MVA2011 IAPR Conference on Machine Vision Applications, June 13-15, 2011, Nara, JAPAN

Stereo Vision System on Automobile for Collision Avoidance

Keiji Saneyoshi
Tokyo Institute of Technology
4259 Nagatsutacho Midoriku Yokohama Kanagawa Japan
ksaneyos@ric.titech.ac.jp

Abstract

There are currently on the market several kinds of sensors for automobile collision avoidance, including those using radar, LIDAR, ultrasonics, monocular vision, and stereo vision. To avoid collisions in a crowded traffic environment, an intelligent sensor must be used that not only detects the distances to obstacles, but acquires other relevant information, such as the areas occupied by the obstacles, the location of the traffic lanes, and the position and motion of other cars and pedestrians. Stereo vision is suitable for this application because of its wide field of vision, simultaneous detection of multiple objects, and ability to measure their sizes, positions, and relative velocities, as well as its ability to detect road shape and lane markings. Nevertheless, stereo vision also has several weak points: (1) the enormous amount of computation required, (2) the problem of mismatching, and (3) its vulnerability to the weather. We have overcome these problems through the use of several techniques: a new hardware system to address point (1), precise rectification for (2), and proper exposure control for (3). Our stereo vision system was first presented at the Tokyo Motor Show in 1991; at that time, the performance was 10 fps with a resolution area of 512 × 200 pixels and depth of 100 pixels. In 1999, the first on board stereo vision system for collision avoidance was put on the market. Recently, we developed a new stereo vision system whose performance was 160 fps with a resolution area of 1312 × 688 pixels and depth of 176 pixels. In this talk, I will present several stereo vision systems and their applications using demonstrations and videos.

1. Introduction

Recently, the number of deaths due to traffic accidents has decreased remarkably in Japan. Ten years ago, the annual number of deaths was more than 8,000 people, which by last year had decreased to less than 5,000. This is not due to improvements in people's driving technique but rather improvements in such safety equipment as safety belts, air bags, and anti-lock break systems (ABS), along with significant advances in emergency medical services. These examples are responsible for reducing deaths in the event of an accident, but not in reducing the accidents themselves. The fact that the number of traffic accidents has not much decreased is evidence for this conclusion. A generation ago, the main cause of traffic accidents was thought to be lack of safety consciousness on the part of the drivers. But the average frequency of causing an accident is only once in 65 years per driver, and that of death due to an accident is only once in 8500 years. Due to human limitations, it is quite difficult to stay consciousness of driving safety over such long periods. Thus, many efforts to prevent traffic accidents by changes in the automobile itself or the road infrastructure have begun.

One of the most important techniques for reducing accidents is detecting obstacles in the road. There are several systems already on the market which can perform such detection, for example those using a laser range finder, milli-wave radar, monocular vision, or stereo vision. In the following section of my talk, I will explain characteristics of these systems and show the superiority of stereovision.

2. Superiority of stereo vision

2.1. Amount of information

The amount of information in one picture is the number of pixels. In the first stages of digital image processing, the typical number of pixels of a digital camera was about 300,000 (VGA) and the camera was very expensive. Today, cameras with more than 1 million pixels are common, and cameras with more than 10 million pixels are easy to find. Considering that each pixel has independent information, it is a marvel that such a large amount of information can be obtained at one time. In the case of laser range finders, the largest amount of information for on-board obstacle detection is probably obtained by the HDL-64 manufactured by VELODYNE corp. [1]. This range finder is relatively expensive but can detect range for every 0.09 horizontal degree, which is comparable to a typical image sensor. Vertically, however, the finder can only detect every 0.42 degree, which is one order of magnitude less than that of the typical image sensor, and one-twentieth of the amount of information obtained by a 1 million pixel camera. Figure 1 shows the difference in the amount of information in the case of a scene including two vehicles at distances of 38 and 64 m. In the case of a laser range finder, the range data of the vehicle at a distance of 64 m consists of only four points. In the case of stereo vision, even the shape of the vehicle, as well as that of the tunnel entrance, can be obtained at a distance of 100 m. Therefore, in terms of amount of information, it is clear that stereo vision is superior to laser range finders.

It is natural to think that a large number of points are not necessary if we can get high-quality information from a low number of data points. In fact, a typical laser range finder can measure distance with a high accuracy of less

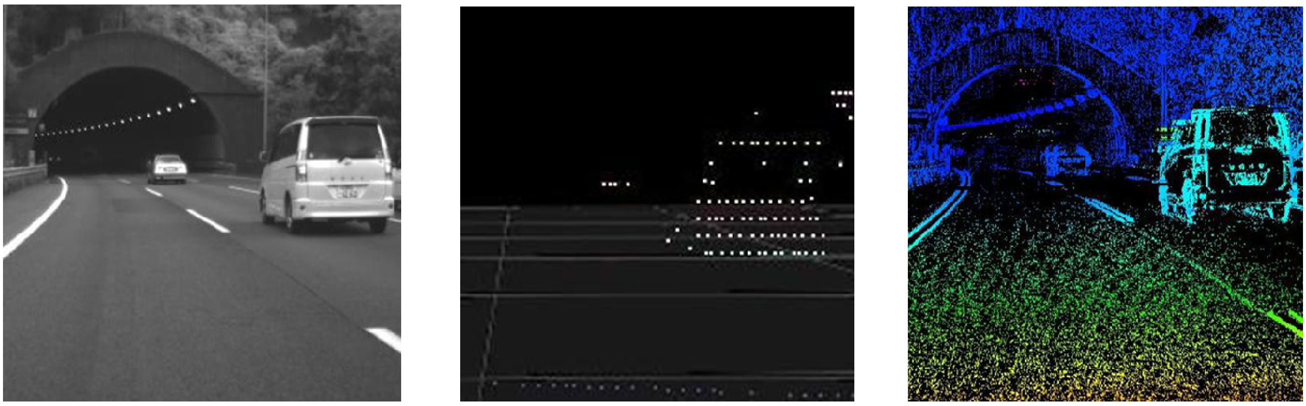


Figure 1. Comparison of the amount of information .
Each pixel has one piece of information.

than 10 cm, independent of distance. In contrast, for stereo vision, accuracy is inversely proportional to distance. As the result the accuracy becomes worse drastically at further distance. But I will show in chapter 5 that this weak point is not so serious problem for obstacle detection. Also the low number of data points for laser range finders is related to their low spatial resolution, which makes it impossible for them to detect objects with small projected area, such as poles, fences, or spokes of wheels. This is because a laser range finder detects surfaces, whereas a vision sensor detects edges of objects or patterns on surfaces. Thus, a vision sensor can detect a pole, fence, or spoke of a wheel very well.

