

RANGE IMAGE QUALITY ASSESSMENT BY STRUCTURAL SIMILARITY

W.S. Malpica

A.C. Bovik

The University of Texas at Austin
Laboratory for Image and Video Engineering
willmalpica@gmail.com

The University of Texas at Austin
Laboratory for Image and Video Engineering
bovik@ece.utexas.edu

ABSTRACT

We propose a new quality metric for range images that is based on the multi-scale Structural Similarity (MS-SSIM) Index. The new metric operates in a manner to SSIM but allows for special handling of missing data. We demonstrate its utility by reevaluating the set of stereo algorithms evaluated in the Middlebury Stereo Vision Page <http://vision.middlebury.edu/stereo/>. The new algorithm which we term Range SSIM (R-SSIM) Index possesses features that make it an attractive choice for assessing the quality of range images.

Index Terms— SSIM, Quality Assessment, Range Images, Structural Similarity

1. INTRODUCTION

Currently there is a lack of methods to objectively assess the quality of range images. Range images of every type, whether they are acquired directly using a range finding device, indirectly through structured lighting [1] or computed from stereo images, can vary significantly in their quality when compared to a ground truth. It is useful to be able to quantify the quality of these range images in order to benchmark range-finding devices and stereo algorithms. In previous work the RMS (root mean squared) error [2] or percentage of bad pixels [2][3][4] has been used to estimate the quality of range images in order to evaluate the performance of stereo algorithms.

Recently there have been important advances in (luminance) image quality assessment. The standard quality metrics used in the past such as mean squared error (MSE) are being replaced by metrics such as SSIM [5], which match better with human subjectivity. Other successful but more complex metrics, such as the Visual Information Fidelity (VIF) Index [6] have also been developed, using models of the human visual system (HVS) and natural scene statistics (NSS).

This paper proposes a new measure, termed R-SSIM Index, which uses of a modified version of the Multi-Scale SSIM [7] Index, but specially designed for range images. Range images, bear both many similarities and differences with luminance images. When applying the SSIM algorithm to range images, the three similarity components of SSIM,

that is, luminance, contrast and structure, find their counterparts in the range domain as depth, surface roughness and 3D structure.

Computational stereo is becoming more prevalent and commonly utilized in applications such as robot navigation and face recognition. In [2] Schartein and Szeliski created a ranking of many recent stereo algorithms. This work has been continued and improved and is available on their website: <http://vision.middlebury.edu/stereo/> [8]. To demonstrate the utility of the R-SSIM Index, the set of stereo algorithms covered in [2] and their website were reevaluated and given a new set of rankings using R-SSIM. We propose R-SSIM as an alternative or supplementary approach to assessing range image quality.

2. THE SSIM ALGORITHM

The Structural Similarity Index was first proposed in [5]. Since its initial publication, the algorithm has gained popularity and acceptance and several variations have been developed. The algorithm's greatest appeal is that it matches human subjectivity. In particular the SSIM Index, like the HVS, highly sensitive to degradations in the spatial structure of image luminances.

The basic SSIM algorithm requires that the two images being compared be properly aligned and scaled so they can be compared point by point. The computations are performed in a sliding $N \times N$ (typically 11×11) gaussian-weighted window. Three similarity functions are computed on the windowed image data: luminance similarity, contrast similarity, and structural similarity, which for two images X and Y are calculated as follows:

$$l(x, y) = \frac{2\mu_X(x, y)\mu_Y(x, y) + C_1}{\mu_X^2(x, y) + \mu_Y^2(x, y) + C_1} \quad (1)$$

$$c(x, y) = \frac{2\sigma_X(x, y)\sigma_Y(x, y) + C_2}{\sigma_X^2(x, y) + \sigma_Y^2(x, y) + C_2} \quad (2)$$

$$s(x, y) = \frac{\sigma_{XY}(x, y) + C_3}{\sigma_X(x, y) + \sigma_Y(x, y) + C_3} \quad (3)$$

where

$$\mu(x, y) = \sum_{p=-P}^P \sum_{q=-Q}^Q w(p, q) X(x + p, y + q) \quad (4)$$

$$\sigma^2(x, y) = \sum_{p=-P}^P \sum_{q=-Q}^Q w(p, q) [X(x + p, y + q) - \mu_X(x, y)]^2 \quad (5)$$

$$\sigma_{XY}(x, y) = \sum_{p=-P}^P \sum_{q=-Q}^Q w(p, q) \cdot [X(x + p, y + q) - \mu_X(x, y)][Y(x + p, y + q) - \mu_Y(x, y)] \quad (6)$$

