

Projected Texture Stereo Vision System in Sunlight

Chiaki AOYAMA*

ABSTRACT

A binocular stereo vision camera system able to measure relative distance to a non-textured floor in real time without contact even under direct sunlight has been developed. A camera that alternately captures images with and without a projected texture at high speed, integrates the respective images within the pixels, and outputs the difference between them outside the pixels was developed for the system. The combination of this camera and a texture projector using a Class 1 laser that is safe for the eyes for pulse light emission has made it possible to measure the distance to a subject with a non-textured surface even under sunlight of 100000 lx if its reflectivity is at least 5%, achieving a measurement accuracy of 1 mm in a range between 0.9 m and 1.5 m.

1. Introduction

To help enable a bipedal walking robot to walk freely, it is necessary to measure the relative distance to the floor in real time.

Honda therefore developed a floor detection sensor for walking indoors⁽¹⁾ (Fig. 1) and applied the unit in its ASIMO robot. The floor detection sensor uses the light-section method, which is one method of realizing stereo vision. In this method, a line of light is projected, and distance is measured from its position on an image by means of triangulation⁽²⁾ (Fig. 2). The use of a projector makes it possible to measure distance even on a non-textured surface. In the normal light-section method, the distance to a flat surface is measured by scanning a single line of light. However, if a single beam is scanned while the robot is walking, the distance to the flat surface changes with the position of the robot during the scan, and thus cannot be measured. Multiple lines of light are therefore projected at the same time and the simultaneity of each line is maintained, making it possible to measure the distance to the flat surface while the robot is walking.

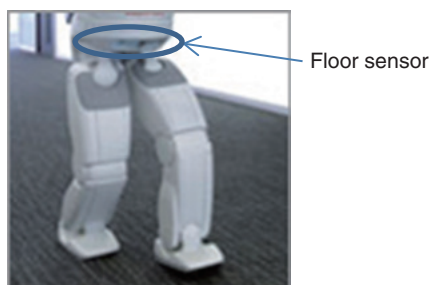


Fig. 1 Floor sensor of ASIMO

To help enable the lines to be detected using simple image processing, everything other than the lines of light is imaged as black, as shown in Fig. 3. A laser emitting narrow-band light is used as the light source, and the camera is fitted with a band pass filter that allows only narrow-band light to pass. Using near-infrared as the wavelength of the laser makes it possible to obtain an image of only the lines of light under fluorescent light, which emits almost no near-infrared light.

However, in sunlight, near-infrared light is extremely strong, and the texture of the floor can be seen, as shown in Fig. 4. If the texture of the floor is linear, the system is unable to distinguish the lines of light and the texture, and it therefore cannot be used close to a window that is admitting sunlight.

Increasing the intensity of the light source reduces the relative effects of environmental light, but this produces safety issues. Safety conditions for the eyes are stringent for light sources such as lasers that concentrate light on a single point when a lens is employed, and increasing their intensity therefore represents a challenge. In the case of light sources other than lasers, it is challenging to project a clear line

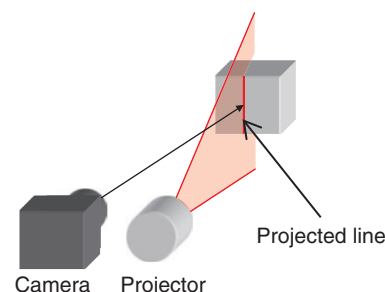


Fig. 2 Light-section method

* Fundamental Technology Research Center

texture regardless of distance, and the light-section method cannot be used.

The binocular stereo vision used in the head section of ASIMO, which does not use projected light, is safe for the eyes, but is unable to measure distance to non-textured surfaces. This is because when triangulation is used to determine distance, clear image characteristics are necessary for the determination of correlations between the images. Because a non-textured or virtually non-textured floor without clear characteristics is not a special case, it is not possible to use binocular stereo vision as a distance sensor in relation to a floor.

The issue of determining distance even when there is no texture can be resolved, if a distance of several meters is available, by employing a texture projector to apply a texture to the surface of an object⁽³⁾ (Fig. 5). However, in systems using conventional cameras, the projected textures become unclear under sunlight, and it is not possible to measure distance. (An example will be discussed below.) If it was possible to obtain images even in sunlight in which a clear texture was projected and that could be used in correlation processing for stereo vision, it would be possible to measure distance even to non-textured surfaces.

Given this, the following targets were established for system development:

- (1) Ability to measure distance to an irregular, non-textured surface within a range from 0.9 to 1.5 m with sufficient accuracy for control purposes
- (2) Ability to be used in sunlight of 100000 lx
- (3) Use of a light source that is safe for the eyes (conforming to Class 1)

This paper will discuss the principle of the system, its configuration, and the measurement results obtained.

