CODING TECNOLOGIES EMPLOYED IN ROCKETRY FOR EFFICIENT RECEPTION WITH DEPTH

*SoumyaV. S
Research Scholar
TKM College of Engineering Kollam
Soumya.4740@gmail.com

Dr Sheeba O
Professor

***Abstract*-3-D video will become one of the most significant video technologies in the next-generation. Inrocketrybandwidth is an essential requirement. Due to the ultra high data bandwidth requirementfor3-Dvideo, effective ecompression technology becomes an essential part in the infrastructure. Thus multiview video coding (MVC)plays a critical role. MVCis an extended versionof H.264/AVC that improves the performance of multiview videos. The entire image is divided into macro blocks. The size of macroblock depends on codecused. Multi-view video coding (MVC) is an ongoing standard in which variable size disparity estimation (DE) andmotion estimation (ME) are both employed to select the best coding mode for each macroblock (MB). The multi-view video plus depth (MVD)coding will give 3D video (3DV).**

***Index Terms*-3D videocoding (3DVC),multi-view video plus depth(MVD), H.264/AVC,multiview video coding(MVC).**

I. INTRODUCTION

WITH the development of the technologyof3DTV and free view pointTV (FTV), MVC attracts moreand more attention. In recent years ,MVC technologyis now being standardized by the Joint Video Team (JVT) asan extensionto H.264[1].

The sensation of realism can be achieved by visual presentations that are based on three-dimensional (3D) images. To generate even more vivid and realistic information ,it is possible to use two or more cameras placed at slightly different view-points. This allows the production of multiview sequences.

The Multi-view video structure consists of several video sequences, which are captured by closely located cameras in most of the applications. The close location of cameras in these applications results in a high redundancy between the sequences from different cameras.

3D video provides a visual experience with depth percaption through the usage of special displays that re-project athree-dimensionalscenefromslightlydifferentdirectionsfor the leftandrighteye.Suchdisplaysinclude stereoscopicdisplays,whichtypicallyshow the twoviews that were originally recorded bya stereoscopic camera system.Here,glasses-basedsystemsare requiredfor multiuser audiences.Especiallyfor 3D home entertainment, newerstereoscopicdisplayscanvarythe baseline between theviewsto adaptto differentviewingdistances.Inaddition,multi-viewdisplaysareavailable,whichshow not only astereopair,butamultitudeofviews(typically20to more than 50 views) from slightly different directions. Eachuserstillperceivesaviewingpairfortheleftand righteye.However,adifferentstereopairis seenwhenthe viewingpositionisvariedbyasmallamount.Thisdoes notonlyimprovethe3Dviewingexperience,butallows the perceptionof 3D videowithoutglasses,alsoformultiuseraudiences.As 3Dvideocontent is mainly producedas stereovideocontent,appropriatetechnologyisrequired forgenerating the additional views fromthestereodata for thistypeof3D displays.Forthispurpose,different3D video formatsor representationshavebeen considered.

Astraightforwardmethodtoencodethe multi-viewsequencesis simulcastcoding,in whicheachviewisencodedindependentlywiththestate-of-artH.264/AVC codec.ThoughtheH.264/AVCcanachieveaveryhighcod- ing efficiencyforeachsingleview,statisticalresults show that therearestill correlationsleftbetweendifferent views[2].

