# ON THE PERFORMANCE OF DIBR METHODS WHEN USING DEPTH MAPS FROM STATE-OF-THE-ART STEREO MATCHING ALGORITHMS

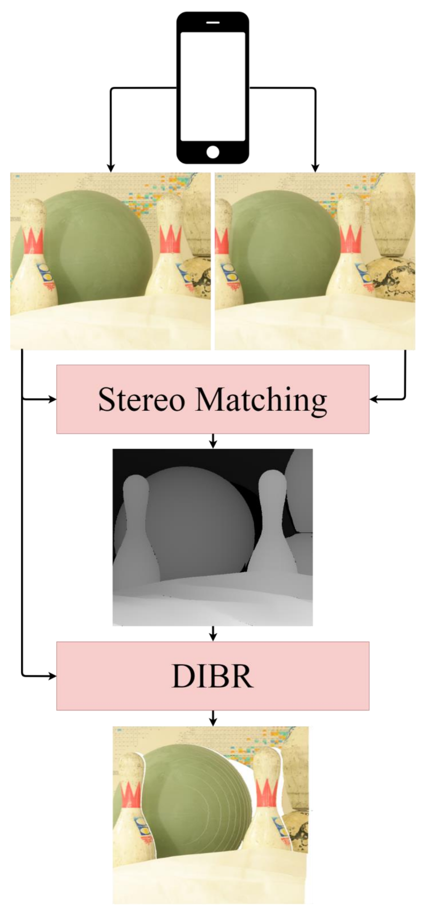

Adriano Q. de Oliveira, Thiago L. T. da Silveira, Marcelo Walter, Cláudio R. Jung

Federal University of Rio Grande do Sul (UFRGS) – Institute of Informatics (INF) – Porto Alegre – RS - Brazil

{aqoliveira, tltsilveira, marcelo.walter, crjung}@inf.ufrgs.br


## Introduction

- Stereo matching (SM) has been applied to several research-linked tasks such as robot navigation, surveillance and obstacle detection [1].
- 3D photography is a promising way for recording and storing view-point changing still images and videos.
- Depth-image-based rendering (DIBR) [2] is a view synthesis model, which uses as input a single color image and its associated depth map, and produces a novel synthesized view.
  - Cracks, ghosts, holes (disocclusions or out-of-field areas).
- There are several methods that address the DIBR problems [3, 4, 5, 6], however, these methods use ground-truth (GT) depth maps for both quantitative and qualitative assessment.




**The present study aims to evaluate the quality of the synthesized views produced by different DIBR approaches when fed with realistic disparity maps produced by SM approaches. Also, "Are the SM and view synthesis evaluation metrics correlated?".**

## Related Work

**Quality Assessment Works**

- Lu and colleagues [7] found that the root mean square (RMS) error of estimated disparity maps may not correlate with the quality of interpolated views.
- Fuhr *et al.* [8] concluded that the "number of bad pixels" in estimated disparity maps is weakly correlated to the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [9] measurements.

**Stereo Matching Algorithms**

- Based on a ranking obtained using a combination of four SM metrics (bad 2.0, avgerr, rms and a95) [10] we selected five algorithms with source code available:

**3rd:** A global optimization method based on Markov Random Field from Taniai *et al.* [11].

**6th:** An extension of standard belief propagation sequential technique applied to SM, developed by Mozerov and Weijer [12].

**21st:** A global SM algorithm that works on a 2D triangulation of the reference view from Zhang and others [13].

**39th:** A dictionary learning data-driven matching cost approach for comparing image patches proposed by Yin *et al.* [1].

