Three-dimensional image sensor system for autonomous mobile robots

Kanehiro Sorimachi

Advanced Robotics, Vol. 4, No. 3, pp. 203–215 (1990)

Review

Three-dimensional image sensor system for autonomous mobile robots

KANEHIRO SORIMACHI

Production Engineering Research Laboratory, Canon Inc., 30-2 Shimomaruko 3-chome, Ohta-ku, Tokyo 146, Japan

Received for *JRSJ* 30 June 1987; English version received 15 August 1988

1. INTRODUCTION

The list of applications for robots would be dramatically increased if a robot capable of moving autonomously, avoiding obstacles and finding its own way could be developed.

One way of exploiting the capability of autonomous movement in a robot would be to give it the vision of a human being or an animal. However, human vision is so complicated that it could not be implemented in robots. Even if a comprehensive understanding could be gained of the domain of visual information processing performed by the nervous system built around the human brain, the advent of a vision system that is practical enough to be installed in robots would still be doubtful.

Even the smallest insects, such as dragonflies and butterflies, avoid obstacles as they fly and land on plants and trees. As these insects act by processing visual information in their nervous systems of limited capacity, knowledge of the visual information they acquire would lend itself to processing by present-day microcomputers.

This seems to suggest that, even though the image of the world as seen by a dragonfly is unknown to us, a vision system that could at least offer visual information needed for moving a robot could be developed.

2. VISUAL INFORMATION NEEDED FOR ROBOT MOVEMENT

A human being perceives two types of visual information: image information and spatial information. Image information is the kind of information that can be recorded in the form of plane images, and includes intensity, colour, and two-dimensional geometry. Spatial information relates to stereoscopic spaces, such as the direction, distance, and movement of visible objects. Image information is of essential importance in distinguishing and recognizing visible objects. We use it for identifying persons or reading characters. In particular, much of the image information reflected in the central pit of the retina is transmitted to the association areas in our brains for object recognition processing.

Spatial information, with its dominant role in controlling human behaviour, is normally accepted with little consciousness. For example, actions such as grasping a cup, jumping over a puddle, or throwing a ball to someone cannot be accomplished unless information on the distance to the objects has been accepted, but usually we are seldom aware of these distances. Probably, one can control movement without being conscious of spatial information because:

(1) recognition processing takes such a long time that it cannot cope with abrupt actions; and
(2) since, in terms of the theory of evolution, movement control is essential to sustaining life, recognition processing was implemented in primitive brains by short-circuit reaction long before the development of association areas in our brains.

Although the image of the world visible to a dragonfly is unknown to us, let us explore normally less-heeded spatial information with reference to case studies to extract the visual information needed for robot movement.

It is possible for a human being to walk along a starlit street at night. The visual information the human being acquires at this time is the presence or absence of objects and their proximity, rather than their colours or geometric details.

Also, if we are reading a book while walking, we perceive images of the scenes in the direction of travel in the peripheral area of our retina by holding the book down. We derive the spatial information needed to achieve movement from the images in the peripheral visual field with less resolution, with our attention being focused mainly on the characters in the book.

When a person is about to collide with another person at the corner of a passage, he or she will stop suddenly or otherwise try to avoid the collision. It is only after he or she has taken such action that he or she recognizes the other person. This behaviour stems from recognition of certain things. The behavioural pattern may be due to learning, but, unquestionably, the spatial information obtained from vision plays a dominant role in triggering the action. On the basis of case studies, nine categories of specifications emerge for a three-dimensional (3-D) image sensor system for autonomous moving robots:

(1) *3-D spatial information.* Since spatial information is more important than image information for movement control, the ability to recognize objects is of no concern.
(2) *Resolution.* Direction and range resolutions may not be critically high.
(3) *Acquisition time.* The system must be capable of perceiving repetitive information in a short time and at short intervals to cope with moving objects.
(4) *Illumination conditions.* No special lighting conditions are required.
(5) *Distances.* Information on objects located at medium distances of 1–15 m is important.
(6) *Non-interference.* Where a single system controls many robots, the system must not exert any adverse effects on the robots which would induce mutual malfunctions.
(7) *Safety.* The system must be free from harmful effects on human beings, animals, or plants located in the same environment.
(8) *Compact and lightweight.* The system, including the information processor, must be compact and lightweight enough to mount on a robot.
(9) *Low power consumption.* The system should consume as little power as possible, so that the robot equipped with it can move without cords.