One more advantage of vision sensors is its ability to detect the white lines on the road surface. A human driving a vehicle uses the white lines to distinguish the lanes of the road. If a wide road with several lanes was not separated by white lines, it would be very difficult to drive without colliding with other vehicles because of difficulty in predicting the motion of the other vehicles. The same is true for a collision avoidance system, and so the sensor must identify the white lines and other regulatory markings on the road. Only vision sensors have this capability.

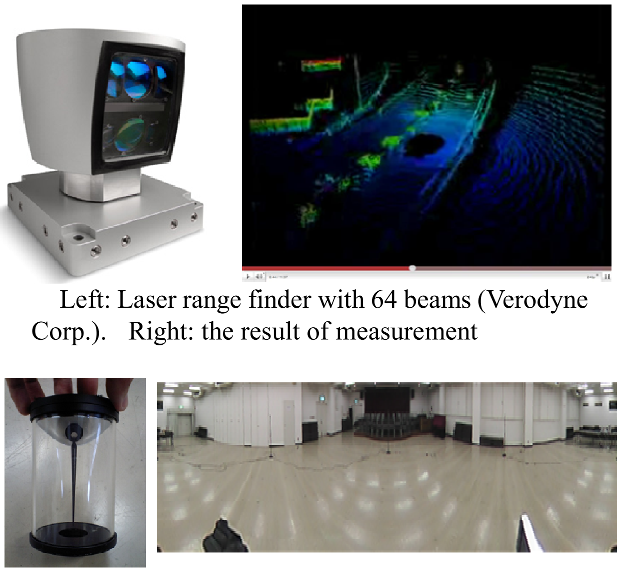
Stereo vision adds important information to that obtained with monocular vision, namely, depth distribution information is added to the brightness distribution information of monocular vision. This information changes the world from two dimensions to three dimensions. Vehicles or mobile robots move in the depth direction, so that the information on depth is essential. Travel using only monocular vision is quite difficult unless the environment is known, that is, there is a well-maintained map.

In the case of milli-wave radar, not shown in Figure 1, precise position information cannot be obtained for distant vehicles because of the large angle of divergence. The radar cannot distinguish between two vehicles in close proximity even if the beam is scanned. Thus, the milli-wave radar can be used only in places where the density of vehicles is low, such as on a highway.

2.2. Field of view

A mobile robot observes a path and judges whether it can proceed before moving. If the path bends, a sensor must have a wide field of view in order to observe as much of the path as possible. Furthermore, the sensor must be able to immediately detect any object which enters suddenly from the side and determine the possibility of collision from the object's relative position and velocity, which also requires the sensor to have a wide field of view. In addition, the sensor must check in all directions to know the motions of other vehicles around the sensor-equipped vehicle when preparing to change lanes.

A laser range finder scans a thin laser beam by rotating a mirror or the finder itself to guarantee a wide horizontal field of view and high spatial resolution. For a wide vertical field of view, some finders use a rotating polygon mirror and others use multiple beams distributed vertically. In either case, these sensors use mechanical systems to scan a beam. The disadvantage of a mechanical scanning system is not only the low durability but also the low information acquisition speed. Milli-wave radar also involves rotating the radar system itself. Another method using milli-wave radar which achieves scanning is a phased array radar using many radar elements. But this type of radar is very expensive and requires a large area to mount. In comparison, a hand vision sensor has a wide field of view of 40 to 50 degrees even using a standard lens. A wider field of view can be easily obtained by using a wide-angle type lens. If information covering all around the vehicle is needed, then an omnidirectional camera [2]



Left: Laser range finder with 64 beams (Verodyne Corp.). Right: the result of measurement

Left: Omni-directional monocular camera. Right: the result of measurement

Figure 2. examples of all-around of view senmay be used. An omnidirectional stereo camera can be constructed by using several omnidirectional cameras. Thus, in the case of vision systems, no mechanical movement is necessary.

2.3. Processing speed

Image processing was once considered a time-consuming task. In particular, stereo matching, as a typical processing step, was thought to make real-time stereo vision impossible to realize. To overcome this difficulty, a variety of ideas were tried, including CENSUS [3]. As a result, it is now possible to perform real-time stereo processing fairly cheaply while remaining gray-scale brightness information. This is dues to the further development of technology, such as that for FPGAs and CPUs. Indeed, in the case of FPGAs, processing speed of the stereo matching process has reached 275 fps for VGA size [4] and, more recently, 150 fps for 1312 × 768 pixels [5]. In the case of CPUs, processing speed reached 30 fps for VGA using a Pentium 4 (Intel Corp.), which is a previous-generation processor [6].

In contrast, the processing time of a laser range finder for one scan is 10 to 30 fps, which is not fast considering the low amount of points. High-speed processing is necessary to track objects moving quickly in the lateral direction, for example, the scenery when the vehicle turns quickly, or a person running into view suddenly from the side. In these cases, a processing speed of more than 30 fps is necessary for ordinary automobiles.

