ITERATIVE DEPTH PROPAGATION FOR MULTI-VIEW IMAGE GENERATION FROM LOW-RESOLUTION DEPTH MAP

Anonymous VCIP submission

Paper ID: 270

ABSTRACT

Various 3D video formats for free-viewpoint television have been proposed. Among them, the video-plus-depth format consisting of a high-resolution color image and a low-resolution depth map receives a lot of attention owing to its advantages, such as high coding efficiency. However, this format needs to enhance the depth image and to synthesize virtual views. In this paper, we propose a method to enhance the resolution of the depth image and generate multi-view images from depth information. We refine the low-resolution depth image using iterative propagation. This method is able to match depth boundaries with color boundaries. After enhancing the initial depth image, we generate multi-view images using 3D warping, color mapping, and hole filling. Experiment results show that the proposed method generates high-quality multi-view images, despite of the low resolution of the depth image relative to the color image resolution.

Index Terms— iterative propagation, depth image enhancement, multi-view image, interpolation

1. INTRODUCTION

The interest on the 3D video service has gradually increased since the applications of the 3D video service are numerous, ranging from movie to robot vision [1, 2]. The current service is based on stereoscopic images which are simultaneously captured by two cameras at different positions. The stereoscopic images provide binocular parallax to viewers, but its viewpoint is limited. In order to resolve this viewpoint problem, multi-view images were proposed, providing various viewpoints.

Although the multi-view provides 3D perception to viewers at various viewpoints, its performance significantly depends on the number of original views. A large number of views guarantee high quality viewing experience, but it is physically impossible to capture multitudinous views in practice. Commercial available cameras are too bulky and too expensive.

Therefore, researchers have developed virtual view synthesis methods. Among various approaches, the most common algorithm is depth image-based rendering (DIBR) [3]. DIBR is associated per-pixel depth information, consisting of the two stages.

At first, points in the image are projected into the 3D space via camera parameters and corresponding depth data. Afterward, these points in the 3D space are re-projected on a desired viewpoint. Since depth data is the basis of DIBR, the quality of depth images is very important. Therefore, one of the issues of DIBR is how to get high quality depth images.

The algorithms of estimating depth data are generally classified into two categories: passive depth sensing and active depth sensing [4-6]. The former estimates depth data indirectly from multiple images and it includes stereo matching and depth from focus. The later usually uses physical sensors such as laser and structured light pattern. The active depth sensing commonly provides more accurate depth images, but its capturing condition is restrictive.

To resolve the disadvantages and to combine the advantages of both approaches, a system coupled with multiple cameras and physical depth sensors have been introduced.

Recently, methods for enhancing the resolution of depth images are attracting much attention. In many cases, the resolution of such depth images is much smaller than that of the corresponding color images. There are mainly two reasons.

The first reason is the low performance of physical sensors, especially depth cameras. Most commercially available depth cameras provide depth images only up to 640×480, while the common resolution of color images is 1280×960 or higher.

The other reason is depth compression. Moving Picture Expert Group (MPEG) has tried to improve depth coding efficiency by down-sampling at an encoder side and up-sampling at a decoder side [7]. Since textures in depth images are relatively simple compared to those of color images, down- and up-sampling do not seriously distorts their original quality.

However, the low-resolution depth maps significantly degrade the quality of multi-view images synthesized by DIBR. To this end, we propose an iterative propagation

method to enhance the depth resolution for multi-view generation. This method refines input depth maps with consideration of discontinuities of the corresponding colors. Afterwards, we generate virtual views at desired viewpoints from the enhanced depth images.

2. RELATED WORK

In this section, we briefly introduce conventional work related to multi-view image generation and depth image enhancement. The common way to get multi-view images is capturing images by multiple cameras. Although this method guarantees the high quality of captured images, the amount of data extremely increases according to the number of cameras.

As an alternative, a multi-view generation method was proposed based on DIBR. Various formats support this method such as single-view color and depth images and multi-view color and depth images. Of course, it is also possible to estimate depth images at a decoder side, but it is difficult and time-consuming for users. Sending depth images with color images can lift the burden at the decoder side.

In such formats, we can additionally reduce the amount of data by sending a low-resolution depth image. As mentioned in the previous section, the quality of depth images affects the viewing quality of synthesized images. Therefore, the performance of depth image enhancement is very important in this format.

