Proceedings of the
IEEE Intelligent Vehicles Symposium 2000
Dearborn (MI), USA • October 3-5, 2000

Reliable Stereo Matching for Highly-Safe Intelligent Vehicles and Its VLSI Implementation

Masanori Hariyama, Toshiki Takeuchi and Michitaka Kameyama

Graduate School of Information Sciences, Tohoku University
Aoba05, Aramaki, Aoba-ku, Sendai 980, Japan
E-mail: hariyama@kameyama.ecei.tohoku.ac.jp

Abstract

To realize highly-safe intelligent vehicles, high-speed acquisition of reliable three-dimensional(3-D) information is essential. Stereo vision is a well-known method for three-dimensional instrumentation. A major issue of stereo vision is to establish reliable correspondence between images. This paper presents a reliable stereo-matching algorithm based on SAD (Sum of Absolute Differences) computation. Reliable correspondence can be established by selecting a desirable window size of the SAD computation based on the uniqueness of a minimum of the SAD graph. A pixel-serial and window-parallel architecture is also proposed to achieve 100% utilization of processing elements. The performance of the VLSI processor is evaluated to be more than ten thousand times higher than that of a general-purpose processor.

1 Introduction

To realize highly-safe vehicles, high-speed acquisition of reliable three-dimensional information is essential. Stereo vision is a well-known method of three-dimensional instrumentation. One important problem on stereo vision is to establish reliable correspondence between images. Another problem is that the correspondence search is time-consuming even if using state-of-art general-purpose processors.

This paper presents a reliable stereo-matching algorithm and high-performance processor architecture suitable for it. In order to determine which pixel in one image(candidate image) matches a given pixel in another image(reference image), we consider a rectangular window centered at the pixel in reference image and compute an SAD for a window centered at each pixel in the candidate image. If the windows exactly matches each other, then the SAD becomes 0. The problem on SAD-based matching is that a window size for SAD computation must be large enough to avoid ambiguity but small enough to avoid the effects of projective distortions[1]. To solve this problem, a stereo matching method with variable window sizes is proposed. The method is based on an idea that an SAD graph has a unique and clear minimum at the reliable matching pixel. A desirable window size that gives the reliable matching pixel is determined at each pixel based on the uniqueness of a minimum of an SAD graph.

In SAD computation, parallelism between pixels in a window changes depending on its window size, while parallelism between windows is predetermined by the input-image size. Based on this consideration, a pixel-serial and window-parallel architecture is proposed to achieve 100% utilization of processing elements. The processing time of the VLSI processor is estimated to be 60*m*sec for input images of a size 512×512. Its performance is more than ten thousand times higher than that of the general-purpose microprocessor(Pentium II 400MHz).

2 Collision Warning System for Highly-Safe Intelligent Vehicles

Figure 1 shows a collision warning system for highly-safe intelligent vehicles. If dangerous situations are detected during driving, then the system makes warning to a driver, so that the driver can avoid an accident. First, input images are acquired by vision sensors. Next, obstacle recognition such as road extraction, human extraction and vehicle extraction are performed. Simultaneously, 3-D coordinates of each object are computed by stereo vision. Using these results, a trajectory of each object is predicted considering the past trajectory. Then, path planning based on collision detection is executed to detect dangerous states.

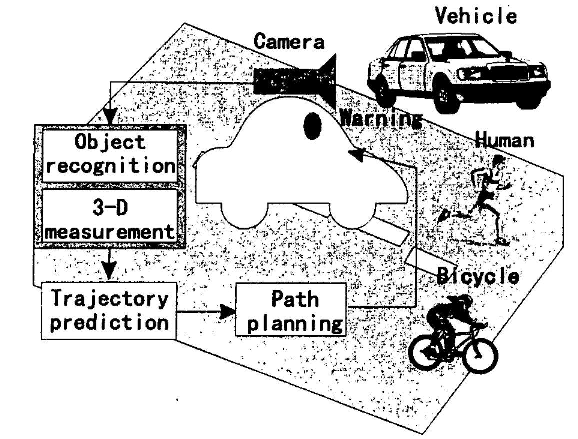


Figure 1: Processings for Highly-Safe Intelligent Vehicles.

