論文の内容の要旨

論文題目

LIGHT FIELD COMPRESSION AND CONVERSION WITH IMAGE-BASED RENDERING

（自由視点画像合成に基づく光線空間情報の符号化と変換）

氏名　田口　裕一

Recent advances in camera, computer, and display technologies have led researchers to develop 3D TV systems, which provide a more natural and intuitive perception of real scenes than 2D TV systems. Such a system captures multi-view images of a scene by using an array of cameras or lenses, transmits them, and presents a free-viewpoint video on 2D displays or a 3D image on 3D displays by using image-based rendering techniques. A 4D function that represents the light rays included in the multi-view image set is called light field.

This dissertation focuses on handling light field data captured with relatively dense, planar multi-view imaging systems, which are used for reproducing an entire scene rather than only a set of objects, and addresses compression and conversion problems of the light field data. Efficient compression techniques are essential for transmission due to the vast amount of data, typically consisting of tens or hundreds of views. Conversion of light field data is also a core technology of 3D TV systems, because the light field data reproduced by displays is different from the data captured by imaging systems in most cases. Image-based rendering can be considered a basic light field conversion, generating a free-viewpoint image from multi-view images.

In both light field compression and conversion, geometry information of the scene plays an important role, because it provides the correspondence of light rays in the light field; it helps compression methods to improve coding efficiency, while enabling conversion methods to enhance the quality of converted views. The first part of this dissertation therefore addresses dense two-frame stereo matching, a fundamental problem for estimating scene geometry from a set of images. We present an over-segmentation-based stereo method that jointly estimates

segmentation and depth to overcome limitations of traditional segmentation-based stereo methods. For mixed pixels on segment boundaries, the method computes foreground opacity (alpha), as well as color and depth for the foreground and background, which gives a more complete understanding of the scene structure than estimating a single depth value.

The next part explores issues of light field compression. In particular, we focus on compression methods that are suitable for image-based rendering. We first present two compression methods that provide a novel scalability, which we call view-dependent scalability. The scalability enables us to render high-quality views around a significant viewpoint even at low bit rates and to improve the quality of views away from the viewpoint with increasing bit rate. One method performs image-based rendering before the encoding process to generate an image at the significant viewpoint, which is located at the head of the encoded bitstream and acts as a reference image for predicting the input multi-view images. The encoded bitstream can be used with three rendering methods depending on the bit rate. The other method uses region of interest (ROI) coding to provide more flexible control of the view-dependent scalability. It is designed for interactive streaming of free-viewpoint videos to compensate smooth movement of the viewpoint. We then explore how we can exploit inter-view correlation in image-based rendering systems while keeping the computational cost low and the system configuration simple. For this purpose, we use a distributed multi-view coding approach, in which the inter-view correlation is exploited only at the decoder, and propose an efficient method that jointly performs decoding and rendering processes in order to directly synthesize novel images without having to reconstruct all the input images.

The last part describes live 3D TV systems using real-time light field conversion. The system presented first in this part performs real-time video-based rendering using an array of 64 cameras and a single PC. The system estimates a view-dependent per-pixel depth map to render a high-quality novel view. The rendering method is fully implemented on the GPU, which allows the system to efficiently perform capturing and rendering processes as a pipeline by using the CPU and GPU independently. We then show a live end-to-end 3D TV system using the 64-camera array and an integral-photography-based 3D display with 60 viewing directions. We present a fast and flexible conversion method from the 64 multi-camera images to the integral photography format. The conversion method first renders 60 novel images corresponding to the viewing directions of the display by using the above rendering method, and then arranges the rendered pixels to produce an integral photography image. All the conversion processes are performed in real time on the GPU of a single PC. The conversion method also allows us to interactively control rendering parameters for reproducing the dynamic 3D scene with desirable viewing conditions.