Stereo Vision Coding for 3D Imaging

C.Surenthiran,M.E.Communication Systems,K.L.N.Collegeofengineering,Pottapalayamm
Surenthiran.active@gmail.com.
P. Karpagavalli, M.E., (PhD), Associate Professor/ ECE, Associate Professor/ ECE
K.L.N. College of Engineering, Pottapalayam. karpagavallimohan@gmail.com
Dr. A. V. Ramprasad, M.E., Ph.D., Principal, K.L.N. College of Engineering, Pottapalayam
avramprasad2002@gmail.com.

Abstract*—This paper presents a different technique for depth-image compression and its implications on the quality of multiview video plus depth virtual view rendering for without requiring 3D glasses in 3D cinema and 3D home entertainment, 3D mobile using a novel depth-image coding algorithm that concentrates on the special characteristics of depth images. Existing method has H.264 intra-coding with depth images and JPEG 2000 were used, it provides low quality and large complexity and large bandwidth. The proposed method has a front end system and back end system. Front end system describes two stereo cameras such as left view and right view cameras are located in different angles. The left view is encoded using the MPEG-4 encoder and Right view has interview prediction performed it makes the correlation between two images after the correlation to find the depth/disparity estimation. After the disparity estimation the right view is encoded by depth image compensation after the input sequences are balanced to compensate for lighting conditions and camera differences, the joint disparity and motion regularization is performed on the Variation of picture (VOP) basis. The left and right view encoded images are given to the transmission channel. The output of the transmission channel contains two bit streams, a left bit stream, which can be decoded by a standard MPEG-4 decoder, and right bit stream decoded by depth image compensation. The decoder outputs are given to the view rendering sequence. Back end of the system has to be converting the 2D plus depth format from view rendering sequence. After that we have to connect the auto stereoscopic display. The result of emerging auto stereoscopic multiview displays emits a large number of views to enable 3D viewing for multiple users without requiring 3D glasses. Experimental results show that the described technique improves the resulting quality of compressed depth images reduced when compared to a JPEG-2000 encoder.

IndexTerms*—Multiview video coding, Interview prediction, Disparity estimation, Depth image based rendering.

I. INTRODUCTION

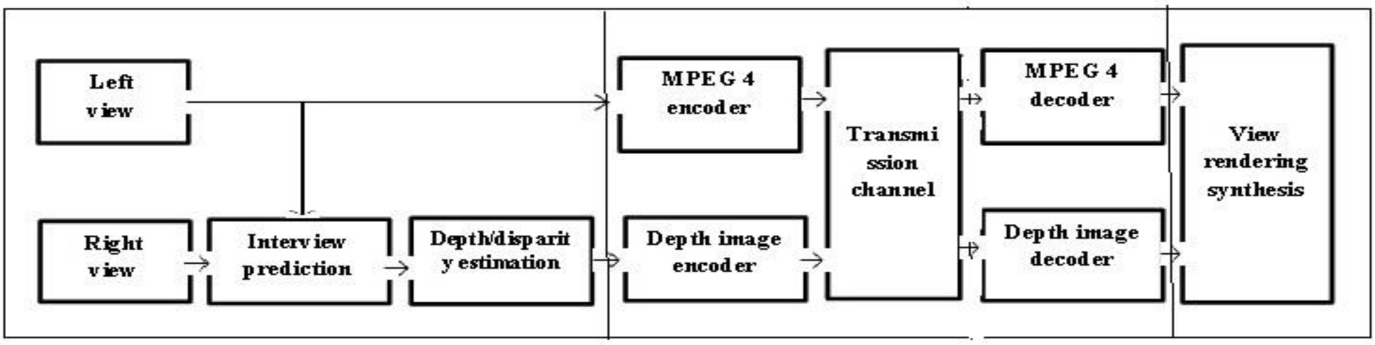
3D video is typically obtained from a set of synchronized cameras, which are capturing the same scene from different viewpoints (multi-view video). This technique enables applications such as free viewpoint video or 3D-TV. Free-viewpoint video applications provide the feature to interactively select a viewpoint of the scene. With 3D-TV, the depth of the scene can be perceived using a multi-view display that shows simultaneously several views of the same scene. Considering the free-viewpoint video application, random access to neighboring views after coding is necessary so that an appropriate coding structure should be adopted. To exploit both spatial (I. e. Interview) and temporal redundancy, it has been proposed [1] to use predefined views as a spatial reference from which neighboring views are predicted.Similarly, only non-central views exploit the spatial interview redundancy. For this reason, by exploiting an appropriate mixture of temporal and spatial prediction, views along the chain of cameras can be randomly accessed. By doing so, we follow recent suggetions [2] in the 3DAV group within MPEG which indicate alternative prediction structures should be investigated.

