most critical, since early ambiguous cost values considerably affect the accuracy of the final results independently, regardless of the stereo matching algorithm. Therefore, obtaining a robust disparity map in real traffic situations require building a cost function that can be effective under radiometric distortions.

In this paper, we propose two new cost functions, which are based on the structural information ( $SSIM$ ),  $C_{SSIM}$ , and its gradient variant, the  $C_{GSSIM}$ . The performance of the proposed costs was evaluated using both aggregation [10] and no aggregation approaches. The local WTA strategy was adopted to generate disparity maps. The experimental results were conducted on two challenging datasets, the real road traffic stereo pairs of KITTI 2012 [12] and KITTI 2015 [13].

The remainder of the paper is organized, as follows: in Section 2, we review the related works to the matching cost functions. In Section 3, we present the proposed cost function. Experimental results and discussions are given in Section 4. Additionally, finally, we draw conclusions in Section 5.

## 2. Related Work

A wide range of cost functions have been proposed in the literature. Of these, the absolute intensity differences, squared intensity differences, cross correlation sum, and normalized cross-correlation. Non parametric cost functions have been introduced for being robust against radiometric distortions [14]. Authors in [15] have proposed a cost function based on the mutual information in order to handle the complex radiometric relationships between images. Several works have focused on enhancing the performance of the traditional cost functions by proposing enhanced costs or by merging multiple cost functions to provide efficient variants of the existed ones. In [16], the authors fused both the absolute difference on image color and gradient along the horizontal direction. Other studies have exponentially fused the absolute difference on image color with the Census Transform (CT) cost function [17]. The authors in [18] have fused three cost functions: the absolute difference on color image, on image gradients, and the CT computed in image gradients using an exponential function. Authors in [19] have proposed an adaptive fusion method of multiple cost matching functions. The efficiency of the state of art cost function has been widely examined in several studies [20–22]. Indeed, the study that is presented in [20] included the comparison of robustness using six cost matching functions in term of photometric distortion and noise. While [21] is more extended and it has included the evaluation of fifteen different cost functions using various optimization schemes. The results have demonstrated that costs that are based on the CT give the best results, particularly for radiometric changes. Recently, authors in [22] have investigated cost functions in stereo matching algorithms for automotive vehicle applications using two different stereo matching algorithms. One is based on global energy optimization (Graph cuts) [7], and the other one uses local adaptive method [10]. The results of this study have proven that the cost function derived from the CT or its variants, as the Cross-Comparison Census (CCC) combined with the mean sum of relative pixel intensity differences within a CT window, provide overly a good performance on the KITTI 2012 benchmark. A variant of CCC cost function [23] was proposed in order to handle better the radiometric distortions. The authors

claim that the proposed cost function outperforms the conventional cost functions on the KITTI 2012 benchmark. These studies have demonstrated that it is quite difficult to address the disparity, with radiometric distortions, relying only on intensity-based cost functions. Some research studies have investigated *SSIM* for stereo matching algorithms. In [24], authors have proposed to compute the final matching cost function using *SSIM* index over filtered left and right patches obtained from the non-local means algorithm [25]. In [26], the *SSIM* index has been introduced for multiview stereo to compute the matching cost function in coarse-to-fine workflow.

### 3. The Proposed Cost Function

#### 3.1. SSIM Based Cost Function ( $C_{SSIM}$ )

When considering the stereo matching problem as a visual issue. Extracting the most adopted information captured by the Human Visual System (HVS) can provide a consistent information in order to accurately describe the considered patch, and facilitate the matching process. In this context, we propose a new cost function based on the structural information [27]. Let  $p(x, y)$  be a pixel in the reference image ( $I_1$ ),  $I_p$  is the intensity value of pixel  $p$  and  $q(x, y-d)$  its hypothetical corresponding, with intensity value  $I_q$  in the target image ( $I_2$ ) at a disparity  $d$ . The  $C_{SSIM}$  between  $p$  and  $q$  is defined, as follows:

$$C_{SSIM}(p, q, d) = [l(p, q, d)]^\alpha \cdot [c(p, q, d)]^\beta \cdot [s(p, q, d)]^\gamma \quad (1)$$

where,  $l(p, q, d)$  is the luminance,  $c(p, q, d)$  the contrast and  $s(p, q, d)$  structure measurements between  $p$  and  $q$ , defined in Equations (2)–(4), respectively.

