

# A Survey: Depth Video Coding for 3D Video System

Shivam S. Agrawal , PG Student

Department of E&TC

STES'S, Smt. Kashibai Navale College of Engineering  
Pune, India

Sonal K. Jagtap, Asso. Professor

Department of E&TC

STES'S, Smt. Kashibai Navale College of Engineering  
Pune, India

*Abstract*—With the recent development of 3D multimedia or display technology and the increasing demand for realistic multimedia, 3D video has gained more attention as one of the most dominant video formats with a variety of application such as 3-dimensional TV or Free view point. In 3D video processing, depth video coding is very essential. The most important part of depth video coding is the boundary regions which affects the visual quality of a synthesized view. Here enhance depth video coding is used to estimate a partition function from the geometric structures of available neighboring blocks, and thereby exploit the high spatial correlation of the depth video signal.

*Keywords*— Depth video coding; 3D video system; H.264 Encoder and decoder; Synthesizer.

## I. INTRODUCTION

Now a days a wide research have been carried out in broadcasting technologies such as 3D-Video (3DV) and Ultra-high definition (UHDTV) which can be expanded beyond the capabilities of existing 2D television, Such as system supports a more involving sense of realism and the function of free new point navigation. There are many applications for free viewpoint video but the most popular application is which allows viewers to select an arbitrary viewpoint and direction within a certain range. Another application is 3D television, which provides viewer with the 3D depth of a capture scene by copying a binocular vision system.

Depth data is used in 3D application such as 3D reconstruction and 3D broadcast with the help of depth image rendering and sensor fusion approaches. Practically fast and accurate depth acquisition is unbendable. However, satisfaction of these two demands still images an open issue. For example, laser scanning can obtain relatively accurate depth maps, but it is not suitable for capturing dynamic scenes.

Alternatively, stereo vision can operate fast, but it suffers from inherent ambiguities such as the lack of texture and occlusion in finding correspondences. From this perspective, active range sensors such as time-of-flight (ToF) camera and Microsoft's Kinect have got considerable attention due to its capability of relatively accurate and fast depth acquisition. However, these also suffer from a critical drawback: low-resolution (LR) of the depth map.

The rest of the paper is organized as follows: In next Section II, different techniques used for depth video coding for 3D video coding. Section III describes the human visual system Finally, Section IV concludes the paper.

## II. LITRATURE SURVEY

In 3D video system for broadcasting depth video is very essential. It content boundary information of the depth image and rendering quality of the video. For doing depth video coding there several methods are available which are elaborated as follow.

### A. Video compression: H.264/AVC standard

Over the few years, digital video compression technologies have become an integral part of the way of creating, communicating and consuming visual information. The rate-distortion performance of modern video compression schemes is the result of an interaction between motion representation techniques, waveform coding of differences, intra-picture prediction techniques and waveform coding of various refreshed regions. The paper starts with an basic concepts of video codec design and then explains how these various features have been integrated into international standards and H.264/AVC standards.

### B. Human visual system properties applied to image segmentation for image compression

The authors describe a gray-level image segmentation method for use in segmentation-based image compression. The method consists of basically two steps: a variation of centroid-linkage region growing to perform the initial segmentation of the image and by nonlinear filtering to eliminate visually insignificant image segments. Both steps have advantage of human visual system properties to improve allocation of image segments. Human visual system experiments have been conducted to determine the interactions and optimum balance between the steps. It is shown that this two-step approach produces substantially better-quality segmented images than region growing used alone.

### C. Perception-oriented video coding based on texture analysis and synthesis

Perception-oriented video coding based on texture analysis and synthesis has gained importance over the past half decades. Hence, the present paper overviews a selection of related approaches that has been proposed over the past few years. They are also uses as content-based video coding (CBVC) methods. In the methods of CBVC, an overview on texture analysis and synthesis is also given. The rules are common to a careful selection of CBVC methods are depicted and the requirements of each of the fundamental modules are

extensively discussed in the context of the limitations of state-of-the-art hybrid video codes such as H.264/AVC.

