2005－AVM－49（8）
2005／7／8

A Proposal on Active Extensible Stereo Camera Array For High Speed Moving Object Detection

Yingdi Xie† Jun Ohya†
†Global Information and Telecommunication Studies, Waseda University
1011 Okuboyama Nishi-Tomida Honjo-shi Saitama 367-0035 Japan
Email: †xieyingdi@gmail.com, †ohya@waseda.jp,

Abstract Recent years, almost motion detection researches have been focused on detecting moving object with a low speed, while only a few have been considered high speed case. In this paper, a new approach for high speed moving object detection based on active extensible stereo camera array, and a novel approach of motion detection are presented. In our approach, the camera array is extensible by increasing or decreasing the number of the stereo camera set to scale the video sequence to a desired frame rate. In the experiment, this approach shows its ability of getting high frame rate images, which is the basic for high speed moving object detection.

Keywords: Motion detection, stereo vision, camera array

1. INTRODUCTION

Motion detection based on active camera, as one of the most popular research topics among robot vision, has been actively researched since decades ago. This technique has been applied to industrial robot vision, military precision guided weapons, ITS (Intelligent Traffic System) and IVS (Intelligent Vehicle System).

Some of the research works achieved a detection function based on a single camera, while others use stereo or multiple camera system. One of the single camera cases [1] applied the image difference method to detect motion, and tracking with blocks matching method. By accounting of the motion vector within each two consecutive frame, the ego-motion of the camera is computed from the histogram of flow vector. Moving object is detected by comparing the difference between each block and the ego-motion. Kim et al.[2] applied an optimal RB (Representative Block) method that achieves higher tracking performance and less computation time than ordinary block matching based methods, but their method requires an automatic pan-tilt mechanism, and their computation speed is limited by a video rate. Another method [3] analyzes the subtraction image obtained by the current frame and the frame estimated from the previous frame and the camera's pan-tilt data so that the motions of moving objects in the scene can be estimated. As a common serious problem of single camera based methods, accurate depth information cannot be obtained from a single frame.

Stereo or multiple camera based method could overcome the above-mentioned limitation existing in the single camera based method. The method developed by Satoh et al. [6] uses stereo cameras both for computing 3D coordinates of feature points and for detecting the motions of moving objects by utilizing the ego-motion of the stereo cameras.

This paper proposes a method that utilizes multiple cameras. In stead of acquiring images simultaneously, a number of stereo pairs acquire images at a short time interval. The image acquisition method is similar to the method [5], which acquires multi-thousand frame-per-second (fps) video based on electronic rolling shutter camera array. Our method aims at achieving scalable high speed motion detection by simply adding up more cameras. We assume our method uses global shutter cameras. In the theoretical analysis, the camera system shows its ability of adapting the motion detection speed to the scene's complexity.

In the next section, the proposed method is explained. In Section 3, a simple experiment on estimating the features coordinates is demonstrated. Section 4 concludes this paper and describes future work.

2. PROPOSED METHOD

As Fig. 1 shows, multiple-image is acquired by a number of camera pairs. The camera pairs number is decided by adaptation of the last frame.

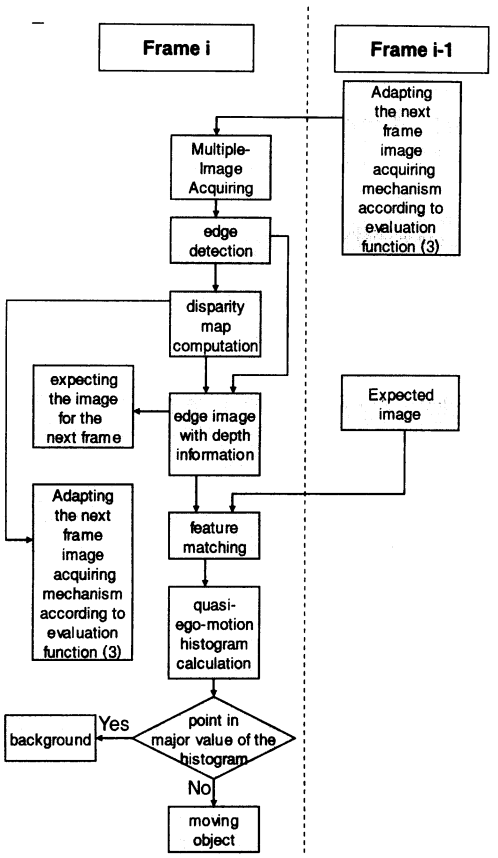

Fig.1 Processing Flow Chart

After applying edge detection for acquired images, disparity map is computed. Here, we assume edge points are feature points. By using the depth information and edge image, the expected image for next frame could be computed. Feature matching between edge image with depth information and the expected image from last frame is carried out. Motion point is detected when point lies in minor value of the histogram of quasi-ego-motion, which will be introduced in 2.2.2.

