Fast Image Inpainting for DIBR View Synthesis Using Distance Transform and Gaussian Filtering

Emil Dumic[1]*, Anamaria Bjelopera*[2]*, Tina Pogac*[3]

[1] *Department of Electrical Engineering, University North*

Ul. 104. brigade 3, 42000 Varaždin, Croatia

[2] *Department of Electrical Engineering and Computing, University of Dubrovnik*

Ćira Carića 4, 20000 Dubrovnik, Croatia

[3] *Department of Multimedia, University North*

Ul. 104. brigade 3, 42000 Varaždin, Croatia

ABSTRACT

This paper presents a fast and lightweight image inpainting method using distance transform and Gaussian lowpass filtering on disoccluded regions for virtual view synthesis. Virtual views are created from one texture-plus-depth image and afterwards a described method is used for the hole filling process. Proposed algorithm is compared with several state-of-the-art image inpainting methods using different no-reference image quality measures (BRISQUE, NIQE, PIQE). Results for the proposed method show competitive performance, while having lower execution time, concluding that the proposed method can be used in different real-time scenarios.

Keywords: distance transform, Gaussian filter, inpainting, view synthesis, DIBR

INTRODUCTION

In the last few years 3D audiovisual content has been in focus of interest for many multimedia applications and services. The goal is to bring realism into visual scene to enrich the human perception of reality. There is a high need of development of 3D technologies in different fields such as entertainment, 3DTV, industry of games, medicine etc. Nowadays, evolution of 3D media technologies includes not only enhancement of existing technology but also creating new content-driven applications (Assunção and Gotchev, 2019), (Domański et al., 2019).
There are two complementary approaches for 3D content creation. One approach is creation of novel three-dimensional material by capturing video and associated pre-pixel depth information. ZcamTM cameras from 3DV Systems which use active range approach in recording of mentioned content can be used **(**Fehn, 2004**)**. Except these new recordings, there is also a need to convert already existing 2D videos into 3D using methods called "structure from motion". The result of 3D content creation methods in all cases is a 2D color video in European digital TV format (720 x 576 luminance pels, 25 Hz, interlaced) and depth-image sequence with same spatio-temporal resolution **(**Fehn, 2004**)**. Each of depth information is stored in an 8-bit gray scale format where 0 denotes the far clipping plane and 255 the near clipping plane of the scene. The greyvalues are normalized to two depth clipping planes to translate depth information to real metric values i.e.near clipping plane Z_near for graylevel 255 (smallest metric depth value Z) and far clipping plane Z_far for graylevel 0 (largest metric depth value Z) (Figure 1).

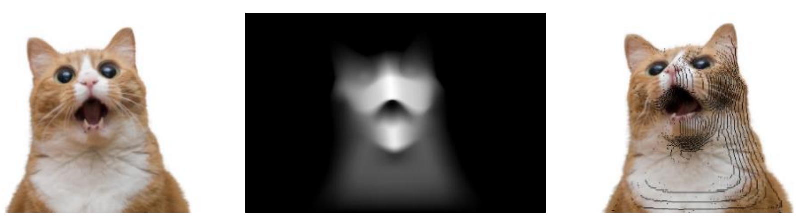
Figure 1: 2D color image (left) and 8-bit depth image (right)

The data formats for depth-based 3D systems include one or several pairs of coupled texture images and depth maps. Depths maps give the information according to the distance of each pixel in the video view related to the view camera position**)**, **(**Domański et al.**,** 2019**)**. "2D + Z" format is a 3D representation of one texture and one depth videos and the one of multiple texture and depth video is known as

Multiview Video plus Depth (MVD) which is a widely accepted data representation format (Zhu et al., 2013).
One of the key technologies essential for several 3D-related applications such as Virtual Reality (VR), Augmented Reality (AR), free viewpoint television (FVV) etc. is the Depth-Image-Based-Rendering (DIBR) technique used for synthesis of virtual views on multistereoscopic display from any subset of stereo or MVD videos (Gautier et al. 2011).
Section 2 introduces DIBR technique including its problems and possible solutions. Section 3 describes the experimental work with the proposed inpainting method and Section 4 brings the conclusion.

