Moving Obstacle Segmentation Using MMW Radar and Image Sequence

Shigeki Sugimoto Hidekazu Takahashi Masatoshi Okutomi
Department of Mechanical and Control Engineering,
Graduate School of Science and Engineering, Tokyo Institute of Technology
(O-okayama 2-12-1, Meguro-ku, Tokyo, 152-8552 Japan,
03-5734-3499, {shige, hidekazu, mxo}@ok.ctrl.titech.ac.jp)

Sensor fusion of millimeter-wave (MMW) radar and a camera is beneficial for advanced driver assistance functions such as obstacle avoidance and Stop&Go. In this paper, we propose a method for detecting a moving obstacle using MMW radar and CCD camera, along with a calibration method for two sensors. Our method is designed for detecting moving obstacles, such as cars, motorcycles, pedestrians, bicycles, and so on in urban areas, observed by a camera mounted on a vehicle. In the proposed method, we detect a moving obstacle by estimating image boundaries which enclose a group of feature points exactly on the obstacle. In order to determine the group, feature points in a whole image are detected at every image frame and tracked over several consecutive frames, and then we apply a motion segmentation technique, so-called subspace factorization, developed in the computer vision domain. For robustness in boundary estimation, we use MMW radar so as to detect the obstacle's rough position which results in an image region where an obstacle would exist; the motion segmentation technique is applied toward the tracks that drop in the region. Note that an obstacle's position detected by MMW radar is remarkably rough because MMW radar has low directional resolution. Nevertheless, the position contributes to rapid and robust estimation for satisfactory image boundaries. We demonstrate the validity of the proposed method through experiments using sensors that are mounted on a vehicle.

Keywords: *millimeter-wave radar, sensor fusion, moving obstacle detection, motion segmentation*

1. Introduction

In recent years, radar-based driver assistance systems such as Adaptive Cruise Control (ACC) have been introduced to the market by several car manufacturers. Most of these systems rely on millimeter-wave (MMW) radar for obtaining information about the vehicle's environment. In general use, a MMW radar is mounted on the front of a vehicle. It measures distance and relative velocity to targets at the front of the vehicle by scanning in a horizontal plane. For maintaining distance to a preceding target, it is necessary to detect lane markings by a camera in order to know whether the target lies in the current lane of the vehicle and to follow the target even if the road curves [1][2][3]. Compared with other long range radars (e.g., laser radar), MMW radar offers advantages of higher reliability in bad weather conditions. Thereby the ACCs work reliably and robustly.

However, most of these systems are designed for high-speed driving. A MMW radar provides relatively high distance resolution, but it has low directional (azimuth/elevation) resolution. Its directional resolution is sufficient for the ACCs for high-speed driving because it can be assumed that the vehicle is cruising in a low traffic density area. Furthermore, the positions of obstacles observed by the radar are limited to the space in front of the vehicle. Many moving obstacles such as automotives, pedestrians, bicycles, and so on exist in a crowded urban area. It is extremely difficult to detect these obstacles and measure their accurate positions by a radar with low directional resolution.

In contrast to MMW radar, a camera provides high spatial resolution but low accuracy in estimation of the distance to an obstacle. The high spatial resolution of the camera can support the low directional resolution of the radar, and the high distance resolution of the radar can support the low accuracy in distance estimation of the camera. Thereby, MMW radar plus camera can be mutually supportive: their sensor fusion offers benefits for more advanced driver assistance functions such as obstacle avoidance and Stop&Go.

Recently, sensor fusion of laser radar and camera(s) has been developed by several researchers (e.g., [4] [5] [6]). Laser radar has higher special resolution than MMW radar; therefore it can detect an occupying area (i.e. an area where an object obstructs on the road). However it is more economical to use existing radar-camera system for more advanced functions.

In this paper, we propose a method for moving obstacle detection using MMW radar and CCD camera. Our method is designed for detecting obstacles, which move across a driver's view in a small range, observed by a camera mounted on a vehicle. As shown in Figure 1, we use a MMW radar for detecting the rough position of an obstacle, and then we decide a candidate image region where the obstacle would exist by projecting the detected position into an image sequence acquired by a camera.

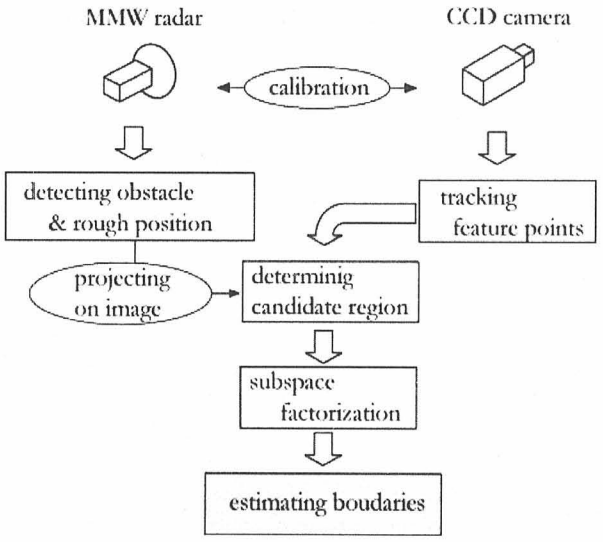


Figure 1. Diagram of proposed method

Before projecting the position, feature points are extracted at every image frame and tracked over several frames. Once the candidate region is decided, we select the tracks that drop in the region, and apply a motion segmentation technique toward the tracks. The motion segmentation technique is the so-called 'subspace factorization' or 'subspace separation' in computer vision domain ([8] [9] [10] [11]). We estimate a linear motion subspace composed by three tracks of the feature points exactly on a single moving obstacle using the least median of squares method (LMedS). Then we select the tracks of the obstacle by evaluating residuals (See Section 4 for details). Finally we decide boundaries which enclose the feature points on the end of the extracted tracks.

