

Low-cost FPGA stereo vision system for real time disparity maps calculation

Paolo Zicari, Stefania Perri, Pasquale Corsonello*, Giuseppe Cocorullo

Department of Electronics Computer Science and Systems, University of Calabria, Rende, Italy

ARTICLE INFO

Article history:

Available online 27 February 2012

Keywords:

Stereo vision
FPGA
Low-cost architecture
VLSI implementation

ABSTRACT

Several applications demand efficient hardware implementations of stereo vision systems in order to furnish real time three-dimensional measurements. This paper proposes a complete fast low-cost stereo vision system that performs stereo image rectification with tangential and radial distortion removal, computes dense disparity maps using the Sum of Absolute Differences as the dissimilarity metric, and, finally, exploits a novel injective consistency check purpose-designed for eliminating unreliable disparity values.

The proposed system has been realized and hardware tested for several images resolutions and disparity ranges. When 1280×720 grayscale images are processed with the disparity range equal to 30, the system allows a frame rate up to 97 fps@89 MHz to be reached. It has been realized on a single low-cost XilinxVirtex-4 XC4VLX60 FPGA chip and it occupies 63 DSPs, 128 BRAMs and 15728 slices.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Real time stereo vision is a very challenging research area involving different application fields like autonomous navigation systems, surveillance systems, people and object tracking. Inspired by the human vision, the distance of a generic point from a stereo camera is measured by calculating the disparity between its projected points into the left and right captured images. Several kinds of software and hardware implementations have been proposed using, respectively, general purpose processors and parallel pipelined circuits [1–11]. Recent examples of efficient implementations of complete stereo vision systems are provided in [3–6,11]. The system proposed in [3] elaborates 640×480 images and achieves a 230 fps frame rate using the Hamming distance of Census transformed images to compute the disparity in a range of 64 pixels. In [4], disparity maps are computed through a local block-based matching technique, implemented on the Cell Broadband Engine (CBE). The CBE by Sony, Toshiba and IBM, is the processor used in the Playstation3(R) console. The Sum of Absolute Differences (SADs) metric was chosen for use in such implementation as it provides a good balance between accuracy and speed, performing better than the sum of squared differences, and having a smaller computational complexity than normalized correlation metrics.

The vision machine presented in [5] operates at more than 20 Hz using a hybrid architecture consisting of one dual-GPU card and one quad-core CPU. A flexible use of GPUs as well as multiple CPU-cores according to the actual structure of the algorithms of the

different sub modules has proved to be an efficient way of lowering processing time and latency with only moderate implementation effort.

In [6], a real-time disparity map computation module is realized exploiting a parallel-pipelined fuzzy inference system. The disparity is computed by SAD and the false correspondences are eliminated by an original fuzzy system. The overall design has been realized on an Altera Stratix III EP3SL340H1152C3.

Finally, in [11] a fast and accurate stereo vision algorithm is proposed for hardware-based systems. The disparity is computed in a range of 60 pixels merging the gradient-based census transform (GCT) performed over 5×5 windows of pixels and the SAD computed over windows of pixels with sizes ranging from 5×5 to 19×19 . The hardware implementation carried out using a Stratix EP1S60 FPGA device for the smallest windows sizes, elaborates 750×400 pixels images at a 60 fps frame rate.

Analyzing stereovision systems existing in literature and previously cited, it can be observed that they were mainly inspired by ever-growing speed performances, thus often disregarding costs. On the contrary, in this paper reducing costs is considered as important as achieving high speed and therefore we propose a low-cost and fast hardware implementation of a complete stereovision system.

The stereovision system presented here is implemented in a single XilinxVirtex-4 XC4VLX60 FPGA chip and reaches very high performances thanks to an accurate hardware design effort which joins the potentialities of the versatile dedicated hardware platform to efficient algorithmic and implementation choices. It is worth noting that, nowadays, the commercial price ratio between XC4VLX200 used in [3], or the Altera Stratix III used in [6], and XC4VLX60 chips used here is about 9:1.

* Corresponding author.

E-mail address: p.corsonello@unical.it (P. Corsonello).

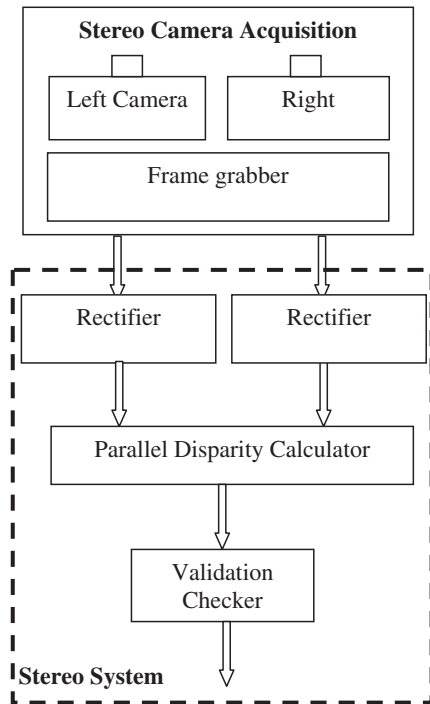


Fig. 1. The stereo system architecture.

