

# No-reference Depth Assessment Based on Edge Misalignment Errors for T+D Images

Sen Xiang, *Student Member, IEEE*, Li Yu, *Member, IEEE*, and Chang Wen Chen, *Fellow, IEEE*

**Abstract**—The quality of depth is crucial in all depth-based applications. Unfortunately, the error-free ground truth is often unattainable for depth. Therefore, no-reference quality assessment is very much desired. This paper presents a novel depth quality assessment scheme that is completely different from conventional approaches. In particular, this scheme focuses on depth edge misalignment errors in texture-plus-depth (T+D) images and develops a robust method to detect them. Based on the detected misalignments, a no-reference metric is calculated to evaluate the quality of depth maps. In the proposed scheme, misalignments are detected by matching texture and depth edges through three constraints: spatial similarity, edge orientation similarity, and segment length similarity. Furthermore, the matching is performed on edge segments instead of individual pixels, which enables robust edge matching. Experimental results demonstrate that the proposed scheme can detect misalignment errors accurately. The proposed no-reference depth quality metric is highly consistent with the full-reference metric, and is also well-correlated with the quality of synthesized virtual views. Moreover, the proposed scheme can also use the detected edge misalignments to facilitate depth enhancement in various practical texture-plus-depth based applications.

**Index Terms**—Depth quality assessment, No-reference, Edge misalignment, 3D Video, Free-viewpoint video.

## I. INTRODUCTION

In recent years, video technologies have advanced rapidly, and many novel applications, such as 3D video (3DV) and free-viewpoint video (FVV), have emerged. 3DV provides depth perception and brings more realistic experiences compared with traditional 2D video. FVV allows audiences to control their viewing positions and angles, making interactions between audiences and content providers possible [1]. In these systems, texture-plus-depth (T+D) is a widely adopted format due to its flexibility in generating arbitrary views by using depth-image-based-rendering (DIBR). In DIBR, the depth is crucial to generating high quality virtual views. Unfortunately, in contrast to texture images, depth maps are usually obtained through calculations rather than direct capture, which often introduces errors [2]. These errors impair subsequent depth processing and further applications. Therefore, it is imperative and important to detect the errors and assess the quality of the original depth maps.

This work was supported in part by the National Natural Science Foundation of China (Grant No. 61231010), 863 High-tech R&D Program (Nos. 2015AA015901, 2015AA015903), US NSF Grant (No. 1405594) and US NIH Grant (No. 5R21AA021924).

S. Xiang and L. Yu are with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074 China. (e-mail: xiangsen@hust.edu.cn, hustlyu@hust.edu.cn).

C. W. Chen is with the Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, NY 14260 USA (e-mail: chenwcw@buffalo.edu).

## A. Image Quality Assessment

Usually, image quality assessment (IQA) can be achieved by using both subjective and objective approaches [3]. Subjective methods require the subjects to grade their perceptual feelings and record their opinion scores. Objective approaches use computational models to predict the perceived image quality. By following this classification, current studies on quality assessment for depth-related videos/images can also be divided into these two categories.

Current subjective approaches to depth quality assessment mainly explore indirectly the relationship between human perception and the quality degradation of stereo or T+D videos. Hewage *et al.* [4] compressed T+D sequences with different quantization parameters, and obtained subjective assessment scores for virtual views. Saygili *et al.* [5] investigated asymmetric coding for stereo videos and conducted subjective tests to determine the distortion threshold for the auxiliary view. Speranza *et al.* [6] studied the coding issues of frame-compatible videos in which the distortions of stereo views were compared with the results of simulcast. Qi *et al.* [7] applied just-noticeable-distortion (JND) to stereoscopic videos, and subjective testing was applied for performance evaluation. Gutierrez *et al.* [8] and Tatematsu *et al.* [9] investigated various factors in the compression and transmission of 3DTV and 3DV-based applications. The influence of these factors on user experiences was evaluated with subjective scores. In general, these studies on subjective evaluation focused on compression and transmission without providing quantitative relationships between human perception and image/video distortion. Moreover, subjective evaluation is rather inconvenient, time-consuming and expensive in practice.

Compared with subjective testing, objective assessment is more convenient in implementation and can be more widely applied. Based on the reference information utilized in the assessment, objective IQA methods can be divided into three categories: full-reference (FR), reduced-reference (RR) and no-reference (NR).

FR-IQA methods usually assess the image quality by comparing the distorted image with its reference in a pixel-by-pixel fashion. The most commonly used metrics, MSE/PSNR and SSIM [10], belong to this category. For depth maps, a widely admitted metric is the percentage of bad matching pixels (PBMP) [2]. To compute PBMP, a pixel is checked with the ground truth, and it will be classified as a bad matching pixel if its disparity error exceeds a one-pixel-width. Because depth maps are not directly displayed but are instead used for view synthesis, researchers have explored the relationship

between the depth quality and the synthesized view quality. Zhao *et al.* [11] proposed the no-synthesis-error model to determine the depth error thresholds. If the depth error of a pixel is within its threshold, then the error will not cause distortion in the synthesized view, and the error can be ignored. In fact, this idea is essentially quite similar to PBMP [2]. Kim *et al.* [12] derived the relationship between depth errors and the corresponding distortions in rendered views. Yuan *et al.* [13][14] proposed a concise distortion model to describe the distortions in given reference views and rendered virtual views. Based on this model, the authors further studied joint texture-depth bit allocation and distortion removal. Later, Fang *et al.* [15] proposed an analytical model to predict the errors in synthesized views. Specifically, this model does not require the actual operation of performing DIBR. By taking advantage of the existing 2D image quality metrics, several researchers have investigated their performances in stereoscopic viewing [16]–[18]. Several researchers improved the existing metrics to make them more suitable for depth. Malpica *et al.* [19] and Ryu *et al.* [20] proposed SSIM-like indexes for depth. Lin *et al.* [21] calculated frequency-integrated metrics for 3D image quality evaluation, which integrated the binocular integration behaviors into the existing 2D objective metrics. From the perspective of the human vision system, perception-based quality metrics for stereo videos were proposed. Lu *et al.* [22] proposed a quality metric for stereo images after investigating frequency dominance on 3D perception. Silva *et al.* [23] developed a stereoscopic structural distortion measure for video compression. Shao *et al.* [24] proposed an FR metric for stereo images based on binocular perception and combination properties. All of these metrics were claimed to be highly consistent with the subjective assessment results.

In contrast to FR-IQA, RR-IQA metrics do not use ground truth directly. Instead, they extract key features from the ground truth. By utilizing these features, RR-IQA can assess distorted images without knowing the original image information. For example, Hewage *et al.* [25] used depth edges to evaluate compression distortions. Similarly, Nur *et al.* [26] utilized a bilateral filter to extract depth features. Malekmo-hamadi *et al.* [27] investigated distortions in compression and transmission, where the distortion model was established by using neighboring and edge properties of the depth.

In general, ground truth is required in FR-IQA and RR-IQA metrics, either directly used as the reference or indirectly used for feature extraction. Unfortunately, because depth maps are obtained by calculation rather than direct capturing, errors cannot be avoided. Therefore, error-free reference depth maps are usually unattainable. For this reason, it is more reasonable to assess depth maps with NR-IQA methods, which do not need references. Su *et al.* [28] proposed a stereoscopic/3D IQA framework, by utilizing both univariate and generalized bivariate natural scene statistics models. Chen *et al.* [29] proposed a model to assess the perceptual quality based on the extracted 2D and 3D features. The model combined features from binocular disparities and estimated neural activities. Park *et al.* [30] developed a model-based neuronal and statistical framework to predict the level of visual discomfort. These models were generally tested with known distortions such as

blur, JPEG, JPEG2K, white noise and video compression. However, in depth estimation, the distortion types are unknown. Furthermore, these schemes focused on the stereo views, either captured or synthesized, but no research was conducted directly on the depth maps.

We also note that most of the existing literature focuses on compression and transmission errors. There is little study in terms of the quality of the original generated depth maps. In fact, in contrast to textures, most depth maps are obtained through calculations [2], and as a result, errors cannot be avoided in the original depth maps. Therefore, it is important and meaningful to detect those errors and improve the depth quality before subsequent operations, such as compression and transmission, are applied. To the best of our knowledge, Milani *et al.* [31] presented the only work on NR quality assessment for depth maps. The basic idea is that segments in depth and texture maps should coincide with each other. Under this assumption, the authors performed image segmentation on both texture and depth maps, and checked the coincidences between the two types of segments. This approach, in reality, is overstrict because a smooth surface can have rich textures inside, which in turn will lead to differences between texture and depth segments.

## B. Overview of the Proposed Work

In this paper, we propose a novel scheme to assess depth images in T+D format. The basic idea of the proposed scheme is completely different from conventional IQA methods, in which the image quality is assessed by checking the intensity/color information. Instead, we believe that depth maps are 2D representations of 3D geometries, and hence, the shape/geometry distortions are more important. Specifically, the misalignment errors between corresponding depth and texture edges are the focus of this paper. The reason lies in the following two aspects. On the one hand, the fundamental principle of depth estimation leads to more errors at boundaries, including edge misalignments. On the other hand, boundary errors are more likely to cause artifacts in virtual views [32]. In our preliminary work [33], depth and texture edges are matched based on their locations and directions. Based on this work, we present in this paper a novel scheme for detecting depth edge misalignments. Afterward, the detected errors are used to calculate an objective metric, which quantifies the depth quality.

In principle, texture maps are directly captured, which enables their edges to be more reliable. By referring to texture edges, we can detect misalignment errors in depth edges. Unfortunately, texture edges not only outnumber depth edges greatly, but also distribute irregularly, which makes it quite challenging to achieve accurate matching between depth and texture edges. To solve this problem, we develop a new scheme called two-step edge-segment-based-matching (two-step-ESBM) in this paper. In the first step, depth edges are matched to texture edges to obtain a coarser solution. In the second step, within the coarser solution, texture edges are matched to depth edges to remove redundancies. In both steps, similarities between edge segments are determined by using three edge features, i.e., their spatial locations, orientations and

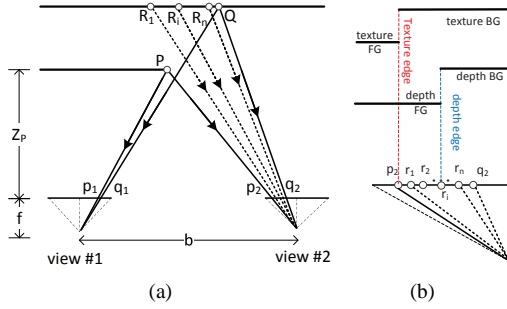


Fig. 1. Sketch of stereo matching. (a) The principle of stereo matching. (b) Enlarged imaging plane of view #2. 'FG': foreground, 'BG': background.

edge lengths. Finally, pixels outside the corresponding edge segments are considered to be bad points, and their proportions are utilized to assess the depth quality.

The contributions of this paper are the following. (1) We propose in this research a no-reference objective metric that is based on edge misalignment errors to assess the depth quality in T+D format. (2) A new scheme called two-step-ESBM is developed to robustly detect misalignments between depth and texture edges. (3) A criterion that consists of edge spatial locations, edge orientations, and edge segment lengths is proposed to evaluate the similarity between edge segments. (4) An adaptive edge segmentation scheme is used to break edges into segments according to their curvature.

The remainder of this paper is organized as follows. In Section II, we analyze the edge misalignment errors for depth maps. The proposed two-step-ESBM scheme is described in detail in Section III. Experimental results are shown in Section IV. Finally, Section V concludes this paper with a summary.

## II. DEPTH EDGE MISALIGNMENT

In depth maps, the boundaries are quite important due to two reasons. On the one hand, it is challenging to obtain accurate boundaries in depth estimation. On the other hand, boundary distortions in depth maps cause errors that are more severe in applications. In this section, we will analyze the boundary properties in detail.

### A. Boundary Errors in Depth Estimation

In Fig. 1, we illustrate the principle of depth estimation with a simple scene, which consists of a foreground and a background. A 3D point  $P$  projects to views #1 and #2, and the projections are  $p_1$  and  $p_2$ , respectively. Given the known positions of  $p_1$  and  $p_2$ , the depth of  $P$  can be calculated as

$$Z(P) = f_c b |\overrightarrow{p_1 p_2}|^{-1} \quad (1)$$

where  $f_c$  and  $b$  are the camera focal length and the baseline between the two views, respectively. Here,  $|\overrightarrow{p_1 p_2}|$  is the disparity, which denotes the location difference between  $p_1$  and  $p_2$ .