2.1 Camera Array System

Each camera of one stereo pair is positioned with a known distance B_x in the horizontal direction and B_y in the vertical direction, where the values B_x *and* B_y denote the baselines in the horizontal and vertical directions, respectively. Note that the "horizontal direction" is parallel to the x axis of the camera's focal plane. Each camera's coordinate axes are aligned so that each baseline is parallel to the camera's x coordinate axis. In case that two cameras of one stereo pair are set in the horizontal direction, then the system consists of a $2m \times n$ camera matrix, as depicted in Fig.2, where m and n are both positive integers. Thus, the image sequence is acquired at a maximum of $30mn$ fps (frames per second).

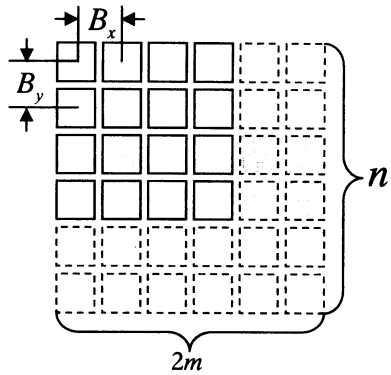

Fig.2 Camera Array

In view of real-time processing, it is preferred to use micro processors, each of which is assigned to a stereo pair, in stead of using a computer that carries out stereo matching for all the pairs. Each stereo camera pair has a micro processor for feature extraction and stereo matching. There is a central-processor for performing image registration and motion detection.

2.2 Processing Steps

A number of camera stereo pairs are controlled to take color images at the same time in order to scale the image sequence to a high frame rate and to obtain disparity map. As an example, the image acquiring strategy is demonstrated as Fig.2. Here, the camera array is assumed to be composed of 8×8 cameras, or 4 stereo pairs. Every 4 cameras, depicted as 4 gray small rectangles at each time in Fig.3, are controlled to take image at each time interval in turn. By computing stereo matching, the depth information of each feature points in the images could be calculated. With this, feature positions within the image for next frame processing can be expected, depicted as gray circles. And, the motion detection is computed from two

consecutive frame images at the same position, in Fig.2, connected by arrowhead of interactive. A more common description of the image acquiring and processing mechanism is given in Fig,3 the last page.

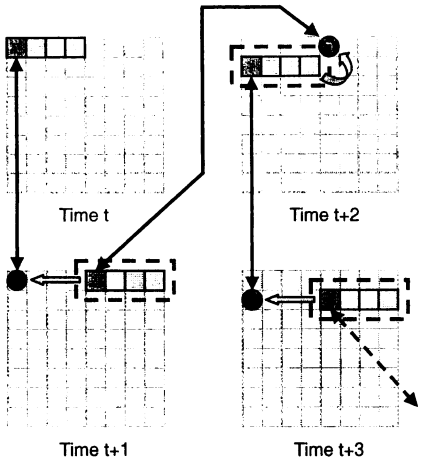

Fig.3 The image taking strategy

Every 4 cameras are controlled to take image at each time interval in turn. Then image for next frame processing, depicted as gray circle, can be expected for further processing.

2.2.1 Stereo Matching and Position Deduction

In this proposal, since the coordinate axis of focus plane for each camera within one stereo pair are fixed to be parallel, the stereo matching problem can be solved by searching corresponding point in one dimension. Not all the pixels in one image apply stereo matching. So firstly, edge detection is carried out and the points on the detected edges are considered as the features for further processing. Here we adopt a noise-against edge detection method [9]. Stereo matching is achieved by using SSSD in inverse distance function to achieve matching [4], in this step, the complexity of the scene is computed for the control of adapting the camera system to the scene, which will be discussed in 2.2.3. After obtaining the disparity map from stereo matching, feature's position within the image for next frame processing could be expected by the following two equations:

$$\hat{x}_{fp_i,t} = x_{fp,t} - \frac{d_x(n_c)F}{z_{fp,t}} \quad (1)$$

$$\hat{y}_{fp,t} = y_{fp,t} - \frac{d_x(n_c)F}{z_{fp,t}} \quad (2)$$

where $(x_{fp,t}, y_{fp,t})$ is the coordinate of the point fp within the original image taken at time t, deep gray rectangle in Fig.2. d_x and d_y are the horizontal and vertical baseline distance between the camera which took the original image, and the camera whose image is to be expected. d_x and d_y are decided by the number of image pairs n_c taken at the same time. F is the camera's focus length. $z_{fp,t}$ is the depth of the point fp at time t.

