A region-based fast stereo matching and tracking

Dong-Jin Seo , Ju-Kyong Jin † , Sang-il Na†† and Dong-Seok Jeong†††

Department of Electronic Engineering, College of Engineering, Inha University, Incheon, Korea

Summary
Stereovision has been seriously studied since the mid of 1990s, and various products have been released in recent years. However, it is difficult to implement stereovision in real time due to complicated calculations and various limit conditions. This paper suggests a high-speed stereovision system that can be used for crime prevention and motion detection by application of a changing pixel region-based algorithm.

Key words:
Stereovision, Motion detecting, camera, sensor

1. Introduction

Stereovision has been greatly noticed in point of extracting distance information using two cameras. However, its use has been limited because of additional costs for cameras, decrease of speed due to excessive complexities, difficulty in real-time implementation and complexities about implementation of embedded systems. Nevertheless, it is possible to use the Stereovision in various fields.[1] Fig. 1 shows stereovision camera that is now on sale.

Fig. 1 Stereovision camera

In a motion detection field, a difference value between frames using one camera is used. It is determined whether or not it is changed according to a difference value that is generated from subtracting each value of pixels of the two successive frames at this time. However, there is no method for checking distance information on a relevant region. Most of crime prevention systems and various uses for motion detection systems are required for distance information. Although a laser sensor is mounted lower part of camera, there is some problem that only distance information on a specific region is perceived. [2]

This paper explains a system for detecting a change pixel value and then estimating distance information on a relevant region at high speed, with this point of view.

2. Stereovision

Stereovision indicates a system for extracting images including distance information using two cameras. Recently, panorama images are made using three or more cameras and various distance information is inserted into relevant images. Generally, three stages are performed to extract distance information using this stereovision.

- Camera calibration

Two cameras are generally used as the same model that has been made by the same company. The purpose is to adjust two cameras as similar as possible. However, a lens and a CCD (or CMS) sensor are completely matched to each other, so internal parameter values should be adjusted. This process, so to speak, indicates camera calibration. In order to perform calibration, a specific pattern is photographed simultaneously from two cameras and then the change matrix is made by using information on the relevant pattern. Moreover, the matrix is used for disparity map calculation. [3]

- Disparity map calculation

After a pixel or a region that is judged to be the same is found from the two images, a distance value between relevant positions on right and left images is calculated. Various methods for finding the same position have been studied. For SSD, SAD, NCC and WMC methods, a mode for comparing all pixel values is used in a specific region.[4] Such mode has so many complexities but can check wider regions. Moreover, a feature-based method has been designed for a faster countermeasure, like an edge-based method, but it is not robust to rotation and flat regions. On the one hand, methods using SIFT, KLT and SURF which are scale and rotation invariant have been designed. But computation complexity of this methods is too heavy.[5] Due to various advantages and disadvantages, in the stereovision system to be actually applied, a variety of functions is combined and used.

Manuscript received April 5, 2010
Manuscript revised April 20, 2010

- Distance calculation

Disparity map is created with a distance value between pixels that are judged to be actually the same on the two images and it is possible to calculate back from this value to the distance through a simple numerical equation. Fig. 2 shows a calculation principle of distance. An image of an object is shown on an image plane after passing through lenses of right and left cameras. At this time, a and b denote the distance between relevant pixels and the center of the image. f denotes focal length of a camera. d denotes distance between two cameras and z denotes actual distance. [1]

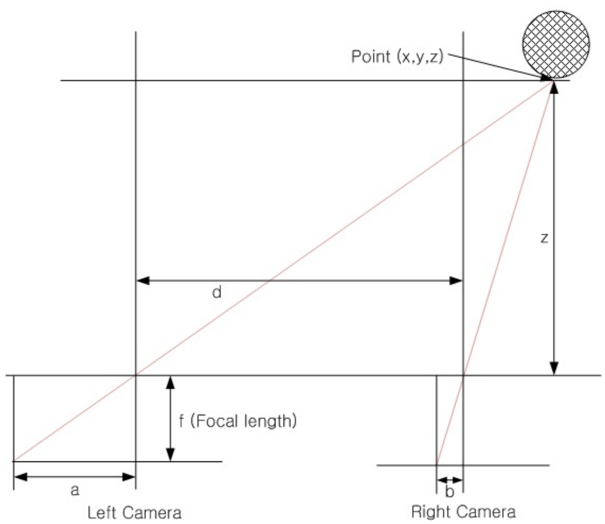

Fig. 2 Distance calculation method

With these parameters, the actual distance z can be calculated. The equation for actual distance is as follows. (1)

$$z = \frac{f\,b}{(a-b)} \qquad (1)$$

3. Motion detection and calculating distance

The most classical method for motion detection is to estimate pixel change inside the image. In other words, it calculates the difference between two successive frames that are inputted from the cameras, or calculates a value that is changed after making of images for background. However, it is difficult to get information on distance from a photographing location.