3. 3-D IMAGE SENSOR SYSTEM

As stated earlier, spatial information is more important than image information for movement control. As a technique of acquiring spatial information, a 3-D image sensor system is currently being studied which perceives the environment as a distribution of the distance to a particular object visible within the visual field. Distance distribution information is easier to comprehend in the form of range images. Given range images, in which distance data are converted into bright images for long distances and dim images for short distances, obstacles can be avoided by guiding the robot in the direction of brighter images. The 3-D image sensor has often been approached not only as a receptor for spatial information, but also as a sensor that recognizes the stereoscopic geometries of objects. Various techniques for achieving the contactless measurement of object distances, including camera auto-focusing, have been studied to date.

Figure 1 shows the major categories of contactless range measurement techniques available. Problems involved in the implementation of these techniques in autonomous moving robots are reviewed below, against the nine requirements listed earlier.

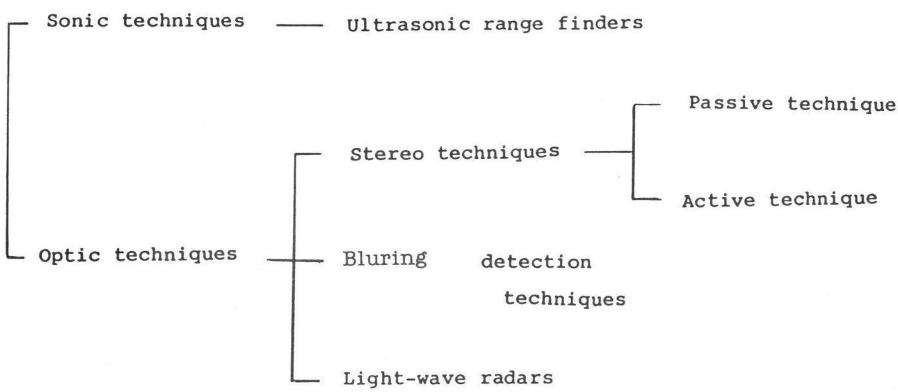

Figure 1. Contactless range measurement techniques.

3.1. *Ultrasonic range finders*

Ultrasonic range finders determine the distance to a remote object by measuring the time it takes for the ultrasonic waves that they radiate to travel to the object and return. These finders, however, are not suitable for measuring distances greater than 3 m, mainly because their lateral resolution is reduced by the difficulty in narrowing ultrasonic beams and because reflected sounds from sources other than the regularly opposed surfaces are hard to obtain as the energy is attenuated in air and sounds are regularly reflected from the many mirror objects involved. Yet ultrasonic range finders are a useful tool in detecting obstacles at short distances, and their use in movement control applications is presently being exploited.

3.2. *Passive stereo technique*

The passive stereo technique detects the parallax between two images of an object pictured with the separation of a baseline distance, like the binocular stereoscopic vision of a human's eyes, to determine the distance to the object on a conceptual extension of triangulation. The technique, though very simple in principle, has the major drawback that it is difficult to locate the imaging positions of the same object point projected in the two images, or the corresponding points. Another weakness is that the distance to a plane with uniform intensity cannot be measured.

3.3. *Active stereo technique*

The active stereo technique determines the distance to an object of interest by projecting optical beams to the object and detecting the position or size projected on the object in images formed from a baseline distance. The technique not only avoids the problem of detectability of corresponding points inherent in the passive stereo technique, but can also measure distances to planes with uniform intensity. When the beam pattern and the direction of projection are to be altered, however, both the input time and the amount of information to be processed increase to input multiple images. Although the active stereo technique can be used in dark places, it involves the projection of a high-energy beam to obtain information on objects at medium distances in bright places.

3.4. *Blurring detection technique*

The blurring detection technique measures the distance to an object from the distance between the lens and the image sensor in effect when the object is in focus with minimum blurring of its image, as the lens is displaced in the direction of the optical axis. Although, theoretically, the differential of an image signal is maximum at the focal point of the lens, better maximum value detection accuracy calls for increases in the focal length and the aperture of the lens because of its depth-of-focus implications. Also, because mechanical motion is involved, research emphasis has now shifted from the blurring detection technique to alternative approaches.

