Reduced Reference Quality Metric for Depth Images

Thanh-Ha Le[1], Seongjo Lee[2], Seung-Won Jung[2], and Chee Sun Won[3]

[1] University and Engineering and Technology, Vietnam National University, Hanoi
[2] Department of Multimedia Engineering, Dongguk University, Seoul, South Korea
[3] Department of Electronics and Electrical Engineering, Dongguk University, Seoul, South Korea

Abstract. In this paper, a new quality metric for depth images is proposed. Unlike the conventional depth metrics which require the additional information such as the ground truth depth image or a stereo image pair, the proposed quality metric demands only a single camera image and its corresponding depth image. In this work, we first empirically observe that the depth distortion is closely related to the local image characteristics. Based on the observation, we introduce a method to assess the local depth distortion for the edge and non-edge regions. Then, the local distortion is adaptively weighted by the Gabor filter and added up to the quality metric for the depth image.

Keywords: Depth image, Gabor filter, image quality assessment, reduced reference, quality metric.

1 Introduction

Depth image, often called the depth map, plays a fundamental role in 3-D video and free viewpoint video applications [1]. In the stereo vision, a right-view image can be reconstructed by using a left-view image and its corresponding depth image. Besides, in the multi-view video, an arbitrary viewpoint image can be generated by interpolating or extrapolating the given images using the depth images. Due to the importance and applicability of the depth image, the estimation of the depth image has been extensively researched in the last decades [2].

In the classical computer vision field, the depth image estimation is formulated as a stereo matching problem. Although the state-of-the-art stereo matching techniques [3, 4] could significantly improve the estimation accuracy of the depth image, inaccurate depth images are often produced because of occlusion and large homogeneous regions. In order to alleviate the inherent difficulty of stereo matching, the depth camera is used to capture the depth of the scene. Owing to the advances in the hardware of the depth camera, an accurate depth image can be achieved but the noise is inevitable and the depth range is limited.

Regardless of the depth image acquisition methods, the quality assessment of the depth image is required. Specifically, to evaluate the performance of the stereo matching algorithms or the depth cameras, the objective depth quality metric (DQM) should be defined. One simple solution is to compare the estimated depth image with the ground truth depth image [2]. This full reference DQM (FR-DQM) can ideally measure the accuracy of the depth image; however, the ground truth depth image is not available in practice. Another possible DQM is to gauge the quality of the

© Springer-Verlag Berlin Heidelberg 2015
James J. (Jong Hyuk) Park et al. (eds.), *Advanced Multimedia and Ubiquitous Engineering*,
Lecture Notes in Electrical Engineering 352, DOI: 10.1007/978-3-662-47487-7_18

reconstructed image obtained by the depth image. For instance, the right-view image is approximated by warping the left-view image using the depth image, and the warped image is compared with the original right-view image. However, such image pairs are not always obtainable in the depth-image-based rendering (DIBR) [5] and the depth camera applications. Although the above two DQMs have a certain limitation, a more general DQM has not been widely studied.

Recently, an FR quality metric for the geometrically distorted images showed that the severe geometric degradation of the image structure is originated from the distortion of the displacement field into the orthogonal direction of the image bars and edges [6]. Motivated by this work, a new reduced reference DQM (RR-DQM) is proposed in this paper. The proposed RR-DQM requires only one camera image and its corresponding depth image. Since the objective is to measure the accuracy of the depth image, the camera image is considered as the side or reduced information of the depth. Throughout the experimental studies, the major sources of depth distortion are found and a suitable measurement is developed using the Gabor filter and the smallest univalue segment assimilating nucleus (SUSAN) detector [7].

The rest of this paper is organized as follows. The proposed RR-DQM is presented in Section 2. The conclusions are given in Section 3.

