

## **A Local Depth Image Enhancement Scheme for View Synthesis**

Wang, Y.; Tian, D.; Vetro, A.

TR2012-070 September 2012

### **Abstract**

The quality of the depth map is crucial for depth image based rendering (DIBR) which enables a variety of advanced 3D video related applications such as perceived depth adjustment for stereoscopic video and intermediate view generation for multiview auto-stereoscopic displays. However, the input depth map for DIBR may suffer from errors and noise, which could seriously impact the rendered view quality. In order to reduce the errors and suppress the noise in the depth map, a local depth image enhancement technique is proposed that leverages trellis-based optimization techniques. A cost function is used to evaluate candidate depth values based on the stereo cost as well as color and depth consistency. Sparse depth features are also used in the enhancement process. The experimental results show notable subjective improvements in terms of rendering quality.

*IEEE International Conference on Image Processing (ICIP)*

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.



# A LOCAL DEPTH IMAGE ENHANCEMENT SCHEME FOR VIEW SYNTHESIS

*Yongzhe Wang, Dong Tian, Anthony Vetro*

Mitsubishi Electric Research Labs  
201 Broadway, Cambridge, MA 02139, USA

## ABSTRACT

The quality of the depth map is crucial for depth image based rendering (DIBR) which enables a variety of advanced 3D video related applications such as perceived depth adjustment for stereoscopic video and intermediate view generation for multiview auto-stereoscopic displays. However, the input depth map for DIBR may suffer from errors and noise, which could seriously impact the rendered view quality. In order to reduce the errors and suppress the noise in the depth map, a local depth image enhancement technique is proposed that leverages trellis-based optimization techniques. A cost function is used to evaluate candidate depth values based on the stereo cost as well as color and depth consistency. Sparse depth features are also used in the enhancement process. The experimental results show notable subjective improvements in terms of rendering quality.

*Index Terms*— depth enhancement, DIBR, 3D video

## 1. INTRODUCTION

Three-Dimensional Video (3DV) is the next milestone for the video industry after the massive deployment of High Definition TV (HDTV). A 3D display is capable of presenting a same scene with a slightly different viewing angle for each eye, thereby provide the binocular depth cue for the viewer. Given the stereoscopic video content, an advanced 3D TV system should be capable of generating virtual views at a different viewing position than the stereo pair. This would enable a variety of applications for 3D TV such as the perceived depth adjustment for the stereoscopic display and the intermediate view generation for multiview auto-stereoscopic displays.

Depth image based rendering (DIBR) is a technique to generate virtual views, which typically requires a depth image of the scene to be available [1]. It is noted that depth images are likely to exhibit noise, either resulting from errors in computing stereo correspondences or through direct acquisition by range sensors; this noise may produce artifacts in the rendered views. Secondly, per-pixel depth images cannot always represent depth discontinuities that typically occur at object boundaries, which is another source of artifacts in the rendered views [2]. Thirdly, compression

artifacts would be introduced when the depth map is coded in a lossy manner and such coding error may affect the depth map quality as well. In addition, depth maps may be even captured or coded at a lower spatial resolution than the corresponding texture, and thus need be upsampled at the receiver side.

Recently, many depth image filtering techniques have been proposed in order to suppress the noise or enhance the quality of the upsampled depth images. Min, et al. [3] proposed a joint global mode filtering technique based on a joint histogram which considers the color similarity and the spatial distances between reference and neighboring pixels. In the case of depth image up-sampling, a joint bilateral filter is designed in [4]. These methods all assume that only the color image at the corresponding view point is available. In [5] a depth image enhancement technique for DIBR using Scale Invariant Feature Transform (SIFT) at the virtual view position has been proposed to improve the consistency among multiple reference depth images.

In this paper, we propose a local depth image enhancement technique which references the color information of the stereo image pair. For a given pixel in the depth map, a set of candidates is first identified before the cost of each candidate is calculated and finally the depth pixel value is updated if the cost of the alternative candidate is smaller than the current depth value. A set of sparse depth features serve as an additional source of candidate depth values. This enhancement technique leverages the trellis-based optimization techniques proposed earlier for robust view synthesis [6], but differs significantly in the underlying principle and manner in which the techniques are applied. In the present work, the ambiguities on depth values are resolved at a much earlier stage and a richer set of information is utilized including sparse depth features. This work also revisits the calculation of costs for each node and considers additional constraints.

The rest of this paper is organized as follows. Section 2 gives an overview of the DIBR system where our algorithm is used. Section 3 describes the proposed depth image enhancement algorithm in detail. Section 4 shows the simulation results using our algorithm. Section 5 concludes the paper.

