Information Fusion 43 (2018) 47–56

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Evaluating virtual image quality using the side-views information fusion and depth maps

Muhammad Shahid Farid[*,a,b], Maurizio Lucenteforte[a], Marco Grangetto[a]

[a] Dipartimento di Informatica, Università degli Studi di Torino, Torino, Italy
[b] College of Information Technology, University of the Punjab, Lahore, Pakistan

ARTICLE INFO

Keywords:
3D image quality assessment
Information fusion
View synthesis
Depth image based rendering
Free-viewpoint TV

ABSTRACT

Three Dimensional (3D) image quality assessment is a challenging problem as compared to 2D images due to their different nature of acquisition, representation, coding, and display. The additional dimension of depth in multiview video plus depth (MVD) format is exploited to obtain images at novel intermediate viewpoints using depth image based rendering (DIBR) techniques, enabling 3D television and free-viewpoint television (FTV) applications. Depth maps introduce various quality artifacts in the DIBR-synthesized (virtual) images. In this paper, we propose a novel methodology to evaluate the quality of synthesized views in absence of the corresponding original reference views. It computes the statistical characteristics of the side views from whom the virtual view is generated, and fuses this information to estimate the statistical characteristics of the cyclopean image which are compared to those of the synthesized image to evaluate its quality. In addition to texture images, the proposed algorithm also considers the depth maps in evaluating the quality of the synthesized images. The algorithm blends two quality metrics, one estimating the texture distortion in the synthesized texture image induced by compression, transmission, 3D warping, or other causes and the second one determining the distortion of the depth maps. The two metrics are combined to obtain an overall quality assessment of the synthesized image. The proposed Synthesized Image Quality Metric (SIQM) is tested on the challenging MCL-3D and SIAT-3D datasets. The evaluation results show that the proposed metric significantly improves over state-of-the-art 3D image quality assessment algorithms.

1. Introduction

Image quality assessment (IQA) has been a major research area since the last three decades due to its vast applicability in modern broadcasting systems and due to the increasing demand of high picture quality by the end user. In the recent years, the advancements in 3D television, cinema and other advanced displays, e.g., free-viewpoint television (FTV) [1] and super multiview (SMV) displays [2], have posed new challenges in terms of quality assessment of 3D pictures. 3D image and video quality assessment is a more difficult and complex problem compared to its 2D counterpart. Due to different nature of acquisition, representation, transmission, and rendering of 3D images, they suffer from different types of quality artifacts [3–6]. Moreover, the additional dimension of depth maps in 3D content also introduces various quality artifacts. Research studies [7–9] suggest that in addition to texture image quality, depth map quality must also be incorporated in the evaluation process to assess the true quality of 3D images.

The latest autostereoscopic and multiview autostereoscopic displays, e.g., FTV, SMV allow the user to enjoy the 3D scene by interactively controlling the viewpoint. Such technologies require a huge number of views to provide a smooth motion parallax. However, capturing, coding, and transmitting such a large number of views is not practical due to various cost, hardware, and bandwidth constraints. Therefore, limited camera views are captured and transmitted and the additional intermediate viewpoints are generated with depth image based rendering (DIBR) techniques [10]. Given a DIBR algorithm, the perceptual quality of the rendered views depends on both the texture image quality and the depth map quality [7,8]. The quality of the synthesized views is important as in a multiview autostereoscopic environment, most of the views presented to the viewer are virtually generated. The quality of these views, thus, has significant impact on the overall user experience. Moreover, being able to predict the quality of virtual views can be exploited also in compression algorithms to drive the rate distortion optimization stage. As an example, the VSO coding tool of 3D-HEVC [11] exploits such prediction to guarantee the quality of the virtual views.

* Corresponding author.
E-mail addresses: shahid@pucit.edu.pk, farid@di.unito.it (M.S. Farid), lucente@di.unito.it (M. Lucenteforte), grangetto@di.unito.it (M. Grangetto).
URLS: http://www.di.unito.it/~farid/ (M.S. Farid), http://www.di.unito.it/~mgrange/ (M. Grangetto).

https://doi.org/10.1016/j.inffus.2017.11.007
Received 25 February 2017; Received in revised form 5 November 2017; Accepted 12 November 2017
Available online 17 November 2017
1566-2535/

The quality of a synthesized image is affected by a number of 2D and 3D artifacts that depend on many factors, such as:

- depth maps are noisy and imperfect since these are generally estimated through stereo-matching algorithms [12], and may cause structural and textural distortion in the synthesized images [13–15]. Moreover, compression of texture images and depth maps introduces additional artifacts [3,16];
- large discontinuities in the depth maps introduce holes in the synthesized images. Holes also appear as a consequence of re-sampling when warped pixel locations need to be mapped to integer coordinates. These holes are filled through inpainting approaches which may introduce blur in the final picture;
- other 3D rendering and viewing artifacts, e.g., binocular rivalry, visual discomfort and ghosting or crosstalk [17–20] have a significant impact on the perceived quality.

These issues make the quality assessment (QA) of 3D synthesized images more complex compared to 2D-IQA. Moreover, in absence of the original reference images, the QA of virtual images becomes more challenging problem. As already recalled, for virtual viewpoints the corresponding ground truth video is generally missing. In Fig. 1, a typical scenario for the quality assessment of a virtual view is shown. An intermediate virtual view V_k', whose corresponding reference view V_k is not available, is generated from the distorted left and right views V_i and V_j through DIBR. In absence of the reference kth view, common full-reference image quality metrics cannot be employed: an algorithm capable to predict the quality of the synthesized image exploiting only the available side views is needed to evaluate the quality of the synthesized view V_k'.

