Transformation of Video Signal Processing Techniques from 2D to 3D: A Survey

Sanjay Koli and Rameez Shamalik

Abstract This paper presents an advanced depth intra-coding approach for 3D video coding based on the High Efficiency Video Coding (HEVC) standard and the multi-view video plus depth (MVD) representation. This paper is motivated by the fact that depth signals have specific characteristics that differ from those of natural signals, i.e., camera-view video. Our approach replaces conventional intra-picture coding for the depth component, targeting a consistent and efficient support of 3D video applications that utilize depth maps or polygon meshes or both, with a high depth coding efficiency in terms of minimal artifacts in rendered views and meshes with a minimal number of triangles for a given bit rate. For this purpose, we introduce intra-picture prediction modes based on geometric primitives along with a residual coding method in the spatial domain, substituting conventional intra-prediction modes and transform coding, respectively. The results show that our solution achieves the same quality of rendered or synthesized views with about the same bit rate as MVD coding with the 3D video extension of HEVC (3D-HEVC) for high-quality depth maps and with about 8% less overall bit rate as with 3D-HEVC without related depth tools. At the same time, the combination of 3D video with 3D computer graphics content is substantially simplified, as the geometry-based depth intra signals can be represented as a surface mesh with about 85% less triangles, generated directly in the decoding process as an alternative decoder output.

Keywords Video coding · 3D video · Multi view with depth · HEVC

1 Introduction

With the latest breakthroughs and ease of access in Internet technology, an exponential increase in data usage can be seen around the world. Internet is basically

S. Koli (✉)
Dr. D.Y. Patil School of Engineering, Pune, India
e-mail: sanjay.koli@dypic.in

R. Shamalik
Bharati Vidyapeeth's College of Engineering for Women, Pune, India

used for two purposes namely information and communication, majority of which is provided to the end user by the means of either images or videos. Videos have easily overtaken images in information and communication areas due to advent of 4G communication and increasing number of devices with high resolution recording as well as more storage space. We are in the era of never before seen data giving rise to new challenges and opportunities to store, handle and process data especially videos to extract useful information out of it.

Video processing and coding techniques are mostly used for video surveillance right from a Person identification to Traffic monitoring for vehicle tracking, in factories or nuclear plants, for military and spying purposes, Underwater activities and archeological expeditions, Medical diagnosis and agricultural activities. The List goes on. Modern day Cinema is a major contributor to video processing and coding techniques due to huge demand in Special effects as well as animated movies and has played a key role in transformation from 2D to 3D videos. 3D videos are proving out to be a cutting edge technology as new paradigms like depth and geometry are included in them to be worked upon.

In this paper we highlight promising techniques used in development of video processing and coding. This paper is divided into four sections. Section 2 discusses the background of various approaches taken for utilization of this technology and overviews the latest 3D coding Techniques and Sect. 3 provides comparison of the techniques followed by Sect. 4 providing conclusion.

2 Background

2.1 2D Techniques

Video-Based Face Recognition on COX Face Database Face recognition from a given video is the need of an hour as we see number of CCTV user multiplying. It surely is bit complicated as compared to age old image based face recognition. A database was created using combination of still images and videos called COX Face Database whose initials comes from the union of three institutions who worked on this project namely Chinese Academy of Sciences (CAS), OMRON Social Solutions Co. Ltd. (OSS), and Xinjiang University [1]. Based on the experiments done a new method was proposed named Pointto-Set Correlation Learning (PSCL) which takes care of correlation between different data types simultaneously. Still a lot of work remains to be done when it comes to Face recognition from a Real-time video.