In the case of milli-wave radars, the processing time is about a few tens of milliseconds, which is not slow. However, the radar cannot detect lateral velocity because precise lateral position cannot be obtained due to the large angle of divergence. Therefore, achieving higher processing time is meaningless for this application. Recently, new milli-wave radars which can detect lateral velocity have been developed by installing two radar antennas on the left and right of the vehicle and angling them toward each other.

2.4. Influence of the environment

The influence of the environment depends on which wavelength is used as a probe. Figure 3 shows the relationship between wavelength and attenuation factor. The least attenuation factor for the atmosphere is achieved in the visible light band. Thus, visible light or a laser using near-infrared rays is a suitable probe for vehicles and mobile robots, to obtain vision similar to that of human beings. The attenuation factor for wavelengths of around 10 μm is reasonably good. Also, detectors sensitive to the 10 μm wavelength can detect a human form because the human body radiates infrared rays at this wavelength. In the case of fog, as shown in Figure 3, visible light or infrared rays are inadequate while the milli-wave band around 80 GHz has good characteristics. The milli-wave radar is often used for this reason. However, in most parts of the world, there are relatively few foggy days and milli-wave band loses this advantage under other bad weather conditions, such as in middle to heavy rain. Therefore, this advantage of milli-wave radar is severely restricted in time and space.

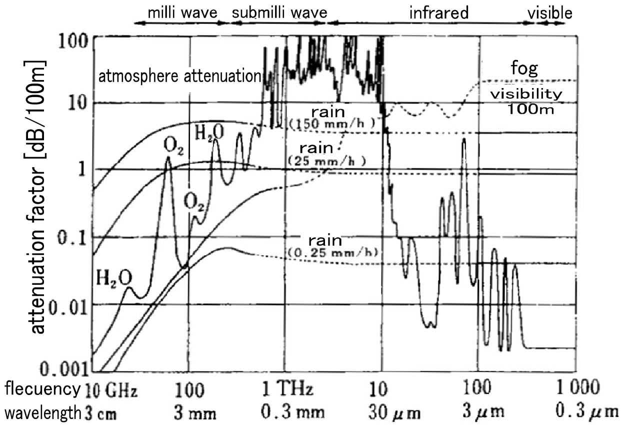


Figure 3. Relationship of wavelength of electro-magnetic wave and attenuation factor.

2.5. Summary

A comparison of the characteristics of laser range finders, milli-wave radar, monocular vision, and stereo vision is summarized in Table 1. Stereo vision surpasses other sensors in the most essential characteristics, such as field of view, ability to detect white lines, and horizontal resolution. The accuracy of the measured distance at a

Table 1. The comparison about monocular vision, stereo vision, laser range finder and milli-wave radar

	Monocular vision	Stereo vision	Laser range finder	Milli wave radar
>100m detection	×	○	◎	◎
Wide field of view	◎	◎	△	△
Accuracy	×	○	◎	◎
Spatial resolution	○	◎	△	△
White line detection	○	◎	×	×
Rain and snow	○	○	○	◎
Fog	△	△	△	◎
Night	○	○	◎	◎
Object dependency	○	○	△	△
Interference	◎	◎	△	△
Safety	◎	◎	△	○
Cost	◎	△	○	△

point 100 m distant by stereo vision is inferior to that by laser range finder or milli-wave radar, as is indicated by a circle in the corresponding columns. However, our experience has revealed that the precise measurement of distance at distant range is not very important, because in the case of distant objects, there is sufficient time to check whether a collision will occur and make appropriate adjustments to the vehicle. The columns concerning safety and interference have a double circle for stereo and monocular vision but a triangle for laser range finders and milli-wave radar because detectors emitting no radiation are the most safe and cause no interference at all. The cost performance of stereo vision may now be reasonable, but its absolute cost is still ten times higher than that of monocular vision or a cheap laser range finder.

3. Real-time stereo vision

3.1. History of stereo vision for vehicles

A review of the development of real-time stereo vision is summarized in the article [8], although it does not include our system. We started developing a real-time stereo vision system for automobiles in 1988 and demonstrated it at the Tokyo Motor Show at 1991, as well as publishing [9]. The system had two 640 × 480 pixel (VGA) cameras and could yield a disparity image every 0.1 second with a depth range of 100 pixels. This would have been the highest performance in the world at that time, according to the report in the article [8]. But our real challenge was just starting then because a system for collision avoidance must operate all the time and anywhere. Thus, we made a great deal of effort to make a system which could operate in rain, in snow, at night, during sunset or sunrise, and under extreme heat or cold. Finally, we put the first automobile with a collision avoidance system using stereo vision on the market in 1999.

3.2 Principle of stereo vision

The developed stereo vision system obtains the distance distribution as follows. As shown in Figure 4, two cameras with the same focal length are arranged horizontally with optical axes parallel to each other. Also, the horizontal axes of the left and right cameras coincide, so that the vertical coordinates of the projected points of any object are the same for the two cameras. The world space coordinates, X, Y, and Z, are obtained from the right camera coordinates of the point, x_r and y_r, and the disparity, $d = x_l - x_r$, by using the following equations:

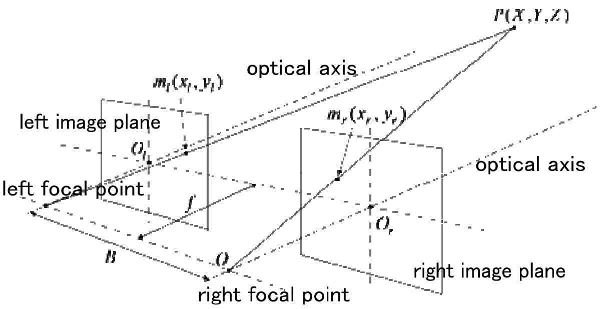


Figure 4. Arrangement of two cameras

$$Z = \frac{Bf}{d}$$

$$Y = \frac{Z}{f} y_r = \frac{B}{d} y_r$$

$$X = \frac{Z}{f} x_r = \frac{B}{d} x_r$$

where B is the base line, which is the length between the optical centers of the cameras, and f is the focal length. The origins of both the world space coordinate system and the right camera coordinate system are set to the optical center of the right camera.