$$l(p, q, d) = \frac{2\mu_p\mu_q + C}{\mu_p^2 + \mu_q^2 + C} \quad (2)$$

$$c(p, q, d) = \frac{2\sigma_p\sigma_q + C}{\sigma_p^2 + \sigma_q^2 + C} \quad (3)$$

$$s(p, q, d) = \frac{\sigma_{(p,q)} + C}{\sigma_p\sigma_q + C} \quad (4)$$

$C$  is a small constant to avoid the denominator being zero.  $\mu_p$  and  $\mu_q$  are the mean values computed in neighborhood  $N_p$  and  $N_q$  of  $p$  in  $I_1$  and  $q$  in  $I_2$ , respectively.  $\sigma_p$  and  $\sigma_q$  are standard deviations of  $p$  and  $q$  respectively. The standard deviations of  $p$  in the support window  $N_p$  is described as follows:

$$\sigma_p = \left( \frac{1}{(|N_p| - 1)} \sum_{p' \in N_p} (I_{p'} - \mu_p)^2 \right)^{1/2} \quad (5)$$

where  $|N_p|$  is the number of pixels in the support window  $N_p$ . The  $\sigma_{(p,q)}$  is the covariance between  $p$  and  $q$ , and can be estimated as:

$$\sigma_{(p,q)} = \frac{1}{(|N_p| - 1)} \sum_{p' \in N_p, q' \in N_q} (I_{p'} - \mu_p)(I_{q'} - \mu_q) \quad (6)$$

Finally,  $\alpha > 0$ ,  $\beta > 0$ , and  $\gamma > 0$  are parameters that allow for controlling the influence of the each of the three components.

#### 3.2. SSIM Gradient Variant ( $C_{GSSIM}$ )

Besides the structural information, the human visual system is capable of extracting the image gradients based structural features, such as (edges and points). Thus, in order to take into account this assumption, the structural information is extracted from image derivatives  $\partial I / \partial x$ ,  $\partial I / \partial y$ , rather

than image intensities. To do so, the luminance (l), the contrast (c), and the structure measurement (s), in Equation (1) will be modified by incorporating the gradient. Therefore, the gradient based structural information cost function  $C_{GSSIM}$  is defined, as follows:

$$C_{GSSIM}(p, q, d) = [l_g(p, q, d)]^\alpha \cdot [c_g(p, q, d)]^\beta \cdot [s_g(p, q, d)]^\gamma \quad (7)$$

where  $l_g$ ,  $c_g$  and  $s_g$  are structural information defined as follows :

$$l_g(p, q, d) = \sum_{p \in \{\frac{\partial I_1}{\partial x}, \frac{\partial I_1}{\partial y}\}, q \in \{\frac{\partial I_2}{\partial x}, \frac{\partial I_2}{\partial y}\}} \frac{2\partial\mu_p\partial\mu_q + C}{\partial\mu_p^2 + \partial\mu_q^2 + C} \quad (8)$$

$$c_g(p, q, d) = \sum_{p \in \{\frac{\partial I_1}{\partial x}, \frac{\partial I_1}{\partial y}\}, q \in \{\frac{\partial I_2}{\partial x}, \frac{\partial I_2}{\partial y}\}} \frac{2\partial\sigma_p\partial\sigma_q + C}{\partial\sigma_p^2 + \partial\sigma_q^2 + C} \quad (9)$$

$$s_g(p, q, d) = \sum_{p \in \{\frac{\partial I_1}{\partial x}, \frac{\partial I_1}{\partial y}\}, q \in \{\frac{\partial I_2}{\partial x}, \frac{\partial I_2}{\partial y}\}} \frac{\partial\sigma_{(p,q)} + C}{\partial\sigma_p\sigma_q + C} \quad (10)$$

$\partial\mu_p$  and  $\partial\mu_q$  are the mean values computed for the neighborhood  $\partial N_p$  in  $\partial I_1$  for  $p$  and  $\partial N_q$  in  $\partial I_2$  for  $q$  and  $q$  in  $\partial I_2$ .  $\partial I_1$  and  $\partial I_2$  are the gradients along  $x$  and  $y$  directions, respectively.  $\partial\sigma_p$  and  $\partial\sigma_q$  are the standard deviations of  $p$  in  $\partial I_1$  and  $q$  in  $\partial I_2$ . The standard deviations of  $p$   $\partial\sigma_p$  is defined, as follows:

$$\partial\sigma_p = \sum_{p \in \{\frac{\partial I_1}{\partial x}, \frac{\partial I_1}{\partial y}\}} \left( \frac{1}{(|N_p| - 1)} \sum_{p' \in \partial N_p} (\partial I_{p'} - \partial\mu_p)^2 \right)^{1/2} \quad (11)$$