In depth estimation, the most crucial step is to identify the projections of a 3D point in stereo views, which is known as stereo matching. For points such as  $P$  and  $Q$  in Fig. 1(a), the relationship is clear because they have projections in both views. However, at the object boundaries, background points such as  $R_1, R_2, \dots, R_n$  in Fig. 1(a) are occluded by the

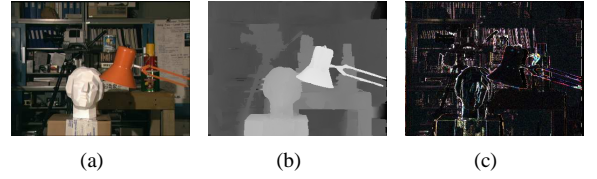


Fig. 2. Errors in the synthesized view. (a)(b) Reference texture and depth maps, respectively. (c) Absolute errors of the virtual view. The errors in (c) are magnified 10 times for display.

foreground. Therefore, these points only have projections in view #2, which are denoted as  $r_1, r_2, \dots, r_n$ . In this case, no corresponding pixel pairs can be found for them in the stereo views, which leads to errors in the estimated depth maps. Specifically, in Fig. 1(b), the texture edge is  $p_2$ , but the corresponding depth edge could be located at  $r_i$ . As a result, misalignments appear between the texture and depth edges, and the pixels between the two types of edges have incorrect depth values. Because occlusions are quite common in stereo vision, misalignment errors widely exist in depth maps. Therefore, it is of great importance to detect edge misalignments.

### B. Error analysis in DIBR

In DIBR, the depth errors cause artifacts in virtual views. Specifically, the relationship between the errors in the depth maps and virtual views is given in [32]:

$$\epsilon(I_v) \approx f'(I_r) \epsilon(d_r) \quad (2)$$

where  $\epsilon(I_v)$  and  $\epsilon(d_r)$  denote the errors in the virtual view and the reference depth, respectively.  $f'(I_r)$  is the first order derivative of the reference texture.

As analyzed in section II-A, boundary pixels are more likely to have errors in depth estimation, which indicates that  $\epsilon(d_r) \neq 0$ . Moreover, colors often change between the objects and  $f'(I_r) \neq 0$ . Therefore, the virtual view error  $\epsilon(I_v) \neq 0$ , which indicates apparent artifacts. In contrast, the situation inside objects is much better. We further consider two cases for the pixels which are inside objects according to the texture distribution. First, many neighboring pixels in the same object have similar colors, which makes  $f'(I_r)$  approximately 0. As a result, the virtual views would be correct, i.e.,  $\epsilon(I_v) = 0$ , even if the depth is inaccurate. Second, in the worst case, texture-rich regions inside objects have depth errors, and hence, the virtual views will suffer from errors. This relationship is exactly true, but the influence of these errors is limited. The reasons are two-fold. On the one hand, in most cases, neighboring pixels have similar colors, and the proportion of texture-rich pixels is quite limited. On the other hand, in stereo matching, rich textures can provide adequate information for detecting corresponding pixels in stereo views, and thus, their depth values are likely to be accurate, i.e.,  $\epsilon(d_r) = 0$ . For these two reasons, the influence of the depth errors in the texture-rich regions is quite limited.

From the above error analysis in DIBR, it can be summarized that depth distortions around the boundaries would dominate over the errors inside the objects. Fig. 2 illustrates an example of the error distribution in DIBR. In Fig. 2(b), errors exist in both the inner parts and the boundaries. In Fig. 2(c), most of the artifacts appear at the boundaries.

Because of the importance of depth edges, various methods [34]-[36] have been proposed to realign the depth and texture edges. However, in these studies, the depth edges were often simply realigned to neighboring texture edges, which is not rigorous or accurate. In our work, we aim to achieve reliable texture-depth edge matching, to enable the distorted depth pixels to be determined accurately. Afterward, further applications can be achieved in high quality.

### C. Texture and Depth Edge Distribution

As analyzed in section II-A, depth distortion can be obtained if the corresponding texture and depth edges are known. However, in practice, this goal is quite challenging. Depth edges indicate only object boundaries, and thus, they are sparse and clear. Nevertheless, texture edges can be caused by a variety of factors, such as object boundaries, shadows and color patterns. As a result, texture edges not only outnumber depth edges greatly, but are also irregularly located. In such a condition, accurate matching between texture and depth edges is difficult.

To achieve accurate edge matching, we analyze the edge distribution in a region and classify the edges into three cases. Case 1: both depth and texture edges exist. This situation occurs at real object boundaries. The foreground and background colors are different and distinguishable. Case 2: only texture edges exist. This case occurs inside the objects, where the texture edges are not caused by object boundaries, but result from other factors such as color patterns and shadows. These texture edges will impair correct texture-depth edge matching and lead to false results. Case 3: only depth edges exist. This situation occurs at object boundaries, and the colors of the foreground and background are similar. This situation rarely occurs because a basic assumption in stereo matching is that pixels in a texture-uniform region are likely to have similar depth values. Therefore, we do not consider this case in our work.

By summarizing the three cases above, depth edges can be regarded as a subset of texture edges. An accurate edge matching algorithm should report misalignments in case 1, and it should also exclude false alarms as discussed in case 2. In this paper, we propose a novel matching scheme that is based on edge segments to achieve this goal.

## III. THE PROPOSED SCHEME

### A. Framework

The framework of the proposed scheme is illustrated in Fig. 3. This scheme is completely different from conventional IQA methods because we focus on edge misalignments rather than intensities/colors. The basic assumption of the proposed two-step-ESBM is that texture and depth edges are projections of the same objects, and thus they have similar properties. Under this assumption, matching between these two types of edges is performed to detect misalignment errors. Based on the detected errors, we obtain a map of bad pixels, together with a metric that quantifies the depth quality.

The proposed scheme focuses on the object boundaries, and thus, edges are first extracted. Afterward, two-step-ESBM is

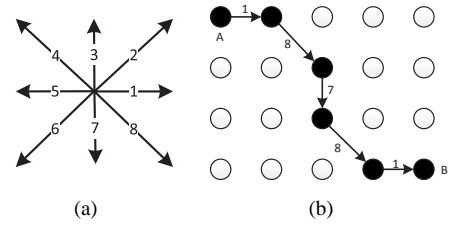


Fig. 4. Sketch of ESDV. (a) Direction of connectivity between two edge units. (b) An example of an edge segment and its direction.

performed to obtain the matching solution between the texture and depth edges. Specifically, the matching is based on the edge segments, instead of the pixels, to improve the robustness. The matching procedure consists of two steps. In the first step, we match depth edges to texture edges to obtain a coarser solution. Afterward, redundant matching pairs are removed from the coarser solution and one-to-one matching results are obtained. In both steps, the matching criterion is based on three factors: spatial similarity, edge orientation similarity, and segment length similarity between the edge segments. With the reported edge pairs, the distorted pixels are determined and treated as bad pixels. Furthermore, based on the bad pixels, a metric that quantifies the depth quality is proposed.

### B. Edge Detection

The proposed scheme is based on texture and depth edges, and thus, robust edge detection is the foundation of the subsequent steps. The Canny edge detector is adopted because it provides reliable edges with only a one-pixel-width. In the Canny detector, the threshold is a crucial parameter which directly determines the edges. We apply Otsu's algorithm [37] to the gradient magnitudes to achieve adaptive thresholding. Otsu's algorithm can minimize intra-class variances for both the edge pixel and the non-edge pixels. Therefore, we can obtain reliable edges from both the depth and the texture maps, which will serve as the inputs in edge segmentation.

### C. Edge Segmentation

Because texture and depth maps have noise and outliers, pixel-based matching between their edges can be easily impaired. As a result, the matching results will be inaccurate and unreliable. In contrast, when individual pixels are grouped into edge segments, the segment structure can provide constraints in the edge matching, which will improve the matching accuracy and robustness. Inspired by this idea, we propose an edge-matching scheme between the depth and texture edge segments rather than individual pixels.

In two-step-ESBM, edge segmentation is an important issue. If an edge is straight, then longer edge segments provide better constraints, which will lead to more accurate results. Whereas for highly curved edges, shorter segments capture local features better, which can enable more desirable matching solutions. Therefore, we introduce an adaptive edge segmentation approach according to the curvature of the edges.

1) *Quantization for Edge Curvature*: to quantify the edge curvature, we introduce the concept of edge segment direction variation (ESDV), which can be calculated as the following.

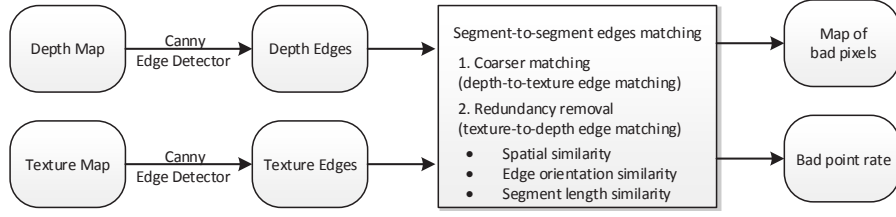


Fig. 3. Flowchart of the proposed two-step-ESBM.

- First, the direction of each edge unit between two pixels is defined in Fig. 4(a). For an edge segment  $s$  with  $n$  pixels, a vector  $\mathbf{eu}=[e_1, e_2, \dots, e_{n-1}]$  represents the edge direction, where the  $i^{th}$  element  $e_i$  denotes the direction of the edge unit between  $p_i$  and  $p_{i+1}$ .
- With  $\mathbf{eu}$ , the direction difference between adjacent edge units  $e_i$  and  $e_{i+1}$  can be calculated as

$$\Delta e_i = \min(|e_i - e_{i+1}|, 8 - |e_i - e_{i+1}|) \quad (3)$$

For  $e_i$  and  $e_{i+1}$ , the direction can change both clockwise and anti-clockwise, and the shorter path is adopted as  $\Delta e_i$  in (3). For an edge vector  $\mathbf{e}$  with  $n-1$  elements, the absolute direction variations constitute a vector  $\Delta \mathbf{e}(s)$  which consists of  $n-2$  elements.

- With the vector  $\Delta \mathbf{e}(s)$ , the ESDV of the edge segment can be calculated as

$$ESDV(s) = \sum_{i=1}^{n-2} \Delta e_i / (n-2) \quad (4)$$

$ESDV(s)$  indicates the average curvature of edge segment  $s$ . In a straight line, the directions of the edge units seldom change and the ESDV value is small. In contrast, the ESDV of a highly curved edge will be very large. Navigating along the edge from A to B in Fig. 4(b), we can obtain  $\mathbf{eu}(s)=[1, 8, 7, 8, 1]$ ,  $\Delta \mathbf{e}(s)=[1, 1, 1, 1]$ , and  $ESDV(s)=1$ .

2) *Adaptive Edge Segmentation*: with ESDV representing the edge curvature, we perform adaptive edge segmentation according to the flowchart in Fig. 5.