As explained above, stereovision is required for so many complexities to extract distance information on the whole region. In this paper, distance information on regions including motions is calculated to satisfy two contents above and defects of the two contents are compensated for each other.

3.1 Motion detection

1) Each background is generated from two cameras and information on changing pixel is extracted by the standard of the generated background. The background used at this time is continuously updated, so it can solve problems of mistaken perception due to background.

Whether an image is changed or not can be understood by calculation of difference from two successive frames. It is possible to express such frame difference using zero-mean Gaussian distribution and its probability density function. If the probability calculated in this way is more than a threshold value, it is judged that there is a change in the inputted image. [6]

$$p(FD|H_0) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{FD^2}{2\sigma^2}\right) \qquad (2)$$

where FD is the frame difference and σ^2 is the variance of the frame difference. H_0 denotes the null hypothesis, that there is no change at the current pixel.

2) If a difference value with background image is generated, each camera performs labeling of a relevant region. The size of changing region that has been generated due to noise is small, so the values less than a certain value are excluded.

3) The barycenter of a region that is more than the reference value is calculated.

3.2 Calculating Region dissimilarity

1) If there are two or more calculated coordinates, a window is moved and SSD and SAD values are calculated. SSD and SAD get to be the standards for judging peripheral similarity, centering around a relevant pixel. According to the equations (3) and (4), when the similarity of two regions gets higher, then the values get smaller.

$$SSD(x, y, d) = \sum_w [R_{x,y} - L_{x+d,y}]^2 \qquad (3)$$

$$SAD(x, y, d) = \sum_w |R_{x,y} - L_{x+d,y}| \qquad (4)$$

At this time, $R_{x,y}$ denotes a pixel value in the right image and $L_{x+d,y}$ a pixel value in the left image. w denotes a fixed size of matching window, centering around a position (x, y) of the right image and a position (x+d, y) of the left image corresponding thereto.

2) The distance of coordinates x and y is calculated, that are judged as the same object. It is possible to calculate distance of a specific pixel in the right and left images easily using the equation (5). At this time, (x_r, y_r) and

(x_l, y_l) denote barycentric coordinates of the right and left images.

$$d = \sqrt{(x_r - x_l)^2 + (y_r - y_l)^2} \qquad (5)$$

3) The calculated distance is substituted for the equation (1) and thus the value z is calculated.

4. Experimental environment and result

4.1 Experimental environment

CPU	DualCore E8600
RAM	DDR2 2GByte
Dev. S/W	Visual studio 2008

This experiment has been performed under the environment of 8600 CPU and 2G DDR2 RAM, and visual studio 2008 has been used as a development environment for the experiment. The resolution of the inputted image is 640x480.

