SSIM BASED RANGE IMAGE QUALITY ASSESSMENT

W.S. Malpica

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

A.C. Bovik

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

ABSTRACT

Although image quality assessment (IQA) has observed much research and advancement in the past, not much progress has been made towards the development of quality assessment metrics for range images. The current metrics being used to assess the quality of range images are reminiscent to the old IQA metrics such as mean squared error that have since been replaced with metrics which better correlate with human perception. This paper presents two new algorithms for range image quality assessment (RIQA). The first, R-SSIM, is based off the Multi-Scale Structural Similarity Index (MS-SSIM), while the other is an extension of the Complex Wavelet SSIM (CW-SSIM) metric for use in RIQA. The utility of these metrics is demonstrated through a re-evaluation of Scharstein and Szeliski's online evaluation of dense two-frame stereo correspondence algorithms.

1. INTRODUCTION

Image quality assessment (IQA) has seen much progress in recent years, but quality assessment for range images has been mostly ignored. Standard IQA algorithms cannot be directly applied to range images due to their fundamental differences. Also referred to as depth images, range images depict a scene, not in terms of luminance or color, but instead each pixel represents the distance between that point in the scene and the camera. Most range images are generated using either, laser scanning, active triangulation or passive stereo.

Laser scanning methods scan a scene with a range finding laser to determine the depth at every pixel in the scene. On the other hand, active triangulation systems project a beam or a sheet of light onto a scene. This light reflecs off the scene and is captured by a sensor, usually a CCD camera set at an offset. Since the position and angles of the source and sensor are known, the distance between the source and the point of reflection can be triangulated. Finally, passive stereo uses two images of the same scene which are taken by two cameras set at an offset from each other. This system calculates the range distances using the same geometric principles as active triangulation, except it triangulates the range from matching pixels in the two images rather than between the known source and a pixel. All of these methods have limitations which cause them to produce regions in the image which contain unknown range data. Laser range finders and active triangulation methods can both suffer from specular reflections where the light from the source does not reflect back to the sensor and hence result in an unknown depth measurement. Active triangulation and passive stereo tend to encounter occlusions, which also generate unknown regions [1].

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. These metrics are similar to the standard quality metrics used in the past for normal luminance images, such as mean squared error (MSE), and they need to be replaced by better metrics. MSE for example has been 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, which uses of a modified version of the Multi-Scale SSIM [7] Index, but specially designed for range images. The paper also introduces an extension of the Complex Wavelet SSIM (CW-SSIM) [8] metric for use in RIQA. Range images, bear 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. R-SSIM also takes special consideration to regions in the images which contain missing data.

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/ [9]. To demonstrate the utility of R-SSIM and CW-SSIM, the set of stereo algorithms covered in [2] and their website were reevaluated and given a new set of rankings using the proposed metrics. We propose these metrics 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 of the algorithm have been developed. The algorithm's greatest appeal is that it matches human subjectivity. In particular, both the SSIM Index and the HVS are 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 NxN (typically 11x11) 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

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

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

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

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

The Multi-Scale SSIM or MS-SSIM Index [7] is one of the most popular variations on the SSIM Index. It utilizes the same basic algorithm except that it operates over several 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.

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

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 [10].

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. Finaly, 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 are the unknown regions or missing data usually found in range maps. 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 these unknown regions.

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 $SSIM(x, y)$ value from its neighboring pixels, but the $SSIM(x, y)$ value 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.

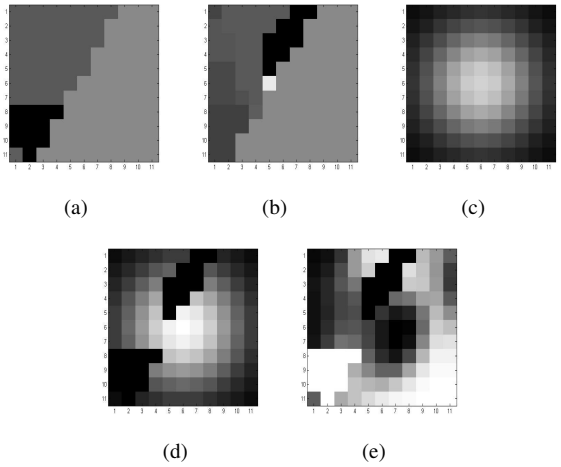


Fig. 1. Explanation of R-SSIM Index. See text for explanation.