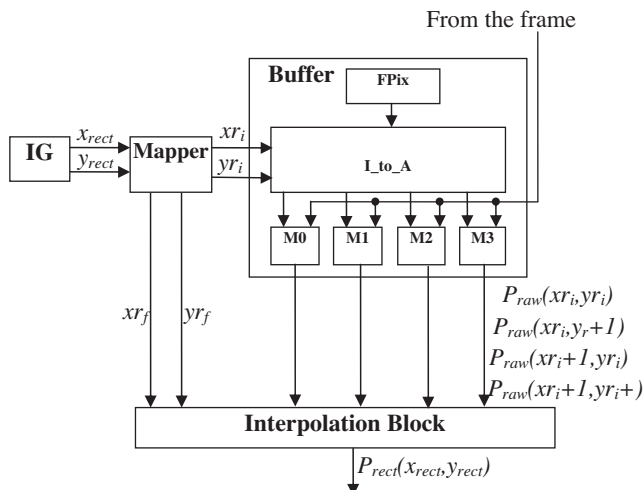


Fig. 2. The rectifier block.

The proposed architecture mainly consists of three stages: the pre-processing, the stereo matching and the post-processing. The pre-processing stage performs the rectification and it is opportunely designed to take into account the effects of the tangential and radial distortions introduced by the camera lenses. A distortion correction [12] (the distortion model considers coefficients up to the 5th order) is applied to the raw images acquired by the stereo camera through intrinsic and extrinsic calibration parameters which are easily calculated by an offline preliminary calibration process according to the Matlab Calibration [13].

In the second stage, disparity values are computed through an SAD-based stereo matching. As deeply discussed and demonstrated in [6], also the post-processing step needed for the consistency check plays a crucial role. With the main objective of agreeing low-cost requirement, as an alternative to the conventional cross-check, a purpose-designed asymmetric injective vali-

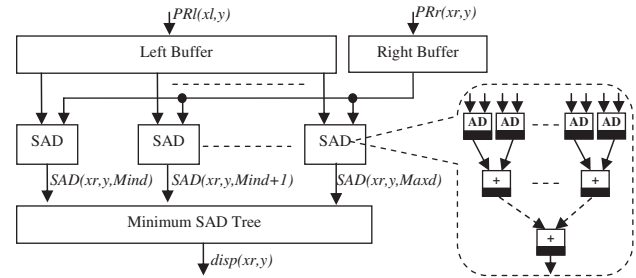


Fig. 3. The parallel disparity computation module.

dation check is implemented. The technique here adopted optimizes the uniqueness check method proposed in [14,15] and requires only a direct matching search to verify that a candidate window is matched by no more than one reference window. In this way, a more area efficient implementation is carried out.

The remainder of the paper is organized as follows: a background of the stereo vision in general and the selected algorithms is reported in Section 2; Section 3 furnishes a description of the implemented hardware stereo architecture; the experimental results are presented and discussed in Section 4; finally, conclusions are given in Section 5.

2. Background

Stereo vision systems create a three-dimensional (3D) reconstruction of a scene by calculating depth information from two $m \times n$ stereo images acquired by a stereo camera. The stereo images are acquired by two distinct cameras placed at a distance b , called baseline, from each other. The preliminary step in the depth reconstruction of any 3D vision system is the camera calibration, an offline operation which calculates the intrinsic and extrinsic parameters representing the model of the used stereo camera [16,17]. Then, in order to compute depth information from the stereo images, the correspondence problem must be solved: for each pixel in the right image its matching point in the left image must be found, or vice versa.

Rectification is usually adopted to simplify the correspondence problem. In fact, it is a projective transformation depending on the calibration parameters that leads to the rectified stereo images, in which corresponding pixels are horizontally aligned [18]. In other words, P being the generic point in the observed scene, the 2D projections of P onto the left and right image, respectively, are the matching pixels $P_{rL}(x_l, y)$ and $P_{rR}(x_r, y)$, with $x_l = 1, \dots, m$, $x_r = 1, \dots, m$ and $y = 1, \dots, n$. Since the matching pixels have the same row index y , searching them becomes a one dimensional problem. By applying the basic principle of triangulation, the disparity, which is the displacement $disp(x_r, y) = x_l - x_r$, and then the distance z of P from the stereo camera can be computed as shown in

$$z(x_r, y) = \frac{b \cdot f}{x_l - x_r} = \frac{b \cdot f}{disp(x_r, y)} \quad (1)$$

Several methods are known in the literature to find matching pixels [19]. Among them, area-based methods compare in terms of some similarity metrics (such as SAD, Cross Correlation (CC), Sum of Squared Differences (SSD), and Hamming distance (HD)) a $W \times W$ reference window centred at the pixel $P_{rR}(x_r, y)$ in the right image to N_c $W \times W$ candidate windows centred at pixels $P_{rL}(x_l, y)$ in the left image, with $x_l = x_r + disp(x_r, y)$, $disp(x_r, y) = Mind, \dots, Maxd$, and $N_c = Maxd - Mind + 1$. The candidate window which is the less dissimilar from the reference window is called the matching candidate window and its central pixel is the matching

Download English Version:

<https://daneshyari.com/en/article/463095>

Download Persian Version:

<https://daneshyari.com/article/463095>

[Daneshyari.com](https://daneshyari.com)