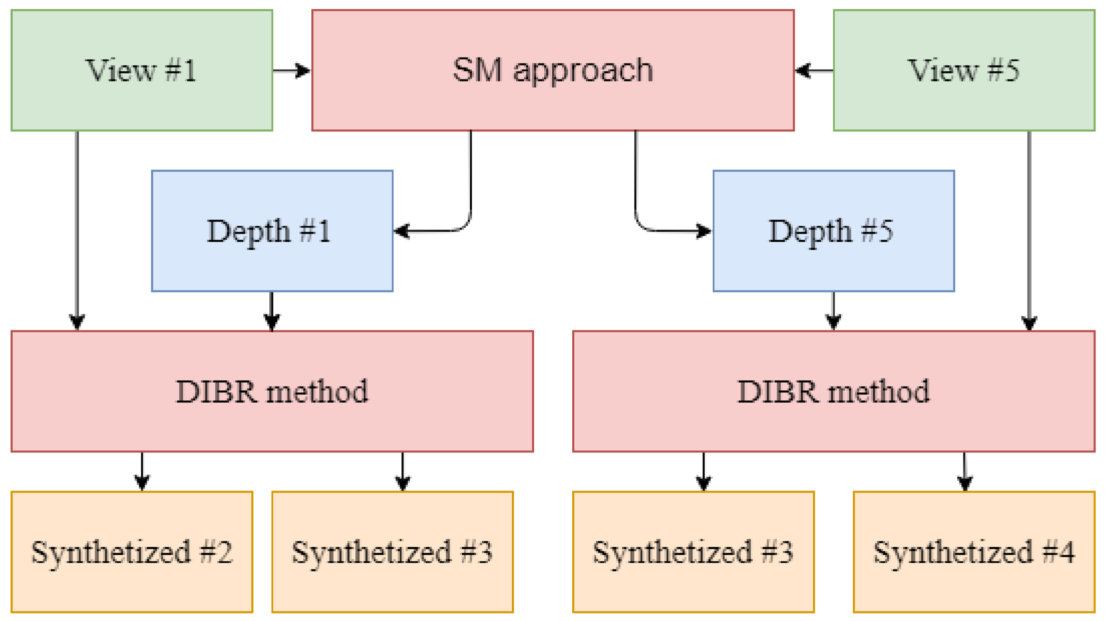
**42nd:** The method that explore the potential of cost filtering and energy minimization from Mozerov and Weijer [14].

**Depth-Image-Based Rendering Methods**

- The hierarchical hole filling algorithm [4] that uses a pyramid-like approach to estimate the hole pixels from lower resolution estimates of the image.
- The work from Ahn and Kim [3] that treats ghosts, empty cracks and holes with specific strategies.
  - The selective hole-filling method proposed in [6] identifies and corrects cracks and ghosts, and tackles larger holes by exploring depth in a patch-based inpainting scheme.
- Oliveira et al. [5] proposed a complete pipeline, with solutions that remove artifacts without excluding image content and apply different filling strategies according to the hole nature.

## Experimental Setup

- We evaluate the performance of all the selected DIBR methods when using depth estimated by all the considered SM algorithms (plus the GT disparity).
- For SM algorithms [1, 11, 12, 13, 14] we used the same metrics employed in the ranking, and for assessing synthesized views, we use PSNR, SSIM and also the context-specific morphological-wavelet PSNR (MW-PSNR) [15] in [3, 4, 5, 6].
- We used the multi-view half-sized image sets from Middlebury 2006 [10].

