

## Depth Map Estimation Techniques for Three-Dimensional Video Processing

Yo-Sung Ho and Sang-Beom Lee

Gwangju Institute of Science and Technology (GIST)  
1 Oryong-dong, Buk-gu, Gwangju 500-712, Korea  
{hoyo, sblee}@gist.ac.kr

Owing to advances in display devices, such as stereoscopic or multi-stereoscopic displays, the three-dimensional television (3DTV) provides users with a feeling of presence from the simulation of reality. In this decade, we expect that the technology will be progressed enough to realize 3DTV including content generation, coding, transmission, and display.

Recently, many researchers have concentrated on acquisition of the depth information. Especially, the depth map is the essential information for three-dimensional video processing including 3DTV [1]. In order to reconstruct intermediate images at virtual viewpoints, we need the depth map. The multiview video can be obtained from multiple cameras directly, whereas the multiview depth map can be estimated by inducing the correspondence between adjacent viewpoints.

As one of the passive 3D depth sensing methods, the stereo matching algorithm is well-known. The key idea of the stereo matching algorithm is finding the corresponding points of pixels between stereoscopic images. Since two images are obtained from slightly different perspectives, the position of a pixel in one view is horizontally displaced in the other view. The amount of displacement allows reasoning about the depth of the scene point.

The stereo matching algorithms are mainly categorized by two groups: pixel-based approach and region-based approach. In many conventional algorithms, the depth information is extracted per pixel. However, the region-based stereo matching algorithms are attractive because of their good performances. These methods assume that a scene has a set of non-overlapping planes in the disparity space and that these planes correspond to at least one homogeneous color segment obtained by image segmentation for the reference image.

Although many conventional methods provide good results, several problems are to be solved yet. Especially, since previous methods perform the depth estimation for each frame separately, we notice that the result has the low temporal consistency. Therefore, we describe how to enhance the temporal consistency of the depth map [2].

### Acknowledgement

This work was supported in part by Information Technology Research Center (ITRC) through Realistic Broadcasting Research Center (RBRC) at GIST (IITA-2008-C1090-0801-0017), and in part by the ETRI FTTH project.

### References

- [1] A. Smolic, K. Müller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, T. Wiegand, "3D Video and Free Viewpoint Video – Technologies, Applications and MPEG Standards," IEEE International Conference on Multimedia and Expo, pp. 2161-2164, July 2006.
- [2] S.-B. Lee and Y.-S. Ho, "Enhancement of Temporal Consistency for Multi-view Depth Map Estimation," ISO/IEC JTC1/SC29/WG11, M15594, July 2008.