SYNTHESIS QUALITY PREDICTION MODEL BASED ON DISTORTION INTOLERANCE

*Seungchul Ryu, Seungryong Kim, and *Kwanghoon Sohn*

Digital Image Media Lab (DIML)
Department of Electrical and Electronic Engineering, Yonsei University, Seoul, Republic of Korea
E-mail: {ryus01, srkim89, *khsohn}@yonsei.ac.kr

ABSTRACT

Free-viewpoint video system will provide viewers with freedom to navigate through the scene at different viewpoints. In the system, arbitrary viewpoints of videos are synthesized by the depth image-based rendering with multi-view plus depth videos. Despite the widespread of technologies for free-viewpoint video system, the field of quality assessment for the free-viewpoint video, especially the quality prediction of a synthesized image, has not yet been thoroughly investigated. This paper analyzes how distortions in color and depth images influence on the quality of a synthesized image. Then, an objective quality prediction model for a synthesized image is proposed based on the concept of intolerance of synthesis distortion. Experimental results show that the proposed model provides outstanding performance in predicting the quality of a synthesized image compared to other models.

Index Terms— synthesis quality prediction model, synthesis distortion intolerance, depth quality model.

1. INTRODUCTION

A promising feature of the future 3DTV is to virtually move through the scene providing a freedom to select a viewpoint. This feature, called free-viewpoint rendering, has become a popular topic in 3D video processing research field due to its wide applications such as free-viewpoint TV (FTV) [1], 3D reconstruction [2], and 3D medical imaging [3]. For providing free viewpoints, an unrealistic number of sequences at every viewpoint should be captured, coded, stored, and transmitted.

To address the impracticality of such an intuitive approach, arbitrary view synthesis using limited numbers of input videos has been considered an alternative approach [1]. Among several view synthesis techniques, depth image-based rendering (DIBR) [4] has attracted much attention due to its bandwidth-efficiency, cost-efficiency, and applicability. In DIBR, novel view videos are synthesized via 3D warping with multiview video plus depth (MVD) format which consists of limited numbers of color and depth videos.

Even though generating high-quality synthesized videos is the most important task for successful FTV system, few efforts have been devoted to studying the quality of synthesized image (QoS) assessment while active researches have been conducted for understanding 2D and 3D image quality [5, 6]. Accordingly, an objective model for measuring QoS is required to be thoroughly studied. In response to this demand, in recent years, several researches have been conducted to develop an objective model for measuring QoS. In [7, 8], the feasibility of 2D objective quality models was evaluated in assessing QoS. The results presented in [7, 8] show that the conventional methods are not sufficient to assess QoS. In [9], edge-based structural distortion indicator was proposed for measuring QoS. In [10], an objective model for QoS was proposed based on the texture-complexity, the diversity of gradient orientations, and the presence of high contrast. These models (referred here as *synthesis quality model*) can assess QoS at the synthesis side in FTV system, and consequently evaluate the performance of the view synthesis process.

However, assessing QoS at the synthesis side in a FTV system is impractical because such a synthesis quality model inevitably requires a heavy view synthesis process. A prediction model with color and depth images (referred here as *synthesis quality prediction model*) is an attractive alternative. A synthesis quality prediction model estimates QoS at arbitrary viewpoint using color and depth images without view synthesis process. While color image quality assessment is actively studied in the literature [11, 12, 13], depth image quality assessment has not been throughly investigated especially in assessing QoS.

The most widely used quality models for depth image are bad pixel percentage (BPP) and root mean square (RMS) presented in [14]. BPP measures the percentage of bad pixels higher than a specific threshold, and RMS measures squared errors between distorted and original depth images. Even though these models were widely used to assess the quality of a depth image, they did not consider image contents or view synthesis process. In [15], Malpical and Bovik proposed a quality model for a depth image based on the multiscale structural similarity. Solh *et al.* proposed full-reference [16] and no-reference [17] quality models based on temporal outliers, temporal inconsistencies, and spatial outliers. In [18], Hewage and Martini proposed a reduced-reference depth quality model based on extracted edge information. These models considered human visual system, but did not take the property of a view synthesis into consideration. In [19], a depth model predicting synthesis distortion was proposed and used for depth video coding. However, the model only a part of view synthesis process is considered, i.e. the effect of spatial complexity of color images on rendering distortions caused by depth distortion.

This paper analyzes the effect of distortions in color and depth images on a synthesized image. Then, we develop an objective synthesis quality prediction model focusing on depth distortions. The remainder of this paper is organized as follows. In Section 2, the proposed model is described with the analysis of synthesis distortions. Section 3 presents the experimental evaluations of the proposed model. Lastly, Section 4 concludes this paper with some discussions.

