

Editorial

Three-Dimensional Image and Video Processing

Peter Eisert,¹ Marc Pollefeys,^{2,3} and Stefano Tubaro⁴

¹ *Fraunhofer Institute for Telecommunications Heinrich-Hertz-Institute, 10587 Berlin, Germany*

² *Department of Computer Science, University of North Carolina at Chapel Hill, NC 27599, USA*

³ *Institute for Computational Science, ETH Zurich, 8092 Zurich, Switzerland*

⁴ *Dipartimento di Elettronica e Informazione, Politecnico di Milano, 20133 Milano, Italy*

Correspondence should be addressed to Stefano Tubaro, stefano.tubaro@polimi.it

Received 31 December 2008; Accepted 31 December 2008

Copyright © 2008 Peter Eisert et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Three-dimensional (3D) TV is expected to be the next revolution in the history of television. It has only recently become feasible to deal with the high processing and bandwidth requirements for real-time acquisition, transmission, and display of high-resolution 3D TV content.

Moreover, capturing 3D scenery by using images and videos taken from different view points is now widely used both in film production for making a seamless fusion of real and virtual scenes and also in industrial applications for quality inspections and goods handling. Also in surveillance applications, the integration of data coming from different cameras are often used to extract 3D information on the imaged scenes/objects/persons.

Although the technical requirements are more and more fulfilled for a commercial success, there is still a need of sophisticated algorithms to handle 3D image information. The amount of data is extremely high, requiring efficient techniques for coding, transmission, and processing. Similarly, the estimation of 3D geometry and other metadata from multiple views, the augmentation of real and synthetic scene content, and the estimation of 3D object or face/body motion has been addressed mainly for restricted scenarios but often lack robustness in a general environment.

This special issue includes six papers dealing with the encoding, visualization, and analysis of 3D scenes. The first five papers address different aspects of a multiview coding system. Multiview coding (MVC) is an extension of the ISO/IEC MPEG4-AVC video coding standard aiming at an efficient encoding of a set of video sequences obtained by viewing the same scene from slightly different viewpoints. An application of MVC is the encoding of video sequences that will be displayed on autostereoscopic multiview displays that are becoming more and more popular and offer a

simultaneous 3D immersive experience to a large group of users. Obviously when we consider the minimum number of views to guarantee the direct perception of 3D information, the multiview coding system reduces to a simple stereo coding system.

The first paper from Peng et al. presents some innovative techniques that can be used for inter- and intraview motion estimations in MPEG4-MVC. The transmission of multiple coded views that are used to feed modern multiview autostereoscopic displays is very demanding due to the fact that usually a large number of video streams is required to guarantee optimal viewing quality from all directions. Therefore, there are several proposals in literature for limiting the number of transmitted video streams. At the receiver side, some video streams can be interpolated from the available ones using disparity or depth information. The second paper by Müller et al. and the third one by Serdar Ince and Janusz Konrad deal exactly with problems connected to view interpolation. The fourth paper of the special issue, authored by Cheng et al., uses a different approach for representing and streaming 3D scenes. They address the encoding and transmission of full 3D scene models. Instead of considering different video streams taken by different view points, explicit object information represented by textured surfaces meshes of the scene are coded. In particular, a proposal for robust (against packet losses on the transmission network) and scalable coding of the object data are presented. The fifth paper by Benoit et al. deals with the problem of quality assessment of 3D-TV contents. Several metrics have been proposed in literature to assess the perceptual quality of two-dimensional images. However, less effort has been devoted to quality assessment of 3D images. The authors review the different issues related to 3D visualization and propose

a quality metric for the assessment of stereo image pairs. The output of the proposed metric is compared with data obtained by subjective tests.

The last paper of the special issue deals with problems more related to the extraction and use of 3D information into surveillance and security systems. The manuscript authored by Wang et al. uses a 3D approach by considering invariant appearance models to track multiple humans into video sequences.

To conclude, we would like to thank the authors for their submissions, the reviewers for their constructive comments, and the editorial team of the EURASIP Journal on Image and Video Processing for their effort in the preparation of this special issue.

We hope that this issue will allow the readers to get some new insights in the recent advances on 3D video processing.

*Peter Eisert
Marc Pollefeys
Stefano Tubaro*