|  |  |
| --- | --- |
| **Joint Collaborative Team on 3D Video Coding Extension Development**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  3rd Meeting: Geneva, CH, 17–23 Jan. 2013 | Document: JCT3V-C0039 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **A New Hybrid View Synthesis Method for View Interpolation and Extrapolation** | | |
| *Status:* | Input Document | | |
| *Purpose:* | Proposal | | |
| *Author(s) or*  *Contact(s):* | Iliya Koreshev, Mahsa T. Pourazad, and Panos Nasiopoulos  2366 Main Mall, Vancouver, BC, Canada V6T 1Z4 | Tel: Email: | {ikoresh, pourazad, panos}@ece.ubc.ca |
| *Source:* | University of British Columbia (UBC), TELUS Communications Inc. | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

This contribution presents a hybrid view synthesizing approach for synthesizing virtual camera views. In this report the proposed method is used to both interpolate and extrapolate virtual views from either one or more available camera views respectively. The performance of the proposed method is evaluated using a series of subjective evaluations of synthesized videos and compared with that of the existing VSRS method. The experimental results indicate that the subjective quality of the synthesized 3D views using this approach is higher than the synthesized views generated by the view synthesis reference software.

# Introduction

The main issue with the synthesizing process is related to estimating the information of occluded areas [1]. A well-practiced solution to this problem, which has been also utilized in the view synthesis reference software [2], is to apply interpolation to estimate the missing texture of the hole pixels. This approach produces a similar looking color to the true background, but fails to reproduce texture that exists in these areas. To avoid creation of holes in the synthesized view, a group of researchers in Disney Research Zurich have proposed to use warping to generate synthesized views from available views [3]. The problem is that due to warping, some deformities can be produced in the generated virtual views especially around foreground objects close to background objects with well-defined vertical edges.

In earlier studies we created a new hybrid approach for hole filling to generate virtual views [4][5]. This approach utilizes some general ideas from the existing VSRS approach and the Disney approach to fill the holes created during the synthesizing process. In this contribution the performance of the proposed hybrid view synthesis method is tested for both interpolation and extrapolation tasks and its performance is compared to that of VSRV. The subjective evaluations show that this hybrid approach generates higher quality views by filling the hole regions by matching texture.

The remainder of this report is organized in the following manner: Section 2 provides an overview on the proposed hybrid view synthesis method, Section 3 elaborates on the subjective performance evaluation, and Section 4 concludes the report.

# Algorithm Description

In the view-synthesizing problem, we have one or more views and we try to synthesize virtual views from the available views. Figure 1 and Figure 2 show the block diagram of the proposed scheme for view-interpolation and view-extrapolation respectively. The first step in our scheme is to create a primary synthesized view based on the closest camera view and its depth map. To do this similar to VSRS, the appropriate shifting amount for different objects in the scene is calculated using the depth and texture information. This shifting process creates holes (pixels with missing color and texture information). The coordinates of the pixels corresponding to these holes are registered by creating a mask with the value of zero for the hole pixels and the value of one for the rest of the pixels. This mask is called “Mask I”.

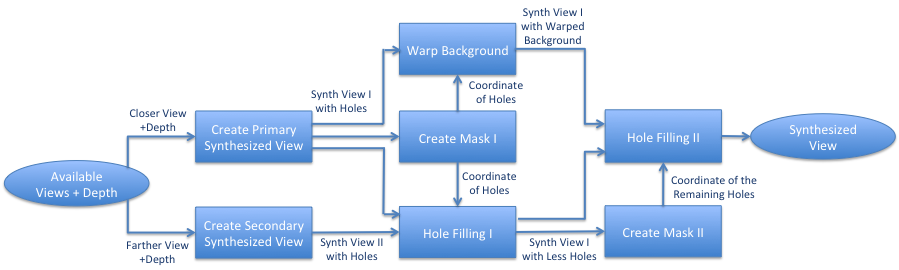


Figure. 1: Flowchart of the hybrid view synthesis technique for interpolation.