The subspace factorization technique is formulated by a pure mathematical theory and produces a robust motion segmentation procedure. In the theory, linear subspaces are formed under an assumption that object motions are observed by a camera modeled by an affine projection (e.g. parallel, weak-perspective or para-perspective projection). In general, the camera should be modeled by the perspective projection; therefore, the assumption can not hold for motions in a real scene (See Subsection 4.1). However, considering only the motion of a single moving obstacle, on the contrary, we can assume that the motion is observed approximately by a camera with an affine projection. If we could select correct tracks which belong to the obstacle, it is possible to estimate the motion subspace that represents the obstacle's motion and to select all feature points exactly on the obstacle. In order to select the correct tracks, we make use of obstacle detection results from the radar. By projecting a rough position detected by the radar into the image sequence, we decide an image region where an obstacle would exist. By limiting image region and tracks to be processed, we can robustly estimate the subspace. Furthermore, we can reduce computational cost by reducing the number of repeated times in LMedS, compared to the case when the region is not decided.

As a previous work, a method for obstacle detection using MMW radar and a single camera is proposed in [13]. This method is based on motion stereo and designed for detecting obstacles which come from far to near. In this case, the relative speed between the vehicle and an obstacle should be large for robustly estimating image boundaries of the obstacle. On the other hand, our method is designed for detecting an obstacle which meanders from outside to inside the course of the vehicle. This is the dangerous situation that obstacles, such as bicycles and pedestrians, suddenly appear in front of the vehicle especially in urban areas.

For sensor fusion of MMW radar and camera, calibration of their locations is also an important issue because flexible sensor locations are required for car design. The calibration method should be simple and easy for mass production. However, in past research for the sensor fusion of radar and camera (e.g., [5] [6] [13]), the sensor locations are constrained strictly and the calibration method is not explicitly mentioned.

In this paper, we also propose a calibration method for the fusion of radar and camera. Generally, calibration of sensor location requires estimation of the transformation between sensor coordinates. In the proposed method, we simply estimate the homography that describes transformation between a radar plane (which is scanned by radar) and an image plane. We use the calibration result for projecting an obstacle's position detected by the radar into the image sequence.

The proposed calibration method can be applied for line-scanning radar with other wave-lengths such as laser and infrared-ray. The method may be useful for a newly-mounted camera and an existing radar system on a vehicle. Once the homography is obtained, we can easily visualize radar reflections by projecting reflection points into an image. We know exactly what causes the radar reflections by looking up the visualized points in the image. This approach can be the basis for the sensor fusion of line-scanning radar and camera.

In the remainder of this paper, the proposed calibration method is described in Section 2. Section 3 describes a simple method for obstacle detection from radar signals. The motion segmentation technique is described in Section 4. Section 5 shows experimental results, and Section 6 concludes this paper.

2. Calibration between Radar and Camera

We use a MMW radar which scans in a plane, called the 'radar plane'. The radar outputs radial distance r, angle θ, relative radial velocity s, and reflection intensity for every reflection; it acquires many data points of radar

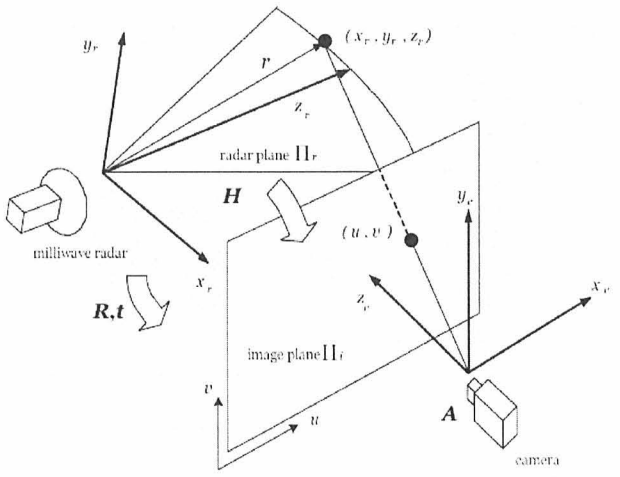


Figure 2. Geometry of radar and camera

reflections for each scan. In general, we can assume that all reflection points exist on the radar plane. Accordingly, we estimate the homography that describes the relationship between the radar plane and the image plane.

2.1. Geometry of radar and camera

As shown in Figure 2, let (x_r, y_r, z_r) and (x_c, y_c, z_c) be the radar and the camera coordinates respectively, and (u,v) be the image plane coordinates. Using homogeneous coordinates, we can describe the equation of transformation between $(x_r, y_r, z_r, 1)$ and $(u,v,1)$ as follows.

$$\omega \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{P} \begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix}, \quad \mathbf{P} = \mathbf{A}[\mathbf{R} \mid \mathbf{t}] \tag{1}$$

In the above equation, the 3×3 matrix **R** and the 3×1 vector **t** denote, respectively, the rotation and translation between the sensor's coordinates; the 3×3 matrix **A** denotes intrinsic camera parameters, and ω is an unknown constant. Generally, calibration between the two sensors requires estimation of the 3×4 matrix **P**, or all of the **R**, **t**, and **A**. On the contrary, we describe the transformation between the radar plane Π_r and the image plane Π_i, as described below.

Considering that all radar data come from somewhere on the radar plane $y_r = 0$, the equation (1) is converted such that

$$\omega \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{H} \begin{bmatrix} x_r \\ z_r \\ 1 \end{bmatrix} \tag{2}$$

where **H** is the 3×3 homography matrix. By estimating **H**, the transformation between the radar plane Π_r and the image plane Π_i is determined without solving **R**, **t**, and **A**.

We use the least squared estimation using more than four data sets of (u,v) and (x_r, z_r) for estimating **H**.

2.2. Determination of Corresponding Data Sets

Generally, a MMW radar has an azimuth/elevation beam width of more than several degrees, which may result from its antenna directivity. It causes low directional resolution of the radar. Therefore, determining accurate reflection positions is a difficult work. However, we can expect the beam center has maximum amplitude; that is, an object in the crossing point of the radar plane encounters maximum reflection intensity.

As shown in Figure 3, we observe radar reflections and acquire frame data while moving a small corner reflector up and down so that it crosses the radar plane. The frame data contain reflections not only from the reflector but also from the environment. For determining the reflector's reflection in each acquired frame, a signal with maximum intensity is extracted (its radial distance r and angle θ are also recorded); thereby we obtain an intensity sequence. With the intensity sequence, we detect local intensity peaks for deciding crossing points of the radar plane. Corresponding radii and angles to the intensity peaks are converted into Cartesian coordinates by $x_r = r\cos\theta$ and $z_r = r\sin\theta$.

