DYNAMIC SCENE ANALYSIS FOR A MOBILE ROBOT IN A MAN-MADE ENVIRONMENT

Saburo Tsuji, Yasushi Yagi and Minoru Asada

Department of Control Engineering
Osaka University
Toyonaka, Osaka 560, Japan

ABSTRACT

Analysis of scene viewed continuously from a robot moving in a man-made environment, such as a building or a plant, yields useful information for the navigation. The knowledge on the environment, richness of the scene in vertical edges and the flatness of the floor, is arranged in constraints for the dynamic scene analysis. The rotational component of camera motion is estimated first from image points invarient from translation. After compensating for movements by the rotation between the consecutive images, the foci of expansion of translational motion of both the robot and moving objects are determined.

1. INTRODUCTION

Mobile robots need much intelligence for performing their tasks in unstructured environments, and many research groups have studied autonomous vehicles as a challenge in Artificial Intelligence research[1-5]. These robots plan routes to the specified destination and navigate in the real indoor environment using visual, ultrasonic and proximity sensors. An important drawback of the vision systems is that the robot closes its eye while it is moving and views the scene only when it stands still. The robot thus wastes much time by iterating frequent moves and stops to obtain information for the navigation.

We, therefore, need a vision system which continuously views the scene and finds the pathway and obstacles there in real time. One important basis for such a system, theories of computing the camera's motion from the optical flow, has been already developed[6-7]. Examples of applying them to real scenes, however, are very few, because finding of the precise optical flow in the time-varying imagery of real scene is, in most cases, difficult and, as a result, the simultaneous estimation of as many as five motion parameters is unreliable. Also the analyses assume the environment is static, and we cannot apply them to a scene containing moving objects.

This work is supported in part by Grant-In-Aid for Scientific Research from the Ministry of Education, Science and Culture, Japanese Government.

To overcome the above-mentioned difficulties, this paper presents a new idea that utilizes the knowledge on the man-made environment as constraints on determining the robot's motion from the time varying imagery. The rotational component is estimated first from the points invarient from any translation such as vanishing points. Compensating for movements between consecutive images by the rotation, we can determine F.O.E.s (focus of expansion) of the translational components of both the robot and objects moving in the scene, which provide us with useful information for navigation.

2. ROBOT AND KNOWLEDGE ON ENVIRONMENT

2.1 HARDWARE SYSTEM

A Heathkit HERO ROBOT, an autonomous vehicle, on which a small CCD TV camera is installed, is used as the robot. An internal sensor, an encoded disk of the front wheel, provides an estimate of distance it moved. However, any good estimate of its yaw angle from outputs of internal sensors is not available, because the robot cannot move straight but along a circular arc even if the front wheel is commanded to be at its center position.

The input image is converted into a 256 by 256 8bit digital picture and analyzed by a microprocessor 6808 augmented by an image processor, a 68000 plus four LSI image processing chips (Hitachi ISP's)[8], which can compute correlation of a 256 by 256 digital image with a 4 by 4 mask at the video rate. An LMI Lambda Machine is used for higher-level processing such as manipulating of the world model and planning.

2.2 PRIOR KNOWLEDGE ON ENVIRONMENT

Since the robot moves in the pathway of a building, we can build the world model which provides the detailed structure and geometry[4,5]. The method proposed in this paper, however, uses properties common to many indoor environments rather than the geometrical world model. Therefore, we can apply it to autonomous vehicles working in most buildings or plants without any detailed knowledge on them. The following two properties are arranged in the constraints for image analysis.

(1) The scene contains many visible vertical edges.

Horizontal edges in scene, appeared as the boundary between the floor and the wall, are not used as primary features of the scene because of difficulties in finding them in image by conventional gradient operators.

(2) The floor is almost flat.

This means that the height of the camera is almost invarient and thus computable from the angles of joints to change the camera direction. Small unevenness of the floor, however, causes the camera to shake and results in a considerable amount of movements in image. Therefore, the conventional motion stereo method which assumes the invarience of camera direction is not applicable to our case.