To fill the holes, in the case of interpolation, as shown in Figure 1, a secondary synthesized view is generated solely based on the farther view by following the same procedure as creating the primary synthesized view. The secondary synthesized view is used to fill the holes in the primary synthesized view (which are registered in Mask I) with the condition that the depth of these areas matches the depth of the neighboring objects to the holes in the primary synthesized view. In the case of extrapolation where a secondary view is not available no other view is generated as shown in Figure 2. Instead the holes in the synthesized view thinner than a set threshold (0.2% of the video width) are filled using nearest neighbor interpolation.

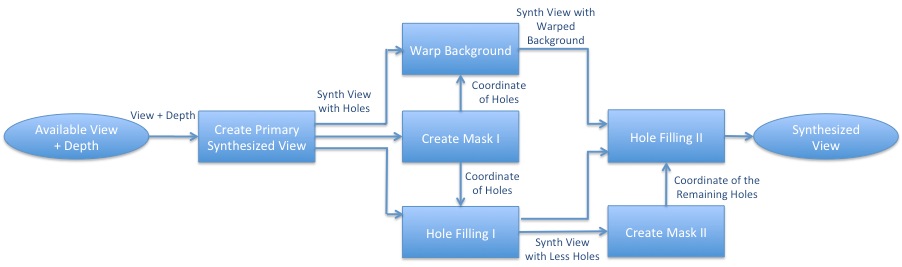


Figure. 2: Flowchart of the hybrid view synthesis technique for extrapolation.

The coordinates of the remaining holes in the primary synthesized view are registered by creating “Mask II” (similar to Mask I) . To fill up the holes, this hybrid approach applies warping to the background area of the primary synthesized view. To do this, Mask I is used to generate the list of warping points. The warping start-points (the points where the hole-areas start with a small overlap towards the background) and the warping end-points (the points where the hole-areas end with a small overlap towards the foreground) are identified using Mask I. To avoid vertical parallax, the warping process for filling the holes should be done in the horizontal direction, so the vertical coordinate of the warping start-point and end-point are equal. The warping process is also restricted to not use the information of the edges of the synthesized image. This is done because there is not enough texture data at the edges to guarantee effective warping. Piecewise cubic Hermite interpolation [6] is then applied to the primary synthesized view which takes a matrix containing all the points in the image as well as a matrix with the corresponding warp points and produces a matrix containing all the interpolated points between the warp points as follows:

|  |  |  |
| --- | --- | --- |
|  | number of nonzero elements in B = m  m < n | (1) |

where *n* is the total number of pixels in the primary synthesized view (including hole pixels), and *m* is the total number of warp points. Matrix A contains (*xi,yi*) which are the coordinate of pixel *i* in the primary synthesized view, matrix B is a sparse matrix containing (*xj’, yj’*) which are either the corresponding coordinate of the warping points (identified based on Mask I) or if no warping points exist then just zero, and matrix C contains (*xi’’*, *yi’’*) which are the interpolated locations of every *x* and *y* point in the warped image. Once the new coordinates in the warped image are obtained, the pixel values of primary synthesized view are mapped accordingly to create the warped image.

To create the final synthesized view as shown in Figure 1 and Figure 2, the hole areas in the primary virtual view (which are marked in Mask II) are filled with the data from the warped image. Once this process is complete, we obtain a virtual view where all the holes are filled.

In summary in the case of interpolation the holes are filled either with data from the secondary virtual viewor from the warped image. In the case of extrapolation the holes are filled either with data from nearest neighbor interpolation or from the warped image.

# Performance Evaluation

In this report the performance of the hybrid view synthesis method is evaluated based on subjective tests. Table 1 shows the test videos and synthesized views used in the interpolation experiment, and Table 2 shows the test videos and synthesized views used in the extrapolation experiment. To evaluate the performance of this method the synthesized views were compared to that of the one generated by the VSRS package (version 3.5) [2].