2. THE PROPOSED PREDICTION MODEL

2.1. Problem statement

In the view synthesis process, a novel view is generated through a geometric warping using a set of color images $\mathbf{C} = \{\mathbf{C_1}, \mathbf{C_2}, ..., \mathbf{C_n}\}$

(*): corresponding author

and the corresponding set of depth images $\mathbf{D} = \{\mathbf{D_1}, \mathbf{D_2}, ..., \mathbf{D_n}\}$, where $\mathbf{n}$ is the number of views used in view synthesis [4]. The QoS depends on the qualities of both color and depth images. To develop a synthesis quality prediction model, both effects of color and depth images on the QoS should be studied.

Distortions in depth images $\delta_\mathbf{D}$ lead geometrical errors in warping process and consequently result in synthesis position errors in a synthesized image $\mathbf{S}$. A distortion in color images $\delta_\mathbf{C}$ causes that a warped intensity itself is distorted. The problem of 3D image quality assessment for QoS can be formulated as predicting distortions in a synthesized image $\delta_\mathbf{S}$ using $\delta_\mathbf{D}$ and $\delta_\mathbf{C}$. The prediction model Φ is defined as follows:

$$\delta_\mathbf{S} = \Phi(\delta_\mathbf{D}, \delta_\mathbf{C}), \tag{1}$$

where $\delta_\mathbf{D} = \{\delta_{\mathbf{D1}}, ..., \delta_{\mathbf{Dn}}\} = \{\mathbf{D_1} - \bar{\mathbf{D}}_\mathbf{1}, ..., \mathbf{D_n} - \bar{\mathbf{D}}_\mathbf{n}\}$, $\delta_\mathbf{C} = \{\delta_{\mathbf{C1}}, ..., \delta_{\mathbf{Cn}}\} = \{\mathbf{C_1} - \bar{\mathbf{C}}_\mathbf{1}, ..., \mathbf{C_n} - \bar{\mathbf{C}}_\mathbf{n}\}$, and $\delta_\mathbf{S} = \mathbf{S} - \bar{\mathbf{S}}$. Here, $\bar{\mathbf{D}}$, $\bar{\mathbf{C}}$, and $\bar{\mathbf{S}}$ are distorted depth images, distorted color images, and distorted synthesized images, respectively.

A general DIBR algorithm [4, 20, 21] consists of three steps: i) warping, ii) blending, and iii) hole-filling as follows.
Step1, *Warping*: The color images of reference views are warped using the corresponding depth maps to the virtual view. In detail, each pixel is projected into the world coordinate, and then mapped into the virtual view point.
Step2, *Blending*: The two (or more) warped images are blended to ensure that foreground objects are visible.
Step3, *Hole-filling*: The holes induced by occlusion are filled by texture inpainting technique.

The view synthesis distortion $\delta_\mathbf{S}$ can be represented by a warping distortion δ_w, a blending distortionδ_b, and a hole-filling distortion δ_h as a function of

$$\delta_\mathbf{S} = \Psi(\delta_w, \delta_b, \delta_h), \tag{2}$$

In order to model the function $\Psi(\cdot)$ in (2), let $\mathbf{S}_v$ denote the image rendered by the original color images $\mathbf{C}$ and depth maps $\mathbf{D}$, and $\bar{\mathbf{S}}_v$ denote the image rendered by the distorted color images $\bar{\mathbf{C}}$ and depth maps $\bar{\mathbf{D}}$. Then, the view synthesis distortion $\delta_\mathbf{S}$ can be approximately decomposed into three distortion components as follows:

$$\begin{aligned}\delta_\mathbf{S} &= \mathrm{E}\left\{\mathbf{S}_v - \bar{\mathbf{S}}_v\right\} \\ &= \mathrm{E}\left\{(\mathbf{S}_w + \Delta\mathbf{S}_b + \Delta\mathbf{S}_h) - (\bar{\mathbf{S}}_w + \Delta\bar{\mathbf{S}}_b + \Delta\bar{\mathbf{S}}_h)\right\} \\ &= \mathrm{E}\left\{(\mathbf{S}_w - \bar{\mathbf{S}}_w) + (\Delta\mathbf{S}_b - \Delta\bar{\mathbf{S}}_b) + (\Delta\mathbf{S}_h - \Delta\bar{\mathbf{S}}_h)\right\} \\ &= \mathrm{E}\left\{\mathbf{S}_w - \bar{\mathbf{S}}_w\right\} + \mathrm{E}\left\{\Delta\mathbf{S}_b - \Delta\bar{\mathbf{S}}_b\right\} + \mathrm{E}\left\{\Delta\mathbf{S}_h - \Delta\bar{\mathbf{S}}_h\right\},\end{aligned} \tag{3}$$

where $\mathbf{S}_w$, $\mathbf{S}_b$, and $\mathbf{S}_h$ are the warped image, blended image, and hole-filled image using $\mathbf{C}$ and $\mathbf{D}$, respectively. $\bar{\mathbf{S}}_w$, $\bar{\mathbf{S}}_b$, and $\bar{\mathbf{S}}_h$ are warped image, blended image, and hole-filled image using $\bar{\mathbf{C}}$ and $\bar{\mathbf{D}}$. E is the averaging operator. The three distortion components in (3) $\mathrm{E}\left\{\mathbf{S}_w - \bar{\mathbf{S}}_w\right\}$, $\mathrm{E}\left\{\Delta\mathbf{S}_b - \Delta\bar{\mathbf{S}}_b\right\}$, and $\mathrm{E}\left\{\Delta\mathbf{S}_h - \Delta\bar{\mathbf{S}}_h\right\}$ are defined as δ_w, δ_b, and δ_h, respectively. In the following sections, we analyze the effect of color and depth distortions on each distortion components.