A first interview prediction technique [3] uses a block-based motion-prediction scheme. Besides compatibility with H.264 coding; a major advantage of this approach is that motion compensation does not rely on the geometry of multiple views, so that camera calibration parameters are not required. However, in the case the baseline distance between cameras is high; it has been reported [3] that a block-based motion-compensation scheme yields a limited codinggain over independent coding of the views. One reason is that the translational motion model employed by the block-based motion compensation scheme is not sufficiently accurate to predict the motion of objects with different depth.

A second, alternative view-prediction scheme [4, 2] is based on a Depth Image Based Rendering algorithm (DIBR). The synthesis algorithm employs a reference texture and depth image as input data. The advantage of the DIBR prediction is that the views can be better predicted even when the baseline distance between the reference and predicted cameras are large, thus yielding a high compression ratio.

Cernigliaroand Jaureguizar, etc. [5], used depth maps to estimatethe likely structure of the motion field for fast Mode decision (MD) in intra-view prediction, and applied the MD results ofneighboring views together with the depth information tomake the results more reliable. But it was complicatedto compute depth maps or expensive to obtain thoseusing depth cameras. Yu and Peng, etc. [6], used global disparityvector (*GDV)* to find the corresponding blocks in otherviews of the current one being encoded, and selectedthe sub-optimal mode according to the modes of thecorresponding blocks for intra-view prediction.

In this paper concentrate the depth map for the intermediate image. And it has a front end system and back end system. Front end system describes two stereo cameras such as left view and right view cameras are located in different angles. The left view is encoded using the MPEG-4 encoder and Right view has interview prediction performed it makes the correlation between two images after the correlation to find the depth/disparity estimation. After the disparity estimation the right view is encoded by depth image compensation after the input sequences are balanced to compensate for lighting conditions and camera differences, the joint disparity and Motion regularization is performed on the Variation of the picture (VOP) basis.The left and right view encoded images are given to the transmission channel. The output of the transmission channel contains two bit streams, a left bit stream, which can be decoded by a standard MPEG-4 decoder, and right bit stream decoded by depth image compensation. The decoder outputs are given to the view rendering sequence.



Multiview images acquisition Coding, transmission and decoding Multiview synthesis

Figure1: Proposed Block Diagram

II.MULTI-VIEW VIDEO CODING (MVC)

A.General MVC System

The MVC system contains the process from the acquisition tothe display of multiple video sequences. Figure 2 showsthe general MVC system.

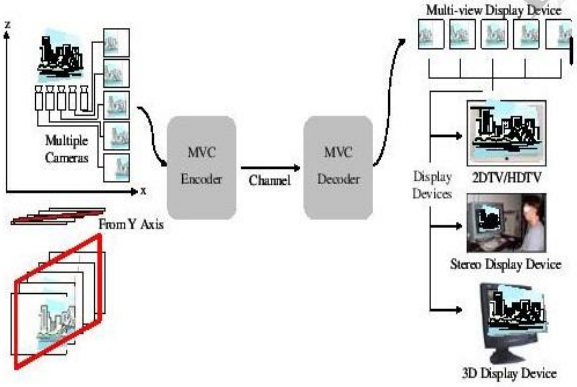


Figure 2: General MVC System

At first, we acquire multi-views, sequences by usingseveral cameras. And then, MVC encoder compresses themulti-view video data. The encoded bit stream istransmitted through the channel. The MVC decoderconverts encoded bit stream to multi-view video sequences.Finally, one display device is chosen by its applicationamong the several devices.

2.3 Interview PredictionStructure

The MVC scheme based on H.264 in [7] is proposed by HHI [8]. To exploit the redundance between differentviews, the methodology of inter-frame prediction in H.264is extended to interview prediction in MVC. Besides; the temporalHierarchical B Picture Prediction Structure (HBPS) proposedin Scalable Video Coding, is also used in MVC to provide temporal scalability.The technologies such as multiple references and variousblock sizes from 16*16 to 4*4 inherited from H.264bring not only high compression performance to MVC,but also high computational complexity, especially in thecontext of the multiplied data of multiview videos

Motion estimation (ME) and mode decision (MD) aretwo main time-consuming processes of video codingas well as MVC. ME can be sped up by fast searchalgorithms or dynamic search range reduction, and MDcan be sped up by technically selecting the most probablemodes partly among which the best one is determined.A number of methods developed to reduce thehigh complexity of ME and MD in single-view videocoding technologies such as H.264 can be directly usedin MVC,

