A Modified Non-local Mean Inpainting Technique for Occlusion Filling in Depth-Image Based Rendering

Lucio Azzari[a], Federica Battisti[b], Atanas Gotchev[a], Marco Carli[b] and Karen Egiazarian[a]

[a]Tampere University of Technology, Korkeakoulunkatu, 10, Tampere, Finland;
[b]Universita' degli Studi Roma TRE, via della Vasca Navale, 84, Rome, Italy

ABSTRACT

'View plus depth' is an attractive compact representation format for 3D video compression and transmission. It combines 2D video with depth map sequence aligned in a per-pixel manner to represent the moving 3D scene in interest. Any different-perspective view can be synthesized out if this representation through Depth-Image Based Rendering (DIBR). However, such rendering is prone to *disocclusion* errors: regions originally covered by foreground objects become visible in the synthesized view and have to be filled with perceptually-meaningful data.

In this work, a technique for reducing the perceived artifacts by inpainting the disoccluded areas is proposed. Based on Criminisi's exemplar-based inpainting algorithm, the developed technique recovers the disoccluded areas by using pixels of similar blocks surrounding it. In the original work, a moving window is centered on the boundaries between known and unknown parts ('target window'). The known pixels are used to select windows which are most similar to the target one. When this process is completed, the unknown region of the target patch is filled with a weighted combination of pixels from the selected windows.

In the proposed scheme, the priority map, which defines the rule for selecting the order of pixels to be filled, has been modified to meet the requirement for disocclusion hole filling and a better non-local mean estimate has been suggested accordingly. Furthermore, the search for similar patches has also been extended to previous and following frames of the video under processing, thus improving both computational efficiency and resulting quality.

The effectiveness of the proposed method is demonstrated by objective and subjective tests.

Keywords: DIBR, inpainting, video processing, 3D-video

1. INTRODUCTION

Usually, what is called a 3D video is a two-channel video, where the channels are associated with the left and right view and hence are separately shown to the left and right eye respectively. In an attempt to reduce the amount of data to be stored and/or transmitted, a new format, informally called 'view plus depth' (V+D)[1] , has been defined. In this format, the left and right channels are replaced by a single video sequence augmented by its *depth map* sequence. Depth map refers to a gray-scale image[2] , where each value is proportional to the distance of the corresponding pixel in the video frame from the camera, as shown in Figure 2 (b). Due to the fact that the depth maps are characterized by uniform regions, delineated by sharp objects borders, they can be easily and efficiently coded. For this reason this format appears to be suitable for 3D video transmission systems with limited bandwidth as the transmission of stereo content reduces to the transmission of 2D video and its relative depth map that represents a little 'overload' of the main video channel.

At the receiver side, the Depth-Image-Based-Rendering (DIBR) technique re-generates the stereo sequence from the color, texture and geometry information in the V+D representation. The use of geometric rules[3] together with the knowledge of the distance of the objects from the camera allow the rendering of a new view of the scene, that is an image representing the same scene recorded from a different point of view as shown in

Further author information: (Send correspondence to Lucio Azzari, E-mail: lucio.azzari@tut.fi)

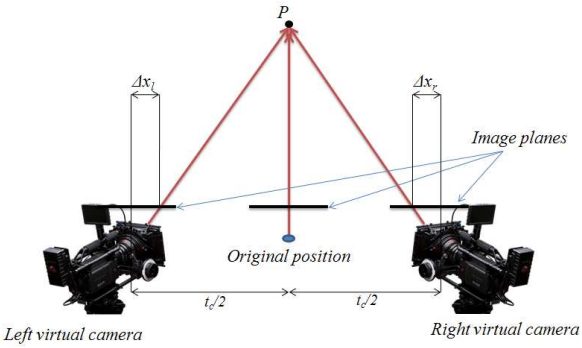

Figure 1. Simulation of camera shift.

Figure 1.
It is possible to represent this procedure by means of the following formula:

$$X_{l/r} = X + \Delta x_{l/r}, \tag{1}$$

where $\Delta x_{l/r}$ is the horizontal shift equal to:

$$\Delta x_{l/r} = \begin{cases} \frac{t_c f}{2Z} & left\ view \\ -\frac{t_c f}{2Z} & right\ view, \end{cases} \tag{2}$$

where Z is the depth value of the pixel in the intermediate image, f is the focal length of the camera and $X_{l/r}$ is the resulting horizontal coordinate of the pixel in the left/right virtual camera[4] . For improving the quality of depth feeling the so-called 'Zero Parallax Setting' (ZPS), a plane in which there is no disparity, is used[5] . With this method it is possible to simulate a virtual shift of the sensors thus adjusting the depth perception[6] . According to Eq. 1, the formula becomes:

$$X_{l/r} = X + \Delta x_{l/r} + h, \tag{3}$$

where the sensors' shift h is:

$$h = \begin{cases} -\frac{t_c f}{2Z_c} & left\ view \\ \frac{t_c f}{2Z_c} & right\ view. \end{cases} \tag{4}$$

In Eq. 4 , Z_c is the *'convergence plane'*, and usually it is set to the intermediate perceived distance. Figure 2 shows the rendered frames using Eq. 3. It can be observed that after their construction (*warping*), the new images present some 'black holes', called disocclusions, which are precisely the regions that become visible after the simulated shifting of the focal point.

The problem of dealing with disocclusion holes has been addressed in several works. Some of them suggest pre-processing of the depth map by using different filters. Since disocclusions are generated by vertical discontinuity in the depth map, smoothing of these discontinuities before the rendering phase reduces the size of the holes and facilitates the filling process. After the rendering process, the disocclusions of size $1-2$ pixels can be filled with a local averaging filter.