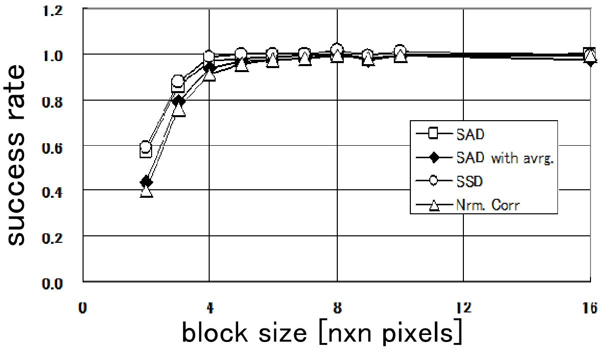


Figure 5. Relationship of matching size and success ratio in the case of adding noise of standard deviation of 3.

Disparity in an image is obtained using the block-matching method. We used a small block of size 4 × 4 pixels. This size was determined on the basis of the success rate of matching in an experiment. We prepared an image pair for stereo matching from a single image. The original image was used for the right image, while the left image was made by shifting the original image 20 pixels to the right and adding noise. A success rate could be obtained easily because the correct disparity was known to be 20 pixels. The results of the matching for various sizes of block are shown in Figure 5. Smaller blocks are better for spatial resolution, and so the smallest usable size was determined, which was 4 × 4 pixels, as shown in Figure 5. In the case of VGA size and a field of view of 40 degrees, the width of the block is equivalent to 0.5 m at a distance of 100 m, which is enough resolution to detect vehicles at that distance. Sum of absolute differences (SAD),

$$S(i,j,d) = \sum_{i,j \in block} \left| IL(i+d,j) - IR(i,j) \right|$$

is used for determining the similarity of the texture of blocks, where i and j are the coordinates of the image and d is disparity. $IL(i,j)$ and $IR(i,j)$ are brightness of the pixel at (i,j) in the left and right images, respectively. Disparity is defined as the value of d which minimizes S in the range of the search. The disparity image, or “depth image”, is obtained by performing this calculation over the entire image. An example is shown in Figure 6. The image is 2048 × 2048 pixels. The disparity is expressed by color;

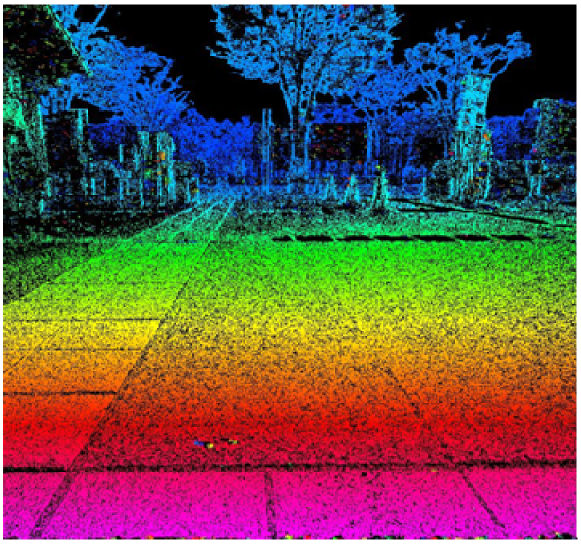


Figure 6. Disparity image: image size is 2048 x 2048 pixels. Disparity is expressed by color.

where large values (close objects) are indicated by red, which changes to yellow, green, and then blue as disparity becomes smaller (distance increases). The extraction of solid bodies from the image can be done quite easily by simply grouping pixels having similar disparities. Disparity is obtained only at the region where a pattern (brightness difference between adjacent pixels) is clearly observed, namely, no disparity data should be used where no pattern is observed. Such regions are indicated by black in Figure 6. In practice, if the brightness difference between a pixel and an adjacent one is less than some threshold value, then the data for that pixel is set to zero. Consequently, a clear disparity image with fine and precise edges, which is essential for collision avoidance, was obtained.

3.3 Distortion compensation and rectification

The search process within the stereo marching process can be performed quickly by using the epipolar constraint, namely, $y_r = y_l$, as shown in Figure 4, because the x direction (scan direction) is sufficient to find the same pattern in the two images. To get the benefit of using this constraint, we must compensate for the distortion in the original image and rectify the image pair. The limit of difference between y_r and y_l for the same object point was determined to 0.1 pixels to obtain a disparity image with high quality. Ordinal compensation methods, such as using a polynomial radial function, are not sufficient to achieve this limit. Therefore, a pixel-to-pixel compensation method using a look-up table (LUT) was adopted. A lattice pattern was used to obtain information about distortion. The intersection point was determined within the accuracy requirement of 0.1 pixels by calculating the positions of two intersecting lines rather than pattern matching the cross pattern around the intersection point. Compensation of pixels not contained in the intersection was done by linear interpolation. The interval of intersections had to be short enough to guarantee the accuracy.

The lattice pattern was taken simultaneously by two cameras for stereo vision. Consequently, seven parameters to rectify images were reduced to only two parameters, as is summarized in Table 2. The remaining two parameters were derived from the position differences between the left and right images for ten small regions with clear patterns and different distances. This method was clearly superior at reducing instability arising from the determination of parameters. A single LUT was made by combining the obtained rectification table with the distortion compensation table in order to shorten the processing time.

Table 2. Seven parameters to rectify images and known parameters after simultaneous distortion compensation of two cameras.