DEPTH IMAGE BASED RENDERING

DIBR improves the 3D visual experience compared to the conventional stereoscopic systems. The idea of DIBR is first to reproject the image points at the original viewpoint into 3D world and second to project these 3D points into the image plane at the virtual target viewpoint. The MVD format combined with DIBR uses only limited number of original views and corresponding depth maps that must be captured, stored, and later transmitted (Tian et al., 2019).
Generally, the creation of novel views from color and depth data using DIBR technique is consisted of image warping and image inpainting. For image inpainting the individual pixels of the original 2D view are shifted according to the corresponding depth (or disparity) map to their new position depending on the camera viewpoint to be synthesized. The image inpainting step estimates the missing image content by appropriate propagation of color and texture information from surrounding image regions to achieve a natural blending between the warped and newly synthesized pixel values **(**Domański et al.**,** 2019**)**.
DIBR technique causes new kinds of distortion compared to 2D distortions (blur, noise etc.) i.e., DIBR artifacts occur in disoccluded areas (Tian et al., 2019) which is shown on Figure 2 where the disoccluded areas are marked in black color. The holes in the novel view can degrade the quality and are called disocclusions.



a) b) c)

Figure 2: Example of DIBR distorted image: a) reference texture b) reference depth, c) synthesized view

The disocclusions problem can be addressed with two strategies on different places of DIBR flowchart: first step is using a low-pass filter to preprocess the depth video before construction to remove disocclusions and second the synthesized view is subsequently processed to fill in major imperfections/gaps in the area with credible, relevant color information, i.e. "hole filling". Some of the existing solutions for view synthesis include View Synthesis Reference Software (VSRS) (Senoh et al., 2017) and also Reference View Synthesizer (RVS) and the Versatile View Synthesizer (VVS) (Rogge et al., 2019).

Preprocessing of Depth Video

The depth video preprocessing reduces the disoccluded regions number and size by smoothing the depth discontinuities most commonly with a Gaussian filter (Zhu et al., 2013). Smoothing the whole depth video causes more damages than applying a correction around the edges. Various adaptive filters are proposed as a solution to this problem. These filters reduce both disocclusions and filtering-induced distortions in the depth video. For example, bilateral filter has edge-preserving capability and compared to Gaussian filters, it operates in spatial and color intensity space which results in better preservation of the sharp depth changes in conjunction with the variation in color space, and in consistent boundaries between texture and depth images (Zhu et al., 2013). Gaussian filter replaces a pixel value with weighted average of its neighbors and bilateral filter replaces pixel values with weighted average of its neighbors in both space and intensity domain.
Filtering is not enough to get satisfying visual quality. Larger disoclussions can remain therefore other post-processing methods like image inpainting must be used, but there is still no ideal solution. but there is still no ideal solution.

Image Inpainting

Image inpainting is one of the most important and useful aspects of image processing that refers to methods used to remove unwanted damages from an image (Vreja and Brad, 2014). It tries to fill pixels in a large missing region of an image with the surrounding information (Zhu et al., 2013). Image inpainting can be classified in three categories: structural inpainting, textural inpainting and hybrid methods (Vreja and Brad, 2014).

Criminisi's algorithm

Criminisi proposed an algorithm which uses exemplar-based texture synthesis to replicate texture and structure. The order in which pixels are restored is also important and each pixel starting with the one at the edge of the occluded area is given a priority. Confidence value is used to compute patch priorities, and to optimize the fill order of the target regions. The confidence in the target patch is updated after the patch has been filled with new values and as filling proceeds, confidence values decay. That implies that the pixel color values near the center of the target region are less reliable (Tian et al., 2019). Criminisi's algorithm preserves edge sharpness, there is no dependency on image segmentation, and it has balanced region filling (Criminisi et al., 2004).