5. CW-SSIM A ROBUST ALTERNATIVE FOR R-SSIM

The complex wavelet SSIM or CW-SSIM [8] is a powerful variation of the SSIM algorithm. It extends the SSIM algorithm into the 'complex wavelet' transform domain. The CW-SSIM algorithm uses two sets of coefficients $c_x = \{c_{x,i} \mid i = 1, \ldots, N\}$ and $c_y = \{c_{y,i} \mid i = 1, \ldots, N\}$ extracted at the same spatial location in the same wavelet subbands of the two images being compared. These coefficients are zero mean, due to the bandpass nature of the wavelet filters, which means that the luminance component of the SSIM algorithm is 1 leaving the SSIM algorithm as:

$$\text{SSIM}(x, y) = \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{10}$$

Converting that formula into the complex wavelet transform domain gives the CW-SSIM formula:

$$\text{CW-SSIM}(c_x, c_y) = \frac{2|\sum_{i=1}^{N} c_{x,i} c_{y,i}^*| + K}{\sum_{i=1}^{N} |c_{x,i}|^2 + \sum_{i=1}^{N} |c_{y,i}|^2 + K} \tag{11}$$

where K is a small constant. To better understand the CW-SSIM index, it can be rewritten as the product of two components:

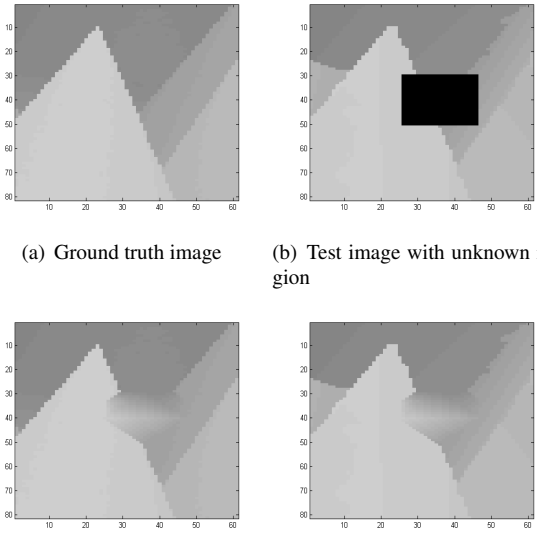
$$\text{CW-SSIM}(c_x, c_y) = \frac{2\sum_{i=1}^{N} |c_{x,i}||c_{y,i}| + K}{\sum_{i=1}^{N} |cx, i|^2 + \sum_{i=1}^{N} |c_{y,i}|^2 + K} \cdot \frac{2|\sum_{i=1}^{N} c_{x,i} c_{y,i}^*| + K}{2\sum_{i=1}^{N} |c_{x,i} c_{y,i}^*| + K} \tag{12}$$

The first component is completely determined by the magnitude of the coefficients, while the second components is solely determined by the consistency of phase changes between $c_{x,i}$ and $c_{y,i}$. Note that the index is not affected by a consistent phase shift.

The structural information of local image features is mainly contained in the relative phase patterns of the wavelet coefficients, therefore a consistent phase shift does not affect the structure of local features. This means the metric becomes more robust to translation, rotation and scaling as long as these distortions are small relative to the size of the wavelet filters. Images do not have to be perfectly registered for this metric to provide a good assessment of the image quality. This is useful in the cases where the ground truth and the range images under evaluation were obtained using different methods, i.e. a ground truth obtained via active triangulation is used to evaluate stereo algorithms. In such a case there may be slight misregistrations between the ground truth range image and the range image being evaluated.