The simple approaches were bilinear, nearest neighbor, and bicubic interpolation methods, which consider only depth images. These algorithms show reliable results, but they lead to errors around depth discontinuities.

In order to handle this problem, Diebel *et al.* proposed a Markov random field(MRF)-based interpolation and defined adaptive weights according to corresponding colors [8]. These weights are used to control smoothness prior in their model. Afterwards, Yang *et al.* suggested a post-processing algorithm based on the bilateral filter [9]. This method iteratively aggregates the costs of input depth images using the filter, and enhances both its spatial resolution and depth precision.

Jung *et al.* proposed a confidence-based MRF model to sharpen depth edges. This method shows better results than others, but inaccurate depth can be propagated to holes sometimes [10].

3. ITERATIVE PROPAGATION

In this paper, we propose a algorithm for generating multi-view images from low-resolution depth images. The proposed method includes depth up-sampling using iterative propagation and view synthesis using 3D warping.

For synthesizing multi-view images, the resolution of depth images should be equal to that of color images. Hence, it is required to enhance the resolution of the input depth images. At first, we generate a gradient map of the color image as shown in Fig. 1(a), and this map is regarded as a guide map during the process. In order to enhance this map, we refine the gradient map using

$$grad(i) = \begin{cases} \alpha \cdot grad(i) & if \quad grad(i) > mgrad(i) \\ \alpha^{-1} \cdot grad(i) & else \end{cases} \quad (1)$$

where $grad(i)$ means the gradient value at a pixel i, and $m\,grad$ represents a mean gradient over a local support. Fig. 1(b) shows the refined gradient map.

(a) initial gradient map (b) refined gradient map

Fig. 1. Gradient maps

Afterwards, we calculate confidence values for neighbors as

$$conf(i, i_n) = \exp(-(w_c(i, i_n) + \lambda(iter) w_g(i, i_n))) . \quad (2)$$

A weight w considers two properties; the pixels in the same object have similar colors and depth values in a local region, and the objects are distinct from each other with noticeable edges. w_c is for the first property and can be expressed as

$$w_c(i, i_n) = \frac{(I(i) - I(i_n))^2}{\gamma_c} \quad (3)$$

where I represents the intensity value and r_c is a constant parameter for color sensitivity. The second weight w_g can be expressed as

$$w_g(i, i_n) = \frac{\max\_grad(i, i_n)}{\gamma_g} \quad (4)$$

where *max_grad* is the gradient value of the color image on the shortest path from i to i_n. r_g is a constant parameter for adjusting gradient sensitivity.

Figure 2 shows the concept of the maximum gradient. Even though n_2 is more close to the current pixel n, its maximum gradient values is greater. Thus, n_1 has a larger weight according to (2).

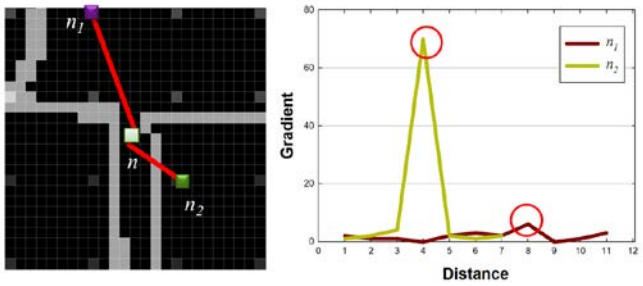


Fig. 2. Maximum gradient values of neighbors

As shown in Fig. 2, we can adaptively allocate proper weights to reference pixels in the local supports. Even though each texture has different shape and color, the allocated weights reflect them well. Especially, although there are many objects with similar colors, the proposed method allocates higher weights to the same object that contains the center position. We can enhance the initial depth map by applying (2) to all pixels.

4. MULTI-VIEW GENERATION

After enhancing the depth map, we generate multi-view images using DIBR. If all camera parameters of an input image are given, we calculate the corresponding pixel position between input and virtual images. When a point **M** in world coordinate is projected to a camera, a pixel **m** in the image can be found by (5). The representations of a single point $\mathbf{M} = [\mathbf{X}\ \mathbf{Y}\ \mathbf{Z}\ \mathbf{1}]^T$ and a projected point $\mathbf{m} = [\mathbf{x}\ \mathbf{y}\ \mathbf{1}]^T$ are the homogeneous form. The relation between m and **M** can be expressed as