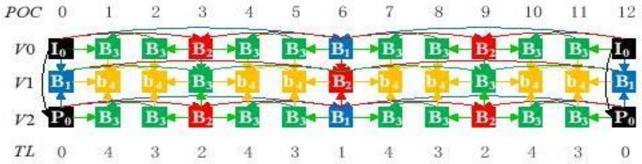


Figure 3: Hierarchical B Picture Prediction Structure for MVC

The four concepts ME, *MV*, DE and *DV* mentioned. For example, in Fig.3, theprocess of searching the best matches in *PV* 1,*POC*0 or*PV* 1,*POC*12 forblocks in *PV* 1,*POC*6 in the same viewis ME, and the vector obtained through ME is *MV*.Accordingly, in different views, if *PV* 0, *POC6* or *PV* 2, *POC6*is the reference pictures, the process is called DE, andthe corresponding vector is *DV*. Especially, the referencepictures in the same view of the current predicted pictureare called *intra-view references* in this paper, and the onesin other views are *interviewing references*.

III.PROPOSED ALGORITHMS

Due to the global disparity, current MVC schemes employ a large search range for view prediction and this makes it difficult to expand the GOP structure for view prediction the proposed algorithm compensates for the global disparity and also expands hierarchical-B picture structure in the spatial prediction. And MPEG-4 video encoder as a replacement for the DCT at boundary-blocks to improve coding efficiency, while retaining backward compatibility. A novel coding algorithm for depth images that concentrates on their special characteristics, namely smooth regions delineated by sharp edges, is compared to H.264 intra-coding with depth images. These two coding techniques are evaluated in the context of multiview video plus depth representations,

A.Fast Disparity and Motion Estimation

a. Search Region Estimation of Disparity for Multiview Video Coding

For multiview video captured by the aligned camera set in which the camera positions are fixed and parallel, as shown in Fig. 4, there exists a strong relationship between the neighboring view videos. Therefore, the disparity between two frames in the neighboring views captured at the same time instance, can be limited to an estimable region.

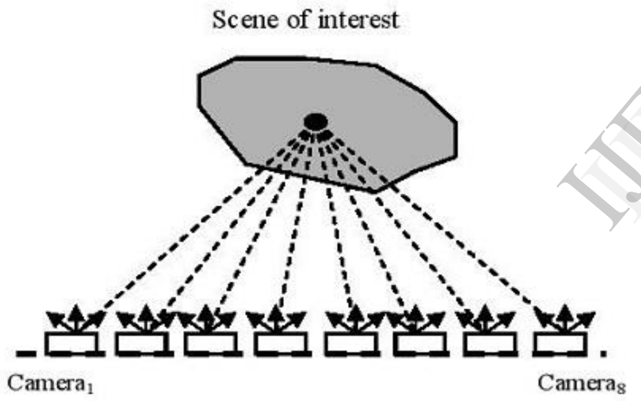


Figure4. Parallel multiview camera setup.

Considering one point *O* projected to the two frames in the neighboring views captured at the same time, as shown in Fig. 5, the positions of pixels *a* and *c* at the screen plane can be denoted as |*ab*| and |*cd*| respectively, where *c* and *a* are the projecting points of *O* captured by the two neighboring cameras. *c*1 and *c*2 represent the positions of the two neighboring cameras. |*OH*1| is the distance between *O* and the screen plane, and |*OH*2| is the distance between *O* andthe camera plane. Then the disparity value of *O* equals to |*cd*|-|*ab*|.

Since the two cameras in Fig. 3 are parallel, we can get that

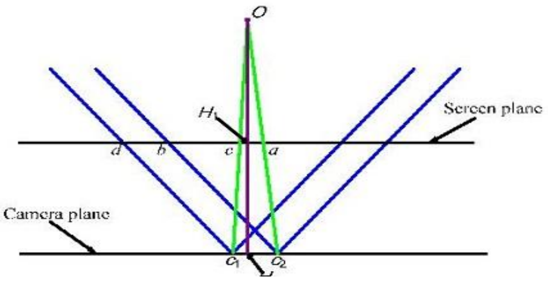


Fig. 5. The relationship between depth and disparity for parallel multiview video.