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Seq. ID** | **Test Sequence** | **Input views** | **View to synthesize** | **Stereo pair** |
| S01 | Balloons | 3-5 | 4 | 3-4 |
| S02 | Kendo | 3-5 | 4 | 3-4 |
| S03 | GT\_Fly | 3-5 | 4 | 3-4 |
| S04 | Balloons | 1-3 | 2 | 2-3 |
| S05 | Kendo | 1-3 | 2 | 2-3 |
| S06 | GT\_Fly | 1-3 | 2 | 2-3 |
| S07 | Balloons | 1-3-5 | 2-4 | 2-4 |
| S08 | Kendo | 1-3-5 | 2-4 | 2-4 |
| S09 | GT\_Fly | 1-3-5 | 2-4 | 2-4 |

Table. 1: Input views and stereo pair for interpolation 2-view test scenario.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Seq. ID** | **Test Sequence** | **Input views** | **View to synthesize** | **Stereo pair** |
| S01 | Balloons | 3 | 4 | 3-4 |
| S02 | Kendo | 3 | 4 | 3-4 |
| S03 | GT\_Fly | 3 | 4 | 3-4 |
| S04 | Balloons | 3 | 2 | 2-3 |
| S05 | Kendo | 3 | 2 | 2-3 |
| S06 | GT\_Fly | 3 | 2 | 2-3 |
| S07 | Balloons | 3 | 2-4 | 2-4 |
| S08 | Kendo | 3 | 2-4 | 2-4 |
| S09 | GT\_Fly | 3 | 2-4 | 2-4 |

Table. 2: Input views and stereo pair for extrapolation 2-view test scenario.

The viewing conditions were set according to the ITU-R Recommendations BT.500-13 [7]. The test sequences were all uncompressed to avoid any coding artifacts. Twenty volunteer subjects, from the age of 18 to 57 participated in the interpolation evaluations, and eighteen subjects from the age of 21 to 28 participated in the extrapolation evaluations. All subjects had none to marginal 3D image and video viewing experience. They all were screened for color and visual acuity (using Ishihara and Snellen charts), and for stereo-vision (Randot test – graded circle test 100 seconds of arc). The evaluation was performed using a 46” Full HD Hyundai 3D TV (Model: S465D) with passive glasses. The TV settings were as follows: brightness: 80, contrast: 80, color: 50, R: 70, G: 45, B: 30. The 3D display and the settings are based on MPEG recommendations for subjective evaluation of the proposals submitted in response to the 3DV CfP [8].

At the beginning of each evaluation session, a short demo sequence “Undo\_Dancer” with different levels of synthesizing artifacts was played for the subjects to become familiar with the artifacts and the testing process. The “Undo\_Dancer” test sequence was omitted from the actual evaluation procedure to maintain the purity of the results.

In the test, the viewers were shown the synthesized stereoscopic test sequences in random order, so that they would watch two different synthesized versions of the same sequence consecutively, without knowing which video was generated by our method or VSRS. Between test videos, a ten-second gray interval was provided to allow the viewers to rate the perceptual quality of the content and relax their eyes. Here, the perceptual quality reflects whether the displayed scene looks pleasant in general. In particular, subjects were asked to rate a combination of “naturalness”, “depth impression” and “comfort” as suggested by [9].

For ranking, there were 10 quality levels, 10 indicating the highest quality and 1 the lowest quality. Three test scenarios were examined: 1) right-view is synthesized, 2) left-view is synthesized and 3) both views are synthesized. Switching the synthesized view between the right and the left eye compensated for the effect of eye dominancy (out of the twenty volunteers for the extrapolation evaluation we had 13 left-eye dominant and 7 right-eye dominant subjects, and out of the eighteen volunteers for the interpolation evaluation we had 8 left-eye dominant and 10 right-eye dominant). After collecting the results outliers were removed (there were 2 outliers in the interpolation evaluation results and none in the extrapolation results).

The performance of this hybrid view synthesis method was compared with the existing VSRS package [2]. Figures 3 and 4 show the mean opinion score (MOS) for the interpolation and extrapolation subjective evaluations respectively. The evaluation results report that the proposed hybrid view synthesis method generates higher quality synthesized views compared to VSRS.