2 Proposed Depth Quality Metric

In general, the depth image is used to render or synthesis images for multi-view and 3-D video applications. Thus, the same local distortion of the depth image does not

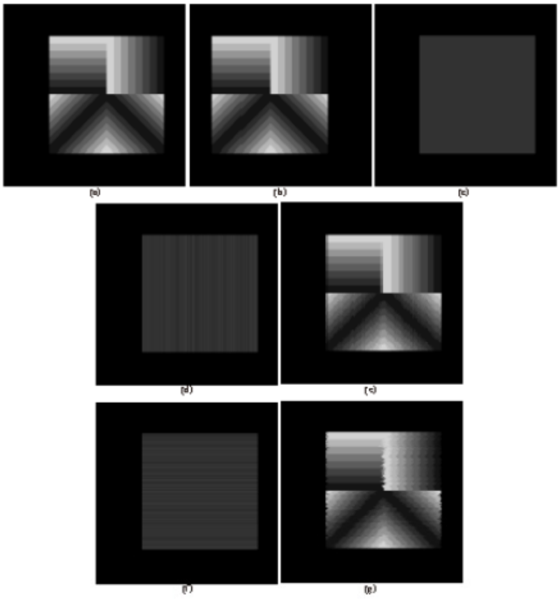

Fig. 1. Synthetic example: (a) left-view image, (b) right-view image, (c) ground truth depth image, (d) depth image with distortion along horizontal direction, (e) compensated left-view image obtained using (b) and (d), (f) depth image with distortion along vertical direction, (g) compensated left-view image obtained using (b) and (f).

equally affect the resultant images. In other words, the local distortion of the depth image should be jointly considered with the local image characteristics.

To demonstrate this supposition, a pair of simple synthetic stereo images of the size 400x400 is generated as shown in Figs. 1(a) and (b). Here, the square of the size 256 x256 containing four representative directional edges as shown in Fig. 1(a) is shifted by 50 pixel distances to the left direction as shown in Fig. 1(b). In other words, the pixels inside the square have the same horizontal disparity as shown in Fig. 1(c). The background is assumed to be located in the infinite distance, and thus there is no horizontal disparity. Since the other directional disparities can be eliminated by using the image rectification technique, the assumption of the horizontal disparity case only is simple but widely adopted [2].

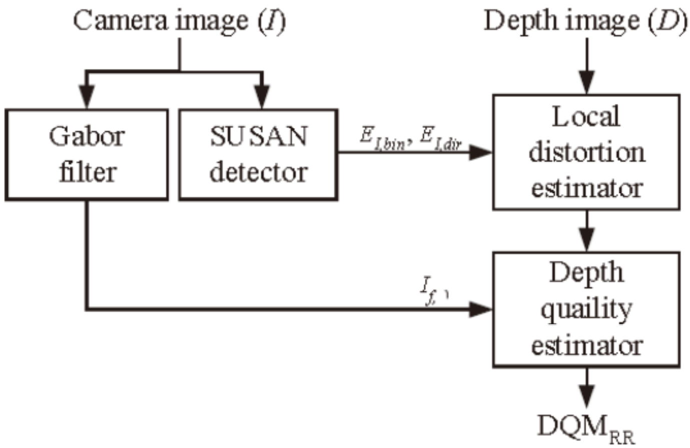

Fig. 2. Flowchart of the proposed method

The disparity values inside the square are then distorted in the horizontal and vertical directions, as shown in Figs. 1(d) and (f), respectively. Specifically, for the depth image with distortion along horizontal direction, one row of the zero-mean uniformly distributed random noise with a variance of 10 is constructed. Its length is the same as the width of the square, 256, and this row is added to the all rows in the square region of the depth image. Thus, the depth values are the same in the vertical direction and only depth distortion exists along the horizontal direction as shown in Fig. 1(d). The depth image with distortion along vertical direction is constructed in a similar manner.