The image sequence is acquired simultaneously by the camera. We extract image frames which correspond to intensity peaks. Then, the reflector's position (u,v) on

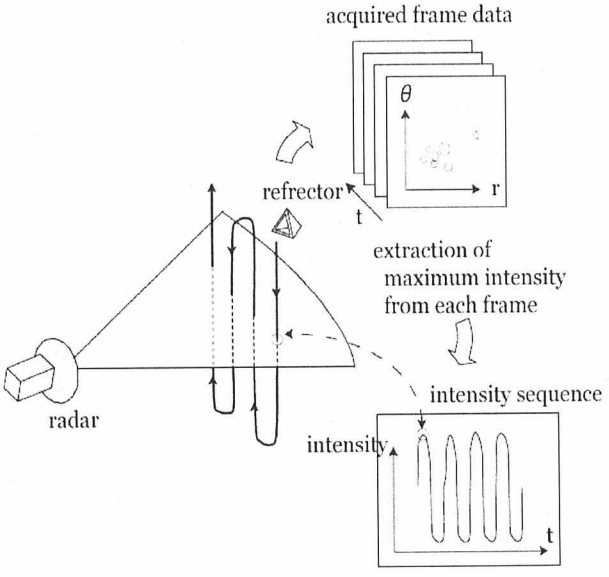


Figure 3. Decision process for the radar plane

each image frame is estimated by a template-matching algorithm. Thereby we obtain the data sets that denote positions on the image plane and the radar plane in the equation (2).

3. Obstacle Detection by Radar

We use the radar in order to detect an obstacle and its rough position that contributes to the motion segmentation technique described in Section 4. In this section, a simple obstacle detection method is presented. We also explain the image region resulting from an obstacle's position detected by the radar.

3.1. Radar Data Segmentation

As mentioned in Section 2, the radar outputs (r,θ,s) along with reflection intensity for every reflection. We simply segment the radar data into clusters using a nearest neighbor algorithm for detecting obstacles.

Data in each radar frame are sparse and spread on the radar plane. They include a lot of errors caused by diffractions, multiple reflections, and Doppler shift calculation failures. For reducing influences of such errors, we first eliminate data which has small reflection intensity. Then we segment the remaining data into clusters by sequentially connecting two data points which have similar (r,θ,s) using pre-defined thresholds for r, θ and s. Finally, isolated data are eliminated again.

3.2. Projection of Radar Data

As shown in Figure 4, we select data in a cluster, and project every data position (x_c,z_c) (converted from r,θ) into the image plane using the homography described in Section 2. The radar and the camera are mounted at the front of the vehicle above and below (See Figure 5). In this case, projected points in the image stand in a nearly horizontal line as shown in Figure 4. Then we draw a rectangle enclosing all projected points. The size of the rectangle is decided as follows; the right border is at 50 pixels to the right of the right-most point; the left border is at 50 pixels to the left of the left-most point; the height is decided by a value which is inversely proportion to the distance r.

We treat the image area enclosed by the rectangle as a candidate region where an obstacle would exist. Although the position and the size of the rectangle are determined roughly and adequately, the region plays an important role for detecting the correct obstacle's boundaries in the image, as described in the following section.

4. Motion Segmentation on Image Sequence

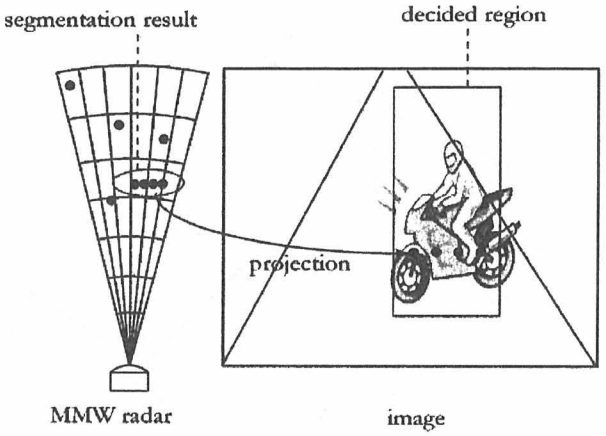


Figure 4. Decision of image region

For detecting the obstacle's boundaries in the image, we apply a motion segmentation technique based on feature tracking. The technique is well known in computer vision domain as 'subspace factorization'. In this section, we roughly describe the motion subspace theorem which is a fundamental of subspace factorization. Then we present details of a motion segmentation process for detecting boundaries.

4.1. Motion Subspace Theorem

Let N be a number of feature points which lie in an object. We track the feature points over M images. Let $(x_{\kappa\alpha}, y_{\kappa\alpha})$ be the image coordinates of the α-th feature point in the κ-th frame. All of the image coordinates are stacked vertically into a $2M$-dimensional vector in the form as follows.

$$\mathbf{p}_\alpha = (x_{1\alpha} \quad y_{1\alpha} \quad x_{2\alpha} \quad y_{2\alpha} \cdots \quad y_{M\alpha})^T, \tag{3}$$

The motion of the α-th point is represented by a single point $\mathbf{p}_\alpha$ in a $2M$-dimensional space: that is, $\mathbf{p}_\alpha$ represents a track of the α-th feature point over M images.

If we assumed that the motion is observed by an affine camera, $\mathbf{p}_\alpha$ has the form [8]

$$\mathbf{p}_\alpha = \mathbf{m}_0 + a_\alpha \mathbf{m}_1 + b_\alpha \mathbf{m}_2 + c_\alpha \mathbf{m}_3, \tag{4}$$

where $(a_\alpha, b_\alpha, c_\alpha)$ are the object coordinates of the αth points, and $(\mathbf{m}_0, \mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3)$ are the $2M$-dimensional vectors composed by the intrinsic parameters of the camera and the geometric relationship between the image and the object coordinates. The equation (4) represents that the N tracks $\{\mathbf{p}_\alpha\}$ belong to the 4-dimensional linear subspace, the so-called 'motion subspace', spanned by the vectors $(\mathbf{m}_0, \mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3)$. In fact, the tracks $\{\mathbf{p}_\alpha\}$ belong to the 3-dimensional motion subspace spanned by the vectors $\{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3\}$, because the vector $\mathbf{m}_0$ can be chosen arbitrarily as the origin of the object coordinates.