3. IMAGE ANALYSIS

3.1 BASIC EQUATIONS

The following equations describing the relation between the motion of a camera in a static environment and the optical flow generated are well known[7].

Let the coordinate system X,Y,Z be fixed to the camera, with the Z-axis pointing along the optical axis (see Fig.1). We denote by $\mathbf{T}=(U,V,W)^T$ the translational component of the motion of the camera and by $\mathbf{w}=(A,B,C)^T$ its angular velocity. Let the instantaneous coordinates of a point P in the environment be (X,Y,Z). Thus we have the equations of the components of velocity as:

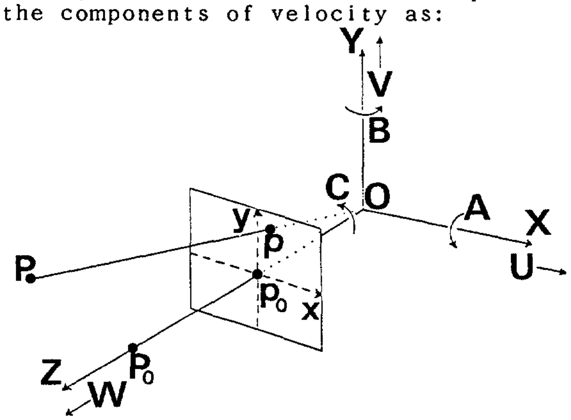


Fig.1 The camera coordinate system

$$X' = -U - BZ + CY$$
$$Y' = -V - CX + AZ$$
$$Z' = -W - AY + BX.$$

Assuming perspective projection between an object point P and the corresponding image point p, we have the equations of the optical flow, denoted by (u,v) at a point (x,y) as

$$u=(-(fU/Z)-fB+Cy)-x(-(W/Z)-(Ay/f)+(Bx/f))$$
$$v=(-(fV/Z)-Cx+fA)-y(-(W/Z)-(Ay/f)+(Bx/f)),$$

where f is the focal length of the camera.

If the accurate flow vectors at five image points are known, we can determine the five motion parameters, A, B, C,(U/W), and (V/W) of each point in space from the above equations. Since the measured optical flow is somewhat inaccurate, the least mean squares method is proposed so as to utilize as much of the available information as possible[7].

3.2 PROPOSED METHOD

Examples of applying the theory described in the previous section to a mobile robot are very few because of the following reasons. Finding of precise and dense optical flow in real dynamic scenes is very difficult and, as a result, the estimates of motion parameters are not reliable. The method needs much computation and seems inadequate for real time processing, the essential need for the mobile robot. Its another drawback is the assumption of static environment. In practical applications, the scene contains moving objects against the background, and we need to find them from a sequence of pictures taken from a robot moving in the scene. The least mean squares method averages over the whole image and is difficult to find F.O.E.s of optical flow generated by the moving objects.

We propose to simplify the problem by utilizing the domain-specific knowledge for the mobile robot in a man-made environment and information available from the internal sensors of robot. The difficulty in finding correspondence between frames is overcomed by another idea that samples and analyzes images so frequently that the movements between consecutive frames are very small. At present our robot is moving very slowly at a speed of 0.3 m/sec, and an image is taken and sampled in a second. Thus, establishing of correspondence of feature points is not difficult. We use feature points distributed sparsely in image rather than the dense velocity map of which computation for real scenes seems difficult.

In stead of solving the simultaneous nonlinear equations with as many as five variables, we decompose the problem into two stages. The rotational component **w** of the camera movement is determined first from points invarient from translation. Compensating for movements in image caused by **w**, we obtain images generated by the pure translation of the camera. The translational component (u_t, v_t) of the optical flow is

$$u_t = (-fU + xW)/Z \qquad v_t = (-fV + yW)/Z,$$

and the location (x_F, y_F) of F.O.E (focus of expansion) is determined as:

$$x_F = (fU/W) \qquad y_F = (fV/W)$$

Utilization of information on F.O.E. will be discussed in Section 3.4.

