Microprocessors and Microsystems 37 (2013) 1144–1154

Contents lists available at ScienceDirect

Microprocessors and Microsystems

journal homepage: www.elsevier.com/locate/micpro

Efficient and high performance FPGA-based rectification architecture for stereo vision

Paolo Zicari

Department of Electronics, Computer Science and Systems, University of Calabria, Arcavacata di Rende 87036, Rende (CS), Italy

ARTICLE INFO

Article history:
Available online 8 October 2013

Keywords:
Stereo vision
Rectification
Calibration
FPGA

ABSTRACT

This paper proposes an efficient and high performance rectification architecture to be used as a preprocessing module in a complete stereo vision system before the matching correspondence calculus. A complete rectification process is implemented in order to remove the radial and tangential distortion effects due to lenses and to align the left and right raw images acquired by a stereo camera for the epipolar constraining. Thus, the epipolar lines are made collinear with each other and with the image scanning lines in order to reduce the complexity of the matching problem to a one-dimension correspondence search.

The image transformation operations required by the rectification process are computed as matrix calculus through a pipelined and efficient hardware design. Unlike the memory mapped rectification function implementations, the proposed solution does not require any external memory block for the storage of pre-computed rectification maps. Moreover, conforming to the camera model adopted by the Stereo MATLAB Calibration Toolbox which is renowned as the most widely used software toolset for estimating the calibration parameters of a stereo camera, the proposed rectification architecture is a ready-to-use hardware solution to be used in stereo vision real-time embedded systems after calibrating the employed stereo camera following the MATLAB Calibration Toolbox procedure.

When implemented in a Xilinx XC4VLX60-12ff1148 FPGA chip, the proposed circuit rectifies 640×480 and 1280×720 stereo images at a frame rate of 367 fps and 120 fps, respectively. The proposed fully pipelined solution uses an efficient raw image buffer system which is opportunely sized in order to store the minimum number of image rows able to guarantee the synchronization between the image buffering and the rectification elaboration without any interruption of the pipelined processing flow. When the proposed rectification system was used for processing the stereo images acquired by the Point Grey Research Bumblebee BB2-03S2 stereo camera, just 32 BRAM blocks were necessary to implement the raw image buffer; thus, after a latency of 136 us (15,387 clock cycles), a continuous flow of left and right rectified image pixels is guaranteed in output, for each inputted left and right couple of raw image pixels, at each clock cycle.

When compared to the other implementations present in literature, the proposed solution offers the advantage of not using any external memory with respect to the memory-mapped rectification solutions while offering a more efficient and complete solution reaching the highest speed performance with respect to the on-the-fly computed rectification implementations present in literature.

1. Introduction

Stereo vision is a common technique for acquiring 3D information from a stereo camera. The depth of the points within a scene is estimated after the stereo correspondence calculus; the latter solves the matching problem for each couple of left and right stereo images. Stereo vision is widely employed in several application areas like robot navigation, obstacle detection for autonomous vehicles, video surveillance, 3D vision assisted surgery equipments, augmented virtual reality. Real time performances demand for ever more efficient stereo vision solutions. For this reason, several on-purpose-designed hardware based systems have been proposed recently. The latter use Digital Signal Processors (DSP), Field Programmable Arrays (FPGA) and Application Specific Integrated Circuits (ASIC). The DeepSea Stereo Vision System is an ASIC architecture proposed in [1,2] for calculating the disparity of 512×480 stereo images through the Census stereo algorithm. Over the last decades, several FPGA implementations have been proposed, exploiting the fast prototyping and immediate test features of the reconfigurable hardware. In [3,4], an FPGA implemented stereo depth system is presented, based on a multi-resolution, multi-orientation phase technique for processing 256×360

E-mail address: zicari@deis.unical.it

0141-9331/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.micpro.2013.09.007

stereo images at 30 fps. In [5], a Sum of Absolute Differences (SAD) core calculates the disparity map of 320×240 stereo images at 425 fps. In [6], the MSVM33 stereo system processes 320×240 images captured by three cameras in order to calculate the disparity maps at 30 fps. In [7], a real time disparity map computation module is implemented in an Altera Stratix-IV FPGA for the computation of 640×480 disparity maps in a range of 80 pixels; the disparity range is minimized by a verging angle control on the stereo setup cameras. In [8], an FPGA based stereo matching circuit processing 512×512 images calculates the minimum SAD over 5×5 correlation windows in a disparity range of 255 pixels at 25.6 fps. In [9], an IP core is proposed and tested on a Stratix EP1S60 device for implementing a novel stereo matching algorithm based on the Modified Census Transform and the sparse Hamming distance calculus. A complete fully parallel and pipelined FPGA stereo vision system is proposed in [10], the latter calculates the Hamming distance of Census transformed images to compute the disparity of 640×480 images at 230 fps.