CTBI

The coherence transport based inpainting (CTBI) is a pixel-based non-iterative method for removing objects and filling regions in images. Inpainting starts from the boundary pixels of the target region. The order in which the pixels in the target region are inpainted is calculated out of their Euclidean distance to the boundary of the target region. The pixel value is estimated from its coherent neighboring pixels and is the weighted average of known pixels within its inpainting radius. Structure tensors are used to estimate the coherence direction (Bornemann and März, 2007).

Proposed algorithm

After the creation of the synthesized view, several pixels are missing due to the different point of view. Those disoccluded areas are filled using distance transform, i.e. every pixel becomes the same as the nearest existing pixel. Distance transform is calculated using OpenCV's "distanceTransform" function. Afterwards, disoccluded areas are dilated using elliptical kernel with the size 5x5. Finally, Gaussian filtering is used over the dilated disoccluded areas, in our experiments with the kernel size 5x5. Kernel size for Gaussian filter and for dilation might be optimized in the future for the different size of disoccluded areas, i.e. it may not be the constant size.

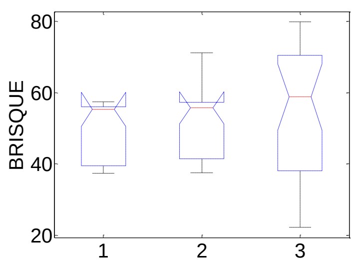
EXPERIMENTAL WORK

In this section we will present results that are obtained using three previously described methods: newly proposed distance transform and Gaussian filtering (using C++ and OpenCV), Criminisi algorithm and CTBI algorithm. Criminisi and CTBI algorithms were tested using Matlab 2020a. Criminisi algorithm can be also found to be implemented in C++, however existing implementations were slower compared to the Matlab implementation, so we used Matlab. Quality of the processed images has been compared using 3 no-reference objective image quality measures: BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator) (Mittal et al., 2012), NIQE (Naturalness Image Quality Evaluator) (Mittal et al., 2013) and PIQE (Perception based Image Quality Evaluator) (Venkatanath et al., 2015) measures, also implemented in Matlab. Lower values of BRISQUE, NIQE and PIQE measures represent better perceptual quality. Unlike BRISQUE and NIQE, PIQE measure does not need to be trained using some image dataset. Figure 3 shows 5 images with corresponding depth maps that were used for the later assessment.

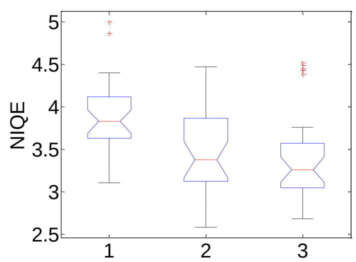
From each image, we calculated 6 arbitrary virtual view in different directions: from upper right direction, from upper direction, from right direction, from left direction, from lower direction and from lower left direction, giving overall (5 images) x (6 views) = 30 images for later comparison. 30 grades were compared for the 3 tested algorithms using ANOVA test and post-hoc multiple comparison test.

Figure 3: Images and corresponding depth maps (Pogac, 2020)

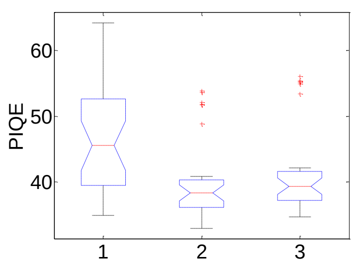
ANOVA test tests if the means of several groups are statistically all equal, against the alternative that they are not all equal. Multiple comparison test provides the information which pairs of means are statistically significantly different, and which are not. ANOVA test and multiple comparison test were calculated using Matlab (Figure 4).



a)



b)



c)

Figure 4: ANOVA test, 1-proposed, 2-Criminisi, 3-CTBI: a) BRISQUE measure, b) NIQE measure, c) PIQE measure

Results for all images and objective quality measures are given in Table 1.
Using multiple comparison test, it can be concluded that for BRISQUE measure, all tested algorithms have statistically similar mean value. For NIQE and PIQE

measures, proposed algorithm has statistically higher mean value (lower quality), while the two other algorithms perform similarly.