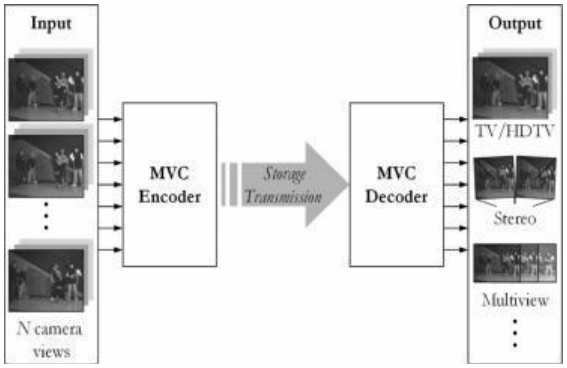

Fig1:Overall structureofanMVCsystem

Stereoscopicvision is basedon theprojectionof an objecton two slightly displacedimageplanesand has anextensiverangeofapplications,suchas3-Dtelevision, 3-D videoapplications,robotvision,virtualmachines, medicalsurgeryand so on. Two picturesofthesamescene takenfromtwonearbypointsformastereopairandcon- tain sufficient information for rendering the captured scenedepth.The above demandingapplicationareasrequirethe developmentofmore efficientcompressiontechniquesofastereoimagepairorastereoimagesequence. Inamonoscopicvideo systemthe compressionis based on the intra-frameandinter-frameredundancy.Typicallythe transmissionor the storageof astereoimagesequencerequirestwice asmuchdata volume asa monoscopic video system.Nevertheless,in astereoscopicsystemamoreefficient coding scheme may be developed if the in- tersequenceredundancy is also exploited.

H.264isthe newestinternationalvideocodingstandard. Comparedtopriorvideocodingstandards,H.264mostly enhancesthecodingefficiency.Soit'smorepossibleto resolvethe problemof stereoscopicstorageandtransmissionusingcodingbasedonH.264.Sincethe multivideo approachcreateslargeamounts ofdatato bestoredor transmittedtotheuser,efficientcompressiontechniques areessentialforrealizingsuchapplications.Thestraightforwardsolution forthis would beto encodeall thevideo signalsindependently using astate-of-the-artvideo codec suchasH.264/AVC[2]–[4]. However,multiviewvideo containsa largeamountof inter-viewstatisticaldependencies,sinceallcamerascapturethe samescenefromdifferentviewpoints. Thesecanbeexploited forcombined temporal/inter-viewprediction, whereimagesarenot only predictedfromtemporally neighboringimagesbut also fromcorrespondingimagesinadjacentviews,referredto asMultiview VideoCoding(MVC).The overallstructure ofMVC definingtheinterfacesis illustratedin Fig.1.

Inthis paper,atypicalstereoscopicvideo compression scenariois mainly studied.Theessential requirementsare describedinSectionII.SectionIIIinvestigatescodingof stereoviews. Theprediction structuresarepresentedin SectionIV.Heretoobtain 3Dviewitrequiresa3-D depthimpressionoftheobservedscenery.SectionVexplainsthe depthcodingapproaches.Finally,SectionVI concludesthis paper.

II. REQUIREMENTS

Thecentralrequirementforanyvideocodingstandard is highcompressionefficiency.Inthespecificcaseof MVC,thismeansasignificant gain compared toindependentcompressionof eachview.Compressioneffi- ciencymeasuresthetradeoffbetweencost(in termsofbit-rate)andbenefit(interms ofvideoquality),i.e.,thequal- ity atacertainbit-rateorthebit-rateatacertainquality. However,compression efficiency is not theonly factorunderconsiderationfor avideocodingstandard.Somerequirementsofavideocodingstandardmayevenbecontradictorysuchascompressionefficiencyand lowdelay in somecases.Thenagoodtradeoffhasto befound.General requirementsforvideo coding suchas minimum resource consumption (memory,processingpower),lowdelay,errorrobustness,orsupportof differentpixelandcolorresolutions, areoften applicableto all videocoding standards.

III.CODINGOF STEREOVIEWS

The maindifferencebetweenclassic videocodingand multiviewvideocodingistheavailability ofmultiplecameraviewsofthesamescene.Ascodingefficiencyof hybridvideocoding dependsonthequality oftheprediction signaltoa greatextent,a codinggaincanbe achievedfor MVCbyadditionalinter-viewprediction.Ifthereisno suchgain,independentlyencodingeachcameraview with temporalpredictionwouldalreadyprovidethe bestpos- sible coding efficiency.

