

Image Segmentation-based Moving Object Detection for Driver Assistance System

Morimichi NISHIGAKI*

Yiannis ALOIMONOS**

ABSTRACT

The detection of objects in front of ego-vehicle is an effective means to help prevent collisions. Since the potential of collision with moving objects, such as other vehicles and pedestrians, is relatively high, it is important to detect these moving objects in particular. Vision sensors are employed to detect the various objects encountered in actual driving situations, but because the camera and the objects are moving independently, it is not easy to detect independently moving objects using vehicle-mounted cameras. The detection of sharp boundaries for moving objects has been especially challenging. We have therefore developed two algorithms to detect sharp boundaries for moving objects. The first of these algorithms detects objects moving in a different direction to the camera by identifying inconsistency between motion in the images and the epipolar constraint. The second algorithm detects objects moving in the same direction as the camera by identifying inconsistency between stereo disparity and motion disparity. The use of image segmentation technology in these algorithms has made it possible to detect sharp boundaries for moving objects. The ability to detect sharp boundaries using these methods was verified in tests using actual vehicles.

1. Introduction

Among Intelligent Transportation Systems (ITS), Advanced Driver Assistance System (ADAS) is a field that is the subject of vigorous research. Sensors for the detection of objects in the vicinity of the subject vehicle, including other vehicles, pedestrians, lane markers, and traffic signals, are key components of ADAS. High expectations have been placed on visual sensors in the detection of these various objects around the vehicle. From the perspective of safe driving assistance, it is essential to detect independently moving objects such as vehicles and pedestrians, because these represent a higher risk of collision with the subject vehicle. However, it is not easy to detect independently moving objects from images. This is because stationary objects also appear to be moving in images captured by a camera mounted on a moving vehicle. The determination of a method of detecting the specific locations of moving objects from moving images of this kind represents an issue for research^{(1), (2)}. When the direction of motion of moving objects differs from the direction of motion of the camera, the moving objects can be detected using a single camera (a monocular camera) if epipolar constraints⁽³⁾ are applied. Epipolar constraints express the constraint conditions in relation to the apparent motion of the stationary objects. When the direction of motion of the moving object is the same as that of the camera, it cannot be detected using a

single camera. This necessitates the use of two cameras (stereo camera), a factor which also increases the complexity of the algorithms employed. The project discussed in this paper therefore developed two algorithms for application in these sensing methods. For automotive applications, it is also necessary to detect the width and size of objects. The method discussed here detects the boundaries of the object using image segmentation technology to divide the image into meaningful regions.

2. Algorithms

Ogale et al. categorized the detection of moving objects from images into three classes, based on the relationship between the movement of the camera and the movement of the object⁽⁴⁾. Objects that move in a different direction to the direction of motion of the camera are defined as belonging to Class 1. In these cases, the object can be identified as a moving object by determining whether or not the movement of the object satisfies the epipolar constraint. The epipolar constraint cannot be employed in the case of objects that move in the same direction as the direction of motion of the camera. Moving objects that move in the same direction as the camera and are partially occluded by stationary objects are categorized as Class 2. In these cases, independently moving objects can be detected by determining whether there is inconsistency between the anteroposterior

* HONDA R&D EUROPE (DEUTSCHLAND) GMBH

** University of Maryland

relationship and motion disparity due to the occluding object. When its direction of motion is the same as that of the camera but no occluding objects are present, the moving object is classified as Class 3. In these cases, moving objects can be detected by determining whether any inconsistency exists between motion disparity and the anteroposterior relationship as determined from stereo disparity.

Depending on the distance of the moving object and the stationary object that partially occludes it, there are cases in which moving objects cannot be detected when using methods for the detection of Class 2 moving objects. However, objects falling into any class can be detected using methods for the detection of Class 3 objects.

Two algorithms for the detection of moving objects using vehicle-mounted cameras were developed based on these classifications and principles of detection. The first of these algorithms is employed in the detection of Class 1 moving objects using a monocular camera. The second algorithm is employed in the detection of more general moving objects, including Class 2 and 3 objects, using a stereo camera. The first algorithm is simpler than the second, but the range of moving objects that can be detected using the algorithm is limited. The second algorithm is able to detect more general moving objects, but necessitates the use of a stereo camera. The two algorithms should be used separately depending on the object for detection. In order to relate the two algorithms to actual traffic conditions, the first algorithm is referred to as an algorithm for the detection of intersecting objects, and the second algorithm is referred to as an algorithm for the detection of preceding objects. This means that the first algorithm, which is simpler and feasible using a monocular camera, can be applied to the systems for the detection of intersecting objects, and the second algorithm, which necessitates the use of stereo cameras and entails complexity in the algorithm itself and system components, can be applied to the systems for the detection of preceding objects.

Sections 2.1 and 2.2 of this chapter will discuss the essential points of the algorithms, while Sections 2.3 and 2.4 will respectively consider the algorithm for the detection of intersecting objects and the algorithm for the detection of preceding objects in greater detail.

2.1. Principle of Detection of Moving Objects

The algorithm for the detection of intersecting objects and the algorithm for the detection of preceding objects employ different principles of detection.

Figure 1 shows a representation of the principle of detection employed by the algorithm for the detection of intersecting objects. The epipolar lines⁽³⁾ when the vehicle is proceeding straight ahead are shown as blue lines on this image. In this case, the epipolar lines show the direction of motion of stationary objects. The point of intersection of the epipolar lines is the focus of expansion (FoE). In the image, all of the stationary objects are imaged as if in motion, following the directions of the epipolar lines outward from the FoE. The epipolar constraint expresses the fact that geometric constraints exist for this apparent motion of the stationary objects in the image. When motion at a specific point in the image follows an epipolar line, it is said to satisfy the epipolar constraint. By contrast, intersecting objects such as the pedestrian in the image do not move following the epipolar lines, and do not satisfy the epipolar constraint. A moving object can therefore be detected by determining whether or not the optical flow for each pixel satisfies the epipolar constraint. The optical flow is the motion at each point in the image expressed as two-dimensional vectors. When the subject vehicle is not restricted to proceeding in a straight line and turning, pitch, and roll rotation are incorporated, determination of whether the epipolar constraint is satisfied becomes more complex. Rectification, which will be discussed below, is therefore applied to cancel the effect of rotational motion by deforming the image, making it possible to detect moving objects as in the case shown in Fig. 1.