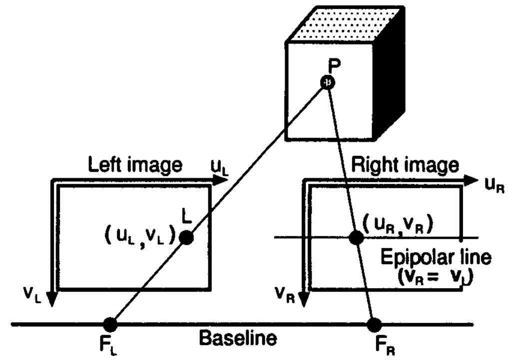


Figure 2: Camera geometry for the stereo system.

The collision warning system is required to quickly respond to a dynamic change of real-world environment. Since computational requirements of the processings exceed the computing power of general-purpose processors, development of special-purpose VLSI processors is desired. Several VLSI processors such as a path planning VLSI processors have been already developed[4]-[6]. One serious problem on the collision warning system is to acquire reliable 3-D information at high speed. Several approaches for high-speed stereo vision have been reported until now[2]-[3]. However, it is desired to compute a depth map of sufficient resolution and reliability.

3 Stereo Matching Algorithm

3.1 Basic Stereo Matching Algorithm

Figure 2 shows a camera geometry of a binocular stereo system. Two cameras with parallel optical axes are arranged along a straight line called a "*baseline*". Points F_L and F_L denote focal points for the left and the right images, respectively. Given a pixel $L(=(U_L, V_L))$ in the left image, let us find a 3-D point P which is projected onto L by perspective projection. It is mathematically guaranteed by imaging geometry that P lies on a line F_1L. The line F_1L is projected onto a line $U_L = U_L$ in the right image, which is called an "*epipolar*" line. The pixel onto which P is projected in the right image is called the corresponding pixel of L. It is also mathematically guaranteed that the corresponding pixel lies on the epipolar line. Once a pixel (U_L, V_L) on the epipolar line in the right image is determined as the corresponding pixel, the 3-D coordinates of P can be computed from the 2-D coordinates U_L, U_L and V_L by triangulation.

To establish the correspondence, a similarity measure must be computed which reflects how well the pixel L matches each pixel on the epipolar line in the right image. One commonly used similarity measure is a sum of absolute differences(SAD). As shown in Fig. 3, let us consider a reference window of a size $W \times W$ centered at $L(=(U_L, V_L))$ in the left image and a candidate window centered at (U_R, V_R) on the epipolar line in the left image. Then, an SAD in a window size W is given by

$$F_W = \sum_{j=-\frac{W-1}{2}}^{\frac{W-1}{2}} \sum_{i=-\frac{W-1}{2}}^{\frac{W-1}{2}} |I_L(U_L+i, V_L+j) - I_R(U_R+i, V_R+j)| \quad (1)$$

where I_L and I_R are intensity values in the left and right images, respectively. If a candidate window exactly matches the reference window, then the SAD becomes 0. Given a reference pixel L in the left image, an SAD is computed for each candidate pixel on the epipolar line in the right image, and an SAD graph is obtained as shown in Fig. 4(a). A pixel where the SAD graph has its minimum is called a "*matching*" pixel. In a straightforward method, a window size is empirically predetermined. The matching pixel in the window size is determined as the corresponding pixel.

The window size is an important parameter in SAD-based stereo matching. If the window size is too small, there exists several possibilities for the choice of the corresponding pixel. Therefore, the window size must be large enough to avoid the ambiguity. On the other hand, if the window size is too large and the window includes pixels whose depths in the scene are different

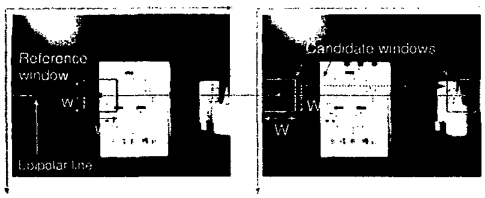


Figure 3: Search for a corresponding point.

from each other, the matching pixel may not be the corresponding pixel due to different projective distortions in the left and the right images.