3.3 DETERMINING OF ROTATIONAL COMPONENTS

We assume that the robot moves almost straight and the coarse estimates of its position and camera angle are available from outputs of the internal sensors. We need, however, to monitor the image to find the deviation of motion increasing gradually from the direction desired.

As mentioned previously, the unevenness of the floor causes the camera to shake. Even a small amount of change in the camera direction results in significant movements in image. A standard lens (16mm) is used for a 3/4 inch image tube of the camera, and the equivalent focal length of the camera system is approximately 600 pixels. A change by 1 degree in the direction causes an image point to move by about 10 pixels. Continuous monitoring of the three components of camera rotation, therefore, is highly desirable to simplify the image analysis.

VANISHING POINT ANALYSIS

Examples of image points invarient from translation are vanishing points. If two vanishing points in consecutive frames are determined as projections on a Gaussian sphere[9], estimation of the rotational component is easy. Since the vertical edges in scene are strong and reliable features, we have exploited to utilize their vanishing point.

Let define the reference of camera orientation such that the Y axis is aligned with the vertical edges and the optical axis (Z axis) is pointed to the destination point of the current movement. Pitch and roll angles of the camera from the reference orientation are determined from the vanishing point of vertical edges found in image. (The rotation of camera between two frames is represented by a set of rotations, a roll about the Z axis, followed by a pitch about the X axis, and finally, a yaw about the Y axis.)

The tilt angle of camera, measured by an internal sensor, provides us with an approximate location of the vanishing point and, therefore, search of the input image for these lines is rather easy. We apply the 3 by 3 Sobel operator to the input images to find edge points and their directions. Since the tilt angle of the camera

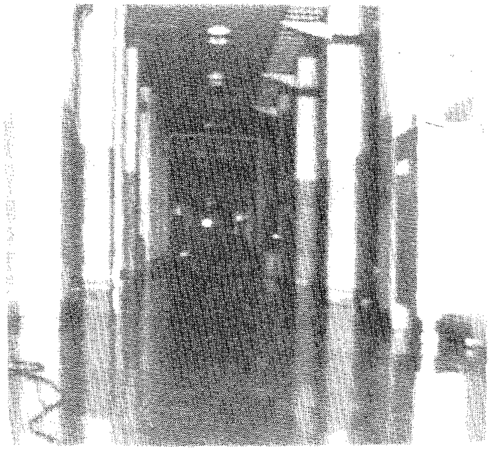


Fig.2 An example of input images.

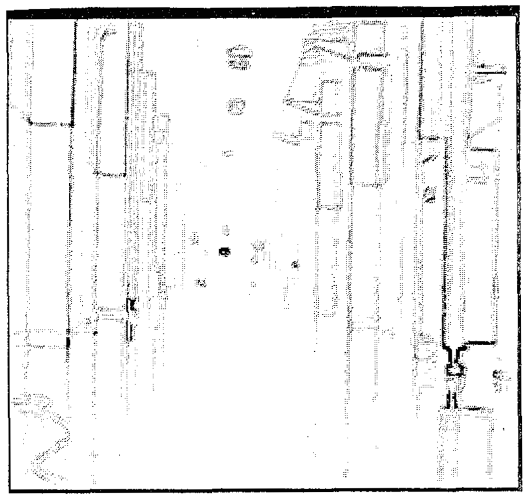


Fig.3 The edge picture of Fig.1.

