

3D IMAGERY IN ROCKETRY

Soumya V. S
M.Tech , Dept of ECE
Marian Engineering College

Subha Varier
Scientist/Engineer SG
Indian Space Research Organization (ISRO)
Thiruvananthapuram.

Abstract - 3-D video will become one of the most significant video technologies in the next-generation. In rocketry bandwidth is an essential requirement. Due to the ultra high data bandwidth requirement for 3-D video, effective compression technology becomes an essential part in the infrastructure. Thus multiview video coding (MVC) plays a critical role. MVC is an extended version of H.264/AVC that improves the performance of multiview videos. The entire image is divided into macro blocks. The size of macroblock depends on codec used. Multi-view video coding (MVC) is an ongoing standard in which variable size disparity estimation (DE) and motion estimation (ME) are both employed to select the best coding mode for each macroblock (MB). A multidirectional spatial prediction method is also employed for each macroblock to reduce spatial redundancy. The multi-view video plus depth (MVD) coding will give 3D video (3DV).

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

I. INTRODUCTION

WITH the development of the technology of 3DTV

and free viewpoint TV (FTV), MVC attracts more and more attention. In recent years, MVC technology is now being standardized by the Joint Video Team (JVT) as an extension to 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 perception through the usage of special displays that re-project a three-dimensional scene from slightly different directions for the left and right eye. Such displays include stereoscopic displays, which typically show the two views that were originally recorded by a stereoscopic camera system. Here, glasses-based systems are required for multi-user audiences. Especially for 3D home entertainment, newer stereoscopic displays can vary the baseline between the views to adapt to different viewing distances. In addition, multi-view displays are available, which show not only a stereo pair, but a multitude of views (typically 20 to more than 50 views) from slightly different directions. Each user still perceives a viewing pair for the left and right eye. However, a different stereo pair is seen when the viewing position is varied by a small amount. This does not only improve the 3D viewing experience, but allows the perception of 3D video without glasses, also for multi-user audiences. As 3D video content is mainly produced as stereo video content, appropriate technology is required for generating the additional views from the stereo data for this type of 3D displays. For this purpose, different 3D video formats or representations have been considered.

A straight forward method to encode the multi-view sequences is simulcast coding, in which each view is encoded independently with the state-of-art H.264/AVC co-dec. Though the H.264/AVC can achieve a very high coding efficiency for each single view, statistical results show that there are still correlations left between different views [2].

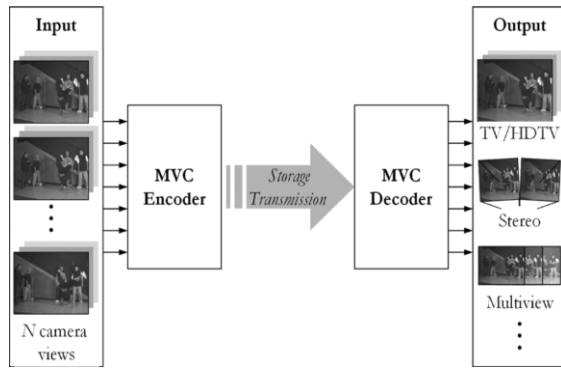


Fig 1: Overall structure of an MVC system

Stereoscopic vision is based on the projection of an object on two slightly displaced image planes and has an extensive range of applications, such as 3-D television, 3-D video applications, robot vision, virtual machines, medical surgery and so on. Two pictures of the same scene taken from two nearby points form a stereo pair and contain sufficient information for rendering the captured scene depth. The above demanding application areas require the development of more efficient compression techniques of a stereo image pair or a stereo image sequence. In a monoscopic video system the compression is based on the intra-frame and inter-frame redundancy. Typically the transmission or the storage of a stereo image sequence requires twice as much data volume as a monoscopic video system. Nevertheless, in a stereoscopic system a more efficient coding scheme may be developed if the inter-sequence redundancy is also exploited.

H.264 is the newest international video coding standard. Compared to prior video coding standards, H.264 mostly enhances the coding efficiency. So it's more possible to resolve the problem of stereoscopic storage and transmission using coding based on H.264. Since the multi video approach creates large amounts of data to be stored or transmitted to the user, efficient compression techniques are essential for realizing such applications. The straight-forward solution for this would be to encode all the video signals independently using a state-of-the-art video codec such as H.264/AVC [2]–[4]. However, multiview video contains a large amount of inter-view statistical dependencies, since all cameras capture the same scene from different viewpoints. These can be exploited for combined temporal/inter-view prediction, where images are not only predicted from temporally neighboring images but also from corresponding images in adjacent views, referred to as Multiview Video Coding (MVC). The overall structure of MVC defining the interfaces is illustrated in Fig. 1.