3.2 Reliable Stereo Matching with Variable Window Sizes

To overcome the problem, variable window sizes for each pixel in the image are used to determine the reliable matching pixel. In the method , as small a window as possible that will still produce the reliable matching pixel is used. As shown in Fig. 4(a), an SAD graph usually has multiple local minima. Let Q_1^W be a matching pixel in a window size W. Let Q_2^W be a pixel where the SAD graph has the second smallest value of all the local minima. Then, a reliability measure R_W of the matching pixel is given by

$$R_W = \frac{(F_W(Q_1^W) - F_W(Q_2^W))}{W^2}.$$

The similarity measure is based on an idea that the matching pixel Q_1^W is reliable if the SAD graph has a unique and clear minimum at Q_1^W. Figure 5 shows a flowchart for determining the corresponding pixel using the similarity measure. First, from an SAD graph in a window size W, the matching pixel Q_1^W and the reliability measure R_W are computed(Fig. 4(a)). Next, it is checked whether the matching pixel Q_1^W is reliable or not. From an SAD graph in a window size $W+2$, the matching pixel Q_1^{W+2} and the reliability measure R_{W+2} are computed(Fig. 4(b)). Only if following two conditions are satisfied, then the matching pixel Q_1^W is reliable and determined as the corresponding pixel.

- The pixel Q_1^W is the matching pixel in both window sizes W and $W+2$.
- The reliability measure increases with the window size.

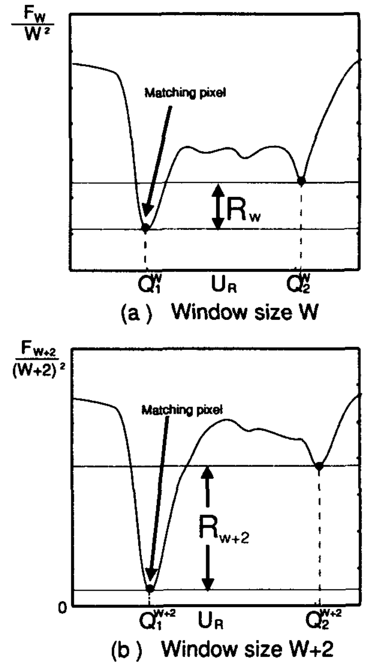


Figure 4: SAD graphs for window sizes W and $W+2$.

Otherwise, W is set to $W+2$, and this step is repeated until the window size $W+2$ is smaller than the maximum window size.

By using our method, a depth map of the front-parallel plane of Fig. 3 is produced as shown in Fig. 6. Its depth error is compared to that of the fixed-window-size method as shown in Fig. 7. The result clearly shows that our method selects reliable matching pixels.

Another advantage of our method is that it does not require any empirically-determined parameter. Therefore, it is suitable for vehicle applications in dynamically changing environment.

4 Pixel-Serial and Window-Parallel Architecture

In the search for the corresponding pixel, there exist pixel-level parallelism and window-level parallelism.

Window-level parallelism. SADs can be computed in parallel for all the candidate windows on the epipolar line as shown in Fig. 8. The number M of candidate windows on the epipolar line is fixed in advance by the input image width. Therefore, it is relatively easy to exploit the window-level parallelism as

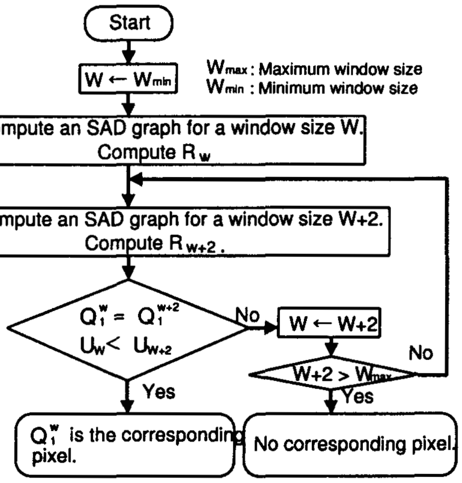


Figure 5: Flowchart of the stereo matching algorithm.