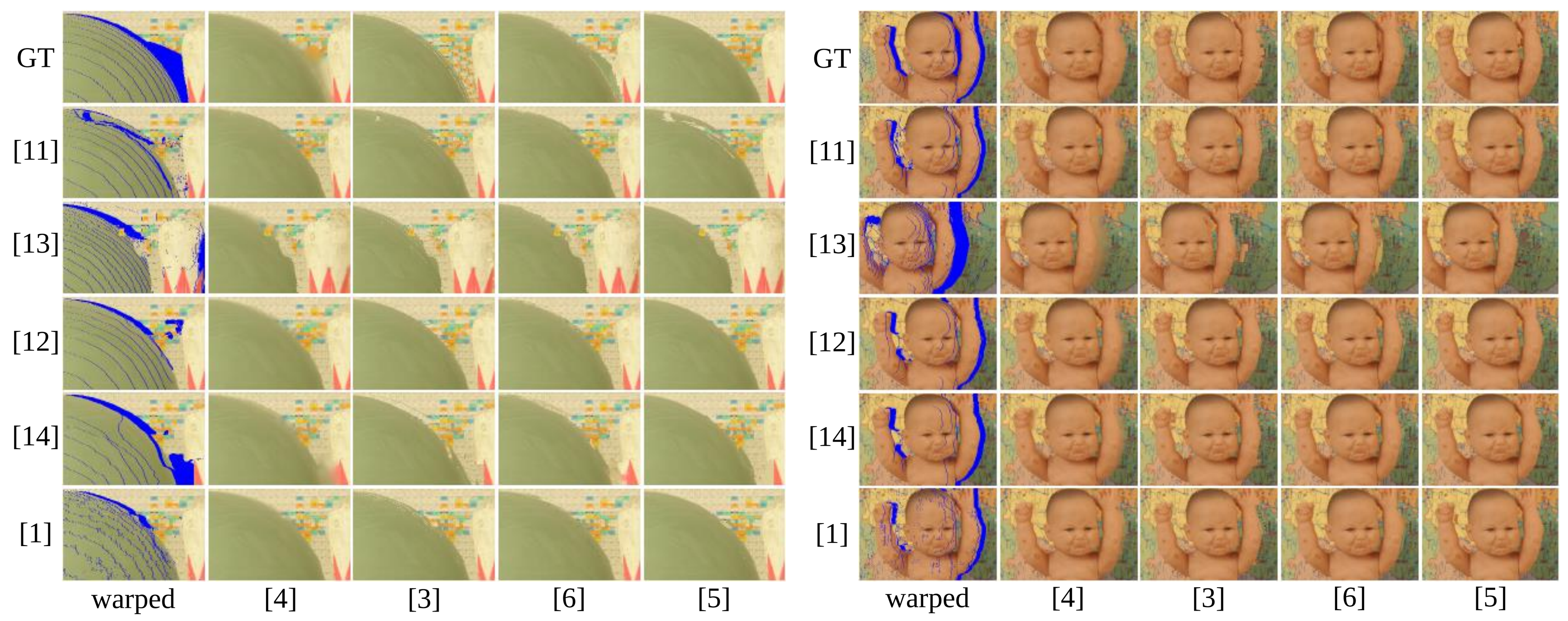



## Results and Discussion

- Based on the pipeline, we are able to compare the results varying (i) SM and (ii) DIBR techniques, being assessed via figures-of-merit such as (iii) bad 2.0, avgerr, rms and a95, and (iv) PSNR, SSIM and MW-PSNR (a four-dimensional hyper-cube).
- Note that using GT disparity maps does not lead to the best synthesized view (similar findings were also reported in [7]).
- The rankings based on PSNR, SSIM and MW-PSNR are inconsistent w.r.t. that based on the depth GT for methods [13], [11, 12, 14] and [1, 13], respectively.
- The relative ranking order between the considered SM algorithms according to bad 2.0, avgerr, rms and the combined score of all analyzed metrics is the same: [11], [12], [13], [1] and [14].
- The ranking order based on metric a95 is different: [12], [11], [13], [14] and [1].
- Spearman correlation [16] indicates that metrics bad 2.0 and MW-PSNR have a fairly strong negative relationship.
- Also the correlation indicates that it is not expected to have necessarily higher SSIM and PSNR values for synthesized views when we choose SM methods that minimize the error metrics bad 2.0, avgerr, rms and a95.

| | [14] | [1] | [13] | [12] | [11] | GT |
|---|---|---|---|---|---|---|
| | 24.334 | 31.960 | 16.883 | 31.143 | 30.918 | 29.273 |
| [4] | 0.7276 | 0.9532 | 0.5331 | 0.9527 | **0.9536** | 0.9472 |
| | 26.258 | **32.926** | 20.207 | 32.593 | 32.025 | 31.475 |
| | 24.548 | 31.335 | 18.254 | 31.024 | 30.128 | 28.713 |
| [3] | 0.7174 | 0.9393 | 0.5258 | 0.9427 | 0.9402 | 0.9290 |
| | 26.225 | 30.892 | 20.363 | 31.419 | 30.478 | 30.075 |
| | 24.685 | 32.196 | 18.230 | 31.793 | 31.811 | 30.066 |
| [6] | 0.7229 | 0.9450 | 0.5314 | 0.9496 | 0.9499 | 0.9422 |
| | 25.881 | 28.854 | 20.080 | 30.050 | 28.951 | 29.649 |
| | 24.859 | **32.626** | 18.292 | 32.052 | 32.041 | 31.721 |
| [5] | 0.7236 | 0.9512 | 0.5368 | 0.9517 | 0.9531 | 0.9522 |
| | 26.435 | 32.848 | 20.432 | 32.717 | 32.330 | 32.609 |

| | bad 2.0 | avgerr | rms | a95 |
|---|---|---|---|---|
| PSNR | $-0.39^{[12,6]}$ | $-0.37^{[12,6]}$ | $-0.34^{[13,6]}$ | $-0.40^{[12,6]}$ |
| SSIM | $-0.40^{[13,6]}$ | $-0.47^{[13,3]}$ | $-0.44^{[13,3]}$ | $-0.33^{[13,3]}$ |
| MW-PSNR | $-0.80^{[11,6]}$ | $-0.74^{[11,6]}$ | $-0.68^{[11,6]}$ | $-0.65^{[11,6]}$ |




- Cracks and ghosts tend to appear more intensely with SM-based depth maps.
- Disocclusion regions are contaminated with over-segmented depth layers due to inaccurate depth estimation.