Assume that the current edge segment  $s$  has  $n$  pixels, and the next pixel is  $p$ . To decide whether  $p$  should be added to  $s$ , we calculate  $ESDV(s+p)$ . If  $ESDV(s+p) > T_{ESDV}$ ,  $p$  is added to  $s$ . Otherwise,  $p$  will not be added to  $s$ , but it will serve as the first pixel of a new edge segment. In straight edges,  $ESDV$  is small, and edge segments can keep growing. However, in highly curved edges,  $ESDV$  exceeds  $T_{ESDV}$  easily, and the edge segments will be short. To make  $T_{ESDV}$  adaptive to different scenes, we calculate  $ESDV$  with all of the edge units in the map and assign the result to  $T_{ESDV}$ . Furthermore, to prevent the edge segments from becoming too long or too short, two length limits are set. When  $s+p$  is shorter than  $L_{min}$ ,  $p$  is added to  $s$  without checking  $ESDV$ . At the same time,  $L_{min}$  is also the lower length bound of the valid edge segments. In other words, if an edge segment is shorter than  $L_{min}$ , it will be treated as an outlier and be ignored in the subsequent matching.  $L_{max}$  is the upper bound of the edge segment length. If  $s$  exceeds  $L_{max}$ , then  $p$  will not be added to  $s$ , but a new edge segment will be created. In addition, both

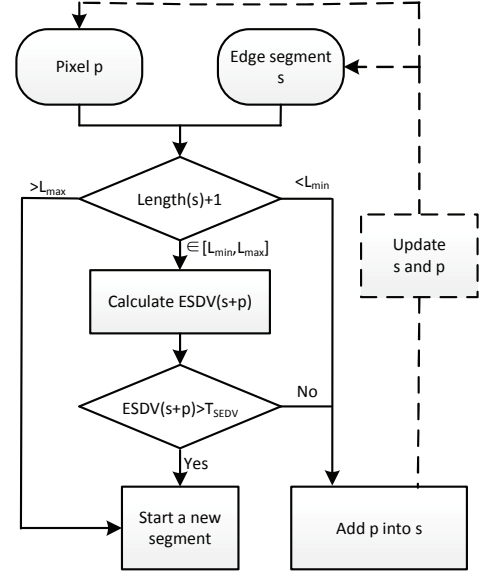


Fig. 5. Flowchart of the adaptive edge segmentation.

$s$  and  $p$  will be updated after the current iteration is ended. Specifically, the current segment  $s$  and pixel  $p$  are combined into the new  $s$ , and the next edge pixel becomes  $p$ .

We should note that edge segmentation is applied only to the ‘source edges’. To be specific, only depth edges are segmented in coarser matching, while in redundancy removal, only those texture edges in the coarser solution are segmented. Based on these edge segments, two-step-ESBM is performed to find the optimal matching pairs.

#### D. Edge-segment-based Matching (ESBM)

After edge segmentation, we detect the corresponding texture and depth edge pairs in two steps. In the first step, the depth edges are matched to the texture edges, and the second step reverses the role of the two types of edges. Without loss of generality, we use source edge and target edge for notation. In each single step, edge-segment-based matching (ESBM) is performed.

For each source edge segment  $s_s$ , all of the target edges in the search window are matching candidates. The search window is defined by dilating  $s_s$  with a  $W \times W$  template. Assuming that there are  $n$  disconnected target edges in the search window, we denote them as  $s_t^1, s_t^2, \dots, s_t^n$ , respectively. Because the corresponding texture and depth edges are projections of the same object, they have similar properties. Motivated by this fact, we use three important edge features: spatial location, orientation and length to evaluate the similarity between two edge segments  $s_s$  and  $s_t^i$ .



1) *Spatial Similarity*: the most basic assumption between the corresponding depth and texture edges is that they should have similar, or even the same, locations. Based on this assumption, the spatial similarity is defined as

$$S_{SP}(s_s, s_t^i) = \exp(-\alpha|\bar{x}_s - \bar{x}_t^i|), \quad (5)$$

where  $\bar{x}_s$  and  $\bar{x}_t^i$  are the average coordinates of pixels in  $s_s$  and  $s_t^i$ , and  $\alpha$  is the attenuation coefficient.  $S_{SP}$  indicates that closer edges have higher similarities. In other words, they are more likely to be the contours of the same object. In contrast, two edges are less likely to be the projections of the same object boundary as their distance increases.

2) *Edge Orientation Similarity*: apart from the spatial locations, the corresponding depth and texture edges should have similar directions. To evaluate the orientation similarity of two edges, we use the edge orientation histogram (EOH), which describes the direction distribution of an edge segment. To be specific, the EOH of edge segment  $s$  is defined as

$$\mathbf{h}(s) = \left[ \frac{\#\{e_1\}}{\#\{e_T\}}, \frac{\#\{e_2\}}{\#\{e_T\}}, \dots, \frac{\#\{e_7\}}{\#\{e_T\}}, \frac{\#\{e_8\}}{\#\{e_T\}} \right] \quad (6)$$

Here,  $\#\{e_i\}$  is the number of edge units in direction  $i$ , as defined in Fig. 4(a), and  $\#\{e_T\}$  denotes the total number of edge units. For example, the histogram of the edge segment shown in Fig. 4(b) is  $\mathbf{h}(s)=[0.4, 0, 0, 0, 0, 0, 0.2, 0.4]$ . With EOH, the orientation similarity between  $s_s$  and  $s_t^i$  can be computed as

$$S_{EO}(s_s, s_t^i) = \frac{\langle \mathbf{h}(s_s), \mathbf{h}(s_t^i) \rangle}{|\mathbf{h}(s_s)| \cdot |\mathbf{h}(s_t^i)|} \quad (7)$$

where  $\langle, \rangle$  is the inner product operator.  $S_{EO}$  ranges between 0 and 1, and larger values indicate higher similarities in the edge directions.

Please note that there is an ambiguity in computing EOH. In Fig. 4(b), if we reverse the direction and track the edge from B to A, both EOH and  $S_{EO}$  will be different. To remove such ambiguity,  $S_{EO}$  is calculated with both histograms and the larger value is adopted.

3) *Segment Length Similarity*: in the third similarity, we inspect the number of pixels in edge segments. Because the corresponding depth and texture edges are assumed to be contours of the same boundary, the number of pixels should be similar. Segment length similarity takes a simple format, as follows:

$$S_{SL}(s_s, s_t^i) = \frac{\min(\#\{s_s\}, \#\{s_t^i\})}{\max(\#\{s_s\}, \#\{s_t^i\})} \quad (8)$$

In  $S_{SL}$ ,  $\#\{s_s\}$  and  $\#\{s_t^i\}$  are the numbers of pixels in  $s_s$  and  $s_t^i$ , respectively. If the texture and depth edges are close in terms of their pixel numbers,  $S_{SL}$  will be close to 1.

4) *Overall Similarity*: the aforementioned similarities are proposed based on different features of edges, and thus, each of them has its own advantages and disadvantages. A robust edge matching scheme should consider all of the three edge features. Therefore, the overall similarity between  $s_s$  and  $s_t^i$  should account for the three similarities, and thus can be defined as

$$S_{all}(s_s, s_t^i) = S_{SP}(s_s, s_t^i) S_{EO}(s_s, s_t^i) S_{SL}(s_s, s_t^i) \quad (9)$$

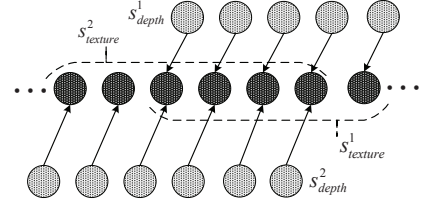


Fig. 6. An example of multiple-to-one matching between depth and texture edge segments in the coarser solution.

Based on  $S_{all}$ , we determine the best match for  $s_s$  from the candidates as follows

$$s_t^* = \arg \max_{s_t^i} S_{all}(s_s, s_t^i) \quad (10)$$

(10) indicates that, with the overall similarity  $S_{all}$ , we can find the optimal target edge segment  $s_t^*$  for the source edge segment  $s_s$ . Especially, among all of the candidates,  $s_t^*$  has the highest similarity with  $s_s$ .

Although (10) gives edge pairs that have the highest similarities, some of them could be unreliable due to having small  $S_{all}$  values. For example, when  $s_t^*$  is the only target edge in the neighborhood of  $s_s$ , it will always be the matching result even if these two edge segments are not similar at all. We use a similarity threshold to eliminate these possible fake matches. To be specific, a matching pair  $s_s$  and  $s_t^*$  will be considered to be reliable only when the similarity exceeds a threshold  $T_S$ .

#### E. Two-step-ESBM

With the aforementioned scheme, we use two-step-ESBM to obtain robust matching pairs. Because the texture edges outnumber the depth edges greatly, we first start from the depth edges and obtain matching results on the texture edges by using the proposed ESBM as introduced earlier in section III-D. In this step, the depth edges are the source edges, and the texture edges are the targets. This step excludes irrelevant texture edges, and a coarser solution that consists of matched edge pairs is obtained. The coarser solution has redundancies with multiple-to-one mappings between the edge segments. As shown in Fig. 6,  $s_{depth}^1$  and  $s_{depth}^2$  are mapped to  $s_{texture}^1$  and  $s_{texture}^2$ , respectively, where  $s_{texture}^1$  and  $s_{texture}^2$  share common pixels.

To remove the redundancy and obtain one-to-one matching solutions, we apply another round of ESBM. This time, the texture edges are used as sources and the depth edges are targets. Furthermore, reverse matching is performed only within the coarser solution because other pixels have already been proven to be irrelevant in the coarser matching. After the redundancies are removed, the whole procedure of the proposed two-step-ESBM is finished, and we can successfully achieve segment-to-segment matching between the texture and the depth edges.

#### F. Depth Quality Metric

After performing the two-step-ESBM, reliable matching results between texture and depth edges are obtained. With these matched edge pairs, the depth quality can be calculated.

As analyzed in section II-A, the texture edges are more reliable, and edge misalignments can indicate depth distortions.

To be specific, pixels between the corresponding texture and depth edges belong to different objects in the texture and the depth maps. Therefore, their depth values are incorrect, and we consider them to be bad pixels. Based on these pixels, the bad point rate (BPR), which quantifies the quality of the depth map, can be calculated as

$$BPR = \frac{\#\{P_B\}}{\#\{P_{test}\}} \quad (11)$$

where  $\#\{P_B\}$  and  $\#\{P_{test}\}$  are the numbers of the detected bad pixels and the tested pixels, respectively.  $BPR$  ranges between 0 and 1, and smaller  $BPR$  values indicate better depth quality.

In an implementation,  $BPR$  can be calculated for specified regions. For example, we can obtain the overall quality of a depth map by calculating

$$BPR_{all} = \frac{\#\{P_B(all)\}}{\#\{P_{test}(all)\}} \quad (12)$$

where  $P_{test}(all)$  denote all pixels in the depth map and  $P_B(all)$  is the set of bad pixels. Similarly, we can also specify the tested and bad pixels within the boundary regions, and thus, the boundary quality can be calculated as

$$BPR_{boundary} = \frac{\#\{P_B(boundary)\}}{\#\{P_{test}(boundary)\}} \quad (13)$$

where  $P_B(boundary)$  and  $P_{test}(boundary)$  are the bad pixels and the tested pixels within the boundary regions, respectively.

#### G. Application: Depth Error Correction

As analyzed in section II-A, depth errors widely exist in stereo matching. In addition to conventional stereo matching, structured light depth cameras, such as Kinect, also face quality problems. On the one hand, depth maps from such devices usually have holes at the object boundaries. On the other hand, structured light depth cameras cannot generate accurate edges either. Therefore, correcting these errors and improving the depth quality are important and imperative.

In depth error correction and hole filling, the joint bilateral filter (JBF) [38] is often applied. The principle of JBF can be represented as

$$d(p) = \frac{1}{K_p} \sum_{q \in \mathcal{N}_p} d(q) f_{SP}(|p - q|) f_I(I(p) - I(q)) \quad (14)$$

In (14),  $\mathcal{N}_p$  is the neighborhood of pixel  $p$ .  $f_{SP}$  is the spatial filter kernel, and  $f_I$  is the range filter kernel, while  $K_p$  is the normalization parameter. The largest advantage of JBF is its capability of preserving boundaries because  $f_I$  can suppress contributions from the pixels that have different colors from that of  $p$ .

With JBF and the proposed two-step-ESBM, we can improve the depth quality as follows. Given a depth map, we first apply JBF to the holes and obtain a complete map. Due to edge misalignments, bad pixels could be taken as reference samples in the filtering, which would lead to error propagation. Fortunately, bad pixels are detected by applying the proposed two-step-ESBM, which is the second step. Finally, we apply

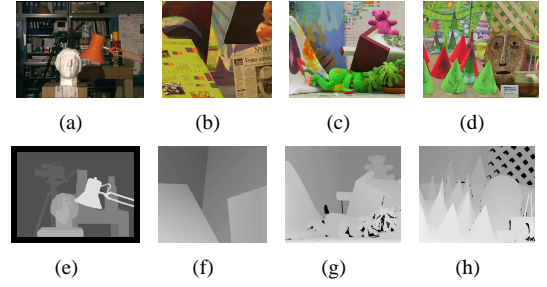


Fig. 7. Test scenes. (a)(b)(c)(d) Texture maps of Tsukuba, Venus, Teddy and Cones respectively. (e)(f)(g)(h) Ground truth depth maps. The black pixels in depth maps do not have valid depth values.