Table 1: Mean values of objective image quality measures for all images (best mean value is bolded for each measure)

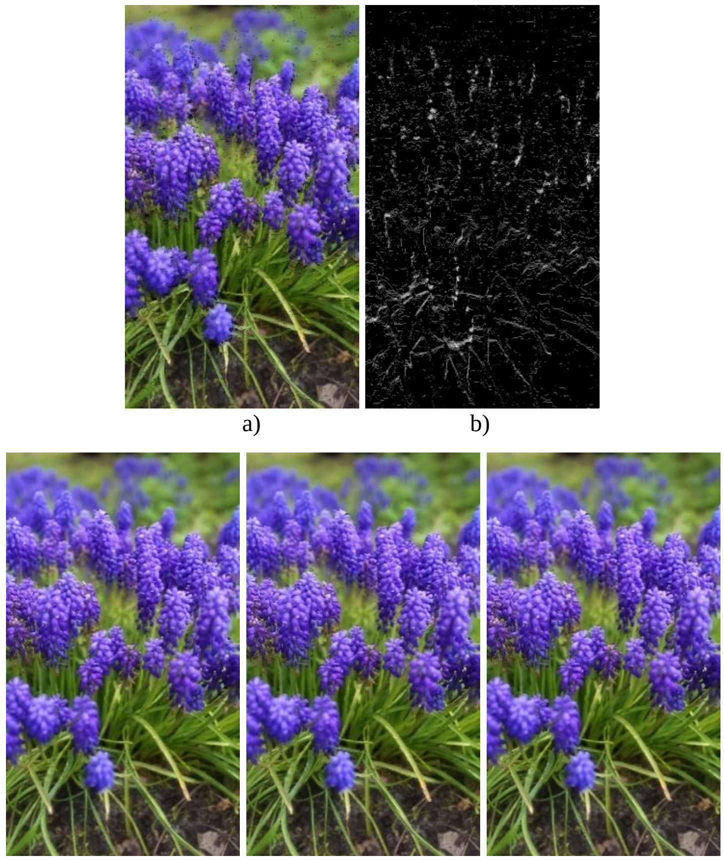
Method	BRISQUE	NIQE	PIQE
Proposed method	**49.3702**	3.8955	47.3187
Criminisi	52.5446	3.4950	**40.2105**
CTBI	54.0188	**3.4326**	41.6561

Timing performance is given in Table 2. Computer setup was Intel i7-4770@3.40 GHz with 32 GB RAM, running on Windows 10. Criminisi algorithm was tested using Matlab 2020a and "inpaintExemplar" function with the default setup. CTBI algorithm was also tested using Matlab 2020a and "inpaintCoherent" function with the default setup. It can be seen that the proposed algorithm works much faster than other tested algorithms. Difference between "proposed" and "proposed+view calculation" is that in the second case, we also calculate time needed to calculate synthesized view and corresponding mask, before applying the proposed algorithm. For other tested algorithms, we used images obtained from the same synthesized view and mask, e.g. time for view calculation was not added.

Table 2: Timing performance for 3 tested algorithms

Algorithm	Time (s) per frame
Proposed	17 ms
Proposed + view calculation	36 ms
Criminisi	~44000 ms
CTBI	~450 ms

Proposed algorithm can be found in the repository (Dumic, 2021), under Applications, View synthesis application. One example view synthesis for each of the tested algorithms can be seen in Figure 5.



c) d) e)

Figure 5: Example image for the upper view: a) synthesized view, b) corresponding mask, c) proposed algorithm, d) Criminisi and e) CTBI