3.5. *Light-wave radars*

Light-wave radars use optical beams, in place of the radio waves used by regular radars, to determine the distance to an object by measuring the time it takes for the beams to travel to the object and return. Since light can be projected in fine beams, the measurement area can be segmented with the benefit of increased direction resolution. However, light transmission is so fast that detection with high time resolution is imperative to increase distance resolution; this requires sophisticated electronics. Also, safety considerations impose constraints on the use of laser beams in specific environments. An experimental system has been developed at the Carnegie-Mellon University as part of its continuing commitment to the research of autonomously moving vehicles. Called NAVLAB, the system moves autonomously on the basis of environmental information obtained from a light-wave radar 3-D image sensor system [1].

These considerations have made me realize the need to develop robots or carriers that can move autonomously in a general environment in which human beings co-exist, and, in view of automatically manoeuvring vehicles of the future, to develop further a 3-D image sensor system by passive approaches. Currently, two image sensor systems based on the passive stereo technique and one system based on the active stereo technique, which can be used as a movement aid or manipulator hand controller, are currently under study. General descriptions of these systems follow.

4. PASSIVE RANGE PATTERN SENSOR (PRPS)

The passive range pattern sensor (PRPS) divides the visual field into 300–1000 directions and acquires data on the distance to the object in each direction by the passive stereo technique. These data are considered the minimum set of information required for a robot to be able to move autonomously.

One of the auto-focusing techniques is that used for cameras and it involves the use of a 128-element CCD line sensor to measure the distance in one direction in a short time, without mechanical motion.

A CCD image sensor is viewed as an assembly of many such line sensors. The left and right images obtained by the stereo technique are split into mosaic form, then projected in an alternate arrangement on the CCD image sensor to form a number of range finders. The optics layout of the PRPS is shown in Fig. 2. Two object lenses with equal focal lengths are placed to have parallel optical axes with the separation of a baseline distance so as to image scenes in the same plane with equal magnifications. After the images are separated through positive and negative mosaic masks on the imaging plane, the images of the scenes formed in the transparent section are synthesized by the reflectors and beam splitter and then re-imaged on the CCD image sensor through the relay lens. Conditions of the images are shown in Fig. 3. As a result, the sensor system is organized:

(1) to project images through the left- and right-eye object lenses on the CCD sensor in an alternate arrangement;
(2) to form the two images of each object point through the left- and right-eye object lenses along a single scanning line on the CCD sensor; and
(3) to project object points imaged on the image sensor through the left-eye lens unless occlusions prevent the object points from being viewed from the other lens position.

The reference visual field is longer in the scanning line direction than the base visual field. Among the multiple scanning lines included in the width of each visual field mask imaged on the image sensor, the image signal only on one scanning line around the centre is used to determine the parallax by image correlation. Assuming that the number of CCD pixels on one scanning line within the base visual field is n, the image illuminance of each pixel is L_{Si}, the number of CCD pixels on the corresponding scanning line in the reference visual field is m, and the image illuminance of each pixel in the scanning direction is L_{Ri}, the value determined by the equation

$$U_j = \sum_{i=1}^{n} |L_{R(i+j)} - L_{Si}|$$

is calculated $m - n + 1$ times from $j = 0$ to $m - n$ by shifting one pixel at a time, so that the position at which the value of U_j is minimum is regarded as the corresponding position. Subsequently, a dedicated processor, with two line memories and two arithmetic processing circuits, processes distances in 300 directions 30 times a second with a delay of 1 ms each, converting the distance data into intensities for bright long distances and dim short distances for display on the CRT.

