

# Chapter 6

## State-of-the Art Motion Estimation in the Context of 3D TV

**Vania V. Estrela**

*Universidade Federal Fluminense, Brazil*

**A. M. Coelho**

*Instituto Federal de Ed., Ciencia e Tecn. do Sudeste de Minas Gerais, Brazil*

### ABSTRACT

*Progress in image sensors and computation power has fueled studies to improve acquisition, processing, and analysis of 3D streams along with 3D scenes/objects reconstruction. The role of motion compensation/motion estimation (MCME) in 3D TV from end-to-end user is investigated in this chapter. Motion vectors (MV's) are closely related to the concept of disparities, and they can help improving dynamic scene acquisition, content creation, 2D to 3D conversion, compression coding, decompression/decoding, scene rendering, error concealment, virtual/augmented reality handling, intelligent content retrieval, and displaying. Although there are different 3D shape extraction methods, this chapter focuses mostly on shape-from-motion (SfM) techniques due to their relevance to 3D TV. SfM extraction can restore 3D shape information from a single camera data.*

### INTRODUCTION

Technological convergence has been prompting changes in 3D image rendering together with communication paradigms. It implies interaction with other areas, such as games, that are designed for both TV and the Internet. Obtaining and creating perspective time varying scenes are essential

for 3D TV growth and involve knowledge from multidisciplinary areas such as image processing, computer graphics (CG), physics, computer vision, game design, and behavioral sciences (Javidi & Okano, 2002). 3D video refers to previously recorded sequences. 3D TV, on the other hand, comprises acquirement, coding, transmission, reception, decoding, error concealment (EC), and reproduction of streaming video.

DOI: 10.4018/978-1-4666-2660-7.ch006

This chapter sheds some light on the importance of motion compensation and motion estimation (MCME) for an end-to-end 3D TV system, since motion information can help dealing with the huge amount of data involved in acquiring, handing out, exploring, modifying, and reconstructing 3D entities present in video streams. Notwithstanding the existence of many methods to render 3D objects, this text is concerned with shape-from-motion (SfM) techniques.

Applying motion vectors (MVs) to an image to create the next image is called motion compensation (MC). This text will use the term “frame” for a scene snapshot at a given time instant regardless of the fact that it is 2D or 3D video. Motion estimation (ME) explores previous and/or future frames to identify unchanged blocks. The combination of ME and MC is a key part of video compression as used by MPEG 1, 2 and 4 in addition to many other video codecs.

Human beings get 3D data from several cues via parallax. In binocular parallax, each eye captures its view of the same object. In motion parallax, different views of an object are obtained as a consequence of head shift. Multi-view video (MVV) refers to a set of  $N$  temporally synchronized video streams coming from cameras that capture the same real world scenery from different viewpoints and it is widely used in various 3D TV and free-viewpoint video (FVV) systems where stereo ( $N = 2$ ) is a special case. Some issues regarding 3D TV that need further developments to turn this technology mainstream are

- Availability of a broad range of 3D content;
- Suitable distribution mechanisms;
- Adequate transmission strategies;
- Satisfactory computer processing capacity;
- Appropriate displays;
- Proper technology prices for customers; and
- 2D to 3D conversion allowing for popular video material to be seen on a 3D display.

Video tracking is an aid to film post-production, surveillance and estimation of spatial coordinates. Information is gathered by a camera, combined with the result from the analysis of a large set of 2D trajectories of prominent image features and it is used to animate virtual characters from the tracked motion of real characters. The majority of motion-capture systems rely on a set of markers affixed to an actor’s body to approximate their displacements. Next, the motion of the markers is mapped onto characters generated by CG (Deng, Jiang, Liu, & Wang, 2008).

3D video can be generated from a 2D sequence and its related depth map by means of depth image-based rendering (DIBR). As a result, the conversion of 2D to 3D video is feasible if the depth information can be inferred from the original 2D sequence (Fehn, 2006; Kauff, et al., 2007).

Light field refers to radiance as a function of location and direction of areas without occlusion. The final objective is to get a time-varying light field that hits a surface and is reflected with negligible delay. An example of a feasible dense light field acquisition system is one using optical fibers with a high-definition camera to obtain multiple views promptly (Javidi & Okano, 2002). For the most part, light field cameras allow interactive navigation and manipulation of video and per-pixel depth maps to improve the results of light field rendering.

Video restoration is crucial to 3D TV and it continues to call for research because of difficulties such as nonstationarities and discontinuities in spatio-temporal signals as well as the need to minimize the error between the unknown original sequence and the restored one. Denoising can benefit from ME to manage large displacements due to camera motion.

At least two video streams captured by different cameras synchronized in time are necessary to show 3D content. A 3D TV system should offer an adequate amount of viewpoints and this trait is called FVV by the MPEG Ad-Hoc Group on 3D Audio and Video (3DAV) (Smolic & McCutchen,

24 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/state-art-motion-estimation-context/73138](http://www.igi-global.com/chapter/state-art-motion-estimation-context/73138)

## Related Content

---

### Protecting Data Confidentiality in the Cloud of Things

Bashar Alohalid and Vassilios G. Vassilakis (2017). *International Journal of Hyperconnectivity and the Internet of Things* (pp. 29-46).

[www.irma-international.org/article/protecting-data-confidentiality-in-the-cloud-of-things/179896](http://www.irma-international.org/article/protecting-data-confidentiality-in-the-cloud-of-things/179896)

### OTDM-WDM System Components Modeling

(2015). *Optical Transmission and Networks for Next Generation Internet Traffic Highways* (pp. 197-244).

[www.irma-international.org/chapter/otdm-wdm-system-components-modeling/117819](http://www.irma-international.org/chapter/otdm-wdm-system-components-modeling/117819)

### The Exploration of Government as a Service Through Community Cloud Computing

Vasileios Yfantis and Klimis Ntalianis (2020). *International Journal of Hyperconnectivity and the Internet of Things* (pp. 58-67).

[www.irma-international.org/article/the-exploration-of-government-as-a-service-through-community-cloud-computing/258104](http://www.irma-international.org/article/the-exploration-of-government-as-a-service-through-community-cloud-computing/258104)

### An Architecture for Big IoT Data Analytics in the Oil and Gas Industry

Ramiz M. Aliguliyev, Rashid G. Alakbarov and Shalala F. Tahirzada (2020). *International Journal of Hyperconnectivity and the Internet of Things* (pp. 25-37).

[www.irma-international.org/article/an-architecture-for-big-iot-data-analytics-in-the-oil-and-gas-industry/258102](http://www.irma-international.org/article/an-architecture-for-big-iot-data-analytics-in-the-oil-and-gas-industry/258102)

### A Concept for Mobile Ad-Hoc Messaging Networks (AMNet)

Stefan Stieglitz, Christoph Fuchß, Sascha Stöckel and Christoph Lattemann (2010). *Networking and Telecommunications: Concepts, Methodologies, Tools, and Applications* (pp. 1569-1580).

[www.irma-international.org/chapter/concept-mobile-hoc-messaging-networks/49826](http://www.irma-international.org/chapter/concept-mobile-hoc-messaging-networks/49826)