$$\frac{|ac|}{|c1c2|} = |OH1|/|OH2|$$

$$= (|OH2| - |H1H2|)/|OH2| \qquad (1)$$

And,

$$|bd| = |c1c2| \qquad (2)$$

Then,

$$disparity(O) = |cd| - |ab| = |bd| - |ac|$$

$$= |bd| - |c1c2| \times (|OH2| - |H1H2|)/|OH2|$$

$$= |c1c2| \times |H1H2|/|OH2| \qquad (3)$$

In (3), |*H*1*H*2| is the intrinsic parameter of the camera, and |*c*1*c*2| is determined by the positions of the two neighboring cameras. So the disparity of the object between two frames in the two neighboring views captured at the same time instance is inversely proportional to |*OH*2|, or generally the object'sdepth. From (3), it can be seen that the disparity is near zero when |*OH*2|, or the depth, is big enough. Considering the reference frame is two dimensions, the angle of the disparity vector can be derived from

$$Angle\big(disparity(O)\big) = \arctan\left(\frac{dsparity\ ver(O)}{disparity\ hor(O)}\right)$$

$$= \arctan\left(\frac{|c1c2\ ver| \times \frac{|H1H2|}{|OH2|}}{|c1c2\ Hor| \times \frac{|H1H2|}{|OH2|}}\right) \qquad (4)$$

$$= arctan(|c1c2\ ver|/|c1c2\ Hor|$$

Where |*c*1*c*2Hor| and |c1c2Ver| represent the horizontal and the vertical distance of the two neighboring cameras respectively. It can be seen that the angle of the disparity vector is only determined by the relative positions of the two neighboring cameras.

B.Global Disparity Estimation

Multi-view video coding uses the multi-view video sequences taken by several cameras. So, there exists a disparity called global disparity between adjacent views.

Figure 6: Global Disparity between Exit_0 and Exit_1

Figure 6 shows the global disparity between Exit_0 and Exit_1. Exit_1 looks like the shifted version of Exit_0 by the shaded area.

a. Global Disparity Calculation

To calculate the global disparity, we can employ one of MAD (Mean Absolute Difference) and MSE (Mean Square Error). Eq. (5) and (6) show the equation for global disparity calculation respectively and Fig. 6 shows related parameters.

$$(gx, gy)MAD = \min_{x,y}\left[\frac{1}{R}\sum_{i,j\in R}|img0(i,j) - img1(i-x, j-y)|\right] \qquad (5)$$

$$(gx, gy)MSE = \min_{x,y}\left[\frac{1}{R}\sum_{i,j\in R}|(img0(i,j) - img1(i-x, j-y))^2|\right] \qquad (6)$$

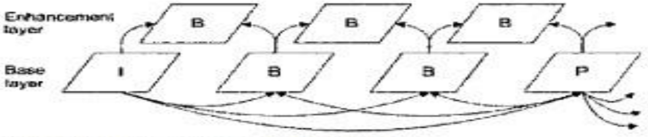
img0 and *img1* in Fig. 7 are two pictures for the global disparity calculation and *R* is the number of pixels in the overlapped area.

Figure 7: Two Reference Frames for Fig. 6

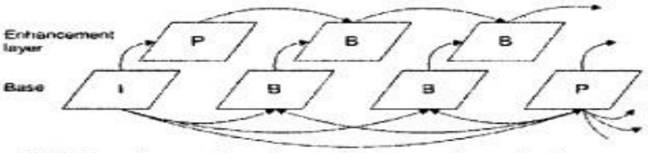
(*gx, gy*) is the displacement vector where the MAD orMSE is minimized and it is chosen as the global disparity of the two views. The global disparity for chrominance components is for 4:2:0 video sequences.

C. MPEG Multiview Profile

The MPEG-2 MVP features a two-layer (base layer and Enhancement layer) video coding scheme. The baselayer video is coded as an MPEG-2 Main Profile (MP) bit stream. The enhancement layer video is coded withtemporal scalability tools and exploits the correlationbetween the two viewing angles to improve thecompression efficiency. There are two configurations in the MPEG-2 MVP. Configuration 1 uses the disparity-compensated prediction. Configuration 2 adopts a disparity-I motion-compensatedprediction scheme, where the enhancement layer uses apicture structure mainly composed of B pictures exceptthat the first picture is a P picture.



(a) Disparity-compensated prediction.



(b) Disparity- and motion- compensated prediction.

