Proceedings of

The International Conference on Intelligent Information Technology (2INTECH 2015) and International Workshop on Advanced Computing and Multimedia Technology (ACMT 2015)

October 21 - 23, 2015

Guam, USA

2INTECH 2015 and ACMT 2015

Table of Contents

A Simplified Rectification Method for Efficient Hardware Implementation

Jongkil Hyun,Byungin Moon

School of Electronics Engineering, Kyungpook National University,Daegu, Korea

bihmoon@knu.ac.kr

Abstract

Rectification is used as a pre-processing step in all stereo matching algorithms. The rectification algorithms are hard to implement in hardware circuit because of the complexity of its mathematical calculations. To solve this problem, this research proposes an efficient calculation method and rectification hardware architecture optimized for stereo matching. The proposed calculation method optimizes previous rectification calculations methods using the linearity of pixel unit calculation and approximation of coordinate translation. The experimental results show that the proposed rectification method has marginal differences in terms of the rectification results compared with the previous rectification method while consuming relatively less hardware resources. The proposed rectification hardware architecture can be implemented in all stereo vision systems and can be applied in various applications such as autonomous vehicles, intelligent robots, and 3D multimedia contents.

Keywords: Stereo vision;stereo matching;rectification; hardware implementation

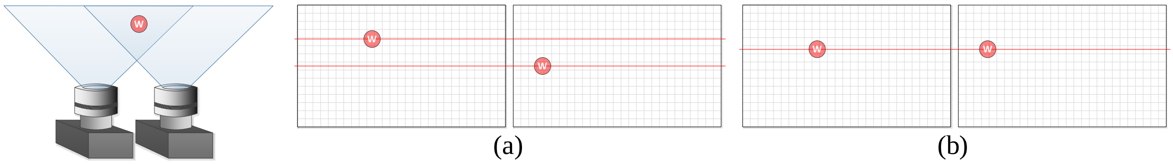
1. Introduction

A stereo vision system calculates the 3D distances by finding the disparity of objects in stereo image pairs captured by a stereo camera. The 3D distance information of a stereo vision system can be applied to various applications such as autonomous vehicles, intelligent robots, and 3D multimedia contents [1-5]. In the field of multimedia, 3D contents generation by stereo vision is also an active research area. For efficient 3D multimedia content creation, efficient real-time performance is required. In order to adapt stereo vision systems for 3D multimedia systems, there has been an abundance of research on hardware implementations of stereo vision systems proposed for real-time performance [6-8]

In general, stereo vision systems for hardware implementation go through pre-process, matching process, and post-process. In the pre-process step, input images from the stereo camera are processed to be easily used the disparity extraction. The matching process step extracts the disparity map by finding correspondence of the same object. The post-process step refines the disparity map by post-correcting the disparity map extracted by the matching process step. The matching process of the stereo vision system includes various algorithms such as the local matching method, the global matching method, and the SGM (semi-global matching) method, and these various matching algorithms [9] must go through the pre-process step known as rectification. The rectification reduces the disparity search range, which is used to discover the disparity in the matching process, to 1D. Thus, rectification is commonly required for all matching algorithms to reduce calculation overhead [10]. Because rectification requires complex matrix calculations such as coordinate translation and image rotation, it causes difficulties in hardware implementation. Thus, this research proposes an optimized rectification calculation method that reduces use of hardware resources by optimizing previous calculation methods.

2. ConventionalRectification Method

The rectification in stereo vision systems is the process of aligning the epipolar line of the epipolar geometry of the stereo camera. By aligning the epipolar line of the epipolar geometry, the height differences between the stereo images taken by the stereo camera are removed. So, the disparity search range for matching isreducedto 1D. This process is shown in Figure 1. As shown in Figure 1-(a), to findpoint (W)in an image, the entire other image must be scanned, but in the case of rectified images, because there is no height difference between the stereo images as shown in (b), the disparity search range to find the correspondence is reduced to one line. This rectification is a critical process in real-time stereo vision systems that extract dense disparity maps.



(a) (b)

Figure 1. Stereo image rectification. (a) Pre-rectification. (b) Post-rectification.

In general the process of rectification with the distortion correction of the lens omitted is as shown in Figure 2. A detailed explanation of this is as follows.