In[7] , a symmetric Gaussian filtering of the depth map has been proposed in order to reduce the size of disoccluded areas. The resulting images contain smaller-size holes for the price of higher blur around edges. In[3] , Zhang *et al.* have tried to avoid blurring artifacts by using an asymmetric Gaussian filter. The resulting images do not contain distortions on vertical edges, while such are still present on the horizontal ones. Park et al.[8] have proposed the use of an edge-dependent filter: it smoothes the edges with different coefficients depending on the value of the gradient in that point. In particular, the smoothing is stronger where the gradient has a large

(a) (b)

(c) (d)

Figure 2. Example of an original image (a) and its depth map (b), and the corresponding left (c) and right (d) rendered images.

value in the horizontal direction. This method improves the rendering results, though disocclusion artifacts are still present. They become more visible when vertical edges are present close to the disocclusion. The above cited methods are based on modification of the original depth map resulting in a possible *depth loss* effect at the rendering phase. However, their main advantage is in the low computational complexity which allows real time application.

In a previous work[9] , the algorithm of Park, that has delivered best results in terms of PSNR of the reconstructed images, has been compared to two inpainting methods opportunely modified. In this paper, a new version of the exemplar-based inpainting algorithm[10] by Criminisi *et al.* is presented. This modification reduces the computational time and facilitates the real time application. The rest of the paper is organized as follows. In Section 2 the proposed approach is motivated and presented. In Section 3 both the objective and subjective experiments performed for assessing the performances of the proposed method are described and the collected results are presented. Finally, in Section 4, the conclusions are drawn.

2. PROPOSED APPROACH

In[10] , Criminisi *et al.* presented an exemplar-based inpainting method. It is aimed at recovering an unknown image region, a *hole*, by using the information from the surrounding regions while maintaining high quality of the textures in the corrected regions. The procedure can be summed up in three steps:

1. Computation of a priority map;
2. Selection of disoccluded areas (holes) and collection of similar patches based on block matching;
3. Hole filling.

In this approach, edges inside the disoccluded regions are propagated first followed by processing the smooth areas.

A priority map P contains, for each pixel of the image, the corresponding priority filling order, which is computed as follows:

$$P = D \bullet C. \tag{5}$$

The term D represents the intensity and direction of the edges surrounding the unknown area. It contains values which are large in the case of a high-contrast edge in the hole direction; C indicates the number of known pixels surrounding the unknown current pixel, and $< \bullet >$ represents the componentwise multiplication of the two matrices.

After the priority map has been computed, the pixel with the highest priority value belonging to the boundary between known and occluded areas is selected, and a target window of size $w \times w$ is centered on it. The known area is used as a template in the similarity matching process. The best match is found and then used to fill the unknown part of the target by substitution.

In the proposed method the three steps have been modified in order to: i) adapt the method to the particular type of holes, i.e. disocclusions, ii) to speed up the block matching step, and iii) to improve the quality of the results.

As can be noticed in Figures 2 (c) and (d), the disocclusions resulting by DIBR techniques are located on the boundaries of objects positioned at different distances from the camera thus resulting in different depth values. By using the classical scheme for computing the priority map, the pixel with the highest priority value may belong to the foreground. In this case the target window is centered on the foreground and regions belonging to the foreground are used for filling the hole. This leads to perceivable artifacts since the disocclusion belongs to the background. To cope with this problem, the use of a modified priority map is proposed. This modification ensures that the filling process is first performed considering areas belonging to the background and then to the foreground.

The depth map contains information about the depth values of all the pixels but for those in the disoccluded areas whose values are unknown. The complete map is obtained after filling the unknown pixel values by some estimation technique. To this aim a rendering algorithm is applied to the depth map. The resulting views present the same disocclusions as the rendered 3D frame, and consequently make it possible to estimate the depth value of the pixels belonging to the boundaries of the disocclusions. The new depth map, $ID_{r\backslash l}$, can be obtained by computing the complement to the maximum luminance value, L_{max}, of $D_{r\backslash l}$. Since the background areas are identified with a higher value than the pixels in the foreground, it is possible to estimate the depth values of the occluded areas by performing a smoothing filtering of $ID_{r\backslash l}$.
In the proposed method, the smoothed views are used as priority maps for generating the best filling order of the pixels in the rendered frame. Figure 3 shows the procedure for the priority map computation.

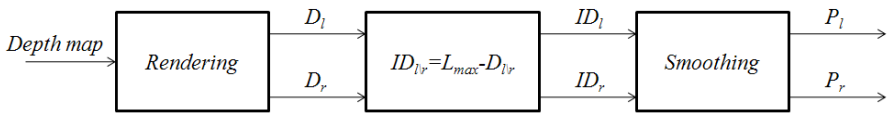

Figure 3. Proposed priority map evaluation scheme.

In the proposed approach, the computational complexity of the block matching algorithm used in[10] , is reduced. Specifically, the distance between the target window and the possible patches has been computed

using only the Y frame component. Another improvement for reducing the computational time is obtained by introducing two thresholds, β and α, as follows:

- The distance between patches, denoted by d, is computed by using only half of the available pixels: if d is with higher value than β, the current patch is discarded and the following patch is considered, otherwise if d is smaller than β, the remaining pixels are used and the distance D is stored and compared to the other distances. The less similar patches are discarded by halving the number of operations.
- After the distance D is computed, for further reducing the computational cost needed to find the most suitable patch, D is compared to the threshold α defining the maximum acceptable difference between the patches. If a value D smaller than α is found, the block matching procedure halts and the current patch is used for filling the disocclusion; otherwise, the next patch is considered.

The matching process is sketched in the flowchart in Figure 4.