Figure 8: Two configurations of the MPEG-2 MVP

One of the twopredictions in enhancement layer B pictures is motioncompensated prediction from decoded immediateprevious enhancement layer picture, and the other Prediction is obtained with respect to immediate previousdecoded base layer picture in display order. Thus, theforward model implies reference with respect to previousdecoded enhancement layer picture (by motion) where asthe backward mode implies reference with respect to thebase layer (by disparity). The prediction structures of thetwo configurations are shown in Figure 8.

a. Proposed Multiview Encoder

The main view isencoded as an MPEG-4 bit stream and the difference isthat the motion vectors for P and B frames in the mainview are provided by the joint disparity and motionestimation module but not by full search blockmatching. The auxiliary view is predicted by jointdisparity and motion compensation from the decodedmain and auxiliary view pictures

In our proposed encoder, user can define the GOPstructure by setting the M (the prediction distance) andN (the intra distance) parameters. The frame structureof I, P and B frames in M PEG provides random access,editability and independently decodability of videosegments [9]. As shown in Figure 9, we retain theframe structure for the main view and introduce newpicture types ID and PD, OD for the auxiliary view,corresponding to disparity-predicted I, P, B pictures. 10pictures are predicted by disparity likes and Po/BD pictures arc predicted jointly by disparity andmotion fields.

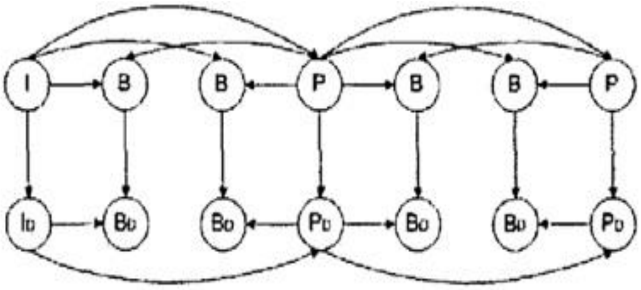


Figure 9: The GOP structure.

D.3D Video Using Depth Maps

We present a novel approach for depth map coding.Here a large number of views for multiview displays is not efficient with videodata only. The efficiency can be drastically increased usingscene geometry information like a depth map. Such atransmission system for 3DV using depth maps is shown inFig. 10. It is assumed that a few cameras, e.g., two or three,are used. The 3DV encoder generates the bit stream, whichcan be decoded at the receiver.

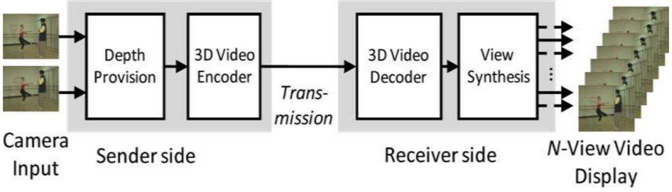


Figure :10 3DV system based on depth-enhanced multiview video

a. Scene Depth Representation

Such a depth-enhanced format for two different viewsis shown in Fig. 11 with color and per-sample depth information.Note that the maximum value for Δs is againlimited for real-world cameras by their aperture angle.

The depth data are usually stored as inverted real-world depth data Id(z), according to

$$\mathrm{Id(z)} = \mathrm{round}\left[255.\left(\frac{1}{\mathrm{z}} - \frac{1}{\mathrm{zmax}}\right) / \left(\frac{1}{\mathrm{zmin}} - \frac{1}{\mathrm{zmax}}\right)\right] \qquad (7)$$

Here, a representation with 8 b/sample and values between0 and 255 is assumed. This method of depth storagehas the following advantages: since depth values are inverted,a high depth resolution of nearby objects isachieved, while farther objects only receive the coarse depth resolution, as shown in Fig. 8. This also aligns with thehuman perception of stereopsis, where a depth impressionis derived from the shift between left and righteye view.

Figure:11Example of depth enhanced format: to view plus a depth format for the Ballet set.

For retrieving the depth values z from the depth maps,the following is applied, which is typically used in synthesisscenarios:

$$z = 1/\left[\frac{Id(z)}{255}.\left(\frac{1}{z} - \frac{1}{zmax}\right) / \left(\frac{1}{zmin} - \frac{1}{zmax}\right)\right] \qquad (8)$$

For this, the original minimum and maximum depthvalue zmin and zmax are required, which have to be signaledwith the 3DV format for a correct geometric displacementin synthesized intermediate views.

b. Depth-Image-Based Rendering

With the provision of per-sample depth data, any number of views within a given range can be synthesized from a few input views. Based on the principles of projective geometry [, arbitrary intermediate views are generated via 3-D projection or 2-D warping from original camera views. This is typically referenced as DIBR [10], [11]. For the presented 3DV solution, the camera views are rectified in a preprocessing step. Thus, the complex process of general DIBR can be simplified to horizontal sample shifting from original into newly rendered views. An example for a fast view generation method with line wise processing and sample shift lookup table can be foundin [12]. The sample shifts are obtained by calculating disparity values d from the stored inversely quantized depth values Id(z)By combining (7) and (8).

$$d = f.\Delta s.\frac{Id(z)}{255}.\left(\frac{1}{z} - \frac{1}{zmax}\right) + \frac{1}{zmax} \qquad (9)$$

Here, the focal length f and camera baseline Δs have to be known. If Δs is given as the spatial distance betweentwo original cameras, d represents the disparity between these cameras and has to be scaled for any intermediateview.