$$\mathbf{m} = \mathbf{A}[\mathbf{R} \mid \mathbf{t}]\mathbf{M} \tag{5}$$

where **A** is the intrinsic camera parameter, and **R** and **t** are the extrinsic camera parameters. With (5), we project a pixel md in the enhanced depth image to the world coordinate using

$$\mathbf{M_d} = \mathbf{R_d}^{-1} \cdot \mathbf{A_d}^{-1} \cdot \mathbf{m_d} \cdot d(\mathbf{m_d}) - \mathbf{R_d}^{-1} \cdot \mathbf{t_d} \tag{6}$$

where the representations of $\mathbf{A_d}$, $\mathbf{R_d}$, and $\mathbf{t_d}$ stand for camera parameters of the input view. $d(\mathbf{m_d})$ is a depth value of the pixel of $\mathbf{m_d}$. After projection of md, we re-project $\mathbf{M_d}$ onto the desired view position using

$$\mathbf{m_t} = \mathbf{A_t}[\mathbf{R_t}|\mathbf{t_t}]\mathbf{M_d} \tag{7}$$

where the representations of $\mathbf{A_t}$, $\mathbf{R_t}$, and $\mathbf{t_t}$ represent camera parameters of the desired view. After that, we map color information to the synthesized depth image and fill the holes caused by viewpoint shifting with the inpainting algorithm.

5. EXPERIMENT RESULTS

In order to evaluate the performance of the proposed system, we experimented on three single-view images, tsukuba, cone, and teddy. We downloaded these test images and their corresponding depth maps from the website of Middleburry. The original depth maps were down-sampled by a factor of 8 for each axis and up-sampled again with the nearest neighbor algorithm.

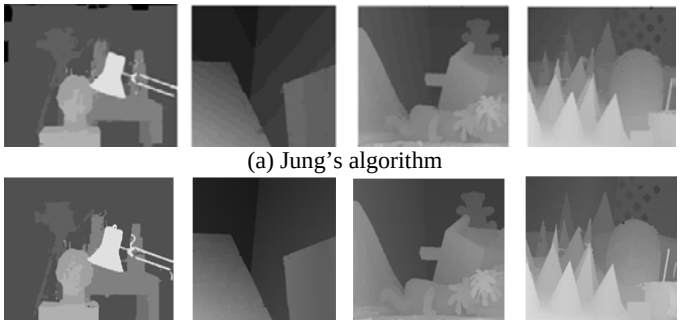
Figure 3 show the enhanced depth maps by Jung's method and the proposed algorithm, respectively. The initial depth maps have very rough boundaries due to their low resolution, but the enhanced depth maps have well matched edges with color images. While the initial depth map causes serious distortions around boundaries, the proposed method improves the viewing quality of the synthesized images.

Table 1 shows the error ratios of various interpolation methods. As shown, the proposed method provides the most accurate results compared to others.

Figure 4 shows the input color images, depth maps (red-boxes), synthesized images, and overlapped images. Even though the input depth maps are much smaller than the color images, the proposed algorithm can generate the natural multi-view images with proper view intervals.

Table 1. accuracy comparison

%	Tsukuba		Teddy		Venus		Cones	
Factor	16	64	16	64	16	64	16	64
Bilinear	8.64	14.98	11.04	18.89	1.63	3.33	14.04	23.61
Bicubic	7.96	13.03	10.42	17.33	1.35	2.77	12.81	22.27
Diebel's	5.12	9.68	8.33	14.50	1.24	2.69	7.52	14.40
Yang's	2.56	6.95	5.95	11.50	0.42	1.19	4.76	11.00
Jung's	1.62	2.81	**5.01**	**7.33**	0.42	1.02	5.59	8.78
Proposed	**1.31**	**2.30**	5.05	7.34	**0.25**	**0.37**	**4.75**	**7.01**



