DTU Library

3D Video Compression and Transmission

Zamarin, Marco; Forchhammer, Søren

Publication date:
2011

[Link back to DTU Orbit](#)

3D Video Compression and Transmission

Marco Zamarin, Søren Forchhammer

Coding & Visual Communications Group, Dept. of Photonics Engineering
Technical University of Denmark, Ørsteds Plads B343, 2800 Kgs. Lyngby, Denmark

{mzam,sofo}@fotonik.dtu.dk

Abstract: In this short paper we provide a brief introduction to 3D and multi-view video technologies - like three-dimensional television and free-viewpoint video - focusing on the aspects related to data compression and transmission. Geometric information represented by depth maps is introduced as well and a novel coding scheme for multi-view data able to exploit geometric information in order to improve compression performances is briefly described and compared against the classical solution based on multi-view motion estimation. Future research directions close the paper.

1. Overview of 3D Video technology

A Three-Dimensional Video (3D Video) system is able to offer to the user a better and deeper experience of video contents providing both the perception of the scene's depth and the ability to freely browse the scene itself. 3D Video has gained significant interest during the last years and due to the advances in acquisition processes and display technologies and it is now wide spreading in medical, entertainment, and communication industries. So, a number of novel 3D video applications are getting feasible [1] due to a certain improvement of capture and display technologies supported by advanced Multi-view Video Coding (MVC) techniques.

3D video applications can be grouped under three main categories: Free-viewpoint video (FVV), Three-Dimensional Television (3DTV), and Immersive teleconferencing. Figure 1(a) illustrates these new trends introducing the end-to-end architectures of different possible applications [2]. In this illustration, a multi-view video is first captured and then encoded by a multi-view video encoder. Then, a server transmits the coded bitstreams to different clients with different capabilities. At the final stage, the coded video is decoded and rendered with different tools according to the application scenario and the capabilities of the receiver.

A FVV system (Fig. 1(a), scenario (a)) provides realistic impressions of a captured 3D scene through an interactive browsing, i.e. the viewer can freely navigate in the scene analyzing it from different viewpoints [3, 4]. Whenever a desired viewpoint is not directly available, a virtual view is interpolated from the available ones through specific rendering algorithms. In this scenario, a limited number of candidate views need to be decoded, and therefore, the decoder can focus its resources only on decoding the right ones. In order to render the FVV content or to synthesize different virtual views, depth information about the considered scene is needed. The so-called "multi-view plus depth" representation permits browsing freely a three-dimensional scene represented by multiple views and depth maps, as if a 3D model was available. At the moment this approach appears to be one of the most effective solutions to the challenges of real-time FVV applications. However, in order to permit a sufficiently flexible browsing of the scene, a large number of views has to be transmitted, and their compression is a fundamental step in the construction of a practical multi-view system.

3DTV refers to extending the traditional TV broadcasts to the transmission of 3D video signals. In this application, more than one view is decoded and displayed simultaneously [4]. A simple 3DTV application can be realized by stereoscopic video. The stereoscopic visualization can be achieved by using particular glasses or other means. However, better 3D perception can be experienced by the end-user through 3D appliances able to render binocular depth cues [6], like autostereoscopic displays. In addition, simultaneous processing of multiple views permits obtaining wide viewing angles (see Fig. 1(a), scenario (b)). Whenever the decoder capability is limited or the transmission bandwidth decreases, the receiver may simply decode and render just a subset of the views still providing 3D display with a narrow view angle, as shown in Fig. 1(a), scenario (c). Since the classical TV or HDTV applications are still dominating the market, it is desirable for the MVC incoming technology to be backward compatible with the current 2D decoders, allowing them to generate a display from MVC bitstreams, as shown in Fig. 1(a), scenario (e).
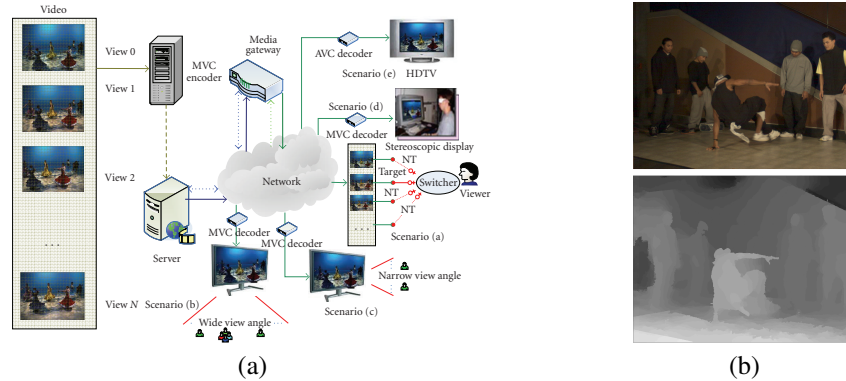
Fig. 1. (a) Overall MVC system architecture (reprinted from [2], copyright 2009 Hindawi). (b) A color image (top) and the relative depth map (bottom) from the "breakdancers" dataset [5].

