Attraction of In-Vehicle Stereo Camera

Toshimichi Takahashi

Keywords ADAS, AD, Image recognition technology

Abstract

Stereo cameras are installed in automobiles as external vision sensors for Advanced Driver Assistance Systems (ADAS) and Autonomous Driving (AD). Compared to ultrasonic sensors, radio wave radars, Light Detection And Ranging (LiDAR), and monocular cameras as external world vision sensors for the same purpose, stereo cameras have higher object detection capabilities and provide useful information for collision determination. Stereo cameras have excellent features for automotive use. In order to maximize the performance of in-vehicle stereo camera, we have developed key technologies for practical use that take real-time processing into consideration. In particular, we are working to improve the accuracy and efficiency of correction and calibration, which is a major problem of stereo camera. We are also working on building a system to evaluate the safety of autonomous car using stereo cameras.

1 Preface

At the Validation Method for Automated Driving (VMAD), which is affiliated with the Subcommittee on Automated Driving (GRVA) of the United Nations Economic Commission for Europe's World Forum for Harmonization of Automotive Standards (WP29), virtual testing using simulation and other methods is being discussed. Among virtual testing, in particular Vehicle-In-the-Loop Simulation (VILS), what is demanded in the market is a system that can evaluate the safety of a fully completed autonomous car without having to modify it. Therefore, in both outdoor and indoor test facilities, we are working on building a system for evaluating the safety of autonomous car using stereo camera.

The elements of recognition, prediction, judgement, and car operation are essential for Advanced Driver Assistance Systems (ADAS) and Autonomous Driving (AD) in automobiles. Cameras are used as sensors at the core of vision technology. In recognizing objects as 3D, there are several possible ways to do this. Monocular camera system such as perspective, shading, and out-of-focus methods used in photography require the assumption that the object is a three-dimensional object in order to recognize a 3D object from a two-dimensional image. On the other hand, with binocular stereopsis, 3D information can be directly obtained by adding one more eye.

Many animals have two eyes, but binocular stereopsis is limited to cats and primates such as humans. You can experience 3D information immediately by doing the following: If you hold one finger in front of you and look into the distance, the finger will appear to be separated into two. This is because the position of the finger reflected on the retina is different between the left and right eyes. The distance between these two fingers is called parallax. When you bring your fingers closer together, the distance between them increases and the parallax increases. Conversely, when you move them farther apart, the parallax decreases. The brain measures this parallax and obtains a sense of distance from it. The overall shape of a three-dimensional object is recognized by measuring the parallax of various parts of the object. This paper introduces the methods and features of artificially achieving binocular stereopsis, including current initiatives.

2 Stereo Method

To artificially achieve binocular stereopsis, it is important to measure parallax accurately and efficiently. Fig. 1 shows parallax and distance. The relationship between the parallax D and the dis-

tance Z to the target object is simply expressed as an inversely proportional equation.

$$Z = \frac{Bf}{D} \quad \cdots\cdots (1)$$

$$D = u_c - u_b \quad \cdots\cdots (2)$$

Where,

Z: Distance to the target object

B: Distance between cameras

f: Camera's focal distance

D: Parallax

u_b: Horizontal coordinates of the reference camera imaging plane of the target object

u_c: Horizontal coordinates of the comparison camera imaging plane of the target object

Since the distance B between the left and right cameras and the focal length f are known constant values for each stereo camera (Bf), once the parallax D is found, it is possible to calculate the distance Z to the object.

The proportionality constant Bf is expressed as the product of the distance B between the cameras and the focal length f of the camera. To facilitate the conversion between distance and parallax, distance is expressed in meters, parallax in pixels, and Bf values are expressed in meters/pixels.

2.1 How to Find Parallax

To find the parallax of a three-dimensional object in an image, find the same small region of the object in one image (hereinafter referred to as "reference image") in the other image (hereinafter referred to as "comparison image") and compare it. Find the difference in coordinates. This method uses pattern matching, which is a type of image processing. The simplest and most commonly used method is to calculate the difference in brightness for each pixel at the same position in the two regions being compared, and then add the absolute values over the entire region to obtain an evaluation value. This evaluation value is called Sum of Absolute Difference (SAD). If the patterns are the same, the brightness at every position will also be the same, so the SAD value will be 0. The less similar the patterns are, the larger the SAD value becomes. The parallax is the deviation from the reference image in the area within the search range where the other comparison image is searched for the area of the reference image and has the smallest SAD value.

