Emergence and Functionality of 3D Videos

Rameez Muneer, S. M. Koli

Abstract: Millions of videos are created and watched every day. Quality of desired video plays a significant role in modern day applications. Deep fake videos have alarmed the established institutions while amused others. A Three-dimensional(3D) stereoscopic film or a 3D Video for that instance is a motion picture that enhances the illusion of depth perception, hence adding a third dimension. 3D videos can effortlessly validate and secure the authenticity of communication and broadcast process due to their auxiliary effect. In this research paper we attempt to observe and analyze numerous measures taken towards initiation and rise of 3D videos along with their efficacy from an end user perspective.

Index Terms: 3D Video, 3D video Coding and Processing, 3D-HEVC, Virtual Reality.

I. INTRODUCTION

Through recent technological breakthroughs and ease of Internet access, world is witnessing tremendous amount of data usage. Internet is broadly utilized for either information or communication, usually consisting of images as well as videos. New possibilities and challenges have aroused not only to handle and process data and videos but also to withdraw favorable deductions from them. The basic purpose behind Video processing or coding techniques are for video surveillance of traffic, factories, theft detection, for medical diagnosis, archeological activities and agricultural experiments.

Omni directional (360-degree) video applications are picking up pace in Video processing and communication. It finds its application in interactive shopping and movies, real time broadcast concerts etc. 360-degree is the main video format used in virtual reality which provides application like video games[1]. Depth image-based rendering (DIBR) also plays a vital role in **Virtual Reality(VR)**[2]. Systematically aligned and synchronized cameras capture scenes from multiple views in a single shot to provide texture based 3D videos[3]. Accordingly an ad-hoc network of cameras which is focused in same direction but located at different angles creates a formidable quality 3D video[4].

In order to achieve perception of 3D video without opting for special glasses a different stereo pair is assigned as approaching angle of spectator can change multiple times[5]. Light field displays and 3D displays can surely benefit edutainment. Also a shopper's choice based marketing is being done as light field video-on-demand [6]. It's quite feasible for any individual to get his hands on VR by using various Head Mounted Displays (HMD) such as Oculus Rift[7] HTC VIVE [8], to enjoy real life 3D effects.

Subjective monitoring with the help of real observers provides best in class viewing impression but falls short on the test of time and being pocket friendly[9]. On the other hand, its objective counterpart which is purely based on mathematical modeling gives acceptable viewing experience prediction and is repeatable and fast as well[10]. **Structure-from-Motion** (SfM) is a popular for multiple angle views and 3D reconstruction of scalable areas [11]
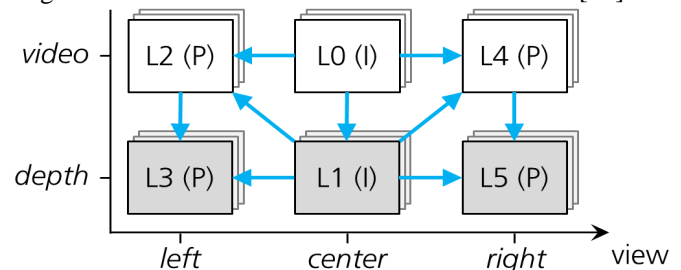

Figure 1:[12] shows three key features to obtain 3D video in a effortless manner. Y axis has video as well as depth while x axis attributes the viewing directions namely Left,Center and Right. Authors introduce a brand new codec 3D-High Efficiency Video Coding(3D-HEVC) which works along with multi-layer coding design of HEVC. It gives Multi-view prognosis between the different angles of either **video or depth**. Interactive holographic byproduct can be seen by multi-component prognosis between the video and depth component of the same view and **integrated multi-view or multi-component** prognosis between video and depth of contrasting angles.

