INTEGRATION, the VLSI journal 51 (2015) 81–91

Contents lists available at ScienceDirect

INTEGRATION, the VLSI journal

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

A new parallel hardware architecture for high-performance stereo matching calculation

Young-Ho Seo, Ji-Sang Yoo, Dong-Wook Kim *

Kwangwoon University, 447-1, Welgye-1Dong, Nowon-Gu, Seoul 139-701, Republic of Korea

ARTICLE INFO

Article history:
Received 13 December 2013
Received in revised form
26 April 2015
Accepted 11 June 2015
Available online 3 July 2015

Keywords:
Stereo vision
Stereo matching
Architecture
Disparity
FPGA
Memory bandwidth

ABSTRACT

In this paper, we propose a hardware (H/W) architecture to find disparities for stereo matching in real time. After analyzing the arithmetic characteristic of stereo matching, we propose a new calculating method that reuses the intermediate results to minimize the calculation load and memory access. From this, we propose a stereo matching calculation cell and a new H/W architecture. Finally, we propose a new stereo matching processor. The implemented H/W can operate at the clock frequency of 250 MHz at least in the FPGA (field programmable gate array) environment and produce about 120 disparity images per second for HD stereo images.

1. Introduction

Stereo vision is a traditional and efficient method used to obtain 3D space information. Recently stereo vision has been used to support automatic movement of a smart car, intelligent robot, human-computer interface, and multi-viewing of 3D display, etc. The most important requirement of stereo vision is the acquisition of reliable depth information in real time. Depth is defined as the distance between a camera and target objects in 3D space. Depth information is acquired through a basic principle based on the human visual system, which recognizes the distance or depth of an object by analyzing and synthesizing the images captured from both eyes in the cerebrum [1–3]. The slightly different images on the retina are called stereoscopic images.

Recently, the researches about stereo vision have included acquisition of high-quality disparities, object or pattern recognition; stereoscopic effect augmentation, 3D display, compression of stereo images, and performance enhancement of stereo matching, etc. In this paper, we focus on a rapid method for performing stereo matching. Since cost function is calculated as many as the search range for all pixels in a stereo image, stereo matching carries a high computational load [4]. Although there is a difference according to the algorithm, it is difficult to get the performance of 1 fps (frame per second) in the case of S/W (software) implementation [5]. Thus, stereo matching should be developed using H/W for configuring a real-time stereo vision system [6–16].

Various H/W-based approaches have been developed to increase the performance of stereo matching [6–16]. There are two approaches for implementing dedicated chips: ASIC (application specific integrated circuit) and FPGA methods. Kimura et al. [6] designed a convolver-based nine-eye stereo machine called SAZAN. It performs real-time acquisition of a dense depth map at 20 MDPS (mil1ion depth-pixels per second). Kuhn et al. [7] proposed a fast and area-efficient implementation of a real-time stereo vision algorithm for spatial depth mapping. It processes 256×192 pixel input streams at rates higher than 50 fps with disparity range of 25. Darabiha et al. [8] implemented a stereo depth measurement algorithm in H/W on FPGAs, which generates 8-bit sub-pixel disparities on 256×360 pixel images at 30 fps. Jia et al. [9] developed a miniature stereo vision machine (MSVM-III) which, running at 60 MHz, can process dense disparity maps with 640×480 pixels in the 64-pixel disparity search range at rates higher than 30 fps, and with 320×240 pixels at 120 fps. Woodfill et al. [10] developed the DeepSea Stereo Vision System, which can run at 200 fps with 512×480 images. Jin et al. [11] proposed a fully pipelined stereo vision system, which includes an entire stereo vision process in FPGA. It generates depth images at 230 fps from a 640×480 input image for the 52-pixel search range. Lygouras et al. [12] implemented a new real-time stereo system which can process stereo-pairs in full VGA resolution at a rate of 25 Mpixels/s and produces 8-bit dense disparity maps. Cocorullo et al. [13] developed a complete fast low-cost stereo vision system. 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. In [14] a fast and accurate stereo vision algorithm is proposed for

* Corresponding author. Tel.: +82 02 940 8362; fax: +82 02 917 8362.
E-mail address: dwkim@kw.ac.kr (D.-W. Kim).

http://dx.doi.org/10.1016/j.vlsi.2015.06.005
0167-9260/

hardware-based systems. 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. Colodro-Conde et al. [15] presented an evaluation of area-based algorithms used for calculating distance in stereoscopic vision systems on FPGA. Tomasi et al. [16] presented a novel architecture for multiple-vision feature extraction, which reached an outstanding computing power of 92.3 GigaOPS at very low power consumption (approximately 12.9 GigaOPS/W). Recently, GPGPU (general purpose graphic processing unit)-based approaches have been widely developed under rapid advancements of GPGPU H/W and related languages [17,18].