After taking and processing the stereo matching and deduction for the next frame, feature matching between two consecutive frames is processed.

2.2.2 Image Registration and Motion Detection

For the sake that the expected image only contains feature's positions, the image registration is based on edge image level. This step can be achieved by adopting block matching [1].

Motion detection is processed by accounting camera's quasi-ego-motion which is computed by every two corresponding features. Here, the so called quasi-ego-motion is the value computed by assuming the feature is only moving in the depth direction. Obviously, if there are objects moving in not only depth information, value of the quasi-ego-motion could not be only one. Thus, motion is detected when the quasi-ego-motion generated by the feature points is the minor part in histogram of quasi-ego-motion. Here, we assume that the feature points of the static objects are occupying major area in the scene.

As described above, at each time interval, one expected edge image with depth information and one edge image with depth information could be computed. For each corresponding feature pair, a quasi-ego-motion is calculated by the following equations:

$$\Delta X_{t,fp} = \frac{1}{F}(z_{fp,t}x_{fp,t} - z_{fp,t+1}\hat{x}_{fp,t+1}) \quad (3)$$

$$\Delta Y_{t,fp} = \frac{1}{F}(z_{fp,t}y_{fp,t} - z_{fp,t+1}\hat{y}_{fp,t+1}) \quad (4)$$

where, $\Delta X_{t,fp}$ and $\Delta Y_{t,fp}$ are quasi-ego-motion. A description of (3) is given in Appendix.

After quasi-ego-motion expected, it is possible to calculate ego-motion by re-computing the quasi-ego-motion when the moving points are excluded.

2.2.3 Adaptation toward scene's complexity

By using SSSD in inverse distance function with fixed multiple baseline has the advantage of eliminate ambiguity while doing stereo matching. But, in the active camera system, the scene changes as the camera array moves. Especially in high speed moving object detection case, excess of the camera number taking picture at one moment will result in reducing detection speed. On the other hand, insufficiency camera pairs will lead false stereo matching. Consequently, a method of camera number decision by adapting toward scene's complexity is developed.

Here, we define the number of same pattern, or ambiguity within the image as the complexity of the scene. In [4], the SSSD in inverse distance function is defined as:

$$e_{\varsigma(12\ldots n)}(x,\varsigma) = \sum_{i=1}^{n} e_{\varsigma(i)}(x,\varsigma) = \sum_{i=1}^{n}\sum_{j\in W}(f_0(x+j) - f_i(x+B_iF\varsigma+j))^2 \quad (5)$$

where, f(x) is the function of the image, W is the dimensional searching window, B_i indicates the ith baseline distance, ς is the inverse distance defined by 1/Z, and, this model is assumed based on independent Gaussian white noise model.

It is showed, according to the experiment in [4], that the more complexity the scene owns, the more minimums the above function has. Based on this conclusion, the adaptation step is achieved by reducing the number the above function's minimum to one, by increasing the camera stereo pair number, and meanwhile, reduce the detection speed. This procedure can be achieved by an iterative way. First judge whether the number of the function's minimum is over one, if so, increase camera stereo pair for taking picture at the same moment, until decreasing the minimum number to one. In the next processing step, retain the stereo pair number while try to calculate the function's minimum with less information, if success, then decrease the stereo pair number and increasing the detection speed. These steps are shown as the following pseudo-code:

```
ADAPTATION()
If FUNCTION'S MINIMUN NUM > 1
  Then INCREASE THE STERO NUM
       ADAPTATION()
Else TRY TO COMPUTE WITH LESS INFO
  If SUCCESS
    Then DECREASE THE STERO NUM
         ADAPTATION()
  Else RETAIN THE STEREO NUM & QUIT
```

2.3 Extensibility and Constraint

In this camera system, the minimum of the camera is 2, general stereo pair. The processing ability, precision and also the detection speed could be improved by simply adding more stereo pairs and the micro processors with them. When increasing cameras, the whole system should be informed of the new camera matrix and each camera's number in order to control each of them.

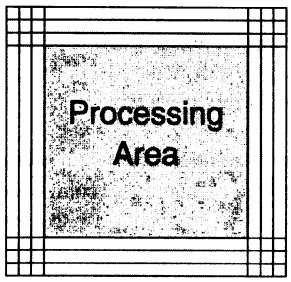

Fig.5, The processing Area is reduced when more cameras participate in the processing for one time interval.