where $w(p, q)$ is a Gaussian weighing function such that $\sum_{p=-P}^P \sum_{q=-Q}^Q w(p, q) = 1$, and C_1 , C_2 and C_3 are small constants that provide stability when the denominator approaches zero. Typically

$$C_1 = (K_1 L)^2, C_2 = (K_2 L)^2 \text{ and } C_3 = C_2/2 \quad (7)$$

where L is the dynamic range of the image and $K_1 \ll 1$ and $K_2 \ll 1$ are small scalar constants. The three similarity functions are then combined into the general form of the SSIM index:

$$\text{SSIM}(x, y) = [l(x, y)] \cdot [c(x, y)] \cdot [s(x, y)] \quad (8)$$

3. MS-SSIM

Since its conception, the SSIM Index has seen various modifications, improvements and applications. One of these is the Multi-Scale SSIM or MS-SSIM Index [7]. It utilizes the same basic algorithm except that it operates over scales. The reference and distorted images are iteratively driven through a low-pass filter and down sampled by factor of two. The resulting image pairs are processed with the SSIM algorithm and multiplied together.

$$\text{MS-SSIM}(x, y) = [l_M(x, y)]^{\alpha_M} \cdot \prod_{j=1}^M [c_j(x, y)]^{\beta_j} \cdot [s_j(x, y)]^{\gamma_j} \quad (9)$$

Here M is the finest scale obtained after $M - 1$ scaling iterations. $l_j(x, y)$, $c_j(x, y)$ and $s_j(x, y)$ are the luminance, contrast and structure components at their different scales. In [7], α_j , β_j and γ_j are set according to scale so they match the contrast sensitivity function of the HVS. For the purposes of this paper and the R-SSIM metric they will be set $\alpha_j = \beta_j = \gamma_j$ and $\sum_{j=1}^M \gamma_j = 1$. The MS-SSIM algorithm compares details across resolutions, providing overall improved image quality assessment, as shown in the massive statistical study detailed in [9].

4. A NEW INDEX FOR RANGE IMAGES

When translating SSIM from intensity images into the range image domain, the three similarity subcomponents find their analog in the range domain. The luminance component becomes a function of mean depth which is a meaningful element in describing a range map. The contrast component can be interpreted as surface roughness. Most importantly, the structure component captures 3-D structure such as discontinuities, depth singularities, detail, 3D shape, and so on.

The one aspect which ordinary SSIM does not handle properly in range maps are unknown regions or missing data. In most applications, range images are output from stereo algorithms, obtained through range finding devices or from structured lighting techniques. These methods often produce unknown regions, e.g. regions where a stereo algorithm fails to compute a depth estimate or where the range finder is shadowed by an obstruction. These unknown regions must be handled appropriately in order to obtain an accurate score from the quality metric. The R-SSIM algorithm is a variation of the MS-SSIM algorithm with the ability to handle unknown regions or missing data.

R-SSIM handles unknown regions differently depending on if they are on the reference image or the distorted image. Pixels in the unknown regions of the reference image are ignored in all R-SSIM calculations. Pixels in the unknown region of the distorted image are ignored when they fall inside the sliding window used to calculate the value $SSIM(x, y)$ of its neighboring pixels, but the value $SSIM(x, y)$ of the unknown pixels themselves are set to zero. Figure 1(a) depicts an 11x11 patch in the reference image where there are some unknown pixels (shown in black). Figure 1(b) shows the same patch in the computed range image which also contains unknown pixels. Figure 1(c) shows the Gaussian weighing function and Figure 1(d) shows it masked by the unknown regions and renormalized. Finally Figure 1(e) shows a map of the SSIM values of that patch, where the pixel in the middle was the one calculated from Figures 1(a), 1(b) and 1(d). In Figure 1(e), the same unknown region from Figure 1(b) is indicated in black with a SSIM score of zero, while the unknown region in Figure 1(a) will be ignored in the final R-SSIM score.

5. EVALUATING STEREO ALGORITHMS USING R-SSIM

Computational stereo is one of the most actively researched fields in computer vision, and new stereo algorithms are being continuously developed. A comparative evaluation is useful in gauging the performance of these algorithms as well as monitoring the progress of the field. Scharstein and Szeliski [2] published a paper performing a taxonomy and evaluation of competitive stereo algorithms. In their evaluation, they used two different quality metrics based on a known ground