E. Advanced View Rendering Synthesis Methods

For any view synthesis, foreground/background objectboundaries are among the most challenging problems. Asimple projection from original views can cause coronaartifacts, as shown in Fig. 12(a) and (c). The reasons forsuch artifacts are certain effects, like incorrect depth valuesand edge samples, which contain a combination offoreground and background color samples. Also, objectedges may be fuzzy and may contain semitransparent content.Therefore, special treatment in such areas has to beapplied. In advanced synthesis methods, a reliability-basedapproach is taken with one or two boundarylayers. Since areas along depth discontinuities in 3DV areknown to produce visual artifacts in the projection process,they are processed separately.

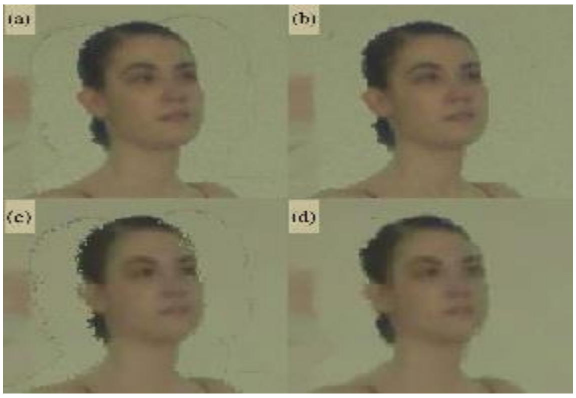


Fig. 12. Example for Comparison of intermediate view quality: (a) and (c) with simple view synthesis and (b) and (d) with reliability-based view synthesis. (a) and (b) using uncompressed data and (c) and (d) using compressed data from the Ballet sequence.

The reliable areas are projected or shifted into the intermediateview first. Then, the unreliable boundary areas aresplit into foreground and background data. Here, foregroundareas are projected next and merged with the reliabledata. Afterwards, the background data are projectedand also merged. The important difference between foreground and background handling is the merging process.The foreground data merge with the reliabledata in a front most sample approach, where the colorsample with the smallest depth value is taken and with thatmost of the important information of the foregroundboundary layer is preserved. In contrast, background informationis only used to fill remaining uncovered areas.Finally, different view enhancement algorithms are applied,including outlier removal, hole filling, and naturaledge smoothing. A more detailed description can be foundin [13]

IV.EXPERIMENTAL RESULT

Here some experiments were conducted; the input images are taken from the two stereo cameras at different angles in the same image. The different view images are such as left view and right view. With these two images we performed the disparities in the intermediate view. By this we improved the time consumption of motion frames between the left and right view images. And also we had been finding the depth map of corresponding the intermediate view image.By this output we can reduce the bandwidth of the image. Here we have used the two data sets such as teddy and lakton stereo images .By using the two dataset we can found the disparity and depth maps the corresponding outputs are shown in the figure 13-17

A.Simulation Result

a.DATASET (teddy)

(a)

(b)