1. Assume a 2D pixelcoordinates of rectified image.
2. The 2Dpixel coordinates of the rectified image are translated to 3D camera coordinates.
3. The translated coordinatesare rotated to the3D cameracoordinates, or in other words the 3D camera coordinates before rectification.
4. The rotatedcoordinatesare projected to 3D homogeneous coordinates of an unrectified image.
5. The projected coordinatesare translated to 2D pixel coordinates of the unrectified image.
6. The pixel values of the calculated coordinatesare mapped to the coordinates of the originally assumed rectified image using linear interpolation.

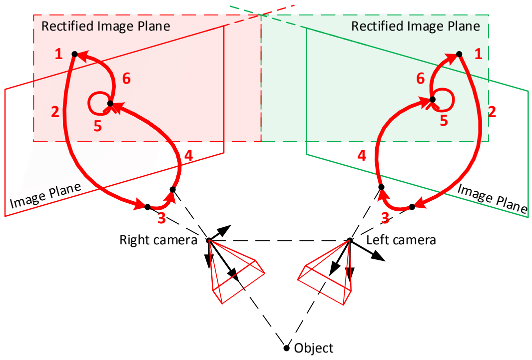


Figure 2. Rectification process.

The formulaic expression of this process is defined as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = A \cdot {R_{rect}}^{T} \cdot {A_{rect}}^{-1} \cdot \begin{bmatrix} x_{rect} \\ y_{rect} \\ 1 \end{bmatrix}, \quad (1)$$

where $[x \quad y \quad z]^T$is the 3D homogeneous coordinates of the unrectified image, and $[x_{rect} \quad y_{rect} \quad 1]^T$is the 3D homogeneous coordinates of the rectified image. A is the intrinsic parameter matrix of the camera, A_{rect} is the intrinsic parameter matrix of the rectified image and R_{rect} is the rectification rotation matrix. These parameter matrices are3×3matricesthat can be calculated using the process of stereo camera calibration [11].

3. Proposed Rectification Method

In general, the process of rectification, as described above, includes complex matrix calculations. Despite the pre-processing step, these complex calculations requires significant hardware resources. To overcome the problem, this research proposes a novel rectification method which requires less hardware resources for hardware implementation by optimizing this general rectification calculation method.

A, A_{rect}, and R_{rect}, which are the parameters of Equation(1), are the unique 3x3 matrix parameters of the stereo camera and these parameters are an inherent property of the stereo camera calibration. Therefore $A \cdot {R_{rect}}^{T} \cdot {A_{rect}}^{-1}$ of Equation(1) can be transposed to one coefficient matrix as

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \equiv A \cdot {R_{rect}}^{T} \cdot {A_{rect}}^{-1}, \quad (2)$$

andEquation(1) can be rewritten as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \cdot \begin{bmatrix} x_{rect} \\ y_{rect} \\ 1 \end{bmatrix} = \begin{bmatrix} c_{11} \cdot x_{rect} + c_{21} \cdot y_{rect} + c_{13} \\ c_{21} \cdot x_{rect} + c_{22} \cdot y_{rect} + c_{23} \\ c_{31} \cdot x_{rect} + c_{32} \cdot y_{rect} + c_{33} \end{bmatrix}. \quad (3)$$

Through the process of Equation(3), the 3D homogeneous coordinates of the unrectified image can be calculated from the 2D pixel coordinates of the rectified image. By dividing both the x and y axis by the z axis of the 3D homogeneous coordinates, the 2D pixel coordinates of the camera image can be calculated. As the commonly used stereo camera sends stereo imagesin pixel units, Equation (3) is further simplified by using accumulation.

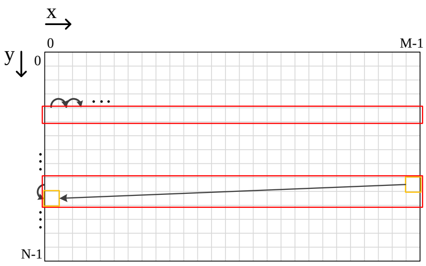


Figure 3. Rules of pixel unit calculations.