In this paper, we propose a new VLSI (very large scale integrated circuit) architecture to implement a dedicated chip for stereo vision. It is a kind of high-performance processor, which can adaptively execute stereo matching according to the user's request and the applied algorithm. Though the SAD (sum of absolute differences) was used for the cost function, other cost functions, instead of SAD, can be used. After analyzing the stereo matching process and scheduling it by using the analyzed results, we derived a new unit H/W cell. The cell can be easily extended for 1D stereo matching. It has a parallel architecture and can be concurrently operated. If the extended H/W is combined with line buffers, which are composed of FIFO (first input first output), it can efficiently calculate 2D stereo matching with high-performance.

The remainder of this paper is organized as follows. Section 2 describes the stereo matching algorithm, and Section 3 develops the H/W architecture that supports the proposed parallel stereo matching algorithm. Section 4 describes the stereo matching processor, and Section 5 describes the implementation results. Section 6 summarizes the proposed and implemented H/W.

2. Stereo matching

2.1. Basic principle

Stereo matching is a process of identifying correspondences between pixels in a pair of stereo images. Before stereo matching, the stereo images are rectified; that is, the epipolar lines are made collinear and to run in parallel with the *x*-axis, which limits the search space to the corresponding rows of the images. Fig. 1 shows an example of stereo matching, in which a region that is similar between the left and right images is searched in an epipolar line. The rectangular region to be used during the searching process is defined as the matching window. The similarity is identified by using a cost function. After calculating all costs in the search range, the most similar regions are selected and the distance between them is determined. This distance is defined as disparity.

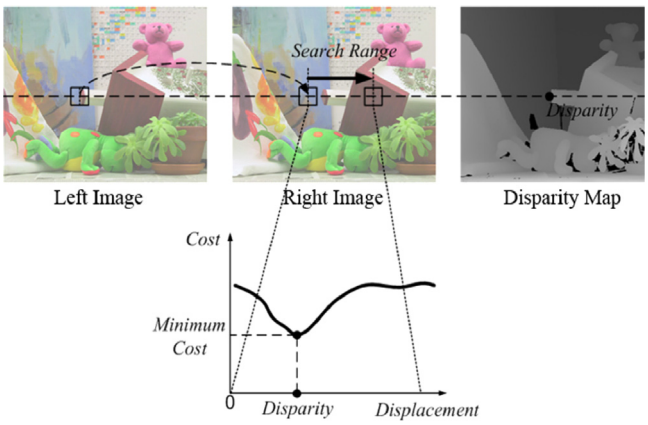

Fig. 1. Similarity measurement using cost function.

Table 1
Comparisons of processing times of cost functions (sec).

Cost function	SAD	SSD	MAD	NCC	MNCC
Teddy	3.24	4.51	5.10	8.54	7.84
Cones	4.51	5.32	5.52	10.52	8.38
Tsukuba	3.62	5.16	5.32	9.48	7.81
Venus	3.87	4.98	5.52	9.84	8.29

Table 2
Comparisons of error ratios of cost functions (%).

Cost function	SAD	SSD	MAD	NCC	MNCC
Teddy	30.8	30.8	31.5	20.2	21.3
Cones	33.5	33.5	32.2	22.2	25.6
Tsukuba	28.5	28.5	27.5	25.5	27.8
Venus	36.4	36.4	35.8	30.5	28.2

2.2. Cost function

The numerical criterion for similarity is defined by a cost function. A cost function can be classified as a minimum cost function such as SAD, SSD (sum of squared differences), and MAD (minimization of absolute deviation), or as a maximum cost function such as NCC (normalized cross correlation), and MNCC (modified normalized cross correlation) [19].

The effect by cost function is slightly different after measuring similarity in stereo matching. Table 1 shows the processing times of the various cost functions. SAD has the shortest execution time, which may change with the test environment. Table 2 shows the error ratios of the estimated disparities, which can vary with the algorithm used for stereo matching. Error ratio was defined as the number of different disparities between the ground truth and the estimated disparity map. A normalization-based cost function has relatively high accuracy, but has a long execution time, so it should be used selectively according to application.

2.3. Matching window

The graph shown in Fig. 2 is the accumulated cost of the 190th scan line in the Teddy image, which is calculated with two matching windows of 3 × 3 and 9 × 9 size. The white line represents the minimum cost values of all pixels in the scan line, and the height corresponds to the disparity of each pixel. Region A and region B have different characteristics. Region A corresponds to the background and has a relatively complex texture. The pixels in this region exist at almost the same distances, so the pixels in the region should have the same disparity. When the 3 × 3 matching window is used, the estimated disparity has many errors; but the 9 × 9 matching window is used, it has little error. The disparities of region B should have a constant rate of change, but with a large matching window, the boundary regions, which comprise the edge of an object, causes disparity error. Thus, it is necessary to decide a suitable matching window according to the local property of the texture in an image.

3. Proposed VLSI architecture

Stereo matching is useful for distance ranging in stereo vision, but it has a limitation for real-life application due to its considerable computational expense. To overcome this limitation, an H/W-based