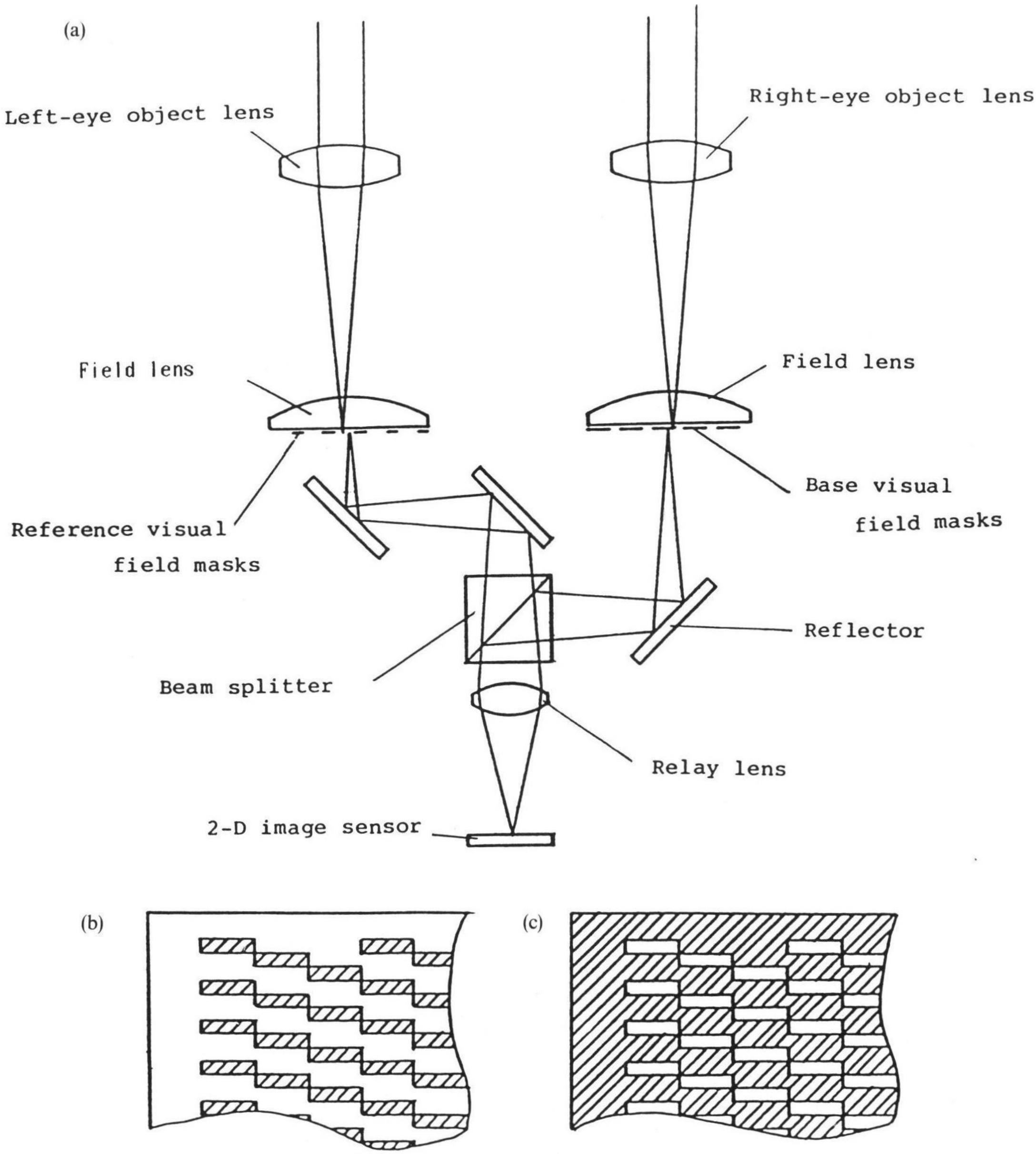

Figure 2. Passive range pattern sensor system optics layout. (a) Optic layout. (b) Reference visual field masks. (c) Base visual field masks.

Figure 4 is a scene composed of panels of stripe patterns. Figure 5 is a range pattern image of this scene. Experiments with this system have yielded the following findings:

(1) It seems that obstacles can be avoided by dividing the visual field into 300–1000 directions, depending on the width of the visual field and the size of the moving robot.

