MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28–30, 2000, The University of Tokyo, Japan

7—5

A Video-Rate Stereo System by Integration of Two Algorithms with/without Occlusion Handling

Yasuyuki Sugaya *
College of Engineering Systems
University of Tsukuba

Yuichi Ohta †
College of Engineering Systems
University of Tsukuba

Abstract

This paper proposes a novel architecture for video-rate stereo machine. It is unfeasible to directly implement a video-rate machine with occlusion handling mechanism, because the computational cost for the correspondence search is higher than that of a simple polynocular stereo algorithm without occlusion handling. Our idea is to integrate two algorithms with/without occlusion handling mechanism in a framework of video-rate stereo machine. By switching two depth maps using motion information, the video-rate depth map is output for dynamic regions, while for static regions is given the dense depth map with sharp object boundaries. We have implemented a prototype system by using a commercially available video-rate stereo machine and a PC with an occlusion detectable algorithm on it.

1 Introduction

Video-rate stereo machines by using multiple-baseline stereo[1] have been developed[2] and some commercial models are available now. They can obtain dense depth map in video-rate. The depth map, however, is often blurred at the object boundaries because they use algorithms without occlusion handling mechanism. We have proposed occlusion detectable stereo algorithms[3][4][5] that can obtain sharp depth map at occlusion boundaries. The algorithms with occlusion handling mechanism, however, need much more computation than those without the mechanism and the direct implementation of a video-rate machine with those algorithms is unfeasible.

This paper proposes a solution for this problem. The idea is to integrate two algorithms with/without occlusion handling in a framework of video-rate stereo machine. For dynamic regions in a scene, the machine outputs the video-rate depth map obtained by an algorithm without occlusion handling mechanism, while for static regions it outputs the depth map obtained by an algorithm with occlusion handling mechanism. When objects do not move, dense and sharp depth map is always obtained. When objects are moving, video-rate depth map is obtained with no latency. Immediately after the objects stop, dense and sharp depth map appears with some latency. This scheme is quite reasonable for the occlusion presentation between the real and virtual objects in the mixed reality application, because an observer is not sensitive to the details of moving objects.

We have implemented a prototype system by using a commercially available video-rate stereo machine and a PC with an occlusion detectable algorithm on it.

2 Stereo system design

2.1 System overview

We have implemented a prototype stereo system by integration of two algorithms with/without occlusion handling in a framework of a video-rate stereo machine. The prototype stereo system consists of four modules. They are a dynamic region extraction module, two stereo modules with/without occlusion handling, and a depth map integration module.

- **The dynamic region extraction module:** The dynamic region extraction module produces dynamic regions in a scene.
- **The stereo module without occlusion handling:** The stereo module without occlusion handling mechanism estimates a depth map for the whole image. It operates on a video-rate stereo machine.
- **The stereo module with occlusion handling:** The stereo module with occlusion handling mechanism estimates the depth map for static regions that are obtained by the dynamic region extraction module. It operates on a PC.
- **The depth map integration module:** In the depth map integration module, two depth

*Address: 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573 Japan. E-mail: sugaya@image.esys.tsukuba.ac.jp

†Address: 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573 Japan. E-mail: ohta@image.esys.tsukuba.ac.jp

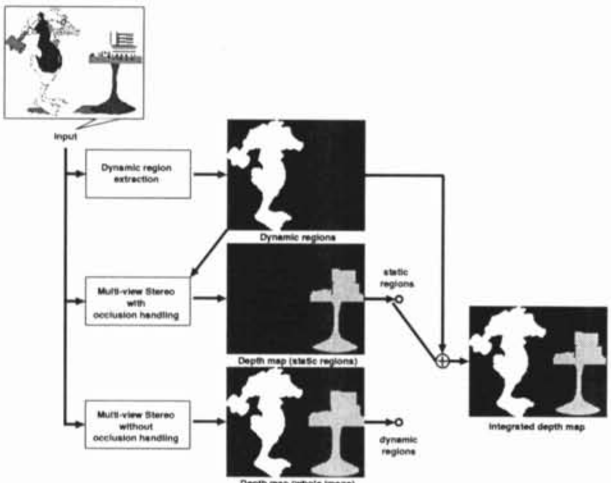

Figure 1: System overview

maps estimated by the two stereo modules are integrated by using dynamic/static information.