A. Disparity-CompensatedPrediction

DE is utilized to exploit inter-view dependencies in MVC. The distance between two points of a superimposed stereo pair that correspond to the same scene point is called disparity. Disparity compensation is the process that estimates this distance (disparity vector or DV), predicts the right image from the left one and produces their difference or residual image (disparity compensated difference or DCD). Theequation that describes disparity compensation, employing the block matching algorithm (BMA), is:

$$\mathrm{SAD}(\mathrm{x},\mathrm{y},\mathrm{d}) = \sum_{i,j \epsilon w(\mathrm{x},\mathrm{y})} |I_L(i,j) - I_R(i-d,j)|$$

where I_L and I_R are pixel intensity functions of the left and right image, respectively. *W*(*x, y*) is square window that surrounds the position (*x, y*) of the pixel. 'd' is the disparity.

Procedure is that out of the two images of a stereo camera one is chosen as the reference image, and the other image slides across it. As the two images 'slide' over one another we subtract their intensity values. Christo Ananth et al. [3] discussed about Reconstruction of Objects with VSN. By this object reconstruction with feature distribution scheme, efficient processing has to be done on the images received from nodes to reconstruct the image and respond to user query. Object matching methods form the foundation of many state- of-the-art algorithms. Therefore, this feature distribution scheme can be directly applied to several state-of- the-art matching methods with little or no adaptation. The future challenge lies in mapping state-of-the-art matching and reconstruction methods to such a distributed framework. The reconstructed scenes can be converted into a video file format to be displayed as a video, when the user submits the query. This work can be brought into real time by implementing the code on the server side/mobile phone and communicate with several nodes to collect images/objects. This work can be tested in real time with user query results.

The minimum difference value over the frame indicates the best matching pixel, and position of the minimum defines the disparity of the actual pixel. All disparity map pixels were obtained using SAD method along the same epipolar lines of the stereo image. Disparity is estimated after segmentation too.

Quality of 3D disparity map depends on square window size, because a bigger window size corresponds to a greater probability of correct pixel disparity calculated from matched points, although the calculation gets slower.

The result of stereo matching process is a grayscale disparity map that indicates the disparity for every pixel with corresponding intensity. Lighter areas are closer to the camera, darker ones further away. Black areas are points, where disparity was unable to be calculated.

The DE algorithm is summarized as below.

(1) For each pixel in the left image (X) take the pixels in the same row in the right image
(2) Separate the row in right image to windows.
(3) For each window, calculate the disparity for each pixel in that window.
(4) Select the pixel in the window which gives minimum SAD with X
(5) Find the pixel with minimum disparity among all windows as the best match to X.

Although temporal prediction is on average the most efficient mode in MVC system, there are many reasons for using both DE and ME to achieve better predictions than using only ME. One main reason is due to complex motion. In general, the temporal motion cannot be characterized in an adequate way, especially when there is non-rigid motion (such as zooming, rotational motion, and deformations of non-rigid objects) or a motion edge. The region with motion edges is usually predicted using small block sizes with large motion vectors and high residual energy, and thus it has low coding efficiency.

On the other side, usually the disparitywhich is mainly determined based on the relative positions of the objects and cameras is more structured than the temporal motion in complex motion region. MBs in region with complex motion are more likely to choose the inter-view prediction mode. So in this thesis one view is compressed using ME and other using DE.

.

B.*Motion HomogeneityDetermined*

Successive video frames may contain the same objects (still or moving). Motion estimation examines the movement of objects in an image sequence to try to obtain vectors representing the estimated motion. Motion compensation uses the knowledge of object motion so obtained to achieve data compression. In inter frame coding, motion estimation and compensation have become powerful techniques to eliminate the temporal redundancy due to high correlation between consecutive frames.

