DTU Library

Wyner-Ziv Coding of Depth Maps Exploiting Color Motion Information

Salmistraro, Matteo; Zamarin, Marco; Forchhammer, Søren

Wyner-Ziv Coding of Depth Maps Exploiting Color Motion Information

Matteo Salmistraro, Marco Zamarin, Søren Forchhammer

Department of Photonics Engineering, Technical University of Denmark, Ørsteds Plads,
2800 Kgs. Lyngby, Denmark

ABSTRACT

Distributed coding of multi-view data and depth maps is an interesting and challenging research field, whose interest is growing thanks to the recent advances in depth estimation and the development of affordable devices able to acquire depth information. In applications like video surveillance and object tracking, the availability of depth data can be beneficial and allow for more accurate processing. In these scenarios, the encoding complexity is typically limited and therefore distributed coding approaches are desirable. In this paper a novel algorithm for distributed compression of depth maps exploiting corresponding color information is proposed. Thanks to the high correlation of the motion in color and depth videos, motion information from the decoded color signal can effectively be exploited to generate accurate side information for the depth signal, allowing for higher rate-distortion performance without increasing the delay at the decoder side. The proposed scheme has been evaluated against state-of-the-art distributed video coding techniques applied on depth data. Experimental results show that the proposed algorithm can provide PSNR improvement between 2.18 dB and 3.40 dB on depth data compared to the reference decoder.

Keywords: Distributed Video Coding, DVC, Depth, OBMC.

1. INTRODUCTION

Recent advances show that Distributed Video Coding (DVC) is becoming a feasible alternative to traditional video coding in scenarios in which the encoding complexity is limited by strict constraints[1,2]. A typical scenario is multi-camera video surveillance, in which power consumption constraints require low-complexity encoders while complexity constraints are relaxed at the decoder side. In a typical DVC setting[3], a subset of the input frames called key frames (KF) is intra-coded (i.e. coded without exploiting temporal redundancy) and transmitted; for the remaining frames called Wyner-Ziv frames parity check data are calculated and made available for transmission. At the decoder side, intra-coded frames are decoded and missing frames are estimated from the available decoded ones. Estimated frames referred as side information (SI) are then corrected, if needed, using parity bits from the encoder. The generation of the SI is of crucial importance as it directly impacts the amount of additional data required to the encoder and therefore the coding efficiency of the whole scheme. Distributed coding of multi-view video has also been proposed in the literature[1]: thanks to the high intrinsic redundancy of multi-view data, it is possible to generate SIs for intermediate views allowing for efficient coding performance. A popular and compact 3D video representation is the so-called video-plus-depth in which depth map stream(s) are associated to mono-/multi-view video(s)[4]. Depth data allows for easy generation of virtual views through depth-image-based-rendering (DIBR) algorithms[4], making this representation suitable for flexible implementations of 3DTV and free viewpoint TV (FTV)[5]. Moreover, depth data can be exploited for other purposes such as activity detection, object tracking and background/foreground separation[6] which might be of interest in a video surveillance scenario. Due to the high correlation of the motion in the color and depth videos, color information can be effectively exploited to predict motion in the depth sequence, like in[7] and more recently in[8]. The same concept can be exploited in a DVC scenario for accurate SI generation, as done in[9] in which multiple decoded color frames are exploited. This paper addresses Wyner-Ziv coding of depth data in a video-plus-depth scenario. In the considered framework color data are available at the decoder and can be

therefore exploited in the Wyner-Ziv decoding of depth data. Specifically, motion vectors for color data are used on corresponding depth images allowing improving the performance of state-of-the-art DVC schemes on depth sequences. Differently from[9], in the proposed scheme the color frame corresponding to the to-be-decoded depth frame is available at the decoder thus allowing for accurate handling of motion. Moreover, a lower delay scheme is adopted as a fixed group of picture (GOP) size of 2 is used. The remaining of this paper is organized as follows: in Section 2 the proposed depth SI generation algorithm is described and discussed. Section 3 presents experimental results highlighting the benefits of the proposed method. Finally Section 4 closes the paper and outlines future work.

