J. Vis. Commun. Image R. 23 (2012) 42–52

Contents lists available at SciVerse ScienceDirect

J. Vis. Commun. Image R.

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

Multiview depth coding based on combined color/depth segmentation

J. Ruiz-Hidalgo [a,*], J.R. Morros [a], P. Aflaki [b], F. Calderero [c], F. Marqués [a]

[a] Department of Signal Theory and Communications, Universitat Politècnica de Catalunya, Jordi Girona 1-3, 08034 Barcelona, Spain
[b] Tampere University of Technology, Tampere, Finland
[c] Universitat Pompeu Fabra, Tànger 122, 08018 Barcelona, Spain

ARTICLE INFO

Article history:
Received 15 September 2010
Accepted 8 August 2011
Available online 18 August 2011

Keywords:
Depth map
Virtual view
Multiview video coding
3DTV
Segmentation-based video coding
Region-merging
Color-depth segmentation
Texture coding

ABSTRACT

In this paper, a new coding method for multiview depth video is presented. Considering the smooth structure and sharp edges of depth maps, a segmentation based approach is proposed. This allows further preserving the depth contours thus introducing fewer artifacts in the depth perception of the video. To reduce the cost associated with partition coding, an approximation of the depth partition is built using the decoded color view segmentation. This approximation is refined by sending some complementary information about the relevant differences between color and depth partitions. For coding the depth content of each region, a decomposition into orthogonal basis is used in this paper although similar decompositions may be also employed. Experimental results show that the proposed segmentation based depth coding method outperforms H.264/AVC and H.264/MVC by more than 2 dB at similar bitrates.

1. Introduction

It is believed that 3D is the next major revolution in the history of TV. Both at professional and consumer electronics exhibitions, companies are eager to show their new 3D applications. In these applications, the 3D effect is obtained by recording the scene from multiple viewpoints, which is referred as multi-view video (MVV) [1]. MVV information is often completed with sample depth information (depth maps) for each view, thus conforming the multi-view video + depth (MVD) format (Fig. 1). The introduction of this MVD format allows the incorporation of new functionalities such as free view point video [2] in current 3DTV applications. MVD representations result in a vast amount of data to be stored or transmitted. In most situations, there is a need to compress these data efficiently without sacrificing visual quality significantly. Previous works presented various solutions for MVV coding (MVC) [3], mostly based on H.264 with combined temporal and inter-view prediction. Also, approaches for depth map coding in MVD have been proposed [4–6].

In 3D applications, depth maps are used for rendering new images (virtual color views) but not to be directly viewed by the end user. Thus, the goal when coding depth maps is to maximize the perceived visual quality of the rendered virtual color views instead of the visual quality of the depth maps themselves [4]. Traditional image compression methods have been designed to provide maximum perceived visual quality, so using these methods to compress depth maps may result in suboptimal performance. The rendering error depends on the quality of the depth map coding and the quality of the coding of the original color view used as a reference, but also on the structure of the depth map. For example: errors in the depth map close to a sharp edge can result in severe rendering artifacts, while errors on a smooth area may have negligible influence on the final quality.

These ideas are exploited, for instance, in [4], where the authors segment the depth map using a quadtree decomposition and then piecewise-linear functions are used to represent the depth information in each resulting block. Even though their experimental results show coding gains comparable to AVC with no temporal prediction (using only I frames) and MVC, costly partition information for the quadtree must be sent which results in a reduction of the performance. Another approach exploiting the previous ideas is presented in [6]. It relies on a shape-adaptive discrete wavelet transform (SA-DWT). The contours of the depth maps are extracted using a Canny edge detector and encoded using a differential Freeman chain-code. SA-DWT greatly reduces the data entropy of the resulting coefficients and the experimental results show significant coding gains with respect to traditional techniques based on DWT. However, a formal comparison against the AVC standard is not provided.

Another region-based representation is presented in [7], where texture, depth map and shape masks are encoded. Binary shape masks are coded by means of a bitmap based method using Context Arithmetic Encoding (CAE) and the depth map is encoded

* Corresponding author. Fax: +34 9 3401 6447.
E-mail addresses: j.ruiz@upc.edu (J. Ruiz-Hidalgo), ramon.morros@upc.edu (J.R. Morros), payman.aflaki@tut.fi (P. Aflaki), felipe.calderero@upf.edu (F. Calderero), ferran.marques@upc.edu (F. Marqués).