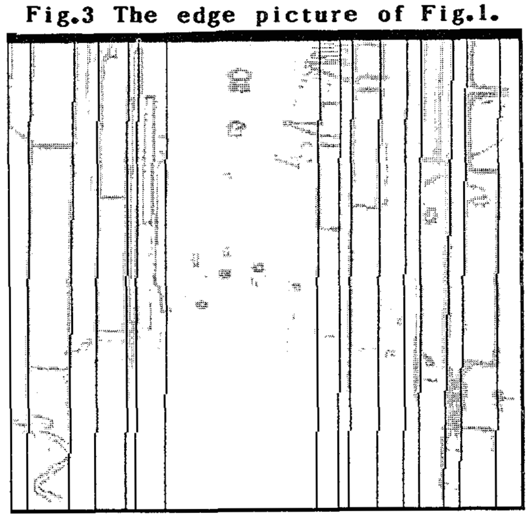


Fig.4 The detected vertical lines.

on our robot is kept to be zero while it navigates, the image lines corresponding to the vertical edges are almost vertical. To find the vertical lines, a histogram of the vertical edge points versus their horizontal positions in image is computed. Each peak in the histogram corresponds to a vertical line of which parameters are then estimated by a least mean squares method. Figs.2-4 show an example of input images, its edge picture and the detected vertical lines, respectively.

Since these lines are nearly parallel, an algorithm for finding vanishing points which accomodates sets of such lines[9] is adopted. For all combinations of the lines detected, we determine vectors that point from the origin O of the coordinates toward their intersections in image. Each of these vectors is mapped onto a Gaussian sphere of which center is located at O. Thus, a point on the sphere represents the orientation of a vector. Each cluster of the intersections projected on the Gaussian sphere corresponds to a vanishing point of lines. If the camera is aligned with the reference orientation, the vanishing point is mapped onto the vertex of the sphere. The pitch and roll angles are determined as deviations of the cluster from the vertex along great circles parallel and perpendicular to the image plane, respectively.

Fig.5 (a) shows an example of the intersections mapped on the pitch-roll plane. The intersections are distributed along a narrow region. This means the estimate of the pitch angle is not so reliable as the roll because of the rather long focal length of the camera used. (We could expect much better results for images taken by the camera with a wide angle lens.) Unreliable vertical lines are therefore discarded by examining whether each line contributes to generate many intersections far from the centroid of the cluster. This process results in a much smaller varience in locations of mapped points as shown in Fig.5 (b). Figs.6 (a) and (b) display the temporal changes of the pitch and the roll in 8 consecutive frames sampled each second while the robot moves at a speed of 0.3 m/sec.

Since most of indoor scenes contain many parallel horizontal edges, the above-

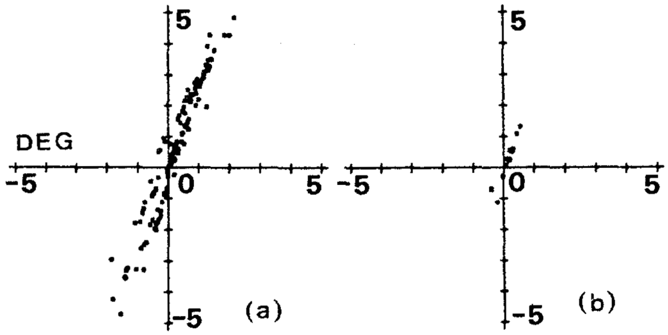


Fig.5 Intersections of vertical lines mapped onto the pitch-roll plane.

mentioned analysis seems promising to determine the yaw angle of camera. The method, however, is not always applicable to real scenes, especially to our robot's environment. The images lack reliable long lines corresponding to these edges and, therefore, it is difficult to precisely locate the vanishing point.

If we can find very distant objects in image, their displacements in consecutive frames give the yaw and pitch angles of camera. The invarience in size of each pattern around the image center in the input image sequence is examined by a simple pattern matching method to know the pattern is of an object far from the camera. If the far object is detected, the yaw angle is easily determined as (u_0/f), where u_0 is the horizontal displacement of the distant object in image. Fig.6 (c) shows the temporal change of the yaw angle estimated by this method.

When the system does not detect the distant objects in images, two variables U/W and B are left unknown. (Since the floor is flat, V/W is almost zero) A method for analyzing simultaneously these two variables will be discussed in the next section.

