Sensible Things that Communicate

PART OF MID SWEDEN UNIVERSITY

Coding of Three-Dimensional Video Content – Diffusion-based Coding of Depth Images and Displacement Intra-coding of Plenoptic contents

Author doctoral thesis: Yun Li

STC Research Centre a part of Mid Sweden University

Abstract:

In recent years, the three-dimensional (3D) movie industry has reaped massive commercial success in the theaters. With the advancement of display technologies, more experienced capturing and generation of 3D contents, TV broadcasting, movies, and games in 3D have entered home entertainment, and it is likely that 3D applications will play an important role in many aspects of people's life in a not distant future. 3D video contents contain at least two views from different perspectives for the left and the right eye of viewers. The amount of coded information is doubled if these views are encoded separately. Moreover, for multi-view displays (i.e. different perspectives of a scene in 3D are presented to the viewer at the same time through different angles), either video streams of all the required views must be transmitted to the receiver, or the displays must synthesize the missing views with a subset of the views. The latter approach has been widely proposed to reduce the amount of data being transmitted. The virtual views can be synthesized by the Depth Image Based Rendering (DIBR) approach from textures and associated depth images. However, it is still the case that the amount of information for the textures plus the depths presents a significant challenge for the network transmission capacity. Compression techniques are vital to facilitate the transmission.

In addition to multi-view and multi-view plus depth for reproducing 3D, light field techniques have recently become a hot topic. The light field capturing aims at acquiring not only spatial but also angular information of a view, and an ideal light field rendering device should be such that the viewers would perceive it as looking through a window. Thus, the light field techniques are a step forward to provide us with a more authentic perception of 3D. Among many light field capturing approaches, focused plenoptic capturing is a solution that utilize microlens arrays. The plenoptic cameras are also portable and commercially available. Multi-view and refocusing can be obtained during post-production from these cameras. However, the captured plenoptic images are of a large size, usually a few times larger than the 1080P High Definition (HD) resolution, and contain significant amount of a redundant information.

An efficient compression of the above mentioned contents will, therefore, increase the availability of content access and provide a better quality experience under the same network capacity constraints.

In this thesis, the compression of depth images and of focused plenoptic contents are addressed. The depth images can be assumed to be piece-wise smooth. Starting from the properties of depth images, a novel depth image model based on edges and sparse samples is presented, which may also be utilized for depth image post-processing. Based on this model, a depth

image coding scheme that explicitly encodes the locations of depth edges is proposed, and the coding scheme has a scalable structure. Furthermore, a compression scheme for block-based 3D-HEVC is also devised, in which diffusion is used for intra prediction. In addition to the proposed schemes, the thesis illustrates several evaluation methodologies, especially the subjective test of the stimulus-comparison method. This is suitable for evaluating the quality of two impaired images, as the objective metrics are inaccurate with respect to synthesized views. For the compression of plenoptic contents, displacement intra prediction with more than one hypothesis is applied and implemented in the HEVC for an efficient prediction. In addition, a scalable coding approach utilizing a sparse set and disparities is introduced for the coding of focused plenoptic images.

The MPEG test sequences were used for the evaluation of the proposed depth image compression, and public available plenoptic image and video contents were applied to the assessment of the proposed plenoptic compression. For depth image coding, the results showed that virtual views synthesized from post-processed depth images by using the proposed model are better than those synthesized from original depth images. More importantly, the proposed coding schemes using such a model produced better synthesized views than the state of the art schemes. For the plenoptic contents, the proposed scheme achieved an efficient prediction and reduced the bit rate significantly while providing coding and rendering scalability. As a result, the outcome of the thesis can lead to improving quality of the 3DTV experience and facilitate the development of 3D applications in general, which can contribute to improving the quality of life.