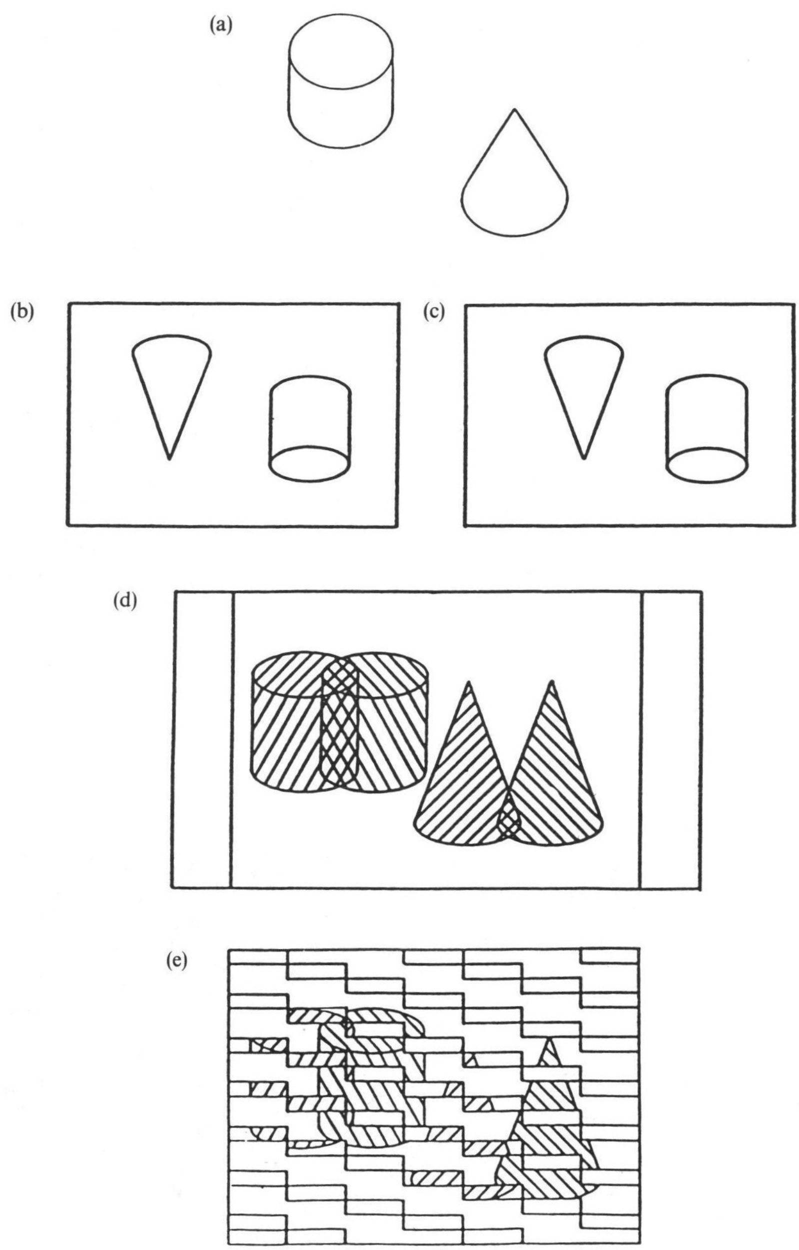

Figure 3. Images formed on the passive range pattern sensor. (a) Scene. (b) Primary imaging through the left-eye object lens. (c) Primary imaging through the right-eye object lens. (d) Unmasked images on the image sensor. (e) Masked images on the image sensor.

(2) Ambiguities of the distance pattern data as they are varied by movement can be alleviated by manipulating the data in a statistical fashion. (At present, the data are processed by the brain of the person watching the CRT.)
(3) Distance pattern data are less prone to the effects of changes in the illumination of scenes, such that image data may be held in a knowledge base to confirm positions

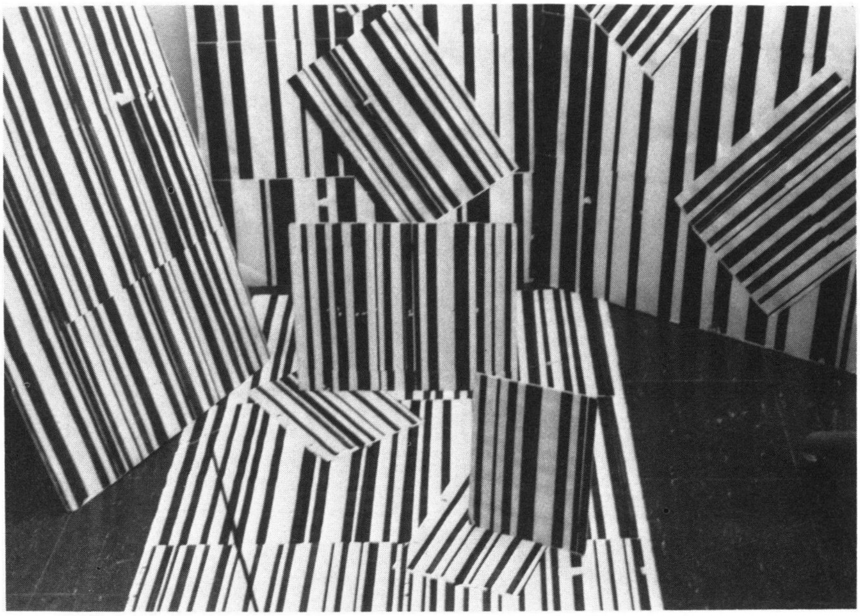

Figure 4. Scene composed of panels of stripe patterns.