The system produces the integrated depth map in the following process as illustrated in Figure 1. The dynamic region extraction module produces dynamic regions by subtracting the background image from the input image (Figure 4.a,b,c). The two stereo modules are executed in parallel. The video-rate stereo produces the depth map for whole image (Figure 4.d), while the stereo module with occlusion handling is applied to the static regions obtained by the dynamic region extraction module (Figure 4.e). The depth map estimated by the occlusion detectable stereo is obtained with some latency, while that by the video-rate machine has no latency. For dynamic regions in a scene, the system outputs the video-rate depth map, while for static regions it outputs the depth map obtained by a PC with occlusion detectable stereo (Figure 4.f).

In the following sections, we describe the details of each module.

2.2 Dynamic region extraction

The dynamic region extraction module produces the dynamic regions in the image captured by the center stereo camera. The dynamic regions are extracted by subtracting the background image from the current image. The background image is captured at the beginning, and is modified at every certain period. The dynamic region extraction and the background modification process are as follows.

1. Subtract the background image from the current image. When the difference by the subtraction is higher than a threshold, the point is decided as dynamic. The threshold is determined by a heuristic method.

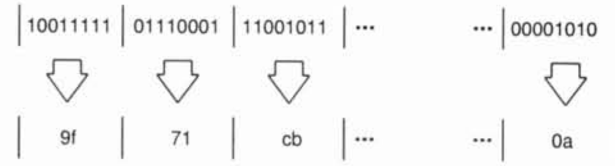

Figure 2: Data compression

2. Increment the counter assigned to each pixel decided to be static.

3. While a certain period, the processes described above are repeated. If a counter become higher than a threshold, the pixel is changed into background.

The result of extraction is represented by 0 or 1 (dynamic regions are 1 and static regions are 0) and every 8 pixels are packed into a byte shown Figure 2. The compressed data is sent to each module through the TCP/IP socket connections.

2.3 Stereo module without occlusion handling

This module uses a commercially available video-rate stereo machine and can produce a dense depth map for whole image in video-rate.

2.4 Stereo module with occlusion handling

This module is applied only to the static regions that are extracted by the dynamic region extraction module. The stereo module with occlusion handling mechanism produces depth map with sharp occluding boundaries by using one of the two occlusion detectable stereo algorithms. The first one is the mask-oriented algorithm[5] that uses "occlusion masks", which represent the typical occlusion patterns occurring in real scene. It is fairly effective to obtain a depth map with sharp occluding boundaries. However, it cannot cope with complex occlusions that are not defined in the "occlusion masks". The other one is the sort-oriented algorithm[6]. It can handle occlusions without "occlusion masks". The sort-oriented algorithm estimates the true disparity by using only the good evaluation values obtained from multiple stereo pairs. This algorithm is effective for complex occlusion patterns, while it is sensitive to noise compared to the mask-oriented algorithm. So in the area without occlusion, the sort-oriented algorithm often produces depth map with lower quality than the mask-oriented one. In order to use these algorithms effectively, we select two algorithms depending on the conditions of target scene.

2.5 Depth map integration module

This module integrates two depth maps obtained by two stereo modules with/without occlusion handling. For dynamic regions, the depth map obtained by the video-rate stereo machine is output without delay, while for static regions the depth map with sharp occluding boundaries is output with some delay. It should be noted that the delay has no problem in static regions.

3 Implementation of a prototype system

We have implemented a prototype video-rate stereo system as illustrated in Figure 3. The input stereo images are captured by the five cameras of the video-rate machine and sent to each module. The dynamic region extraction module and the stereo module with occlusion handling were implemented on PCs (Intel PentiumII 450MHz). The stereo module without occlusion handling is a commercially available video-rate stereo machine. The depth map integration module and data control process run on an SGI workstation.