Translation to z axis (depth)	known
Rotation around y axis	Known
Rotation around x axis	Known
Aspect ratio	Known
Focal length ratio	Known
Translation to y axis	Unknown
Rotation around z axis	Unknown

3.4 Real-time operation

Real-time operation is essential to detect obstacles for on-board collision avoidance systems. Stereo processing was once considered a time consuming task. But now stereo matching and related processing can be performed in a short time thanks to remarkable advancements in CPU and FPGA technology. Figure 7 illustrates the flow of the process. The 10-bit digitized brightness signals from two cameras are sent to the addresses of the first LUTs by Camera Link cables. The frequency of the system clock is 80 MHz, and the two data signals for each camera are sent simultaneously. The LUT is used not only to balance the sensitivities of the two cameras but also to transform the signal data into logarithmic data of 8 bits to expand the effective dynamic range. The converted signals are stored in buffers temporarily. The size of the buffer is 128 lines and the second LUTs for distortion compensation and rectification extract the brightness data, so that the second LUT can compensate the distortion lying up to width of 64 lines.

The corrected data is stored in multiple buffers of four lines because several stereo matching units operate in parallel and the matching block size is 4 × 4 pixels. One stereo matching unit consists of 16 subtractors, 16 absolute value calculators and 15 adders, as shown in Figure 8. These calculators operate in parallel and are pipelined so that one SAD result is output every one clock cycle and a disparity value is obtained just at the end of a depth search. The disparity is examined using a peculiar point reduction filter: the filter operates on 3 × 3 pixel blocks to determine whether the disparity at the center is different by the threshold value from that of the surrounding disparities. The filter operates parallel to the stereo matching process.

4. Recognition

First, the road surface must be detected because everything that should be detected is near the road. Therefore,

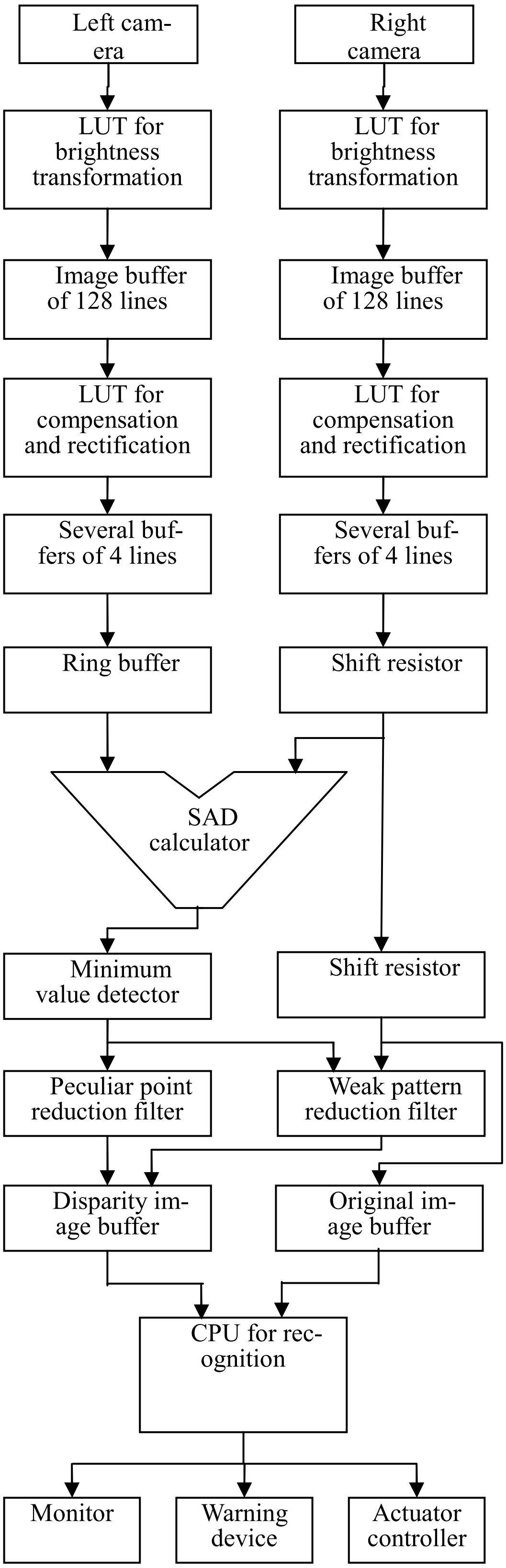

Left camera
Right camera
LUT for brightness transformation
LUT for brightness transformation
Image buffer of 128 lines
Image buffer of 128 lines
LUT for compensation and rectification
LUT for compensation and rectification
Several buffers of 4 lines
Several buffers of 4 lines
Ring buffer
Shift resistor
SAD calculator
Minimum value detector
Shift resistor
Peculiar point reduction filter
Weak pattern reduction filter
Disparity image buffer
Original image buffer
CPU for recognition
Monitor
Warning device
Actuator controller

Figure 7. Block diagram for real time processing.

the position of the road in the disparity image is very important. Then, white line markings and solid bodies are recognized from the obtained disparity image.