2. PROPOSED METHOD

The DVC decoder used in this work is based on[10]. Specifically, the SI generation and the residual estimation modules have been taken as starting point and improved in order to exploit color motion information for depth coding. When decoding the depth frame at instant t (referred as Dt), depth KFs at instant t-1 and t+1 (called Dt-1 and Dt+1, respectively) can be exploited for SI generation. Coding of depth maps using DVC is challenging as smooth and regular areas may mislead the Motion Estimation (ME) scheme employed by the SI generation module at the decoder. In the considered scenario color frames at instant t-1, t and t+1 (called Ct-1, Ct and Ct+1, respectively) are also available at the decoder (see Fig. 1).
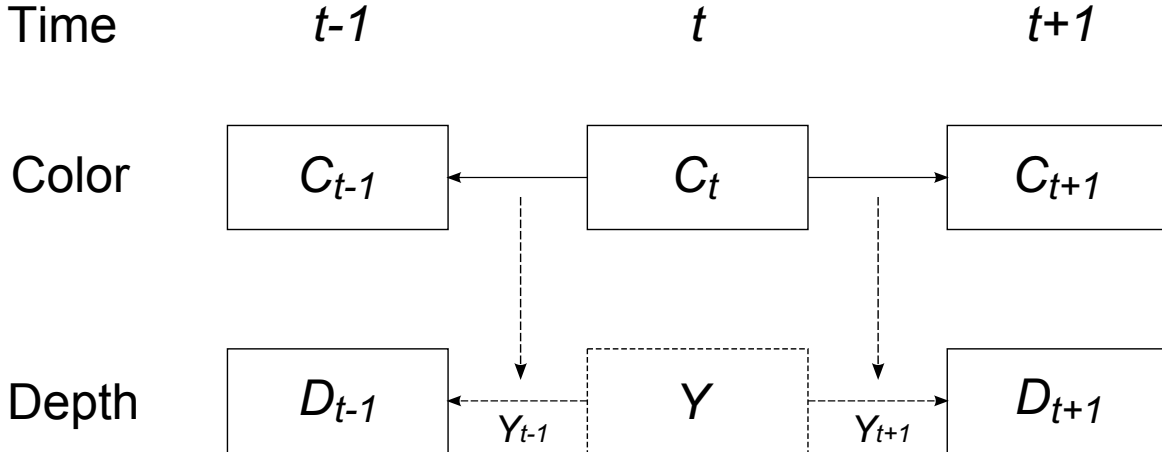

Figure 1. Depth SI generation.

If depth and color edges are perfectly aligned, the corresponding Motion Vectors (MVs) present similar behavior[9]. We exploit this fact in order to effectively predict Dt: first MVs between frames Ct and Ct-1, and between Ct and Ct+1 are calculated. Then, the two motion fields are used to compensate the depth frames Dt-1 and Dt+1 thus originating two depth SIs Yt-1 and Yt+1. Finally, the average Y = (Yt-1 + Yt+1)/2 of these two images is used as SI for the current to-be-decoded depth frame Dt. Together with the SI, an estimation of the residual signal i.e. the difference between the original frame Dt and the SI Y has to be calculated. The residual is important in order to calculate the reliability of the SI: a good estimation is crucial for an efficient parity bit allocation. Since Dt is not accessible at the decoder, the residual has to be estimated. In our case the absolute difference between Yt-1 and Yt+1 has been used as residual signal. The argument behind this simple but effective choice is that if a depth region in Dt is well predicted in Y, there is a high chance that the corresponding values in Yt-1 and Yt+1 are similar. Instead, regions of Dt that are not well predicted will most likely present significant differences between Yt-1 and Yt+1 in the corresponding regions. The performance of the ME module are also crucial: minor deviations in the color frames could create noticeable depth mismatches at the decoder (e.g. regions having similar texture might belong to objects with different distances from the camera plane), reflecting in low quality SIs. After various tests, the so-called Adaptive Rood Pattern Search ME algorithm proposed in[11] has been adopted. Even though this method is not providing the lowest mean squared error (MSE) between Y and Dt, it is the one which generates fewer artefacts in Y thanks to its two sequential search stages, hence allowing to use less parity bits compared to classic ME methods.

3. RESULTS

The proposed SI generation method has been evaluated on the Breakdancers and Ballet[12] DIBR sequences (view 4, 100 frames for both color and depth) with QCIF resolution at 15 fps. KFs have been intra-coded by means of H.264/AVC as in[13]. The following QPs have been used for depth data: 31, 34, 37, 40, with the following quantization matrices: Q7, Q6, Q4, Q1[13]. A fixed GOP of 2 has been used in all the cases. The proposed method has been compared against the scheme proposed in[10] and the reference DISCOVER decoder[14], both applied to depth map sequences. The rate-distortion curves for the Wyner-Ziv frames are shown in Fig. [INSERT IMAGES!!!!] for both on-line and off-line (i.e. supposing Dt is known at the decoder) residual estimations. Off-line results give an insight of the best performance achievable by the proposed scheme. Results show that the proposed method significantly outperforms the DISCOVER codec and provides a gain also over the method[10] for both test sequences. The presented approach suffers in case of high motion data and when significant occlusions between frames occur, like in the Breakdancers sequence. Better performance can probably be obtained if specialized ME methods and SI fusion algorithms are considered. Rate and distortion differences using Bjøntegaard metric[15] have also been evaluated for the proposed method with on-line residual: in the case of Breakdancers the PSNR gain is 0.30 dB and the bit-rate saving is 3.71% when compared to[10], and 2.18 dB and 35.08% when compared to DISCOVER. As for Ballet, the PSNR gain is 1.62 dB and the bit-rate saving is 24.44% when compared to[10], and 3.40 dB and 46.39% when compared to DISCOVER.