The  $\partial\sigma_{(p,q)}$  is defined, as follows:

$$\begin{aligned} \partial\sigma_{(p,q)} &= \sum_{p \in \{\frac{\partial I_1}{\partial x}, \frac{\partial I_1}{\partial y}\}} \frac{1}{(|N_p| - 1)} \\ &* \sum_{p' \in \partial N_p, q' \in \partial N_q} (\partial I_{p'} - \partial\mu_p)(\partial I_{q'} - \partial\mu_q) \end{aligned} \quad (12)$$

In contrast to the Equation (1), this enables to compute the new structural features on image principal derivatives with respect to  $x$  and  $y$  coordinates.

#### 4. Experimental Results

In this section, we evaluate the ability of the proposed cost functions to discriminate stereo correspondences. We explore the proposed costs for stereo matching through two different algorithms: a stereo matching algorithm without aggregation stage and a fast local adaptive aggregation technique. These cost functions are then compared to the top cost functions  $C_{DIFFCensus}$  [22] and  $C_{GCC}$  [23]. The optimal parameter values that were proposed in [22,23] were retained. Experiments were conducted on the KITTI 2012 [12] and KITTI 2015 [13] training datasets in order to evaluate the proposed approach in the context of intelligent vehicles applications.

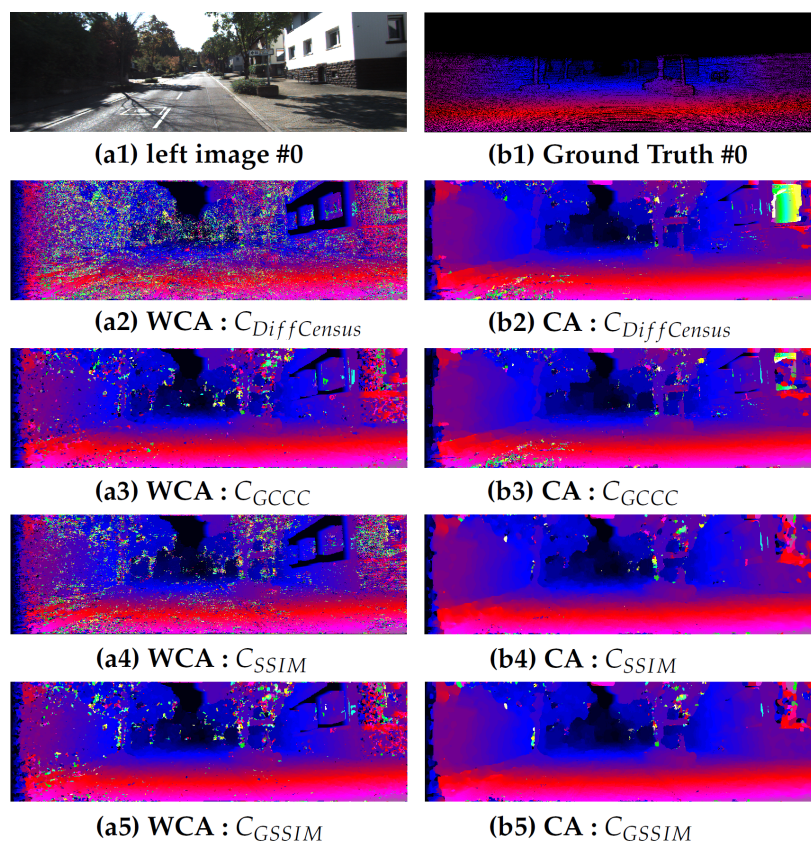
- The KITTI 2012 is divided into two sets, training one which contains 194 stereo pairs and 195 stereo pairs in the testing one.
- The KITTI 2015 dataset contains 200 training stereo pairs and 200 testing pairs.