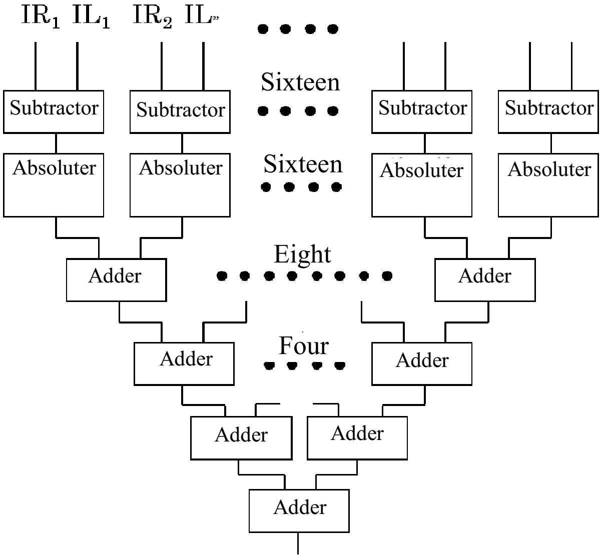


Figure 8. Block diagram of SAD calculator

4.1 Road surface detection

Recently, the noise level of image signals has decreased remarkably, so that disparity at a place with a weak pattern, such as a road, has become calculable, as shown in Figure 6. About one hundred disparity data points on the road surface region were selected, and a plane was fitted using the least squares method under the assumption that the surface neither inclined nor curved. This assumption is mostly correct for a highway. There was a possibility that the selected data included mismatched data and data for solid bodies above the road. Thus, the least squares method was applied twice. The disparity data far from the plane determined by the first least squares calculation were removed, and then the second least square calculation was performed for the remaining data. Although this method is not so robust compared to other methods, such as RANSAC [10], the processing time is short and strong robustness is not necessary, because most of the disparity data in front of the vehicle on a highway includes that of the road surface. Consequently, we chose the above-described method.

4.2 White lane mark detection

An original image was converted to an edge image with direction information using a Sobel filter. Then, the edge data around the road surface detected in the disparity image was extracted as shown in Figure 9. Combining the information in the original and disparity images is very simple because the data from the two images determine exactly the same position. The extracted edge data was further refined using the following conditions:

1) The width between the edge pair is from 12 to 25 cm,

2) the brightness between the edge pair is higher than outside of the edge pair, and

3) the width between the two pairs of edges satisfying the above conditions is from 3.4 m to 3.9 m.

Conditions 1 and 2 are based on the highway standards of Japan. These widths can be measured absolutely regardless of distance by using world coordinates obtained

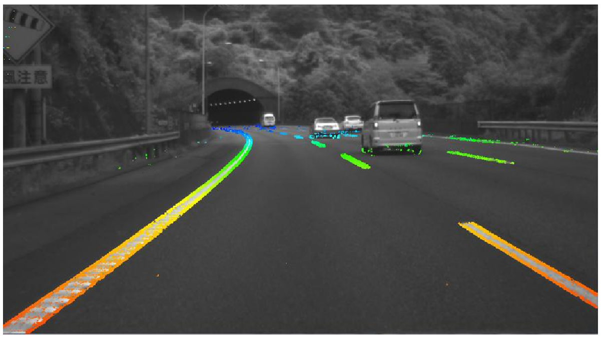


Figure 9. the edge data around the road surface detected in the disparity image

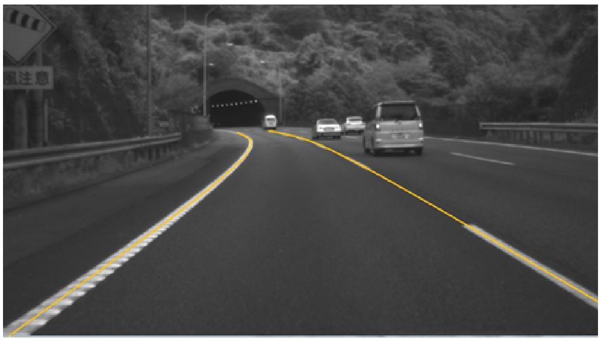


Figure 10. The result of white line detection

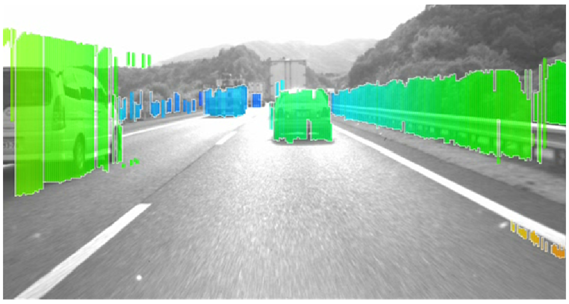


Figure 11. Rectangles placing to the positions which were calculated from the disparities given to the rectangles

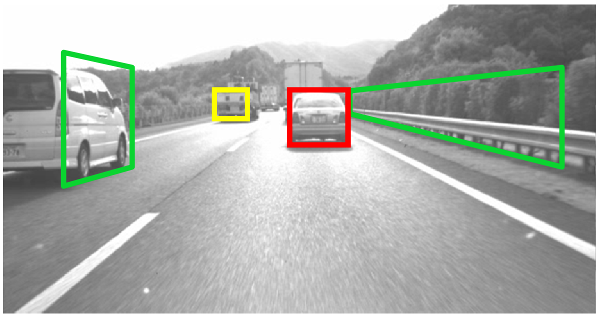


Figure 12. Result of recognition of solid bodies

from the disparity information. The brightness in condition 2 is examined by using edge directions of the edge pair.

The selected edge data is projected onto a horizontal plane in world coordinates as a lane boundary. Data satisfying only conditions 1 and 2 is projected onto the same plane taking dashed lines into consideration. Curved lines are fitted on dense data groups near the vehicle. The lines are expanded to distant according to the directions of the lines and modified if edge data was found. It is not necessary to consider drastic changes in direction because the curvature of the world coordinates of a lane does not change drastically. If one side of the lane is bordered by a dashed line, a predicted line is produced from the solid line on the other side. The result of white line detection is shown in Figure 10. Consequently, the lane bordered by smooth curved lines is obtained with sufficient robustness by using disparity information.

4.3 Solid body detection

Solid body detection uses the property that the disparities of a thin pole standing perpendicularly on the ground are almost constant from the top to the bottom. The disparity image was divided lengthwise into thin rectangles. The width of the rectangles was the same as that of the matching block in stereo matching. A frequency distribution of the disparity within the rectangle was taken and the disparity at the maximum frequency was assigned to that rectangle. More specifically, disparities at the maximum frequency and adjacent disparities were averaged with sub-pixel precision. Solid bodies standing perpendicularly on the ground could be assumed on the basis of groups these rectangles. Figure 11 illustrates this fact. This figure was drawn by placing rectangles to the positions which were calculated from the disparities given to the rectangles. It was clear from the figure that a solid body could be recognized as a group of rectangles. Thus, we detected a solid body by grouping adjacent rectangles with similar disparities. The disparity image of an automobile was very often separated into left and right groups because there are few perpendicular edges inside the outline of an automobile body. Thus we fitted a plane to a small area where a weak pattern appeared using disparities at both side of the area. The result of detecting solid bodies is shown in Figure 12.