3.4 ANALYSIS OF TRANSLATIONAL COMPONENT DETERMINING F.O.E. OF OPTICAL FLOW GENERATED BY TRANSLATION

If all components of the rotational component **W** are determined, we can compensate for movements of image points by **W** to obtain the optical flow generated by the pure translation. The location of F.O.E. is estimated from the cluster of intersections of the optical flow.

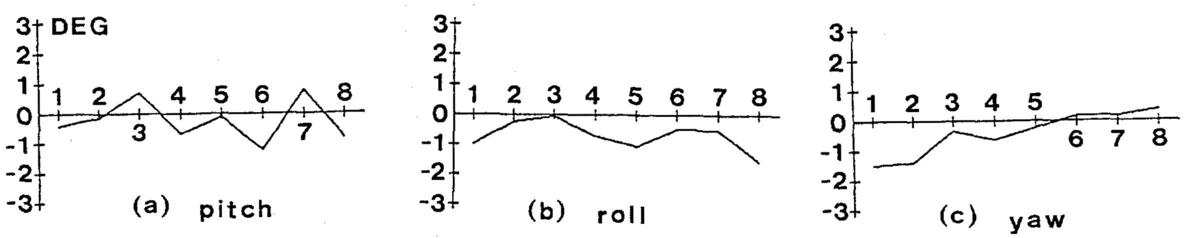


Fig.6 Temporal changes of the rotational components.

Feature points in image are extracted in the first frame by applying the Moravec's interest operator[3] to the neighbourhoods of the vertical lines detected already. The corresponding points in the consecutive frames to these feature points is established by the S.S.D.A. Fig.7 displays an example of the results of this process, the selected feature points and their optical flow between the first and eighth frames.

Compensation for the movement by the rotational component is then made, and we get Fig.8 (b), the estimate of optical flow caused by the pure translation, from Fig.8 (a), the optical flow obtained from the image sequence. Finally, the F.O.E is determined as the intersections of the translational components.

The F.O.E. gives useful information for navigation. Since the approximate velocity of the robot is known from the internal sensors, we can estimate distance to each static object in the scene from the robot. The error in the estimate in the example illustrated in Fig.8 is about 10 percent if the objects are not far from the robot. Computation of the time and distance of the object when it comes closest to the robot is also possible, thus we can avoid collision.

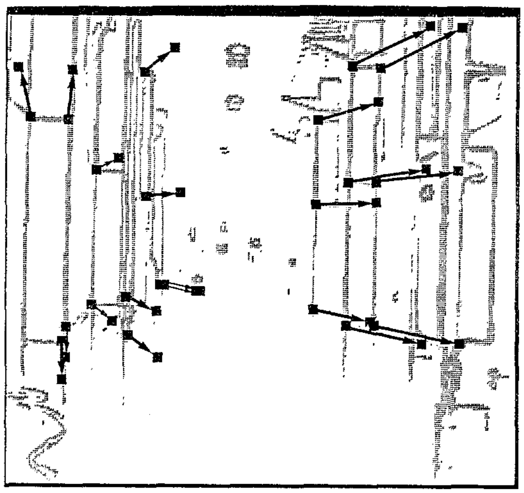


Fig.7 Disparity map of feature points.

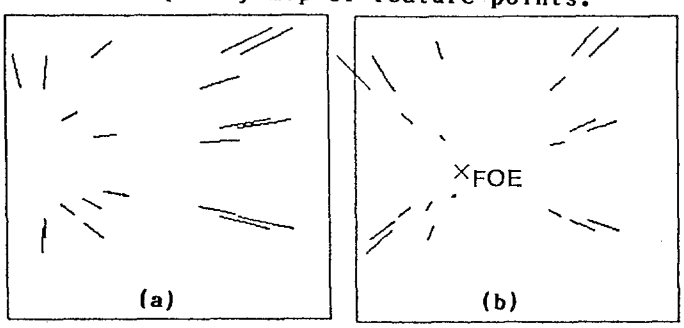


Fig.8 (a) Optical flow from image sequence.
(b) Optical flow by pure translation.