When the moving object precedes the subject vehicle, the principle of detection using the epipolar constraint cannot be employed and another principle should be used despite the fact that the object is moving, because of the effect of motion following the direction of the epipolar lines. Figure 2 shows a conceptual diagram of the principle of detection for the algorithm for the detection of preceding objects. This diagram shows the same scene, incorporating both stationary and moving objects, captured by a stereo camera (right) and a moving camera (left). Using the stereo camera, the distances to objects can be measured using

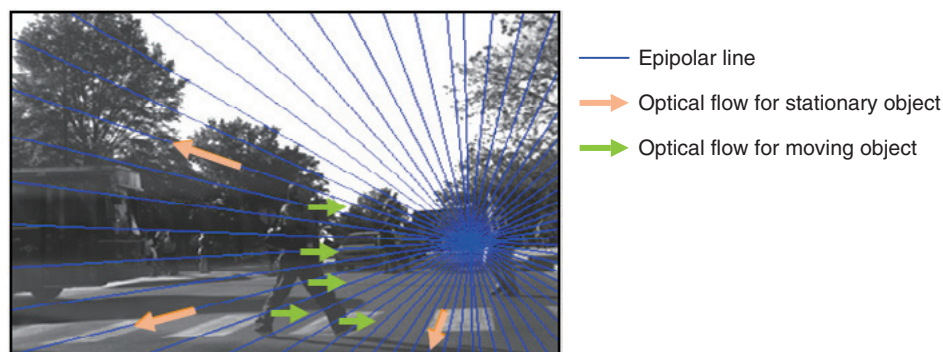


Fig. 1 Principle of detection of moving objects for intersecting objects

triangulation, irrespective of whether the objects are moving or stationary. By contrast, in the case of the moving camera, when the distances to objects are measured by means of triangulation using the images captured at time t and time $t+1$, the distance to stationary objects can be measured correctly, but the distance to moving objects cannot. This is because the object itself moves and changes position during the time that the camera is moving. As Fig. 2 shows, both the stereo camera and the moving camera measure identical distances to the stationary object (shown as the solid lines), but measure different distances to the moving object (shown as the broken lines). Therefore, if the distances to an object as measured by stereo and moving cameras differ in a comparison, the object can be identified as a moving object. Up to this point, stereo cameras and moving cameras have been discussed as separate cameras, but if a stereo camera is considered as being mounted on a vehicle and in a state of motion, one of the cameras (for example the camera on the right-hand side of the vehicle) could be considered to be a moving camera, and this principle could be applied.

2.2. Achievement of Sharper Boundaries via Image Segmentation

The research discussed here attempted to make it possible to detect sharp boundaries of moving objects using image segmentation technology. In the case of the algorithm for the detection of intersecting objects, image segmentation was employed to increase the accuracy of the optical flow; in the case of the algorithm for the detection of preceding objects, image segmentation was used to increase the accuracy of

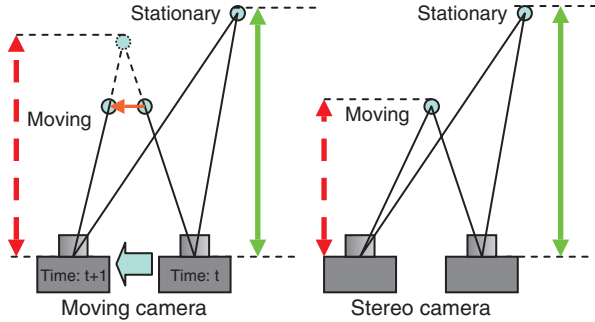
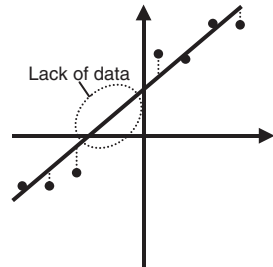
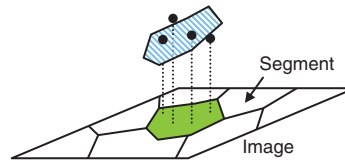


Fig. 2 Principle of detection of moving objects for preceding objects



(a) 1D model fitting



(b) Segmentation-based model fitting on image

Fig. 3 Model fitting based on segmentation

stereo disparity in addition to the accuracy of the optical flow. The concept on which the use of image segmentation is based is identical for both optical flow and stereo disparity, and a brief overview will be provided here. The background concept is model fitting, as shown in Fig. 3(a). In the figure, measured values are shown as points. It is assumed that the measured values are affected by noise, and that some values will be missing, having for various reasons been unable to be measured. Given that in principle the measured values accord with a model expressed as a linear expression, they can be assimilated to a straight line, as shown in the figure. Leaving aside the measured values themselves and taking the straight line resulting from model fitting as representing new measured values makes it possible to reduce the effect of noise and obtain more appropriate values even for missing values. Figure 3(b) shows a diagram produced by the application of this concept to measured values in an image. As the figure shows, the segmentation process divides the image into segments, and measured values can be obtained within the shaded segment (as shown by the black points). For example, if the measured values are assumed to accord with a plane model, model fitting using the measured values within the segment makes it possible to obtain the hatched plane that represents measured points. In the research discussed here, motion models were used for optical flow, and disparity models were used for stereo disparity. Conventional methods of calculating the optical flow or the stereo disparity use texture in the vicinity of the pixels to be calculated, assuming that the optical flow or the stereo disparity is uniform. Therefore, calculation could not be performed in regions with less texture and it was a challenge to determine the boundaries of objects due to increased errors in regions in which the assumption did not hold. Segmentation-based model fitting identifies regions in which the optical flow and stereo disparity change uniformly according to models in advance. Reducing the effect of noise and obtaining more appropriate values even for the regions with less texture by means of model fitting using image segmentation has made it possible to accurately determine the changes in optical flow and stereo disparity at the boundaries of the object.