The evaluation for the KITTI 2012 datasets is measured by computing the percentage of disparity errors with respect to the ground truth. While, for the KITTI 2015  $D1$ —all error measure is computed, it represents the percentage of pixels for which the estimation error is larger than three pixels and larger than 5% of the ground truth disparity at each pixel. For the parameters sets of both cost functions,  $C_{SSIM}$  and  $C_{GSSIM}$ , were experimentally set as:  $\alpha = 0.9$ ,  $\beta = 0.1$  and  $\gamma = 0.2$  to minimize the overall

error rate. Parameter  $C$  is set to the smallest value to prevent dividing by zero. In the aggregation stage, the spacial and color similarity thresholds were fixed at  $L = 9$  and  $\tau = 20$ , respectively. The local WTA strategy was adopted in order to generate disparity results. We used the highest matching cost instead of the lowest one, as the proposed costs are built upon similarity measurement.

#### 4.1. Evaluation of the Discriminative Ability of the Proposed Costs

In this section, the effectiveness of the proposed cost functions is studied on both KITTI datasets without using any cost aggregation method. Figure 1 shows a visualization of the output disparity results for each cost functions using both two stereo algorithms is presented. Column one shows the results that were obtained without using an aggregation method, while column two shows the results obtained with based on adaptive aggregation method. The output results for the #0 stereo pair from the KITTI 2012 training dataset are presented. The presented figure illustrates, in both cases, that the proposed cost functions lead to promising results, while the conventional costs provide highly noisy disparity results.



**Figure 1.** Disparity maps of #0 stereo pair from the KITTI 2012 dataset . The first column corresponds to left stereo (a1) with its corresponding ground truth disparity map (b1). The computed disparity maps are listed in the following lines corresponding to cost functions  $C_{DiffCensus}$ ,  $C_{GCCC}$ ,  $C_{SSIM}$  and  $C_{GSSIM}$ , respectively. First column (a) correspond to the output obtained without the use of an aggregation method (WCA), while second column (b) are the output based on the adaptive aggregation method (CA).

Tables 1 and 2 present the mean error rate on all of the stereo pairs in different regions (non–occluded and all) on both datasets, KITTI 2012 and KITTI 2015, respectively. According to results, our both cost functions achieves notable results. In addition, the  $C_{GSSIM}$  provides the lowest rate of error in both datasets.



The presented results demonstrate the discrimination power of the proposed costs without considering aggregation costs, which proves the effectiveness of the *SSIM* information for capturing reliable local information for stereo matching. The next section investigates the efficiency of these costs while using aggregation techniques.

**Table 1.** Percentage of erroneous disparities of stereo matching without an aggregation method for KITTI 2012 training database.

Cost Functions	3-px Threshold	
	Non-Occluded	All
$C_{DIFFCensus}$ [22]	54.25	55.30
$C_{GCCC}$ [23]	30.78	32.34
$C_{SSIM}$	20.94	22.73
$C_{GSSIM}$	18.00	19.86

**Table 2.** Results on the KITTI 2015 training datasets.

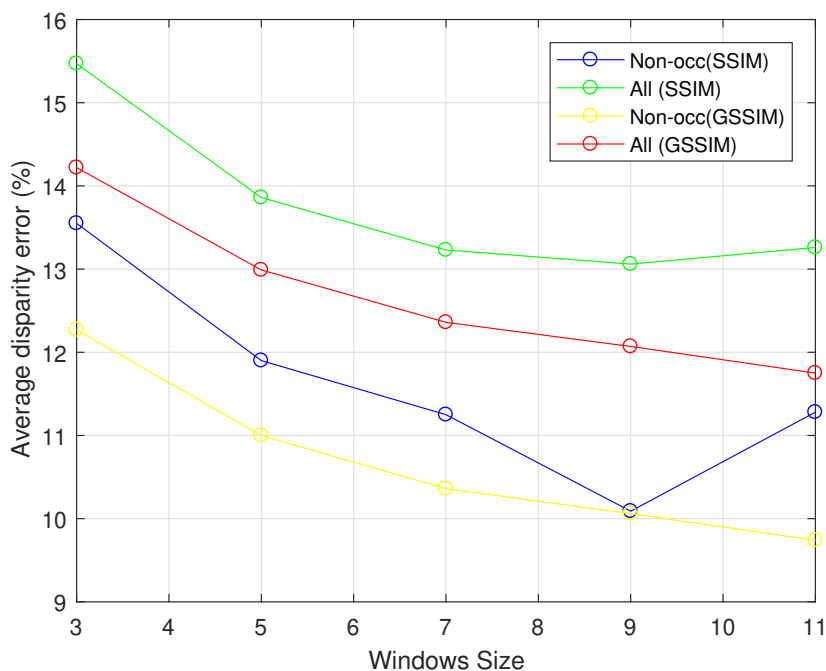
Cost Functions	No Aggregation Method		Aggregation Method	
	D1—All (Non-Occluded)	D1—All (All)	D1—All (Non-Occluded)	D1—All (All)
$C_{DIFFCensus}$ [22]	50.74	49.87	18.63	20.05
$C_{GCCC}$ [23]	27.52	28.76	14.63	16.05
$C_{SSIM}$	15.23	16.69	11.08	13.06
$C_{GSSIM}$	15.38	16.83	10.06	12.07

