J. Vis. Commun. Image R. 21 (2010) 513–522

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Depth and depth–color coding using shape-adaptive wavelets ☆

Matthieu Maitre [a], Minh N. Do [b,*]

[a] Windows Experience Group, Microsoft, Redmond, WA, USA

[b] Department of Electrical and Computer Engineering, The Coordinated Science Laboratory, and the Beckman Institute, University of Illinois at Urbana-Champaign, Urbana IL, USA

ARTICLE INFO

Article history:
Received 17 August 2009
Accepted 9 March 2010
Available online 21 March 2010

Keywords:
3D-TV
Free-viewpoint video
Depth map
Depth image
Shape-adaptive wavelet transform
Lifting
Boundary wavelets
Edge coding
Rate-distortion optimization

ABSTRACT

We present a novel depth and depth–color codec aimed at free-viewpoint 3D-TV. The proposed codec uses a shape-adaptive wavelet transform and an explicit encoding of the locations of major depth edges. Unlike the standard wavelet transform, the shape-adaptive transform generates small wavelet coefficients along depth edges, which greatly reduces the bits required to represent the data. The wavelet transform is implemented by shape-adaptive lifting, which enables fast computations and perfect reconstruction. We derive a simple extension of typical boundary extrapolation methods for lifting schemes to obtain as many vanishing moments near boundaries as away from them. We also develop a novel rate-constrained edge detection algorithm, which integrates the idea of significance bitplanes into the Canny edge detector. Together with a simple chain code, it provides an efficient way to extract and encode edges. Experimental results on synthetic and real data confirm the effectiveness of the proposed codec, with PSNR gains of more than 5 dB for depth images and significantly better visual quality for synthesized novel view images.

Published by Elsevier Inc.

1. Introduction

Free-viewpoint three-dimensional (3D-TV) video provides an enhanced viewing experience in which users can perceive the third spatial dimension (via stereo vision) and freely move inside the 3D viewing space [1]. With the advent of multi-view autostereoscopic displays, 3D-TV is expected to be the next evolution of television after high definition. Three-dimensional television poses new technological challenges, which include recording, encoding, and displaying 3D videos. At the core of these challenges lies the massive amount of data required to represent the set of all possible views – the plenoptic function [2] – or at least a realistic approximation.

The depth image-based representation (DIBR) has recently emerged as an effective approach [3], which allows both compact data representation and realistic view synthesis. A DIBR is made of pairs of and *depth* and *color* images (see an example in Fig. 1), each of which provides a local approximation of the plenoptic function. At the receiver, arbitrary views are synthesized from the DIBR using image-based rendering with depth information [4].

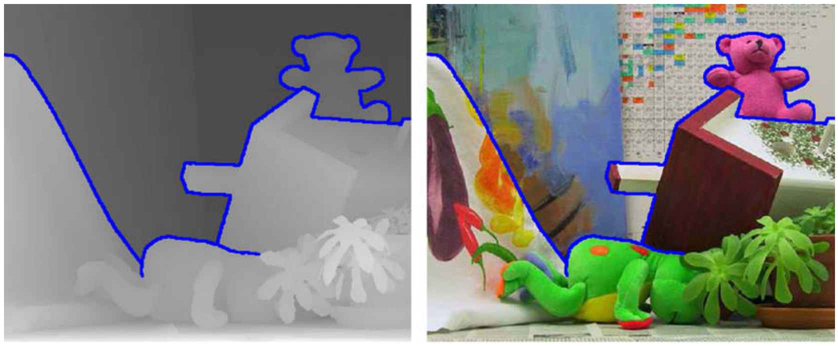
Depth information can be obtained by stereo match or depth estimation algorithms [5]. However, these algorithms are usually complicated, inaccurate and inapplicable for real-time applications. Thanks to the recent developments of lower-priced, fast and robust range sensors [6] which measure time delay between transmission of a light pulse and detection of the reflected signal on an entire frame at once, depth information can be directly obtained in real-time from depth cameras. This makes the DIBR problem less computationally intense and more robust than other techniques. Furthermore, depth information helps to significantly reduce the required number of cameras and transmitting data for 3D-TV. Thus, depth measurements will likely become ubiquitous in visual communication as they provide perfect complementary information to the traditional color measurements in capturing 3D scenes.

Depth inputs can be considered as monochrome images and therefore encoded using classical codecs like MPEG-2, H.264/AVC, or JPEG-2000 with only minor modifications [1,3]. However, the transforms used to decorrelate the data, for instance discrete wavelet transforms (DWT) or discrete cosine transforms (DCT), loose their decorrelation power along edges. This issue is paramount with depth images, which tend to exhibit much sharper edges than regular images. Context-adaptive entropy coding [7] can limit the rate-distortion (RD) impact of poor data decorrelation, but our experimental results shall show that designing better decorrelation transforms can still lead to significant RD gains.