Motion compensation is an algorithmic technique employed in the encoding of video data for video compression, for example in the generation of MPEG-2 files. Motion compensation describes a picture in terms of the transformation of a reference picture to the current picture. The reference picture may be previous in time or even from the future. When images can be accurately synthesized from previously transmitted/stored images, the compression efficiency can be improved.

Motion estimation is the process of determining motion vectors that describe the transformation from one 2D image to another; usually from adjacent frames in a video sequence. It is an ill-posed problem as the motion is in three dimensions but the images are a projection of the 3D scene onto a 2D plane. The motion vectors may relate to the whole image (global motion estimation) or specific parts, such as rectangular

blocks, arbitrary shaped patches or even per pixel. The motion vectors may be represented by a translational model or many other models that can approximate the motion of a real video camera, such as rotation and translation in all three dimensions and zoom. The motion vector information of a MB is only available after ME is performed. We also know that motion vectors of a MB in one view are strongly related to those of the corresponding MB in the previously coded views.

In real video scenes, motion can be a complex combination of translation and rotation. Such motion is difficult to estimate and may require large amounts of processing. However, translational motion is easily estimated and has been used successfully for motion compensated coding.

Most of the motion estimation algorithms make the following assumptions:

1. Objects move in translation in a plane that is parallel to the camera plane, i.e., the effects of camera zoom, and object rotations are not considered.
2. Illumination is spatially and temporally uniform.
3. Occlusion of one object by another, and uncovered background are neglected.

There are two mainstream techniques of motion estimation: pel-recursive algorithm (PRA) and block-matching algorithm (BMA). PRAs are iterative refining of motion estimation for individual pels by gradient methods. BMAs assume that all the pels within a block has the same motion activity. BMAs estimate motion on the basis of rectangular blocks and produce one motion vector for each block. PRAs involve more computational complexity and less regularity, so they are difficult to realize in hardware. In general, BMAs are more suitable for a simple hardware realization because of their regularity and simplicity. BMA technique is used in this thesis.

Efficient motion estimation reduces the energy in the motion-compensated residual frame and can dramatically improve compression performance. Motion estimation can be very computationally intensive and so this compression performance may be at the expense of high computational complexity. Key performance issues are:

- Coding performance
- Complexity
- Storage and/or delay
- Side information
- Error resilience

The motion estimation creates a model by modifying one or more reference frames to match the current frame as closely as possible.The current frame is *motion compensated* by subtracting the model from the frame to produce a motion compensated residual frame.This is coded and transmitted, along with the information required for the decoder to recreate the model (typically *a set of motion vectors*) in NAL format.

At the same time, the encoded residual is decoded and added to the model to reconstruct a decoded copy of the current frame(which may not be identical to the original frame because of coding losses). Different prediction structures are there such as IBBPBBI…,IBPBI…In this work B structure is not considered and prediction is made by using IPPP…. structure.

Motion estimation is computationally expensive since

- search is done at every pixel position
- over different reference frames

There are several different fast integer search methods out of which full search is implemented.

In a typical full search algorithm ,each frame is divided into Macro blocks. Each MB must be compared with all the Macro Blocks of next frame. Since an MB is 8x8 pixels, there are1728 MBs in one frame for a 288x384 pixel video.

Each Macro block in the present frame is matched against candidate blocks in a search area on the reference frame. These candidate blocks are just the displaced versions of original block. The best (lowest distortion, i.e., most matched) candidate block is found and its displacement (motion vector) is recorded. In a typical inter frame coder, the input frame is subtracted from the prediction of the reference frame. Consequently the motion vector and the resulting error can be transmitted instead of the original block; thus inter frame redundancy is removed and data compression is achieved.

At receiver end, the decoder builds the frame difference signal from the received data and adds it to the reconstructed reference frames. The summation gives an exact replica of the current frame. The better the prediction the smaller the error signal and hence the transmission bit rate.

Advantages are

- Good accuracy.
- Practical for real-time applications such as in medical field.