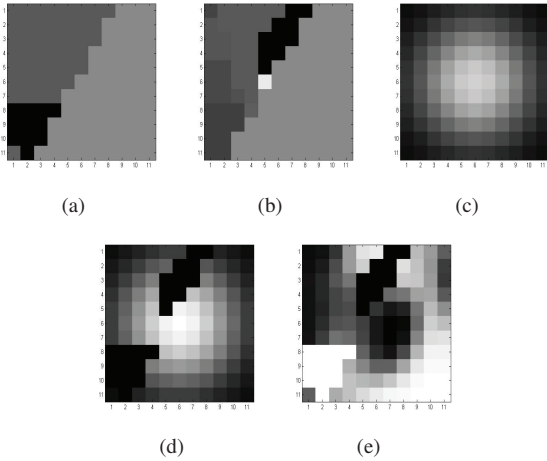


Fig. 1. Explanation of R-SSIM Index. See text for explanation.

truth. The RMS (root mean squared) error computed between the disparity map d_C and the ground truth map d_T :

$$R = \left(\frac{1}{N} \sum_{(x,y)} |d_C(x,y) - d_T(x,y)|^2 \right)^{1/2} \quad (10)$$

where N is the total number of pixels, and the percentage of bad matching pixels:

$$B = \frac{1}{N} \sum_{(x,y)} (|d_C(x,y) - d_T(x,y)| > \delta_d) \quad (11)$$

where δ_d is a disparity error tolerance, which is set to δ_d in this paper.

The authors continue to evaluate new algorithms on their website: <http://vision.middlebury.edu/stereo/> which now contains 34 algorithms. In their online evaluation they run each stereo algorithm on four different image pairs and compare the results against a ground truth. They only use percentage of bad pixels as a quality measure. The percentage of bad pixels is evaluated for each of the four images in three different regions, as seen in Figure 2. The first region covers all regions known in the ground truth, the second region covers all regions which are not occluded and the third region covers all the areas which are near depth discontinuities and near occlusions. The results of the evaluation ranked the algorithms according to their performance. The average ranking was taken from the mean rank from the different regions and images.

In order to demonstrate the utility of the R-SSIM algorithm, the same stereo algorithm results were evaluated using the R-SSIM Index instead of the percentage of bad pixels. The results are shown in Table 1. Correlating the results, displayed in Figure 3 it can be observed that the two metrics correlate well. This demonstrates that the R-SSIM Index is

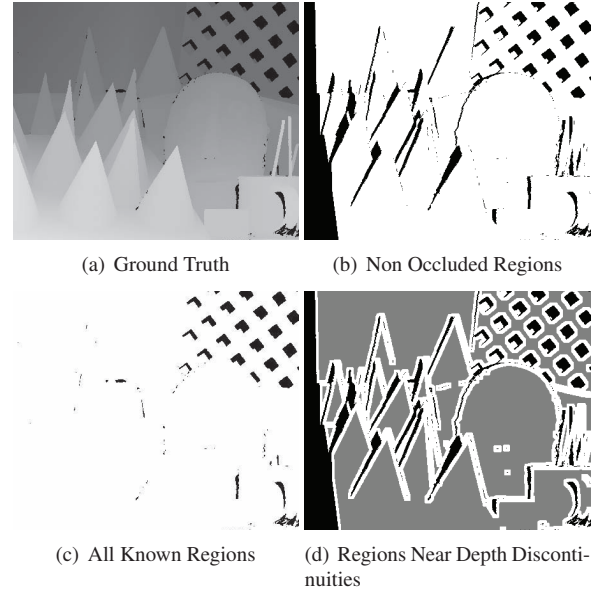


Fig. 2. Images from Middlebury stereo data set. See text for explanation.

sensitive to the distortions that the Middlebury rankings assess. However, the R-SSIM Index measures more than loss of depth values, since it also is sensitive to errors in depth, roughness, and 3-D surface structure, which can only be measured from local image patches, as opposed to single pixels.

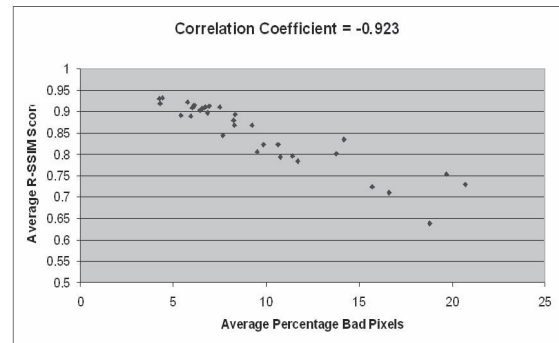


Fig. 3. Correlations between R-SSIM Index values and percentage of bad pixels on the Middlebury dataset.

Figure 4 shows that the two methods can give very different rankings to the same images. The two algorithms being assessed are graded in nearly reverse rank-order by the two assessment methods. Visual inspection of the two images suggests that in this instance, the R-SSIM algorithm delivers a more meaningful assessment of the quality of the computed range maps.