4.4 Evaluation of a collision risk

We chose the closest vehicle located in the same lane and in front of the sensor-equipped vehicle as that which must be paid attention to by using the information on white lane marks and solid bodies described above. Then, the risk of collision with the vehicle was estimated.

Stereo vision cannot measure a great distance with high accuracy. For example, a typical stereo vision with Bf value of 300 pixel-meters for outdoor use results in a disparity of 3 pixels for the distance 100 m. The accuracy of disparity corresponds to plus or minus 0.25 pixels at most. Thus, the obtained value of 100 m from a disparity of 3 pixels actually means from 92.3 to 109.1 m. A relative velocity between the front (i.e., leading) and sensor-equipped vehicles is essential to the estimation of collision risk. It might seem impossible to calculate relative velocity using a distance with such poor accuracy. However, in reality, the accuracy can be improved by using disparity data obtained before collision because the great distance means that there is sufficient time to obtain large amounts of data before a collision can occur. Also, there is sufficient time to make a correct deceleration. This fact will be proved in the following calculation.

First, the rate of disparity change was obtained using the least squares method with *n* times series disparity data. The time period was determined as the period during which disparity increased by one. Then, the obtained rate of disparity change, *dd/dt*, was transformed into a relative velocity, *v*, by the following equation:

$$v = -\frac{Bf}{d(d+1)}\frac{dd}{dt}$$

The propagation of errors by the least squares method was expressed by the following equations:

$$\sigma_a = \frac{\sigma'}{\Delta t\sqrt{\sum_{i=1}^{n} i^2}}$$

for disparity change, and

$$\sigma_b = \frac{\sigma'}{\sqrt{n}}$$

for disparity itself.

Here *Δt* was the inverse of the frame rate, for which 30 fps or 150fps was used in this calculation. The error distribution of one measurement was assumed as Gaussian with standard deviation *σ'*. The error accompanying disparity was estimated to be plus or minus 0.25 pixels for each measurement. Thus, this value was used for the standard deviation. The propagated error of the disparity change was also transformed to that of the relative velocity.

As the vehicle becomes closer, the number of data, *n*, decreases. To avoid accuracy worsening, *n* was fixed to 4 if *n* decreased to less than 4, which corresponds to an ordinary measurement with constant measuring time period. The number of 4 was determined by try and error.

The distance, z_e, and the relative velocity, v_e, used for the prediction of the collision were calculated using twice the standard deviation as follows:

$$z_e = -\frac{Bf}{d + 2\sigma_b}$$

$$v_e = -\frac{Bf}{(d + 2\sigma_b)(d + 2\sigma_b + 1)}\left(\frac{dd}{dt} - 2\sigma_a\right)$$

These values were determined as the worst case of the measured data, namely, further away and slower comparing the real distance and velocity. The calculation was applied to an automatic collision avoidance system based on braking. The deceleration is always calculated even if the relative velocity is very low. The system starts to brake with the calculated deceleration when the deceleration passes the limit 0.4 G. If the calculated deceleration is less than 0.4 G, braking would interfere with the normal operation by the human driver. The deceleration was renewed every measuring time period as described above. If the calculated value of the deceleration exceeds 0.8 G, the real deceleration was restricted to 0.8 G for the safety of the driver and passengers.

Two results of the calculations are shown in Figures 13 and 14. Figure 13 shows the result under typical stereo vision, namely, Bf is 300 pixel-meters, the frame rate is 30 fps, and the initial velocity is 100 km/h. It was assumed that the system could identify the leading vehicle at a distance of more than 120 m, corresponding to a disparity of 2.5 pixels. When a vehicle approached within 120 m, the system started measuring the relative velocity and obtained the relative velocity at a disparity of 3.5 pixels, or a distance of 85.7 m, and the next measurement started simultaneously. The number of data used for the calculation was 38, which decreased the standard deviation to about one-sixth that for one measurement. The obtained relative velocity was 25.1 m/s, which was 10% less than

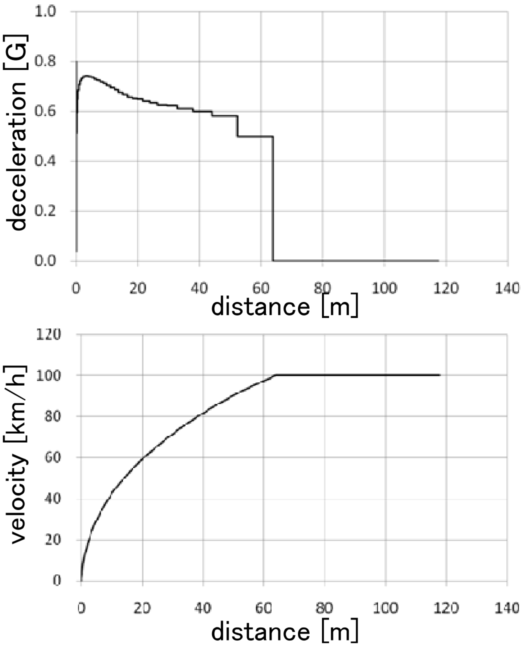


Figure 13. Calculated deceleration (upper) and velocity (lower) under the conditions: Bf = 300, 30fps, initial velocity is 100km/h