2.2. Analysis of warping distortion

This section analyzes the warping distortion δ_w induced by $\delta_\mathbf{C}$ and $\delta_\mathbf{D}$. As depicted in Fig. 1, a depth distortion of one pixel leads two warping distortions at (i) the original warping point $\mathbf{x}_v$ and (ii) the distorted warping point $\bar{\mathbf{x}}_v$. The warping distortions δ_w are approximated by summing the two warping distortions $\delta_{\mathbf{x}_v}$ at the original warping point $\mathbf{x}_v$ and $\delta_{\bar{\mathbf{x}}_v}$ at the distorted warping point $\bar{\mathbf{x}}_v$, and

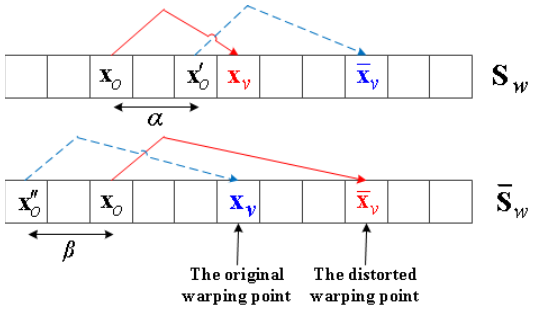

Fig. 1. Warping distortions induced by one distorted pixel

formulated as follows:

$$\begin{aligned}\delta_w &= \mathrm{E}\left\{\mathbf{S}_w - \bar{\mathbf{S}}_w\right\} \\ &\approx \mathrm{E}\left\{\delta_{\mathbf{x}_v} + \delta_{\bar{\mathbf{x}}_v}\right\} \\ &= \mathrm{E}\left\{|\mathbf{s_w}(\mathbf{x}_v) - \bar{\mathbf{s}}_w(\mathbf{x}_v)|\right\} + \mathrm{E}\left\{|\mathbf{s_w}(\bar{\mathbf{x}}_v) - \bar{\mathbf{s}}_w(\bar{\mathbf{x}}_v)|\right\}\end{aligned}, \tag{4}$$

where $\mathbf{s_w} \in \mathbf{S}_w$ is the warped pixel using $\mathbf{C}$ and $\mathbf{D}$ (referred as original warping), and $\bar{s}_w \in \bar{\mathbf{S}}_w$ is the warped pixel using $\bar{\mathbf{C}}$ and $\bar{\mathbf{D}}$ (referred as distorted warping).

Let $\mathbf{x}_o$ denote the pixel position warped to $\mathbf{x}_v$ during the original warping process, $\mathbf{x'}_o$ denote the pixel position warped to $\bar{\mathbf{x}}_v$ during the original warping process, and $\mathbf{x''}_o$ denote the pixel position warped to $\bar{\mathbf{x}}_v$ during the distorted warping process. Then, (4) is rewritten as follows:

$$\delta_w = \mathrm{E}\left\{\left|\mathbf{C}(\mathbf{x}_o) - \bar{\mathbf{C}}(\mathbf{x''}_o)\right|\right\} + \mathrm{E}\left\{\left|\mathbf{C}(\mathbf{x'}_o) - \bar{\mathbf{C}}(\mathbf{x}_o)\right|\right\}. \tag{5}$$

First, let us assume that there is no color distortion, i.e., $\mathbf{C} = \bar{\mathbf{C}}$, to analyze the effect of depth distortion $\delta_\mathbf{D}$ on the warping process. Then, (5) is modified as follows:

$$\delta_w = \mathrm{E}\left\{\left|\mathbf{C}(\mathbf{x}_o) - \mathbf{C}(\mathbf{x''}_o)\right|\right\} + \mathrm{E}\left\{\left|\mathbf{C}(\mathbf{x'}_o) - \mathbf{C}(\mathbf{x}_o)\right|\right\}. \tag{6}$$

The warping distance $\Delta\mathbf{x}$ between the original pixel $\mathbf{x}_o$ and the warped pixel $\mathbf{x}_v$ is generally formulated as (7) under the assumption that the cameras are rectified and linearly arranged [19].