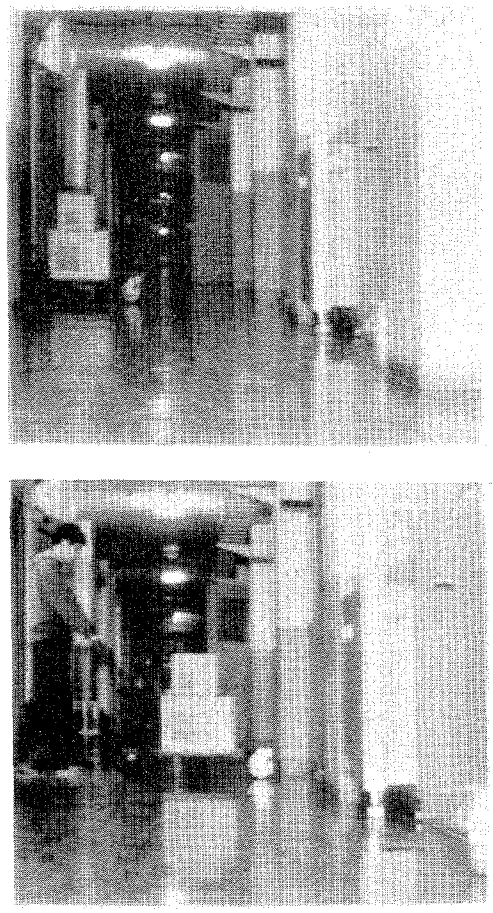
Objects moving against the static background could be found as containing feature points whose optical flow clusters around F.O.E.s different from that of the background, if their motion vectors in space have the directions different from the robot's motion. Separation of objects moving along the similar direction as that of the robot from the background is not possible, unless we do not use heuristics such that the objects are on the floor.

DETERMINING OF YAW AND TRANSLATION

If any feature point of a distant object is not found in image, simultaneous estimation of both the yaw angle and the translational component is needed. If we assume the images are taken with a long focus lens, say f=600 pixels for our robot, then

$$u=-(fU/Z)+(Wx/Z)-fB$$
$$v=(Wy/Z)$$

are approximately valid. The optical flow of feature points of objects at Z, generated by the yaw and the translation, therefore, forms a cluster around



(a)

(b)

Fig.9 Input images of a scene containing a moving object,
(a) the first frame, (b) the eighth frame.

((fU/W)+fB(Z/W),0) in image. If we assume a yaw angle, we can compensate for movements by the yaw and obtain the clusters of the optical flow. By changing the yaw, the actual rotational component is detected as of obtaining the sharpest clusters. The cost of computation is not expensive since the possible range of the yaw angle is known a priori. This method has the advantage over the least mean squares method that it could accept the scene containing moving objects against the background. Figs.9 show the first and eighth frames of

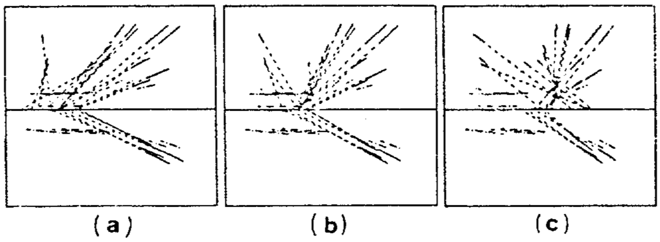


Fig.10 The estimation process of the yaw angle.
The estimate of yaw angle is changed by 0.5 degree in each step. (a),(b) and (c) are examples showing clusters of the compensated optical flow. (b)is selected as the best estimate.