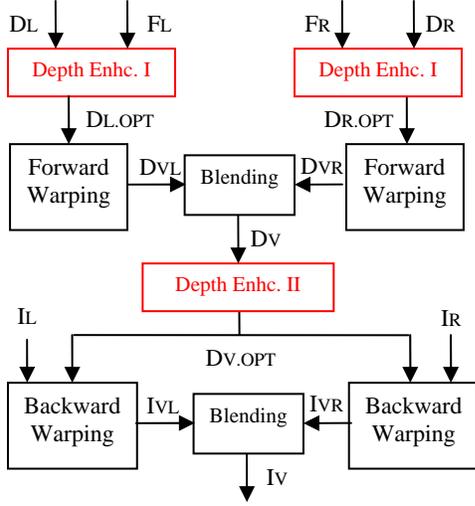


Fig. 1. Proposed depth enhancement technique integrated in a DIBR system

## 2. SYSTEM OVERVIEW

Our proposed depth map enhancement technique is designed and integrated in a DIBR system. There are two categories of color image warping methods for DIBR in general: forward warping and backward warping [7]. With forward warping, the color sample values are mapped to the virtual view directly; with backward mapping, the depth images that correspond to the reference view are first warped to the virtual view, before the warped depth images are used to find a corresponding color sample value in the reference view for each sample location in the virtual view. Backward warping is assumed in this paper for DIBR, while forward warping can be easily adapted to our algorithm as well.

Fig. 1 shows the structure of the backward warping DIBR system integrated with our proposed depth enhancement technique. In the figure, the symbol I, D, and F represent the color image, depth image, and the sparse depth feature, respectively; the subscript L, R and V represent the left view, right view and virtual view respectively; and OPT indicates the optimized (enhanced) depth image.

In this system, the depth image that corresponds to the reference view is not used directly in the backward warping process of the color image, and there might be additional errors introduced in the forward warping process of the depth image to the virtual view. So, our proposed depth enhancement technique consists of two stages: Stage I is conducted on the depth maps DL and DR that correspond to the reference view positions, while Stage II is carried out on the warped depth image DV that corresponds to the virtual view position. Stage I has the advantage of having more information to reference, specifically each pixel in the depth map has the collocated pixel in the color image to reference.

In addition, Stage I also has the sparse depth features FL and FR as input given by the sparse depth estimation. Sparse depth features are a set of feature points with associated depth values. Given a stereo color image pair, the sparse depth estimation algorithm outputs a set of sparse depth features with their associated depth value. In this paper, we use a sparse depth estimation algorithm based on KLT [8]. The complexity is low relative to the dense depth estimation process since only the correspondences of a sparse set of pixel need to be calculated; the reliability is high considering the correspondences are found on distinctive corners rather than texture-less areas.

Note that depth and disparity can be converted to each other given the camera parameters. In this paper, we use disparity with proper scaling as the depth image format.

## 3. PROPOSED DEPTH IMAGE ENHANCEMENT ALGORITHM

Each stage of the proposed depth image enhancement approach consists of the same steps with slight variations, as will be introduced in the following subsections. For each pixel in the depth image, a set of depth candidate is first selected; then the cost for each candidate is calculated; finally, the depth value with the lowest cost is assigned to the current pixel. A flow chart is shown in Fig. 2.

### 3.1. Depth candidate selection

In our proposed method, the input depth value specified in the dense depth image is always selected as the first candidate,  $d_1$ . This value may be incorrect and thus lead to artifacts or inconsistencies with respect to the input images. Therefore, additional candidates are considered.

Different strategies may be used to determine the additional depth candidates. For instance, neighboring depth values were considered in [6]. In this work, we propose a more sophisticated and efficient method of selecting candidates that makes use of the color information. Specifically, alternative depth values are selected based on the minimal collocated color intensity difference of the neighboring pixels. The problem is stated as follows:

$$q^* = \arg \min_{q \in Q} f(I_{N(p)}, I_{N(q)}) \quad (1)$$

where  $p$  is the current pixel position,  $q$  is a candidate pixel position in  $Q$ ,  $N(\cdot)$  represents a neighborhood around the current pixel or candidate pixel position,  $I$  is the color image, and the function  $f(\cdot)$  is a dissimilarity metric. In this work, the Mean Absolute Difference (MAD) is used as the dissimilarity metric and the neighborhood size is  $3 \times 3$ . Then the position  $q^*$  is mapped to a corresponding depth or disparity value with respect to the position of the current pixel  $p$ .

In Stage I, the candidate depth values are selected as follows. The first alternative depth,  $d_2$ , is obtained by

constraining  $Q$  to be the five neighboring positions to the left of the current pixel in the same line. The second alternative depth,  $d_3$ , is obtained by constraining  $Q$  to be the five neighboring positions above the current pixel in the same column. The last alternative depth,  $d_4$ , is the associated depth value of the nearest sparse depth feature.