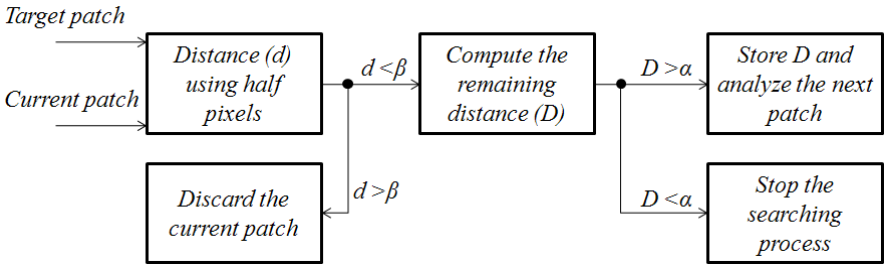

Figure 4. Flowchart of the proposed block matching algorithm.

The selection of the thresholds α and β is critical for the method's performances. To this aim, the block matching based method proposed in[9] , has been used to fill a set of 5 3D-videos with variable window's size (5×5 pixels, 7×7 pixels and 9×9 pixels). The average per-pixel distance between each original and best match patch has been evaluated to analyze the overall error trend. From Figure 5 it can be noticed that the error relative to the 95^{th} percentile is below 5. According to the performed test β has been set equal to 5 and α equal to 1.

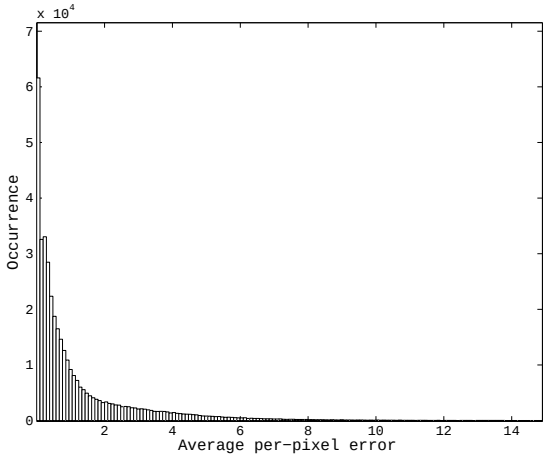

Figure 5. Occurrence of the average per-pixel error for 389524 block matching results.

For further reducing the computational complexity, the search region used in the similarity matching procedure, is reduced to a window of size $M \times M$ pixels around the target in the current frame and it is extended to the

previous and successive N frames in the temporal domain. In this way, moving objects are considered and the disocclusions replacement can be performed by also using the information revealed by the objects' movements. A further improvement is achieved by increasing from 1 to k the number of similar patches exploited for recovering the occluded areas. A weighted non-local mean of the k patches is performed for estimating the filling part, as shown in Eqs. 6 and 7. The coefficients are proportional to the distance between target and patches:

$$\psi_{t'} = \frac{\sum_{i=1}^{k} w(\psi_i)\psi_i}{\sum_{i=1}^{k} w(\psi_i)}, \tag{6}$$

and

$$w(\psi_i) = e^{-\frac{d(\psi_t, \psi_i)}{h}}, \tag{7}$$

where $\psi_{t'}$ is the region of the target window to be recovered, ψ_i are the most similar k patches and $w(\psi_i)$ are their relative weights. In order to adapt the algorithm to the use of the weighted mean of more patches, it is necessary to modify the threshold β by taking into account the fact that the error relative to the 95^{th} percentile increases; in the performed experiments, in which k has been selected equal to 5, β has been chosen equal to 7.

3. EXPERIMENTAL RESULTS

To verify the effectiveness of the proposed techniques, experimental tests have been performed. As described in the previous Section, the raw impaired videos have been processed with different algorithms and the quality of the resulting data assessed by means of both objective and subjective experiments.
Twenty sequences were drawn from a pool of uncompressed stereo videos: *Horse*, *Car*, *Butterfly*, *Bullinger*, and *Quest*[11] . The video content has been chosen to provide a range of different situations in motion, detail, color, contrast, brightness, and depth perception, as described below:

- *Horse*: outdoor scene with slow motion and two main depth levels;
- *Car*: outdoor moving scene with constant moderate motion;
- *Butterfly*: cartoon video with fixed camera, different depth levels and moving objects;
- *Bullinger*: news type of scene with a speaker in foreground and still background (two mail depth levels);
- *Quest*: cartoon video with varying depth levels.

The original videos are in the format of two-channel stereo, which have been used to estimate the depth maps aligned with the left channel[11] . In order to mitigate possible flickering effects in rendered videos, all estimated depth maps have been further processed with a modified bilateral Gaussian filter presented in[12] . The test set is composed by 5 original 3D videos and 15 processed sequences. The processed sequences are obtained by rendering the right channel using the original left channel and the estimated associated depth followed by processing the reconstructed sequences with three occlusion-handling algorithms denoted as: Oliveira[13] , Park[8] , and the proposed one. All video sequences are 10 s long and the size of each frame is 427×240 pixels. The frame rate for the sequence *Bullinger* is 30 fps, while for the other videos is 15 fps. Sample frames that have been extracted from the original left sequences are shown in Figure 6.

3.1 Objective tests

In this work, the objective quality of the rendered sequences has been evaluated by using state of the art quality assessment methods. Given a luminance frame $I(x, y)$ and its impaired version $I'(x, y)$, the objective quality can be rated by using:

(a) (b) (c)

(d) (e)

Figure 6. Sample left frames extracted by the videos used in the test: *Horse* (a), *Car* (b), *Butterfly* (c), *Bullinger* (d), and *Quest* (e).

- Peak Signal to Noise Ratio (PSNR):

$$\text{PSNR(dB)} = 10 \log_{10} \left(\frac{\max \left((I(x,y))^2 \right)}{\text{MSE}} \right), \tag{8}$$

  where $\text{MSE} = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} [I(x,y) - I'(x,y)]^2$.