☆ This work was supported by the National Science Foundation under Grant ITR-0312432. The work was performed while at the University of Illinois at Urbana-Champaign, Urbana IL, USA.

* Corresponding author.
E-mail addresses: mmaitre@microsoft.com (M. Maitre), minhdo@uiuc.edu (M.N. Do).

1047-3203/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.jvcir.2010.03.005



(a) Depth image and edges (b) Color image and edges

Fig. 1. Input data of the DIBR representation with shared edges superimposed over a depth image (a) and color image (b). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

A large research effort has been spent to design more efficient image transforms. Wavelet footprints [8] reduced scale correlation by representing edges explicitly, through their locations and a set of wavelet coefficients. Geometric wavelets [9–13] reduced space correlation by taking into account the 2D nature of images using non-separable wavelets. In [14] for example, major RD gains over JPEG-2000 were reported on depth images using a representation based on platelets [11].

Several issues limit the appeal of geometric wavelets and wavelet footprints for coding of depth images. These methods tend to have large computational requirements due to their reliance on complex RD optimizations [10,11,13,14] instead of established tools like fast filter banks, quantizers, and entropy coders in common codecs. Some of these methods also tend to rely on redundant data representations [8,9,12], which reduces RD performances.

Unlike typical color or grayscale images, depth images do not contain texture. In fact, as seen in Fig. 1(a), depth images are piecewise smooth with distinct edge around object boundaries that give sharp discontinuity in depth measurements. Therefore, unlike for traditional color and grayscale images, edges in depth images can be detected robustly and accurately (assuming that depth images have good quality). Moreover, for many image-based rendering algorithms with depth information, the accuracy of input depth image around object boundaries is very critical for the quality of the synthesized novel views [15]. A slight shift of an edge point in the depth image would lead to a large distance change in 3D of the corresponding point on the object boundary.

These observations lead us to consider *explicitly encoding locations of edges in depth images*. With encoded edge information, we can use the shape-adaptive discrete wavelet transform (SA-DWT) [16] to obtain invertible representation with small coefficients in both smooth regions and along encoded edges. The result is a simple, yet highly effective codec for depth images or depth–color image pairs.

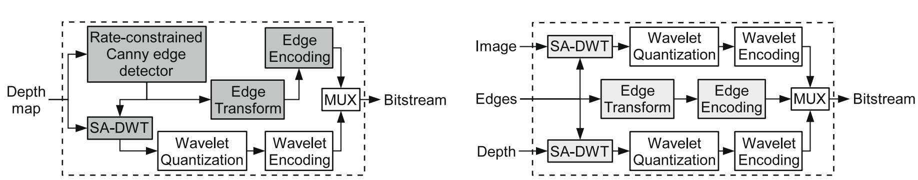
The remainder of the article is organized as follows. Section 2 presents an overview of our proposed codec. Section 3 presents the SA-DWT based on lifting with a novel border extension scheme. Section 4 details the encoding of edges with a novel rate-constrained edge detection. Section 5 presents experimental results. Some preliminary results have been presented in [17,18].

2. Proposed codec

Fig. 2(a) shows our proposed codec for depth images. The proposed codec takes advantage of the SA-DWT [16] implemented by lifting [19,7]. Lifting provides a procedure which is fast, in place, simple to implement, and trivially invertible. The locations of the main edges are encoded explicitly and the regions on opposite sides of these edges are processed independently, effectively preventing wavelet bases from crossing edges. As a result, the transform generates small wavelet coefficients both in smooth regions and along the encoded edges.

In order to handle signals with finite-domain, SA-DWT schemes extend the signals using usually zero-padding, periodic extension, or symmetric extension [20]. As a consequence, the transform looses vanishing moments near edges and tends to generate larger high-pass coefficients there. Boundary wavelets have been proposed to address this issue [21]. Here, we introduce a novel scheme which only adds trivial computations to the lifting scheme: a short linear filter is applied along edges, which provides a polynomial extrapolation of the signal with arbitrary order and results in transforms with as many vanishing moments near edges as away from them.

In our proposed codec, depth edges are detected using a novel rate-constrained version of the Canny edge detector [22]. The proposed edge detector replaces the two hysteresis thresholds of the original detector by a series of decreasing thresholds and associated edge significance bitplanes. These biplanes serve the same goal as wavelet significance bitplanes in codecs based on SPIHT and EBCOT [7]: they allow the most important edges to be encoded first, which leads to higher RD performances. Edges are encoded using a simple differential Freeman chain code [23].



(a) (b)

Fig. 2. Overview of the proposed encoder. It explicitly encodes the locations of the main edges and applies a shape-adaptive discrete wavelet transform (SA-DWT) to the depth and color images (gray boxes). (a) Depth only. (b) Depth and color.