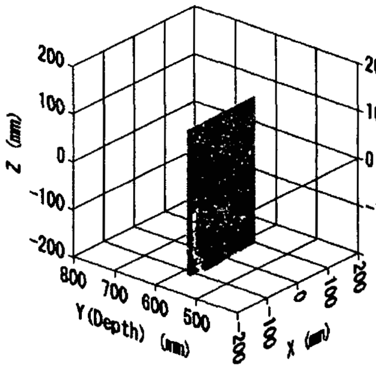


Figure 6: Depth map of the front-parallel rectangular plane.

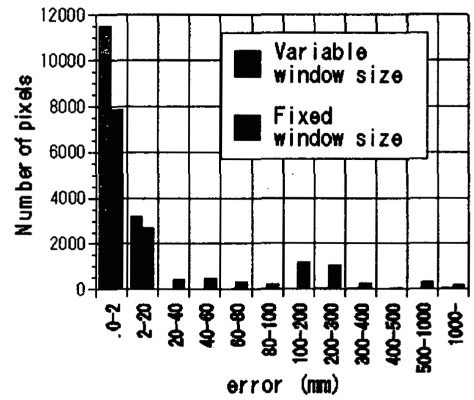


Figure 7: Evaluation of depth error.

described later.

Pixel-level parallelism. Absolute differences(ADs) in Eq. (1) can be computed in parallel for all the pixels in a candidate window as shown in Fig. 9 (a). The pixel-level parallelism is not suitable for the stereo matching with variable window sizes since the number of ADs computed in parallel is changed depending on the window size W. On the other hand, in a pixel-serial computation for an SAD, a single AD is computed in each step as shown in Fig. 9(b).

Based on the consideration on the parallelism, a pixel-serial and window-parallel architecture is proposed. As shown in Fig. 10, a stereo matching unit consists of identical n processing elements(PEs) to compute SADs for n candidate windows in parallel. Pixels in the right image are stored in a memory module ML. For parallel access, pixels in the right image are equally distributed among memory modules $MR_i (i = 1, \cdots, n)$. As mentioned above, since the number M of candidate windows on the epipolar line is fixed in advance by the input image, SAD computation for M/n candidate windows can be mapped in advance onto each PE so that all the PEs are 100% utilized. Fig. 11 shows a block diagram of the PE with a single AD unit and a single adder for summing ADs. The AD unit and the adder in the PE can be 100% utilized since an AD and an addition are computed in each step in the pixel-serial computation.

The major problem in designing the stereo matching unit is to find simple interconnection networks which supports fast and efficient communication. In the pixel-serial and window-parallel computation, ADs for the same pixel in a reference window can be computed in the same control step. This scheduling results in a simple shared bus to transfer a single pixel in each control step from the M_L to all the PEs. Moreover, if all the ADs for a candidate window are computed in the same PE, there is no need to transfer intermediate results from one PE to other PEs. By this functional unit allocation, complexity of an interconnection network between PEs is minimized. Finally, since only one pixel in a candidate window is performed in each control step as shown in Fig. 9(a), it is possible to store all the pixels in a candidate window in one memory module. By this memory allocation and the functional unit allocation, a memory module MR_i is

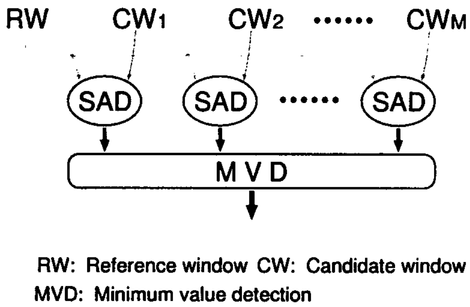
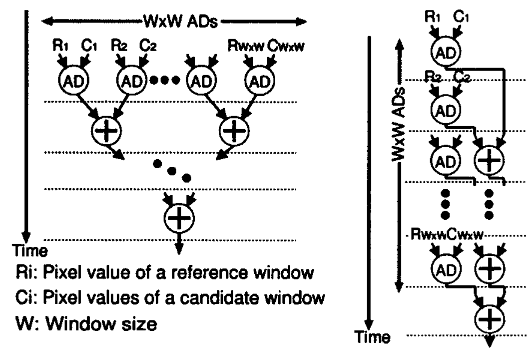


Figure 8: Data-flow graph of correspondence matching.