Vehicle Detection by Adaboost Algorithm Traffic monitoring is the need of an hour especially in fast growing cities. Vehicle detection plays a key role in it where it is divided into two steps off-line training and on-line identification [2]. Haar feature is used to extract the image from a large pile of vehicle samples for off-line training process. Further processing is done by Adaboost algorithm to recognize the required
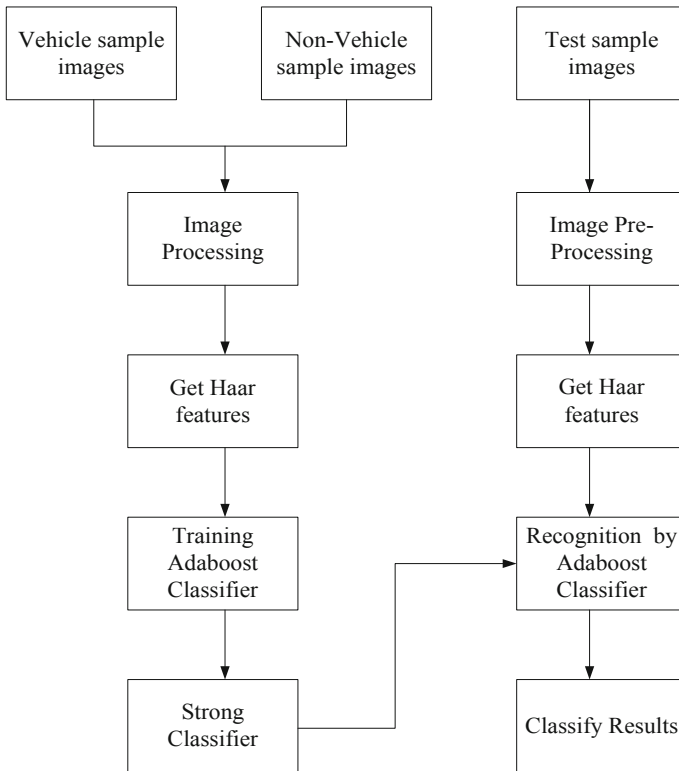
Fig. 1 Adaboost algorithm application [2]

sample given by Haar Classifier as shown in Fig. 1. This technique works only when the data is already saved.

Real-Time Video Processing using Adaptive Edge Detection on FPGAs Edge Detection is most tried and tested algorithm when it comes to Real time video and image processing. It comes really handy when the sample taken from the video is blurred or the video stream itself is anti-aliased. Neoh et al. [3] uses Canny edge detection to serve the above purpose. In Canny algorithm an outline is highlighted between object of interest and background. Thus providing maximum localization and minimum error rate. Although FPGA provides an economical and apt alternative to run the real time video processing giving good enough computational speed for canny algorithm, as the complexity of the algorithm increases computational speed decreases.

Video Surveillance Using Guassian Mixture Modelling (GMM) and Camshift A novel monitoring system to improve safety for personnel in nuclear plant is proposed by Jorge et al. [4]. There is a possible Radiation Hazard in the environment for working personnel. Hence background modeling and subtraction method GMM

is used. Further Camshift technique is used for individual tracking using color histogram. Camshift is an improved version of Mean shift method specially used for video processing. It proves helpful as cumulative tracking of region of interest is done in a distributive manner. Citing the criticality of a nuclear plant this method should be used as a redundant technique.

2.2   3D Techniques

As discussed earlier 3D videos are next generation of entertainment for the viewers. Hence video coding techniques such as Multi View Coding (MVC) and High Efficiency Video Coding (HEVC) were developed over the years.

As shown in Fig. 2, 3D video coding can be explained in four steps. Firstly a video is captured using multi view camera to have multiple frames. Then processing is done on the frames collectively to produce a 3D video, Further the 3D video is coded with the latest coding techniques for transmission and finally at the receiver side the video is rendered with depth estimation.

In this section we summarize the complementary techniques associated with the above mentioned coding standards for presenting the best possible experience to the end user.

Motion Vector Inheritance (MVI)   The basic differentiation between a 2D and 3D video is the visibility of depth inside the frames. When it comes 3D video coding MVC has limitations as does not support the coding of associated depth information. It basically encodes all the views together hence diluting the all-important depth data while decoding which gives 3D video superiority. In MVI a prediction model is proposed for efficient 3D video coding in which motion vectors are repeatedly used and effective block partitioning of video signal is done [5]. The only difference while using MVI is the increased complexity in encoder side.