Unfortunately CW-SSIM cannot ignore unknown regions common in range images like R-SSIM can. The wavelet coefficients used by CW-SSIM describe the whole image and cannot be calculated while ignoring unknown regions in the image. Instead of ignoring them, the proposed solution is to fill them in via interpolation, but in order to better match the images, the interpolation must also be performed on the opposing image in the same region as seen in Figure 2. The interpolation step, will inherently affect the accuracy of the quality assessment, and therefore a study was performed to determine the magnitude of the error introduced by the interpolation step. The study utilized several image pairs with no unknown regions and calculated the similarity using CW-SSIM. Unknown regions were then introduced, interpolated and followed by a second application of the CW-SSIM algorithm. The difference between the CW-SSIM scores showed that the greater the percentage of the image that is covered by unknown regions, the greater the difference between the scores. In the same manner, the larger the size of the unknown regions, the more

difficult it is for the interpolation to properly estimate for the missing pixel values and hence a greater difference between the scores. Overall though, the effect is negligible as long as the unknown regions filled in are relatively small. In the study performed, even when replacing ten percent of the image with unknown regions, the difference between the CW-SSIM score did not read 0.02 [11].



(a) Ground truth image (b) Test image with unknown region

(c) Ground truth image with interpolated region (d) Test image with interpolated region

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

6. EVALUATING STEREO ALGORITHMS USING R-SSIM AND CW-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 stereo algorithms. In their evaluation, they used two different quality metrics based on a known ground truth. The RMS (root mean squared) error computed between the disparity map d_C and the ground truth map d_T:

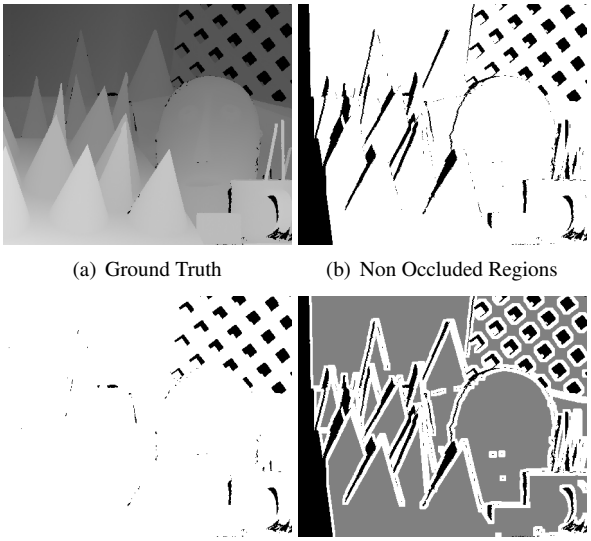
$$R = (\frac{1}{N} \sum_{(x,y)} |d_C(x,y) - d_T(x,y)|^2)^{1/2} \qquad (13)$$

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

$$B = \frac{1}{N} \sum_{(x,y)} (|d_C(x,y) - d_T(x,y)| > \delta_d) \qquad (14)$$

where δ_d is a disparity error tolerance, which is set to 1 in this paper.

The authors continue to evaluate new algorithms on their website: http://vision.middlebury.edu/stereo/ which now contains 39 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 3. 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.



(a) Ground Truth (b) Non Occluded Regions

(c) All Known Regions (d) Regions Near Depth Discontinuities

Fig. 3. Images from Middlebury stereo data set. See text for explanation.

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 4 it can be observed that the two metrics correlate well. This demonstrates that the R-SSIM Index is 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.

In the same manner CW-SSIM was also used to rank

all the stereo algorithms in the Middlebury evaluation. The only difference is that it was only run for one area: all known regions. CW-SSIM was not run on the other two areas because they would introduce too many unknown regions for the interpolation to handle appropriately. The results are shown in Table 1. Figure 5 shows a scatter plots comparing the CW-SSIM and percentage of bad pixels and a correlation coefficient between the metrics. The high value of correlation coefficient indicates that, like the R-SSIM metric, the CW-SSIM metric is also sensitive to the distortions that the Middlebury ranking assesses.