Figure 13: Left(a) and Right(b) view input images

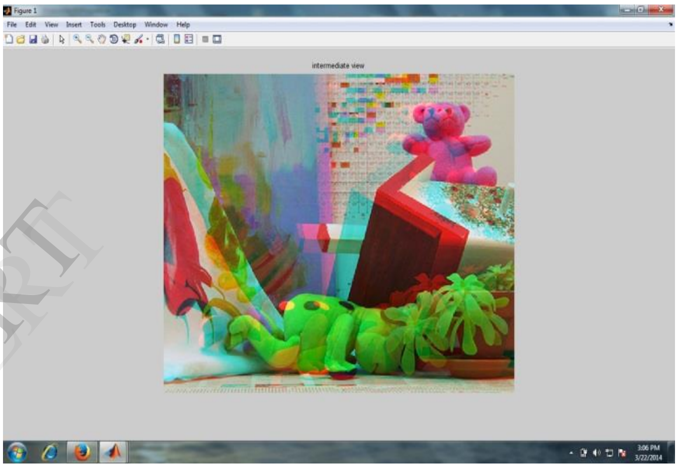


Figure: 14 Intermediate view of left and right images

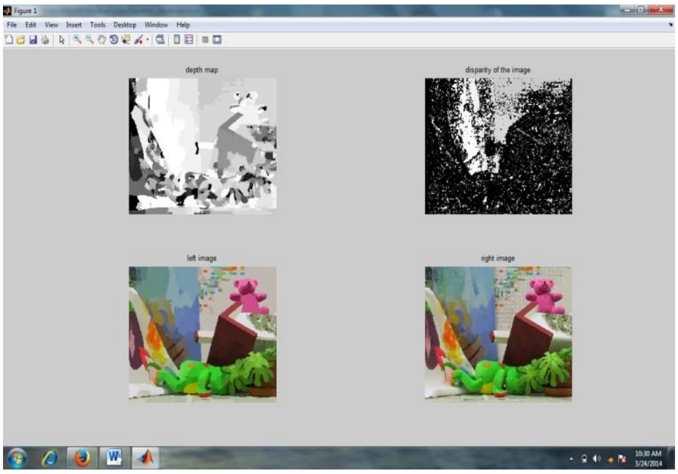
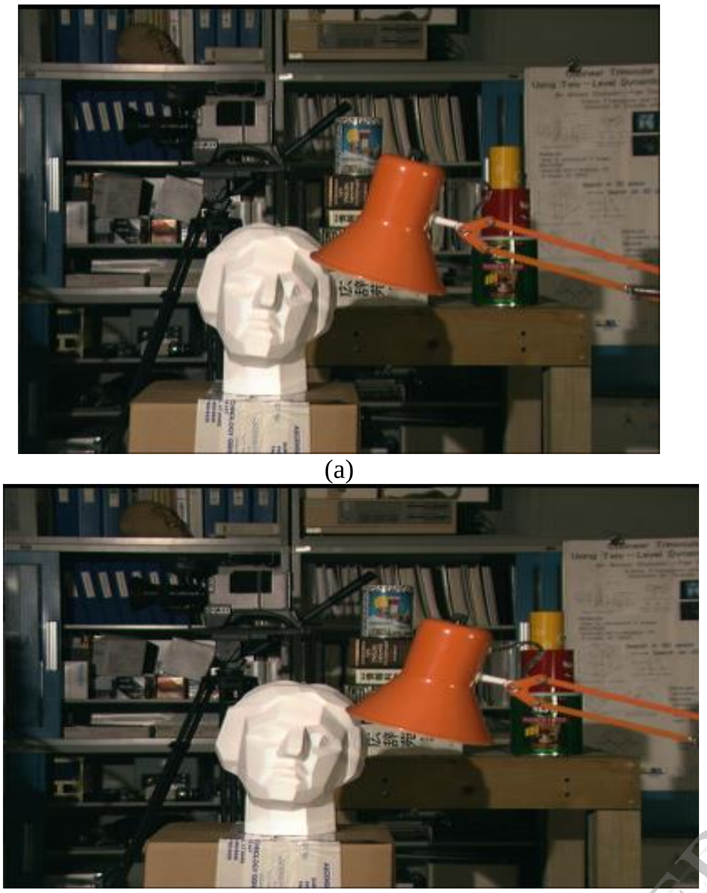


Figure: 15disparity and depth map of corresponding intermediate view *b. DATASET (lakton stereo)*



(a)

(b)