In this paper, a typical stereoscopic video compression scenario is mainly studied. The essential requirements are described in Section II. Section III investigates coding of

stereo views. The prediction structures are presented in Section IV. Here to obtain 3D view it requires a 3-D depth impression of the observed scenery. Section V explains the depth coding approaches. Finally, Section VI concludes this paper.

II. REQUIREMENTS

The central requirement for any video coding standard is high compression efficiency. In the specific case of MVC, this means a significant gain compared to independent compression of each view. Compression efficiency measures the tradeoff between cost (in terms of bit-rate) and benefit (in terms of video quality), i.e., the quality at a certain bit-rate or the bit-rate at a certain quality. However, compression efficiency is not the only factor under consideration for a video coding standard. Some requirements of a video coding standard may even be contradictory such as compression efficiency and low delay in some cases. Then a good tradeoff has to be found. General requirements for video coding such as minimum resource consumption (memory, processing power), low delay, error robustness, or support of different pixel and color resolutions, are often applicable to all video coding standards.

III. CODING OF STEREO VIEWS

The main difference between classic video coding and multiview video coding is the availability of multiple camera views of the same scene. As coding efficiency of hybrid video coding depends on the quality of the prediction signal to a great extent, a coding gain can be achieved for MVC by additional inter-view prediction. If there is no such gain, independently encoding each camera view with temporal prediction would already provide the best possible coding efficiency.

A. Disparity-Compensated Prediction

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).

As a first coding tool for dependent views, the concept of disparity-compensated prediction (DCP) has been added as an alternative to motion-compensated prediction (MCP). Here, MCP refers to inter-picture prediction that uses already coded pictures of the *same* view at *different* time instance, while DCP refers to inter-picture prediction

that uses already coded pictures of *other* views at the *same* time instance.

B. Motion Homogeneity Determined

A region with homogeneous motion means that the motions in the region have homogenous spatial property, and the corresponding motions in a spatial window are with consistence. A uniform motion vector field at 4x 4 block level can be generated for the calculation of motion homo-geneity in each MB. A Block Matching Algorithm is a way of locating matching blocks in a sequence of digital

video frames for the purposes of motion estimation. The purpose of a block matching algorithm is to find a match-ing block from a frame *i* in some other frame *j*, which may appear before or after *i*. This can be used to discover tem-poral redundancy in the video sequence, increasing the ef-fectiveness of the interframe video compression and tele- vision standards conversion. Block matching algorithms make use of an evaluation metric to determine whether a given block in frame *j* matches the search block in frame *i*.

IV. PREDICTION STRUCTURES

To the fact that current existing prediction structures lack have low coding efficiency a Diagonal Interview Pre-diction (DIP) is presented in this paper, which performs the interview prediction from the reference pictures of dif-ferent time slots to the encoding picture. By introducing the DIP, a MVC prediction structure can support the 3d view of rocketry, while raising the coding efficiency. In comparison, the traditional interview prediction, in which the reference picture of the coding picture, is noted as Normal Interview Prediction (NIP). Figure 2 gives ex-amples of different prediction structures.

Figure 2(a) shows a simple DIP case, in which the en-coding picture is predicted from two reference pictures of the previous time slot, in which one is a temporal refer-ence picture, and another one is an spatial reference picture. Figure 2(b) shows a NIP case, in which the encod-ing picture is then predicted from a temporal reference picture and a spatial prediction reference picture but at the same time slot to the encoding picture. In figure 2(c), the coding picture is predicted from only one temporal refer-ence picture, and views are encoded independently, such a coding structure is called Simulcast coding.

In Figure 2(b) structure, the decoding of the current view has one picture decoding delay compared with the reference view, i.e. the decoding of picture (T,V) has to wait until the decoding of picture (T,V-1) is finished.