2.3. Detection of Intersecting Objects

Ego-motion estimation^{(5), (6)} and optical flow estimation were the essential elements in realizing the algorithm for

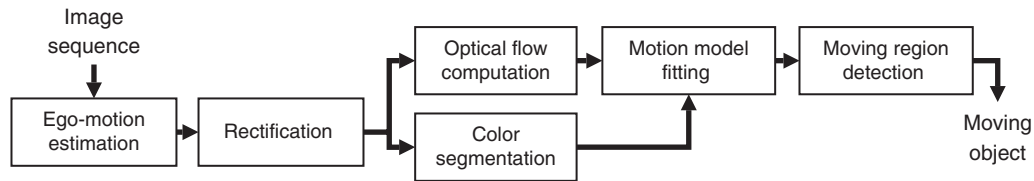


Fig. 4 Block diagram of intersecting object detection

the detection of intersecting objects. Ego-motion estimation was necessary for rectification, which will be discussed below. A method of calculation of the optical flow based on color segmentation was developed in order to obtain dense optical flows with sharp boundaries for the moving objects. Color segmentation is a method of image segmentation that divides images based on color homogeneity. Figure 4 shows the composition of the algorithm. Two continuous image frames represent the input. The motion of the camera is estimated during “Ego-motion estimation,” and “Rectification” processing is applied in order to cancel the effect of the rotation of the camera on the image. During the “Optical flow computation” process, the initial values of the optical flow are calculated, and “Color segmentation” is applied to the first image. Next, during the “Motion model fitting” process, motion model parameters that express the motion of each segment obtained from color segmentation on the image are determined from the initial optical flow by model fitting. During the “Moving region detection” process, the region occupied by the moving object is detected by identifying the pixels for which the optical flow does not follow the epipolar lines. The main elements of the algorithm will be discussed in more detail below.

2.3.1. Rectification

This sub-section will discuss the “Rectification” processing that is applied to cancel the effect of the rotation of the camera after the camera motion parameters have been determined during “Ego-motion estimation” in the processing flow shown in Fig. 4.

The optical flow is obtained from two continuous image frames. When the translational and rotational motion of the camera are determined by ego-motion estimation, it is possible to deform the images in order to shift the epipoles⁽³⁾ of the two images to the centers of the images and cancel

the motion in the images resulting from the rotational motion of the camera. As shown in Fig. 5, by this means, despite the fact that the motion of the camera incorporates both translational and rotational motion, the images can be deformed to make it appear that the camera displays only translational motion. This deformation of the images makes it easier to apply processing in the later stages to determine whether or not the optical flow follows the epipolar lines. In other words, doing away with the effect of the rotational motion of the camera makes it possible to more easily detect the moving object, as in the case shown in Fig. 1. Figure 6(b) shows an image to which rectification processing has been applied. In this example, rectification has slightly distorted the entire image in Fig. 6(b) in relation to the original image shown in Fig. 6(a).

2.3.2. Motion model fitting using segmentation

Again looking at the processing flow shown in Fig. 4, this sub-section will discuss the method of estimation of the motion model parameters for “Motion model fitting” using the initial optical flows calculated in the “Optical flow computation” process and the segments obtained in the “Color segmentation” process.

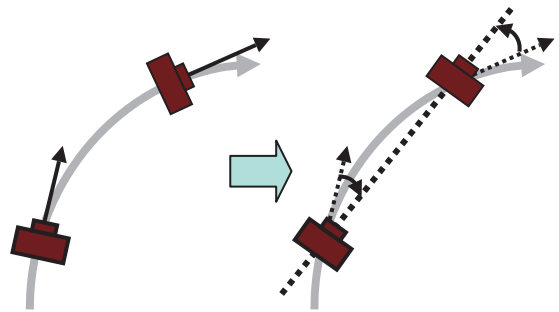


Fig. 5 Concept of rectification

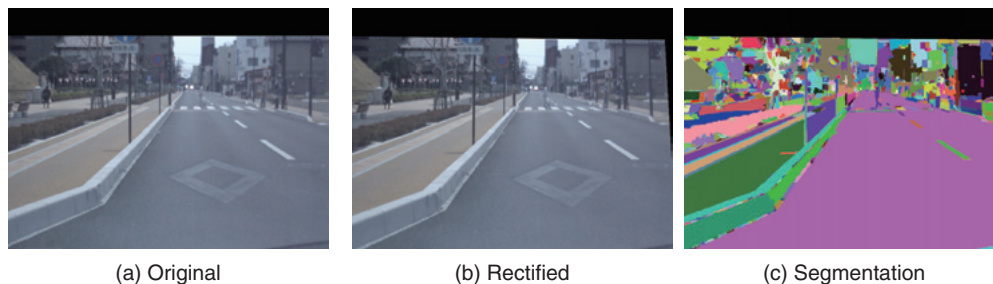


Fig. 6 Examples of rectification and segmentation

When viewed in terms of distance from a vehicle driving on the road, the objects that appear in the traffic environment – the road surface, other vehicles, signs, and so on – can be approximated to a collection of planes. When the motion in the image is determined, at a certain distance each section of even non-rigid bodies such as pedestrians can be approximated as planes. Because the motion of the planes in the two image frames is expressed as homography transformations, it is logical to model the motion on the images of each plane making up the moving object using homography transformations. A method of calculating the optical flow based on color segmentation was therefore developed, in which the images are divided into segments based on color homogeneity. Each segment is treated as part of a plane, and it is assumed that the motion of the segments can be approximated using homography transformations. First, the initial optical flows are calculated by means of a block matching method using two continuous rectified image frames. The initial optical flows are only calculated for regions within the blocks in which sufficient texture exists. Next, the first image of the two frames is segmented based on color homogeneity using the mean-shift method⁽⁷⁾, one of the color segmentation methods. Figure 6(c) shows an example of segmentation. Pixels painted the same color belong to the same segment.