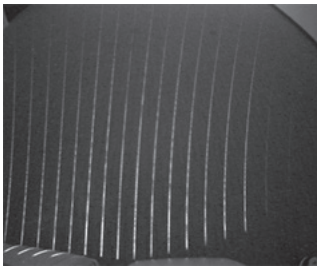


Fig. 3 Good image for floor sensor

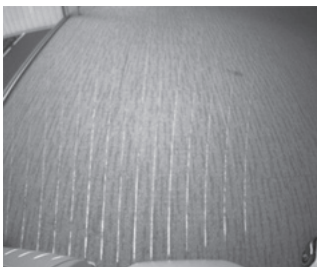


Fig. 4 Bad image for floor sensor

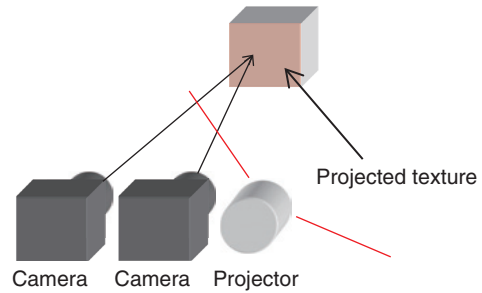


Fig. 5 Projector stereo camera system

2. Principle of System

The research discussed in this paper developed a projected texture binocular stereo vision system using two cameras and one or two texture projectors. The project also developed a camera and projector that would make it possible to see a texture under sunlight while maintaining safety for the eyes.

This section will discuss the principles underpinning the realization of increased resistance to environmental light in each of the elements of the system.

2.1. Difference Camera

In order to mitigate the effect of environmental light, the system compares the difference between images in which a texture has been projected and images in which a texture has not been projected, and does away with environmental light by canceling the shared environmental light in the images (Fig. 6). If, hypothetically, two images featuring only an environmental light component, with no projection, were taken consecutively and the differences between the images compared, there would be a difference of zero if no movement had occurred. In other words, a completely black image would be obtained.

If movement had occurred, because the position of the subject would have changed due to the time difference between the images, when the difference between the images was compared, it could not rigorously be judged as zero. In the case of images taken while a robot is moving, if the images are taken at intervals of 10 μ s or less, at the robot's walking pace the error due to misalignment of the images would be equivalent to or less than noise even at the boundaries between black and white sections, and could thus be ignored. However, in order to obtain images of sufficient brightness at an exposure of only 10 μ s, a light source able to instantaneously output an extremely bright light is necessary. It is challenging to install a light source of this type in a robot.

In order to obtain images of sufficient brightness while realizing a time difference approaching zero, the research therefore adopted a method in which a number of short-exposure images are integrated. Using this method, if the integration period is 5 ms, the error due to movement will be equivalent to or less than noise and can be ignored if

the time difference is 50 μs or less. The project was able to obtain a projector with a light emission interval of 100 μs and a light emission period of several μs , making it possible to obtain images of sufficient brightness to be integrated 50 times in a 5 ms period.

However, it is challenging to read and integrate images in 100 μs or less each time a short-exposure image is taken. This is equivalent to reading 100000 images per second, and such high-speed reading is only possible using a large, high-speed camera.

The research therefore focused on realizing the ability to integrate images within the pixels at high speed and without noise, conducting charge reading and integration operations within the pixels and outputting differences outside the pixels.

Figure 7 shows a simple model of the structure of a pixel, and Fig. 8 shows the flow of integration of the electrons within a pixel. In the center of the simplified pixel there is a photoelectric conversion section (a photodiode), and to the left and right there are charge accumulation sections (capacitors). Transfer gates control the transfer of electrons between the photoelectric conversion section and the charge accumulation sections.

Figure 8 shows the potential of the path between A and A' in the structure of the pixel. The higher in the figure, the higher the potential. Electrons travel in the direction of lower potential, that is, downwards. If there are areas of high potential on the transfer path, the electrons accumulate near these areas without crossing them.

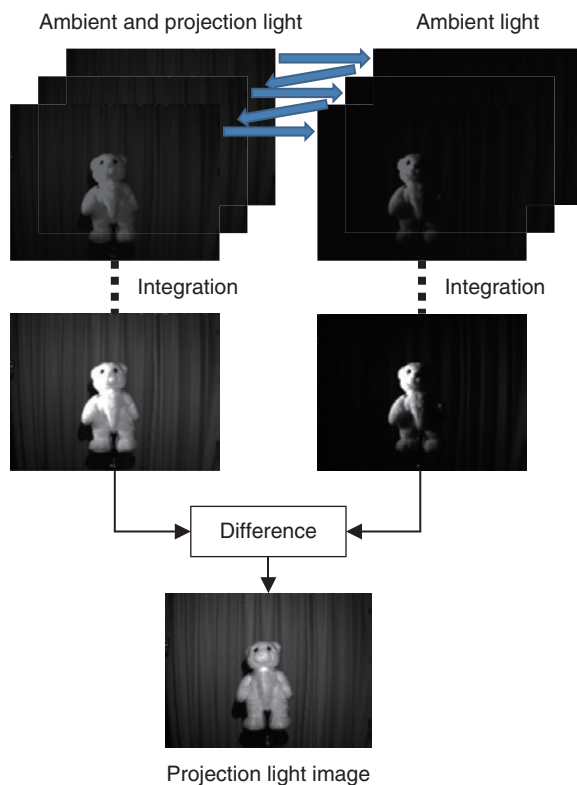


Fig. 6 Integration difference method

The series of integration operations commences with processing that discharges the electrons from the charge accumulation section, making the charge zero. Following this, the left and right transfer gates are opened alternately, and, depending on their state, the electrons produced in the photoelectric conversion section are transferred alternately to the left and right, and are integrated in the charge accumulation sections. After the integration processing has been repeated a number of times, a pixel signal is obtained by reading the level of the accumulated electrons. The integration of the electrons by alternation between the transfer gates can be repeated 100000 times or more per second, but the integrated pixel signals for the entire screen can only be read a maximum of 50 times.