Multi-view Video with Depth Data   In this technique Müller et al. [6] proposes addition to HEVC standard in terms of inter-view motion parameter and inter-view residual prediction. Authors developed 3D video coding (3DVC) extension for established HEVC standard. Backward compatibility with 2D video coding is maintained by separately coding multi view video and depth. As shown in Fig. 3, an alternative coding tools are introduced for coding the dependent views and the depth data with inter-component prediction techniques that work on data from already coded components simultaneously, as indicated by the grey arrows. At the encoder side depth-enhanced formats using view synthesis optimization with block-wise synthesized view distortion change is achieved while at decoder side Depth image based rendering(DIBR) is done for the required final view.

Multiview Video Plus Depth Coding With Depth-Based Prediction Mode
Predictoin plays an important role in high quality video transfer by not only saving the required Data but also increasing the Bit rate for faster communication. Bal
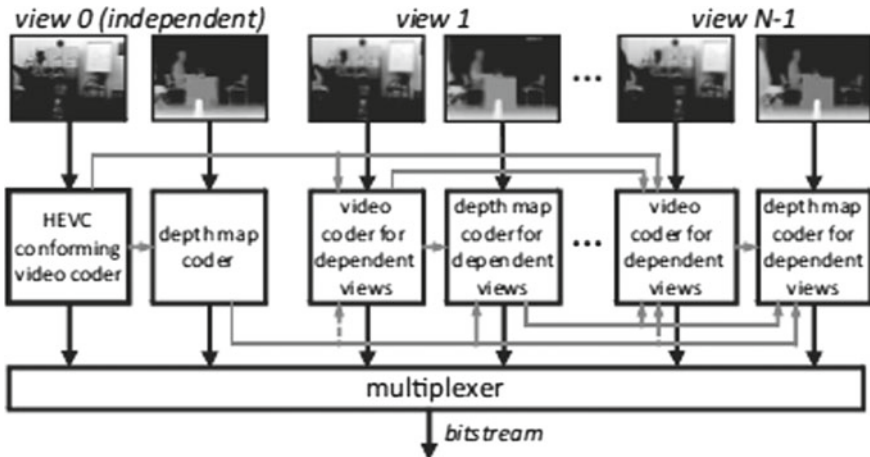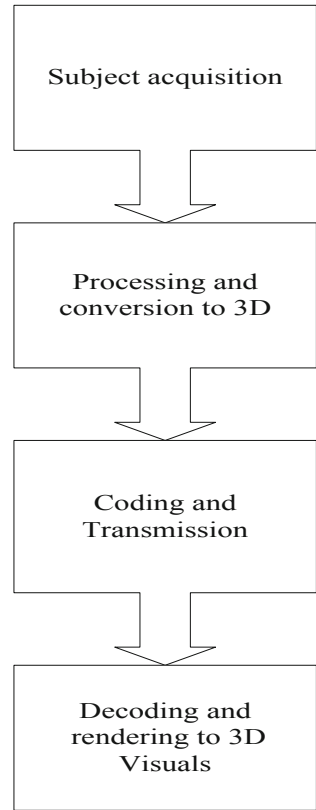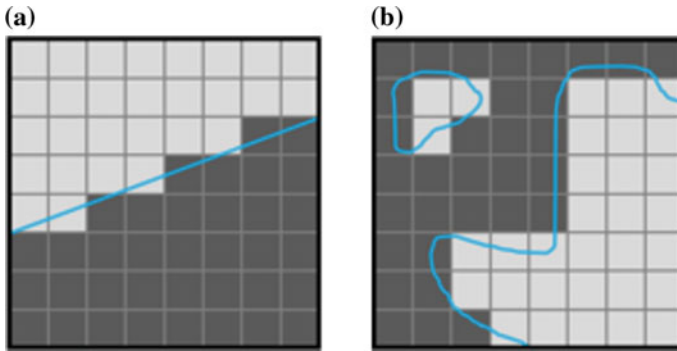
Fig. 2 3D video coding
steps

Fig. 3 Encoder structure with inter-view and inter-component prediction [6]

Fig. 4 a Wedgelet partition, **b** contour partitions [8]

et al. [7] propose a depth-based prediction mode (DBPM) technique named multiview video plus depth (MVD) which covers video as well as depth data. As the prediction method focuses on depth only, accuracy of encoder increases and efficiency of system to show 3D video increases at a very low cost. Few minor changes needs to be done in the syntax of MVC format.