2.2 Search Range

Fig. 2 shows the parallel search direction and scanning direction. This is the ideal camera arrangement for searching for actual comparison images. In this arrangement, the focal lengths of the two cameras are the same, the optical axes are parallel to each other, the two imaging planes are perpendicular to the optical axis and on the same plane, and the scanning directions of each imaging plane are two different. This coincides with the direction connecting the optical centers. This arrangement is called a parallel and equal arrangement. Parallel and equal positions can be found by one-dimensionally scanning the same height of the corresponding comparison image in any area of the reference image. If they are not parallel, the search direction and scanning direction do not match, and complex processing is required. For this reason, in-vehicle stereo cameras that require high-speed performance generally use parallel and co-located cameras.

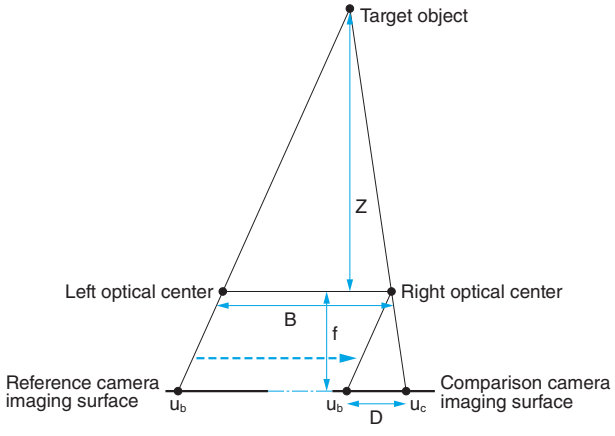

Fig. 1 Parallax and Distance

The relationship among distance to the object, position appearing in the camera, and parallax is shown.

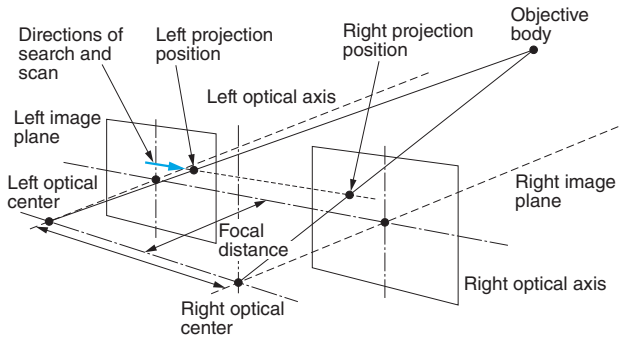

Fig. 2 Parallel Search Direction and Scanning Direction

The relationship between direction of search and that of scan in the parallel and coordinate layout is shown.

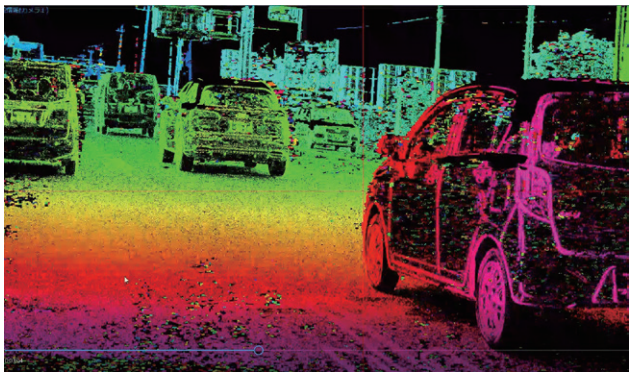

Fig. 3 Parallax Image Taken with 4K Stereo Camera

From a large parallax, color is changed from red to yellow, green, cyan, and blue. Accordingly, the red color corresponds to a nearer position.

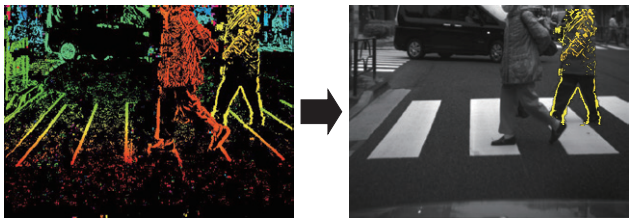

Fig. 4 Object Extraction from Parallax Image

By making parallax grouping based on the parallax image, a pedestrian on the right side can be extracted without any knowledge and assumption. Accordingly, this method is more reliable compared with monocular camera which is a competing technology of a stereo camera.

The generated image is a parallax image that has not yet been converted into distance. Conversion to distance involves division, which increases the processing load. For this reason, noise removal, deletion/correction of mismatched parallaxes, and subsequent object detection are performed on the parallax images, and if necessary, partial conversion to distance is performed to reduce the processing load. Fig. 3 shows an example of a parallax image taken with a 4K stereo camera.

3 Features of Stereo Method

Fig. 4 shows object extraction from parallax image. The features of the generated parallax images are as follows.

(1) In a parallax image, the same object becomes a cluster of similar parallaxes. Therefore, one 3D object is extracted as a block by virtue of parallax grouping. Since images are extracted simultaneously from various parallaxes, multiple three-dimensional objects can be extracted simultaneously and easily, along with their shapes. Since the parallax is known, the distance can be easily calculated.