Although we could acquire high-speed video by simply adding more cameras, it has two limitations. The first one is the processing area problem. As shown in Fig.5, the processing area is the overlapped part for all of the cameras in this system. However, increasing camera will result in processing area's reduction. Another limitation is the Processing Speed. In spite of adding stereo will also increase the micro processor with it, scaling up to a high detection speed will shorten the time for the central processor to calculate.

3. THE EXPERIMENT

We have done a simple experiment on

deducing the features coordinates by using equation (1) and (2). Fig.5. in the last page shows the images with features in circle. In this experiment, we use 4 cameras, or 2 stereo pairs, allocated in one line. In order to simplify the stereo matching, we adopt Harris Corner Detector to detect feature instead of edges. To get a good view, the images with titles, are showed as a (2,2) matrix. File names are ordered by the camera from right to left.

4. CONCLUSION AND FUTURE WORKS

In this paper, a new method of high speed motion detection based on active camera and a novel approach of motion detection are presented. With this approach, a detection extensible camera system can be achieved. Its attraction is the ability of automated adapting the trade-off between motion detection speed and the precision according to the complexity of the scene. And the system can be scaled up by simply adding more general cameras.

And now we are taking experiment on other parts of the approach and developing adopting other stereo matching method, such as DP(Dynamic Programming) [7], and other method described in [10], aims at computationally efficiency. And we are also thinking of developing a coarse to fine method by adopting Markov Chain with the camera array's moving information, and record object's position in the database, like human memory, in order to prepare for further computation.

APPENDIX

Different from ego-motion, quasi-ego-motion is calculated by assuming the feature point only has a movement in the depth direction. Here, only the motion in x direction is described, the motion in y direction is the same. So, we have:

$$\frac{z_{fp,t}x_{fp,t}}{F} - \Delta X_{fp,t} = \frac{z_{fp,t+1}x_{fp,t+1}}{F}$$

$$\Delta X_{t,fp} = \frac{1}{F}(z_{fp,t}x_{fp,t} - z_{fp,t+1}x_{fp,t+1}) \quad (3)$$

REFERENCE

[1] 羽下哲司，鷲見和彦，八木康史，“時間平均シルエットを用いた能動カメラによる人の追跡”，電子情報通信学会論文誌，vol.J88-D-II, no.2, pp.291-301, 2005

[2] Wan-Cheol Kim, Cheol-Ho Hwang, and Jang-Myung Lee, “efficient tracking of a moving object using optimal representative blocks” Proceedings of the 2003 IEEE/ASME

[3] Don Murray and Anup Basu, “Motion Tracking with an Active Camera” IEEE Transactions on Pattern Analysis and Machine Intelligence, VOL.16, NO.5, May 1994.

[4] Masatoshi Okutomi and Takeo Kanade, “A Multiple-Baseline Stereo” IEEE Trans. on Pattern Analysis and Machine Intelligence, VOL.15, NO.4, April 1993.

[5] Bennett Wilburn, Neel Joshi, Vaibhav Vaish, Marc Levoy, Mark Horowitz: “High-Speed Videography Using a Dense Camera Array.”, CVPR2004: 294-301

[6] Y.Satoh, T.Nakagawa, T.Okatani, K.Deguchi "A Motion Tracking by Extracting 3D Feature of Moving Objects with Binocular Cooperative Fixation", Proceedings of the 2002 IEEE/RSJ intl. Conference on Inteligent Robots and Systems, EPFL, Lausanne, Swizerland, October, pp.13-18, 2002

[7] Sven Forstmann, Yutaka Kanou, Jun Ohya, Sven Thuering, Alfred Schmitt, “Real-Time Stereo by using Dynamic Programming”, CVPR2004.

[8] Agrawal, M and Konolige, K and Iocchi, L. Real-time detection of independent motion using stereo, in Proceedings IEEE workshop on visual motion, 2005.

[9] Fabrizio Russo and Annarita Lazza, “Color Edge Detection in Presence of Gaussian Noise Using Nonlinear Prefiltering”, IEEE Transactions on Instrumentation and Measurement, VOL. 54, NO. 1, FEBRUARY 2005

[10] Myron Z. Brown, Darius Burschka, Gregory D. Hager, “Advances in Computational Stereo”, IEEE Transactions on Pattern Analysis And Machine Intelligence, VOL. 25, NO.8, AUGUST 2003.