Figure 16: Left(a) and Right(b) view input images

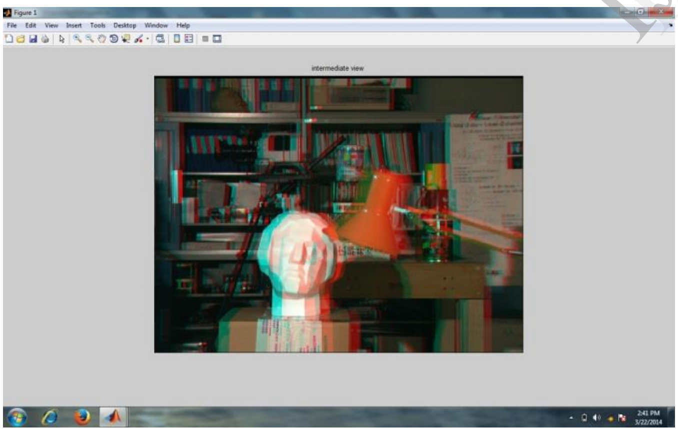


Figure: 17 Intermediate views of left and right images

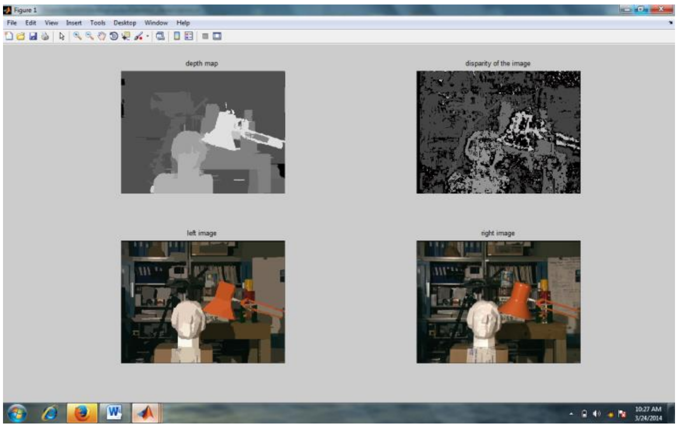


Figure: 18 disparity and depth map of corresponding intermediate view

V.CONCLUSION

In this work disparity estimation and finding the depth map technique is proposed. It is able to process non-rectified input images from uncelebrated stereo cameras and at the same time retain low computational complexity. The hierarchical search scheme is based on the MPEG 4 motion estimation algorithm, initially developed for video coding. The proposed algorithm searches for stereo correspondences inside $D \times D$ search blocks requiring, however, significantly less computations than a typical full search. Future work of the system is to convert the depth map image into 2D plus depth format by using 2D plus depth conversion software then it connect to the auto stereoscopic display we can get 3D output image.

REFERENCE

[1] E. Martinian, A. Behrens, J. Xian, A. Vetro, and H. Sun, “Extensions Of h. 264/avc for native video compression,” in IEEE Int. Conf. on Image Proc., Atlanta, USA, October 2006.

[2] P. Merkle, K. Mueller, A. Smolic, and T.Wiegand, “Efficient compression of multi-view video exploiting inter-view dependencies based on H.264/MPEG4-AVC,” in Int. Conf. on Mult. and Expo, ICME 2006, Toronto, Canada, 2006, vol. 1, pp. 1717–1720.

[3] M. Magnor, P. Ramanathan, and B. Girod, “Multi-view coding for image based rendering using 3-D scene geometry,” IEEE Trans. On CSVT, pp. 1092–1106, November 2003.

[4] G. Cernigliaro, F. Jaureguizar, A. Ortega, J. Cabrera, and N. Garcia, “Fast mode decision for multiview video coding based on depth maps,” in *Visual Communications and Image Processing, Proceedings of SPIE*, San Jose, USA, Jan. 2009.

[5] X. Z. Xu and Y. He, “Fast disparity motion estimation in mvc based on range prediction,” in *Image Processing, IEEE International Conference on*, San Diego, USA, Oct. 2008.

[6] ISO/IEC JTC1/SC29/WG11 M12542, “Multi-view Video Coding based on Lattice-like Pyramid GOP Structure,” October 2005.

[7] G. J. Sullivan, T. Wiegand, and H. Schwarz, “Editors’ draft revision to itu-t rec. h.264 — iso/iec 14496-10 advanced video coding - in preparation for itu-t sg 16 aap consent (in integrated

form),” in *30th Meeting of Joint Video Team (JVT)*, Geneva, Switzerland, Jan. 2009, doc.JVT-AD007.

[8] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, “Efficient prediction structures for multiview video coding,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 11, pp. 1461–1473, 2007.

[9] Luo Van, Zhang Zhaoyang, and An Ping, "Stereo video coding based 011 frame estimation and inter polation", IEEE Trans. on Broadcasting, vo1.49, no.1 , pp. 14-21, Mar. 2003.

[10] P. Kauff, N. Atzpadin, C. Fehn, M. Mu¨ller, O. Schreer, A. Smolic, and R. Tanger, BDepth map creation and image based rendering for advanced 3DTV services providing interoperability and scalability,[Signal Process., Image Commun., Special Issue on 3DTV, vol. 22, no. 2, pp. 217–234, Feb. 2007.

[11] A. Redert, M. O. de Beeck, C. Fehn, W. Ijsselsteijn, M. Pollefeys, L. Van Gool, E. Ofek, I. Sexton, and P. Surman, BATTESTVAdvanced three-dimensional television system techniques,[in Proc. Int. Symp. 3D Data Process. Visual. Transm., Jun. 2002, pp. 313–319.

[12] P. Merkle, Y. Wang, K. Mu¨ller, A. Smolic, and T. Wiegand, BVideo plus depth compression for mobile 3D services,[in Proc. IEEE 3DTV Conf., Potsdam, Germany, May 2009, DOI: 10.1109/3DTV.2009.5069650.

[13] K. Mu¨ller, A. Smolic, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, BView synthesis for advanced 3D video systems,[EURASIP J. Image Video Process., vol. 2008, Special Issue on 3D Image and Video Processing, 2008, article ID 438148, DOI: 10.1155/2008/438148