- **The dynamic region extraction module:** The dynamic region extraction module captures the center stereo camera image and produces the compressed dynamic regions. The extracted dynamic regions are sent to an SGI WS through a TCP/IP socket connection. This module is implemented by using the MMX architecture of Pentium processor on a PC and can process 18 frames per second with 360×243 image size. The background image is modified about every 5 seconds.
- **The stereo module without occlusion handling:** We employed a commercially available video-rate stereo machine. It can produce dense depth maps at 30 frames per second. The depth map is sent to the SGI WS as digital video signal.
- **The stereo module with occlusion handling:** The occlusion detectable stereo module works on a PC. This module is applied to the static regions sent from the SGI WS through a TCP/IP socket connection. The depth map is sent to tne SGI WS in the same way. When the whole scene is static, i.e., when the module should process the whole image, it takes 5 to 8 seconds to produce the depth map of a frame with disparity search range 32. This will become much faster when we implement the algorithm by using the MMX architecture of the Pentium processor. Furthermore by using the occlusion information detected by the stereo module without occlusion handling, the area to which this expensive stereo module should be applied is considerably reduced[6].
- **The depth integration module:** The depth integration module switches two depth maps based on the dynamic/static information. The data control process controls the data sent from the dynamic region extraction module and from the stereo module with occlusion handling mechanism by using double buffers. These processes are implemented on the SGI WS.

The performance of the dynamic region extraction module or the stereo module with occlusion handling are not 30 frames per second. However, the total system can output the integrated depth map at 30 frames per second.

4 Conclusion

We have proposed a novel architecture for video-rate stereo machine and have developed a prototype stereo system by integration of two algorithms with/without occlusion handling mechanism on a video-rate stereo machine and a PC. The system outputs the depth map obtained by the video-rate stereo with no latency, while it outputs the depth map obtained by stereo module with occlusion handling with 5 to 8 seconds delay. The prototype system totally can produce the integrated 360×243 depth map at 30 frames per second.

References

[1] M.Okutomi, T.Kanade, :"A Multiple-baseline stereo", Trans. Pattern Anal. & Mach. Intell., vol.15, no.4, pp.353-363, 1993

[2] T.Kanade, A.Yoshida, K.Oda, H.Kano, M.Tanaka: "A Stereo Machine for Video-rate Dense Depth Mapping and Its New Applications",Proc. CVPR'96, pp.196-202, 1996

[3] K.Satoh, Y.Ohta: "Passive depth acquisition for 3D image displays", IEICE Trans. on Information and Systems, vol.E77-D,no.9,pp.949-957, 1994

[4] Y.Nakamura, T.Matsuura, K.Satoh, and Y.Ohta: "Occlusion detectable stereo -Occlusion Patterns in Camera Matrix-", Proc.CVPR'96, pp.371-378, 1996

[5] K. Satoh, Y. Ohta: "Occlusion Detectable Stereo -Systematic Comparison of Detection Algorithms", Proc. ICPR'96, pp.280-286, 1996

[6] Y. Sugaya and Y. Ohta:"Stereo by integration of two algorithms with/without occlusion handling", Proc.ICPR2000, pp.109-113, 2000