- PSNR_{HVS_M}[14] that is a modified version of PSNR that takes into account both the Contrast Sensitivity Function (CSF) and the between-coefficient contrast masking in the Discrete Cosine transform (DCT) domain. This metric has been demonstrated to rank impaired images closely to the human judgment[14] .

- Weighted Peak Signal to Noise Ratio (WPSNR)[15] , a pixelwise comparison metric as the PSNR, whose ranking is modified according to the amount of texture in the image:

$$WPSNR(dB) = 10 \log_{10} \frac{\max \left((I(x,y))^2 \right)}{\| NVF \left(I'(x,y) - I(x,y) \right) \|^2}, \tag{9}$$

  where NVF is the Noise Visibility Function whose value is 1 in flat regions and 0 in textured regions and edges[15];

- Video Quality Metric (NTIA-VQM)[16] is a Full Reference standardized system for quantifying the perceptual quality degradation in video systems using compression. The scores are reported on a nominal range of [0, 1], where zero indicates excellent quality. Based on good correlation with subjective human rating, it has been adopted as standard and international recommendation[17,18] .

- NRMos[19] is a No Reference metric for assessing the quality of impaired videos. It is based on the analysis of the inter-frame correlation measured at the output of the rendering application. It does not require information on the errors, delays, and latencies affecting the link and on the countermeasures introduced by decoders in order to face the potential quality loss. The quality is assessed with a score in the range [0, 5], where 5 stands for excellent quality.

The obtained results have been averaged on the video database and are reported in Table 1. All adopted quality metrics show an improvement in the processed video frames, with respect to the raw ones (raw denotes rendered channels with no holes filling). The Criminisi based method, slightly outperforms the other ones.

Table 2 shows the values obtained by the quality evaluation of each sequence in the database.

PSNR_{HVS_M} Matlab code is freely available for download at http://www.ponomarenko.info/psnrhvsm.htm

Algorithm	PSNR	$PSNR_{HVS_M}$	WPSNR	NTIA-VQM	NRMos
Proposed	**27.78**	**25.52**	**36.26**	**0.31**	**4.90**
Oliveira based	26.78	24.14	34.74	0.38	4.84
Park based	27.56	25.15	35.99	**0.31**	4.85
Raw	23.47	20.88	31.58	0.53	4.42

Table 1. Average objective results computed by using state of the art metrics.

Bullinger	PSNR	$PSNR_{HVS_M}$	WPSNR	NTIA-VQM	NRMos
Proposed	26.35	24.21	34.84	0.33	5.00
Oliveira based	26.33	23.53	34.00	0.37	4.90
Park based	26.30	23.80	34.69	0.34	5.00
Raw	23.74	20.84	31.04	0.57	4.56

Butterfly	PSNR	$PSNR_{HVS_M}$	WPSNR	NTIA-VQM	NRMos
Proposed	27.84	24.55	36.14	0.35	4.80
Oliveira based	25.68	21.34	32.81	0.42	4.94
Park based	27.53	23.94	35.56	0.36	4.83
Raw	16.6	12	23	0.89	3.59

Car	PSNR	$PSNR_{HVS_M}$	WPSNR	NTIA-VQM	NRMos
Proposed	29.43	27.75	37.12	0.30	4.94
Oliveira based	29.48	27.86	37.18	0.30	4.96
Park based	29.24	27.51	36.88	0.30	4.96
Raw	25.88	24.83	34.29	0.47	4.66

Horse	PSNR	$PSNR_{HVS_M}$	WPSNR	NTIA-VQM	NRMos
Proposed	25.92	25.10	36.11	0.24	4.89
Oliveira based	25.98	25.08	36.05	0.24	4.90
Park based	25.44	24.75	35.89	0.24	4.66
Raw	22.09	20.27	31.79	0.49	4.37

Quest	PSNR	$PSNR_{HVS_M}$	WPSNR	NTIA-VQM	NRMos
Proposed	29.37	25.99	37.07	0.32	4.86
Oliveira based	29.00	25.57	36.63	0.39	4.86
Park based	29.27	25.76	36.92	0.34	4.83
Raw	26.46	23.75	34.78	0.44	4.60

Table 2. Objective results computed by using state of the art metrics for the sequences: *Bullinger*, *Butterfly*, *Car*, *Horse*, and *Quest*.

3.2 Subjective tests

The methods for disocclusion filing have been compared also against quality judgments of persons in subjective tests.

Video clips were presented to the viewers on a portable 3.1" autostereoscopic display with horizontally double-density pixel arrangement created by NEC Technologies[20] .

The viewers ranked the test sequences according the Single Stimulus Continuous Quality Evaluation SSCQE[21] approach.

In total, 33 subjects with age between 22 and 40 participated to the tests.

The test has been divided into 5 sections, namely:

- *Preliminary vision test*: this is used to check for possible vision disabilities of the test persons, e.g. binocular blindness;
- *Training test*: this is used to train the observers to rank the videos for their quality on the given scale;
- *quality test*: this is the core of the test. In this phase each user is asked, after each video, to rank the quality of each sequence, as follows:
  - **Acceptance of the video**: each observer has to say if the quality of the video is acceptable (yes/no);
  - **Overall video quality**: each observer has to judge the overall video quality by expressing his/her opinion with a number in the range $[0, 10]$ where 0 corresponds to the lowest quality and 10 to the highest;
  - **Perceived depth**: each subject evaluates the perceived sensation of depth with a number in the range $[0, 10]$ where 0 corresponds to the lowest level of perceived quality and 10 to the highest.

The overall test per subject has been about 40 minutes long.