The candidate depth values for Stage II are selected based on the depth values only. In particular,  $d_2$  is selected as the median value from preceding five pixels of the same line and  $d_3$  is selected as the median from the same column.

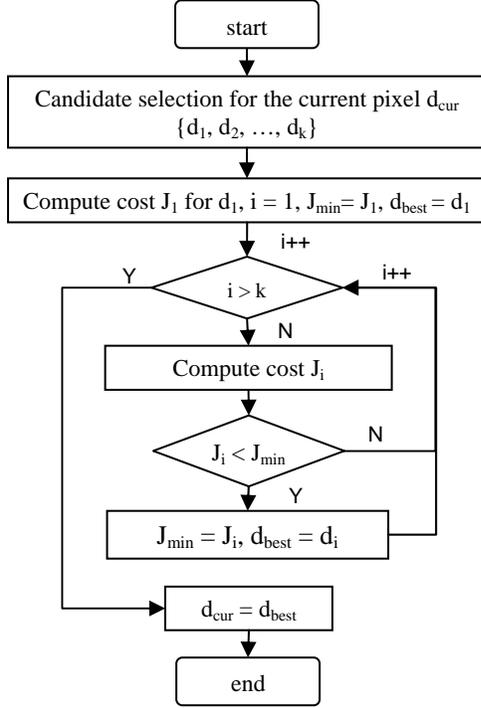


Fig. 2. Flow chart of the depth enhancement technique

### 3.2. Cost calculation

Once the depth value candidates are selected, the cost of each of them is calculated. For Stage I, the cost function is comprised of three measures:

- stereo cost: measures the consistency in color intensity between two views;
- color difference cost: measures the color intensity difference between the current location and the candidate pixel location; and
- depth difference cost: measures the depth value difference between the current pixel and the candidates.

Eq. (2) is defined considering the above measurements.

$$J_i = J_{stereo} + J_{color\_diff} + J_{depth\_diff} \quad (2)$$

The first sub-cost function  $J_{stereo}$  is defined as follows:

$$J_{stereo} = mrMAD[I_{N(p)}, I'_{N(p')}] \quad (3)$$

where  $p$  is the current pixel position,  $p'$  is the corresponding position in the other view  $p' = p + d_i$ ,  $I$  and  $I'$  are the color

images that correspond to the current view position and the other view position, respectively.  $N(\cdot)$  is a  $9 \times 9$  window, and  $mrMAD$  is the mean-removed mean absolute difference [9].

The second term  $J_{color\_diff}$  imposes an additional color consistency constraint on the alternative depth value candidates:

$$J_{color\_diff} = k_c * MAD[I_{N(p)}, I_{N(q)}] \quad (4)$$

where  $q$  is the position of the alternative depth candidate, and  $k_c$  is a constant weighting factor.

The third term  $J_{depth\_diff}$  is defined by:

$$J_{depth\_diff} = k_d * |d_1 - d_i| \quad (5)$$

where  $k_d$  is a weighting factor that controls the weight of this sub-cost function. The values of the weighting factors are determined empirically.

The cost function is slightly modified to adapt to Stage II. The first sub-cost function is replaced by:

$$J_{stereo} = MAD[I'_{N(p)}, I''_{N(p'')}] \quad (6)$$

where  $I'$  and  $I''$  are the left and right color image,  $p'$  and  $p''$  are the corresponding positions of the current pixel position  $p$  in the reference views.  $p'$  and  $p''$  can be obtained by:

$$\begin{aligned} p'_u &= p_u + l' * d_u \\ p''_u &= p_u - (1 - l') * d_u \end{aligned} \quad (7)$$

where  $u$  can be replaced by  $x$  or  $y$  indicating whether it is a horizontal or a vertical component of the position,  $l'$  represents the virtual view position relative the left view in-between the stereo pair.

We use  $mrMAD$  instead of  $MAD$  in Stage I since we have observed that the mean-removed MAD is more robust to illumination variations, while yielding matches that better maintain structural information. In this way,  $mrMAD$  can be viewed as a looser condition so that the disturbance from transparency or illumination conditions would be limited and hence correct depth values can be selected and carried to the next stage. The stronger  $MAD$  condition is applied in Stage II to refine the depth value. The second sub-cost function,  $J_{color\_diff}$ , is omitted from the cost calculation of  $J_i$  in Stage II since there is no corresponding color image at the virtual view.

## 4. EXPERIMENTS

The proposed depth enhancement approach is implemented within a proprietary system that estimates depth from a stereo pair. This method is compared to view synthesis using the original dense depth images. It is noted that when choosing the depth candidates for a next sample, the referenced depth values from previous samples may have been modified by the proposed approach.