(a) Monocular camera (b) Stereo camera

Fig. 5 Pictures by Cameras Placed Close to Mt. Fuji

For (a), the distance is essentially unclear with a monocular camera. For (b), however, with a stereo camera, distance can be determined from parallax, so this photo can be recognized as a photo placed nearby.

(2) Information on the distance to any object can be obtained as long as it appears on the screen. For example, if you take an image of Mt. Fuji from Tokyo and assume that the distance is 400 m when the parallax is 1 pixel, you can see that Mt. Fuji is farther than 400 m because the parallax is almost 0 pixel. This is important information for collision avoidance, but as shown in Fig. 5, when using a monocular camera, even a photograph of Mt. Fuji placed nearby looks the same, which makes it impossible to determine whether it is nearby or far away.

(3) Distance information to objects displayed on the screen is obtained as a one-to-one correspondence with pixels. The combination of a monocular camera, Light Detection And Ranging (LiDAR), or millimeter-wave radar is commonly used as sensor fusion. However, since the coordinate systems of the objects detected by each sensor are different, alignment and changes over time are difficult. This is often a problem.

(4) Fig. 6 shows the distance measurement accuracy of a VGA-sized stereo camera. In the example of a VGA stereo camera, the distance accuracy is ±4 cm at 10 m and ±1 cm at 5 m. The closer the distance from the object, the higher the accuracy, allowing more accurate collision avoidance. Although LiDAR has a constant accuracy regardless of distance, there is a concern that even if the accuracy is high at long distances, the control may be diminished at short distances where collision judgements must be made quickly.

(5) Fig. 7 shows a stereo camera that can detect lateral movement. LiDAR and millimeter-wave radar detect the distance to a point within a plane, so if the plane moves laterally, it is difficult to detect that movement. A stereo camera can continuously

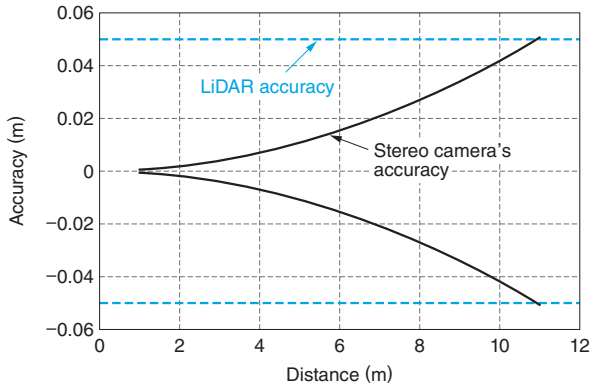

Fig. 6 Distance Measurement Accuracy of VGA-Sized Stereo Camera

This shows that stereo camera has better ranging accuracy within a short distance.

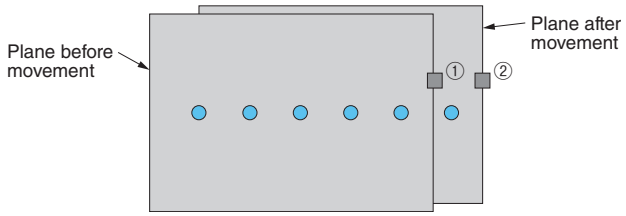

Fig. 7 Stereo Camera that can Detect Lateral Movement

Since the LiDAR repeatedly measures only the distance to the blue spots all the time, sidewise movements cannot be identified even though the plane is moved. The stereo camera can measure the edge position and its parallax (distance).

detect boundaries that change in position on a pixel-by-pixel basis and in short time intervals, so it can accurately detect relative velocity vectors that include the lateral movement of objects. ① represents the 3D position of the edge detected in a certain frame. ② represents the 3D position of the same edge detected in the next frame. The relative velocity in the lateral direction can be calculated from the difference between ① and ②. This greatly affects its ability to predict collisions with an object that jumps out from the side and also affects the tracking performance when the vehicle turns.

4 Correction and Calibration

The stereo camera has many features, but when actually taking pictures, the images are obtained through a lens, which inevitably causes distortion. As mentioned above, in the case of a parallel and coordinate layout, the search range coincides with the scanning direction. When distorted, the search range does not form a straight line, which

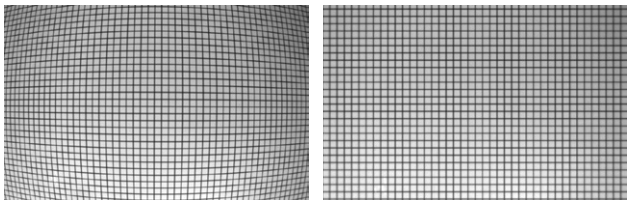

Fig. 8 Grid Image before and after Correction

The left figure shows an image with distortion and the right figure shows an image after correction. After correction, it is understood that the grid image is clearly corrected as a result of reduction of distortion.

makes it difficult to search for the area. Even if it is found within the search range, the exact parallax cannot be determined. This is the most difficult problem in putting stereo cameras into practical use.