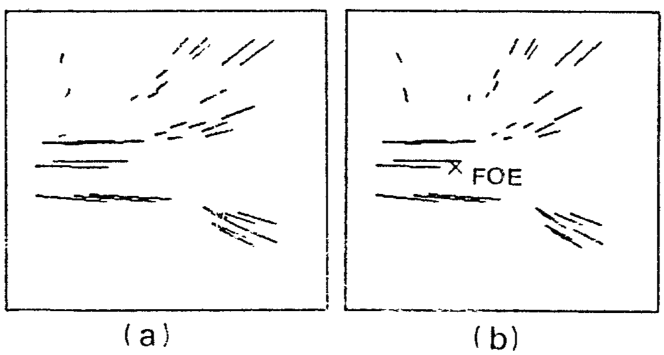


Fig.11 (a) Optical flow in the input image sequence.
(b) Optical flow by pure translation.

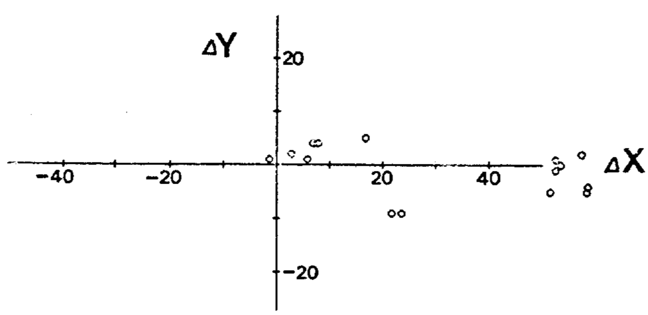


Fig.12 The flow vectors which do not cluster around the F.O.E. of the background.

an input scene in which an object is moving from left to right. Figs.10 show the process of finding the sharpest cluster by changing the yaw angle. Adopting the yaw for which the majority (the background) of optical flow forms the sharpest cluster, we obtain the translational components of the optical flow (Fig.11 (b)) from the flow vector field (Fig.11(a)). The flow vectors which are considered as of moving objects or unreliable ones are further analyzed to find moving objects and their velocity. Fig.12 shows most of these vectors are almost identical and of a same moving object.

4. CONCLUSIONS

This paper presents methods for analyzing the image sequence taken by a TV camera mounted on a moving robot. Since the direct application of the basic theories for passive navigation do not provide us with reliable estimates of motion parameters, we utilize a priori knowledge on the man-made environment to simplify the analysis. The motion parameters obtained are not very precise but enough for giving information for the passage in the building.

REFERENCES

[1]Nillson, N.G., A mobile automaton, **Proc. 1st Int. Joint Conf. Artificial Intell.**, pp.509-520, 1969.
[2]Giralt, G., Sobek, R. and Chatila, A multi-level planning and navigation system for a mobile robot, **Proc. 6th Int. Joint Conf. Artificial Intell.**, Tokyo, pp.335-337, 1979.
[3]Moravec, H.P., The Stanford Cart and The CMU Rover, **Proc. IEEE**, Vol.71, no.7, pp.872-884, 1983.
[4]Tsuji, S., Monitoring of a building environment by a mobile robot, **Proc. 2nd Int. Symp. Robotics Research**, Kyoto, 1984.
[5]Yachida, M., Ichinose, T. and Tsuji, S., Model-guided monitoring of a building environment by a mobile robot, **Proc. 8th Int. Joint Conf. Artificial Intell.**, Munich, pp.1125-1127, 1983.
[6]Prazdney, K., Determining instantaneous direction of motion from optical flow generated by a curvilinearly moving observer, **Computer Graphics and Image Processing**, Vol.17, pp.238-24, 1981.
[7]Bruss, A. R. and Horn B.K.P., Passive navigation, **Computer Vision, Graphics and Image Processing**, Vol.21, no.1, pp.3-20, 1983.
[8]Fukushima, T. et al., An image signal processor, **IEEE Intn'l Solid-States Circuit Conf. Digest of Tech. Papers**, pp.258-259, 1983.
[9]Magee, M.J. and Aggarwal, J.K., Determining vanishing points from perspective images, **Computer Vision, Graphics, and Image Processing**, Vol.26, pp.256-267, 1984.