Wide Angle Multi-Shift Stereo Camera with Monocular Vision

Masayuki Kobayashi\textsuperscript{1}, Kazuyoshi Yamazaki\textsuperscript{1},
Felipe Gomez Caballero\textsuperscript{2}, Takuma Osato\textsuperscript{2}, Takeshi Endo\textsuperscript{2}, Masayuki Takemura\textsuperscript{2},
Takeshi Nagasaki\textsuperscript{3}, and Takeshi Shima\textsuperscript{3}

\textsuperscript{1} Central Research Laboratory, Hitachi Ltd., Kokubunji, Tokyo, Japan,
\textsuperscript{2} Hitachi Research Laboratory, Hitachi Ltd., Hitachi, Ibaraki, Japan,
\textsuperscript{3} Hitachi Automotive Systems Ltd., Hitachi, Ibaraki, Japan
masayuki.kobayashi.tx@hitachi.com

\textbf{Abstract—} In recent years, following the development of automated driving (AD) and advanced driver assistance systems (ADAS), stereo cameras which are able to measure distance are advancing accordingly by employing wide-angle field of view to detect various objects. At the same time, demand for low cost stereo cameras which can detect far objects also exists. Also, detection distance performance degradation due to low resolution CMOS image sensors and distortion from wide-angle lenses remain as problems when increasing the stereo camera field of view (FOV). In order to solve such problems, we developed a multi-shift stereo camera. In the multi-shift method, the left and right CMOS image sensors are shifted their center with respect to each lens optical axis center. By using this method, far object detection performance can be preserved while obtaining a wider FOV without the use of an expensive image sensor. We also developed a monocular vision detection employing top view image subtraction method, enabling detection of pedestrians and cyclists in the monocular vision region which is created at the outer region of the camera FOV. Furthermore, by using the disparity information from the center region to estimate the camera posture, we can achieve high accuracy for detection and distance measurement at the monocular region.

\textbf{Keywords—} stereo camera; stereo vision; safety driving; distance estimation

I. INTRODUCTION

A century after Karl Friedrich Benz invented the automobile, our life has become prosperous and economy developed thanks to such convenient transportation mean. On the other hand, many people die in traffic accidents every year. Recently, advanced driver assistance systems (ADAS) had developed to achieve both convenience and safety while automated driving systems (AD) are being developed around world. Our research focuses on the creation of next generation stereo cameras that can contribute to improve safety of AD/ADAS.

A stereo camera system uses a set of two cameras from which obtained images can be used to measure the distance to a target object by calculating the object’s disparity between the two images. Since distance information can be obtained, they have been widely used on automotive, surveillance, and robotics applications. Several car manufacturers sell cars with stereo camera systems that support automatic braking system (AEB) and adaptive cruise control (ACC) \cite{1} \cite{2}.

In order to increase the safety performance, there is a need for an extended field of view (FOV) stereo camera that can detect crossing cyclists and pedestrians while turning at intersections \cite{3} \cite{4} \cite{5}. On the other hand, detection and distance measurement of far preceding vehicles is also required for high-speed range ACC. If the stereo camera FOV is simply extended, resolution of CMOS image sensors becomes insufficient. In addition, when employing a wide-angle stereo camera, it is difficult to ensure calibration accuracy due to distortion correction associated with wide-angle lenses.

We focus on integrating stereo vision and monocular vision to solve the CMOS image sensors resolution and lens distortion problems in order to realize wide-angle stereo camera. Object detection at the outer area of the camera’s FOV is required for objects such as crossing cyclist or pedestrian at intersection turning. It is also worth noting that only a 30-m detection range is required for such scenarios; hence, high-precision distance measurement by stereo vision is not required. In this paper, we describe the developed multi-shift stereo camera system which combines stereo and monocular vision, as well as the moving object detection method implemented in the monocular vision region which employs road disparity.

II. WIDE ANGLE BY MULTI-SHIFT METHOD

Fig. 1 shows the concept of the FOV for a stereo camera using the multi-shift method.

![Fig. 1. FOV of multi-shift stereo camera.](image)

It can be seen above that the field of view for each camera is extended to the outer side. Therefore, the center area of the field of view becomes a stereo vision region since both left and right camera images are available, while the right and left areas of the field of view are monocular vision regions.
Fig. 2 shows example images obtained by the multi-shift stereo camera. The left view camera image has a monocular vision at left of stereo vision, and the right camera view image has a monocular vision at right of stereo vision.

Fig. 2. Example of multi-shift stereo camera image set.

Fig. 3 depicts a diagram illustrating the positional relationship between the CMOS image sensors and the lenses for the multi-shift method. A conventional stereo camera, shown in Fig. 3 (a), is designed to have the same field of view for both left and right cameras (A in Fig. 3 (a)) by aligning the center of the CMOS image sensor and the lens optical axis center. On the other hand, for the multi-shift method shown in Fig. 3 (b), each CMOS image sensor center is shifted to the left or right, relative to each lens optical axis center. By shifting each CMOS image sensor in the opposite direction, the field of view of left and right cameras, respectively, becomes wider to the left or right, achieving a wide FOV (B in Fig. 3 (b)).

In order to perform stereo matching in the horizontal direction on the center field of view to create a stereo vision, it is necessary to correct the images of the central projection without distortion which have epipolar lines in parallel. If distortion in original image is low, residual error after distortion correction is smaller. Thus, the center field of view is used for stereo vision in order to reduce disparity errors (i.e., wrong distance estimations) and blank disparity pixels (i.e., no distance information available) which are typically caused by residual errors after distortion correction. Also, both optical axis centers of the lenses are aimed towards the vanishing point in order reduce distortion and achieve high accuracy distance measurement.