Given the depth image and the right-view image, the left-view image can be reconstructed. Specifically, the pixels in the left-view image are found from the pixels in the right-view image in which the pixel positions are determined according to the horizontal disparity values in the depth image. From the compensated left-view images as shown in Figs. 1(e) and (g), we can first see that the horizontal image edges are not visually deteriorated. Since only the horizontal disparity is assumed, the different types of distortion can only change the start and end positions of the horizontal edges. Thus, the local distortion of the depth image in the horizontal edge regions can be disregarded. It is also confirmed that the distortion in the compensated images is prominent when the depth value varies along the image edges. For instance, the vertical image edges are severely damaged when the depth image has distortion along vertical direction as shown in Figs. 1(f) and (g).

From the above observations, it is found that the effect of the local depth distortion is strongly dependent on the local image characteristics. Thus, the relation between the depth distortion and image characteristics should be exploited to measure the accuracy of the depth image. Figure 2 shows the flowchart of the proposed RR-DQM. In the proposed method, Gabor filter is used to differently weight the local image structures. In addition, the SUSAN edge detector [7] is employed to attain the edge information of the image. In particular, the SUSAN detector is known to robustly estimate image edges and their edge direction.

Let $E_{I,bin}$ and $E_{I,dir}$ denote the binary edge map and the edge direction map of the image I obtained by the SUSAN detector, respectively. For the simplicity, $E_{I,dir}$ is quantized to represent only the horizontal, vertical, left diagonal, and right diagonal directions. Then, for the non-edge positions, the local depth distortion is measured by the average difference of the depth values in the local neighborhood. On the other hand, for the edge positions, the depth variation along the edge direction is assessed to take the edge distortion or deformation into account.

Let $D_{I,D}$ represent the depth distortion map obtained from the binary edge map $E_{I,bin}$ and the depth image D, defined as

$$D_{I,D}(x,y)=\begin{cases}\frac{1}{n(N)}\sum\limits_{(u,v)\in N}\left|D(x,y)-D(x+u,y+v)\right| & ;\text{ if } E_{I,bin}(x,y)=0\\ \left|D(x,y)-\frac{1}{2}\left(D(x+u_1,y+v_1)+D(x+u_2,y+v_2)\right)\right| & ;\text{ otherwise}\end{cases}, \quad (1)$$

where N is the set of 8-neighborhood positions and $n(N)$ is the cardinality of N. Thus, for the non-edge positions, the mean absolute difference (MAD) of the depth values is used to describe the local depth distortion. For the edge positions, the average of the two neighboring depth values along the edge direction is compared with the depth value of the current position. Here, (u_i,v_i) is set according to the edge direction. For instance, (u_1,v_1) = (0,1) and (u_2,v_2) = (0,-1) for the vertical edge. Since the gradual change along edge direction tends to be the actual depth variation, the central difference is adopted to assess the local depth distortion.

Until now, the methods of obtaining the Gabor energy and the depth distortion map have been described. The rationale behind the proposed quality metric of depth images is that depth discontinuities along the image edges are the main factor of the depth distortion. Thus, both the local image characteristics (encoded in the Garbor energy) and the local depth distortion (encoded in the depth distortion map) are used to define the final distortion map, $\overline{D}_{I,D}$. To this end, $\overline{D}_{I,D}$ is defined by combining the depth distortion map and the Gabor energy as follows:

$$\overline{D}_{I,D}(x,y)=\sum_{\theta\in\Theta} w_\theta\cdot I_{f,\theta}(x,y)\cdot D_{I,D}(x,y), \quad (2)$$

where θ={0°, 45°, 90°, 135° } and w_θ is the weight of the direction θ. Similar definition for the image geometric distortion can be found in [6]. Figure 3 shows the final distortion maps obtained by using Figs. 1(d) and (f), where w_θ = { 1,0.5,0,0.5} for four directions in the same order of θ. Since the vertical edge is the most sensitive and the horizontal one is less vulnerable to the local depth distortion as we confirmed, w_θ is determined accordingly. By comparing Figs. 1 and 3, it can be seen that the final distortion maps correspond with the visual geometric deterioration resulted from the depth distortion.