D. Segmentation of textured images using a multi-resolution Gaussian autoregressive model

It presents a new algorithm for segmentation of textured images using a multi resolution Bayesian approach. The new algorithm uses a multi resolution Gaussian autoregressive (MGAR) model for the pyramid representation of the observed image, and assumes a multi scale Markov random field model for the class label pyramid. The models used in this method incorporate correlations between different levels of both the observed image pyramid and the class label pyramid. The technique used for segmentation is the minimization of the expected value of the number of misclassified nodes in the multi resolution lattice. The estimate which satisfies this criterion is referred to as the "multi resolution maximization of the posterior marginal" (MMPM) estimate and "maximization of the posterior marginal" (MPM) estimate. Previous multi resolution segmentation techniques have been based on the maximum a posteriori (MAP) estimation criterion, which is not appropriate for segmentation than the MPM criterion. This gives the number of distinct textures which is known in the observed image. The parameters of the MGAR model including the means, prediction coefficients, and prediction error variances of the different textures are unknown. A modified version of the expectation-maximization (EM) algorithm is used to estimate these parameters. The parameters of the Gibbs distribution for the label pyramid are assumed to be known. Experimental results giving the performance of the algorithm are obtained.

E. Textural feature for image classification

Texture is one of the important characteristics utilized in identification of objects in an image, whether the image is a photomicrograph, an aerial photograph, or a satellite image. This paper summarizes some easily computable textural features based on gray-tone spatial dependencies, and performs their application in category-identification tasks of three different kinds of image data: photomicrographs of five kinds of sandstones, 1:20 000 panchromatic aerial photographs of eight land-use categories, and Earth Resources Technology Satellite (ERTS) multispecialty imagery having seven land-use categories. Two kinds of decision rules are used: one for which the decision regions are convex polyhedra (a piecewise linear decision rule), and another for which the decision regions are rectangular parallelepipeds (a min-max decision rule). In each experiment the data set was divided into two parts, a training set and a test set. Test set identification accuracy is 89 percent for the photomicrographs, 82 percent for the aerial photographic imagery, and 83 percent for the satellite imagery. These obtained results show that the easily computable textural features probably have a general applicability for a broad variety of image-classification applications.

F. Comparison of texture feature based gabor filter

Texture features are compared that are based on the local power spectrum obtained by a bank of Gabor filters. The

features differ in the type of nonlinear post-processing which is applied to the local power spectrum. The following features are considered: Gabor energy, complex moments, and grating cell operator features. The ability of the corresponding operators to produce different feature vector clusters for different textures is compared using two methods: the Fisher (1923) criterion and the classification result comparison. Both methods give robust results. The grating cell operator performs with the best discrimination and segmentation results. The texture detection capabilities of the operators and their robustness to non-texture features are also compared. The grating cell operator is the only one that selectively responds only to texture and does not give false response to non-texture features such as object contours

G. Tracking human motion in structured environments using a distributed camera system

This paper presents a comprehensive framework for tracking finer human models from sequences of synchronized monocular gray scale images in multiple camera coordinates. It establishes the feasibility of an end-to-end person tracking system using a unique combination of motion analysis on 3D geometry in different camera coordinates and other existing techniques in motion detection, segmentation, and pattern recognition. The system initiates with tracking from a single camera view. When the system predicts that the active camera will no longer have a good view of the subject of interest, tracking will be switched to another camera which provides a better view and requires the least switching to continue tracking. The non-rigidity of the human body is addressed by matching points of the middle line of the human image, spatially and temporally, using Bayesian classification schemes. Multivariate normal distributions are employed to model class-conditional densities of the features for tracking, such as location, intensity, and geometric features. Limited degrees of occlusion are tolerated within the system. Experimental results using a prototype system are presented and the performance of the algorithm is evaluated to demonstrate its feasibility for real time applications.

### III. HUMAN VISUAL SYSTEM

A significant property of the human visual system (HVS) is that motion detection is more reliable than colour vision, shape recognition, or other visual stimulus. In particular, the HVS gives priority to track fast and unpredictable objects ("noticeable motion") when compared to slow or predictable motion ("non-noticeable motion") Therefore, for the case of multiple motion types in a scene, noticeable motion attracts more attention.

Besides, assume that for background or non-noticeable motion objects the viewer only perceives the semantic meaning of the displayed objects keeping his/her attention on noticeable objects.

In human visual system extracted frames are used for motion classification in which noticeable and non-noticeable objects are separated out. Noticeable objects include foreground

objects and non-noticeable objects possess background objects.

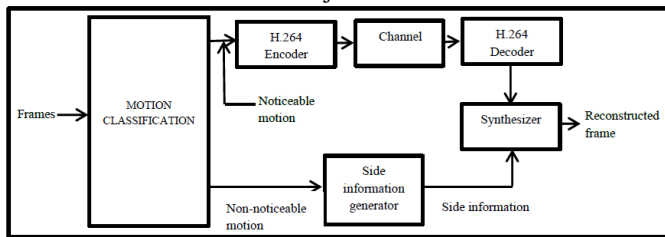


Fig. 1: Methods for human visual system

Later, noticeable objects are encoded which are then transferred to decoder through a channel. The non-noticeable objects are non-encoded but they are kept as side information. Finally, decoded objects and side information are combined in a synthesizer to give output of reconstructed frames.