$$\Delta\mathbf{x} = \mathbf{D}(\mathbf{x}) \cdot \frac{a \cdot \sigma_x}{255}\left(\frac{1}{Z_{far}} - \frac{1}{Z_{near}}\right), \tag{7}$$

where a is the first element of the intrinsic matrix, σ_x is the first element of the translation vector, Z_{far} and Z_{near} are the range of depth values. Since all the values except for $\mathbf{D}(\mathbf{x})$ are constant, (7) can be rewritten as $\Delta\mathbf{x} = c \cdot \mathbf{D}(\mathbf{x})$ where $c = \frac{a \cdot \sigma_x}{255}\left(\frac{1}{Z_{far}} - \frac{1}{Z_{near}}\right)$. Then, the difference α between $\mathbf{x}_o$ and $\mathbf{x'}_o$, and the difference β between $\mathbf{x}_o$ and $\mathbf{x''}_o$ are approximately computed as follows:

$$\begin{aligned}\alpha &= \mathbf{x'}_o - \mathbf{x}_o \\ &= (\bar{\mathbf{x}}_v - c \cdot \mathbf{D}(\mathbf{x'}_o)) - (\bar{\mathbf{x}}_v - c \cdot \bar{\mathbf{D}}(\mathbf{x}_o)) \\ &\approx (\bar{\mathbf{x}}_v - c \cdot (\mathbf{D}(\mathbf{x}_o) + d')) - (\bar{\mathbf{x}}_v - c \cdot \bar{\mathbf{D}}(\mathbf{x}_o)) \\ &= c \cdot (d - \delta_\mathbf{D}(\mathbf{x}_o))\end{aligned}, \tag{8}$$

$$\begin{aligned}\beta &= \mathbf{x}_o - \mathbf{x''}_o \\ &= (\mathbf{x}_v - c \cdot \mathbf{D}(\mathbf{x}_o)) - (\mathbf{x}_v - c \cdot \bar{\mathbf{D}}(\mathbf{x''}_o)) \\ &\approx (\mathbf{x}_v - c \cdot \mathbf{D}(\mathbf{x}_o)) - (\mathbf{x}_v - c \cdot (\bar{\mathbf{D}}(\mathbf{x}_o) + d'')) \\ &= c \cdot (\delta_\mathbf{D}(\mathbf{x}_o) + d'')\end{aligned}, \tag{9}$$

where $d' = \mathbf{D}(\mathbf{x'}_o) - \mathbf{D}(\mathbf{x}_o)$ and $d'' = \bar{\mathbf{D}}(\mathbf{x''}_o) - \bar{\mathbf{D}}(\mathbf{x}_o)$ are proportional to depth gradient at $\mathbf{x}_o$ (if $\mathbf{x'}_o$ and $\mathbf{x''}_o$ are near $\mathbf{x}_o$). Thus, both α and β are variables that depend depth distortion and

depth gradient.

From these observations, (6) can be modified as follows:

$$\delta_w = \sum_{\mathbf{x}_o \in \Omega_\mathbf{D}} |\mathbf{C}(\mathbf{x}_o) - \mathbf{C}(\mathbf{x}_o + \alpha)| + \sum_{\mathbf{x}_o \in \Omega_\mathbf{D}} |\mathbf{C}(\mathbf{x}_o) - \mathbf{C}(\mathbf{x}_o - \beta)| \quad (10)$$

where $\Omega_\mathbf{D}$ is a set of depth distorted pixels. The properties inferred from (10) are as follows: i) As referred above a depth distortion at one pixel $\delta_\mathbf{D}(\mathbf{x}_o)$ leads warping distortions at two pixels $\delta_w(\mathbf{x}_o+\alpha)$ and $\delta_w(\mathbf{x}_o - \beta)$, and the number of warping distortions is linearly proportional to the number of depth distortions, ii) The degree of warping distortions $|\delta_w|$ is not directly influenced by the degree of depth distortions $|\delta_\mathbf{D}|$, and rather influenced by spatial gradient of color images $\mathbf{C}$ around $\mathbf{x_o}$, iii) α and β increase when the degree of depth distortion $|\delta_\mathbf{D}|$ and spatial gradient of depth image $\mathbf{D}$ around $\mathbf{x_o}$ are large. The incremented α and β possibly increase, respectively, $|\mathbf{C}(\mathbf{x}_o) - \mathbf{C}(\mathbf{x}_o + \alpha)|$ and $|\mathbf{C}(\mathbf{x}_o) - \mathbf{C}(\mathbf{x}_o - \beta)|$, and consequently $|\delta_w|$.

Next, we analyze the effect of color distortion $\delta_\mathbf{C}$ on the warping process. Let us assume that there is no depth distortion, i.e., $\mathbf{x}_o = \mathbf{x'}_o = \mathbf{x''}_o$ in (5). Thus (5) is modified as follows:

$$\delta_w = 2\mathrm{E}\left\{\left|\mathbf{C}(\mathbf{x}_o) - \bar{\mathbf{C}}(\mathbf{x}_o)\right|\right\}. \quad (11)$$

This means that δ_w is directly influenced by the amount of color distortions.