First of all the acceptance of the videos has been evaluated. The collected results are expressed in percentage and are shown in Figure 7. All sequences but *Quest* have reached acceptance rates above 50%. As of the *Quest* sequence, the accuracy of the estimated depth around foreground objects borders has been rather low. It could not be corrected with the post-filtering technique applied. The low-quality depth caused artifacts in the rendered foreground objects and various hole-filling algorithms made no difference for the evaluating subjects who rejected the corresponding videos. Note that the results of the objective comparisons did not detect that problem, as the rendering artifacts are in relatively small areas and will negligible effect on the overall score. Due to the low acceptance rate, the quality results of the *Quest* sequence are excluded from the analysis thereafter. The acceptance rate averaged over the rest of four sequences is depicted in Figure 10 (c). The figure shows quite good acceptance of all of the methods compared.

The quality of the rendered sequences and the perceived depth have been evaluated by means of the Mean Opinion Score (MOS). The obtained results and the confidence intervals at 95% for individual sequences are shown in Figures 8 and 9.

The quality and the perceived depth score averaged over four sequences are shown in Figure 10 (b) and (c).

As seen in Figure 8, the proposed method performs in the most consistent way along different sequences. It is the best for three of the sequences and equally good to the Parks technique for the sequence *Butterfly*, which represents a synthetic content. While there is similar trend for the different contents, there are higher scores for the sequences *Butterfly* and *Horse*. In qualitative description after the tests, most of the subjects mentioned that those contents were more appealing in terms of likable foreground objects and vivid colors. The overall score calculated over four contents, as shown in Figure 10 (a), also confirms the superiority of the proposed method. As far as the perception of depth is concerned, Figure (9), there is higher uncertainty among techniques and contents manifested by the wider confidence intervals. Comparing Figure 8 with Figure 9, one could conclude that the disocclusion effects are mostly perceived as 2D type of artifacts influencing the overall quality, as exemplified by Figure 10 (a), while their effect to the depth perception is rather marginal (Figure 10 (b)).

4. CONCLUSIONS

In this paper a modified version of the inpainting technique by has been presented, aiming at efficient and high-quality disocclusion handling in DBIR. To this purpose a weighted mean of the k most similar patches has been used for recovering the disoccluded regions in a more efficient way. Improvements have been suggested for speeding up the search for similar patches and improving the quality by proper prioritization and non-local mean weighting. In order to prove the feasibility of the method, objective and subjective tests have been performed, comparing the technique with the technique by Park[8] and the modified version of the inpainting technique by Oliveira[9] . The developed technique shows high and consisting ranks over various video contents with reduced computational cost.

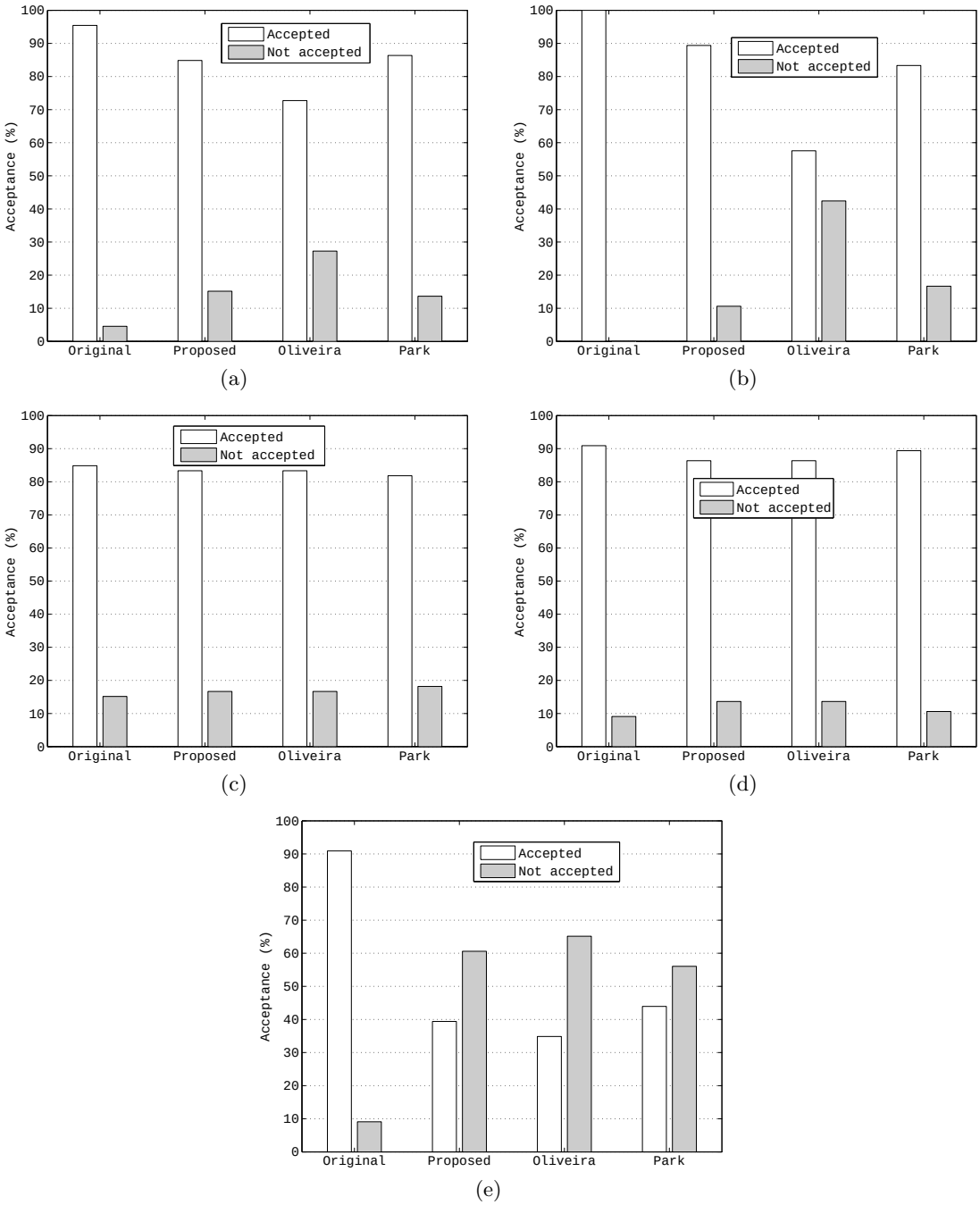

Figure 7. Acceptance evaluations for *Bullinger* (a), *Butterfly* (b), *Car* (c), *Horse* (d) and *Quest* (e) sequences.