As shown in Figure 3, when processing resolution $(M \times N)$ images in pixel units, 2 types of processing operation are required. When the position of a pixel increases in the x-axis direction, the position of y-axis is fixed and the position of x-axis is increasedfrom 0 to $M-1$, but when the position of a pixel increases in the y-axis direction, the position of x-axis decreases from$M-1$ to 0, and the position of y-axis increases from 0 to $N-1$.By applying this uniformity of pixel unit calculation fromEquation(3), the $[x \quad y \quad z]^T$ of the 3D homogeneous coordinates of the unrectified image can be optimized as an accumulation according to the increase in the x-axis and the increase in the y-axis. Therefore, Equation(3) can be rewriten as

$$\begin{bmatrix} x_n \\ y_n \\ z_n \end{bmatrix} = \begin{bmatrix} x_{n-1} \\ y_{n-1} \\ z_{n-1} \end{bmatrix} + \begin{cases} \begin{bmatrix} c_{11} \\ c_{21} \\ c_{31} \end{bmatrix}, when\ increase\ column, \\ \begin{bmatrix} c_{12} - (M-1) \cdot c_{11} \\ c_{22} - (M-1) \cdot c_{21} \\ c_{32} - (M-1) \cdot c_{31} \end{bmatrix}, when\ increase\ row. \end{cases} \left(\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} = \begin{bmatrix} c_{13} \\ c_{23} \\ c_{33} \end{bmatrix} \right) \quad (4)$$

When calculating along the x-axis, because there is no change in rows, the only position of the x-axis is variable and$[x \quad y \quad z]^T$ can be each expressed as linear equations that contain gradient $[c_{11} \quad c_{21} \quad c_{31}]^T$. When calculating along the y-axis, the change in the x-axis point is always from $N-1$to 0 andthe position of the y-axis increases from0 to $M-1$. Therefore when calculating along the y-axis direction, the gradient of the x-axis is constant, and the position of y-axis can be expressed as a linear equation withthe gradient $[c_{12} - (M-1) \cdot c_{11} \quad c_{22} - (M-1) \cdot c_{21} \quad c_{32} - (M-1) \cdot c_{31}]^T$. Therefore acomplex matrix operationlike Equation (1)can be optimized using the simple accumulation of the constant coefficient values according to the increase in rows and columns from the initial values as shown in Equation(4).The initial values $[x_0 \quad y_0 \quad z_0]^T$inthe Equation(4) are the 3D homogenous coordinates of the unrectified image that correspond to the $(0,0)$in the rectified image coordinates, where$[x_n \quad y_n \quad z_n]^T$isthe position of the 3D homogeneous coordinates of the unrectified image correspondingto the currently calculated position of the rectified image coordinatesand $[x_{n-1} \quad y_{n-1} \quad z_{n-1}]^T$ isthe position of the3D homogeneous coordinates of the unrectified image correspondingto the previouslycalculated position of the rectified image coordinates.

$[x_n \quad y_n \quad z_n]^T$is the 3D homogeneous coordinates of the unrectified image.So these coordinates must be translated into 2D pixel coordinates. In order to do this translation, the x-axis and y-axis are each divided by the z-axis. From Equation (3), $z = c_{31} \cdot x_{rect} + c_{32} \cdot y_{rect} + c_{33}$. The variables that influence the value of z are the 2D pixel coordinates of the rectified image and the 3rd row vector $[c_{31} \quad c_{32} \quad c_{33}]^T$ of the coefficient matrix of Equation(2). The 3rd row vector of the coefficient matrix of Equation(2) can be rewritten as

$$[c_{31} \quad c_{32} \quad c_{33}] = \left[{}^{r_{31}}/_{f_x} \quad {}^{r_{32}}/_{f_y} \quad r_{33}\right] \cong [0 \quad 0 \quad 1].(5)$$

$$\left(f_x \gg r_{31}, f_y \gg r_{32}, r_{33} \approx 1\right)$$

where f_x and f_y are the focal lengths of the camera, and $[r_{31} \quad r_{32} \quad r_{33}]$ is the 3rd column of the rectification rotation matrix. According to the characteristics of the camera intrinsic parameter and rectification rotation matrix, z can be simplified as 1 as shown in Equation (5). First, because the rectification rotation matrix takes the form of a symmetric matrix, the diagonal elements are mostly 1 and the rest of the elements are smaller than 1. Because the focal length is a pixel unit coefficient that is the intrinsic parameter of camera,it is much larger than 1. Therefore the z value from Equation(3) can be considered as 1 as in Equation (5), which means that when translating the 3D homogeneous coordinates to 2D pixel coordinates in the process of Equation (4), the division by z can be omitted. Therefore, the process of Equation(4) can be simplified as