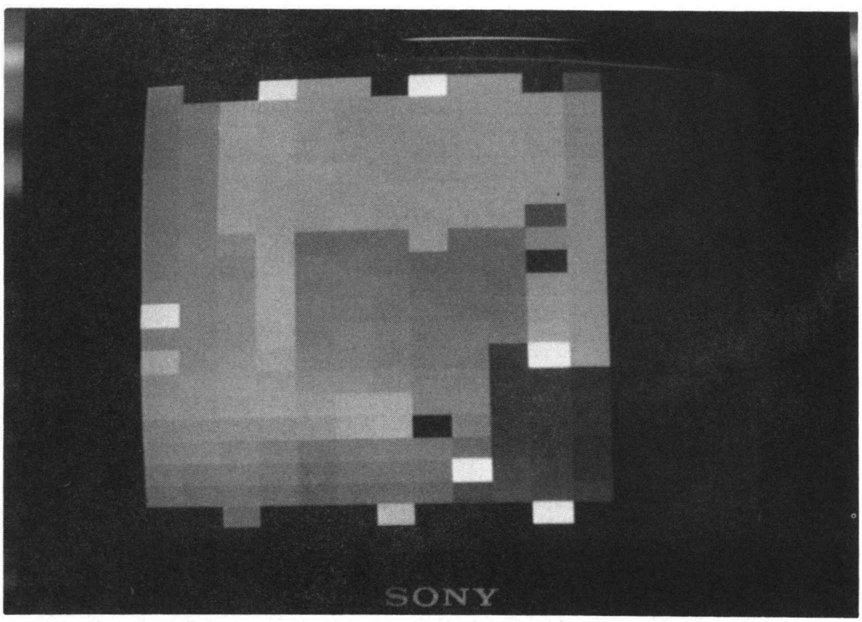

Figure 5. Range pattern image of the center of the scene shown in Fig. 4. Long distances are displayed bright, short distances dim.

by pattern matching, as 8-bit-per-direction records on as many as 1000 directions amount only to 1 kB per screen load.

(4) The distance to planes with uniform intensity cannot be measured without surrounding interpolation. Particularly, road surfaces often resist distance measurement, calling for interpolation and knowledge information processing.

(5) As images at close distances have an increased magnification, the objects falling within the base visual field often have a uniform intensity. Hence, the combined use of the active stereo technique may be considered.

(6) Since even auto-focus cameras could fail to measure distances to objects coming within their range, it will be necessary to design a dedicated optics system.

Many factors for discussion still remain, but I am convinced of the technical practicability of a PRPS for use in the image sensor system for an autonomous moving robot.

5. ACTIVE RANGE PATTERN SENSOR (ARPS)

The active stereo technique is suited to measuring the distances to objects nearby that have increased image magnifications and a good chance of having uniform-intensity planes, because it requires less light projection during range measurement. Since short-distance information on 1 m or shorter distances is often needed in the implementation of movement control and this information is also useful in controlling a manipulator hand, I have examined an image sensor system called the active range pattern sensor (ARPS).

The idea of the ARPS is that distance measurement is made possible by radiating a light flux at each object point within the base visual field in the PRPS described previously and thus producing artificial contrast in the images in the base visual field.

Two lenses with equal focal lengths are placed to have parallel optical axes with the separation of a baseline distance. One lens has a mask drilled with many small windows mounted on its focal plane and is illuminated backwards to radiate many beams of spotlight against the object of interest. The distance to each spotlighted point is measured by picking up the object in the image sensor installed on the focal plane of the other lens and locating the spot image. The layout of the optics system is shown in Fig. 6.

Observation of the images in the spotlight received by the image sensor on the monitor screen indicates that the spot images are displaced in the direction of lens formation with changes in the object distance, but the sizes of the images and their positions at right-angles to the direction of lens formation do not change. This fact can be easily understood, as projecting the light flux, shown in Fig. 6, in the direction of lens formation causes the rays of light through the left and right lenses to overlap at the same position because of the equal perspectives of the two lenses. Position detection can be facilitated by aligning the direction of the scanning lines of the image sensor with the direction of lens formation.