The 3-dimensional subspace is formed by 3 bases of the subspace; we can estimate the 3 bases using more than three tracks of the feature points (We do not have to estimate $a_\alpha, b_\alpha, c_\alpha, \mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3$). Once the subspace is estimated, we can select the tracks that belong to the subspace (i.e. the object) by evaluating residuals of individual tracks with respect to the subspace.

A motion segmentation algorithm based on the equation (4) had been presented in [8] first. Since then, various extensions have been proposed for robustness and reduction of computational cost (e.g. [9] [10] [11]). These methods are basically designed for segmenting multiple objects in an image sequence. In contrast, our approach is rather simple.

In order to estimate a motion subspace of a single moving obstacle, we need to select three tracks of the feature points exactly on the obstacle. In general, however, it is difficult to do it from an entire image sequence without any clues. In our case, the radar helps us to detect obstacle's rough position. If we know the image region where the obstacle exists, it is much easier to select the three tracks for estimating the 3-dimensional subspace. For example, we can select three tracks which drop in the region (i.e. the ends of the tracks exist in the region). However, as mentioned in Section 3, the candidate region is decided roughly. There are outliers when we randomly select such three tracks from the candidate region. Therefore we use LMedS as a robust estimator, as described in the next subsection.

On the other hand, tracks of moving obstacles observed by a moving camera are, in general, not represented by the equation (4), because it cannot be assumed that motions are observed by a camera modeled by an affine projection. For example, in the case of background motion observed by a moving camera, the distances between individual 3-D points on the background and the optical center of the camera are generally distributed in a wide range. In such a case, the camera should be modeled by the perspective projection. Limited on a moving obstacle, however, the distances are distributed in a relatively small range. In this case, the camera can be modeled by an affine projection; i.e. the tracks of the obstacle are approximately represented by (4). Therefore we can expect that the motion segmentation technique based on the linear motion subspace works effectively.

4.2. Motion Segmentation and Boundary Detection

Details of the method for estimating the boundaries of the obstacle detected by the radar are described in this subsection. We estimate the motion subspace and the boundaries by the following steps.

1) We extract feature points at each frame. Then the points are tracked over five frames and their image coordinates in every frame are preserved. For extracting and tracking feature points, we use the KLT tracker in OpenCV Library available from [14]. Note that tracking should be done before the radar detects obstacles. That is, we immediately use preserved tracks (over just five frames) when an obstacle is detected by the radar.

2) After deciding the candidate region described in Section 3, we take the preserved tracks and select a set of tracks which drop in the region. Before that, tracks which have no motion are eliminated because we are interested only in the moving obstacle. In addition, such tracks often engender errors in estimation of a subspace in Step 4) and 6). For each track, we calculate the Euclid distance between two coordinates on the beginning and the end of the track. If the distance is smaller than 7 [pix] the track is eliminated.

3) Three tracks are randomly selected from the set.

4) A subspace is fitted to the selected three tracks in the following process.

Let $(\mathbf{p}_a, \mathbf{p}_b, \mathbf{p}_c)$, $(1 \le a,b,c \le N, a \ne b \ne c)$ be the selected three tracks that represent 3 points in the $2M$-dimensional space (M=5 in this case). Also, let $\mathbf{p}_G$ be the center of the selected three tracks in the space. All tracks in the set are shifted as follows.

$$\bar{\mathbf{p}}_i = \sum_{i=1}^{N} (\mathbf{p}_i - \mathbf{p}_G) \tag{5}$$

The $2M \times 2M$ moment matrix $\mathbf{M}$ is calculated by

$$\mathbf{M} = \sum_{i=a,b,c} (\bar{\mathbf{p}}_i \bar{\mathbf{p}}_i^T) . \tag{6}$$

We compute its eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_{2M}$ and corresponding eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_{2M}$. The 3 vectors $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ are an orthonormal basis of a 3-dimensional subspace which is a candidate for the motion subspace of the moving obstacle detected by the radar.

5) We calculate individual residuals ε_i^2 of the tracks in the set with respect to the subspace fitted in Step 4) as follows.

$$\varepsilon_i^2 = \sum_{j=4}^{2M} (\bar{\mathbf{p}}_i^T \mathbf{u}_j)^2, \quad i \ne a,b,c \tag{7}$$

Then the median $\varepsilon_{(a,b,c)}^2 = \mathrm{median}(\varepsilon_i^2)$ is preserved. We use the median so as to evaluate the selected three tracks $(\mathbf{p}_a, \mathbf{p}_b, \mathbf{p}_c)$. Note that we assume that more than 50% of the tracks in the set belong to the obstacle. If all of $\mathbf{p}_a, \mathbf{p}_b, \mathbf{p}_c$ are exactly on the obstacle, the median has a small value, otherwise a large value.

6) By repeating Step 3) ~ 5) a number of times n

mentioned below, we obtain n values $\{\varepsilon_{(a,b,c)}{}^2\}$ as a collection of medians. Then we take the smallest value from the $\{\varepsilon_{(a,b,c)}{}^2\}$ and the corresponding three tracks. We decide that the 3 eigenvectors $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ corresponding to the three tracks with the smallest $\varepsilon_{(a,b,c)}{}^2$ span the motion subspace for the moving obstacle.

7) After that, we evaluate individual residuals of all tracks in the image sequence (for five frames) with respect to the subspace estimated in Step 6). Every track with smaller residual than a pre-defined threshold is selected.

Finally, we extract the feature points on the end of the selected tracks. The obstacle's boundaries are decided by drawing a rectangle which encloses all of the extracted feature points.

For estimating the correct subspace resulting from the obstacle's motion, we need to repeat Step 3) ~ 5) a sufficient number of times. We decide the repeated number n by the following criteria.

We decide the repeated number n so that the correct three tracks of the obstacle are selected at least one time by a 99.99% probability. The number of times n can be calculated as follows.

$$\left(\frac{{}_N C_3 - {}_{N_0} C_3}{{}_N C_3}\right)^n < 0.0001, \tag{8}$$

where N is the number of tracks in the set mentioned in Step 2), and N_0 is the number of the tracks of the obstacle.