Fisheye lenses are used for the multi-shift stereo camera. A fisheye lens is an ultra-wide-angle lens, for which its distortion increases along with the view angle. In this way, images captured when using a fisheye have higher resolution in the center areas compared to the resolution of the outer areas. Therefore, in comparison with the outer-angle area, the center field of view has many CMOS image sensor pixels assigned, which ensures that required resolution for far detection is available. By combining the fisheye lens and the multi-shift method, it is possible to achieve both far-distance detection and wide FOV for the stereo camera.

III. MONOCULAR DETECTION WITH STEREO VISION

Monocular object detection for the multi-shift stereo camera is described in this section. We developed a monocular detection technique employing a top view image subtraction method combined with a camera posture estimation by road disparity obtained from the stereo vision. The top view image subtraction method refers to a background image subtraction method using top view images. Fig. 4 shows the top view image subtraction method detection concept. In the top view image subtraction, the camera images of current and previous time frames are affine-transformed to create top view images, then the resulting images are subtracted to obtain a differential of the top view images. The created top view images can be moved in a way that the road on both images overlap, resulting on an image where only the differentials between the three-dimensional objects and the moving objects standing on the road surface remain. Thus, making detection possible for vehicle mounted systems as the point of view moves along with the own vehicle. Further, distance measurement becomes possible since the top view image is projected image on the road surface, and hence the distance on the top view image between the camera and contact point of the object is proportional to the distance from the camera to the object.
In order to create an appropriate top view image in situations where the camera posture changes dynamically, it is important to estimate the posture of the camera at any given time. Fig. 5 (a) shows the concept of camera posture estimation using disparity from the road surface. Fig. 5 (b) shows the concept of road disparity on the camera image. As described in [7], it is possible to specify road surface pixels in disparity image. Assuming a flat road surface can be estimated from each distance and position value, it is possible to estimate the camera posture parameters corresponding to camera height, tilt and roll angles. By using the estimated camera posture to create appropriate top view images, it is possible to perform object detection and distance measurement with high accuracy.

Fig. 6 shows the processing flow and related images of the monocular detection. Where, Fig. 6 (a) is a top view image, Fig. 6 (b) is a top view differential result, Fig. 6 (c) is the detection result on the camera image and Fig. 6 (d) is the process flow of the monocular detection. The process is as follows. First, it specifies the disparity information of the road surface by stereo vision. Next, the camera posture estimated from the disparity of the road surface is used to create an affine table, which is used to perform affine-transform that produce top view image (Fig. 6 (a)). Differential of the top view image of the current and previous frames (Fig. 6 (b)) is then calculated, and detection on the monocular image is performed (Fig. 6 (c)).

Examples of monocular detection of the stereo camera are shown in Figs. 7-9. Figs. 7-9 (a) depicts a camera image which represents a combined left and right camera images showing the correspondent detection result as rectangles. In the same manner, the top view image (Figs. 7-9 (b)), represents a combined left and right camera image. Also, the image in Figs. 7-9 (c) shows a combined left and right image representing the top view differential, where subtraction differentials appear as white pixels.

Fig. 7 shows a detection result for a scenario with a crossing cyclist on the left side and pedestrians sitting on the right side (Fig. 7 (a)). The own-vehicle which attached the camera travels towards a collision point. Both targets appear in the middle area of the top view image (Fig. 7 (b)), and differentials occurred in the same area in the differential image (Fig. 7 (c)). We were able to confirm that the cyclist...
and pedestrians were detected properly as a result that the differentials were calculated properly (Fig. 7 (a)).

Fig. 8. Experimental result of monocular detection, a pedestrian with parking lot paints.

Fig. 8 shows the detection result for a scenario with a crossing pedestrian from left side. The own-vehicle is parked. The pedestrian and parking space paint appear in the top view image (Fig. 8(b)), the pedestrian differentials appear in the differential image while the parking space paint differentials appear only to a small extend (Fig. 8(c)). We were able to confirm that only the crossing pedestrian was detected without interference from the parking space paint (Fig. 8(a)).

Fig. 9. Experimental result of monocular detection, pedestrians at intersection turning.

Fig. 9 shows the detection result of a scenario with a crossing and a still pedestrian at an intersection. The own-vehicle was turning the intersection. Although there is some data from the road paint in the top view image (Fig. 9(b)) and the own-vehicle was turning, only the differentials for the pedestrians remains in the differential image (Fig. 9(c)). We were able to confirm that the pedestrians were detected properly at intersection.

V. CONCLUSIONS

We developed a stereo camera employing the multi-shift method which can achieve both far-distance detection and wide-angle FOV. Also, we developed a monocular detection technique for the monocular vision field of view of the multi-shift stereo camera with camera posture estimation using road disparity.

In the multi-shift approach, CMOS image sensors are moved from center of lenses to the outside. This creates monocular vision on the outer part of the stereo camera FOV, expanding its field of view widely. In the multi-shift method, it is possible to suppress the resolution reduction of the center field of view by using fisheye lenses, having many pixels for the stereo vision field, preserving the detection distance performance in the stereo vision field. The multi-shift stereo camera achieved both far distance detection performance and wide-angle FOV without employing expensive high-resolution CMOS image sensors.

In addition, the developed monocular detection technique using top view image subtraction method with camera posture estimation was able to achieve high-precision detection in the monocular vision field by appropriately estimating the camera posture using disparity of the road surface.

In this way, and avoiding cost increase, it is possible to preserve the convenience of preceding vehicle follow-up (ACC) since the far detection performance is not impaired. Also, it became possible to realize automatic emergency braking (AEB) for wide-angle FOV targets. We expect that the spread of this low-cost, wide-angle multi-shift stereo camera, will contribute to further reduce traffic accidents.

REFERENCES