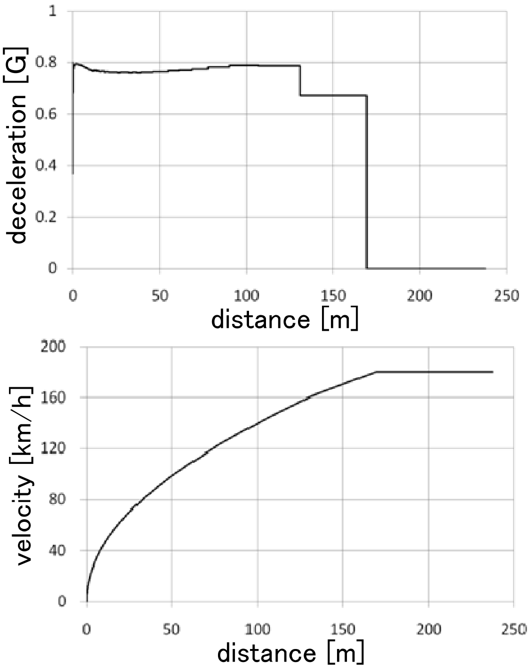


Figure 14. Calculated deceleration (upper) and velocity (lower) under the conditions: Bf = 600, 150fps, initial velocity is 180km/h

the correct velocity of 27.8 m/s because of underestimation from using the data of the worst case. Therefore the calculated deceleration was also underestimated. The accuracy of the measured relative velocity as well as that of the measured distance became gradually better because of repeated measurement. Thus, the calculated deceleration increased gradually to control the relative velocity and distance. In this case, the sensor-equipped vehicle stopped at a distance of 2.8 cm from the leading vehicle and no collision occurred.

Figure 14 shows the result using an advanced stereo vision system that we developed, namely, Bf is 600 pixel-meters and the frame rate is 150 fps. We increased the initial velocity to 180 km/h. The measurement of the relative velocity was at a distance of 240 m and the result was obtained at a distance of 171 m. The obtained relative velocity was 47.5 m/s, which was 2.5 m/s less than the real velocity, 50 m/s. The number of data used for the calculation was up to 206, so that the relative accuracy was twice that of a typical stereo vision system. In this case, the sensor-equipped vehicle also stopped, this time at a distance of 0.3 cm from the leading vehicle, and no collision occurred.

Thus, it was proved that high accuracy for one distance measurement at a great distance is not necessary. In practice, a human driver will avoid a collision with the leading vehicle in such a high-speed situation by steering rather than by braking because the width of the vehicle was so small relative to the distance. The starting point for taking action by steering is much closer than that for braking. But if the solid body in question were a long wall, then collision avoidance by braking would still be necessary.

5. Summary

One of the most important techniques for collision avoidance is to detect obstacles on the road. A stereo vision is one of the most suitable techniques for this purpose because of its large amount of information, wide field of view, high processing speed and low influence of environment. The first real time stereo vision for obstacle detection on an automobile was demonstrated in 1991. The system was put on the market at 1999. Recently we have developed a new high speed and high resolution real time stereo vision system. This system executes most of stereo process including stereo matching and rectification by FPGA thanks to the remarkable development of FPGA. The system achieved 160fps with the size of 1312x688 pixels and depth length of 176 pixels by using parallel and pipeline processing technique. White lane marks and solid bodies are detected robustly using information obtained both the disparity image and original image.

To check the ability for an application to the automatic braking system the calculation was performed. As the result even a typical specification of stereo vision such as VGA size, 30fps and Bf of 300 pixels times meters could detect a relative velocity of 100km/h within the error of 10% at the distance of 85.7m. Consequently the sensor-equipped vehicle stopped without collision by braking with deceleration of within 0.8G. This result indicates that the weak point of stereo vision, namely, the low accuracy at a long distance is not so serious problem for application to the collision avoidance system.

It is natural and essential to recognize traffic environment using the information of depth direction because an automobile moves to depth direction. The three-dimensional recognition using two eyes is the high grade method which animals obtained finally after a long process of evolution. Our stereo vision was inspired from this superior method. In the near future new age will come where many kinds of robots working in cooperation with human being. I believe that a stereo vision will be the most widely used sensor for three-dimensional recognition.

Acknowledgement: I express my gratitude to Fuji Heavy Industries Co., Ltd for giving the opportunity to study a stereo vision in early 1990's when most of researchers had thought impossible to use a stereo vision on an automobile for obstacle detection. The part of this work was supported under "Development of Energy-saving ITS Technology" project of New Energy and Industrial Technology Development Organization.

References

[1] http://www.velodyne.com/lidar/hdlproducts/hdl64e.aspx

[2] http://www.vstone.co.jp/products/sensor_camera/spec.html

[3] R. Zabih and J. Woodfill, "Non-Parametric Local Transform for Computing visual correspondence" Proc. Third European Conf. Computer Vision, pp. 150-158, 1994

[4] C. Georgoulas, L. Kotoulas and G. Siracoulis, "Real-time Disparity Map Computation Module", Microprocessors and Microsystems, 32 (2008) pp. 159-170

[5] K. Saneyoshi, H. Iwata and K. Oshida "High Resolution Real Time Stereo Camera for Automobile", Proc. 2010 JSAE Annual Congress (Spring), 20105347, May, 2010

[6] T. Tanzawa, Yamanashi Univ., Private Communication. 2001

[7] http://www.fujitsu-ten.co.jp/release/2009/03/20090327_01.html

[8] Myron Z. Brown, Darius Burschka, Gregory D. Hager, "Advances in Computational Stereo", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 25, N0. 8, pp.993-1008, 2003

[9] K. Saneyoshi, K. Hanawa, , "Image Recognition System for Active Drive Assist", Proc. INt. Symp. AVEC, 1992, Sep. Yokohama, Japan, p280 (1992)

[10] Sunglok Choi, Taemin Kim, and Wonpil Yu, "Performance Evaluation of RANSAC Family". In Proceedings of the British Machine Vision Conference pp. 1-12, 2009