II. RELATED WORK

Jamali et al.[13] compared the execution of contrasting projections in 3D and multi-view bit rate which utilises disparity estimation that comes out dull when evaluated with simulcast coding as more time spent for inter-view prediction however it gives maximum compression but later stands comfortable not only in quality and coding efficiency but also in Complexity. Kim et. al[14] came up with a hole filling method which ensures global optimization and make use of the information from pre-filled coherent view and previous frame. An economical model for label propagation is extended which serves the purpose of maintaining the spatio-temporal consistency and symmetry between two coaxial frames in synthesized 3D videos at many virtual viewpoints.

Retrieval Number: C6513029320/2020©BEIESP
DOI: 10.35940/ijeat.C6513.029320
Journal Website: www.ijeat.org

4319

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

This model propagates measurable labels available in neighboring view and last frame for the target image to be overwhelmed for desired throughput.
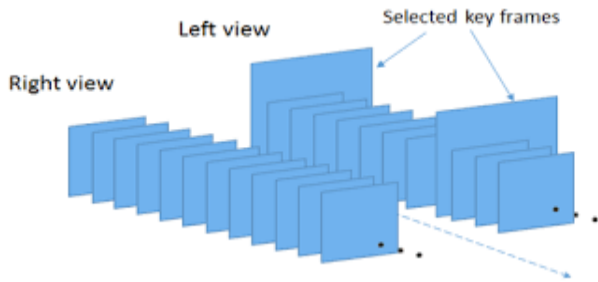

Figure 2: [15] highlights an array of frames in which Malik et. al for achieving HEVC based spacial coverage scaling mixed resolution coding presents a prototype for multi view videos containing interleaved frames. Real size of basic frames is kept constant while information in neighboring frames of the center and coherent views are down-sampled. Muller et. al[16] came up with new 3D video standard(3D-HEVC).This codec seves as a supplementary addition to the highly used HEVC standard. This standard facilitates stereoscopic multi-view platforms to provide multi-view video plus depth. Concept like disparity-compensated prediction, inter-view motion parameter and residual prediction are further extended and combined for coding of the individual videos. The HEVC extension further inculcates depth coding, new intra coding modes, a modified motion compensation and motion vector coding also motion parameter inheritance.

(a)

(b)

Figure 3:[17] shows two variants of the same frame proposed by Ploshkin et. al for a temporal-shift estimation algorithm which highlights geometric distortions between stereoscopic views with sub frame accuracy. As it is assumed constant throughout each scene , it makes the process quite strong. Motion and Disparity maps are obtained using Block-Matching algorithm. Tamboli et. al[18] assessed the quality of visuals with respect to light field video frames. Same distortions to the source images were given while retaining the default parameter settings of the objective

metric to evaluate the quality; they concluded that the distortions were separated in metric space decently.

A database is created by Wu et. al[19] with various formats of projections for Efficient VR quality assessment. subjective scores are given based on the assessment quality of self-built. Traditional image quality assessment metrics are used for the database Validation. At the end, a 3D convolutional neural network is designed to predict the VR video quality without reference VR video. Keeping in view features like **spatial resolution** and **perceptually significant depth level,** Reduced Reference (RR) metric is developed for depth perception of the 3D video by Gokce[20]. Abstraction filter is effectively used for development of the proposed metric. Structural Similarity metric (SSIM) is utilized to predict the depth perception degraded with the change in the perceptually important levels of the compressed depth maps having dissimilar spatial resolutions and relative depth levels.

(a)   **(b)**

Figure 4: [21] shows two different versions of same image along with associated depth image in which darkest spot shows the far distance of pixel from the user while the brightest one highlights the close proximity. The 3D video quality metrics are characterized with respect to a normal reference which ioncludes Full , No and Reduced Reference (NR) of which NR depth perception prediction metric is chosen for advanced 3D video services. For depth estimation of a unique image Zhou et. al[22] chose Deep Learning approach. A Convolutional Neural Network is used to generate a depth map with the help of an RGB image which converts a video shot with a single camera to Stereoscopic content tailored for viewing in a VR environment. Winken M et al. [23] opted for Motion Vector Inheritance (MVI), a prognosis prototype which is introduced for efficient 3D video coding in which motion vectors are rotationaly used for resulting block partitioning of video signal.