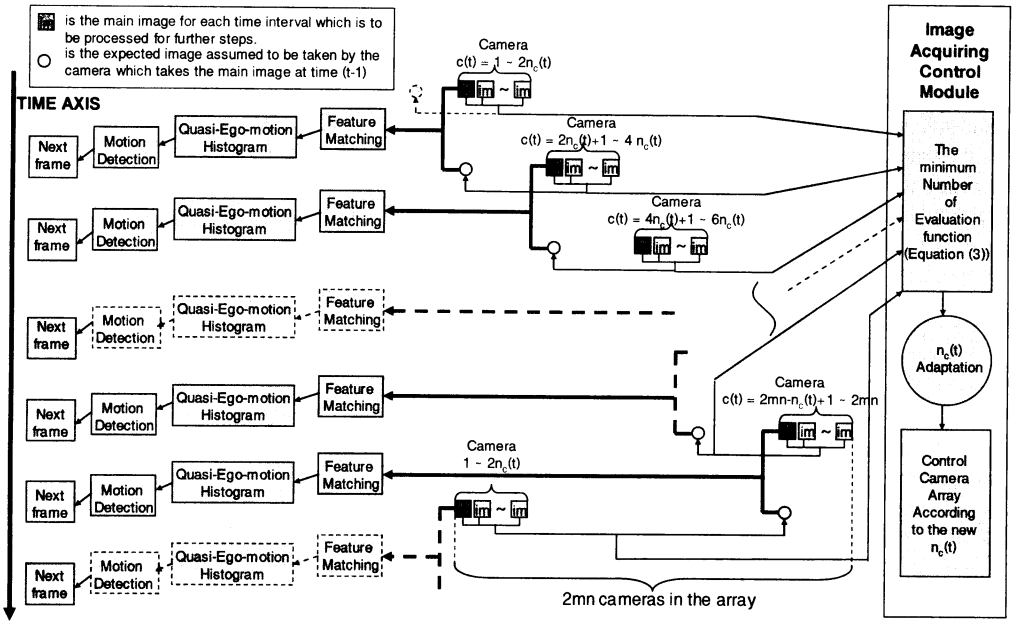

Fig.3 The time sequence relation between image taking and processing

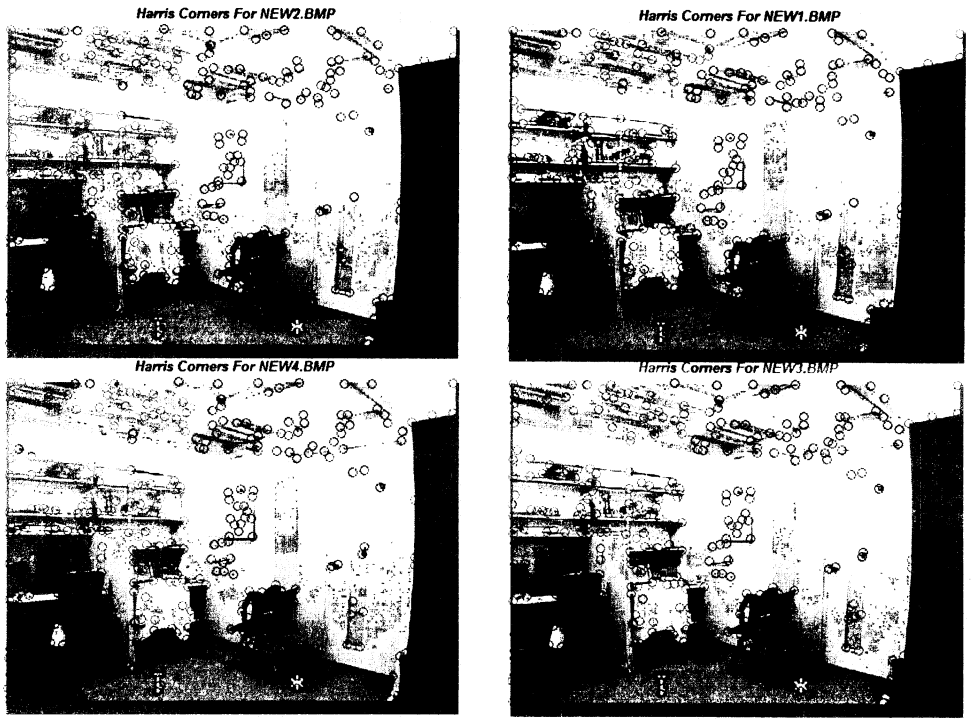

Fig. 6 Experiment on feature position deduction
The file name is inverse ordered by the camera matrix.