4.1 Distortion Correction

In order to solve the problem for practical application, it is possible to measure distortion and incorporate a correction processing circuit. Correction methods can be broadly divided into two methods: one is to parameterize the distortion and correct it by calculation, and the other is to directly determine the corrected coordinates for each pixel. In both cases, the patterns, which have patterns such as checkers and lattices, are photographed, and corrections are made based on the positional relationships of the feature points obtained from the images.

Parameterization methods are widely used. For example, such methods are built into Open source Computer Vision library (OpenCV) and it is easy to use. However, since it is based on symmetry, it is difficult to deal with distortions that break symmetry, such as when the central axis of the lens is misaligned. For this reason, stereo cameras that require high accuracy require a method to correct each pixel. Fig. 8 shows the grid images before and after correction.

4.2 Calibration

Even if the image distortion is corrected, it will not become parallel and equidistant. This is a problem with the relative positional relationship of the stereo camera. It is necessary to adjust the stereo camera position so that it is parallel and equidistant to the object. The accuracy of the adjustment must be such that the positional deviation between the reference image and comparison image is within 0.1 to 0.2 pixels. The size of each pixel in a

Complementary Metal Oxide Semiconductor (CMOS) sensor is approximately 3 μm, and 0.1 to 0.2 pixels are smaller than 1 μm. The camera's positional shift has a translational component and a rotational component, but since the camera uses a lens to reduce the landscape to a factor of several thousand, the effect on the translational component can be ignored. On the other hand, the rotational component has a focal length of the lens of about 5 mm, and if the camera is tilted by 1/10,000, a shift of about 0.2 pixels will occur. These rotational components are caused by twisting or warping of the base that connects the two cameras. It is difficult to mechanically suppress this so that it does not change over time. Therefore, it is adjusted electronically so that they are parallel and equidistant. This adjustment is also done by photographing the pattern images, which include patterns such as checkers and lattices. Image misalignment due to minute rotational components appears in the form of image translation or rotation. Therefore, this correction can be done with simple processing since it is only necessary to consider the translational movement and rotation of the entire image. If the rotation becomes large, the optical axis will no longer be perpendicular to the imaging plane, so the magnification will differ at both ends of the screen. This case is a little more complicated, but since it can be calibrated using linear transformation, it can be handled with light processing.

4.3 Electronic Adjustment

Electronic adjustment for correction and calibration can accommodate deviations of about 50 pixels before adjustment, so it can be kept within the tolerances of the parts. Therefore, by assembling the parts without adjusting the position and then making the final adjustments electronically and unattended, manual adjustment work can be eliminated. If corrections and calibrations are combined in a table in advance, the computational load will be light, and it can be used for real-time processing. Therefore, there is no difference in processing time or load between the parameterization method and the direct pixel-by-pixel correction method. In addition, by applying electronic adjustment, camera misalignment can be constantly checked from parallax images during operation and adjustments can be made automatically. Electronic adjustment makes it possible to eliminate the need for stereo camera adjustments.

An in-vehicle stereo camera focuses at infinity, so in order to capture the pattern for correction, the pattern must be placed far away so that it is in focus. In order to place a pattern far away, it is necessary to prepare a pattern with very large dimensions. To avoid this situation, there is a method (1) that places the pattern close to the stereo camera and uses a conversion lens to focus on it.

5 Postscript

Since a car moves in the depth direction, it is completely natural for it to perceive its surroundings as objects with depth. 3D image recognition using two eyes is a method of binocular stereoscopic vision that primates use on a daily basis. It is an advanced method of recognizing the external world that was finally acquired at the end of the animal's evolution. The stereo camera developed by ITD Lab Corporation. also took inspiration from this idea.

We live in an era where mobile robots and various robots that assist humans are active. Such robots, not just cars, are required to have the same ability to recognize objects and the surrounding environment as humans. We are confident that stereo cameras, which recognize objects in the same way as humans, will continue to play an active role as the most important sensor.

Lastly, we would like to express our deep gratitude to ITD Lab Corporation for their cooperation.

《**Reference**》

(1) Masao Nakagawa, Hiroyuki Yamamoto, Toshimichi Takahashi, Keiji Saneyoshi: "Study on validation of stereo camera using display and conversion lenses", Journal of the Japan Society for Precision Engineering, 2022, Vol.88 No.10, pp.789-794