Like other methods based on image segmentation⁽⁸⁾, it was assumed that the boundaries of the segments incorporate the object boundaries, and that each segment images part of a plane. Parameters are estimated by fitting in a motion model in which the initial optical flows for each segment are expressed as homography transformations. When points are imaged on a plane, the motion of a point at a specific position (x_1, y_1) in a frame when the point moves to another position (x_2, y_2) in the next frame is modeled using the following equation:

$$w \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix} \quad (1)$$

where w is a quantity that represents the scale of homography transformation, obtained by determining the third factor after the transformation as 1, and $\{h_{ij}\}$ is the homography transformation parameters to be estimated. Given that the destination position (x_2, y_2) for each point at an initial position (x_1, y_1) can be determined from the initial optical flow, the substitution of the corresponding points

makes it possible to estimate the parameters. In order to reduce the effect of errors, parameters were estimated using the RANSAC method⁽⁹⁾, a robust estimation method. When the estimated parameters are used to calculate the motion of each pixel in the segments, more accurate optical flows with sharper boundaries than the initial optical flows are obtained because of using color segmentation.

2.3.3. Moving region detection

This sub-section will discuss the “Moving region detection” process in the processing flow shown in Fig. 4, in which the moving region is identified following the determination of more accurate optical flows in the “Motion model fitting” process.

Moving regions in the images are identified by determining that the direction of the optical flow for each pixel differs from the direction of the corresponding epipolar line. Because rectification has been applied to cancel the rotational motion of the camera and shift the epipoles to the centers of the images, the epipolar line corresponding to any specific point is found as a straight line connecting that point and the center of the image. If the direction of the optical flow for each pixel does not follow the epipolar lines, the object can be identified as a moving object.

2.4. Detection of Preceding Objects

In order to detect objects moving in the same direction as the subject vehicle, independently moving points are identified by comparing their distance as estimated from their motion in the image and their distance as measured by the stereo camera, and they are classified as belonging to a moving object. It is possible to identify moving objects because the distances match in the case of stationary objects, but differ in the case of moving objects. The use of color segmentation in conjunction with this method helps to enable sharp boundaries to be obtained.

Figure 7 shows the entire algorithm as a block diagram. The additional elements in relation to the algorithm for the detection of intersecting objects are stereo disparity computation based on image segmentation and distance estimation based on motion. In addition, because of the difference in the principle of detection, the method of identification of the moving object region in the image differs from that employed in the case of detection of intersecting objects. The main elements of the algorithm are discussed in greater detail below.

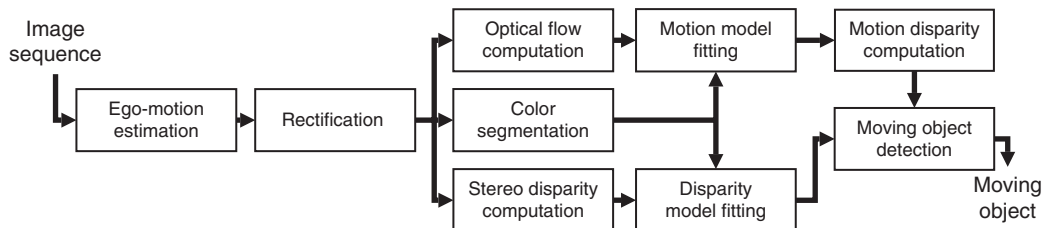


Fig. 7 Block diagram of preceding object detection

2.4.1. Stereo disparity model fitting based on image segmentation

This sub-section will discuss the method of estimating stereo model parameters during the process of “Disparity model fitting” using the initial stereo disparity as calculated during the “Stereo disparity computation” process and the image segments obtained from the “Color segmentation” process, as shown in the processing flow in Fig. 7.

As in the case of calculation of the optical flow, model fitting is employed using the initial stereo disparity in order to estimate the disparity model parameters for each image segment. Because the initial stereo disparity is determined by block matching, it can only be calculated for regions with sufficient texture. In addition, in order to minimize error due to occlusion of objects, the reference image is replaced and the disparity of the right image in relation to the left image and of the left image in relation to the right image is determined; the figures obtained are employed as disparity values only in cases in which there is a match between the two. The mean-shift method is employed to segment one of the stereo images based on color homogeneity, and the disparity model parameters are estimated by fitting initial disparity maps to each segment. At this time, the stereo images are rectified to render them as parallel stereo images. For parallel stereo images, the corresponding points on the images from the left and right cameras (x_1, y_1) and (x_2, y_2) for a plane in three-dimensional space (X, Y, Z) expressed by $\alpha X + \beta Y + \gamma Z = 1$ satisfy the following equation.

$$\begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha' & \beta' & \gamma' \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix} \quad (2)$$

Here, α' , β' and γ' are defined as follows.

$$\alpha' = B\alpha + 1 \quad (3)$$

$$\beta' = \frac{f_x}{f_y} B\beta \quad (4)$$

$$\gamma' = f_x B \left(\gamma - \frac{\alpha c_x}{f_x} - \frac{\beta c_y}{f_y} \right) \quad (5)$$

B is the length of the baseline between the stereo cameras, and f_x and f_y are the x and y components of the cameras' focal lengths in units of pixels. (C_x, C_y) is the principal point, at which the image plane and the optical axis intersect. Because (x_1, y_1) and (x_2, y_2) are determined when the initial stereo disparity is calculated, the disparity model parameters α' , β' and γ' can be obtained for each segment using Eq. (2). The calculation of disparity for each pixel using the model parameters that have been determined makes it possible to obtain more accurate disparity maps than the initial disparity, and also increases the clarity of the boundaries of the disparities.