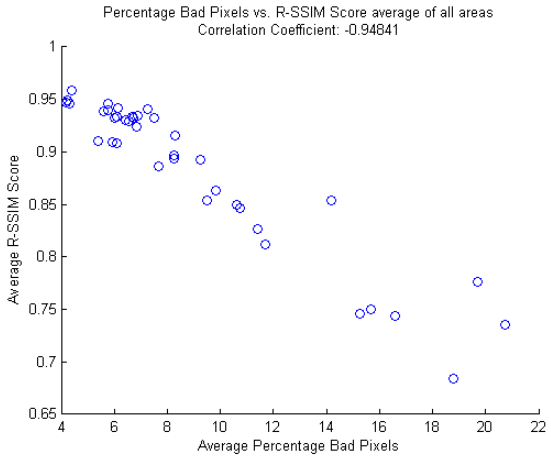


Fig. 4. Correlations between R-SSIM Index values and percentage of bad pixels on the Middlebury dataset.

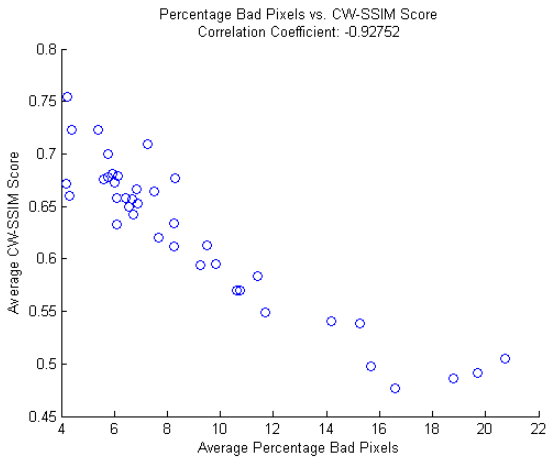
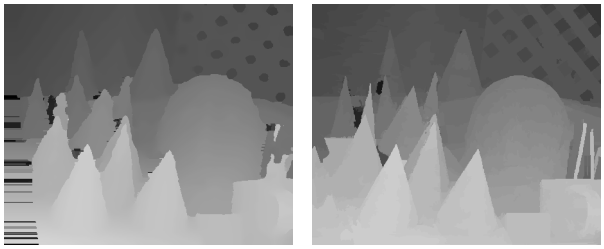


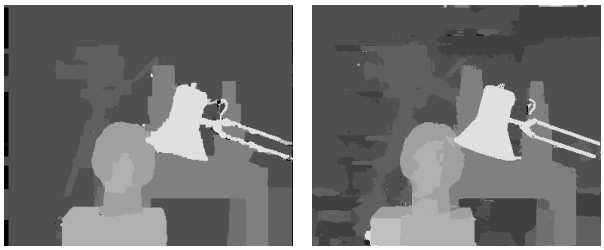
Fig. 5. Correlations between CW-SSIM Index values and percentage of bad pixels on the Middlebury dataset.

Figure 6 shows that the R-SSIM, CW-SSIM and percentage of bad pixels can give very different rankings to the same images. In the case of these particular stereo algorithms, the SSIM based metrics provide a completely different ranking than the Middlebury rankings. Visual inspection of the two images suggests that in this instance, the SSIM based metrics deliver a more meaningful assessment of the quality of the computed range maps.



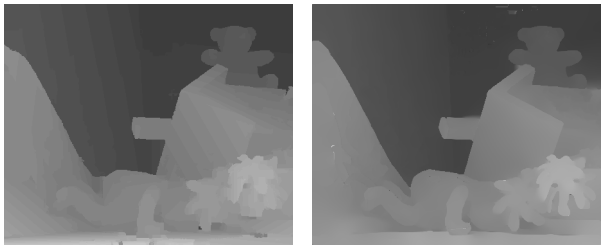
(a) CostRelax Algorithm
Middlebury rank/score: 18/10.2
R-SSIM rank/score: 23/0.893
CW-SSIM rank/score: 23/0.640

(b) RegionTreeDP Algorithm
Middlebury rank/score: 25/11.9
R-SSIM rank/score: 14/0.893
CW-SSIM rank/score: 17/0.664