#### 4.2. Evaluation of the Proposed Costs Using the Adaptive Aggregation Technique

To further reduce noise and construct refined cost functions, the adaptive aggregation method [10] was performed. This choice is motivated by the fact that this method is fast and accurate, which is suitable for real time applications.

The effectiveness of the proposed method was firstly evaluated with respect to the support window size on KITTI 2012 training datasets. Figure 2 presents the mean error rate, in both non-occluded and all regions, computed at the default 3 pixels threshold for all of training set images. It can be noted that the size of the support window impacts highly the performance of the algorithm of both cost functions. Indeed, significant improvement in the performance of the local stereo matching algorithm can be obtained as the size of the support window increases. More precisely, the improvement is by a factor of 1.65% for the non-occluded and by 1.61% for occluded zones, for the  $C_{SSIM}$  cost function when the size window passed from 3 to 5, for example.

In the following, we evaluate the robustness of the proposed cost functions based on adaptive aggregation method against the state-of-the-art cost function. Tables 2 and 3 present the average percentage of erroneous pixels with both non-occluded and all regions. In Table 3, the errors were calculated at three different pixels error thresholds, while in Table 2 the D1—all error was computed. The obtained results indicate that the proposed  $C_{GSSIM}$  cost functions outperform the others ones by a significant margin. Indeed, the  $C_{GSSIM}$  provides the lower mean disparity errors on both datasets, followed by the proposed  $C_{SSIM}$  cost function under different scenarios. Indeed, in Table 3 at the default three pixel threshold, the improvement obtained by  $C_{GSSIM}$  is of the order 2.23, 3.47 for non-occluded region and of 2.84, 3.4 for other zones, with respect to  $C_{DIFFCensus}$  and  $C_{GCCC}$  costs. Besides, from Table 2, we can see clearly that the performance of our methods are significantly better than all other cost functions in both regions. For example, the improvement obtained by  $C_{GSSIM}$  is of the order 1.87, 2.91 for non-occluded region and of 1.82, 2.84 for other zones, with respect to  $C_{GCCC}$  and  $C_{DIFFCensus}$  costs.



**Figure 2.** The disparity results obtained with respect to support window size for both cost functions, for KITTI 2012 training sets.

**Table 3.** Percentage of erroneous disparities in non-occluded and regions for the KITTI 2012 training set.

Cost functions	2 px Threshold		3 px Threshold		5 px Threshold	
	Non-Occluded	All	Non-Occluded	All	Non-Occluded	All
$C_{DIFFCensus}$ [22]	20.08	21.88	12.97	14.91	9.70	11.63
$C_{GCCC}$ [23]	20.47	22.27	11.93	13.89	10.19	12.18
$C_{SSIM}$	16.39	18.29	11.08	13.06	8.18	10.17
$C_{GSSIM}$	14.06	16.00	10.06	12.07	7.83	9.83

This evaluation shows that the proposed  $C_{GSSIM}$  cost function is more appropriate for the real outdoor disparity computation than the top performers  $C_{DiffCensus}$  and  $C_{GCCC}$ .

