J Real-Time Image Proc
DOI 10.1007/s11554-013-0338-1

ORIGINAL RESEARCH PAPER

Scalable hardware architecture for disparity map computation and object location in real-time

Pedro Miguel Santos · João Canas Ferreira · José Silva Matos

Received: 31 July 2012 / Accepted: 12 March 2013

Abstract We present the disparity map computation core of a hardware system for isolating foreground objects in stereoscopic video streams. The operation is based on the computation of dense disparity maps using block-matching algorithms and two well-known metrics: sum of absolute differences and Census transform. Two sets of disparity maps are computed by taking each of the images as reference so that a consistency check can be performed to identify occluded pixels and eliminate spurious foreground pixels. Taking advantage of parallelism, the proposed architecture is highly scalable and provides numerous degrees of adjustment to different application needs, performance levels and resource usage. A version of the system for 640 × 480 images and a maximum disparity of 135 pixels was implemented in a system based on a Xilinx Virtex II-Pro FPGA and two cameras with a frame rate of 25 fps (less than the maximum supported frame rate of 40 fps on this platform). Implementation of the same system on a Virtex-5 FPGA is estimated to achieve 80 fps, while a version with increased parallelism is estimated to run at 140 fps (which corresponds to the calculation of more than 5.9×10^9 disparity-pixels per second).

P. M. Santos (✉)
Faculdade de Engenharia, Universidade do Porto, Porto, Portugal
e-mail: pedro.miguel.santos@fe.up.pt

J. C. Ferreira · J. S. Matos
INESC TEC and Faculdade de Engenharia,
Universidade do Porto, R. Dr. Roberto Frias,
4200-465 Porto, Portugal
e-mail: jcf@fe.up.pt

J. S. Matos
e-mail: jsm@fe.up.pt

Keywords Dense disparity map · Reconfigurable embedded system · Real-time image processing

1 Introduction

Many applications require an efficient way to compute the distance of objects in a scene to a camera or image sensor, as is the case of automobile crash-avoidance systems [18], or human-computer interface systems such as the one described in [9, 11], which inspired the present work. In that system, the user's hand position is detected and its coordinates sent to a computer so that the hand itself serves as pointing device, much like a computer mouse.

A typical stereoscopic setup uses two side-by-side cameras for capturing images of a scene from two slightly different viewpoints. The position of a given object will exhibit a relative displacement in the two images, which is inversely proportional to the object's closeness to the cameras. This displacement is called disparity: objects with a large disparity are close to the cameras, while those with small disparity are farther away. Image processing techniques, such as block-matching algorithms, can be applied to stereo image pairs to compute the distance of all points in a scene to the cameras, thus generating dense disparity maps from which depth maps can be calculated. Disparity maps can be used as a means to obtain a segmentation of the scene into objects, by aggregating pixels that have similar disparities (or in other words, points at similar distance from the cameras). Identification of foreground objects is a direct application of this technique.

This work presents the implementation of a system for isolating foreground objects, which is capable of processing pairs of 640 × 480 images (8-bit pixels) at a frame rate of 40 frames per seconds (fps), detecting a