As mentioned in Step 5), we assume that more than 50% of the tracks in the set exist on the obstacle. This assumption is reasonable because the set of tracks are decided from the candidate region. In experiments, the number of tracks (feature points) in the set is no more than 100. However the number of repeated times is about 70 which enable real-time processing for estimating obstacle boundaries.

Note that the candidate region described in Section 3 does not need to completely enclose the correct obstacle's region in the image. Our method works even when the radar observes a part of reflections from the obstacle because, as mentioned in Step 6), we extract all

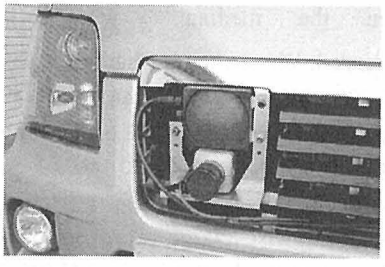


Figure 5. Car-mounted sensors

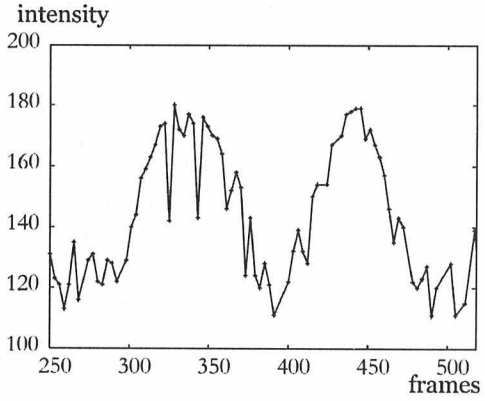


Figure 6. Intensity sequence

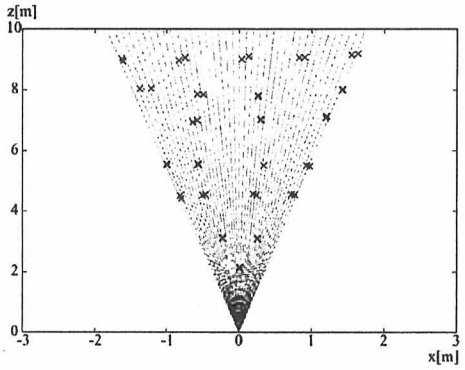


Figure 7. Calibration points (radar)

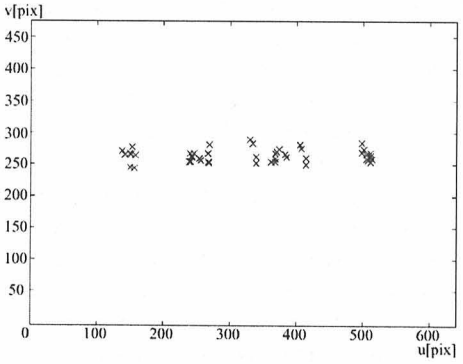


Figure 8. Calibration points (image)

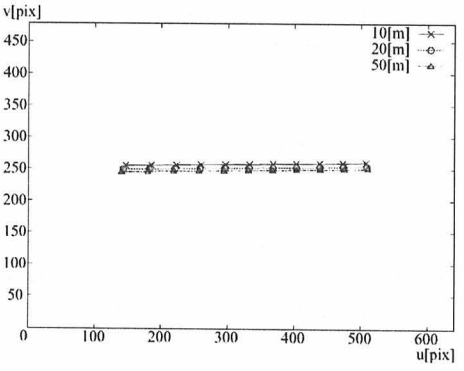
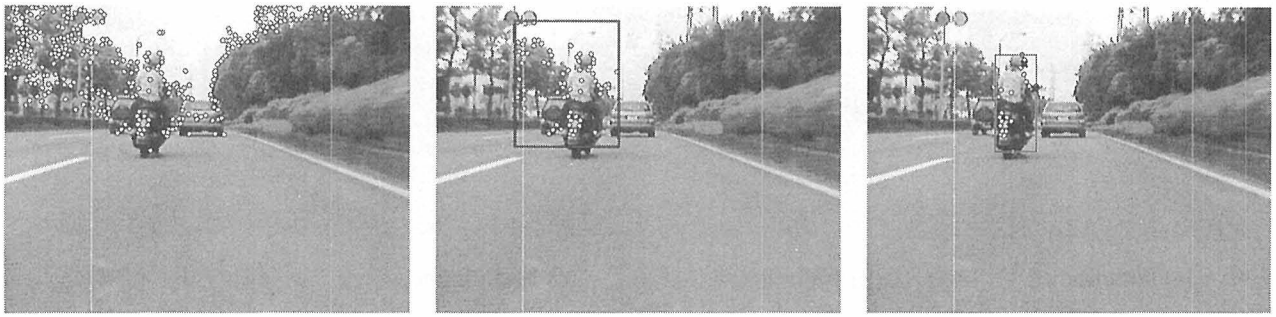


Figure 9. Calibration result



(a) Tracked feature points (b) Candidate image region (c) Detected boundaries

Figure 10. Obstacle Detection

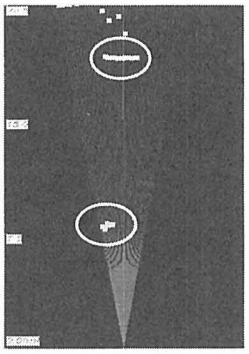


Figure 11. Radar data segmentation

feature points that belong to the motion subspace.

5. Experimental Results

In this section, we show a calibration result and obstacle detection results using real radar/image frame sequences observed in urban areas.

We mounted the radar and the camera at the front of the vehicle as shown in Figure 5. We used a line-scanning MMW radar which was offered by a radar manufacturer for experimental uses. The detection range was 50[m] and the scanning angle was 20.9[deg] horizontally. The radial and lateral resolution were 0.1[m] and 0.55[deg], respectively. Its frame rate was 10 [fps]. The color camera was SONY DFW V500 with 640x480[pix] and 30 [fps].

5.1. Calibration

Figure 6 shows an example of the intensity sequence described in Section 2. The 46 data sets, which represent positions on the radar plane and the image plane, are shown in Figure 7 and Figure 8, respectively. We estimated the homography matrix **H** using these data sets. Figure 9 shows transformed positions (the radius between 10-50m and the angle between -10-10 degree) on the radar plane to the image plane using **H**.