## Conclusions

- DIBR methods can generate better results if using SM-based depth maps, instead of the ground-truth.
- DIBR techniques are ranked differently when fed by depth maps generated with SM algorithms or ground-truth depth.
- SM methods that minimize SM error measures do not necessarily result in better synthesized views according to SSIM and PSNR.
- MW-PSNR has a strong negative correlation to SM metrics and may be more useful for assessing DIBR methods than PSNR and SSIM.
- SM-based depth maps contain errors that mislead DIBR techniques, indicating that they may not be prepared for real scenario applications.

## References


[1] J. Yin, H. Zhu, D. Yuan, and T. Xue, "Sparse representation over discriminative dictionary for stereo matching," Pattern Recogn., vol. 71, pp. 278 – 289, 2017.
[2] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3DTV," in Stereoscopic Displays and Virtual Reality Systems XI, 2004, vol. 5291, pp. 93–104.
[3] I. Ahn and C. Kim, "A novel depth-based virtual view synthesis method for free viewpoint video," IEEE Trans. Broadcast, vol. 59, no. 4, pp. 614–626, 2013.
[4] M. Solh and G. AlRegib, "Hierarchical hole-filling for depth-based view synthesis in FTV and 3D video," IEEE J. Sel. Topics Signal Process., vol. 6, no. 5, pp. 495–504, 2012.
[5] A. Q. de Oliveira, M. Walter, and C. R. Jung, "An artifact-type aware DIBR method for view synthesis," IEEE Signal Process. Lett., vol. 25, no. 11, pp. 1705–1709, 2018.
[6] A. Q. de Oliveira, G. Fickel, M. Walter, and C. Jung, "Selective hole-filling for depth-image based rendering," in IEEE ICASSP, 2015, pp. 1186–1190.
[7] J. Lu, Q. Yang, and G. Lafruit, "Interpolation error as a quality metric for stereo: Robust, or not?," in IEEE ICASSP, 2009, pp. 977–980.
[8] G. Fuhr, G. P. Fickel, L. P. Dal'Aqua, C. R. Jung, T. Malzbender, and R. Samadani, "An evaluation of stereo matching methods for view interpolation," in IEEE ICIP, 2013, pp. 403–407.
[9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Trans. Image Process., vol. 13, no. 4, 2004.
[10] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in IEEE CVPR, 2007, pp. 1–8.
[11] T. Taniai, Y. Matsushita, Y. Sato, and T. Naemura, "Continuous 3D label stereo matching using local expansion moves," IEEE Trans. Pattern Anal. Mach. Intell., pp. 1–1, 2018.
[12] M. G. Mozerov and J. van de Weijer, "One-view occlusion detection for stereo matching with a fully connected crf model," IEEE Trans. Image Process., vol. 28, no. 6, pp. 2936–2947, 2019.
[13] C. Zhang, Z. Li, Y. Cheng, R. Cai, H. Chao, and Y. Rui, "Meshstereo: A global stereo model with mesh Alignment regularization for view interpolation," in IEEE ICCV, 2015, pp. 2057–2065.
[14] M. G. Mozerov and J. van de Weijer, "Accurate stereo matching by two-step energy minimization," IEEE Trans. Image Process., vol. 24, no. 3, pp. 1153–1163, 2015.
[15] D. Sandié-Stankovié, D. Kukolj, and P. Le Callet, "Multi-Scale Synthesized View Assessment Based on Morphological Pyramids," J. Ele. Eng., vol. 67, no. 1, pp. 3–11, 2016.
[16] C. Spearman, "The proof and measurement of association between two things," The American Journal of Psychology, vol. 15, no. 1, pp. 72–101, 1904.



This work was supported by: CAPES CNPq


ICASSP 2019