$$\begin{bmatrix} x_n \\ y_n \end{bmatrix} = \begin{bmatrix} x_{n-1} \\ y_{n-1} \end{bmatrix} + \begin{cases} \begin{bmatrix} c_{11} \\ c_{21} \end{bmatrix}, when\ increase\ column, \\ \begin{bmatrix} c_{12} - (M-1) \cdot c_{11} \\ c_{22} - (M-1) \cdot c_{21} \end{bmatrix}, when\ increase\ row. \end{cases} \left(\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} c_{13} \\ c_{23} \end{bmatrix}\right) \quad (6)$$

4. Proposed Rectification HardwareArchitecture

Previouslyproposed rectification hardware architectures are broadly divided into two types. The first is the LUT-based method and the second is the computational logic-based method. Because the LUT-based method must store all the correspondence of coordinate translation for the rectification images, this method requires a significant amount of memory. Also, as the resolution of the image increases, the amount of memory alsoincreased proportionally. On the other hand, thecomputational logic-based method uses a minimum amount of memory and has the advantage of being able to calculate the correspondence coordinates for the rectification images. However, the computational logic-based method rectification has a complex logic due to complex matrix operations. Therefore in this research, the complex operations of rectification are optimized using the linearity of pixel unit calculation and approximation of coordinate translationanda novel hardware architecture is proposed which uses a proposed rectification calculation method. Figure 4 shows both the optimized rectification hardware architecture and also the conventional rectification hardware architecture.

(a) (b)

Figure 4. Rectification hardware architecture. (a) Conventional rectification. (b) Proposed rectification.

The proposed rectification optimizes the process of calculating the correspondence coordinates of the camera image. As shown in Figure 4, when compared to conventional rectification hardware architecture, the multipliers and the dividers from the correspondence coordinates calculating part have been removed.

5. ExperimentalResults

The proposed rectification hardware architecture was designed using Verilog HDL and implemented using Xilinx FPGA Virtex-7. Figure 5 is a comparison of the results of the conventional rectification and the proposed rectification, where Table 1 is the pixel coordinate error of the proposed rectification based on the results of the conventional rectification. As shown in Table 1, the average pixel coordinate error is under 1 pixel, which is an error that occurs due to the omitted process, division by z when, translating the 3D homogeneous coordinates of the unrectified image to 2D pixel coordinates. However,as shown in Figure 5, this error is marginal and the proposed rectification also finely aligns with the epipolar line.

(a)
(b)
(c)

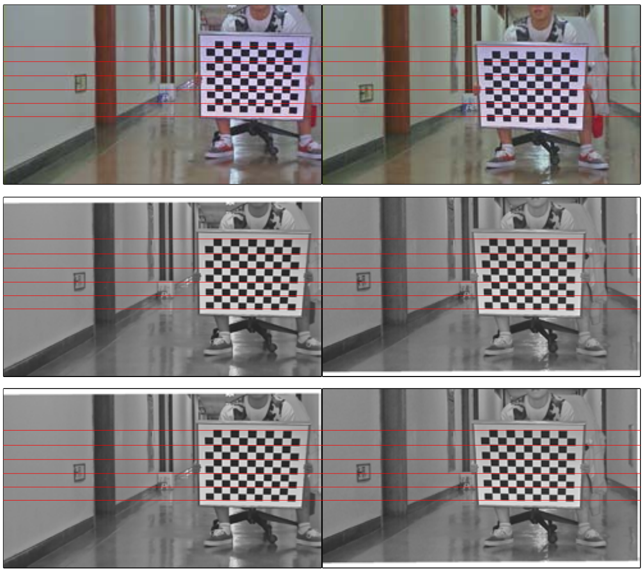


Figure 5. Rectification results. (a) Original left and right images. (b) Conventional rectified images. (c) Proposed rectified images.

Table 1. Number of pixel coordinate errors of the proposed rectification.