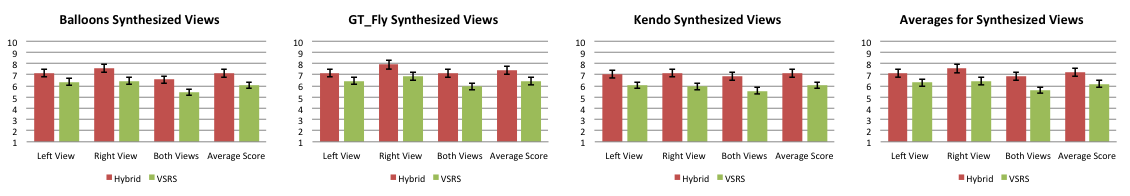


Figure. 3: MOS for individual interpolated scenes and the average MOS for all scenes combined.

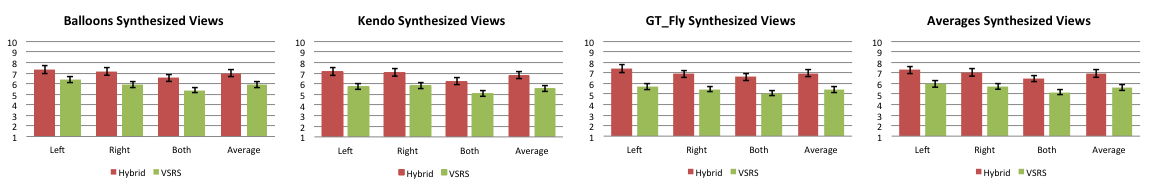


Figure. 4: MOS for individual extrapolated scenes and the average MOS for all scenes combined.

# Conclusion

In this document we present an overview of our hybrid view synthesis method. The performance of the proposed method was verified using a series of subjective tests and was compared with VSRS. Experimental results show that the proposed hybrid view synthesis method generates higher quality synthesized views compared to VSRS.

# References

1. ISO/IEC JTC1/SC29/WG11 “Description of Exploration Experiments in 3D Video Coding,” N11630, Oct. 2010.
2. ISO/IEC JTC1/SC29/WG11 “View Synthesis Software Manual,” MPEG, release 3.5, Sept. 2009.
3. M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, “Nonlinear disparity mapping for stereoscopic 3D,” ACM SIGGRAPH 2010 papers (SIGGRAPH '10), Hugues Hoppe (Ed.). ACM, New York, NY, USA, Article 75, 10 pages, 2010.
4. I. Koreshev, M. T. Pourazad, and P. Nasiopoulos, "Hybrid View-Synthesizing Approach for Multiview Applications", the 3DTV Conference (Zurich), Oct. 2012.
5. I. Koreshev, M. T. Pourazad, and P. Nasiopoulos, "A New Hybrid Approach for View Extrapolation and Hole Filling", the IEEE International Conference on Consumer Electronics (Las Vegas), Jan. 2013.
6. F. N. Fritsch and R. E. Carlson, "Monotone Piecewise Cubic Interpolation," SIAM J. Numerical Analysis, Vol. 17, 1980, pp.238-246.
7. ITU-R, “Methodology for the subjective assessment of the quality of television pictures,” ITU-R, Tech. Rep. BT.500- 13, 2012.
8. ISO/IEC JTC1/SC29/WG11 MPEG, Document N12036, “Call for proposals on 3D video coding technology,” 96th MPEG meeting, Geneva, March 2011.
9. Q. Huynh-Thu, P. L. Callet, and M. Barkowsky, “Video quality assessment: from 2D to 3D - challenges and future trends,” in Proceedings of IEEE International Conference on Image Processing (Institute of Electrical and Electronics Engineers, New York, 2010), pp. 4025–4028.

# Patent rights declaration(s)

**TELUS/UBC may have current or pending patent rights relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation | ISO/IEC International Standard (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).**