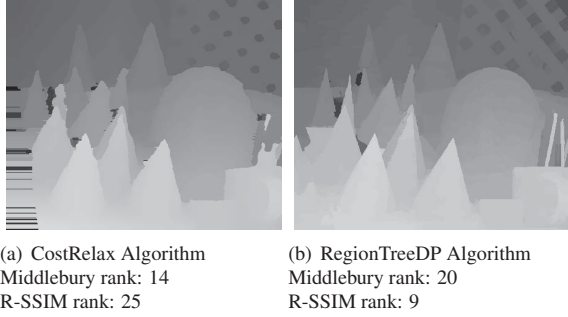


Fig. 4. Comparison of the results of two computational stereo algorithms and their Middlebury and R-SSIM rankings. In this case the rankings diverge significantly.

6. CONCLUSION

We have proposed R-SSIM as a new and needed quality metric for range images. We have demonstrated its utility by evaluating the 34 stereo algorithms in the Middlebury Stereo Vision Page. Through their evaluation and comparison of stereo algorithms Scharstein and Szeliski [2] have continue to provide a valuable resource to the field of computer vision. We believe that the R-SSIM algorithm is an effective method for range image fidelity assessment which complements their evaluations in a beneficial way.

7. ACKNOWLEDGMENTS

The research in this paper was sponsored by the Air Force Research Laboratory (AFRL).

8. REFERENCES

[1] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," *CVPR*, vol. 1, no. 7, pp. 195–202, June 2003.

[2] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, , no. 47, April-June 2002.

[3] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," *ICCV*, vol. II, pp. 508–515.

[4] C.L. Zitnick and T. Kanade, "A cooperative algorithm for stereo matching and occlusion detection," *IEEE Trans Pattern Anal Machine Intell*, vol. 22, no. 7, pp. 675–684, 2000.

[5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to struc-

Table 1. Calculated R-SSIM Index values and rankings and Middlebury rankings on stereo image dataset

Stereo Algorithms	Avg. Rank Middlebury	Avg. Rank R-SSIM	Avg. Score Middlebury	Avg. Score R-SSIM
AdaptWeight	12.90	11.33	6.66	0.91
AdaptingBP	2.10	4.42	4.24	0.93
C-SemiGlob	9.10	5.58	5.76	0.92
CostRelax	22.30	22.67	10.61	0.82
DP	28.30	22.83	14.18	0.83
DistinctSM	10.10	9.17	6.14	0.92
DoubleBP	3.20	7.67	4.30	0.92
EnhancedBP	11.90	10.92	6.70	0.91
GC+occ	18.20	16.92	8.27	0.87
GC	24.70	25.67	11.42	0.80
GenModel	21.20	25.75	9.50	0.81
ImproveSubPix	13.80	8.83	6.90	0.91
Infection	32.60	30.42	20.74	0.73
Layered	18.40	16.75	8.24	0.88
MultiCamGC	18.80	13.83	8.30	0.89
OverSegmBP	10.60	7.75	6.11	0.91
PhaseDiff	31.90	32.42	18.77	0.64
ProgrEdgeBased	27.00	23.83	13.78	0.80
RealTimeGPU	21.30	24.83	9.83	0.82
RealtimeBP	17.00	20.83	7.68	0.84
RegionTreeDP	11.00	11.75	6.56	0.91
ReliabilityDP	23.40	27.67	10.73	0.79
SO+borders	8.80	11.42	6.03	0.91
SO	31.30	31.25	16.60	0.71
SSD+MF	30.00	30.58	15.68	0.72
STICA	30.90	28.67	19.68	0.75
SegTreeDP	13.20	14.42	6.82	0.90
Segm+visib	8.80	12.08	5.40	0.89
SegmentSupport	10.50	13.33	6.44	0.90
SemiGlob	14.90	8.92	7.51	0.91
SubPixDoubleBP	4.00	3.67	4.39	0.93
SymBP+occ	8.00	13.17	5.92	0.89
TensorVoting	21.20	19.00	9.23	0.87
TreeDP	23.80	26.67	11.68	0.78

tural similarity," *IEEE Trans. Image Processing*, vol. 13, pp. 600–612, Apr 2004.

[6] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Processing*, vol. 15, pp. 430–444, Feb 2006.

[7] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers*, Nov 2003.

[8] "Middlebury stereo vision page," October 2007, <http://vision.middlebury.edu/stereo/>.

[9] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans Image Processing*, vol. 15, pp. 3440–3451, Nov 2006.