Table I Various techniques for 3D-HEVC[24]

	Technique	Abbr.	Description	Dependency	Ref.	Pred.	Sec.
Texture	Neighboring block disparity vector	NBDV	Derives predicted disparity information (PDI) for a CU	V, A, I	M, D	D	IV-A1
	Extended TMVP[1] for merge	-	Extends TMVP to also operate for inter-view prediction	V, A	M	M	IV-C1
	Inter-view motion prediction[2, 3]	IV, IV$_S$	Uses PDI for inter-view prediction of merge candidates	V	M	M	IV-C4
	Disparity information merge cand.[3]	DI, DI$_S$	Uses PDI directly as a merge candidate	-	-	M	IV-C2
	Residual prediction[3]	RP	Predicts residual from a different view or AU	V, A, A+V	S, M[4]	S, R	IV-D1
	Illumination compensation	IC	Adapts an inter-view sample prediction to the current view	V, I	S	S	IV-D2
	Depth refinement[3]	DR	Improves PDI using disparity from a depth map	C+V	S	D	IV-A2
	View synthesis prediction[2, 3]	VSP	Derives merge candidates from samples of a depth block	C+V, I	S	M	IV-C6
	Depth based block partitioning[3]	DBBP	Subdivides an inter-pic. predicted CB based on a depth block	C+V, V, A	S	S	IV-D3
Depth	Inter-view motion prediction	IV	Uses a default disparity value to predict a merge candidate	V	M	M	IV-C4
	Full sample motion accuracy	-	Reduces ringing artifacts and complexity	-	-	-	IV-D4
	Intra_Wedge mode	-	Subdivides a PB by a straight line and predicts DCs	I	S	S	IV-E1
	Intra-pic. skip / Intra_Single mode	-	Signals an angular mode or a single value for the entire CB	I	S	S	IV-E2
	DC offsets / DC-only mode	-	Codes residual DC explicitly / skips transform coefficients	-	-	-	IV-F1
	Depth look-up table	DLT	Quantizes the residual DC non-linearly	-	-	-	IV-F2
	Quadtree limitation	QTL	Derives a depth CB partitioning from texture	C	P	P	IV-B
	Texture merge candidate[2]	T	Derives a merge candidate from texture	C	M	M	IV-C5
	Intra_Contour mode	-	Predicts a PB subdivision from texture and predicts DCs	C, I	S	S	IV-E1

Dependency from picture in different: V) View; C) Component A) AU; A+V) AU and View; C+V) Component and View; I) Intra-picture.
Reference and predicted information: M) Motion; D) PDI; S) Sample; R) Residual; P) Partitioning syntax.
[1]Temporal motion vector prediction. [2]Using sub-block partition (SBP) motion accuracy (IV-C3). [3]Depends on NBDV. [4]For RP from a different AU only.

Table I provides a bird's eye view of the 3D-HEVC texture and depth coding techniques. enhanced compression operation is produced due to novel IV prediction schemes and texture coding tools.Many of the texture coding measures are deducted disparity for IV prognosis from different available samples of a pre-decoded depth layer. Texture-only coding methods outperforms depth-dependent ones.

III. RESULT AND DISCUSSION

We can work upon a non linear regression model to produce variety of parameters with extra coefficients. We can work on coefficients like gain, offset and gamma using the given formula,

$$P_{ref} = C_{gain} \{P_{tar} /(^{2bitdepth}-1)\}^c \; gamma \; (^{2bitdepth}-1) + C_{offset}$$

The results of which are shown in Table II[25]

Coefficient	Red	Green	Blue
Gain	0.5456	0.1136	0.3702
Offset	13.9680	0.46	0.37
Gamma	0.42	0.39	0.38

As shown above all three parameters are maximum for color red and in a reverse way minimum for color blue. Almost every screen which is used on daily basis gives us a 2D video; however 3D video adds depth and validity to the video processing and coding techniques. 3D glasses were compulsory to experience a 3D video but now state of the art cameras recording same frame with different angle is capable of delivering the same 3D effect to the end user. Available Video Processing techniques are mostly binary as in outcomes are in terms of either positive or negative. Techniques are complex including multidisciplinary fields, can be replaced by a simple model with clear objectives. Some of the techniques include training a Neural Network which can be a lengthy process. Datasets created for Processing can be divided according to application. Estimations are limited to certain human facial expression and gestures.