2. Related work

Various stereoscopic image quality assessment algorithms have been proposed in literature [21–26], however the quality assessment of DIBR-synthesized images is relatively less investigated. To quantify the structural distortion in synthesized view due to DIBR, Bosc et al. [27] compared the edges of the original and the warped images. However, this metric is limited to structural distortion estimation and cannot be used to represent the overall quality of the virtual image as it does not compute the color related artifacts. CSED (Color and Sharpness of Edge Distortion) [28] is another full reference quality metric that targets the hole regions to assess the color distortion and uses the edge sharpness of the reference and virtual images to assess structural distortion. The algorithm in [29] compares the regions of high spatial frequency of the stereopair images to estimate the contrast and luminance changes.

Kim et al. [30] proposed to apply a weighting map to the conventional 2D quality metrics such as SSIM [31] and peak signal to noise ratio (PSNR). The weighting map is computed by combining the depth maps and the motion information estimated from the texture images. You et al. [32] proposed to use the disparity information to improve the performance of 2D quality metrics on 3D data. The 3D-IQA algorithm in [33] estimates the structural distortion in the synthesized image using Hausdorff distance and combines it with SSIM score. 3DSwIM [34] metric detects the human skin regions in the virtual image and finds the corresponding regions in the reference images; the two corresponding regions are compared to determine the quality of the virtual picture.

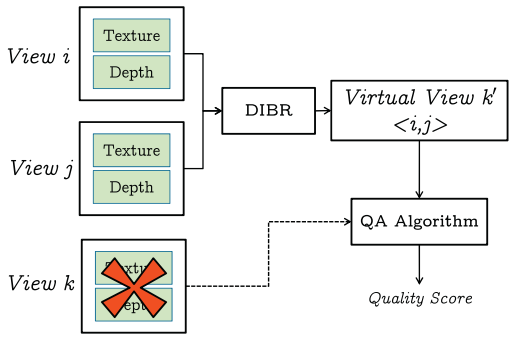

Fig. 1. A typical DIBR-Synthesized view quality assessment scenario: the quality of intermediate virtual view V_k' is subject to evaluation whose corresponding reference view V_k is not available.

De Silva et al. [25] proposed a learning based Stereoscopic Structural Distortion (StSD) metric to evaluate the quality of 3D videos. The quality assessment algorithm proposed in [35] estimates the quality of the synthesized image by measuring how classes of image contours change due to synthesis process. The no-reference synthesized image quality metric proposed in [36] exploits simple morphological operators to predict the quality of the virtual image. They use the opening and closing morphological operations to remove the synthesis distortions in the virtual image. This filter image is then compared with the synthesized image to estimate its quality. The 3D quality estimator proposed in [37] combines SSIM and C4 [38] and also exploits the disparity to estimate the quality. View Synthesis Quality Assessment (VSQA) [39] combines SSIM with three weighting functions derived from contrast, orientation and texture maps of the reference and synthesized views to assess the quality of virtual pictures. A good literature on 3D-IQA and various 3D quality artifacts can be found in [40,41].

Most existing quality assessment algorithms for DIBR synthesized images are full-reference and rely on the conventional 2D-IQA algorithms. As described earlier, in modern 3DTV and FTV applications few sparse viewpoints are captured and DIBR is exploited to obtain a large number of intermediate views to support smooth horizontal parallax. The existing 3D-IQA metrics cannot be used to evaluate the quality of these novel views as the corresponding ground truth videos are not available. In this paper, we propose a new methodology to assess the quality of virtual views in absence of the corresponding references. We propose to use the original texture and depth images from whom the virtual image is generated as references to estimate the quality of the virtual image. This concept is novel and has not been explored in the past for 3D image quality assessment. Based on this concept, we present a novel 3D-IQA algorithm to estimate the quality of the synthesized images in absence of corresponding reference images. The major contributions of this paper are as follows:

- The overall contribution of this paper is the proposal of a novel 3D-IQA algorithm to evaluate the quality of virtual images obtained with depth image based rendering (DIBR) techniques. The proposed metric considers not only the texture images but also the corresponding depth maps in assessing the quality of the synthesized image. The two quality scores are combined to obtain the overall synthesized image quality.
- A novel Texture Distortion Metric (TDM) is proposed to assess the quality of the synthesized texture image exploiting information fusion theory. The novelty of TDM lies in the use of cyclopean eye theory and divisive normalization transform (DNT) in context of DIBR synthesized image quality assessment. In particular, the histograms of the DN transformed texture input images are fused together to estimate the statistical characteristics of the mental (cyclopean) image. The histogram of the DNT of synthesized image is also computed and compared with that of the mental image to estimate the quality of the synthesized texture image. Such approach to texture quality evaluation has been previously introduced in [42], where it has been analyzed in presence of HEVC compression only using objective metrics.
- In addition to texture images, depth maps are also considered in quality evaluation; indeed, distortion in depth maps significantly affects the sharp boundaries in the depth map which in turn causes structural distortion in the synthesized image. Based on this