4.3. Sensitivity of the Cost Functions in the Presence of Radiometric Distortions

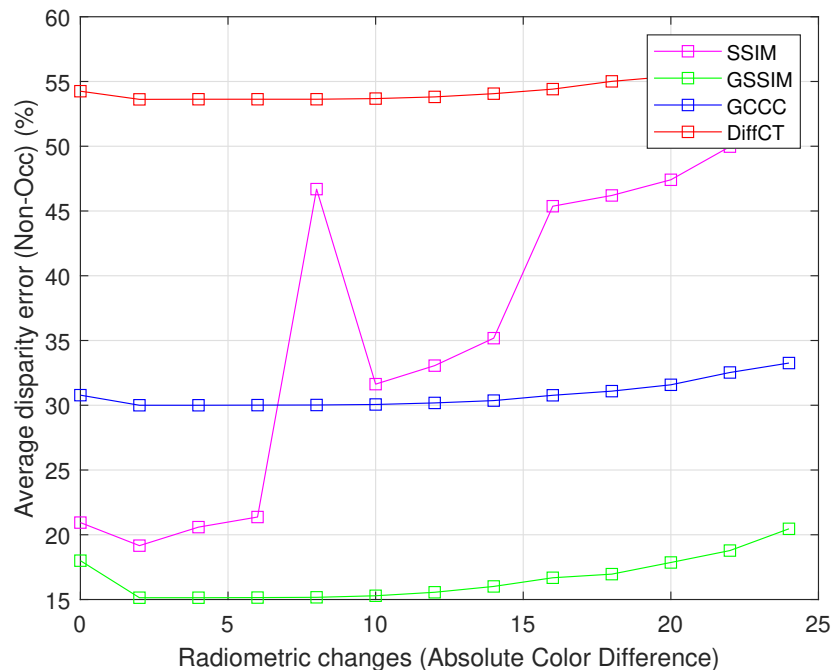
In this section, we study the impact of radiometric distortions on different cost functions. These distortions are generated while using the absolute color difference between corresponding pixels [22]. At each level of radiometric distortion, we compute the mean disparity errors for all KITTI training set for  $C_{SSIM}$ ,  $C_{GSSIM}$ ,  $C_{DIFFCensus}$  [22], and  $C_{GCCC}$  [23] cost functions. It can be visualized from the Figure 3 that the proposed cost  $C_{GSSIM}$  give the lowest error rate at all radiometric distortion levels.

4.4. Discussion

In the literature, it has been proven that cost functions based on pixel intensities are very sensitive to radiometric changes. In this paper, new intensity based cost functions have been proposed. It takes the local intensity, luminance, and contrast into account, which provide a significant local information to describe the considered pixel within a support window. This new consideration provides the ability of the proposed cost function to deal with radiometric changes (see Figure 3). The results described in Tables 2 and 3 demonstrate that the proposed cost functions outperform the top performer, in both KITTI 2012 KITTI 2015 datasets, compared to  $C_{DiffCensus}$  and  $C_{GCCC}$  costs. Although these latter



promise better results with aggregation techniques, the aggregation costs proposed have led to the best results (see Tables 2 and 3). It must be noted that the overall performance of the proposed cost functions depends on support window size. It can be seen that both cost functions performs well as the size of the support region increases, as shown in Figure 2. This is trivial since large support regions hold sufficient information to more accurately describe the considered patch, and then lead to good accurate initial cost functions.



**Figure 3.** The disparity results obtained with respect to radiometric distortions for all of the presented cost functions, for KITTI 2012 training sets.

## 5. Conclusions

In this paper, we presented a new stereo matching algorithm with a new structural information based cost functions for the cost computation step. Thus, two cost functions were proposed and evaluated using real road scenes from the challenging KITTI 2012 and KITTI 2015 training datasets. The obtained results have demonstrated that both cost functions lead to the lowest disparity mean errors as compared to the top performer in this data set under different scenarios, which has proven that our cost functions are more robust to radiometric distortions than conventional cost functions. The evaluation of the proposed local stereo matching algorithm using the best performing cost function over the current state-of-the-art algorithms has demonstrated the potential merits of the proposed stereo similarity measurement.

**Author Contributions:** Conceptualization, O.Z., M.R.; funding acquisition, O.Z.; investigation, O.Z., M.R., A.A. and C.D.; methodology, O.Z., A.A.; project administration, M.R. and C.D.; supervision, A.A., M.R. and C.D.; writing—original draft, Z.O; writing—review—editing, O.Z., M.R., A.A. and C.D. All authors have 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. Foix, S.; Alenya, G.; Torras, C. Lock-in time-of-flight (tof) cameras: A survey. *IEEE Sens. J.* **2011**, *11*, 1917–1926. [[CrossRef](#)]
2. Schwarz, B. Mapping the world in 3D. *Nat. Photonics* **2010**, *4*, 429–430. [[CrossRef](#)]