Disadvantages are

- All pixel points must be examined. So large amount of computations.
- Takes more time.

A Full search (FS) approach algorithm is as follows:

1. Divide current frame into 9x9 Macro blocks (search area);

2. Calculate Block Distortion (BD) by moving the search area along the reference frame.

3. Choose one that gives the minimum distortion as the best matching block and block distance between pixels is the final solution of the motion vector.

By exhaustively testing all the candidate blocks within the search window, full search (FS) algorithm gives the global optimum solution (i.e., the minimum matching error point over the search window) to the motion estimation, while a substantial amount of computational load is demanded. To overcome this drawback, many fast block-matching algorithms (BMA's) have been developed, for example, 2-D logarithmic search (LOGS) , three-step search (TSS) , conjugate direction search (CDS) , cross search (CS) , new three-step search (NTSS) ,four-step search (4SS)etc. These fast BMA's exploit different search patternsand search strategies for finding the optimum motion vector with drasticallyreduced number of search points as compared with the FS algorithm.So in order to avoid this complexity, we should reduce search positions . Fast Block Matching Algorithm Diamond Search can be used.The advantages are fewer search points compared to Full search (FS), Lesser amount of computations,Fastest method., Application is Live match on TV .The disadvantage is it is less accurate for medical applications.The full search method should pre-encode all the pixels in frames, which takes more time.

IV.DEPTH PERCEPTION

People can see depth because they look at the 3D world from two slightly different angles (one from each eye). Our brains then figure out how close things are by determining how far apart they are in the two images from our eyes. The idea here is to do the same thing with a computer. As 2D videos do not normally have sufficient true depth information for stereo conversion, depth (z) of each pixel can be computed from the disparity values according to the relation

$$z = \frac{f*l}{d}$$

where d is the disparity, f is the focal length and l, distance between two cameras. Although depth estimation algorithms have been improved considerably in recent years, they can still be erroneous in some cases due to mismatches, especially for partially occluded image and video content that is only visible in one view.

Another method for depth provision is the use of special sensors, like time-of-flight cameras, which record low resolution depth maps. Here, post processing is required for interpolating depth for each video sample Such sensors currently lack accuracy for larger distances and have to be placed at slightly different positions than the video camera. It is therefore envisioned that in the future a recording device would capture high precision depth together with each color sample directly in the sensor.

It is obtained that the implementation of the stereo matching method based on the SAD algorithm using the segmentation provides much better depth map gives a nice final result, than the implementation of the SAD algorithm without segmentation. Studies were done on different segmentation processes such as hard thresholding, canny edge detection and texture based segmentation. The algorithm for depth estimation is that estimate it from disparity using the equation mentioned above.

V.RESULTS

Simulation is done using Matlab. As we all know it is a technical computing method. In the experiment the stereoscopic right and left view of "tsukuba" is used for encoding. Inter and inter view predictions were made on it and depth is estimated. Depth estimation has been done without segmentation and with segmentation. The SAD algorithm using the segmentation provides much better depth map gives a nice final result, than the implementation of the SAD algorithm without segmentation. The results of inter prediction are shown in figure 4.

Figure 2: Results of Inter Prediction (a) Left Image (b) Right Image (c)ResidualImage (d)Reconstructed Image

The performance of this method is also evaluated. Here prediction is made using 9x9 blocks. From the figure we could understand that residual image took only lesser number of bits for its representation. So by transmitting a residual image compression can be achieved. The reconstructed image after inter prediction preserves quality too.

The result obtained after disparity estimation i.e. disparity map, segmented image and extracted depth are shown in figure.5. Segmented disparity gives better result compared to process without segmentation. Different segmentation processes were done and one which gives best depth map is selected. Texture based segmentation, canny edge detection; hard thresholding, etc were performed. Of these hard thresholding gives better result compared to canny edge and texture based segmentation. The depth map result obtained after hard thresholding is shown below.