2.4.2. Estimation of distance from motion

This sub-section will discuss the estimation of distance from motion during the “Motion disparity computation” process following the determination of high-accuracy optical flows by “Motion model fitting” in the processing flow shown in Fig. 7.

By contrast with the detection of intersecting objects, for the detection of preceding objects it is necessary to estimate distance from motion in the images in order to compare distances in the stereo images. Because rectification has been applied to cancel the rotational motion of the cameras, the distance Z can be estimated as follows.

$$Z = T \frac{x}{v_x} = T \frac{y}{v_y} \quad (6)$$

Here, T is the degree of translational motion, and is an unknown parameter. (v_x, v_y) is the optical flow at the point (x, y) . The distance Z can be determined from Eq. (6) using the x or y components of the optical flow, but the scale factor, T , is indefinite. Figure 8 shows examples of the estimation of distance from the x and y components of the optical flow. In this figure, changes in distance are rendered as changes in shading, with brighter regions representing closer distances, and darker regions further distances. Pitch-black represents regions with no calculation result. While it is possible to estimate distance using either component, there is a significant error in the vertical line passing through the FoE obtained when the x component is used and in the horizontal line passing through the FoE obtained

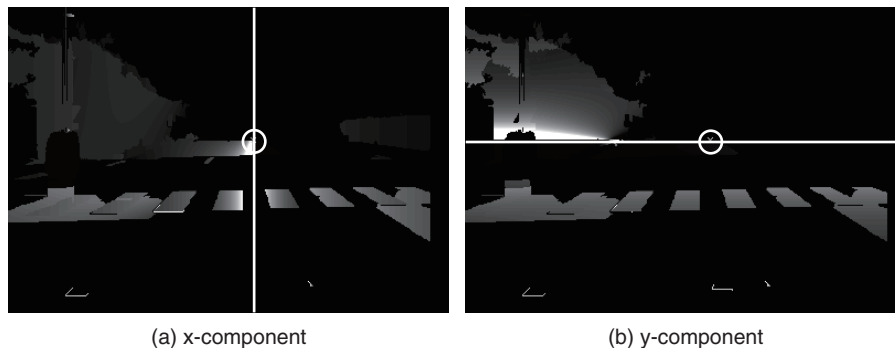


Fig. 8 Intermediate motion disparity computation results

when the y component is used. The circles in the figure show the position of the FoE. This error occurs because of the small values of each component of the optical flow. The orientation and magnitude of the optical flow can be used to minimize this error. However, making use of the fact that the errors occur in specific regions of the image, for the sake of simplicity error is minimized by taking a weighted average of two distances in order to reduce the weight of the regions in which the error is greatest for each component. The distance $Z(x, y)$ at point (x, y) in the image is calculated as follows.

$$Z(x, y) = w_x(x, y)Z_x(x, y) + w_y(x, y)Z_y(x, y) \quad (7)$$

Here, $w_x(x, y)$ and $w_y(x, y)$ are weights at position (x, y) in the image, set such that $w_x(x, y) + w_y(x, y) = 1$. $Z_x(x, y)$ and $Z_y(x, y)$ are the distances at (x, y) estimated from the x and y components of the optical flow, respectively. Figure 9 shows the distribution of the weights and the distance map obtained by weighted averages. The significant errors that occurred in the vertical and horizontal lines passing through the FoE in the case of distance maps calculated using each component, as shown in Fig. 8, have been minimized in the distance maps shown in Fig. 9.

2.4.3. Identification of moving regions

This sub-section will discuss the identification of moving regions in the “Moving object detection” process in the processing flow shown in Fig. 7, using the distance derived from motion in the “Motion disparity computation” process and the distance derived from the high-accuracy stereo disparity in the “Disparity model fitting” process.

Because the scale factor is unknown for distance estimated from motion, it cannot be directly compared with distance estimated from stereo disparity. Based on the assumption that the majority of the regions in the scene shown in the images belong to stationary objects, the

unknown scale factor, T , is calculated as follows.

$$T = \text{median}_{(x,y)} \left(\frac{D_s(x, y)}{D_m(x, y)} \right) \quad (8)$$

Here, $D_s(x, y)$ and $D_m(x, y)$ are the distances at position (x, y) in the images as derived from stereo disparity and motion, respectively. Medians are calculated for all points for which D_s and D_m are valid. Moving regions can be detected by identifying the pixels that satisfy the following condition.

$$|D_s - T \cdot D_m| > \theta \quad (9)$$

Here, θ is a positive threshold.

3. Experiments

Employing image segmentation technology, the method discussed here has increased performance in the estimation of optical flows and stereo disparity against the conventional block matching method. This has made it possible to detect the boundaries of objects during the process of moving object detection. This chapter will consider the results of experiments concerning optical flow and stereo disparity estimation performance using image segmentation, after which the results of moving object detection will be discussed.