JBF again to the bad pixels, which will compute correct depth values for them. Furthermore, if the given depth map itself is complete, the hole filling in the first step is skipped. By following this approach, we can detect and attempt to correct depth edge errors, which in turn will improve the depth quality.

#### IV. EXPERIMENTAL RESULTS

Extensive experiments have been conducted to verify the proposed scheme. We first present in detail both the subjective and the objective results of the proposed depth quality assessment. Afterward, the relationship between the proposed depth quality metric and the DIBR results is presented. Then, the performance of the depth error correction is presented. Finally, we will analyze the influence of the parameters.

##### A. Experiment Design

To verify the proposed scheme, different depth maps of the same scene are needed, and the proposed scheme should distinguish their quality differences. The Middlebury [39] database provides depth maps that are estimated with various algorithms, which meets the requirements as stated earlier. We use four scenes: Tsukuba, Venus, Teddy and Cones, and their texture and ground truth depth maps are shown in Fig. 7. For each scene, depth maps estimated with nine stereo matching algorithms [40]–[48] are used in the experiments. From [40] to [48], the overall accuracy of the depth estimation algorithms decreases.

There are several parameters in the proposed scheme, and their values are determined based on the following concerns.  $W$  is obtained by analyzing the range of misalignments. As shown in Fig. 1, to match the depth edge  $r_i$  with the texture edge  $p_2$ , the search range  $sr$  should be larger than  $|\vec{r_i p_2}|$ , and its upper bound  $sr_{ub}$  equals  $|\vec{p_2 q_2}| - 1$ . Given the foreground disparity  $disp(FG)$  and the background disparity  $disp(BG)$ , there exists the equation  $|\vec{p_2 q_2}| = |disp(FG) - disp(BG)| + 1$ . In addition, the search window should extend to both sides of the depth edge, so  $W = 2sr_{ub} + 1 = 2|disp(FG) - disp(BG)| + 1$ . Combining this conclusion and the depth/disparity maps, we set  $W$  to 15.  $L_{min}$  excludes short edges from the matching, and it is set to 5 because edge segments shorter than that are believed to be caused by noise and outliers but not object boundaries.  $L_{max}$  is set to 30 and  $T_s$  is set to 0.125. Their values are determined empirically with one scene and applied to all of the tested scenes. Moreover, their values are determined based on the principle of generating accurate

and robust bad pixels. In section IV-F, we will investigate the influence of the parameters on the bad point rate in detail, where the issue of parameter selection will be further discussed.

### B. Subjective Results

Figs. 8 to 11 present the detected edge pairs and bad points<sup>1</sup>. In these figures, the marked depth and texture edge pairs are marked in blue and red, respectively. Other edges are not labeled in the final results for the following two reasons. On the one hand, as discussed in the second case of section II-C, most of the texture edges are not caused by object boundaries. Therefore, they are located far away from the depth edges and are not matched at all. On the other hand, a small number of the matched edge pairs do not exceed the similarity threshold  $T_s$ , and are rejected from the results. Based on the detected texture and depth edge pairs, bad points in between them are marked in green.

For each scene in Figs. 8 to 11, when the depth quality degrades, as shown from subfigure (a) to subfigure (i), more pixels are marked in green, which indicates a growing number of detected bad pixels. In the worst cases, as shown in subfigures (i), where the depth maps are severely distorted, the proposed scheme can report dense bad points along the depth edges. These matching results demonstrate that the proposed scheme can distinguish depth quality differences.

The different properties among the scenes also have effects on the performance of the proposed scheme. In particular, the proposed scheme performs better on Tsukuba and Venus than on Cones and Teddy. The reasons for the performance difference are two-fold. First, Tsukuba and Venus have fewer depth layers and larger color differences between the objects. Therefore, their depth maps are of better quality, and the depth edges are also clearer and simpler. In contrast, Teddy and Cones have more depth layers, color-like objects and highly slanted planes. These factors degrade the depth quality and also impair the proposed edge matching. As a result, fewer errors are reported in Teddy and Cones than in Tsukuba and Venus. Second, besides geometry differences, depth errors also distribute differently in the four scenes. In the depth maps of Tsukuba and Venus, the majority of errors are caused by edge misalignments. However, in Teddy and Cones, depth errors are more widely distributed. Specifically, the stereo images of Teddy and Cones are captured with large baselines, and thus, some of the objects are recorded in only one view. As a result, these objects cannot obtain correct depth values. For example, in Fig. 11, all of the estimated depth maps do not have the cones which should be located near the left borders of the images. In this case, the proposed scheme loses its foundation of detecting depth errors.

Although the proposed scheme is effective in detecting misalignments in most cases, it still encounters failure cases. First, distortions in texture-homogeneous regions cannot be detected because there are no texture edges in these regions.

<sup>1</sup>To illustrate the details more clearly, we provide enlarged figures at [http://ei.hust.edu.cn/lab/mcnc-lab/doc/enlarged\\_results.pdf](http://ei.hust.edu.cn/lab/mcnc-lab/doc/enlarged_results.pdf)

TABLE III  
PEARSON CORRELATION BETWEEN BPR/ERR AND PBMP

	Tsukuba	Venus	Teddy	Cones
Err_all [33]	0.84	0.55	<b>0.79</b>	0.75
PBR_all	<b>0.96</b>	<b>0.93</b>	0.77	<b>0.88</b>
Err_boundary [33]	0.91	<b>0.98</b>	<b>0.97</b>	0.78
PBR_boundary	<b>0.98</b>	0.96	0.83	<b>0.80</b>

Fortunately, according to (2), these errors do not have significant influence on virtual views. The second kind of failures occur at the weak edges, which separate the pixels with similar colors/intensities. For example, the doll and the leaves in Teddy have similar colors and depth values. Therefore, both the texture and the depth edges are difficult to extract. In addition to these two cases, some parts in the depth maps, such as the missing cones in Fig. 11 and the lamp holder in Fig. 8, are too distorted to provide useful geometry information. As a result, the proposed scheme is not applicable to these extreme cases.

### C. Objective Results

Based on the detected bad points, the objective quality of the depth is computed. The results are provided with three metrics, the percentage of bad matching pixels (PBMP) proposed in [2], the error rate (Err) proposed in [33] and the proposed bad point rate (BPR). All of the three metrics are computed based on the proportion of bad pixels. For Err and BPR, bad points are detected through edge matching between the texture and depth maps. Err and BPR do not require ground truth, and thus, they belong to no-reference metrics. For PBMP, the estimated depth map is compared with the ground truth to calculate the disparity errors. If the disparity error of a pixel is larger than a one-pixel-width, this pixel is regarded as a bad pixel. Because the ground truth is utilized, PBMP is an accurate and reliable full-reference metric. Moreover, as addressed in section III-F, we compute both the overall quality and the boundary quality<sup>2</sup> of the depth maps. The results are shown in Tables I and II, respectively. In both tables, the BPR and Err values increase with growing PBMPs. This trend demonstrates that both the proposed scheme and the method in [33] can detect bad points.

In Table I, we note the performance differences among the scenes. Because many bad pixels are not located along the depth edges in Teddy and Cones, which is analyzed in section IV-B, both the proposed scheme and [33] cannot detect these errors. At the same time, by utilizing the ground truth, PBMP can count every bad pixel. Therefore, large differences exist between PBMP and BPR/Err. In contrast, more errors are along the object boundaries in Tsukuba and Venus, which can be detected with the proposed scheme and [33]. As a result, the reported BPR/Err values are closer to PBMP in these two scenes. We also find that the values of Err are larger than those of BPR. This difference arises because the proposed scheme is more rigorous than [33]; therefore, the reported bad pixels are fewer in number but could be more accurate. For example, the redundancy removal step is included in the proposed scheme but not in [33]. Checking the values of Tsukuba and Venus

<sup>2</sup>In the implementation, the boundary region consists of occluded region and depth-discontinued region, both of which are defined in the database [39].



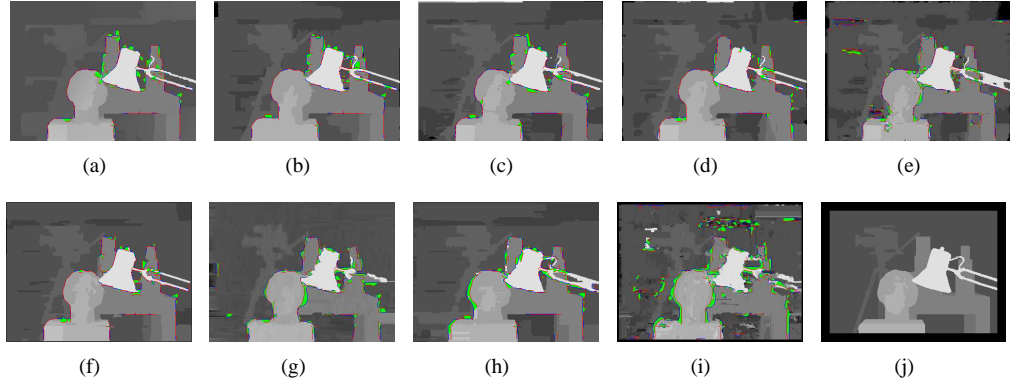


Fig. 8. Detected bad points in Tsukuba. (a) to (i) Depth maps estimated with [40] to [48], respectively. The detected bad pixels are marked in green. (j) Ground truth depth map.

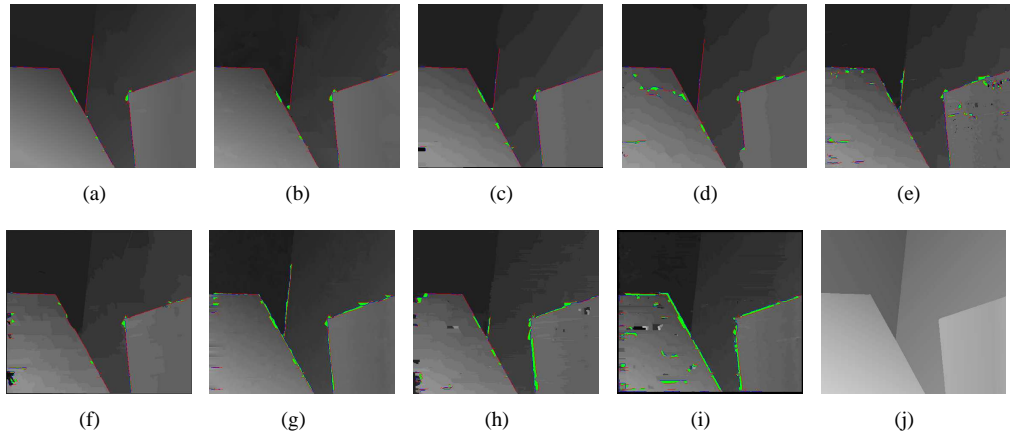


Fig. 9. Detected bad points in Venus. (a) to (i) Depth maps estimated with [40] to [48], respectively. The detected bad pixels are marked in green. (j) Ground truth depth map.

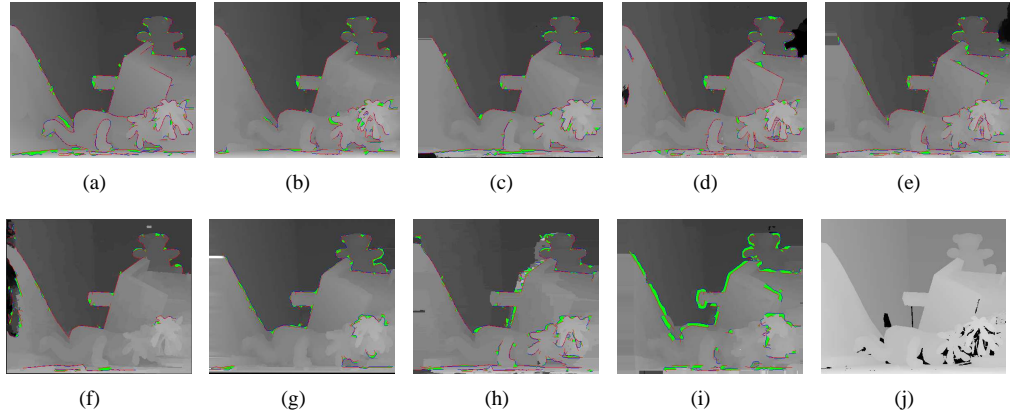


Fig. 10. Detected bad points in Teddy. (a) to (i) Depth maps estimated with [40] to [48], respectively. The detected bad pixels are marked in green. (j) Ground truth depth map.

in Table I, we note that the Err values are even larger than those of PBMP, in some cases. This finding indicates that the method in [33] classifies correct pixels as bad pixels. In contrast, the BPR values are smaller than BPMP values, which demonstrates that the proposed scheme detects bad pixels more accurately.