IV. CONCLUSION

360 degrees screen also facilitate the same effect with DIBR to give a Virtual reality experience. Deep learning algorithms efficiently takes care of depth in a video frame especially when a Neural network is trained for it. Neural networks can be bit complex at times and does not guarantee reliability on a repetitive basis. Block partitioning contribute to geometric realisation of 3D videos while prediction adds pace to the whole process. Both interview and inter-component prediction gives satisfactory results. Guassian Mean Models and its types are extensively used for proper estimation of videos. Video Processing is done on database available from various sources. Histograms also play a vital role for edge estimation of videos. New standards have been introduced for 3D video processing. In 3D Video Processing, major work is done on gesture and human activity detection. Depth and Geometry have added new paradigms for 3D video processing whereas prediction adds value to the whole process. For future scope real time

processing based on direct application of 3D video processing can be proposed.

REFERENCES

1. L. Argyriou, D. Economou, V. Bouki, and I. Doumanis, "Engaging immersive video consumers: Challenges regarding 360-degree gamified video applications," in 15th International Conference on Ubiquitous Computing and Communications, pp. 145–152, Dec 2016.
2. P. Ndjiki-Nya, M. Koppel, D. Doshkov, H. Lakshman, P. Merkle, K. Muller, and T. Wiegand, "Depth image-based rendering with advanced texture synthesis for 3D video," IEEE Trans. Multimedia, vol. 13, no. 3, pp. 453-465, Jun. 2011.
3. A. Smolic, "3D video and freeviewpoint video: From capture to display," in Pattern recognition, vol. 44, no 9, pp. 1958-1968, 2011.
4. Matthijs Douze, J´erˆome Revaud, Jakob Verbeek, Herv´e J´egou, and Cordelia Schmid, "Circulant temporal encoding for video retrieval and temporal alignment," International Journal of Computer Vision, vol. 119, no. 3, pp. 291–306, 2016.
5. P. Benzie, J. Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, and C. V. Kopylow, "A survey of 3DTV Displays: Techniques and technologies," IEEE Trans. Circuits Syst. Video Technol., vol. 17, no. 11, pp. 1647–1658, Nov. 2007.
6. A. Cserkaszky, A. Barsi, P. A. Kara, and M. G. Martini, "Towards display-independent light-field formats," International Conference on 3D Immersion (IC3D), Dec 2017, pp. 1–7,2017.
7. Oculus,Rift,Online,https://www.oculus.com/rift/#oui-csl-rift-games,2018l
8. HTC Valve, Online, https://www.vive.com/us/product/vive-virtual-reality-system/, 2018.
9. Q. Huynh-Thu, P. Le Callet, and M. Barkowsky, "Video quality assessment: From 2d to 3d; challenges and future trends," in 2010 IEEE International Conference on Image Processing, pp. 4025–4028,Sept 2010.
10. P. Lebreton, A. Raake, M. Barkowsky, and P. Le Callet, "Evaluating depth perception of 3d stereoscopic videos," IEEE Journal of Selected Topics in Signal Processing, vol. 6, no. 6, pp. 710–720, Oct 2012.
11. Sameer Agarwal, Noah Snavely, Ian Simon, Steven M Seitz, and Richard Szeliski, Building rome in a day," in 2009 IEEE 12th ICCV. IEEE, pp. 72–79,2009.
12. H. Schwarz, T. Wiegand, Inter-view prediction of motion data in multiview video coding, Picture Coding Symposium (PCS 2012), pp.101-104, 7-9 May 2012.
13. Mohammadreza Jamali et. al,"Comparison of 3D 360-Degree Video Compression Performance Using Different Projections", IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), pp. 58-64,5-8 May 2019.
14. Hak Gu Kim, Yong Man Ro," Multi-view Stereoscopic Video Hole Filling Considering Spatio-Temporal Consistency and Binocular Symmetry for Synthesized 3D Video", IEEE Transactions on Circuits and Systems for Video Technology, Volume: 27 , Issue: 7 , pp. 1435 – 1449, July 2017.
15. Bruhanth Mallik, Akbar Sheikh Akbari, Ah-Lian Kor," Mixed-resolution HEVC based multiview video codec", 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV- CON),pp.104-108, 2017.
16. Karsten Müller, Heiko Schwarz, Detlev Marpe et. al," 3D High-Efficiency Video Coding for Multi-View Video and Depth Data", IEEE Transactions on Image Processing, Volume: 22 , Issue: 9, pp.3366 3378,Sept.2013.
17. Aleksandr Ploshkin and Dmitriy Vatolin," Accurate Method Of Temporal-Shift Estimation For 3d Video", 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON),pp.214-218,2018.
18. Roopak R. Tamboli, Peter A. Kara ,Aron Cserkaszky , et. al," 3d Objective Quality Assessment Of Light Field Video Frames", 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON),pp.405-409,2018.
19. Pei Wu,Wenxin Ding, Zhixiang You,And Ping An,"Virtual Reality Video Quality Assessment Based On 3d Convolutional Neural Networks",IEEE International Conference on Image Processing (ICIP),pp.
20. 623628,2019.
20. Gokce Nur Yilmaz,"A Depth Perception Evaluation Metric For Immersive 3d Video Services". 3DTV Conference: The True