The spot images projected on the object vary in blurring size with changes in the object distance. The ARPS has a halogen lamp filament placed at one focal position of the elliptical reflector, with the entrance pupil of the projection lens at the other focal position. In this set-up, the direct light from a small-area light source is suppressed, with only the light reflected upon the reflector entering the lens through the windows

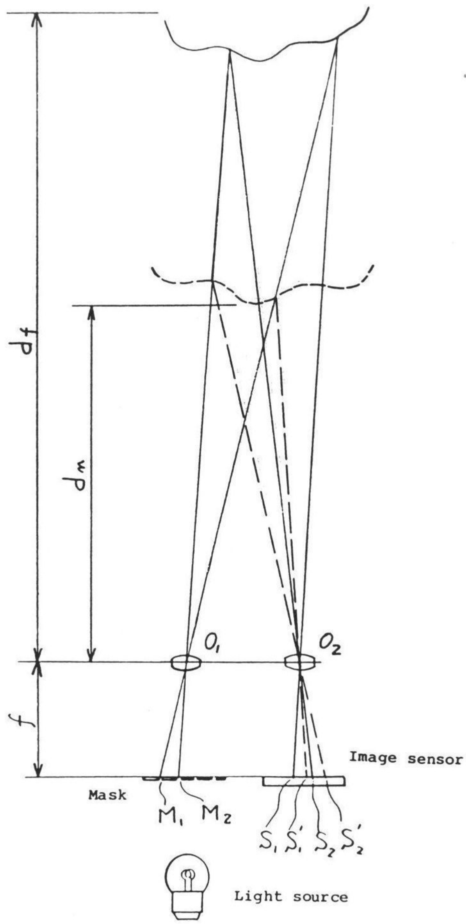

Figure 6. Active range pattern sensor optics layout.

on the mask, thereby minimizing changes in the blurring size with changes in the object distance and loss of the quantity of light in the peripheral area.

Knowledge of the correspondence between the windows on the mask and the positions of the spot images is essential in measuring the distance to the spotlighted point on the object. To this end, the ARPS has the windows on the mask (Fig. 7) spaced wider than the distance of travel of the spot images on the image sensor. Though this limits the number of windows placed in the direction of lens formation to more or less than 10, the window positions have a one-to-one correspondence with the spot images, permitting range measurement in 500–1000 directions with a single image input. The ARPS also permits distances to be essentially processed on a real-time basis, making itself an aid to movement or manipulator hand controller.

6. PASSIVE RANGE IMAGE SENSOR (PRIS)

The previous section introduced the notion that robot movement can be controlled with information on object distances in up to 1000 directions. Implementation of this

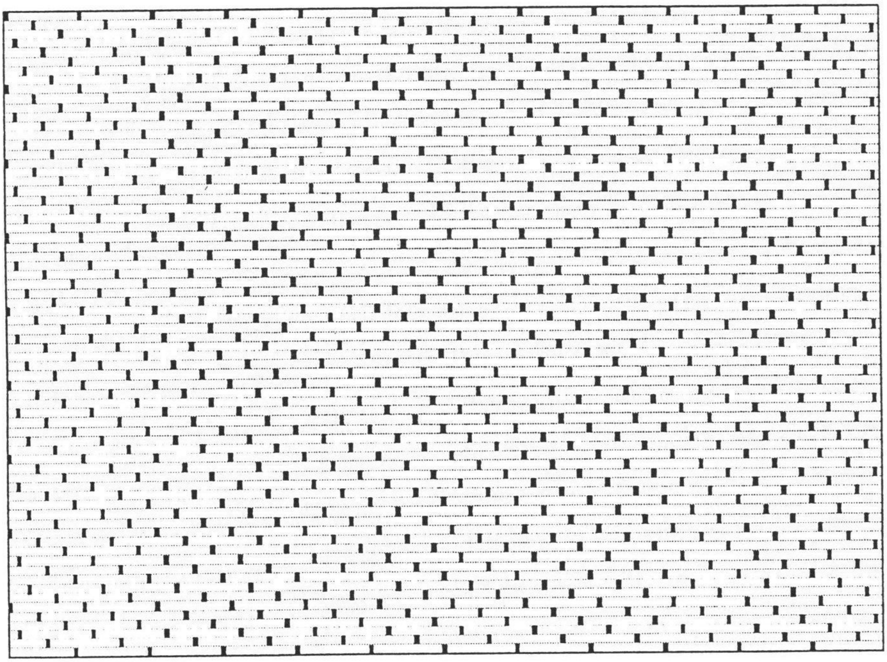

Figure 7. Active range pattern sensor mask.

notion requires range images with a higher resolution to complement object recognition information. One of the functions of the human eyes is convergence. I have developed a technique, called plane convergence, that interlocks lens shifting and focusing in photography with each other to simulate the effect of convergence. The plane convergence technique suggested the practicability of acquiring range images of the edges of scenes in 2–3 s with an equivalent of the pixel density of the image sensor by detecting the corresponding points in the passive stereo technique [2]. Although the vision system operating on the principles of the plane convergence technique, called the passive range image sensor (PRIS), has only gone through the simulation stage as yet, it is briefly described below.