In Table II, the data are obtained only within boundary regions, which can better demonstrate the performances of the edge matching schemes. In the table, PBMP is larger than BPR/Err in most cases. The reason lies in the different approaches to the bad pixel detection. By referring to the ground truth, PBMP can detect all of the bad pixels. However,

BPR/Err can only report a portion of the bad pixels with edge matching because no reference is utilized. As mentioned earlier, PBMP is a reliable and accurate full-reference metric. Therefore, getting closer to PBMP indicates higher accuracy in detecting the bad pixels. In Table II, BPR gets closer to PBMP than Err does in most cases. This relationship demonstrates that the proposed scheme outperforms the method in [33].

With the data in the two tables, we further present the scatter plots between BPR/Err and PBMP, which are shown in Figs. 12 and 13, respectively. In these figures, the horizontal axis represents PBMP, and the two vertical axes denote the BPR and Err values, respectively. Linear regression is performed

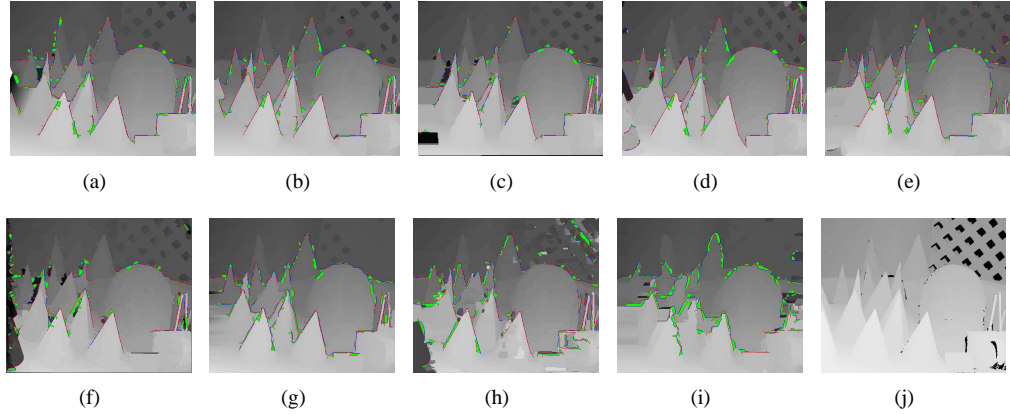


Fig. 11. Detected bad points in Cones. (a) to (i) Depth maps estimated with [40] to [48], respectively. The detected bad pixels are marked in green. (j) Ground truth depth map.

TABLE I  
PBMP, ERR AND BPR OF THE DEPTH MAPS (ALL)

	Tskuba			Venus			Teddy			Cones		
	PBMP (%)	Err (%)	BPR (%)	PBMP (%)	Err (%)	BPR (%)	PBMP (%)	Err (%)	BPR (%)	PBMP (%)	Err (%)	BPR (%)
AdaptBP [40]	1.37	5.37	1.47	0.21	1.09	0.18	7.06	4.62	1.52	7.92	2.91	1.53
WarpMat [41]	1.35	6.20	1.23	0.24	2.44	0.23	9.30	3.78	1.04	8.47	3.78	1.48
P-LinearS [42]	1.67	8.55	1.28	0.89	1.29	0.29	12.00	2.85	1.23	8.44	4.33	1.27
VSW [43]	1.88	8.58	1.08	0.81	5.07	0.52	13.3	2.67	0.96	8.85	4.20	1.50
BPcompressed [44]	3.63	9.97	2.06	1.89	3.43	0.86	13.9	6.60	1.07	9.85	6.95	1.62
Layered [45]	1.87	5.73	0.87	1.85	3.06	0.32	14.3	3.44	0.73	14.70	4.80	1.63
SNCC [46]	6.08	7.09	2.43	1.73	2.52	1.09	11.10	4.30	1.26	9.02	7.26	1.90
ReliabilityDP [47]	3.39	11.25	1.91	3.48	2.64	1.13	16.90	7.90	1.59	19.90	10.73	2.37
Infection [48]	9.54	23.05	4.95	5.53	5.11	2.16	25.10	10.74	3.95	21.30	7.98	3.1

TABLE II  
PBMP, ERR AND BPR OF THE DEPTH MAPS (BOUNDARY REGIONS)

	Tskuba			Venus			Teddy			Cones		
	PBMP (%)	Err (%)	BPR (%)	PBMP (%)	Err (%)	BPR (%)	PBMP (%)	Err (%)	BPR (%)	PBMP (%)	Err (%)	BPR (%)
AdaptBP [40]	19.24	10.00	16.28	12.22	1.72	11.98	22.43	3.71	12.08	15.29	3.63	8.6
WarpMat [41]	18.58	10.42	15.66	12.68	3.31	11.74	22.24	3.76	9.26	18.14	4.72	8.88
P-LinearS [42]	20.56	13.14	15.55	18.9	5.61	13.54	20.03	4.24	9.65	13.32	5.23	7.26
VSW [43]	18.50	11.56	14.01	16.93	5.25	12.27	21.66	3.18	10.38	18.26	4.95	8.41
BPcompressed [44]	26.93	13.78	19.52	23.11	7.23	16.35	23.91	7.73	9.43	19.36	7.35	9.13
Layered [45]	18.47	9.48	13.10	16.19	6.12	10.17	21.34	2.88	4.94	22.64	5.6	6.68
SNCC [46]	35.16	14.40	22.06	29.32	13.31	24.88	25.23	8.61	9.93	17.39	9.05	9.88
ReliabilityDP [47]	28.23	17.99	19.29	32.57	16.43	21.25	35.48	11.3	12.12	29.1	11.62	9.75
Infection [48]	45.54	29.93	29.09	45.2	29.03	29.95	51.05	19.72	17.5	45.38	11.8	13.02

between PBMP and BPR/Err, and the goodness-of-fit coefficient  $R^2$ , is adopted to measure the accuracy of the fitting. In Fig. 12,  $R^2_{BPR}$  exceeds 0.6 and reaches as high as 0.92. Similarly, the  $R^2_{BPR}$  values in Fig. 13 are larger than 0.63 and reach as high as 0.96. These results demonstrate that the proposed BPR is highly consistent with the full-reference metric, PBMP. Compared with Err, BPR has larger or similar  $R^2$  values in most cases. One exception is in Fig. 13(c), where  $R^2_{Err}$  is substantially larger than  $R^2_{BPR}$ . However, checking the data of Teddy in Table. II, we find that the PBMP values are large (more than 21), but the Err values are very small (approximately 3). These findings indicate that Err does not detect bad pixels accurately even though the  $R^2$  values are quite large.

Similarly, we also compute the Pearson coefficients between BPR/Err and PBMP in Table III. The results are similar to the  $R^2$  values in Figs. 12 and 13. The correlation exceeds 0.77 and reaches as high as 0.98, which demonstrates that the proposed scheme can detect depth distortions accurately.

#### D. Results of Synthesized Views

Usually, depth maps are not displayed directly to the viewers, but instead are used for subsequent applications such as depth-image-based-rendering (DIBR) and 3D reconstruction. Therefore, it is important to consider the relationship between the depth quality and the results of an application. Taking DIBR as an example, we investigate the quality of the depth and the synthesized views.

In the implementation, we render virtual views for the four scenes: Tsukuba, Venus, Teddy and Cones. In sections IV-B and IV-C, depth maps generated with the nine stereo matching algorithms [40]-[48] are utilized for depth quality assessment. Here, the same depth maps, together with the undistorted texture map, are used to render virtual views by performing DIBR. As a result, we obtain nine synthesized virtual views for each scene. The quality of these virtual views is evaluated with MSE and denoted as  $MSE_{syn}$ . To explore the relationship between the depth quality and the virtual view

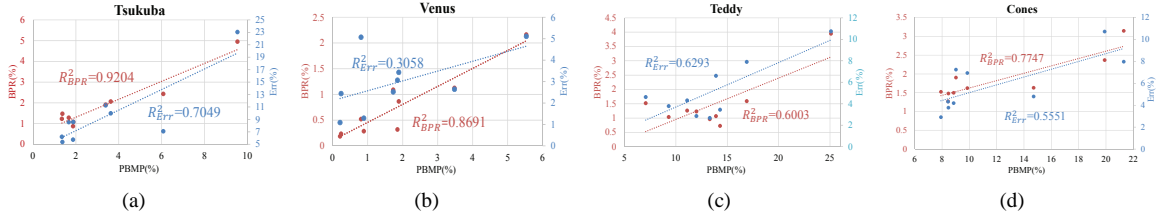


Fig. 12. Scatter plot of BPR, Err [33] Vs. PBMP [2] calculated with all pixels in the depth maps. (a) Results of Tsukuba,  $R^2_{BPR} = 0.9204$ ,  $R^2_{Err} = 0.7049$ . (b) Results of Venus,  $R^2_{BPR} = 0.8691$ ,  $R^2_{Err} = 0.3058$ . (c) Results of Teddy,  $R^2_{BPR} = 0.6293$ ,  $R^2_{Err} = 0.6003$ . (d) Results of Cones,  $R^2_{BPR} = 0.7747$ ,  $R^2_{Err} = 0.5551$ .

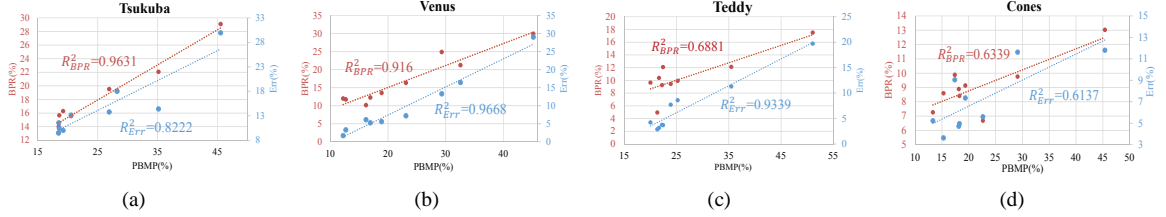


Fig. 13. Scatter plot of BPR, Err [33] Vs. PBMP [2] calculated with the boundary pixels. (a) Results of Tsukuba,  $R^2_{BPR} = 0.9631$ ,  $R^2_{Err} = 0.8222$ . (b) Results of Venus,  $R^2_{BPR} = 0.9616$ ,  $R^2_{Err} = 0.9668$ . (c) Results of Teddy,  $R^2_{BPR} = 0.6881$ ,  $R^2_{Err} = 0.9339$ . (d) Results of Cones,  $R^2_{BPR} = 0.6339$ ,  $R^2_{Err} = 0.6137$ .

quality, we investigate the  $BPR^3$  values of the tested depth maps and the corresponding  $MSE_{syn}$  values. In addition, we also reproduce the results of Zhao's depth no-synthesis-error model [11] and Fang's depth analytical model [15], which are used for comparison. Zhao *et al.* claimed that minor depth distortions can be ignored because they do not lead to errors in disparity or in the synthesized views. In fact, this idea is essentially quite similar to PBMP [2]. Therefore, by taking PBMP as the depth quality metric, the performance of Zhao's model [11] can be evaluated from the relationship between PBMP and  $MSE_{syn}$ . Fang's analytical model [15] predicts the quality of the synthesized views with full-reference texture and depth errors. With  $MSE_{pred}$  denoting the predicted error, the accuracy of the prediction model can be evaluated by comparing  $MSE_{pred}$  and  $MSE_{syn}$ . Therefore, we investigate the relationship between  $MSE_{syn}$  and PBMP/ $MSE_{pred}$ /BPR, and the scatter plots are shown in Figs. 14, 15 and 16, respectively. As mentioned before, every scene has nine virtual views. Correspondingly, in every scatter plot, there are nine markers.