2. 3D Video, what do we need?

There exist many different formats for 3D video data [7], mainly divided in two categories: image-based (e.g. light-fields and multi-view videos) and geometry-based (e.g. 3D mesh models). One of the most common ones is a hybrid one: the already mentioned "multi-view plus depth" video format. This representation includes two different types of data: color and geometry. Color information is provided by a number of synchronized video streams (typically between 2 and 16) of the same scene from different viewpoints while geometric information is described by depth maps. A depth map is a grey-scale image in which pixel values describe the distance between the depth camera plane and the relative 3D point in the real scene. Figure 1(b) shows an example of a color image with the associated depth map. A depth map can be used to project a view to a different viewpoint (operation called "*3D warping*") in order to obtain a predicted view and use it for coding purposes.

3. Compression of Multi-View Video

A lot of different schemes have been proposed in order to extend the up-to-date video coding techniques to multi-view texture data compression, like the H.264 MVC from ITU-T [8] that applies motion compensation techniques - typically adopted to reduce temporal redundancy - in order to predict one view from another relative to a different viewpoint. The Rate-Distortion (R-D) improvement obtained with such a technique is significant, but it is reasonable to think that better results - concerning both R-D and scalability - can be achieved exploiting the information relative to the 3D nature of the scene.

Our research direction focuses on novel lossy compression schemes for multi-view data able to fully exploit the inter-view redundancy in order to achieve better compression performances. Some preliminary results [9] have already been achieved through a coding scheme that presents a number of novelties compared to H.264 MVC. First, the inter-view motion prediction stage is replaced by a 3D warping procedure based on depth information. Then, the traditional 2D-DCT is replaced by a multi-dimensional transform, namely a 3D-DCT. The 3D-DCT is applied on disparity compensated data in order to better exploit view-redundancy in the transform stage. The transform coefficients are then coded by standard entropy coding algorithms. Figure 2(a) summarizes the whole scheme. The scheme has then been extended in order to include geometry compression [10]. Specifically, the algorithm encodes a single depth map through a standard H.264/AVC lossy encoder setting the quality in order to allocate around 25% of the available bandwidth for geometric information, as suggested in [7]. In this extended scheme the multi-dimensional transform is replaced by a standard H.264/AVC coder in order to better compress the views stack at the expense of computational complexity. Experimental results show that a preliminary version of the proposed scheme is able to outperform the state-of-the-art H.264 MVC encoder at low bit rates. Figure 2(b) shows the R-D performance comparison of the proposed scheme (red curves) against H.264 MVC (blue curves) for the "breakdancers" dataset (refer to [10] for all the details). The comparison is made both with and without considering the bit rate relative to the geometric information. As it is possible to notice, the proposed scheme performs better than MVC at low bit rates while at the high ones MVC performs better due to some limitations of the 3D warping process during the views stacking/unstacking operations.
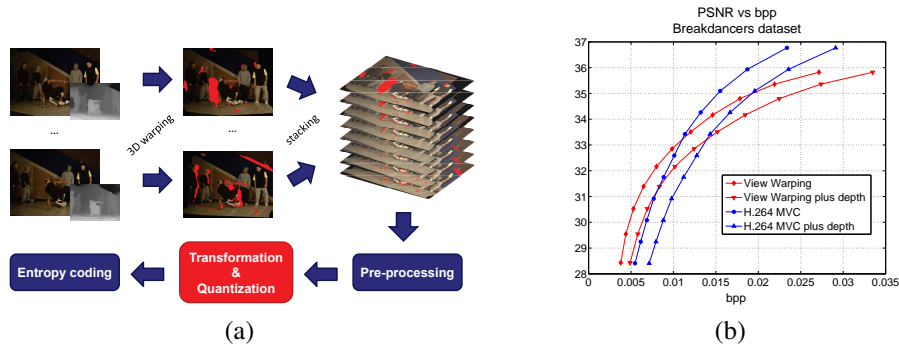
Fig. 2. (a) The proposed scheme (reprinted from [9], copyright 2010 Elsevier). (b) Coding performance comparison on the "breakdancers" dataset (reprinted from [10], copyright 2010 ACM).