4. CONCLUSIONS

In this work we explored applying a modern DVC decoder to the depth map coding problem exploiting color motion information. The main contribution of the paper is a novel depth SI generation method able to exploit motion information from corresponding decoded color frames. Experimental results show that the proposed algorithm can outperform both the method of Huang et. al. [Huang2012] and the reference DISCOVER codec. Future work includes multi-hypothesis decoding and development of specialized ME algorithms for more efficient depth motion prediction. More experimental results will be included and evaluated as well.

REFERENCES

[1] Guillemot, C., Pereira, F., Torres, L., Ebrahimi, T., Leonardi, R., and Ostermann, J., "Distributed monoview and multiview video coding," *Signal Processing Magazine, IEEE* **24**, 67 –76 (sept. 2007).

[2] Pereira, F., "Distributed video coding: basics, main solutions and trends," in [*Proceedings of the 2009 IEEE international conference on Multimedia and Expo*], *ICME'09*, 1592–1595, IEEE Press, Piscataway, NJ, USA (2009).

[3] Girod, B., Aaron, A., Rane, S., and Rebollo-Monedero, D., "Distributed video coding," *Proceedings of the IEEE* **93**, 71 –83 (jan. 2005).

[4] Kauff, P., Atzpadin, N., Fehn, C., Müler, M., Schreer, O., Smolic, A., and Tanger, R., "Depth map creation and image-based rendering for advanced 3dtv services providing interoperability and scalability," *Signal Processing: Image Communication, Special issue on three-dimensional video and television* **22**(2), 217 – 234 (2007).

[5] Tanimoto, M., Tehrani, M., Fujii, T., and Yendo, T., "Free-viewpoint tv," *Signal Processing Magazine, IEEE* **28**, 67 –76 (jan. 2011).

[6] Mehrotra, S., Zhang, Z., Cai, Q., Zhang, C., and Chou, P., "Low-complexity, near-lossless coding of depth maps from kinect-like depth cameras," in [*Multimedia Signal Processing (MMSP), 2011 IEEE 13th International Workshop on*], 1 –6 (oct. 2011).

[7] Oh, H. and Ho, Y.-S., "H.264-based depth map sequence coding using motion information of corresponding texture video," in [*Proceedings of the First Pacific Rim conference on Advances in Image and Video Technology*], *PSIVT'06*, 898–907, Springer-Verlag, Berlin, Heidelberg (2006).

[8] Winken, M., Schwarz, H., and Wiegand, T., "Motion vector inheritance for high efficiency 3d video plus depth coding," in [*Picture Coding Symposium (PCS), 2012*], 53 –56 (may 2012).

[9] Petrazzuoli, G., Cagnazzo, M., Dufaux, F., and Pesquet-Popescu, B., "Wyner-ziv coding for depth maps in multiview video-plus-depth," in [*Image Processing (ICIP), 2011 18th IEEE International Conference on*], 1817 –1820 (sept. 2011).

[10] Huang, X. and Forchhammer, S., "Cross-band noise model refinement for transform domain wynerziv video coding," *Signal Processing: Image Communication* **27**(1), 16 – 30 (2012).

[11] Nie, Y. and Ma, K.-K., "Adaptive rood pattern search for fast block-matching motion estimation," *Image Processing, IEEE Transactions on* **11**, 1442 – 1449 (dec 2002).

[12] Zitnick, C. L., Kang, S. B., Uyttendaele, M., Winder, S., and Szeliski, R., "High-quality video view interpolation using a layered representation," in [*ACM SIGGRAPH 2004 Papers*], *SIGGRAPH '04*, 600–608, ACM, New York, NY, USA (2004).

[13] "Discover project test conditions," (December 2007). http://www.img.lx.it.pt/ discover/test_conditions.html.

[14] Artigas, X., Ascenso, J., Dalai, M., Klomp, S., Kubasov, D., and Ouaret, M., "The DISCOVER codec: Architecture, Techniques and Evaluation," in [*Picture Coding Symposium (PCS'07)*], (2007).

[15] Bjontegaard, G., "Calculation of average psnr differences between rd-curves," in [*VCEG Meeting*], *VCEG Meeting* (April 2001).