CONCLUSION

DIBR techniques are one of the key solutions and promising tools which support advanced 3D video services by synthesizing new views. In this paper we proposed new algorithm based on distance transform and Gaussian filtering and compared it with two existing image inpainting methods using 3 no-reference objective quality measures. BRISQUE measure did not show statistically significant differences in the results for all three algorithms. However, NIQE and PIQE measures have similar results for Criminisi's and CTBI algorithms, while the proposed algorithm has statistically somewhat lower scores. From the subjective point of view, Figure 5, all algorithms perform similarly. However, it might be expected that Criminisi's and CTBI algorithms would produce somewhat better results for larger disoccluded areas, for example from more different points of view. Regarding the time comparison, results show that Criminisi’s and CTBI algorithms need significantly more time to calculate the images, while the proposed method calculates images in real time. With the improvement and special adaptation of the proposed method and with the creation of more precise and detailed depth maps, distorted images and videos may be computed with even better objective and subjective results.

REFERENCES

Assunção, P. A., Gotchev, A. (2019) "3D Visual Content Creation, Coding and Delivery." Springer International Publishing.

Bornemann, F., März, T. (2007) "Fast Image Inpainting Based on Coherence Transport." *Journal of Mathematical Imaging and Vision*, Vol. 28, pp. 259–278.

Criminisi, A., Perez, P., Toyama, K. (2004) "Region filling and object removal by exemplar-based image inpainting." *IEEE Trans. on Image Processing*, Vol. 13, No. 9, pp. 1200–1212.

Domański, M. et al. (2019) "Emerging Imaging Technologies: Trends and Challenges" in: Assunção P., Gotchev A (eds) 3D Visual Content Creation, Coding and Delivery. Signals and Communication Technology. Springer, Cham.

Dumic, E. Proposed algorithm materials under Application, View synthesis application. Available at: http://msl.unin.hr, [Accessed 30th April 2021].

Fehn, C. (2004) "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV." Proc. SPIE 5291, Stereoscopic Displays and Virtual Reality Systems XI, pp. 93-104.

Gautier, J., Le Meur, O., Guillemot, C. (2011) "Depth-based image completion for view synthesis." 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), Antalya.

Mittal, A., Moorthy, A. K., Bovik, A. C. (2012) "No-Reference Image Quality Assessment in the Spatial Domain." *IEEE Transactions on Image Processing*, Vol. 21, No. 12, pp. 4695- 4708.

Mittal, A., Soundararajan, R., Bovik, A. C. (2013) "Making a Completely Blind Image Quality Analyzer." *IEEE Signal Processing Letters*, Vol. 22, No. 3, pp. 209–212.

Pogac, T. (2020) "Rekonstrukcija virtualnih pogleda dobivenih pomoću DIBR algoritama." Master thesis, University North, Varaždin.

Rogge, S. et al. (2019) "MPEG-I Depth Estimation Reference Software." 2019 International Conference on 3D Immersion (IC3D), pp. 1-6.

Senoh, T., Yamamoto, K., Tetsutani, N., Yasuda, H., Wegner, K. (2017) "View Synthesis Reference Software (VSRS) 4.2 with improved inpainting and hole filling." ISO/IEC JTC1/SC29/WG11 MPEG2013, M40657, Hobart, Australia.

Tian, S., Zhang, L., Morin, L., Déforges, O. (2019) "A Benchmark of DIBR Synthesized View Quality Assessment Metrics on a New Database for Immersive Media Applications." *IEEE Transactions on Multimedia*, Vol. 21, No. 5, pp. 1235-1247.

Venkatanath, N., Praneeth, D., Chandrasekhar, Bh. M., Channappayya, S. S., Medasani, S. S. (2015) "Blind Image Quality Evaluation Using Perception Based Features." Proceedings of the 21st National Conference on Communications (NCC). Piscataway, NJ.

Vreja, R., Brad, R. (2014) “Image Inpainting Methods Evaluation and Improvement.” *The Scientific World Journal*, Hindawi, Volume 2014.
Zhu, C., Zhao, Y., Yu, L., Tanimoto, M. (2013) “3D-TV System with Depth-Image-Based Rendering: Architectures, Techniques and Challenges.” Springer-Verlag New York.