Figure 6 indicates that the radar fails to acquire the correct reflection intensity of the reflector at some frames, which represents the lack of stability of radar observation. Extracted points on both planes are influenced by this lack of stability. However, the calibration result in Figure 9 reasonably indicates the actual sensor arrangement, i.e. the radar is located above the camera, and scanning directions of the radar are nearly parallel to the y_c axis of the image.

5.2. Moving Obstacle Detection

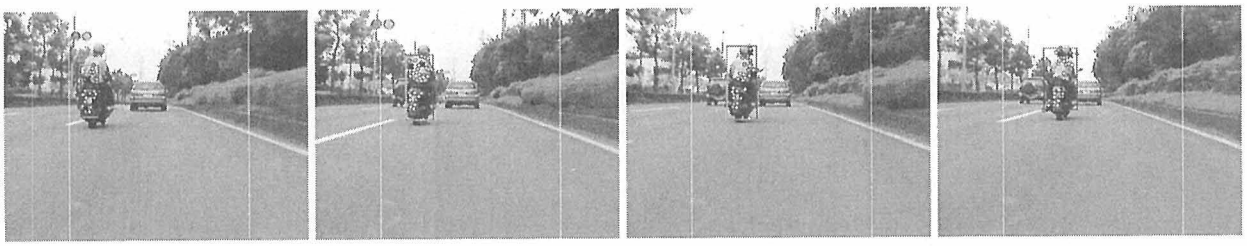
Figure 10 and 11 show an acquired image and a radar frame in an urban area, respectively. In the scene, a motorcycle was moving from the left to the center in the front area of the vehicle (it moved in a relative range from 7 [m] to 8[m] in the scene).

In Figure 10 (a), we show the feature points tracked over the last five frames. Only the latest image coordinates are shown by the green points. Two vertical lines on the left and right parts of the image frame indicate the right-most and the left-most limits of the radar's scanning angle, respectively. (We estimated the limits from the calibration result.)

The corresponding radar frame is shown in Figure 11. As a result of the obstacle detection method described in Section 3, two obstacles (a motorcycle about 8[m] away from the vehicle and a car about 20[m] away) are detected by the radar.

Figure 10 (b) shows the estimated candidate region that corresponds to the motorcycle. The region is shown by the rectangle. The tracks that drop in the region are shown by the small circles. After eliminating motionless tracks, 62 feature points remained.

Figure 10 (c) shows the result of the motion segmentation method described in Section 4. The detected boundaries of the motorcycle are shown by a rectangle. The three 'x' points indicate the three tracks (note that only the last image coordinates are shown) that have the smallest median value; that is, the three tracks determine the subspace of the motion. The number of the feature points on the motorcycle is 34 (indicated by small



(a) 90th frame (b) 99th frame (c) 108th frame (d) 117th frame

Figure 12. Detected boundaries on image sequence

circles in the rectangle).

Figure 10 (c) indicates the validity of the proposed method. In this case, the proposed method using subspace factorization works perfectly. In Figure 10 (b), we can see that several feature points out of an obstacle's area exist in the candidate region. However, all outliers are eliminated and the feature points exactly on the obstacle remain in Figure 10 (c).

Figure 12 shows an image sequence of the same scene as Figure 10. At some frames, we failed to detect some feature points around the driver's head; heights of the rectangles at individual frames are different. However, we can see that the proposed method satisfactorily caught the obstacle's boundaries.

Figure 13 shows an image sequence of a dusky scene. In this case, the vehicle turned to the right after a car passes (the car was in a relative range from 8[m] to 10[m]). In spite of the darkness, feature points are extracted well and we could detect boundaries of the car satisfactorily.

For evaluating the proposed method, we count a number of image frames where valid boundaries are detected.

We firstly select consecutive image frames in which a moving obstacle is observed in an image area where the radar can detect it (The area is indicated by the two vertical lines in Figure 10 (a)). We refer to the selected frame as the 'base frame'. The base frame is decided by the following steps.

(1) We select image frames where a target (the motorcycle in the case of Figure 12) is observed in the image area. For each frame, we manually set a rectangle which just encloses the target in the current frame. We refer to this rectangle as the 'base rectangle' (See Figure 14).

(2) For each frame, the feature points in the 5th frame before are tracked to the current frames, and then we remove every track which has a small motion distance (less than 7 [pix]) between the coordinates in the first frame and in the last. If more than 10 tracks dropped in the base rectangle, we select the image frame.

In the case of the scene Figure 12, 49 consecutive frames (for about 1.6 [sec]) were selected as the base frames. In the case of Figure 13, 80 frames (for about 2.6[sec]) were selected.

We need a valid result of the obstacle detection by the radar in order to succeed in the moving obstacle segmentation. We check the candidate regions estimated from the obstacle's positions detected by the radar. As shown in Figure 14, we compare the candidate region (rectangle) with the base rectangle. If the overlapped area occupies more than 50% of the base rectangle, we conclude that we have succeeded in estimation of the candidate region; we do not care how big the candidate region is. Note that the frame rate of the radar was 10[fps] (the camera had 30[fps]); therefore, we use the candidate region estimated by radar frames toward the consecutive three image frames.

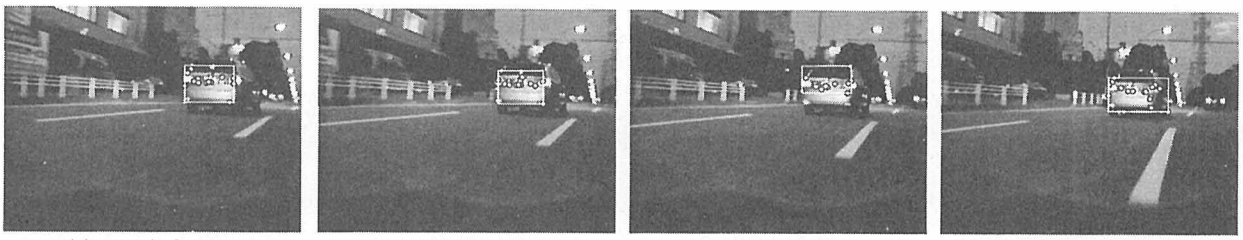
Finally we evaluate the proposed method by counting the number of frames where detection of boundaries has succeeded. For evaluation, the motion segmentation method is done for the frames where the candidate region detection has succeeded. We decide the success in the moving obstacle segmentation by the following criteria.