Depth Intra Coding based on Geometric Primitives The focus is shifted from just depth to geometric analogy in block segmentation. The approach is to approximate an object of interest by reducing the block sections, in turn lowering the load on decoder with giving efficient throughput. New intra and inter-component prediction modes, using wedgelet and contour block partitions, and a complementary residual adaptation method in the spatial domain are introduced shown in Fig. 4. A gap between 3D video and 3D computer graphics is reduced by introducing a 3D-HEVC extension to highly used standard of HEVC [8] Further work can be extended in temporal and inter-view consistency of surface meshes.

3   Comparative Study

After presenting various video signal processing techniques spanning from 2D to 3D, it becomes necessary to compare them in terms of their advantages and disadvantages as shown in Table 1. 2D video signal processing techniques have advantage in terms of simplicity and economical alternatives but as the complexity of the application increases their usability is restricted on the other hand 3D video signal processing techniques stand apart as per the current user requirements an crucial applications but are associated with complex techniques and algorithms with new coding standards and extensions. Based on the comparative study we strongly suggest to work on 3D video signal processing techniques to come up with state of the art coding techniques to fulfil requirements of end user and critical applications.

Table 1 Comparative study

Ref. no.	Technology	Advantages	Disadvantages
[1]	Face recognition on COX face database	Combination of still images and videos with new method PSCL	Work can only be done on available database and not on real-time videos
[2]	Haar Classifier-Adaboost algorithm	Vehicle identification with improved speed	Result may vary for different vehicle models
[3]	Canny edge detection on FPGA	Supports real-time image processing with FPGA as economical Alternative	Complexity of the algorithm is proportional to measuring Cost
[4]	GMM and Camshift	Small percentage errors due to color confusion	Prone to error for people with similar clothes, height and background regions
[5]	Motion vector inheritance (MVI)	Hybrid video coding approach with block partitioning	Small increase in encoder efficiency
[6]	Multiview video with depth	Increased compression and coding efficiency By 3D-HEVC	Unable to handle large arrangements
[7]	Depth-based prediction mode	Complexity analysis insures high encoder accuracy	Tradeoff among depth map resolution and quantization parameter
[8]	Multiview video with geometry	3D-HEVC developed with backward compatibility	High complexity

4   Conclusion

A brief collection of key video signal processing techniques is presented. Journey of video signal processing from 2D to 3D techniques is highlighted with their applications and limitations. Comparative study with respect to the pros and cons of the presented techniques is done. Based on which we conclude that 3D video signal processing is the new emerging and promising technique to be worked upon. The latest work done in 3D video coding iterates its importance in the era of unlimited data. Standards like HEVC and 3D-HEVC are really effective in coding multi view video with depth and geometry respectively. A new displays can be built to show 3D videos without user needing any glasses as Multi view videos with depth and geometry are being developed.

References

1. Huang Z (2015) A benchmark and comparative study of video-based face Recognition on COX face database. IEEE Trans Image Process 24(12)
2. Cao J (2016) Research on urban intelligent traffic monitoring system based on video image processing. Int J Signal Process, Image Process Pattern Recognit 9(6):393–406
3. Shan H, Hazanchuk NA et al (2005) Adaptive edge detection for real-time video processing using FPGAs. Altera Corporation
4. Jorge CAF et al (2013) Improved people detection in nuclear plants by video processing for safety purpose. In: International nuclear Atlantic conference—INAC, ISBN: 978-85-99141-05-2
5. Winken M et al Motion vector inheritance for high efficiency 3D video plus depth coding. In: Picture coding symposium, Kraków, Poland, 2012
6. Müller K et al (2013) 3D high-efficiency video coding for multi-view video and depth data. IEEE Trans Image Process 22(9)
7. Bal C et al (2014) Multiview video plus depth coding with depth-based prediction mode. IEEE Trans Circuits Syst Video Technol 24(6)
8. Merkle P, Müller K, Marpe D, Wiegand T (2016) Depth intra coding for 3D video based on geometric primitives. IEEE Trans Circuits Syst Video Technol 26(3)