REFERENCES

[1] Redert, A., Beeck, M. O. D., Fehn, C., Jsselsteijn, W., Pollefeys, M., Gool, L. V., Ofek, E., Sexton, I., and Surman, P., "Attest: Advanced Three-dimensional Television System Technologies," in [*Proc. 1st International Symposium on 3D Data Processing Visualization and Transmission (3DPVT 2002)*], (2002).

[2] Smolic, A., Wang, Y., Müller, M., and Kauff, P., "Report on generation of video plus depth data base." available at http://sp.cs.tut.fi/mobile3dtv/results/tech/D2.3_Mobile3DTV_v1.0.pdf.

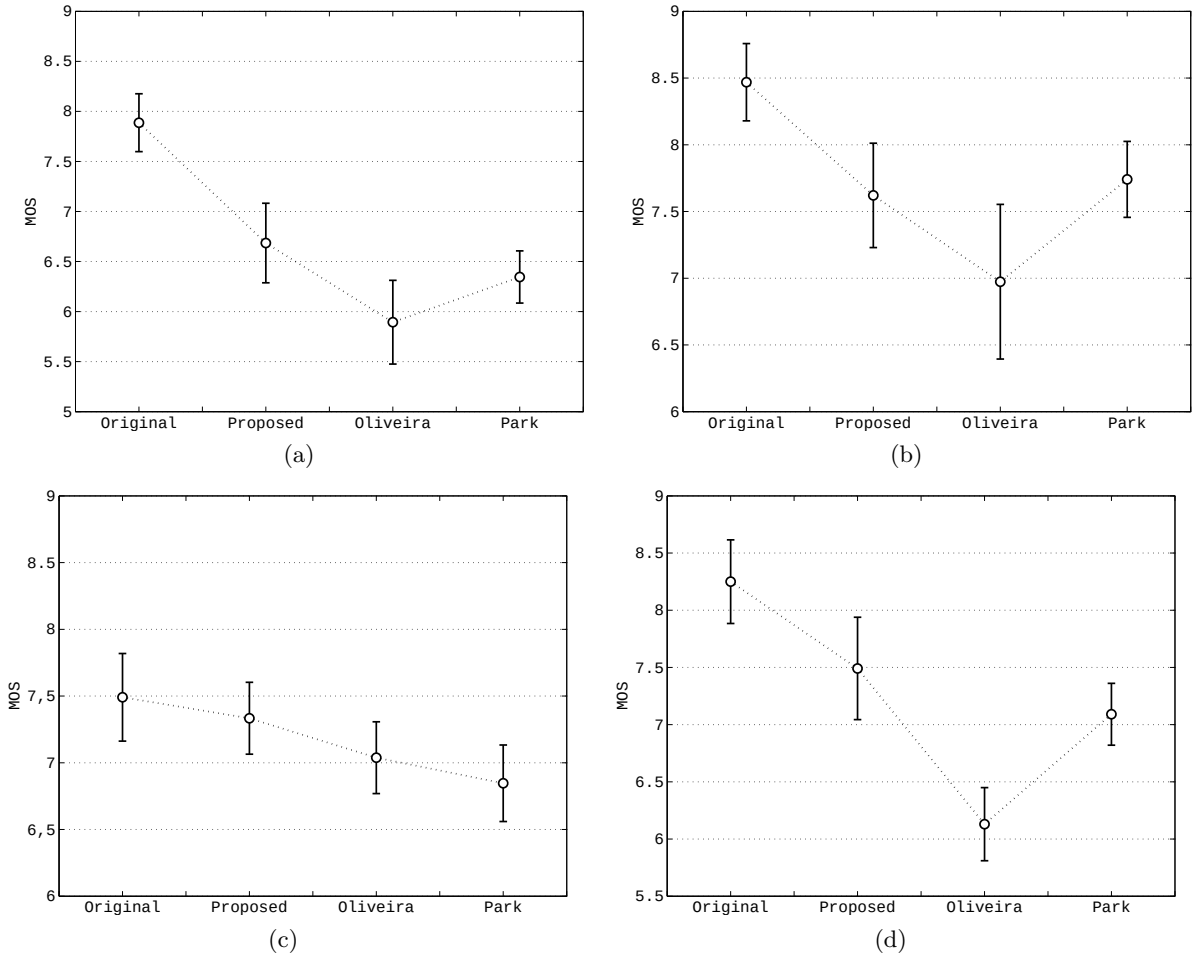

Figure 8. Quality MOS for *Bullinger* (a), *Butterfly* (b), *Car* (c), and *Horse* (d) sequences.

[3] Zhang, L. and Tam, W. J., "Stereoscopic image generation based on depth images for 3D TV," in [*IEEE Transactions on Broadcasting*], **51**, 191 – 199 (2005).