Because view synthesis is performed based on depth maps, a depth map of lower quality leads to more errors in the synthesized view. This relationship indicates that  $MSE_{syn}$  should increase with growing PBMP/ $MSE_{pred}$ /BPRs. In the presented scatter plots, despite some outliers, the samples typically follow this rule. To quantify the accuracy of the three metrics, we fit  $MSE_{syn}$  and PBMP/ $MSE_{pred}$ /BPRs with linear regressions, and the coefficient  $R^2$  indicates the fitting accuracy.

In Fig. 14, Zhao's model is accurate, with  $R^2$  reaching 0.9. There are two reasons for this high consistency. First, PBMP is a full-reference metric obtained by referring to the ground truth, which makes it accurate. Second and more importantly, the metric itself is synthesized-view-quality oriented. More specifically, only pixels that cause disparity errors in the virtual views are classified as bad matching pixels and get counted in PBMP. Therefore, the definition of PBMP is highly related

to the quality of the synthesized views. In Fig. 15, the  $R^2$  values between  $MSE_{pred}$  and  $MSE_{syn}$  are larger than 0.58 and can reach as high as 0.9. This finding demonstrates that the analytical model [15] can accurately predict the quality of the virtual views. Moreover, in [15], the errors in the reference views are also reliably computed with the ground truth.

Finally, we investigate the relationship between  $MSE_{syn}$  and BPR. BPR indicates the proportion of the detected bad pixels in the depth maps. Therefore,  $MSE_{syn}$  is expected to increase with growing BPR. In Fig. 16, the samples match with our analysis. The fitting accuracy,  $R^2$ , between  $MSE_{syn}$  and BPR exceeds 0.6 for Tsukuba, Venus and Cones. More specifically,  $R^2$  reaches 0.83 for Tsukuba, which demonstrates that BPR can predict well the virtual view quality. We also note that the  $R^2$  of Teddy is only 0.32. The reason is that many bad pixels in Teddy are in the highly slanted plane or around the green leaves. As analyzed in section IV-B, the proposed scheme cannot detect such depth errors. Unfortunately, these undetected bad pixels still cause errors in the virtual views, which degrades the fitting accuracy greatly. In contrast, both PBMP and  $MSE_{pred}$  are obtained with ground truth, and thus, they can take the effect of every bad pixel into account. Therefore, PBMP and  $MSE_{pred}$  can correlate better with  $MSE_{syn}$  than BPR does.

In conclusion, the proposed BPR is highly correlated with the quality of the synthesized views. To a large degree, BPR can predict the quality of the virtual views. Although it does not perform as well as full-reference metrics in some cases, the performance of BPR can still be considered quite satisfactory because it does not utilize any reference at all.

We should also note that, the relationship between the depth quality and virtual view quality has been studied, and we take DIBR as only an application example. Nevertheless, in different applications, the standards and approaches in evaluating the depth quality also differ. For example, in DIBR, the depth errors do not cause virtual view errors in some cases [13]-[15][32]. However, in 3D reconstruction [49], any depth error will introduce distortions into the final results. From

<sup>3</sup>In section IV-D and IV-F, BPR means  $BPR_{all}$  calculated with (12).

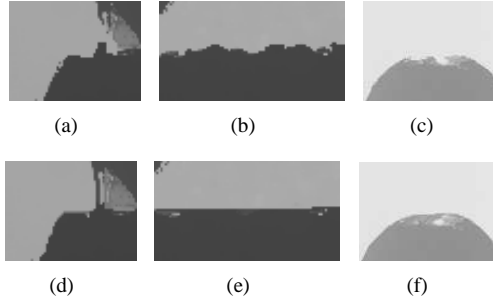


Fig. 18. Zoomed figures of the depth maps. (a)(d) Region A of doll. (b)(e) Region B of doll. (c)(f) Region A of girl. The upper and lower rows illustrate the depth maps before and after error correction, respectively.

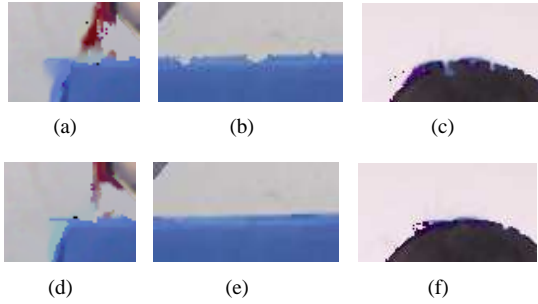


Fig. 19. Zoomed results of the virtual views. From (a) to (f), each figure is rendered with the corresponding depth map in Fig. 18.

this point of view, the proposed scheme can be regarded as an application-independent depth quality assessment approach because it focuses on the depth maps themselves.

#### E. Results of Depth Error Correction

By utilizing the proposed scheme to detect bad pixels, we can correct the depth errors and improve the depth quality. As introduced in section III-G, the total procedure consists of three steps: hole filling, error detection and error correction. In Fig. 17, the results of every step are presented. For a better illustration, several regions are enlarged and the details are shown in Fig. 18. It can be observed that, with the proposed error corrections, the depth distortions are reduced and depth quality is thereby improved. We further render virtual views with the original and the corrected depth maps. DIBR is applied, and the occluded regions in the virtual views are filled with background textures. We present some of the details in Fig. 19, which are rendered with the corresponding depth shown in Fig. 18. It can be observed that the modified depth maps improve the quality of the virtual views significantly, especially at object boundaries.

In addition to depth maps from Kinect, we have also implemented depth error correction on the Middlebury database. Due to limited space, we can only take Teddy as an example here, and the results are shown in Fig. 20. Specifically, we attach the ground truth depth edges to the original and the corrected depth maps with red lines. In the first row of Fig. 20, the estimated depth maps are distorted. For example, the foreground in regions A and B are expended severely. After error correction, the modified depth maps are shown in the second row of Fig. 20. It can be observed that the majority of distorted pixels have been modified, and the corrected depth maps are more similar to the ground truth.

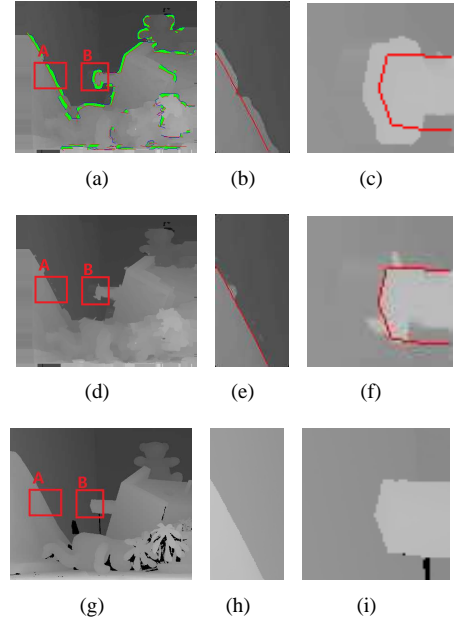


Fig. 20. Error correction results for Teddy, where the depth estimation algorithm is [48]. (a)-(c) Estimated depth maps. (d)-(f) Corrected depth maps. (g)-(i) Ground truth depth maps. The second and third columns show the details of region A and B in the first column. The red lines in (b)(c)(e)(f) are true depth edges extracted from the ground truth.

Please note that the results of the error correction are obtained by performing error detection and correction only once. In fact, if this procedure is implemented several times recursively, the depth quality can be further improved. For example, the remaining errors in Fig. 20(e) and 20(f) can be corrected.

#### F. Influence of the Parameters

In the proposed scheme, there are mainly four parameters, the search window  $W$ , the limits of the segment length  $L_{min}$  and  $L_{max}$ , and the similarity threshold  $T_s$ . These parameters affect the matching results, and thus, analyzing their influence will help us to determine reasonable parameters in assessing the depth quality.

1) *Influence of  $W$* : we set  $W$  to be in the range of 5 to 23, with an interval of 2, and the BPR values are plotted in Fig. 21. When  $W$  increases, more candidate edge segments are included for edge matching. As a result, an increasing number of misaligned texture-depth edge pairs are detected. Correspondingly, BPR becomes larger as  $W$  increases. Nevertheless, after a certain point, most of the valid matching pairs have been obtained, and then, enlarging  $W$  will not add new bad pixels to the results. In the figure, the curves have turning points, after which the BPR values become stable.

2) *Influence of  $L_{min}$* : we set  $L_{min}$  to be from 3 to 29 with an increasing step of 2, and the BPR values are plotted in Fig. 22. Edge segments shorter than  $L_{min}$  are excluded from the matching because they are believed to be noise or outliers and not real object boundaries. As  $L_{min}$  increases, fewer edge pairs can be matched. As a result, fewer bad pixels are reported, and BPR decreases. When the  $L_{min}$  values are small, BPR is stable. In contrast, when the  $L_{min}$  values are large, BPR decreases significantly with increasing  $L_{min}$ .



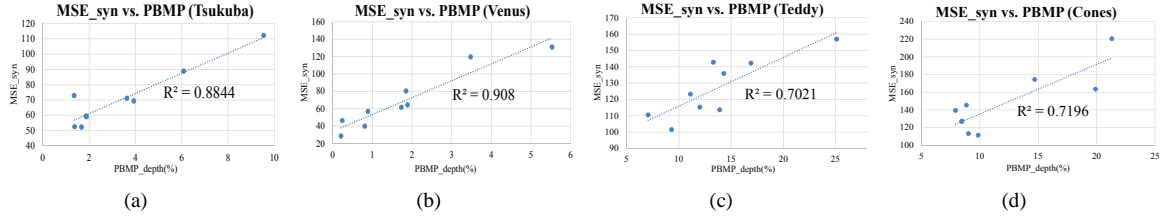


Fig. 14. Scatter plot of  $MSE_{syn}$ -PBMP [2]. The idea of Zhao's depth no-synthesis-error model [11] is essentially quite similar to PBMP, both of which are synthesized-view-quality oriented. (a) Scatter plot of Tsukuba,  $R^2=0.8844$ . (b) Scatter plot of Venus,  $R^2=0.908$ . (c) Scatter plot of Teddy,  $R^2=0.7021$ . (d) Scatter plot of Cones,  $R^2=0.7196$ .

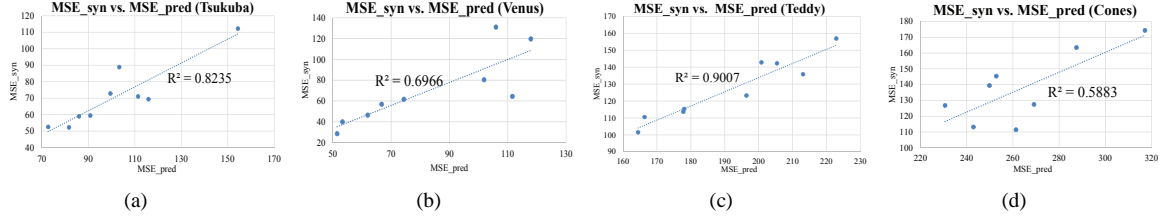


Fig. 15. Scatter plot of  $MSE_{syn}$ - $MSE_{pred}$  [15]. (a) Scatter plot of Tsukuba,  $R^2=0.8235$ . (b) Scatter plot of Venus,  $R^2=0.6966$ . (c) Scatter plot of Teddy,  $R^2=0.9007$ . (d) Scatter plot of Cones,  $R^2=0.5883$ .

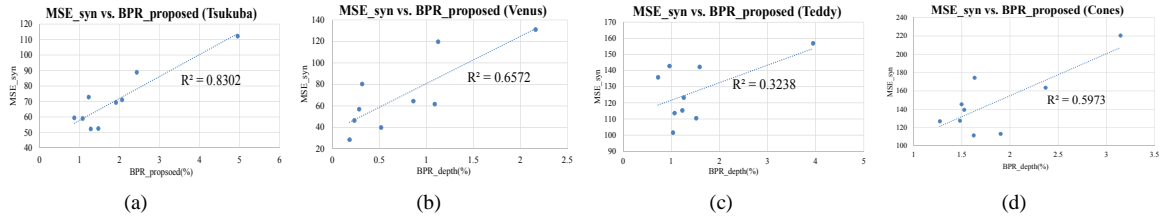


Fig. 16. Scatter plot of  $MSE_{syn}$ -BPR. (a) Scatter plot of Tsukuba,  $R^2=0.8302$ . (b) Scatter plot of Venus,  $R^2=0.6572$ . (c) Scatter plot of Teddy,  $R^2=0.3238$ . (d) Scatter plot of Cones,  $R^2=0.5973$ .