#### A. foreground-background segmentation

The foreground-background extraction problem aims at estimating the camera movement (global motion) that affects both the moving and stationary points in a scene. By global motion we mean the apparent motion induced by the camera in the video sequence. Extracting the background from a video sequence is an open problem because one needs to take into consideration various issues such as illumination changes, background object displacements or non-static backgrounds. Moreover, the method has to be computationally efficient yet robust. Various effects may degrade the performance including moving object shadows that appear on background objects or non-static background.

#### B. Iterative Foreground Extraction

In the foreground extraction there are the basic four steps which will be extract the foreground object

1. Once the eight parameters are obtained, apply the global motion model, to all central point's belonging to background blocks of the reference frame. Thus, we obtain representing the ideal positions of the central points in the current frame. Using both sets of points we obtain the difference between and form  $\delta_k$ .

2.  $\delta_k$  is compared with a sequence dependent threshold  $D_{th}$ . Macro-blocks with  $\delta_k$  greater than  $D_{th}$  are considered to be foreground, otherwise they are background.

3. Once the new set of background macroblocks is obtained, we remove the isolated macroblocks with a simple morphological closure operation.

4. Finally, obtain an updated version of the elements belonging to  $(X_k, Y_k)$  and consequently to  $(X_k', Y_k')$  that we use to estimate the eight-parameter model again using LS.

## IV. CONCLUSIONS

The comparative study of depth video coding techniques for 3D videos focuses on the human visual system (HVS) with advancement in the high coding efficiency. The HVS method has low data rate giving high perceived visual quality of 3D videos as compared to other existing methods. An area such as slow/global motion objects can be synthesized with acceptable perceptual quality which provides a very reasonable data rate savings.

## REFERENCES

- [1] T. Sikora, "Trends and perspectives in image and video coding," *Proc. IEEE*, vol. 93, no. 1, pp. 6–17, Jan. 2005.
- [2] G. Sullivan and T. Wiegand, "Video compression: From concepts to the H.264/AVC standard," *Proc. IEEE*, vol. 93, no. 1, pp. 61–79, Jan. 2005.
- [3] Katsaggelos, L. Kondi, L. Meier, J. Ostermann, and G. Schuster, "MPEG-4 and rate-distortion-based shape-coding techniques," *Proc. IEEE*, vol. 86, no. 6, pp. 1126–1154, Jun. 1998.
- [4] H. Peterson, "Image segmentation using human visual system properties with applications in image compression," Ph.D. dissertation, Purdue Univ., West Lafayette, IN, May 1990.
- [5] Y. Matsushita, E. Ofek, W. Ge, X. Tang, and H.-Y. Shum, "Full-Frame Video Stabilization with Motion Inpainting," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 7, pp. 1150–1163, July 2006.
- [6] H. A. Peterson, S. Rajala, and E. J. Delp, "Human visual system properties applied to image segmentation for image compression," in *Proc. IEEE Global Telecomm. Conf. (GLOBECOM)*, Phoenix, AZ, Dec. 2–5, 1991, pp. 91–95.
- [7] E. J. Delp, R. L. Kashyap, and O. Mitchell, "Image data compression using autoregressive time series models," *Pattern Recognit.*, vol. 11, pp. 313–323, Jun. 1979.
- [8] P. Ndjiki-Nya, B. Makai, A. Smolic, H. Schwarz, and T. Wiegand, "Improved H.264/AVC coding using texture analysis and synthesis," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Barcelona, Spain, Sep. 2003, pp. 849–852.
- [9] P. Ndjiki-Nya, T. Hinz, C. Stuber, and T. Wiegand, "A content-based video coding approach for rigid and non-rigid textures," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Atlanta, GA, Oct. 2006, pp. 3169–3172.
- [10] P. Ndjiki-Nya, T. Hinz, C. Stuber, and T. Wiegand, "A content-based video coding approach for rigid and non-rigid textures," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Atlanta, GA, Oct. 2006, pp. 3169–3172.
- [11] M. L. Comer and E. J. Delp, "Segmentation of textured images using a multiresolution Gaussian autoregressive model," *IEEE Trans. Image Process.*, vol. 8, no. 3, pp. 408–420, Mar. 1999.
- [12] Z. Farbman and D. Lischinski, "Tonal stabilization of video," *ACM Trans. on Graphics*, vol. 30, no. 4, p. 89, 2011.