1047-3203/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jvcir.2011.08.001

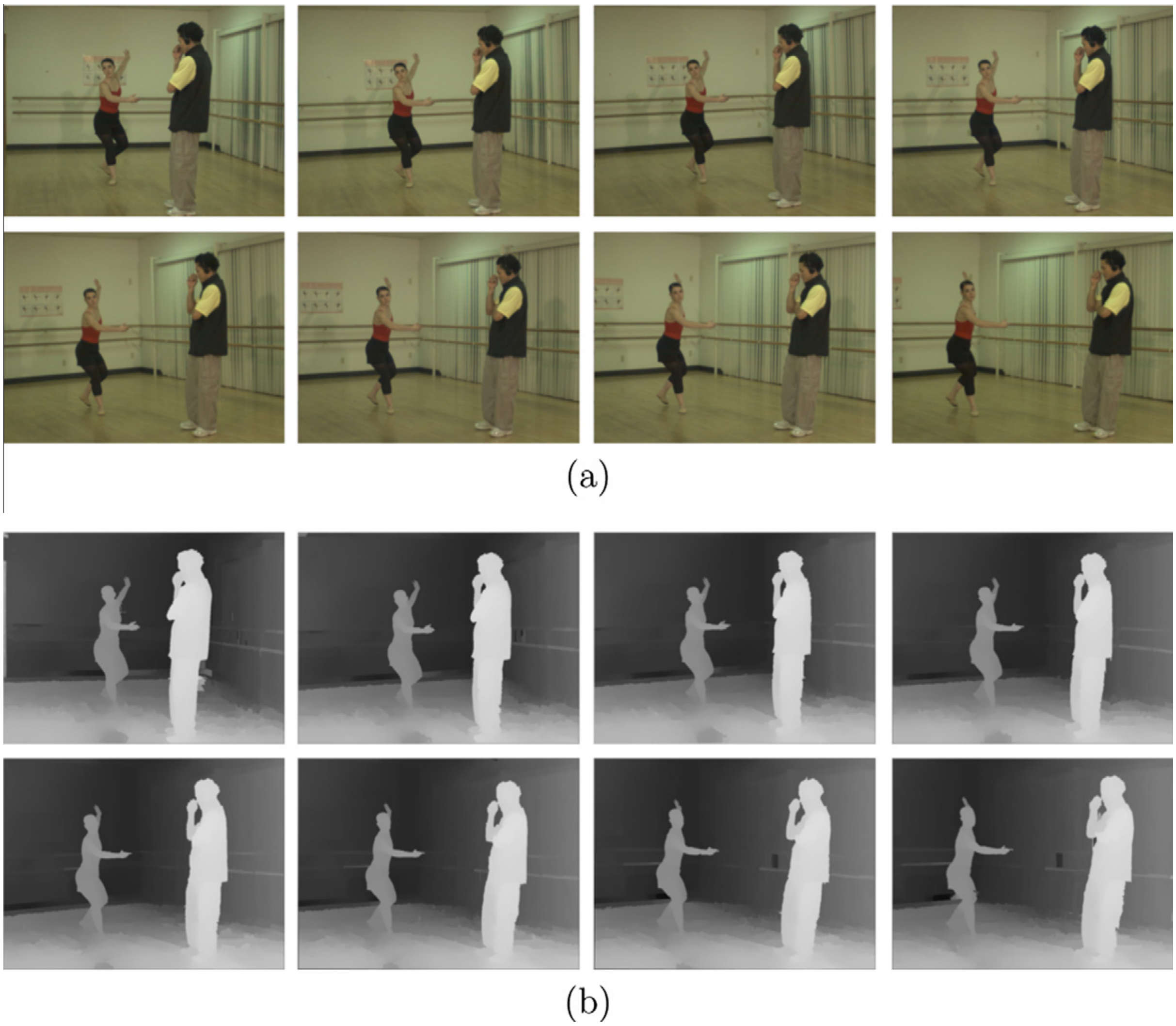

Fig. 1. Multi-view video + depth (MVD) ballet sequence composed of 8 camera views. (a) Color view for all cameras at frame $t = 0$. (b) Corresponding depth information at the same time instant.

block-by-block using planar models functions. Due to the high cost required to transmit the shape mask, it is difficult to achieve efficient coding.

In [5], the authors propose a combined representation, called layered depth image (LDI), for color views and depth in MVD sequences. As the LDI representation aims to reduce the redundant information between camera views, it is very useful for coding purposes. However, the number of layers for each element in the representation must be encoded lossless or near-lossless which limits its coding efficiency. Also, the LDI representation is more suited to scenarios with cameras close to each other.

A temporal prediction scheme is presented in [8] that exploits the similarity between texture and depth motion vectors to reuse part of the texture vectors, adapting them to fit into the depth map coding. The authors claim gains of 1 dB over H.264, but only for low bit-rates. In [9], a method is proposed to exploit the fact that depth map distortion does not translate directly into virtual view distortion (relationship is indirect: errors in the depth map cause geometric errors in the interpolated view, which translate to luminance/chrominance errors in the rendered views). They propose a method to estimate the rendered view distortion from the depth map distortion and use the rendered view distortion as a metric for RD optimization. The authors obtain coding gains up to 1.6 dB over applying RD based on conventional depth distortion metrics. In [10], a new prediction scheme is presented, aiming at reducing the rate devoted to blocks with arbitrary edge shapes. From the H.264 intra prediction framework, they construct an edge-aware prediction scheme based on a graph representation of pixels in a block and pixels from previously coded blocks. This way, edge information can be preserved. The method allows for a reduction up to 29% of the bit-rate for depth sequences for a fixed PSNR.

Our proposal is inspired on segmentation-based coding systems (SBCS) [11]. These systems partition images (in this case, depth maps) into a set of arbitrarily shaped homogeneous regions. The separate coding of each region allows reducing the coding artifacts due to sharp edges and, hence, it is very appropriate to encode depth maps. As the need to send the partition information (shape of the regions) to the decoder is one of the major efficiency drawbacks of SBCS, we present a novel solution that greatly reduces the cost associated to depth partition coding using the fact that the depth partition can be estimated using the decoded image color partition. Experimental results in Section 3 will show that the proposed method outperforms AVC and MVC when encoding MVD sequences.

This paper is organized as follows. Section 2 describes the approach proposed to encode depth maps in multiview sequences using a segmentation based technique. Section 3 presents the