Stereo Images	Average number oferror pixels		Max number of error pixels	
	x-axis	y-axis	x-axis	y-axis
Left image	0.4465	0.2255	2	2
Right image	0.5028	0.2880	3	2

Thehardwareresourcesconsumedwhentheconventionalrectificationandproposedrectificationwereimplementedu singFPGAareshowninTable 2. The proposed rectification hardware architecture uses less than 1/3of the hardware resources compared with that of conventional rectification hardware architecture because of the simplified rectification calculation.

Table 2. Synthesis results of the FPGA hardware implementations.

Method	Number of Slice Registers	Number of Slice LUTs
Conventional rectification	12837	9508
Proposed rectification	593	3718

6. Conclusions

For real-time performance of stereo vision which is used for 3D multimedia system, this research proposed a novel rectification method by optimizing previous rectification calculations methods using the linearity of pixel unit calculation and approximation of coordinate translation and also proposedrectification hardware architecture using the proposed rectification calculation method. The experimental results show that when comparing the proposed rectification to the conventional rectification there were marginal differences in the results while reducing the use of hardware resources by a significant amount. Because the rectification is necessary for stereo matching algorithms, to adapt to the stereo vision based real-time 3D multimedia system, it must be considered when designing the rectification hardware architecture for real-time performance and hardware resource minimization. The proposed rectification hardware architecture shows marginalerrorsand uses less than 1/3 hardware resourceswith compared with conventional rectification hardware architecture. So, the proposed rectification method can be efficiently applied to real-time multimedia system based on stereo vision.

Acknowledgements:

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the C-ITRC (Convergence Information Technology Research Center) (IITP-2015-H8601-15-1002) supervised by the IITP (Institute for Information & communications Technology Promotion).

Corresponding Author:
Associate professorByungin Moon
School of Electronics Engineering
Kyungpook National University
Buk-gu, Daegu 41566, Korea
E-mail:bihmoon@knu.ac.kr

References

1. Bertozzi M. GOLD: A Parallel Real-Time Stereo Vision System for Generic Obstacle and Lane Detection.Image Processing, IEEE Transactions on Jan. 1998; 7(1):62-81.
2. Bensrhair A, Bertozzi M, Broggi A, Fascioli A, Mousset S, Toulminet G. Stereo Vision-based Feature Extraction for Vehicle Detection. IEEE Intelligent Vehicle Symposium, Paris, France Jun. 2002; 2:465–470.
3. Uchida N, Shibahara T, Aoki T, Nakajima H, Kobayashi K. 3-D Face Recognition Using Passive Stereo Vision. IEEE International Conference on Image Processing 2005 (ICIP 2005), Genoa, Italy Sep. 2005; 2:950–953.
4. Murray D, Jennings C. Stereo Vision-Based Mapping and Navigation for Mobile Robots.IEEE International Conference on Robotics and Automation, Albuquerque, NM Apr. 1997;2:1694–1699.
5. Zhang L, Vazquez C, Knorr S. 3D-TV Content Creation: Automatic 2D-to3D Video Conversion. Broadcasting, IEEE Transactions on Jun. 2011; 57(2);372-383.
6. Jin S, Cho J, Pham XD, Lee KM, Park SK, Kim M, Jeon W. FPGA Design and Implementation of a Real-Time Stereo Vision System. Circuits and Systems for Video Technology, IEEE Transactions on Jan. 2010; 20(1):15-26.
7. Darabiha A, Rose J,Maclean WJ. Video-Rate Stereo Depth Measurement on Programmable Hardware. IEEE Computer computer Society Computer Vision and Pattern Recognition, Madison, WIJun. 2003; 1:203–210.
8. Jia Y, Zhang X, Li M, An L. A Miniature Stereo Vision Machine (MSVM-III) for Dense Disparity Mapping. 17th International Conference Pattern Recognition, Cambridge, UK Aug. 2004; 1:728–731.

9. Lazaros N, Christou Sirakoulis GC. Gasteratos A. Review of Stereo Vision Algorithms: From Software to Hardware. International Journal of Optomechatronics, Nov. 2008; 2(4):435-462
10. Pantilie CD, Haller I, Drulea M. Nedevschi S. Real-Time Image Rectification and Stereo Reconstruction System on the GPU. 10th International Symposium on Parallel and Distributed Computing (ISPDC), Cluji Napoca, Romania Jul. 2011; 79-85
11. Bouget JY. Camera Calibration Toolbox for MATLAB, Available: http://www.vision.caltech.edu/bouguetj/calib_doc.