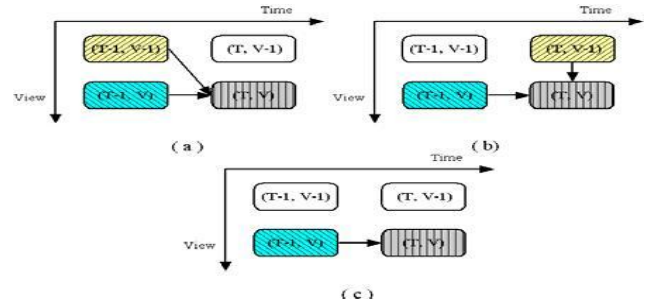


Figure.2 Diagonal Inter-View Prediction Test Mode. (a) The Diagonal inter-view prediction test mode. (b) Nor - mal inter-view prediction test mode. (c) Simulcast test

But for the structure of DIP in Figure 2(a), the two views can be decoded simultaneously, as the DIP reference pictures are always been decoded at the previous time slot. When the number of views becomes very large, the NIP will cause large decoding delay. As a result, the DIP or the Simulcast coding mentioned above is a good structure on the point of decoding delay removing and parallel comput-ing.

Besides the fast algorithm described above, the motion estimation process in the prediction stage can be further speed up based on the motion correlation of different frames. By considering two consecutive frames of same view motion estimation can be done.

V. DEPTH PERCEPTION

In the MVC reference software JMVC, different mode sizes including 16 x 16, 16 x 8, 8 x 16, 8 x 8, 8 x 4, 4 x 8, and 4 x 4 are used in the prediction procedures. Large sizes are usually selected for the macroblocks (MB) in the regions with homogeneous motion, while small sizes are selected for the MBs with complex motion. This technique achieves the highest possible coding efficiency, but results in extremely large encoding time which obstructs it from practical use.

A depth map represents a relative distance from a camera to an object in the 3D space, it can be regarded as a grayscale image using dark and bright values to represent far and close object, and the object depth not only repres-ents the physical object position in 3D space but also in-dicates the motion activity of the object itself on the image plane. Under the condition that cameras are set up in a close parallelized structure, the depth maps are correlated to the texture video motion fields.

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 determ-ining how far apart they are in the two images from our

eyes. The idea here is to do the same thing with a computer. The algorithm is based on "Segment-Based Stereo Matching Using Dissimilarity Measure".

The first step is to get an estimate of the disparity at each pixel in the image. A reference image is chosen, and the other image slides across it. As the two images 'slide' over one another we subtract their intensity values. Additionally, we subtract gradient information (spatial derivatives). We record the offset at which the difference is the smallest, and call that the disparity.

Next we combine image information with the pixel disparities to clean up the disparity map. First, we segment the reference image. Then, for each segment, we look at the associated pixel disparities. Here assign each segment to have the median disparity of all the pixels within that segment. This gives depth.

VI. CONCLUSION

In rocketry bandwidth is an essential requirement. To achieve good coding efficiency redundancy within a frame and redundancy between views are exploited. Here DE is utilized to exploit inter-view dependencies in MVC.

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 motion edge. For the former, the ME based on the translational rigid motion model of blocks fails for zooming, rotational motion and deformation of non-rigid objects, and thus it produces poor prediction results. For the latter, 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 disparity which 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. Thus, the region with homogeneous motion is more likely to select temporal prediction mode where inter-view prediction is not needed, and the region with complex motion is more likely to select inter-view prediction mode. The comparative experimental results show that the proposed algorithm not only significantly reduces the complexity of MVD coding while improves the coding performance, but also maintain the rendering quality.

REFERENCES

- ISO/IEC/JTC1/SC29/WG11, "Multiview Coding Using AVC," Bangkok, Thailand, Jan. 2006.
- [1] U. Fecker, and A. Kaup, "Statistical Analysis of Multi-Reference Block Matching for Dynamic Light Field Coding", Proc. 10th International Fall Workshop Vision, Modeling, and Visualization, pp. 445-452, Erlangen, Germany, Nov. 2005.
- [2] "Advanced Video Coding for Generic Audiovisual Services, Version 3," ITU-T Rec. & ISO/IEC 14496-10 AVC, 2005.
- [3] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Lu-thra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560-560, Jul. 2003.
- [4] G. Sullivan and T. Wiegand, "Video compression—From concepts to the H.264/AVC standard," *Proc. IEEE, Special Issue on Advances in Video Coding and Delivery*, vol. 93, no. 1, p. 18, Jan. 2005.