The rectification preprocessing stage is fundamental in a stereo vision system in order to reduce the distortion effects due to the camera lenses and to perfectly align the stereo images so that the complexity of the matching problem can be reduced to a one-dimension correspondence search. The calibration of a stereo camera furnishes the intrinsic and the extrinsic parameters to be used by the rectification calculus in order to execute the proper image transformations which can be considered as a mapping function for reconstructing the rectified pixels from the acquired raw image pixels. The hardware stereo vision system implementations presented in [7,8,11–13] do not include any rectification preprocessing stage because their main focus is on the stereo matching problem for the disparity map computation, thus assuming that already rectified stereo images are the input of the stereo vision system. Two different approaches are proposed in literature for the implementation of the rectification in hardware. The rectification function is implemented as a memory mapped function after an offline computation for all the possible coordinates of the rectified image, like in [14–18]. In this case, external memories are necessary to store the big quantity of rectification mapping information, except for [15,18] where a reduction approach is used in order to decrease the memory occupation. Conversely, on-purpose-designed circuits map each rectified image pixel in its corresponding raw image pixels through the execution of matrix transformation operations, like in [19,10,20]. Unfortunately, simplified solutions are often provided in order to reduce the complexity of the calculus. In fact, in [10], a simple inverse matrix transformation was used as a reverse mapping for rectification without taking into account any distortion effects. In [19], the simplified bicubic polynomial function was used to approximate the rectification homograph transformation. Finally, in [20], a matrix calculation block generates the address of the transformed pixel using the homograph matrix taking into account the radial distortion of lenses.

The proposed rectification design is a complete and general solution for rectifying raw images acquired by a stereo camera. No external memory is required for the mapping function because the matrix calculus is computed through a pipelined and efficient hardware circuit. The proposed rectification design is aimed to reduce the radial and tangential distortion effects in the raw stereo images and to align them in order to satisfy the epipolar constraint by using the calibration parameters previously calculated by following the Stereo MATLAB Calibration Toolbox software procedures in [21]. Thus, a ready-to-use rectification module is offered to be included in the preprocessing stage of the complete stereo vision systems after calibrating the employed stereo camera with the MATLAB Calibration Toolbox, and extracting the corresponding calibration parameters. Implemented as a fully parallel and pipelined architecture, the proposed rectification architecture furnishes two rectified left and right pixels for each inputted left and right raw image pixel at each clock cycle in a row order, after a fixed number of latency clock cycles depending on the distortion of the stereo camera.

The remainder of the paper is organized as follows: a background of the stereo vision and the rectification is reported in Section 2; related works are analyzed in Section 3. Section 4 furnishes a description of the proposed approach and architecture; the experimental results are reported in Section 5. Finally, the conclusions are discussed in Section 6.

2. Background

Stereo vision is used for 3D reconstruction of the real scene. The depth information is extracted by opportunely processing the couple of stereo images acquired by two distinct cameras at a distance b, called baseline. The stereo matching problem for the correspondence search is the main task in a stereo vision, furnishing a map of disparity values as a result. The disparity represents the position displacement between the corresponding pixels Ir and Il in the right and left images respectively, which are the projection of the same point P in the real scene. The distance z of the point P from the stereo camera is calculated from the disparity d through triangulation computations taking into consideration the baseline b and the focal length f of the stereo camera as reported in (1).

$$z = \frac{b \cdot f}{d} \tag{1}$$

Several stereo matching algorithms are present in literature for calculating the disparity maps: the local methods compute disparity values based on local information around certain positions of pixels, including area based (SAD, SSD, Census), phase based and the feature based (matching on features like lines, corners and edges) methods. Conversely, the global methods such as graph cuts, belief propagation and dynamic programming are more complex because they attempt to minimize an energy function computed on the whole image area. Quality and performance comparisons are evaluated through benchmark images provided with accurate ground truth disparity maps. Local methods can be implemented very efficiently if the correspondence search is executed along just one dimension of the image area, thus reducing the complexity. The one-dimension search space is possible only if the rectification process is executed on the acquired stereo images before solving the stereo matching in order to satisfy the epipolar constraint. As shown in Fig. 1, the conjunction of the two focal points Fr and Fl of the right and left camera respectively intersects the left and right

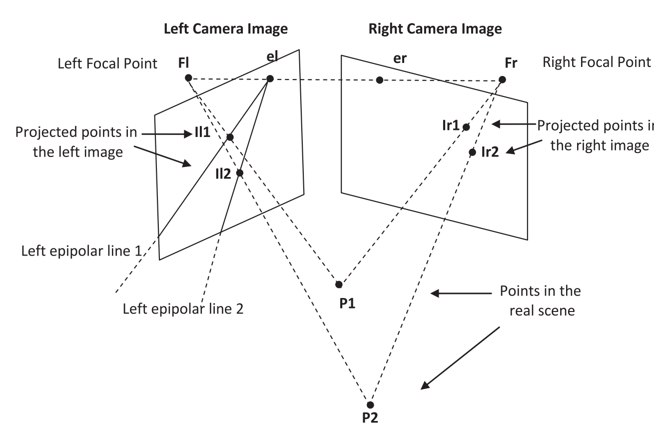

Fig. 1. Epipolar geometry.