Vision-Capture, Transmission and Display of 3D Video (3DTV-CON),pp.1-4, 2017.

21. Gokce Nur Yilmaz A, Federica Battisti," Depth Perception Prediction Of 3d Video For Ensuring Advanced Multimedia Services", International Conference on Information Networking (ICOIN),pp.146-149,2018.

22. Xichen Zhou, Bipin C. Desai, Charalambos Poullis," Automatic 2d To Stereoscopic Video Conversion For 3d Tvs", 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON),pp.112-116,2017.

23.
Martin Winken_, Heiko Schwarz_, and Thomas Wiegand,"Motion Vector Inheritance for High Efficiency 3D Video plus Depth Coding",IEEE 2012 Picture Coding Symposium,pp65-69,2012.

24.
Gerhard Tech et. al, "Overview of the Multiview and 3D Extensions of High Efficiency Video Coding "IEEE Transactions on circuits and systems for video technology, vol. 26, no. 1, pp.35-49,January 2016.

25. Yo sung ho,"3D Video Processing Techniques for AR/VR Contents Generation",online, http://vclab.gist.ac.kr/,Dec 2017.

AUTHORS PROFILE

Rameez Muneer Shamalik, received his B. E. & M. E. degree from the Department of Electronics and Telecommunications Engineering of Savitribai Phule Pune University in 2008 & 2013 respectively. He has about ten years' experience in teaching including research. He is currently pursuing Ph. D. with the department of Electronics and Telecommunication Engineering, G. H. Raisoni College of Engineering and Management, Pune India. He works as assistant professor in Bharati Vidyapeeth's College of Engineering for Women, Pune India. His areas of interest are Signal processing and Wireless devices and communication**..**

Dr. S. M. Koli has received his B.E. degree from the Department of Electronics Engineering from Shivaji University, Kolhapur, Maharashtra, India in 1997, M. Tech. degree from the Department of Electronics and Telecommunication Engineering from Dr. Babasaheb Ambedkar Technological University, Lonere, Maharashtra, India in 2006 and Ph.D. degree in the Department of Engineering and Technology, from Sant Gadage Baba Amravati University, Amravati, Maharashtra, India in 2015. He is currently working as Professor in E&TC at the Department of Electronics and Telecommunication Engineering at Dr. D. Y. Patil School of Engineering, Pune, India and research supervisor at G. H. Raisoni college of Engineering and Management, Pune, India . His research interests are mainly in the wireless domain and video signal processing.