3.1. Accuracy of Estimation of Optical Flows using Image Segmentation

The increase in the accuracy of estimation of optical flows against the initial optical flows due to the use of image segmentation was quantitatively evaluated. Here, optical flows obtained using image segmentation refer to optical flows calculated for each pixel using the model parameters obtained from the process of motion model fitting discussed

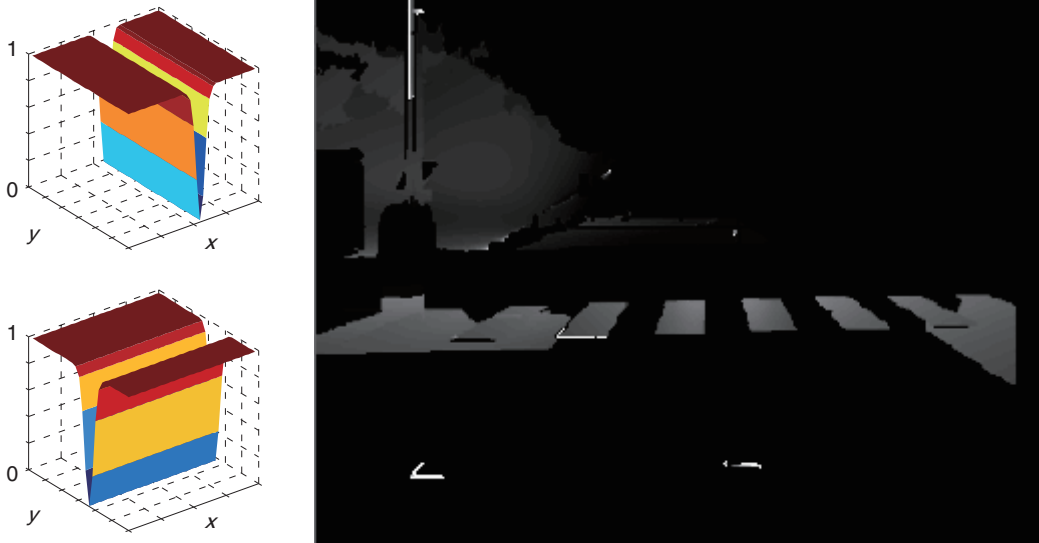


Fig. 9 Weight maps and motion disparity computation result

Table 1 Error in segmentation-based optical flow estimation

Average angular error [deg]	Sequence No.1	Sequence No.2	Sequence No.3
Initial optical flow (Block matching)	11.25 (112330)	15.57 (54162)	10.84 (81210)
Segmentation-based optical flow	8.64 (219010)	9.32 (153671)	4.60 (176364)

Average end-point error [pixel]	Sequence No.1	Sequence No.2	Sequence No.3
Initial optical flow (Block matching)	0.17 (112330)	0.22 (54162)	0.17 (81210)
Segmentation-based optical flow	0.13 (219010)	0.14 (153671)	0.07 (176364)

above. Table 1 shows the angular error and the endpoint error for initial optical flows obtained using block matching and optical flows obtained using image segmentation for each image sequence, as determined using a Middlebury data set⁽¹⁰⁾. Endpoint error is error in the length of the vectors that express the optical flows. The numbers in Table 1 represent angular error and the endpoint error in degrees and numbers of pixels, respectively. In certain cases, valid optical flows could not be obtained by block matching due to the existence of textureless regions, or could not be obtained by the image segmentation-based method due to a high degree of model fitting error. The numbers in parentheses in Table 1 show the number of pixels for which valid optical flows were obtained in each image sequence. The figures for errors in Table 1 are average values for the number of pixels in the parentheses. As the figures for error in the table make clear, estimation of optical flows based on image segmentation helps to enable more accurate optical flows to be obtained than the conventional method using block matching. In addition, the fact that there are a high number of valid

pixels indicates that it was possible to supplement the optical flows in textureless regions. This demonstrates that the initial optical flows calculated during the “Optical flow computation” process in Figs. 4 and 7 displayed increased accuracy following the “Motion model fitting” process shown in the same figures. The estimation of optical flows using image segmentation has increased the sharpness of the boundaries of motion, and has thus helped to enable clearer boundaries to be determined for objects in the images.

3.2. Accuracy of Estimation of Stereo Disparity using Image Segmentation

Figure 10 shows examples of an initial stereo disparity map obtained using block matching and a stereo disparity map obtained using image segmentation. Here, the stereo disparity based on image segmentation is disparity calculated for each pixel using the model parameters obtained during the disparity model fitting process discussed above. Figures 10(a) and (b) show the left and right images from the stereo cameras, and (c) shows the result of color

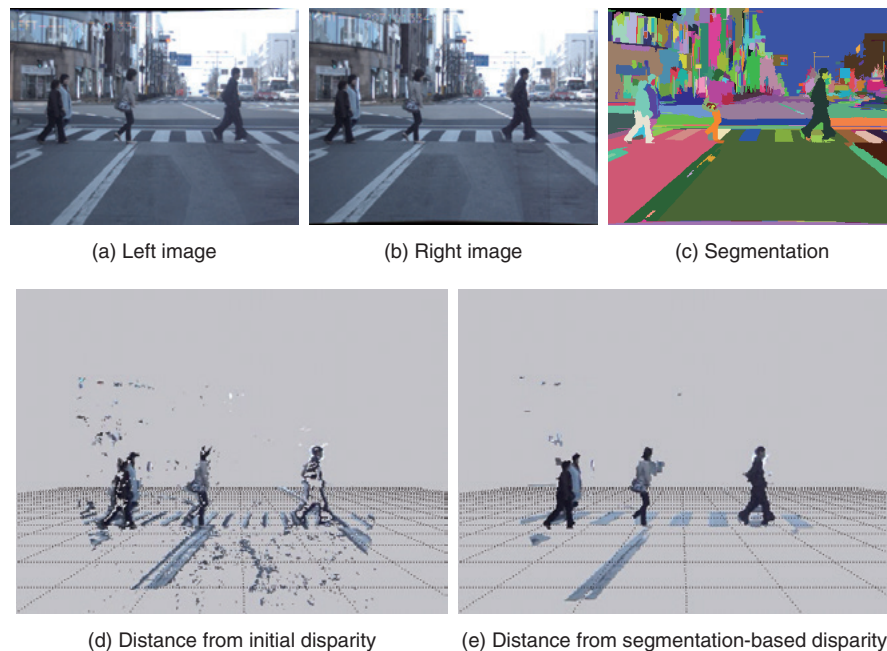


Fig. 10 Comparison of initial and segmentation-based disparity

segmentation of the right image. Figure 10(d) shows the results of estimation of the initial stereo disparity from block matching, and Fig. 10(e) shows the results of estimation of the stereo disparity from image segmentation. There is a high level of noise in the initial stereo disparity. As the torsos of the pedestrians indicate in particular, block matching is unable to measure disparity in textureless regions. In the image segmentation-based method, the use of a disparity model makes it possible to supplement the textureless regions, and the reduction of the effect of noise sharpens the boundaries of objects. This indicates that the initial stereo disparity calculated in the “Stereo disparity computation” process shown in Fig. 7 has increased in accuracy following the “Disparity model fitting” process shown in the same figure. Because there are few parameters and fitting is simple