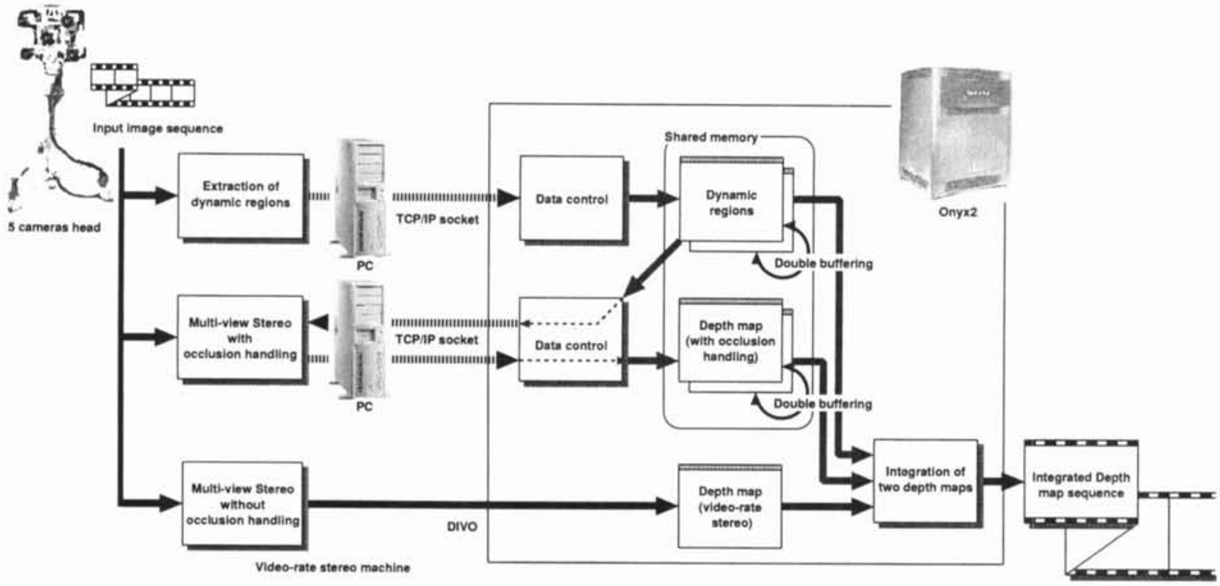

Figure 3: Prototype video-rate stereo system

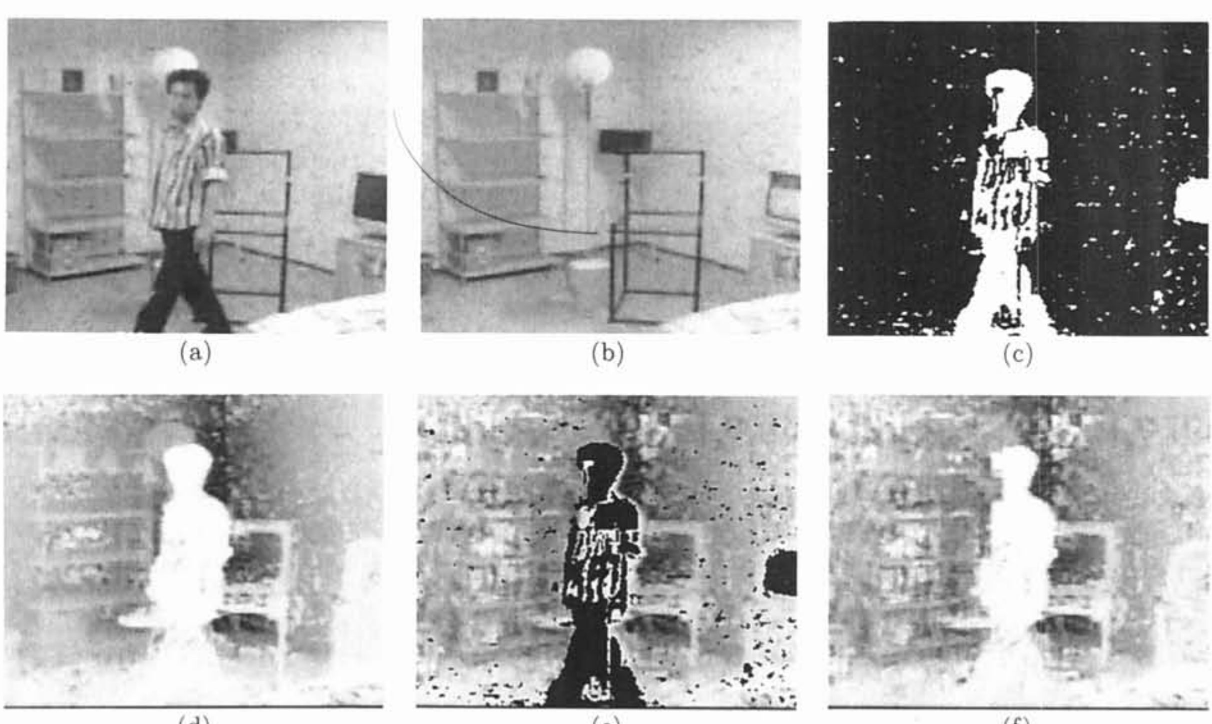

Figure 4: Experimental results

(a) input image (b) background image (c) extracted dynamic regions (d) depth map obtained by stereo without occlusion handling (e) depth map obtained by stereo with occlusion handling (f) integrated depth map