(a) Jung's algorithm

(b) The proposed algorithm

Fig. 3. Comparison of the enhanced depth maps: Tsukuba, Venus, Teddy, and Cones

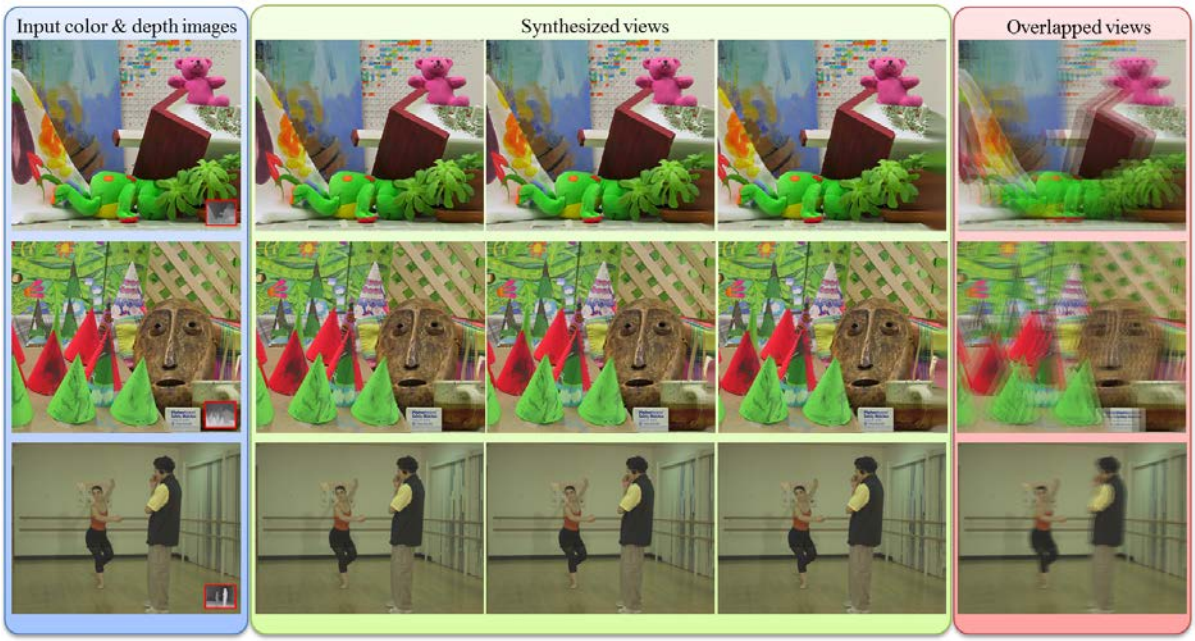


Fig. 4. Synthesized multi-view images: Teddy, Cones, and Ballet

6. CONCLUSION

In this paper, we have presented a new approach to generate a multi-view image from a low-resolution depth map. We have enhanced the resolution of the depth images with consideration of colors and gradient values. Thereafter, we have generated the multi-view image by depth map warping, color mapping, and hole filling. Experimental results have shown that our scheme can provide more accurate high-resolution depth maps compared with the previous methods. It is expected that the proposed algorithm can reduce the amount of data while keeping image quality. Therefore, our proposed method is very useful for various 3D multimedia applications and display devices.

7. ACKNOWLEDGEMENT

8. REFERENCES

[1] Z. Yuhua and Z. Tong, "3D Multi-view Autostereoscopic Display and Its Key Technologie," *Asia-Pacific Conference on Information Processing*, pp. 31-35, 2009.

[2] A. Smolic, K. Mueller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, "3D Video and Free Viewpoint Video - Technologies, Applications and MPEG Standards," *IEEE International Conference on Multimedia and Expo*, pp. 2161-2164, 2006.

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

[4] D. Fofi, T. Sliwa, and Y. Voisin, "A Comparative Survey on Invisible Structured Light," *SPIE Machine Vision Applications in Industrial Inspection*, pp. 90-98, 2004.

[5] S. B. Gokturk, H. Yalcin, and C. Bamji, "A Time-Of-Flight Depth Sensor - System Description, Issues and Solutions," *IEEE Conference on Computer Vision and Pattern Recognition Workshop*, pp. 35-35, 2004.

[6] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms," *International Journal of Computer Vision*, vol. 47, no. 1, pp. 7-42, April 2002.

[7] A. Smolic and D. McCutchen, "3DAV Exploration of Video-based Rendering Technology in MPEG," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 3, pp. 348-356, March 2004.

[8] J. Diebel and S. Thrun, "An Application of Markov Random Fields to Range Sensing," *Advances in Neural Information Processing Systems*, vol. 18, pp. 291-298, 2006.

[9] Q. Yang, R. Yang, J. Davis, and D. Nister, "Spatial-Depth Super Resolution for Range Images," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2007.

[10] J.I. Jung and Y.S. Ho, "Depth Image Interpolation Using Confidence-based Markov Random Field," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 4, pp. 1399-1402, Nov. 2012.