for models in the case of stereo disparity, a quantitative evaluation like the one for the optical flows was not conducted, but in other scenes also, as in the case shown in Fig. 10, the use of the segmentation-based method helped to enable the achievement of better results for distance estimation than conventional block matching.

3.3. Experiment for Detection of Intersecting Objects

The method of detection of intersecting objects discussed above was tested using images collected on ordinary roads. The experiment used a scene in which the subject vehicle was driving towards a pedestrian crossing the road in a pedestrian crossing. Sufficient distance was maintained, and the images were captured under normal safety conditions. Figure 11 shows the results of detection

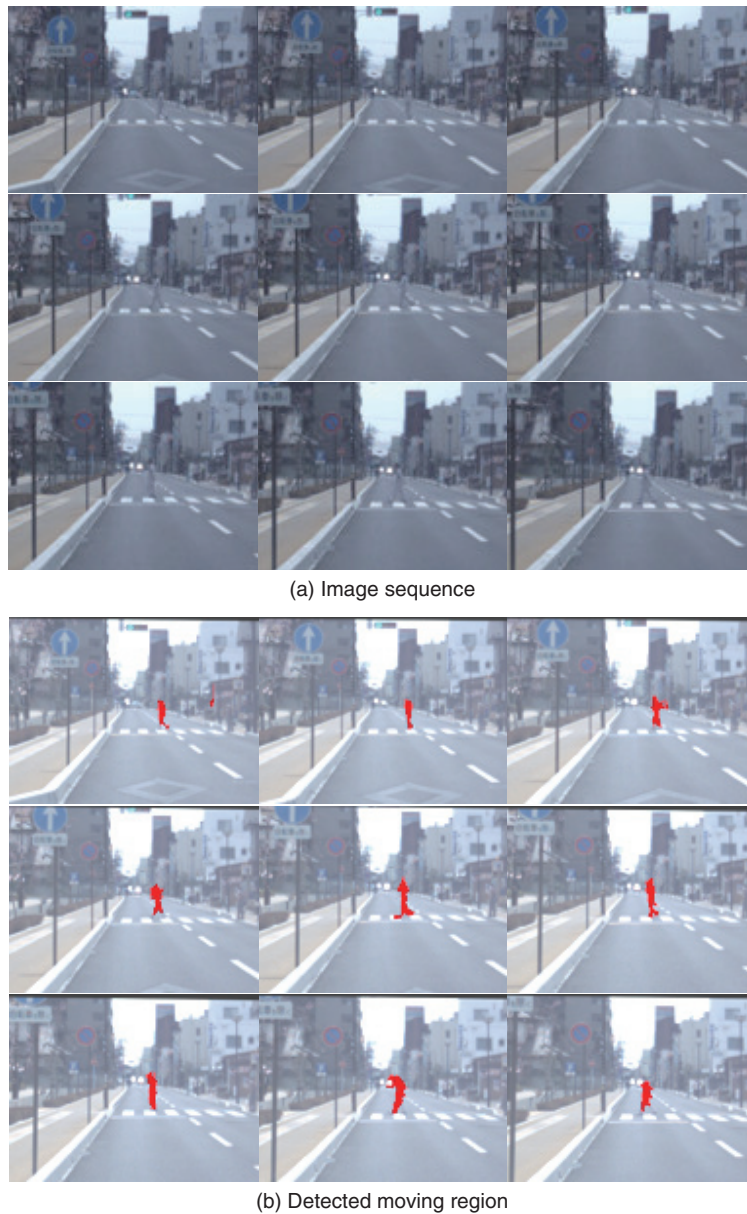


Fig. 11 Detection results for intersecting object

of the pedestrian. This figure shows the results of detection for each frame captured. The captured images are shown in Fig. 11(a). In the first frame, the pedestrian begins to enter the path of the subject vehicle, and reaches the center of the vehicle's path in the final frame. The regions painted out in Fig. 11(b) are the regions detected as belonging to a moving object. The pedestrian has been detected with no major instances of misdetection, and the boundaries of the pedestrian have also largely been correctly detected. For the sake of simplicity in this experiment, the region of the moving object was judged from horizontal motion only, and the moving objects that could be detected were limited to objects moving towards the FoE in the image. In addition,

in order to reduce error, post-processing that judged whether or not an object was moving based on size and position in the image was also employed.