Simulations were conducted on several sequences, including *GT\_fly* and *Undo\_dancer* from Nokia, and *Poznan\_hall* from Poznan University. For each sequence, a virtual view is synthesized at 0.4 of the baseline distance from the left view in-between the stereo pair. Fig. 3 shows the sample images of the synthesized results with Stage I or/and II. It can be observed that each stage reduces some of the artifacts and the combination of the both stages yields the best result. Figs. 4-5 show sample images of the synthesized results without and with the full proposed depth enhancement technique. It can be observed that the artifacts are significantly reduced, especially along the object boundaries. In addition, vertical edges are well preserved which are sensitive to the human eyes.

### 5. CONCLUSIONS

In this paper, a novel depth image enhancement technique for view synthesis is proposed. It consists of two stages, one on the depth images that correspond to the reference views, and another on the warped depth image that corresponds to the virtual view. In each stage, a set of candidates is first identified for each pixel. Sparse depth features are also utilized in determining candidates depth values. Then, the cost of each candidate is calculated by referencing the depth map itself and the stereo color image pair. The pixel value is updated by the candidate with the lowest cost. The experimental results shows that the rendering artifacts can be reduced and the vertical structures can be better preserved in the virtual view image.

### 6. REFERENCES

[1] H. Y. Shum, S. B. Kang, and S. C. Chen, "Survey of Image-Based Representations and Compression Techniques," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 11, November 2003.

[2] J. Shade, S. Gortler, L.W. He, and R. Szeliski, "Layered Depth Images," *Proc. SIGGRAPH'98*, Orlando, FL, July 1998.

[3] D. Min, J. Lu, M. N. Do, "Depth Video Enhancement Based on Weighted Mode Filtering," *IEEE Transactions on Image Processing*, vol. 21, no. 3, March 2012.

[4] J. Kopf, M. F. Cohen, D. Lischinski and M. Uyttendaele, "Joint Bilateral Upsampling," *ACM SIGGRAPH 2007*.

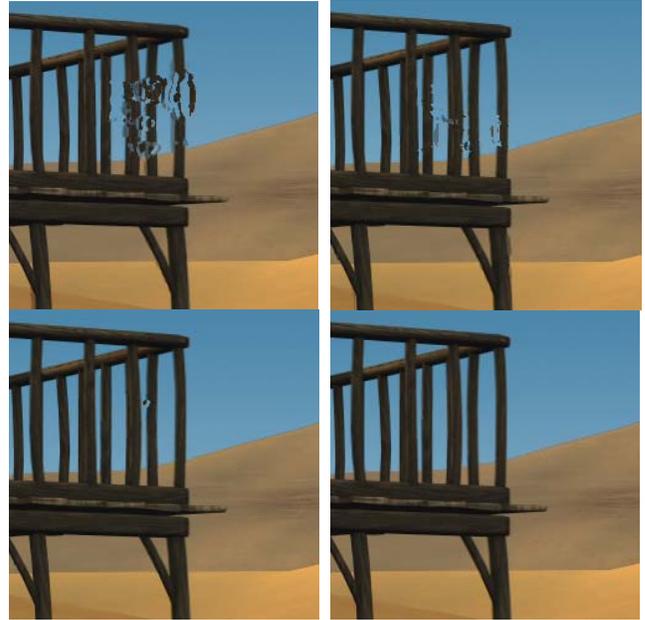
[5] H. Li, and F. Markus, "Sift-based Improvement of Depth Imagery," *Proc. IEEE International Conference on Multimedia and Expo, Barcelona, Spain, July 2011*.

[6] D. Tian, A. Vetro, and M. Brand, "A Trellis-based Approach for Robust View Synthesis", *Proc. IEEE International Conference on Image Processing*, Brussels, Belgium, Sept. 2011.

[7] D. Tian, P. L. Lai, P. Lopez, and C. Gomila, "View Synthesis for 3D Video," *Proc. SPIE*, vol. 7443, August 2009.

[8] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *International Joint Conference on Artificial Intelligence*, pages 674-679, 1981.

[9] D. S. Turaga, and T. Chen, "Estimation and Mode Decision for Spatially Correlated Motion Sequences," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 10, October 2001.



Upper Left: w/o depth enhancement, Upper Right: w/ Stage I, Lower Left: w/ Stage II, Lower Right: w/ Stage I and II  
Fig. 3. Sample view synthesis results of *GT\_fly*



Left: w/o depth enhancement Right: w/ depth enhancement  
Fig. 4. Sample view synthesis results of *Undo\_dancer*



Left: w/o depth enhancement Right: w/ depth enhancement  
Fig. 5. Sample view synthesis results of *Poznan\_hall*