2.3. Analysis of blending and holefilling distortions

This section analyzes the blending distortion δ_b and holefilling distortion δ_h induced by $\delta_\mathbf{C}$ and $\delta_\mathbf{D}$. The blended image $\mathbf{S}_b$ is commonly derived using the weighted averaging of left warped image $\mathbf{S}_{wL}$ and right warped image $\mathbf{S}_{wR}$ as follows:

$$\mathbf{S}_b = w_L\mathbf{S}_{wL} + w_R\mathbf{S}_{wR}, \quad (12)$$

where w_L and w_R are the weighting factors of the left and right warped images, respectively. Then, the blending distortion δ_b is computed as follows:

$$\begin{aligned}\delta_b &= \mathrm{E}\left\{\mathbf{S}_b - \bar{\mathbf{S}}_b\right\} \\ &= \mathrm{E}\left\{(w_L\mathbf{S}_{wL} + w_R\mathbf{S}_{wR}) - \left(w_L\bar{\mathbf{S}}_{wL} + w_R\bar{\mathbf{S}}_{wR}\right)\right\} \\ &= \mathrm{E}\left\{w_L\mathbf{S}_{wL} - w_L\bar{\mathbf{S}}_{wL}\right\} + \mathrm{E}\left\{w_R\mathbf{S}_{wR} - w_R\bar{\mathbf{S}}_{wR}\right\} \\ &= w_L\delta_{wL} + w_R\delta_{wR}\end{aligned}, \quad (13)$$

where $\bar{\mathbf{S}}_{wL}$ and $\bar{\mathbf{S}}_{wR}$ are respectively left and right warped images using $\bar{\mathbf{C}}$ and $\bar{\mathbf{D}}$. The blending distortion δ_b is the weighted average of the warping distortions δ_{wL} and δ_{wR}. That is, the blending distortion δ_b is mainly determined by the selection of weighing factors w_L and w_R. Note that the rendered view, i.e., a distance of rendered view from the reference view (referred as σ_x in Section 2.2), possibly can increase the constant value c in (8) and (9) and consequently $\bar{\mathbf{S}}_{wL}$ and $\bar{\mathbf{S}}_{wR}$. But, it is negligible compared to the other influences.

The hole-filling distortion δ_h is modeled as follows:

$$\delta_h = \mathrm{E}\left\{\mathbf{S}_h - \bar{\mathbf{S}}_h\right\}. \quad (14)$$

where $\mathbf{S}_h$ is obtained by the hole filling method Γ as follows:

$$\mathbf{S}_h = \Gamma\left(\mathbf{R}_h, \mathbf{S}_b \,|con_1, ..., con_n\right). \quad (15)$$

The hole-filling model Γ computes $\mathbf{S}_h$ with hole-region $\mathbf{R}_h = \{\mathbf{s_b} \in \mathbf{S}_b \,|\mathbf{s_b} \text{ is hole}\}$ and $\mathbf{S}_b$ under the constraints con_i such as depth, edge, and object constraints. Hole-filling distortions generally depend on the quality of blended image near $\mathbf{R}_h$ and the number of holes. The number of holes is related to the amount of depth gradient and warping and blending distortions near holes cause severe hole-filling distortions.

2.4. The proposed synthesis quality prediction model

The properties analyzed in Sections 2.2 and 2.3 are summarized as follows:
(P-i) The warping distortion δ_w is linearly proportional to the number of depth distortions.
(P-ii) The warping distortion $\delta_w(\mathbf{x})$ is influenced by gradient of color images near $\mathbf{x}$
(P-iii) The warping distortion $\delta_w(\mathbf{x})$ is possibly influenced by depth gradient and depth distortion near $\mathbf{x}$.
(P-iv) The warping distortion $\delta_w(\mathbf{x})$ is highly influenced by $\delta_\mathbf{C}(\mathbf{x})$.
(P-v) The blending distortion δ_b is mainly determined by view-weighting factors w in (12).
(P-vi) δ_w and δ_b near hole can cause severe hole-filling distortions δ_h, and the holes generally occur near pixels having large depth gradients.
Based on these properties, we design a synthesis quality (distortion) prediction model $Q_s$as follows:

$$Q_s = \sum_i w_i \cdot (Q_{\mathbf{Di}} + Q_{\mathbf{Ci}}), \quad (16)$$

where $Q_{\mathbf{Ci}}$ and $Q_{\mathbf{Di}}$ are synthesis distortions induced by $\delta_\mathbf{C}$ and $\delta_\mathbf{D}$ at viewpoint i, respectively. w_i is weighting factors for view i. The most widely used weighting factor selection method is based on the rendered position, in which the method weights on closer position among reference views. According to this, w_i is computed as $w_i = \frac{d_i}{\sum_i d_i}$ where d_i is the distance between the rendered position and the i^{th} reference view position. Since $Q_\mathbf{C}$ is directly influenced by $\delta_\mathbf{C}$, we can use color image quality model for estimating $Q_\mathbf{C}$, such as [22, 23, 24]. $Q_\mathbf{D}$ is estimated as follows:

$$Q_\mathbf{D} = \left\{\frac{1}{M}\sum_\mathbf{x}\left(\mathrm{T}(\mathbf{x})\cdot\tau(\mathbf{x})\right)^\rho\right\}^{1/\rho}, \quad (17)$$

where $\mathrm{T}(\mathbf{x}) = \begin{cases} 1 & \delta_\mathbf{D}(\mathbf{x}) > thd \\ 0 & \delta_\mathbf{D}(\mathbf{x}) \le thd \end{cases}$, M is the number of pixels, ρ is Minkowski parameter, $thd = 2$ is employed here, and $\tau(\mathbf{x})$ is *synthesis intolerance* which indicates how severe depth distortion at $\mathbf{x}$ is. A synthesis intolerance $\tau(\mathbf{x})$ is composed of three factors: $\tau(\mathbf{x}) = \lambda_{g_\mathbf{C}}\tau_{g_\mathbf{C}}(\mathbf{x}) + \lambda_{g_\mathbf{D}}\tau_{g_\mathbf{D}}(\mathbf{x}) + \lambda_{\delta_\mathbf{D}}\tau_{\delta_\mathbf{D}}(\mathbf{x})$ where $\lambda_{g_\mathbf{C}}$, $\lambda_{g_\mathbf{D}}$, and $\lambda_{\delta_\mathbf{D}}$ are controlling parameters for weights three factors (Here $\lambda_{g_\mathbf{C}} = 0.4$, $\lambda_{g_\mathbf{D}} = 0.4$, and $\lambda_{\delta_\mathbf{D}} = 0.2$ are used). $\tau_{g_\mathbf{C}}(\mathbf{x})$ is a texture-intolerance caused by spatial gradients of color images (P-ii) and computed using morphological gradient of $\mathbf{C}$ near $\mathbf{x}$ as follows:

$$\tau_{g_\mathbf{C}}(\mathbf{x}) = \max\Omega(\mathbf{x}) - \min\Omega(\mathbf{x}), \quad (18)$$

where $\Omega(\mathbf{x}) = \left\{\mathbf{C}(\mathbf{x}+\mathbf{\Delta})|\mathbf{\Delta} \in \mathbf{N^2}, |\mathbf{\Delta}| \le \mathbf{1}\right\}$. $\tau_{g_\mathbf{D}}(\mathbf{x})$ is a hole-intolerance[1] caused by spatial gradients of depth images (P-iii & P-vi) and computed using Canny edge detector [25] and dilation operator as follows:

$$\tau_{g_\mathbf{D}}(\mathbf{x}) = [\![e(\mathbf{x}),]\!] \quad (19)$$

[1] we named $\tau_{g_\mathbf{D}}(\mathbf{x})$ as 'hole-intolerance' because hole-filling distortions are more severe than warping distortions caused by depth gradient

Table 1. Performance comparisons of $Q_{\mathbf{D}}$

Model	PCC	SROCC
PSNR [11]	0.31	0.17
MS-SSIM [15]	0.24	0.08
RMS [14]	0.27	0.17
BPP [14]	0.33	0.62
AM [19]	0.67	0.75
Proposed method	**0.80**	**0.81**

where e($\mathbf{x}$) is Canny edge and $[\![\cdot]\!]$ is dilation operator. $\tau_{\delta_{\mathbf{D}}}(\mathbf{x})$ is a depth distortion-intolerance caused by depth distortions (P-iii) and computed as the average depth distortion of neighboring pixels as follows:

$$\tau_{\delta_{\mathbf{D}}}(\mathbf{x}) = K_N \cdot \sum_{\mathbf{x}\in\mathbf{R}_N} \delta_{\mathbf{D}}(\mathbf{x}), \tag{20}$$

where K_N is normalization factor, and $\mathbf{R}_N$ is a set of neighboring pixels.

An isolated distortion in a depth image results in an isolated warping position error. The isolated warping position error causes two possible distortions as described in Fig. 1, an isolated hole and a duplicated mapping. Generally, an isolated hole is well filled with hole-filling method [26] and duplicated mapping can be suppressed well by competition-rule in blending process. Thus, we considered the isolated distortion as a tolerable distortion, and eliminated the isolated distortion from (17) as follows:

$$Q_{\mathbf{D}} = \left\{ \frac{1}{M} \sum_{\mathbf{x}} (\Upsilon(\mathbf{x}) \cdot \mathrm{T}(\mathbf{x}) \cdot \tau(\mathbf{x}))^{\rho} \right\}^{1/\rho}, \tag{21}$$

where $\Upsilon(\mathbf{x})$ is 1 when the number of distorted depth in neighboring pixels $\mathbf{R}_i = \{\mathbf{x} + \Delta | \Delta \in N^2,\ |\Delta| \leq 5\}$ is larger than $thd2$ ($thd2 = 3$ here).