4. Conclusion and Future work

In this short paper an alternative approach for multi-view image compression based on 3D warping and multi-dimensional transform coding is described. Preliminary experimental results show how the proposed approach can outperform the standard MVC scheme at low bit rates. Performances at high bit rates are not as satisfactory as the ones at low bit rates, but many improvements and optimization steps are possible. Future developments will extend the proposed scheme to multi-view video sequences including R-D optimization strategies in order to improve the compression performances at high data rates. Protection mechanisms against transmission errors will be developed as well. Moreover, compression schemes specifically targeted to geometry data will be investigated.

Another research direction focuses on live browsing of 3D contents. A few works about the optimal bit rate partitioning between geometry and texture data in order to maximize the Quality of Service (QoS) perceived by the final user have been proposed in the literature and the research on this topic is just at the beginning. 3D data streaming becomes even more critical in case of unreliable connections since packet losses could significantly degrade the quality of the reconstructed scene. Joint and reliable coding and transmission of both color and geometric data is still a widely unexplored field in which the possibilities for research are numerous.

References

1. A. Smolic, H. Kimata, and A. Vetro, "Development of MPEG standards for 3D and free viewpoint video," in *Three-Dimensional TV, Video, and Display IV, Proceedings of SPIE*, vol. 6016, Oct. 2005.
2. Y. Chen, Y. K.Wang, K. Ugur, M.M. Hannuksela, J. Lainema, and M. Gabbouj, "The emerging MVC standard for 3D video services," *EURASIP J. Appl. Signal Process.*, vol. 2009, no. 1, pp. 1-13, 2009.
3. C. Fehn, "Depth-Image-Based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TV," in *Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems XI*, pp. 93-104, 2004.
4. H. Kimata, M. Kitahara, K. Kamikura, Y. Yashima, T. Fujii, and M. Tanimoto, "System Design of Free Viewpoint Video Communication," *Computer and Information Technology, Int. Conf. on*, vol. 0, pp. 52-59, 2004.
5. L. C. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 600-608, 2004.
6. C. Fehn, and R. Pastoor, "Interactive 3-DTV-Concepts and Key Technologies," *Proceedings of the IEEE*, vol. 94, pp. 524-538, Mar. 2006.
7. H. Ozaktas, and L. Onural, *Three-Dimensional Television: Capture, Transmission, Display* (Springer, 2008).
8. ISO/IEC MPEG & ITU-T VCEG, "Joint Draft 8.0 on Multiview Video Coding," Jul. 2008.
9. M. Zamarin, S. Milani, P. Zanuttigh, and G.M Cortelazzo, "A Novel Multi-View Image Coding Scheme based on View-Warping and 3D-DCT," *Elsevier JVCI*, vol. 21, iss. 5-6, pp. 462-473, Jul.-Aug. 2010.
10. M. Zamarin, S. Milani, P. Zanuttigh, G.M Cortelazzo, and S. Forchhammer, "A Joint Multi-View Plus Depth Image Coding Scheme Based on 3D-Warping," in *1st Int. Workshop on 3D Video Processing (3DVP 2010)*, pp. 7-12, Oct. 2010.