The remaining problem is to determine the metric, RR-DQM, of the depth image distortion. The RR-DQM is defined by pooling the all distortion values in $\overline{D}_{I,D}$ except for the outlier regions,

$$DQM_{RR} = \frac{1}{n(\Psi_1)}\left(\sum_{(x,y)\in\Psi_1} \left|\overline{D}_{I,D}(x,y)\right|\right), \tag{3}$$

where ψ_1 is a set of all pixel positions in the image excluding outlier regions and $n(\psi_1)$ is the cardinality of ψ_1.

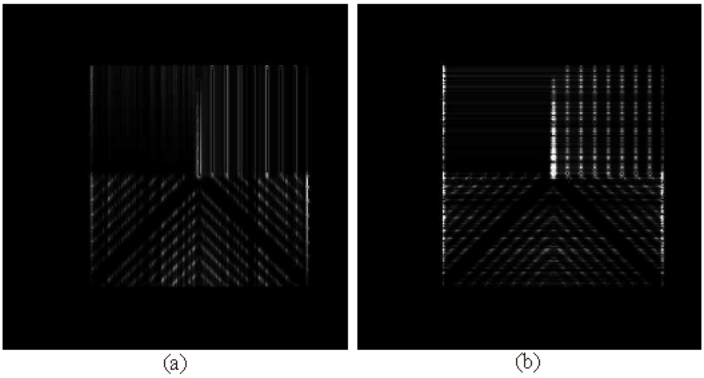

Fig. 3. Final distortion maps corresponding to (a) Fig. 1(d) and (b) Fig. 1(f)

3 Conclusion

We proposed a depth quality assessment technique that does not require the ground truth depth image or the stereo image pair. Based on the analysis using the synthetic image, the strong correlation between the local depth distortion and the local image characteristic is verified. Then, the depth distortion is measured depending on the edge directions. In addition, the Gabor filter is used to adaptively weight the local depth distortion. The experimental results show that the proposed metric closely approximates the conventional depth quality metrics that necessitate the additional information.

Acknowledgement. This work was supported by the basic research projects in natural science in 2012 of the National Foundation for Science \& Technology Development (Nafosted), Vietnam (102.01-2012.36, Coding and communication of multiview video plus depth for 3D Television Systems). This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2015-H0301-15-1021) supervised by the NIPA (National IT Industry Promotion Agency).

References

1. Smolic, A., Mueller, K., Merkle, P., Kauff, P., Wiegand, T.: An overview of available and emerging 3D video formats and depth enhanced stereo as efficient generic solution. In: Proc. 27th Conference on Picture Coding Symposium, pp. 389–392 (2009)
2. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comput. Vision 47(1-3), 7–42 (2002)
3. Klaus, A., Sormann, M., Karner, K.: Segment-based stereo matching using brief propagation and a self-adapting dissimilarity measure. In: Proc. IEEE Conference on Pattern Recongnition, pp. 15–18 (2006)
4. Wang, Z.-F., Zheng, Z.-G.: A region based stereo matching algorithm using cooperative optimization. In: Proc. IEEE Conference on Computer Vision and Pattern Recongnition, pp. 1–8 (2008)
5. Fehn, C.: Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. In: Proc. SPIE, vol. 5291, pp. 93–104 (2004)
6. D'Angelo, A., Zhaoping, L., Barni, M.: A full-reference quality metric for geometrically distorted images. IEEE Trans. Image Process. 19(4), 867–881 (2010)
7. Smith, S.M., Brady, J.M.: SUSAN-a new approach to low level image processing. Int. J. Comput. Vision 23(1), 45–78 (1997)
8. Scharstein, D., Szeliski, R., `http://vision.middlebury.edu/stereo`
9. Yang, Q., Yang, R., Davis, J., Nister, D.: Spatial-depth super resolution for range images. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)