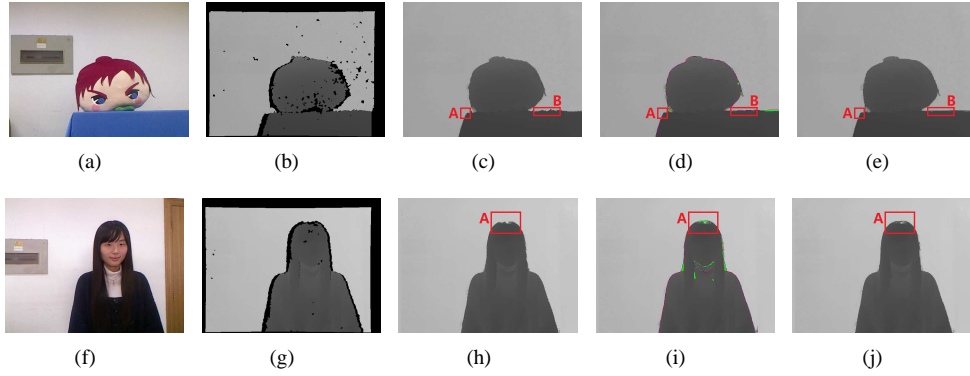


Fig. 17. Results of depth error correction with the proposed scheme. (a)(f) Texture maps. (b)(g) Depth maps captured with Kinect. (c)(h) Filled depth maps by using JBF. (d)(i) Detected misalignment errors and bad pixels. (e)(j) Depth maps after error correction.

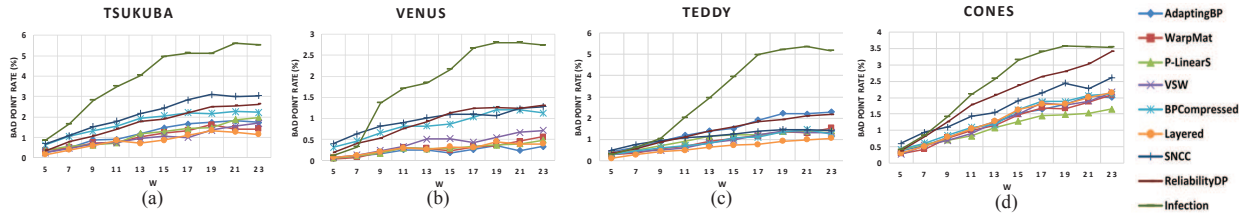


Fig. 21. Influence of search window. (a)(b)(c)(d) BPR- $W$  curves for Tsukuba, Venus, Teddy and Cones, respectively.

This relationship demonstrates that most edge segments are relatively long, and only a small number of edges are caused by noise and outliers. Therefore, a small  $L_{min}$  can preserve the valid matching pairs more completely, which will benefit the accuracy of the results. Moreover, Fig. 22 also shows that when  $L_{min}$  is small, the results are more robust and

stable. This pattern indicates that small changes in  $L_{min}$  do not influence the matching results greatly. In contrast, a large value of  $L_{min}$  will reject many reasonable matching edges, thus leading to poor performance of the proposed scheme. Therefore, a small value of  $L_{min}$  brings better matching results than large ones do.



3) *Influence of  $L_{max}$* : in the test,  $L_{max}$  ranges between 20 and 80 with a step of 4, and the BPR values are plotted in Fig. 23.  $L_{max}$  is the maximum allowable length for the edge segments. As  $L_{max}$  increases, some of the edge segments become longer, which increases the difficulty of edge matching. Therefore, fewer matching pairs will be detected and BPR decreases in Fig. 23. Similar to  $L_{min}$ , the curves in Fig. 23 also indicate that small  $L_{max}$  values have advantages over large values. On the one hand, smaller  $L_{max}$  values lead to more complete matching results. On the other hand, when the  $L_{max}$  values are small, the BPR values are stable, which demonstrates more robust performance.

In the figures, we note two special cases. The first one is that the BPR of Venus remains stable in Fig. 23(b). The reason is that Venus has very straight edges, and thus, edge segmentation does not affect the matching results greatly. The other case is that the BPR values of Tsukuba increase irregularly when  $L_{max}=76$ . This irregularity is caused by a long depth edge, which is located on the left of the plaster portrait. As  $L_{max}$  changes, the segments of this long edge, together with the detected bad pixels, are also different. When  $L_{max}$  is in the middle of the valid range, i.e., 60, this curved long edge is broken into several segments. Unfortunately, due to disturbances from invalid texture edges, some of the depth edge segments fail in the edge matching, and as a result, the BPR is small. When  $L_{max}$  is 76, this edge forms a single edge segment. As a whole unit, this edge successfully obtains the corresponding texture edges, together with the bad pixels. Therefore, the BPR increases. We should mention that this irregular increase of BPR in Tsukuba is an exception. First,  $L_{max}$  is only the upper bound of the edge segments, and most of the edge segments are shorter than that. Therefore,  $L_{max}$  cannot directly determine the results of edge segmentation or matching. Moreover, the matching results are also affected by many other factors, such as the distribution of texture and depth edges, and this irregular case occurs only under special circumstances. Among hundreds of edges in the four scenes, this case appears only once in Tsukuba.

4) *Influence of  $T_s$* : in the test,  $T_s$  varies from 0 to 0.9 with an interval of 0.1, and the corresponding BPR values are plotted in Fig. 24.  $T_s$  rejects unreasonable matched texture and depth edge pairs. As  $T_s$  increases, an increasing number of matched edge pairs are removed and BPR decreases. In fact, because the three similarity constraints have already excluded most of the unqualified edge pairs, a large value of  $T_s$  will reject too many correct edge pairs. As a result, many bad pixels will be left out by mistake, which results in poor performance of the proposed scheme. In contrast, a small value for  $T_s$  is sufficient to reject the fake edge pairs and preserve the reasonable results, which can benefit the accurate detection of edge misalignment errors.

Finally, we note that although BPR indicates the depth quality, its values do not directly demonstrate the performance of the proposed scheme. Therefore, the parameters cannot be determined by simply making BPR values large or small. Instead, a good set of parameters should bring robust matching results. Based on this idea and the conclusions of the parameter analysis mentioned above, we determine the parameters for

the proposed depth quality assessment, and the experimental results presented earlier demonstrate that these parameters achieve accurate and robust performances.

## V. CONCLUSION

Error-free depth maps are often unattainable, which makes conventional full-reference and reduced-reference IQA approaches inapplicable to depth quality assessment. In this paper, we presented a novel scheme, which is completely different from conventional IQA methods, for assessing the quality of depth maps in a no-reference fashion. This scheme, called two-step edge segment based matching, focuses on the misalignment errors of depth edges. To detect these errors, depth and texture edges were matched through spatial, orientation, and segment length similarities. The matching results can detect bad depth pixels, based on which a metric, the bad point rate, can be calculated to quantify the depth quality.

The proposed scheme is motivated by the idea that the quality of depth maps can be assessed from their intrinsic physical principles, instead of treating them simply as images. In this study, texture and depth are considered to be different forms of representations of the same scenes. Therefore, the relations between them can be properly utilized. Meeting the challenging constraint of reference-free assessment, the proposed scheme is proven to be effective. We hope that this new line of research, which is based on physical principles, will lead to more accurate depth quality assessment approaches for further applications.

Experimental results demonstrated that the proposed scheme can detect depth edge errors effectively. The proposed metric, BPR, is highly consistent with the full-reference metric, for which the Pearson correlation coefficient reaches 0.98 and the  $R^2$  coefficient reaches 0.96. Moreover, the proposed metric is also highly correlated with the quality of synthesized views obtained via DIBR. Furthermore, the proposed scheme only requires T+D maps, making it widely applicable. Because depth is a fundamental element in contemporary 3D video data, the proposed scheme has a great potential to facilitate many practical applications. For example, in depth-based 3DV systems, the proposed scheme can detect depth edge distortions and help with depth enhancement. With depth enhancement, the quality of the virtual views can be further improved and audiences can enjoy better viewing experiences.

## REFERENCES

- [1] 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," in *Proc. IEEE Int. Conf. on Multimedia and Expo*, Jul. 2006, pp. 2161-2164.
- [2] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, no. 1-3, pp. 7-42, Apr. 2002.
- [3] Z. Wang and A. Bovik, "Modern image quality assessment," *Synthesis Lectures on Image, Video, and Multimedia Processing*, vol. 2, no. 1, pp. 1-156, Jan. 2006.
- [4] C. Hewage, S. T. Worrall, S. Dogan, and A. M. Kondoz, "Prediction of stereoscopic video quality using objective quality models of 2-d video," *Electron. Lett.*, vol. 44, no. 16, pp. 963-965, Jul. 2008.
- [5] G. Saygili, C. G. Grler, and A. M. Tekalp, "Quality assessment of asymmetric stereo video coding," in *Proc. IEEE Int. Conf. on Image Process.*, Sep. 2010, pp. 4009-4012.

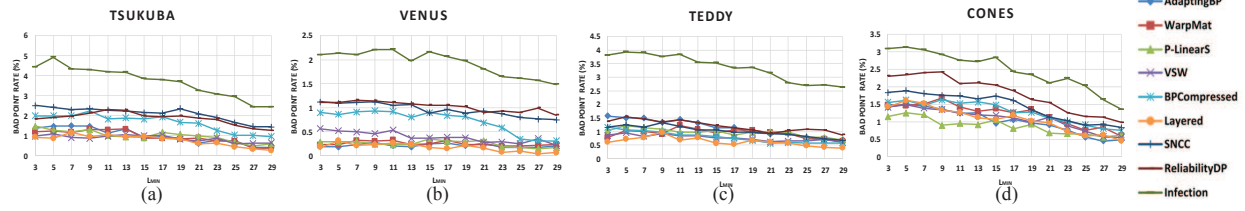


Fig. 22. Influence of  $L_{min}$ . (a)(b)(c)(d) BPR- $L_{min}$  curves for Tsukuba, Venus, Teddy and Cones, respectively.

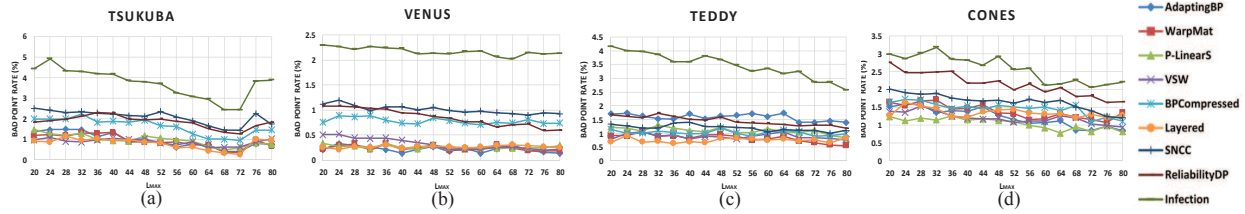


Fig. 23. Influence of  $L_{max}$ . (a)(b)(c)(d) BPR- $L_{max}$  curves for Tsukuba, Venus, Teddy and Cones, respectively.

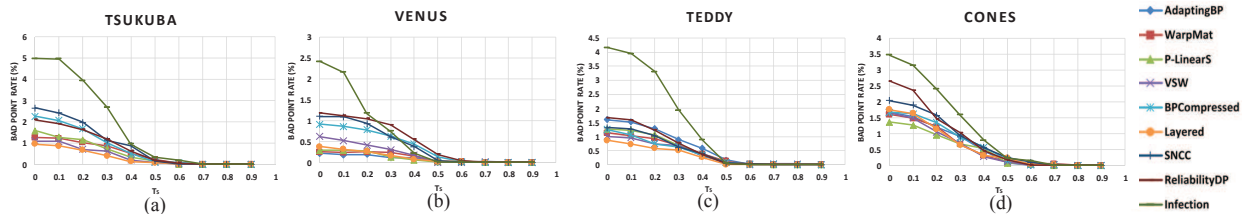


Fig. 24. Influence of  $T_s$ . (a)(b)(c)(d) BPR- $T_s$  curves for Tsukuba, Venus, Teddy and Cones, respectively.