The human eyes exercise convergence control as they focus on an object point of interest and intersect their viewing axes at the gaze point at the same time. With a particularly high resolution in the central pit in the retina, only the object of interest can be viewed clearly, as images other than those at the gaze point are blurred and displaced in the lateral direction in the retina, even when objects at different distances are seen in the same direction.

The plane convergence technique is a planar extension of the effect of convergence. As shown in Fig. 8, object lenses O_1 and O_2 are set up to have parallel optical axes. Object point P is then brought into focus by translating one lens, O_2, to $O_{2'}$ and changing the distance between the lenses and the image sensors from f to f' at the same time. When the relation $fb = f'b'$ (where b is the fixed distance between the two image sensors, b' is the distance between the lenses, and f is the focal length of the lenses) holds, all the object points visible in the visual field in the plane perpendicular to the optical axis through P are in focus and their relative image positions on the left

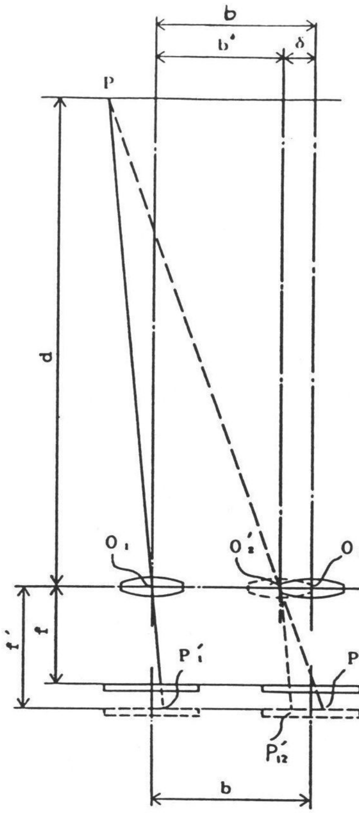

Figure 8. Passive range image sensor optics layout.

and right image sensors match. This plane is called the convergence plane. When the values of f' and b' are varied while maintaining the relation $fb = f'b'$, the convergence plane moves back and forth in the visual space, allowing one-dimensional scanning. Concurrently with this scanning, the left and right image signals are output synchronized, then differentiated to obtain differentials. Analysis of the timing changes in the differentials reveals characteristic signal changes in the pixels of matched images in focus. Since the distance d between the lenses and the convergence plane can be determined by calculating $d = fb/\delta$, assuming that $b - b'$ is δ, a single distance image can be obtained by scanning the visual field in the convergence plane by one stroke while allocating distance d calculated from the value of δ where the left and right images in focus are matched.

The range resolution depends on the number of frames or fields from the image sensors that are generated during each stroke of scanning. The rate measurement accuracy depends on the measurement accuracy of δ.

7. CONCLUSION

The characteristics required of a vision sensor system and the method of acquiring spatial information, which is needed for a robot to move autonomously, with reference to the information man obtains when walking, for example, along a starlit street, have been discussed. A brief description of a vision system under study to meet these requirements has also been given.

Considering the fact that even the smallest animals, with nervous systems of limited capacity, control their movement by processing necessary survival information, the visual system can be developed at a more rational level of refinement, allowing robots to move autonomously, if the information required to meet specific purposes can be skilfully identified.

REFERENCES

1. M. Hebrt and T. Kanade, "3-D vision for an autonomous vehicle," *IEEE, 87 THO* 166-9, pp. 375–380, 1987.
2. K. Sorimachi, T. Endo and S. Yamada, "A new range image sensor system," *IEEE, 87 THO* 166-9, pp. 182–185, 1987.

ABOUT THE AUTHORS

Kanehiro Sorimachi (M' 83) joined Canon Inc. on graduating from Chiba University in 1985. He has been engaged in mechanical design of optical machines. He is a Member of the SICE, JSME and JSPE.