3. PERFORMANCE EVALUATIONS

In order to evaluate the performance of the prediction model Q_s described in Section 2.4, the correlation between scores of model Q_s and $\delta_{\mathbf{S}}$ is measured. A four parameter logistic function, as recommended in [27], is used for non-linear regression before calculating the performance measures. The used logistic function is as follows:

$$S_{pi} = \frac{\beta_1 - \beta_2}{e^{(S_i - \beta_3)/|\beta_4|} + 1} + \beta_2, \tag{22}$$

where β_1, β_2, β_3, and β_4 are model parameters, S_{pi} is the predicted score, and S_i is the score of the model. The values of β_1, β_2, β_3, and β_4 are first obtained by fitting to the corresponding e_s, and then the predicted score, S_{pi}, is calculated. The predicted score values are used in calculating the performance measures, Pearsons correlation coefficient (PCC, which indicates the prediction accuracy) and the Spearman rank-order correlation coefficient (SROCC, which indicates the prediction monotonicity). Note that for a well-defined model, the values of PCC and SROCC should be high.

To construct evaluation databases, MPEG 3DV test sequences [28]: Undo_Dancer, GT_Fly, Poznan_Hall2, Cafe, Lovebird, and Newspaper are used. 18 pairs of original depth and color images were randomly selected (three pairs from each sequence) and distorted using five distortions, Gaussian blurring, median filtering, Gaussian noise, scatter noise, and quantization, with five steps for each distortion. Totally, 450 test pairs were constructed excluding original images.

Table 2. Performance comparisons of Q_s

$Q_{\mathbf{C}}$	$Q_{\mathbf{D}}$	PCC	SROCC
MS-SSIM [22]	BPP [14]	0.75	0.68
	AM [19]	0.82	0.73
	Proposed method	**0.86**	**0.81**
VIF [23]	BPP [14]	0.78	0.69
	AM [19]	0.84	0.75
	Proposed method	**0.90**	**0.87**

In the first experiment, the performance of $Q_{\mathbf{D}}$ is evaluated using database-1 where intermediate view images are synthesized at three rendered positions using the original color images $\mathbf{C}$ and distorted depth images $\bar{\mathbf{D}}$. View Synthesis Reference Software (VSRS) 3.5[29] was used for view synthesis. The average mean squared errors (MSE)[2] of three synthesized images is defined as δ_s. For performance evaluation of the proposed depth quality model $Q_{\mathbf{D}}$, five models: Peak Signal to Noise Ratio (PSNR) [11], MultiScale-Structural SIMilarity (MS-SSIM) [15], Root Mean Squared error (RMS) [14], Bad Pixel Percentage (BPP) [14], and Autoregressive Model (AM) [19] were compared in terms of predicting QoS. The PSNR is the most widely used image quality metric. RMS and BPP are models presented in [14] for measuring the quality of depth image. MS-SSIM is used for measuring the quality of depth image with consideration of unknown regions in [15]. However, in general, the unknown regions do not exist in depth image. Also, in our database, unknown regions do not exist. Thus, the original MS-SSIM model is used for comparison. AM [19] measures depth image quality considering local complexity of $\mathbf{C}$.

As shown in Table 1, objective performance measures (in terms of both PCC and SROCC) show that the proposed depth quality model presents the best performance, while all the other models do not provide competitive performance. PSNR, RMS, and MS-SSIM do not consider any effects of depth distortions on synthesis distortions, while BPP considers only (P-i) into the model. AM considers (P-ii) and partially (P-i), and thus shows moderate performance as shown in Table 1. Nonetheless, the proposed depth quality model outperforms the other methods.

In the second experiment, the performance of Q_s is evaluated using database-2 where intermediate view images are synthesized at three rendered positions using distorted color images $\bar{\mathbf{C}}$ and distorted depth images $\bar{\mathbf{D}}$. Similar to the first experiment, average MSE of three synthesized images is used as δ_s. In this paper, MS-SSIM [22] and VIF [23] are employed as $Q_{\mathbf{C}}$. Table 2 presents the performance comparisons of the proposed Q_s. As shown in Table 2, the combination of VIF for $Q_{\mathbf{C}}$ and the proposed model for $Q_{\mathbf{D}}$ outperforms the other combinations. The results show that the proposed synthesis quality prediction model can address the lack of existing models and the performance is promising.

4. CONCLUSION

In this paper, a synthesis quality prediction model is proposed. The proposed model estimates QoS using color and depth images based on the synthesis intolerance without view synthesis process. Experimental analysis shows that the proposed model provides consistent performance in predicting QoS. The results also show that the proposed model outperforms the compared models. Future work includes developing a no-reference synthesis quality prediction model owing to its wide applicability.