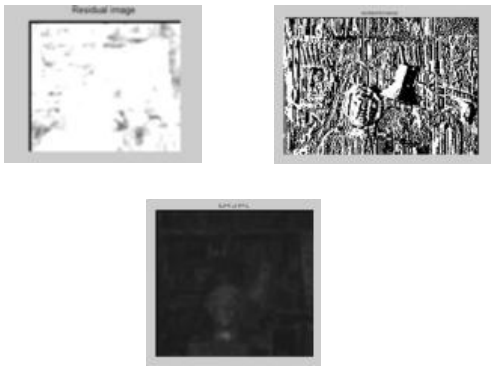

Figure 3: Results obtained are (a) disparity map (b) segmentation map (c) depth map

Evaluation of video coding algorithms is done using peak signal-to-noise-ratio (PSNR) of thesignal is;$\mathrm{PSNR} = 10.\log_{10}\frac{(255)^2}{MSE}$ with MSE being the mean squared error between the original and decoded video samples. The performance measurement done using PSNR is shown in below given table.

Table I Performance evaluation of stereoscopic coding methods

Coding Method	**Resolution**	**PSNR**
H.264/MPEG-4	288x384	32
H.264.MVC (without segmentation)	288x384	96
H.264.MVC (without segmentation)	288x384	82

VI.CONCLUSION

Inrocketry bandwidth is an essential requirement.To achievegoodcoding efficiencyredundancywithina frame and redundancy between views are exploited. Herefull search algorithmisutilizedtoexploitinter-viewdependencies in MVC.

This paper provides a 3 D view of a launched rocket from stereoscopic cameras. In old days 3 D view were obtained from multiple views. For this multiple cameras are required. Since cameras are on board in a rocket the number of cameras must be limited (usually 3). Today the each separation events are viewed in 2D by pacing separate cameras on each plane. This paper presents a new method to obtain the 3D full view of the whole events using two stereo cameras by preserving quality. From these stereo cameras 3D view is obtained on an autostereoscopic display developed by DIMENCO, marketed by Philips. There 2D-plus-Depth converted to 28 different views and interwoven into a stunning 3D format with a field of 150 degrees.

To achieve compression efficiency one of the view is compressed using H.264/MPEG-4 and other using H.264/MVC. The depth map extraction is a challenge nowadays. Here a method is simulated by which depth data can be extracted from disparity using the distance between stereo camera position and the given scene geometry information. It has been found that diamond search is the best algorithm for having ME in this application. In future a modified version of MPEG-4 must be developed by which depth can be extracted during transmission itself.

REFERENCES

[1]ISO/IEC/JTC1/SC29/WG11,"MultiviewCoding Using AVC,"Bangkok,Thailand, Jan. 2006.

[2]U. Fecker,andA. Kaup,"Statistical AnalysisofMulti-ReferenceBlockMatching forDynamicLightField Coding",Proc.10th InternationalFallWorkshopVision, Modeling,andVisualization, pp. 445-452, Erlangen,Germany, Nov.2005.

[3] Christo Ananth, M.Priscilla, B.Nandhini, S.Manju, S.Shafiqa Shalaysha, "Reconstruction of Objects with VSN", International Journal of Advanced Research in Biology, Ecology, Science and Technology (IJARBEST), Vol. 1, Issue 1, April 2015, pp:17-20

[4]T. Wiegand,G. J.Sullivan, G.Bjøntegaard,andA. Luthra,"OverviewoftheH.264/AVC video coding standard,"*IEEETrans. Circuits Syst.VideoTechnol.*, vol. 13, no. 7, pp. 560–560, Jul. 2003.

[5]G. Sullivan andT. Wiegand,"Videocompression—Fromconceptsto theH.264/AVC standard,"*Proc.IEEE, SpecialIssueon AdvancesinVideoCoding and Delivery*, vol. 93, no. 1, p. 18, Jan. 2005.