3. Hsieh, Y.; McKeown, D.; Perlant, F. Performance evaluation of scen registration and stereo matching for artographic feature extraction. *IEEE Trans. Pattern Anal. Mach. Intell.* **1992**, *14*, 214–238. [[CrossRef](#)]
4. Vincent, E.; Laganière, R. Detecting and matching feature points. *J. Vis. Commun. Image Represent.* **2005**, *16*, 38–54. [[CrossRef](#)]
5. Veksler, O. Stereo correspondence by dynamic programming on a tree. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–26 June 2005; Volume 2, pp. 384–390.
6. Sun, J.; Zheng, N.N.; Shum, H.Y. Stereo matching using belief propagation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 787–800.
7. Kolmogorov, V.; Zabih, R. Computing visual correspondence with occlusions using graph cuts. In Proceedings of the Eighth IEEE International Conference on Computer Vision, Vancouver, BC, Canada, 7–14 July 2001; Volume 2, pp. 508–515.
8. Rhemann, C.; Hosni, A.; Bleyer, M.; Rother, C.; Gelautz, M. Fast cost-volume filtering for visual correspondence and beyond. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 504–511.
9. Yoon, K.J.; Kweon, I.S. Adaptive support-weight approach for correspondenc search. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 650–656. [[CrossRef](#)] [[PubMed](#)]
10. Zhang, K.; Lu, J.; Lafruit, G. Cross-based local stereo matching using orthogonal integral images. *IEEE Trans. Circuits Syst. Video Technol.* **2009**, *19*, 1073–1079. [[CrossRef](#)]
11. Scharstein, D.; Szeliski, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vis.* **2002**, *47*, 7–42. [[CrossRef](#)]
12. Geiger, A.; Lenz, P.; Urtasun, R. Are we ready for autonomous driving? the kitti vision benchmark suite. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2020; pp. 3354–3361
13. Menze, M.; Geiger, A. Object scene flow for autonomous vehicles. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3061–3070.
14. Zabih, R.; Woodfill, J. Non-parametric local transforms for computing visual correspondence. In Proceedings of the Third European Conference on Computer Vision, Stockholm, Sweden, 2–6 May 1994; Volume 2, pp. 151–158.
15. Hirschmuller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 328–341. [[CrossRef](#)] [[PubMed](#)]
16. Klaus, A.; Sormann, M.; Karner, K. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In Proceedings of the 18th International Conference on Pattern Recognition, Hong Kong, China, 20–24 August 2006; Volume 3, pp. 15–18.
17. Mei, X.; Sun, X.; Zhou, M.; Jiao, S.; Wang, H.; Zhang, X. On building an accurate stereo matching system on graphics hardware. In Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, Spain, 6–13 November 2011; pp. 467–474.
18. Stentoumis, C.; Grammatikopoulos, L.; Kalisperakis, I.; Karras, G. On accurate dense stereo-matching using a local adaptive multi-cost approach. *ISPRS J. Photogramm. Remote Sens.* **2014**, *91*, 29–49. [[CrossRef](#)]
19. Saygili, G.; van der Maaten, L.; Hendriks, E.A. Adaptive stereo similarity fusion using confidence measures. *Comput. Vis. Image Underst.* **2015**, *135*, 95–108. [[CrossRef](#)]
20. Hirschmuller, H.; Scharstein, D. Evaluation of cost functions for stereo matching. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8.
21. Hirschmuller, H.; Scharstein, D. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 1582–1599. [[CrossRef](#)] [[PubMed](#)]
22. Miron, A.; Ainouz, S.; Rogozan, A.; Benschair, A. A robust cost function for stereo matching of road scenes. *Pattern Recognit. Lett.* **2014**, *38*, 70–77. [[CrossRef](#)]
23. Zeglazi, O.; Rziza, M.; Amine, A.; Demonceaux, C. Accurate dense stereo matching for road scenes. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 720–724.
24. Xu, Y.; Long, Q.; Mita, S.; Tehrani, H.; Ishimaru, K.; Shirai, N. Real-time stereo vision system at nighttime with noise reduction using simplified non-local matching cost. In Proceedings of the 2016 IEEE Intelligent Vehicles Symposium (IV), Gothenburg, Sweden, 19–22 June 2016; pp. 998–1003.

25. Buades, A.; Coll, B.; Morel, J.-M. A non-local algorithm for image denoising. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; pp. 60–65.
26. Fei, L.; Yan, L.; Chen, C.; Ye, Z.; Zhou, J. OSSIM: An Object-Based Multiview Stereo Algorithm Using SSIM Index Matching Cost. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6737–6949. [[CrossRef](#)]
27. 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](#)]



© 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/>).