(c) MultiCamGC Algorithm
Middlebury rank/score: 18/1.99
R-SSIM rank/score: 1/0.945
CW-SSIM rank/score: 2/0.863

(d) AdaptDispCalib Algorithm
Middlebury rank/score: 4/1.42
R-SSIM rank/score: 23/0.875
CW-SSIM rank/score: 24/0.649



(e) SymBP+occ Algorithm
Middlebury rank/score: 6/10.7
R-SSIM rank/score: 11/0.966
CW-SSIM rank/score: 11/0.701

(f) InteriorPtLP Algorithm
Middlebury rank/score: 9/11.9
R-SSIM rank/score: 4/0.974
CW-SSIM rank/score: 5/0.741

Fig. 6. Comparison of the results of six computational stereo algorithms and their Middlebury, R-SSIM and CW-SSIM rankings.

7. CONCLUSION

We have proposed R-SSIM as a new and needed quality metric for range images. We have also presented CW-SSIM as another alternative for quality assesment of range images, in particular when image robustness against slight rotation and translation is required. We have demonstrated their utility by evaluating the 39 stereo algorithms in the Middlebury

Stereo Vision Page. Through their evaluation and comparison of stereo algorithms Schartein and Szeliski [2] have continue to provide a valuable resource to the field of computer vision. We believe that the R-SSIM and CW-SSIM algorithms are an effective method for range image fidelity assessment which complements their evaluations in a beneficial way.

8. ACKNOWLEDGMENTS

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

Table 1. Average Middlebury, R-SSIM and CW-SSIM rankings on Middlebury stereo image dataset

Algorithm	Middlebury	R-SSIM	CW-SSIM
AdaptDispCalib	11.2	19.9	20.8
AdaptOvrSegBP	9.5	9.4	12.8
AdaptWeight	16.8	14.5	15.8
AdaptingBP	2.8	6.7	3.3
C-SemiGlob	11.8	7.9	8.0
CostRelax	26.9	27.8	27.5
DP	32.9	28.7	31.8
DistinctSM	13.5	11.3	13.0
DoubleBP	4.8	9.0	15.5
DoubleBP2	2.8	7.8	12.0
EnhancedBP	15.7	14.6	19.0
GC+occ	22.7	20.8	21.8
GC	29.3	29.2	25.5
GenModel	25.8	28.3	23.0
ImproveSubPix	16.9	12.2	17.3
Infection	37.4	35.4	34.8
InteriorPtLP	16.7	9.6	5.8
Layered	23.1	22.5	23.0
MultiCamGC	23.4	17.4	15.8
OverSegmBP	13.7	11.5	17.0
PhaseBased	34.2	35.4	29.5
PhaseDiff	37	37.0	32.5
PlaneFitBP	10.4	11.1	12.0
RealTimeGPU	26.1	28.5	27.3
RealtimeBP	21.5	23.5	22.5
RegionTreeDP	14.8	15.3	18.8
ReliabilityDP	27.9	30.1	29.8
SO+borders	12.2	15.0	11.8
SO	36.3	35.8	36.8
SSD+MF	34.6	34.8	35.3
STICA	35.8	33.3	35.0
SegTreeDP	16.7	17.5	16.8
Segm+visib	11.5	15.8	4.5
SegmentSupport	14.4	15.8	15.8
SemiGlob	18.2	12.3	15.3
SubPixDoubleBP	5.5	3.2	5.3
SymBP+occ	10.6	17.0	10.0
TensorVoting	25.9	23.1	26.8
TreeDP	28.6	31.3	31.8

9. REFERENCES

[1] D. A. Forsyth and J. Ponce, *Computer Vision, A Modern Approach*, Prentice Hall, 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 structural 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, "Multiscale structural similarity for image quality assessment," *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers*, Nov 2003.

[8] Z. Wang and E. P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," *IEEE Int. Conference on Acoustics, Speech, and Signal Processing*, pp. 573–576, 2005.

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

[10] 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.

[11] W. Malpica, "Range image quality assessment by structural similarity," M.S. thesis, University of Texas at Austin, 2008.