[2]MSE is employed as an index for $\delta_{\mathbf{S}}$ to measure absolute synthesis distortions rather than perceptual distortions

5. REFERENCES

[1] M. Tanimoto, "FTV: Free-viewpoint television," *Signal Processing: Image Communication*, vol. 27, no. 6, pp. 555–570, Jul. 2012.

[2] H. Kim and K. Sohn, "3D reconstruction from stereo images for interaction between real and virtual objects," *Signal Processing: Image Communication*, vol. 20, no. 1, pp. 61–75, Jan. 2005.

[3] D. Ruijters and S. Zinger, "IGLANCE: transmission to medical high definition autostereoscopic displays," in *Proc. IEEE 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, Potsdam, Germany, May 2009.

[4] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," in *Proc. SPIE 5291*, San Jose, CA, USA, Jan. 2004.

[5] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*, Morgan & Claypool, USA, 2006.

[6] S. Ryu, D. Kim, and K. Sohn, "Stereoscopic image quality metric based on binocluar perception model," in *Proc. IEEE International Conference on Image Processing (ICIP)*, Orlando, FL, USA, Sep. 2012.

[7] E. Bosc, M. Koppel, R. Pepion, M. Pressigout, L. Morin, P. Ndjiki-Nya, and P. Le Callet, "Can 3D synthesized views be reliably assessed through usual subjective and objective evaluation protocols?," in *Proc. IEEE International Conference on Image Processing (ICIP)*, Brussels, Belgium, Sep. 2011.

[8] E. Bosc, R. Pepion, P. L. Callet, M. Koppel, P. Ndjiki-Nya, M. Pressigout, and L. Morin, "Towards a new quality metric for 3-D synthesized view assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 7, pp. 1332–1343, Nov. 2011.

[9] E. Bosc, P. L Callet, L. Morin, and M. Pressigout, "An edge-based structural distortion indicator for the quality assessment of 3d synthesized views," in *Proc. Picture Coding Symposium (PCS)*, Krakow, Poland, May 2012.

[10] P. H. Conze, P. Robert, and L. Morin, "Objective view synthesis quality assessment," in *Proc. SPIE 8288*, Burlingame, CA, USA, Jan. 2012.

[11] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*, Morgan & Claypool, USA, 2006.

[12] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 165–182, Jun. 2011.

[13] S. Ryu and K. Sohn, "No-reference quality assessment for stereoscopic images based on binocular quality perception," *IEEE Transactions on Circuits and Systems for Video Technology*, (in press).

[14] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 7–42, Apr. 2002.

[15] W. S. Malpica and A. C. Bovik, "Range image quality assessment by structural similarity," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, Apr. 2009.

[16] M. Solh, G. AlRegib, and J. M. Bauza, "3VQM: a vision-based quality measure for DIBR-based 3D videos," in *Proc. IEEE International Conference on Multimedia and Expo (ICME)*, Barcelona, Spain, Jul. 2011.

[17] M. Solh and G. AlRegib, "A no-reference quality measure for DIBR-based 3D videos," in *Proc. IEEE International Conference on Multimedia and Expo (ICME)*, Barcelona, Spain, Jul. 2011.

[18] C. T. E. R. Hewage and M. G. Martini, "Reduced-reference quality metric for 3D depth map transmission," in *Proc. IEEE 3DTV-Conference: The True Vision Capture, Transmission and Display of 3D video (3DTV-CON)*, Tampere, Finland, Jun. 2010.

[19] W. Kim, *3-D video coding system with enhanced rendered view quality*, Ph.D. thesis, University of Southern California, 2011.

[20] A. Smolic, K. Muller, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, "Intermediate view interpolation based on multiview video plus depth for advanced 3D video systems," in *Proc. IEEE International Conference on Image Processing*, San Diego, CA, USA, Oct. 2008.

[21] Y. Mori, N. Fukushima, T. Yendo, Y. Fujii, and M. Tanimoto, "View generation with 3D warping using depth information for FTV," *Signal Processing: Image Communication*, vol. 24, no. 1-2, pp. 65–72, Jan. 2009.

[22] E. P. Simoncelli Z. Wang and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, Nov. 2003.

[23] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[24] S. Ryu and K. Sohn, "No-reference sharpness metric based on inherent sharpness," *Electronics Letters*, vol. 47, no. 21, pp. 1178–1180, Oct. 2011.

[25] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, Nov. 1986.

[26] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, Sep. 2004.

[27] VQEG, "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment Phase II," Aug. 2003.

[28] ISO/IEC JTC1/SC29/WG11, "Call for proposals on 3D video coding technology," *ISO/IEC JTC1/SC29/WG11 Doc. N12036*, Mar. 2011.

[29] ISO/IEC JTC1/SC29/WG11, "Reference softwares for depth estimation and view synthesis," *ISO/IEC JTC1/SC29/WG11 Doc. M15377*, Apr. 2009.