

Chapter 16

Low-Complexity Stereo Matching and Viewpoint Interpolation in Embedded Consumer Applications

Lu Zhang
IMEC, Belgium

Ke Zhang
IMEC, Belgium

Jiangbo Lu
Advanced Digital Sciences Center, Singapore

Tian-Sheuan Chang
National Chiao-Tung University, Taiwan

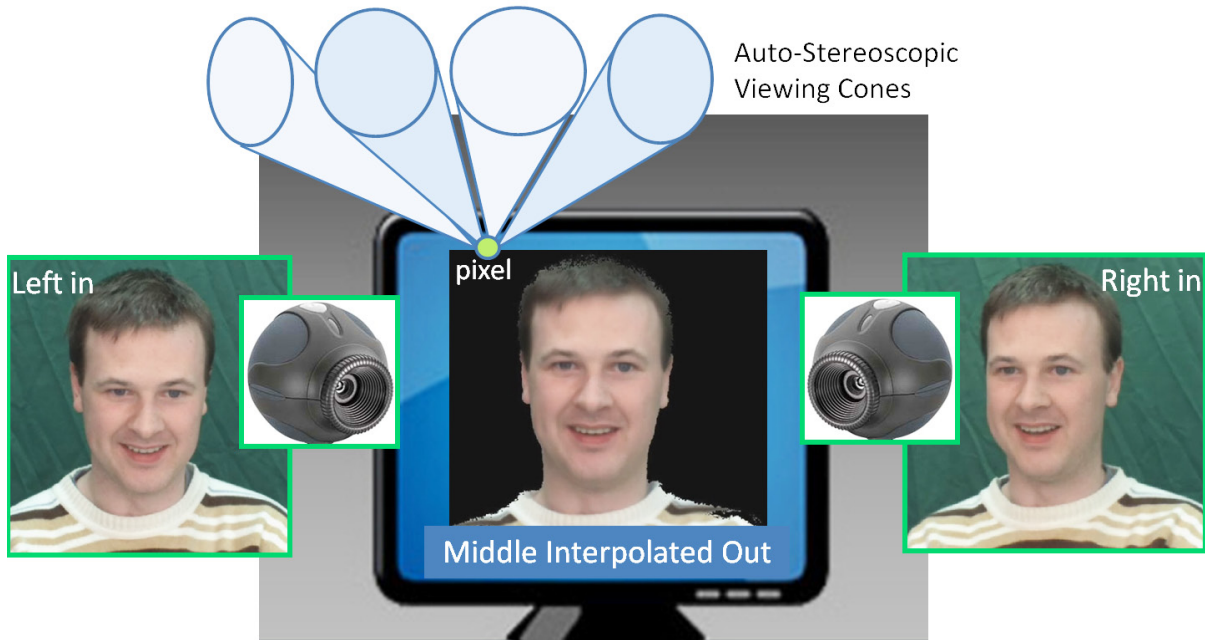
Gauthier Lafruit
IMEC, Belgium

ABSTRACT

Viewpoint interpolation is the process of synthesizing plausible in-between views - so-called virtual camera views - from a couple of surrounding fixed camera views. To make viewpoint interpolation possible for low/moderate-power consumer applications, a further quality/complexity trade-off study is required to conciliate algorithmic quality to architectural performance. In essence, the inter-dependencies between the different algorithmic steps in the processing chain are thoroughly analyzed, aiming at an overall quality-performance model that pinpoints which algorithmic functionalities can be simplified with minor global input-output quality degradation, while maximally reducing their implementation complexity w.r.t. arithmetic and line buffer requirements. Compared to state-of-the-art CPU and GPU platforms running at several GHz clock speed, our low-power 100 MHz FPGA implementation achieves speedups with one to two orders of magnitude, without impeding on the visual quality, reaching over 100 frames per second VGA high-quality, 64-disparity search range stereo matching and enabling viewpoint interpolation in low-power, embedded applications.