[4] Jung, K. H., Park, Y. K., Kim, J. K., Lee, H., Yun, K. J., Hur, N. H., and Kim, J. W., "2D/3D Mixed Service in T-DMB System Using Depth-Image Based Rendering," in [*Proc. 10th International Conference on Advanced Communication Technology*], **3**, 1868 – 1871 (2008).

[5] Lipton, L., "Foundations of the Stereoscopic Cinema - A Study in Depth," in [*Van Nostrand Reinhold*], (1982).

[6] Woods, A., Docherty, T., and Koch, R., "Image Distortions in Stereoscopic Video Systems," in [*Proc. SPIE Stereoscopic Displays and Applications IV*], 36 – 48 (1993).

[7] Zhang, L., Tam, W. J., and Wang, D., "Stereoscopic image generation based on depth images," in [*Proc. IEEE International Conference on Image Processing (ICIP 04)*], 2993–2996 (2004).

[8] Park, Y. K., Jung, K., Oh, Y., Lee, S., Kim, J. K., Lee, G., Lee, H., Yun, K., Hur, N., and Kim, J., "Depth-Image-Based Rendering for 3DTV service over T-DMB," in [*Signal Processing: Image Communication*], **24**, 122–136 (2009).

[9] Azzari, L., Battisti, F., and Gotchev, A., "Comparative Analysis of Occlusion-Filling Techniques in Depth Image-Based Rendering for 3D Videos," in [*ACM MoViD*], (2010).

[10] Criminisi, A., Perez, P., and Toyama, K., "Object Removal by Exemplar-Based Inpainting," in [*IEEE Transactions on Image Processing*], (2004).

[11] Smolic, A., Tech, G., and Brust, H., "Report on generation of stereo video data base." available at http://sp.cs.tut.fi/mobile3dtv/results/tech/D2.1_Mobile3DTV_v3.0.pdf.

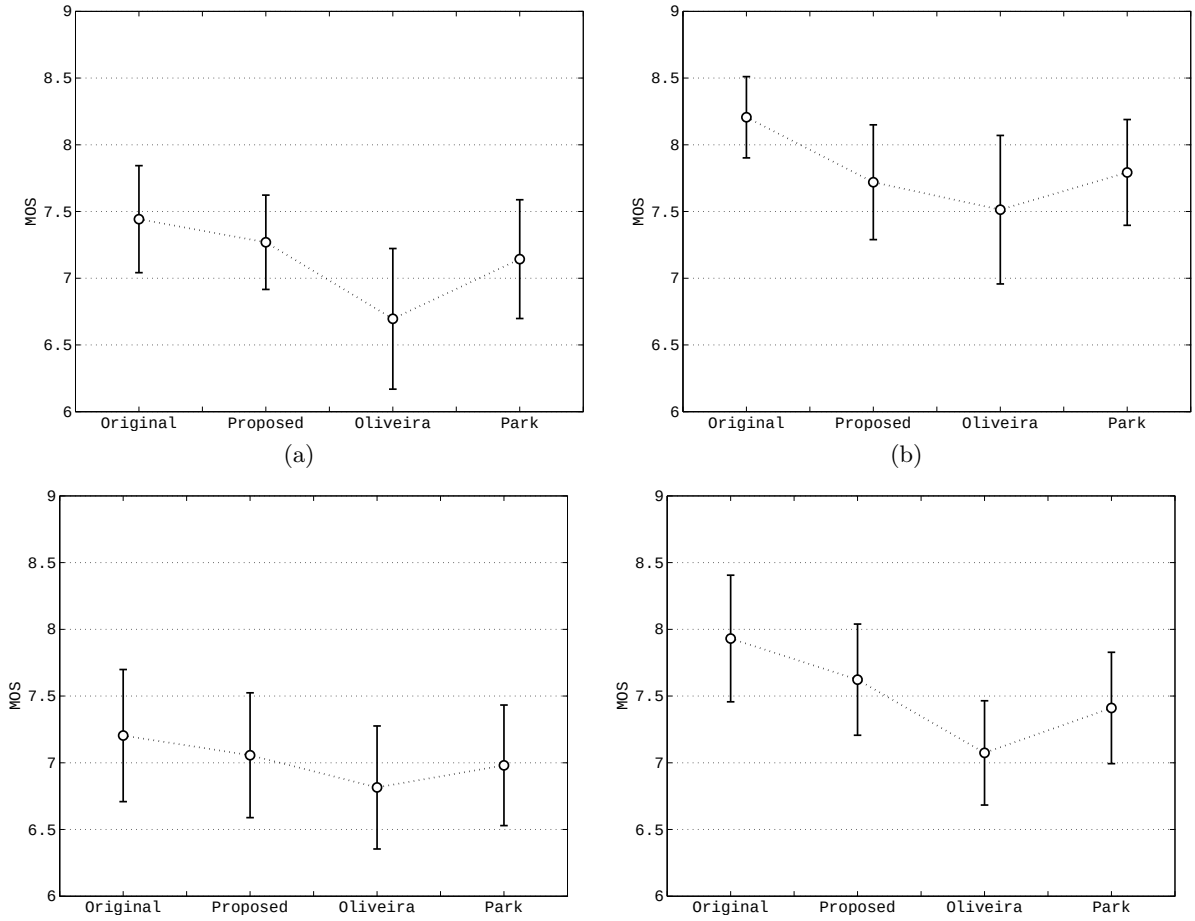

Figure 9. Perceived depth MOS for *Bullinger* (a), *Butterfly* (b), *Car* (c), and *Horse* (d) sequences.

[12] Smirnov, S., Gotchev, A., and Egiazarian, K., "A memory-efficient and time-consistent filtering of depth map sequences," in [*Proceedings of the SPIE*], **7532**, 753217–753217–12 (2010).