(1) The feature points that construct the obstacle's boundaries (See Section 4.2, Step 6) are composed of only points exactly on the target in the image.

(2) The overlapped area in Figure 15 occupies more than 65% of the base rectangle.

We expect the estimated boundaries should be fully enclosed by the base rectangle. We carefully check both the boundaries and the feature points within the boundaries. The area rate in (2) is decided by considering that the height and width of the boundaries should be more than 80% of the length of the base rectangle.

Table.1 shows evaluation results for the two scenes shown by Figure 12 and 13. As mentioned in Section 3, radar data contains a lot of errors. In addition, slanted or small obstacles might be missed because of the weak reflections. These errors and misses affect the obstacle detection by the radar, and we failed at some frames in detection of the valid candidate as shown in Table 1(b). On the other hand, motion segmentation succeeded in many frames. In the case of scene 1 (Figure 12), correct



(a) 120th frame (b) 129th frame (c) 138th frame (d) 147th frame

Figure 13. Detected boundaries on image sequence (dusky scene)

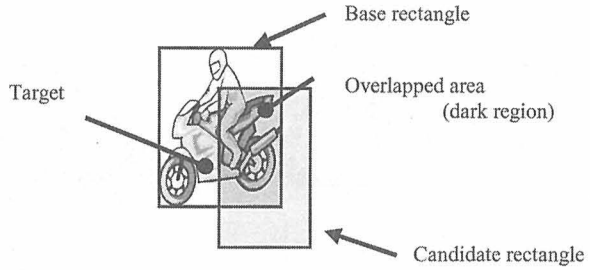


Figure 14. Overlapped area for evaluating candidate region.

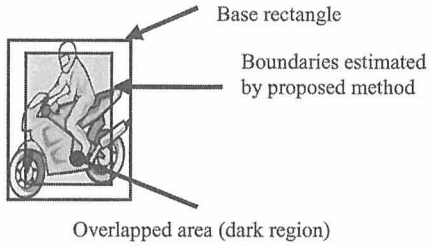


Figure 15. Overlapped area for evaluating obstacle boundaries.

	(a)Number of base frames	(b)Number of image frames where estimation of valid candidate region is succeeded.	(c)Number of image frames where boundary detection by the proposed method is succeeded (rate for (b)).
Scene 1 (Figure11)	49	39	37(97.44 %)
Scene 2 (Figure12)	80	60	50(83.33 %)

Table 1. Evaluation results

boundaries are estimated at 97% for the number of frames that succeeded in candidate detection. The rate is smaller in the case of scene 2 (Figure 13). This is mainly because darkness in the scene and highlighted feature points cause incorrect feature tracking. However, boundary detection is succeeded in more than 80% of frames after obstacle detection by the radar.

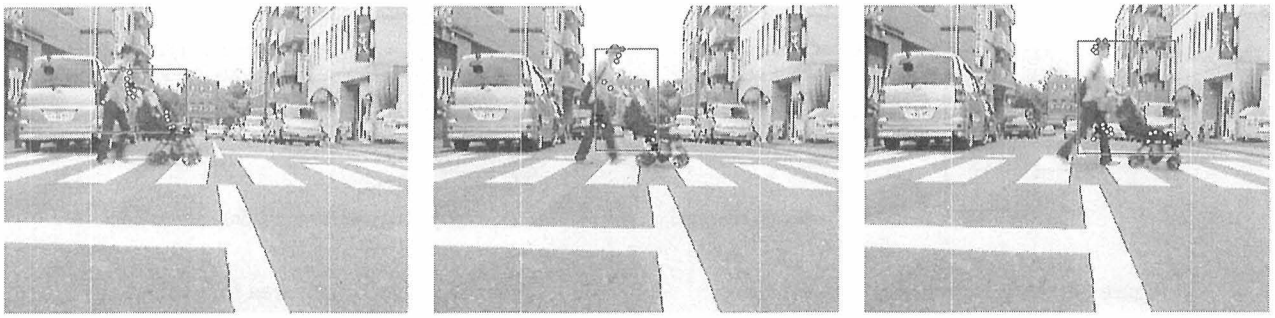
Figure 16 shows examples of boundary detection by the proposed method for a pedestrian and a baby carriage. The intervals among the left, the middle and the right column are 9 frames (0.3[sec]). In figure 16(a), a pedestrian walked across from the left to the right on the scene (in a range from 4[m] to 6[m]). In Figure 16(b) a pedestrian with a baby carriage walked across from the left to the right (in a range about 5[m]). We adjusted sensitivity of the radar outputs because radar reflections from human body and cloth material are very weak. Detection of the candidate region is missed especially in the case of a pedestrian when the radar has low sensitivity.

It is important to note that estimation of the correct boundaries of a pedestrian by the subspace theorem is difficult, because the theorem described in Section 4 is valid only for rigid objects. The human body is not rigid and changes its shape while it walks on. In the case of Figure 15, we hardly detect the whole of the pedestrian's body. However, in most of frames, we could estimate the region of the center of the body, and extract the correct feature points that belong to the pedestrian.

Through the experiments, we eliminate the tracked feature points that have small motion (less than 7 [pix] per 5 frames). In the case of Figure 15, a pedestrian moved 30 [pix] per 5 frames at a range of 5[m]. The value 7 [pix] is small enough to detect a pedestrian which moves across in front of the vehicle at a range in real scene.

In the experiments, a PC with a Pentium IV 3GHz was used. The detection processes described in Section 3 and 4 were implemented in C-language programs. We use OpenCV libraries [14] for not only tracking feature points but also calculating matrixes and eigenvectors through the experiments. According to the experiment,

(a). Example of detection results for a pedestrian.



(b). Example of detection results for a baby carriage.