DOI: 10.4018/978-1-61350-326-3.ch016

Figure 1. Interpolation of Left/Right camera views into a rendered virtual viewpoint for eye-gaze correction in video teleconferencing (bottom), possibly augmented with auto-stereoscopic 3D displays where each pixel projects multi-directional viewing cones from which two are captured by the viewer's eyes (top)



INTRODUCTION

Figure 1 shows a typical eye-gaze correcting video conferencing application where virtual camera viewpoint interpolation restores straight eye contact to video tele-conference participants by interpolating surrounding views of the user/viewer/participant captured through cameras all around the display. This principle can be extended towards rendering multiple, adjacent, interpolated viewpoints for auto-stereoscopic, shutter-glasses-free 3D displays, where depth impression is obtained by rendering – for each pixel - up to ten different images in different viewing cones (see Figure 1 (top)), two of which being captured by the viewer's eyes. Ultimately, these dozens of views are calculated through viewpoint interpolation from a single pair of cameras, capturing stereoscopic content.

An essential DSP kernel in this process is the extraction of depth from the stereo cameras.

Though we humans do not experience the difficulty of perceiving depth from our binocular view on the outside world, this depth extraction – also called stereo matching – is an incredibly complex processing step that only recently has been ported to embedded platforms (Woodfill, 2004; van der Horst, 2006) at the expense of the quality of the extracted depth image (also called dense depth map) in targeting near-to-real-time performances.

Figure 2 confirms we achieve competitive, real-time processing (over 100 frames per second at VGA resolution, including frame buffer access latency), while preserving high-quality standards, as confirmed by the very low Bad Pixel Error Rate (BPER) reported in Figure 2(d), following the definition of (Scharstein, 2002), i.e. the average difference between calculated and ground truth disparities over all pixels in the image (cfr. Figure 2(b)), using the test images of <http://vision.middlebury.edu/stereo/>. The black arrows refer

22 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/low-complexity-stereo-matching-viewpoint/60272

Related Content

Cognitive Brain Tumour Segmentation Using Varying Window Architecture of Cascade Convolutional Neural Network

Mukesh Kumar Chandrakar and Anup Mishra (2021). *International Journal of Computer Vision and Image Processing* (pp. 21-29).

www.irma-international.org/article/cognitive-brain-tumour-segmentation-using-varying-window-architecture-of-cascade-convolutional-neural-network/288383

Relevant Feature Subset Selection from Ensemble of Multiple Feature Extraction Methods for Texture Classification

Bharti Rana, Akanksha Juneja and Ramesh Kumar Agrawal (2015). *International Journal of Computer Vision and Image Processing* (pp. 48-65).

www.irma-international.org/article/relevant-feature-subset-selection-from-ensemble-of-multiple-feature-extraction-methods-for-texture-classification/151507

Deep Convolutional Sequence Approach Towards Real-Time Intelligent Optical Scanning

Mansi Mahendru, Sanjay Kumar Dubey and Divya Gaur (2021). *International Journal of Computer Vision and Image Processing* (pp. 63-76).

www.irma-international.org/article/deep-convolutional-sequence-approach-towards-real-time-intelligent-optical-scanning/288386

Mathematical Foundations Modeled after Neo-Cortex for Discovery and Understanding of Structures in Data

Shubha Kadambe (2013). *Developing and Applying Biologically-Inspired Vision Systems: Interdisciplinary Concepts* (pp. 224-241).

www.irma-international.org/chapter/mathematical-foundations-modeled-after-neo/72031

Accelerating Deep Action Recognition Networks for Real-Time Applications

David Ivorra-Piqueres, John Alejandro Castro Vargas and Pablo Martinez-Gonzalez (2019). *International Journal of Computer Vision and Image Processing* (pp. 16-31).

www.irma-international.org/article/accelerating-deep-action-recognition-networks-for-real-time-applications/226242