- [6] F. Speranza, R. Renaud, A. Vincent, and W. J. Tam, "Perceived picture quality of frame-compatible 3d tv video formats," in *Proc. IEEE Int. Conf. on Multimedia and Expo*, Jul. 2012, pp. 640-645.
- [7] F. Qi, T. Jiang, X. Fan, S. Ma, and D. Zhao, "Stereoscopic video quality assessment based on stereo just-noticeable difference model," in *Proc. IEEE Int. Conf. on Image Process.*, Sep. 2013, pp. 34-38.
- [8] J. Gutierrez, P. Prez, F. Jaureguizar, J. Cabrera, and N. Garcia, "Subjective assessment of the impact of transmission errors in 3d tv compared to hdtv," in *Proc. 3DTV Conf.*, May 2011, pp. 1-4.
- [9] A. Tatematsu, Y. Ishibashi, N. Fukushima, and S. Sugawara, "Qoe assessment in tele-operation with 3d video and haptic media," in *Proc. IEEE Int. Conf. on Multimedia and Expo*, Jul. 2011, pp. 1-6.
- [10] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600-612, Apr. 2004.
- [11] Y. Zhao, C. Zhu, Z. Chen, and L. Yu, "Depth No-Synthesis-Error Model for View Synthesis in 3-D Video," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2221-2228, Aug. 2011.
- [12] W. S. Kim, A. Ortega, P. Lai, D. Tian, and C. Gomila, "Depth map coding with distortion estimation of rendered view," in *Proc. SPIE Vis. Inform. Process. and Commun.*, Jan. 2010, pp. 75430B-75430B.
- [13] H. Yuan, Y. Chang, J. Luo, F. Yang, and Z. Lu, "Model-based joint bit allocation between texture videos and depth maps for 3-D video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 4, pp. 485-497, Mar. 2011.
- [14] H. Yuan, J. Liu, H. Xu, Z. Li, and W. Liu, "Coding distortion elimination of virtual view synthesis for 3-d video system: Theoretical analyses and implementation," *IEEE Trans. Broadcast.*, vol. 58, no. 4, pp. 558-567, Dec. 2012.
- [15] L. Fang, N. Cheung, D. Tian, A. Vetro, H. Sun, and O. C. Au, "An Analytical Model for Synthesis Distortion Estimation in 3D Video," *IEEE Trans. Image Process.*, vol. 23, no. 1, pp. 185-199, Jan. 2014.
- [16] E. Bosc, R. Pepion, P. Le Callet, M. Pressigout, and L. Morin, "Reliability of 2d quality assessment methods for synthesized views evaluation in stereoscopic viewing conditions," in *Proc. 3DTV Conf.*, Oct. 2012, pp. 1-4.
- [17] P. Hanhart and T. Ebrahimi, "Quality assessment of a stereo pair formed from decoded and synthesized views using objective metrics," in *Proc. 3DTV Conf.*, Oct. 2012, pp. 1-4.
- [18] C. Hewage, S. T. Worrall, S. Dogan, S. Villette, and A. M. Kondoz, "Quality evaluation of color plus depth map-based stereoscopic video," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 304-318, Apr. 2009.
- [19] W. S. Malpica and A. C. Bovik, "Range image quality assessment by structural similarity," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Apr. 2009, pp. 1149-1152.
- [20] S. Ryu, D. Kim, and K. Sohn, "Stereoscopic image quality metric based on binocular perception model," in *Proc. IEEE Int. Conf. on Image Process.*, Sep. 2012, pp. 609-612.
- [21] Y. Lin and J. Wu, "Quality Assessment of Stereoscopic 3D Image Compression by Binocular Integration Behaviors," in *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1527-1542, Apr. 2014.
- [22] F. Lu, H. Wang, X. Ji, and G. Er, "Quality assessment of 3d asymmetric view coding using spatial frequency dominance model," in *Proc. 3DTV Conf.*, May 2009, pp. 1-4.
- [23] V. De Silva, H. Arachchi, E. Ekmekcioglu, and A. Kondoz, "Toward an impairment metric for stereoscopic video: A full-reference video quality metric to assess compressed stereoscopic video," in *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3392-3404, Sep. 2013.
- [24] F. Shao, W. Lin, S. Gu, G. Jiang, and T. Srikanthan, "Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics," in *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1940-1953, May 2013.
- [25] C. Hewage and M. G. Martini, "Reduced-reference quality evaluation for compressed depth maps associated with colour plus depth 3d video," in *Proc. IEEE Int. Conf. on Image Process.*, Sep. 2010, pp. 4017-4020.
- [26] G. Nur, and G. B. Akar, "An abstraction based reduced reference depth perception metric for 3D video," in *Proc. IEEE Int. Conf. on Image Process.*, Sep./Oct. 2012, pp. 625-628.
- [27] H. Malekmohamadi, A. Fernando and A. Kondoz, "A new reduced reference metric for color plus depth 3D video," *J. Vis. Commun. and Image Representation*, volume 25, no. 3, pp. 534-541, Apr. 2014.
- [28] C. Su, L. K. Cormack, and A. C. Bovik, "Oriented Correlation Models of Distorted Natural Images With Application to Natural Stereopair Quality Evaluation," in *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1685-1699, May 2013.
- [29] M. Chen, L. K. Cormack, and A. C. Bovik, "No-Reference Quality Assessment of Natural Stereopairs," in *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3379-3391, Sep. 2013.
- [30] J. Park, H. Oh, S. Lee and A. C. Bovik, "3D Visual Discomfort Predictor: Analysis of Disparity and Neural Activity Statistics," in *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1101-1114, Mar. 2015.
- [31] S. Milani, D. Ferrario, and S. Tubaro, "No-reference quality metric for

- depth maps,” in *Proc. IEEE Int. Conf. on Image Process.*, Sep. 2013, pp. 534-541.
- [32] H. Deng, L. Yu, B. Feng, and Q. Liu, “Structural similarity-based synthesized view distortion estimation for depth map coding,” *IEEE Trans. Consumer Electron.*, vol. 58, no. 4, pp. 1338-1344, Nov. 2012.
- [33] S. Xiang, J. Xue, L. Yu, and C. W. Chen “No-reference depth quality assessment for texture-plus-depth images,” in *Proc. IEEE Int. Conf. on Multimedia and Expo*, Jul. 2014, pp.1-6.
- [34] Y. Zhao, C. Zhu, Z. Chen, D. Tian, and L. Yu, “Boundary artifact reduction in view synthesis of 3d video: From perspective of texture-depth alignment,” *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 510-522, Jun. 2011.
- [35] X. Xu, L. Po, K. Ng, L. Feng, K. Cheung, C. Cheung, and C. Ting, “Depth map misalignment correction and dilation for DIBR view synthesis,” *Signal Process.: Image Commun.*, vol. 28, no. 9, pp. 1023-1045, Oct. 2013.
- [36] D. Miao, J. Fu, Y. Lu, S. Li, and C. W. Chen, “Texture-assisted kinect depth inpainting,” in *Proc. IEEE Int. Symp. on Circuits and Syst.*, May 2012, pp. 604-607.
- [37] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62-66, Jan. 1979.
- [38] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama, “Digital photography with flash and no-flash image pairs,” *ACM Trans. Graph.*, vol. 23, no. 3, pp. 664-672, Aug. 2004.
- [39] <http://vision.middlebury.edu/stereo/eval/>
- [40] A. Klaus, M. Sormann, and K. Karner, “Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure,” in *Proc. Int. Conf. on Pattern Recognition*, Aug. 2006, vol. 3, pp. 15-18.
- [41] M. Bleyer, M. Gelautz, C. Rother, and C. Rhemann, “A stereo approach that handles the matting problem via image warping,” in *Proc. IEEE Conf. on Comput. Vis. and Pattern Recognition*, Jun. 2009, pp. 501-508.
- [42] L. De-Maeztu, S. Mattoccia, A. Villanueva, and R. Cabeza, “Linear stereo matching,” in *Proc. IEEE Int. Conf. on Comput. Vis.*, Nov. 2011, pp. 1708-1715.
- [43] W. Hu, K. Zhang, L. Sun, J. Li, Y. Li, and S. Yang, “Virtual support window for adaptive-weight stereo matching,” in *Proc. IEEE Vis. Commun. and Image Process.*, Nov. 2011, pp. 1-4.
- [44] T. Yu, R. Lin, and B. Tang, “Efficient message representations for belief propagation,” in *Proc IEEE Int. Conf. on Comput. Vis.*, Oct. 2007, pp. 1-8.
- [45] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, “High-quality video view interpolation using a layered representation,” in *Proc. ACM SIGGRAPH*, Aug. 2004, pp. 600-608.
- [46] Nils Einecke and Julian Eggert, “A two-stage correlation method for stereoscopic depth estimation,” in *Proc. IEEE Int. Conf. on Digital Image Computing: Techniques and Applicat.*, Dec. 2010, pp. 227-234.
- [47] Q. Yang, C. Engels, and A. Akbarzadeh, “Near real-time stereo for weakly-textured scenes,” in *Proc of British Mach. Vis. Conf.*, Sep. 2008, pp. 1-10.
- [48] G. Olague, F. Fernández, C. Pérez, and E. Lutton, “The infection algorithm: An artificial epidemic approach for dense stereo correspondence,” *Artificial Life*, vol. 12, no. 4, pp. 593-615, Fall 2006.
- [49] S. Izadi, D. Kim, O. Hilliges, *et al.* “KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera,” in *Proc. ACM Symp. on User Interface Software and Tech.*, Oct. 2011, pp. 559-568.



**Sen Xiang** (S'11) received the B.S. degree in telecommunication engineering from Huazhong University of Science and Technology, Wuhan, in 2010. He is currently working toward the Ph.D. degree with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan.

His research interests include depth image processing, 3D video, structured light depth acquisition and related areas.



**Li Yu** (M'08) received the B.S., M.S., and Ph.D. degrees in electronic and information engineering from Huazhong University of Science and Technology, Wuhan, China, in 1992, 1995, and 1999, respectively. She is a Professor with the School of Electronic Information and Communications and is the Director of the Research Center of Broadband Wireless Communication and Multimedia Communication, Huazhong University of Science and Technology.

Her research interests include multimedia communication, wireless network, and signal processing in communications.



**Chang Wen Chen** (F'04) is a Professor of Computer Science and Engineering at the University at Buffalo, State University of New York. He has been Allen Henry Endow Chair Professor at the Florida Institute of Technology from July 2003 to December 2007. He was on the faculty of Electrical and Computer Engineering at the University of Rochester from 1992 to 1996 and on the faculty of Electrical and Computer Engineering at the University of Missouri-Columbia from 1996 to 2003.

He has been the Editor-in-Chief for IEEE Transactions on Multimedia since January 2014. He has also served as the Editor-in-Chief for IEEE Transaction on CSVT from 2006 to 2009. He has been an editor for several other major IEEE Transactions and Journals, including the proceedings of the IEEE, IEEE journal of selected areas in communications, and IEEE journal on emerging and selected topics in circuits and systems. He has served as conference chair for several major IEEE, ACM and SPIE conferences related to multimedia video communications and signal processing. His research is supported by NSF, DARPA, Air Force, NASA, Whitaker Foundation, Microsoft, Intel, Kodak, Huawei, and Technicolor.

He received his B.S. from University of Science and Technology of China in 1983, MSEE from University of Southern California in 1986, and Ph.D. from University of Illinois at Urbana-Champaign in 1992. He and his students have received eight best paper awards or best student paper awards over the past two decades. He has also received several research and professional achievement awards, including the Sigma Xi Excellence in Graduate Research Mentoring Award in 2003, Alexander von Humboldt Research Award in 2009, and the State University of New York at Buffalo Exceptional Scholar-Sustained Achievement Award in 2012. He is an IEEE Fellow and a SPIE Fellow.