(a) Scheduled DFG of pixel-parallel SAD computation. (b) Scheduled DFG of pixel-serial SAD computation.

Figure 9: Data-flow graph of SAD computation.

dedicated to only one PE.

The stereo matching unit with 512 PEs is designed using a 0.5μm CMOS process and the performance is evaluated. The time required to produce a depth map estimated to be 60$m$sec for input images of a size 512 × 512. The performance of the VLSI processor is more than 10 000 times faster than the general-purpose processor(Pentium II 400MHz).

5 Conclusion

The variable-window-size method determines the reliable matching pixel using only SAD graphs. Therefore, to increase reliability of stereo matching, it can be easily combined with the multi-baselise stereo that uses a sum of SADs as a similarity measure[7].

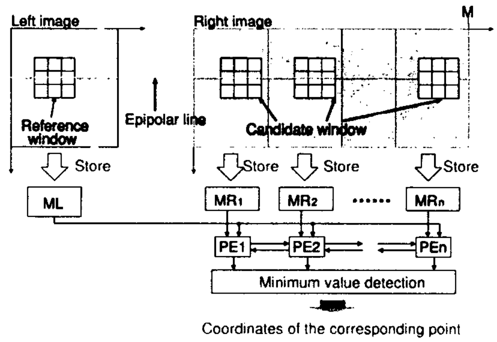


Figure 10: Block diagram of the stereo matching unit for the window-parallel computation.

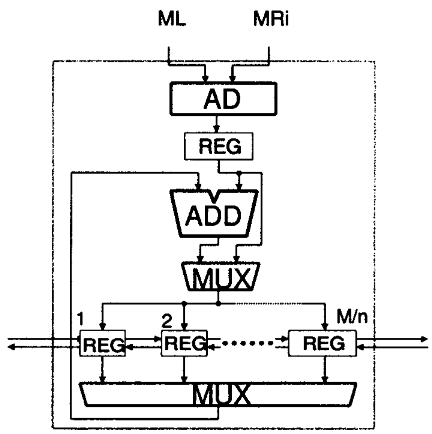


Figure 11: PE for the pixel-serial SAD computation.

References

[1] S.T.Barnard and M.A.Fischler, "Stereo vision, " in Encyclopedia of Artificial Intelligence. New York: John Wiley, pp.1083-1090, 1987.

[2] T. Ozaki, M. Ohzora and K. Kurahashi, "An Image Processing System for Autonomous Vehicle," in Proc. SPIE Vol. 1195 Mobile Robots IV, 1989.

[3] I. Masaki, S. Decker, A. Gupta, B.K.P.Horn, H-S. Lee, "Cost-Effective Vision Systems for Intelligent Vehicles," in Proc. Intelligent Vehicles Symposium, pp.39-43, 1994.

[4] M. Hariyama and M. Kameyama, "Collision Detection VLSI Processor for Intelligent Vehicles Based on a Hierarchical Obstacle Representation, " Proc. IEEE Conference on Intelligent Transportation Systems, 1997.

[5] Masanori Hariyama and Michitaka Kameyama, "Design of a Collision Detection VLSI Processor Based on Minimization of Area-Time Products", in Proc. IEEE International Conference on Robotics and Automation, pp.3691-3696,1998.

[6] S. Lee , M. Hariyama, M. Kameyama, "A Three-Dimensional Instrumentation VLSI Processor Based on a Concurrent Memory-Access Scheme," IEICE Trans. Electron, vol.E80-C, No.11, pp.1491-1498,1997.

[7] Masatoshi Okutomi, Takeo Kanade, "A Multiple-Baseline Stereo," IEEE Trans. PAMI, vol.15, no.4, 1993.