4.2 Result image

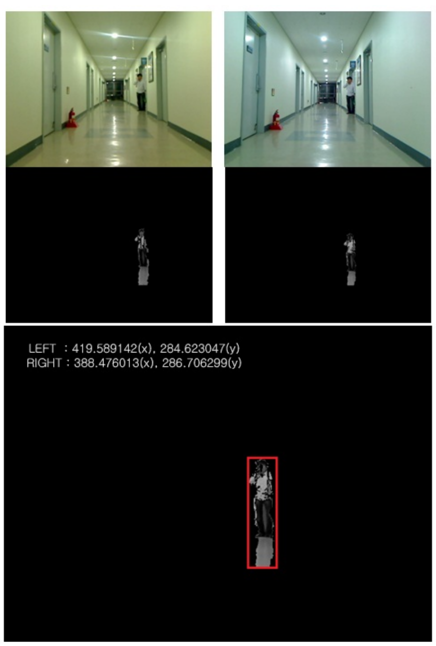

Fig. 3 The result of single object detection

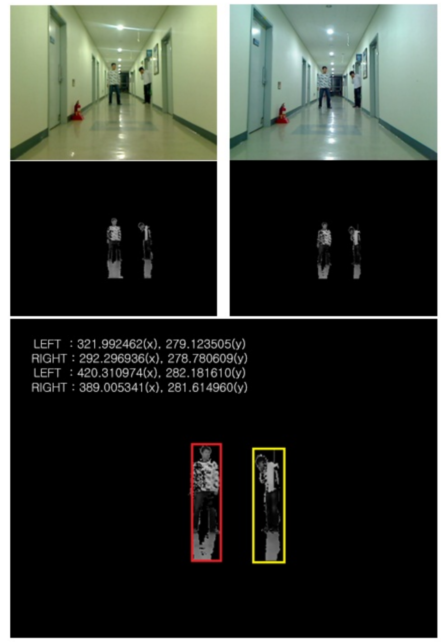

Fig. 4 The result of multi-objects detection

4.3 Result analysis

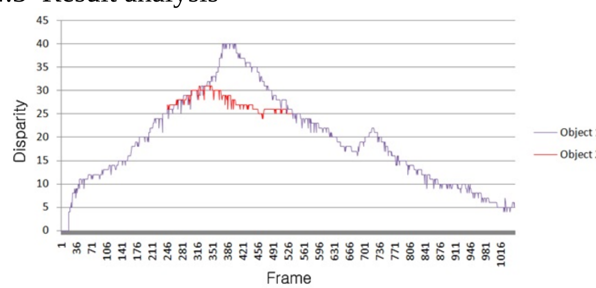

Fig. 5 Disparity value calculated for each frame

Fig. 3 shows a detection image when a single object appears. As shown in the figure, in spite of difference of white valance, a region is normally extracted by the standard of brightness information. If a disparity value is displayed in the right image by the standard of the extracted central point, there is a difference value amounting to about 30.

Fig. 4 shows results from the case of appearance of two objects. Even though the number of regions in which pixel values are changed is two, each barycenter is calculated and compared and then the side that is matching therewith is calculated.

Fig. 5 shows the variance of disparity value by frame. This shows that object 2 was detected from 246^{th} frame and two objects were normally tracked.

Total photographing time required for the experiment was 35.12Sec, and complexities for 1048 frames were performed during this time. In other words, 29.84 frames were processed per second.

5. Conclusion

The present paper is focusing on perform stereo matching at high speed and tracking an object that suddenly appears at high speed. For this, adaptive background is generated and regions that show change of pixel values are rapidly calculated, then only the regions are compared. As the number of regions to be calculated is decreased as many as possible, it makes the increase in calculation speed and the real-time calculation.

If moving paths of detected regions are forecasted and matching of two images is performed according to such forecast, it will produce desired and better results.

References

[1] David A.Forsyth and Jean Ponce, "Computer vision", Prentice Hall, pp.234-250, 2002

[2] Huang C.-C. and Zheng Y.F.,"Integration of vision and laser displacement sensor for efficient and precise digitizing", *IEEE Industrial Electronics Society*, vol 43, pp.364-371, 1996

[3] Jean-Philippe Tarel and Jean-Marc Vezien, "A complete software Solution for Camera Calibration", *INRIA*, pp.5-10, 1996

[4] Takeo Kanade and Masatoshi Okutomi, "A stereo matching algorithm with an adaptive window", *IEEE Transactions on pattern anaylsis and machine intelligence*, Vol 16, No 9, 1994

[5] Tola E. and Lepetit V. and Fua P.,"An Efficient Dense Descriptor Applied to Wide-Baseline Stereo", *IEEE Computer Society*, vol 32, pp.IEEE Computer Society, 2010

[6] T. Aach, A. Kaup, and R. Mester, "Statistical model-based change detection in moving video," *Signal Processing*, vol. 31, pp. 165–180, Mar.1993.

Dong-Jin Seo received the B.S. and M.S. degree in Electronic Enginee-ring at the School of Electronic Engineering from INHA University, Incheon, Korea, in 2002 and 2004. He has been received the Ph.D. course in Electronic Engineering at the School of Electronic Engineering from INHA University, Incheon, Korea from 2006. His main research interests include stereo vision, machine vision and image/video coding.

Ju-Kyong Jin received the B.S. and M.S. degree in Electronic Enginee-ring at the School of Electronic Engineering from INHA University, Incheon, Korea, in 2002 and 2004. He has been received the Ph.D. course in Electronic Engineering at the School of Electronic Engineering from INHA University, Incheon, Korea from 2006. His main research interests include machine vision, image/video coding, MPEG-7 and DRM.

Sang-il Na received the B.S. and M.S. degree in Electronic Enginee-ring at the School of Electronic Engineering from INHA University, Incheon, Korea, in 2002 and 2004. He has been received the Ph.D. course in Electronic Engineering at the School of Electronic Engineering from INHA University, Incheon, Korea from 2006. His main research interests include video motion estimation and image/video coding.

Dong-Seok Jeong received the B.S. degree from Seoul National University, Korea, in 1977. And he worked as research engineer in Agency for Defense Development. He received the M.S. degree and Ph.D degree from Virginia Tech, USA, in 1985 and 1988, respectively. Currently he is an associate professor of the Department of Electronic Engineering at Inha University. He is a senior member of IEEE. His research interests include machine vision, image/video coding, MPEG-7 and DRM.