[13] Oliveira, M. M., Bowen, B., and McKenna, R., "Fast Digital Image Inpainting," in [*Proc. International Conference on Visualization, Imaging and Image Processing*], 261–266 (2001).

[14] Ponomarenko, N., Silvestri, F., Egiazarian, K., Carli, M., Astola, J., and Lukin, V., "On between-coefficient contrast masking of DCT basis functions," in [*Proc. 3rd International Workshop on Video Processing and Quality Metrics for Consumer Electronics*], (2007).

[15] Voloshynovskiy, S., Pereira, S., Iquise, V., and Pun, T., "Attack modelling: Towards a second generation watermarking benchmark," in [*Signal Processing, Special Issue on Information Theoretic Issues in Digital Watermarking*], **81**, 1177 – 1214 (2001).

[16] NTIA, "General video quality metric (VQM)." available at http://www.its.bldrdoc.gov/vqm/.

[17] T1.801.03, A., "American national standard for telecommunications - digital transport of one way video signals - parameters for objective performance assessment." American National Standard Institute. Available at http://www.ansi.org (2003).

[18] ITU-T144, "Recommendation j.144 (rev.1) - objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference." Recommendation of ITU, Telecommunication Standardization Sector. Available at http://www.itu.org (2004).

[19] Neri, A., Carli, M., Montenovo, M., Perrot, A., and Comi, F., "No reference quality assessment of Internet multimedia services," in [*Proc. 14th European Signal Processing Conference*], (2006).

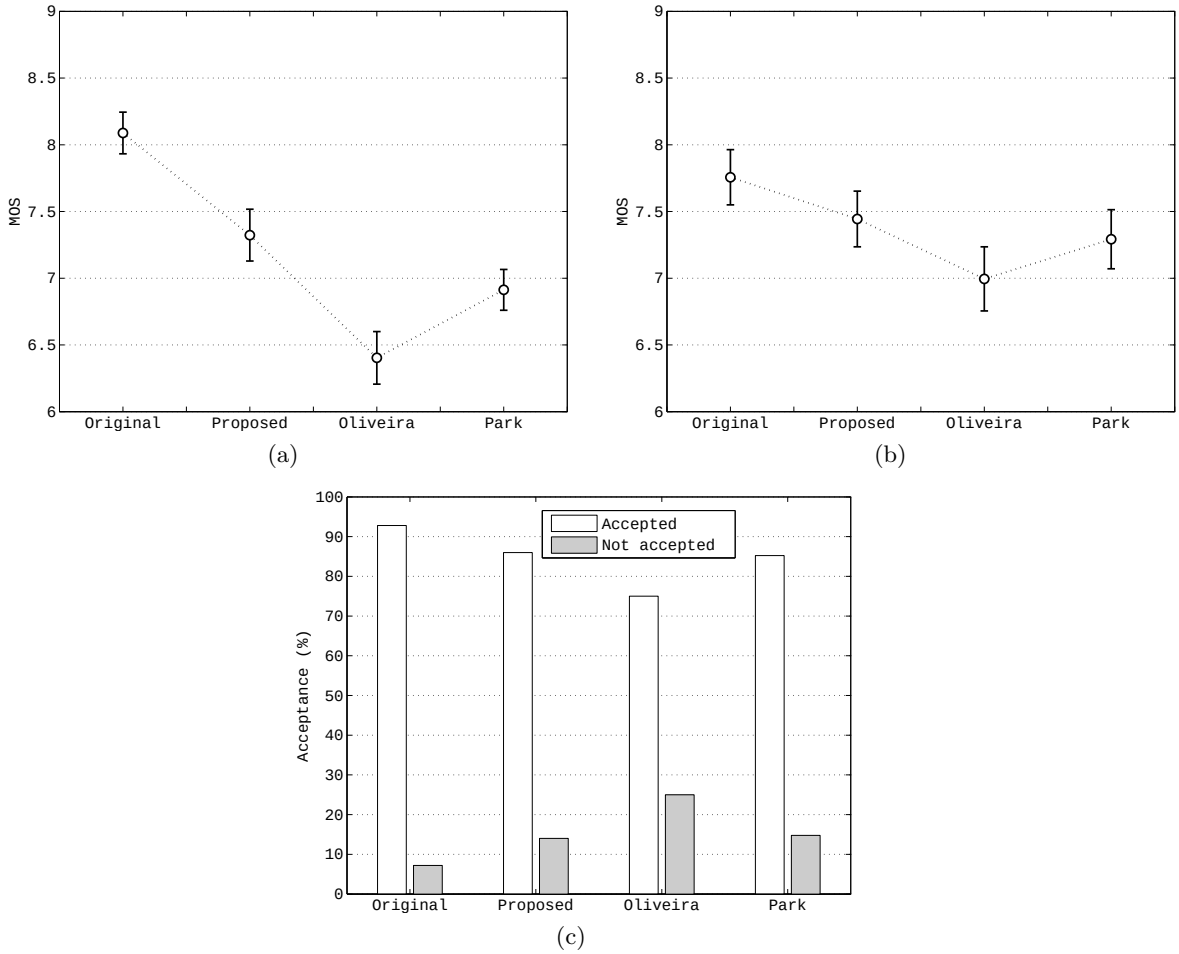

Figure 10. Average results of overall quality MOS (a), perceived depth MOS (b) and percentage of acceptance (c).

[20] Uehara, S., Hiroya, T., Kusanagi, H., Shigemura, K., and Asada, H., "1-inch diagonal transflective 2D and 3D LCD with HDDP arrangement," in [*Proc. SPIE-IST Electronic Imaging: Stereoscopic Displays and Applications XIX*], **6803** (2008).

[21] "Methodology for subjective assessment of the quality of television pictures," tech. rep., ITU Recommendation BT.500-11 (2002).