3.4. Experiment for Detection of Preceding Objects

Next, the method of detection of preceding objects was applied to a scene captured on an expressway. In the scene used in this experiment, an overtaking vehicle was present in the lane next to the subject vehicle, and there was also another vehicle in front of the subject vehicle (Fig. 12). Figure 12(a) shows every eighth frame from the recorded images. Regions detected as belonging to moving objects are painted out in Fig. 12(b). The method

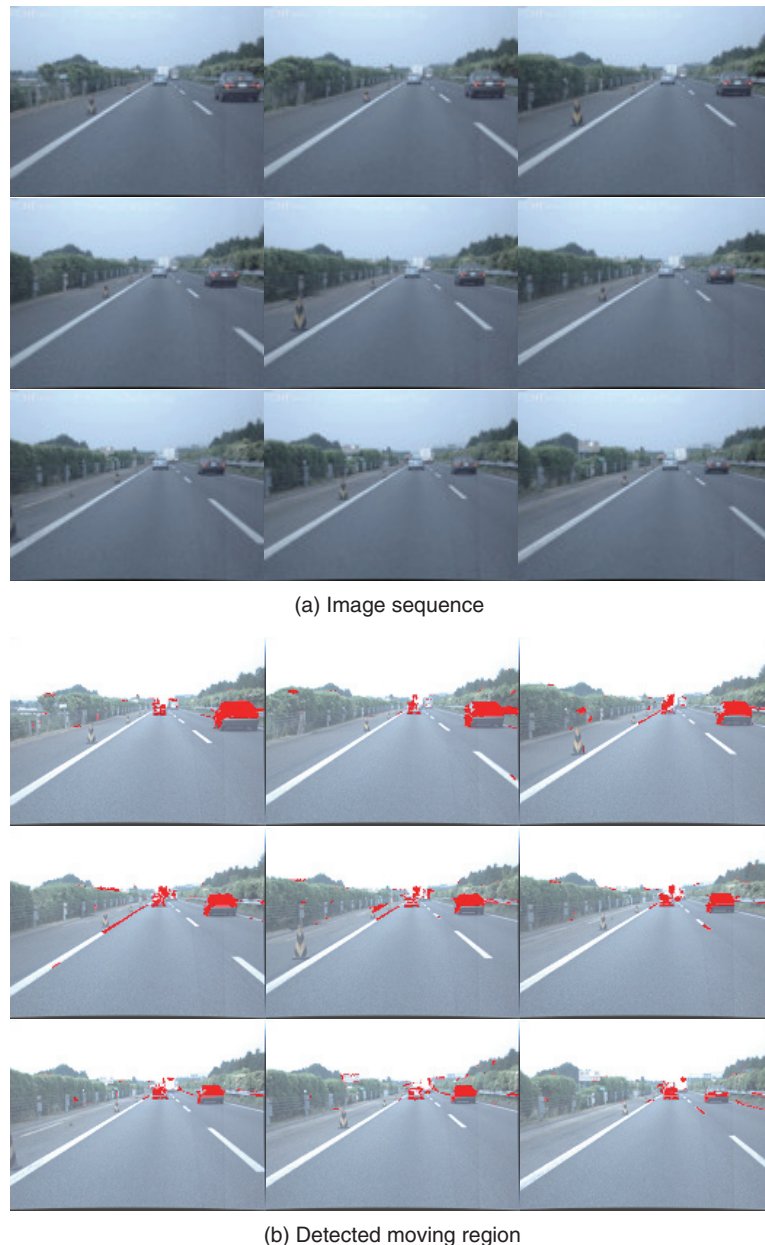


Fig. 12 Detection results for preceding object

employed made it possible to detect both the preceding vehicle and the vehicle in the next lane. In addition, other than in the shadowed areas close to the road surface, the method was able to detect the boundaries of the overtaking vehicle. In this experiment, judgment was made regarding the presence of moving objects for each frame and each pixel, and no post-processing to reduce misdetection was applied. Some misdetection can be observed, but it occurs intermittently, and it would be possible to reduce it through the employment of post-processing such as clustering or tracking.

4. Conclusion

Two methods of detecting independently moving objects from images captured by cameras mounted on vehicles were developed. The first of these algorithms is based on the use of a single camera, and can be employed when the camera and the object for detection (for example a pedestrian crossing the road or an intersecting vehicle) are moving in different directions. The second algorithm is based on the use of stereo cameras. This makes the system more complex, but enables detection of more general moving objects, including preceding objects, by comparing distance derived from stereo disparity and distance derived from motion. Experiments showed that by using this method, it was possible to obtain sharp boundaries for a moving object, and that as a result it was possible to detect the moving object.

The programs necessitate several tens of seconds of computation time for the processing of each frame. The reduction of this processing time will be an issue for the future. In addition, when employing color segmentation in the calculation of optical flows and stereo disparity, it has been assumed that the boundaries of the segments include the boundaries of the objects. However, there are cases in which misdetection or non-detection may occur when this assumption does not hold. Resolving this will also represent an issue for the future.

References

- (1) Agrawal, M., Konolige, K., Iocchi, L.: Real-time detection of independent motion using stereo, IEEE workshop on Motion and Video Computing (WACV/MOTIONS) (2005)
- (2) Vaudrey, T., Wedel, A., Rabe, C., Klappstein, J., Klette, R.: Evaluation of moving object segmentation comparing 6d-vision and monocular motion constraints, IVCNZ08, p. 1–6 (2008)
- (3) Hartley, R. I., Zisserman, A.: Multiple View Geometry in Computer Vision, Second Edition, Cambridge University Press (2004)
- (4) Ogale, A. S., Fermüller, C., Aloimonos, Y.: Detecting independent 3D movement, Handbook of Geometric Computing Applications in Pattern Recognition,

Computer Vision, Neural computing, and Robotics, E. Bayro-Corrochano, Springer Verlag (2005)

- (5) Hartley, R. I.: In defense of the eight-point algorithm, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 6, p. 580–593 (1997)
- (6) Domke, J., Aloimonos, Y.: A probabilistic framework for correspondence and egomotion, ICCV Workshop on Dynamic Vision (2005)
- (7) Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 5, p. 603–619 (2002)
- (8) Bleyer, M., Gelautz, M., Rhemann, C.: Colour segmentation-based computation of dense optical flow with application to video object segmentation, ÖGAI-Journal, Vol. 24, No. 1, p. 11–15 (2005)
- (9) Fischler, M. A., Bolles, R. C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM, Vol. 24, No. 6, p. 381–395 (1981)
- (10) Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M. J., Szeliski, R.: A database and evaluation methodology for optical flow, ICCV2007 (2007)

■ Author ■



Morimichi NISHIGAKI



Yiannis ALOIMONOS