Figure 16. Examples of detection results.

the maximum computational time was 0.15 [sec/frame] for detecting boundaries of an obstacle (including tracking feature points over five frames). If we do not use the radar, and if we detect obstacles using only image sequence by applying the subspace factorization technique without any clues, the computational time is more than a few seconds without satisfactorily results. Thereby radar and the subspace factorization technique for image sequence are a proper choice for detecting obstacles in respect to robustness and computational efficiency.

7. Summary

We proposed a calibration method for sensor fusion of MMW radar and camera, and an obstacle detection method using the subspace factorization technique. For detecting boundaries of an obstacle, we used the radar for deciding the obstacle's rough position in the image sequence. We showed the validity of the proposed method through the experiments.

Through this work, we confirm that the motion observed by a car-mounted camera can be approximately treated as a motion observed by an affine camera. This means that several extensions of the subspace separation technique can be applied for advanced driver assistance functions. If moving obstacles are segmented from the background perfectly, we can analyze and recognize the obstacles using various recognition algorithms.

Our method is mainly designed for detecting the obstacles that move to block the course of the vehicle. For such an objective, the proposed method seems to work very well. However, motion segmentation results depend not only on obstacle's motion but motion of the background. Therefore we will analyze the motion subspace theorem and evaluate segmentation results considering more various movements in the background in real scenes, along with results for non-rigid objects such as a human body.

It is also important to detect obstacles which lie in the background such as walls, poles, guard rails, and so on. For detecting these obstacles, a segmentation method based on a distance measure is effective. Accordingly, in the future work, we will research a method for applying the proposed method to stereo vision.

8. Acknowledgements

The authors would like to thank Hayato Tateda for discussions and supports through the experiments.

9. References

[1] D. Langer and T. Jachem, "Fusion Radar and Vision for Detecting, Classifying and Avoiding Roadway Obstacles", in Proceedings of the IEEE Symposium on Intelligent Vehicles, 1998, pp.333-338.

[2] D. Langer, "An Integrated MMW Radar System for Outdoor Navigation", PhD Thesis, Robotics Institute, Carnegie Mellon University, 1997.

[3] M. Beauvais and S. Lakshmanan, "CLARK: A Heterogeneous Sensor Fusion Method for Finding Lanes and Obstacles", in Proceedings of the IEEE International Conference on Intelligent Vehicles, 1998, pp.475-480.

[4] J. Laneurit, C. Blanc, R. Chapuis, and L. Trassoudaine, "Multi-Sensorial Data Fusion for Global Vehicle and Obstacles Absolute Positioning", in Proceedings of the IEEE Intelligent Vehicle Symposium, 2003, pp128-143.

[5] R. Aufrere, C. Mertz, and C. Thorpe, "Multiple Sensor Fusion for Detecting Location of Curbs, Walls, and Barriers", in Proceedings of the IEEE Intelligent Vehicle Symposium, 2003, pp.126-131.

[6] S. Mockel, F. Scherer, and P. F. Schuster, "Multi-Sensor Obstacle Detection on Railway Tracks", in Proceedings of the IEEE Intelligent Vehicle Symposium, 2003, pp.42-46.

[7] N. Furui, H. Miyakoshi, and M. Noda, "Development of a Scanning Laser Radar for ACC", in Proceedings of the SAE International Congress: Electrical/Electronics, 1998, pp.71-76.

[8] J. P. Costeria and T. Kanade, "A Multibody Factorization Method for Independently Moving Objects", International Journal of Computer Vision, Vol.29, No.3, 1998, pp.159-179.

[9] K. Kanatani, "Motion Segmentation by Subspace Separation: Model Selection and Reliability Evaluation", International Journal of Image and Graphics, Vol.2, No.2, 2002, pp.179-197.

[10] N. Ichimura, "Motion Segmentation Based on Factorization Method and Discriminant Criterion", in Proceedings of the IEEE International Conference on Computer Vision, 1999, pp.600-605.

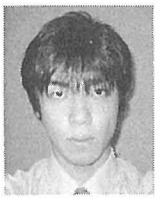
[11] C. W. Gear, "Multibody Grouping from Motion Images", International Journal of Computer Vision, Vol.29, No.2, 1998, pp.133–150.

[12] N. Ohta, "Structure from Motion with Confidence Measure and Its Application for Moving Object Detection", Transactions of the Institute of Electronics, Information and Communication Engineers, Vol.J76-D-II, No.8, 1993, pp.1562-1571, in Japanese.

[13] T. Kato, Y. Ninomiya, and I. Masaki, "An Obstacle Detection Method by Fusion of Radar and Motion Stereo", IEEE Transactions on Intelligent Transportation Systems, Vol.3, No.3, 2002, pp.182-188.

[14] Open Source Computer Vision Library (OpenCV) for Intel® architecture, Available: http://www.intel.com/research/mrl/research/opencv/

Shigeki Sugimoto received the B.Eng. and the M.Eng. degree in control system engineering from Tokyo Institute of Technology, Tokyo, Japan, in 1995 and 1997, respectively. He joined the Fu's Lab. Ltd. Tokyo, Japan in 1998. He is currently a research scientist for the promotion of industry-government-academic cooperation at the Tokyo Institute of Technology since 2003.



Hidekazu Takahashi received the B.Eng. degree from Tokyo Institute of Technology, Tokyo, Japan, in 2003. He is currently a graduate student at the Department of Mechanical and Control Engineering at the Tokyo Institute of Technology, Tokyo, Japan.

Masatoshi Okutomi received the B.Eng. degree in mathematical engineering and information physics from the University of Tokyo, Tokyo, Japan, in 1981 and the M.Eng. degree in control engineering from the Tokyo Institute of Technology, Tokyo, Japan, in 1983. He joined the Canon Research Center, Canon Inc., Tokyo, Japan, in 1983. From 1987 to 1990, he was a visiting research scientist with the School of Computer Science at Carnegie Mellon University, Pittsburgh, PA. He received the Ph.D. degree for his research on stereo vision from the Tokyo Institute of Technology, in 1993. In 1994, he joined the Tokyo Institute of Technology, where he is currently a Professor in the Graduate School of Science and Engineering. His current research interests include both theoretical and practical aspects of computer vision and image processing.

Received: 11 May 2004
Revised: 19 August 2004
Accepted: 3 September 2004
Editor: Shunsuke Kamijo