Correction is applied to the signals that have been read to help ensure that inputs of the same light intensity are

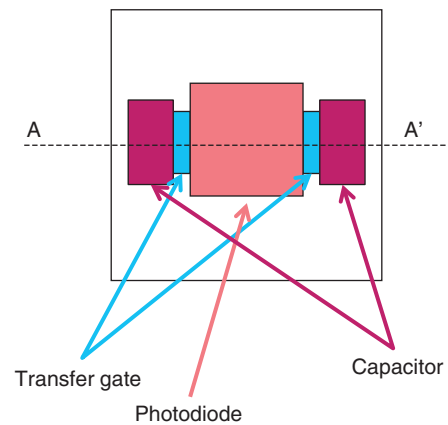


Fig. 7 Simple pixel model

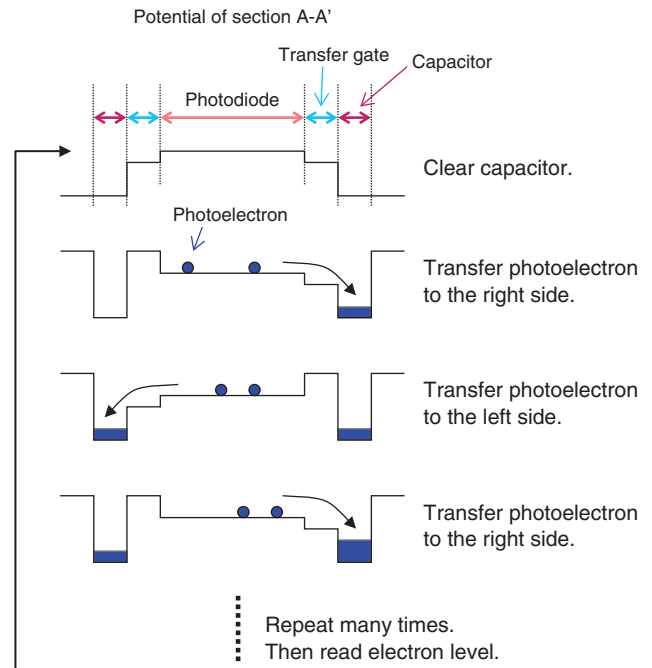


Fig. 8 Integration sequence

output at the same intensity. Following this, the difference between the images is identified. It is necessary to apply correction to the pixel signals, because if the signals were left uncorrected, differences in sensitivity and offset would result in differences in output for each pixel even if light of the same intensity had been incident on the entire screen. The application of correction means that output would be zero if there were no changes in the environment in which the images were taken.

It was possible to identify differences but not to apply correction within the pixels, and differences were therefore not identified within the image sensor.

2.2. Projector

Given that a light source able to emit narrow-band light at high power in a short period was necessary for the projector, a laser was selected for use. A diffuser was positioned in front to produce a broad light source despite the use of a laser, and to help ensure that light would not be concentrated on a single point even when concentrated by a lens. The use of the diffuser increases safety by a factor of several tens by helping to ensure that light is not concentrated on the retina.

The moiré method was employed as the texture projection method (Fig. 9). The use of a parallel grid pattern separated at

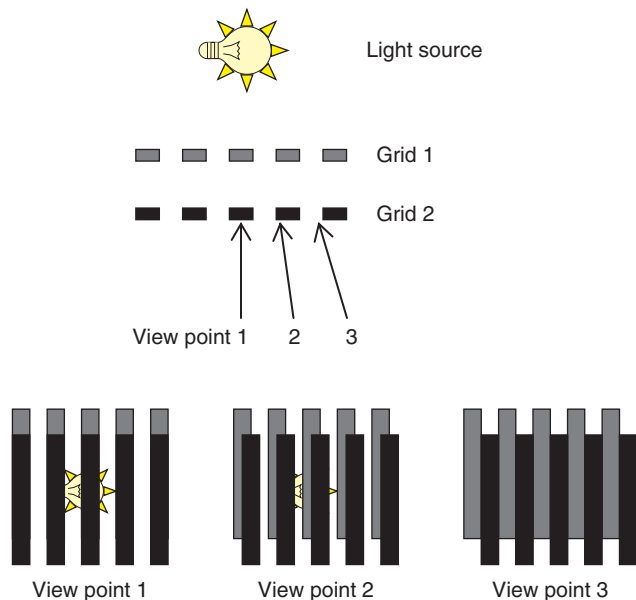


Fig. 9 Moiré method



Fig. 10 Moiré image

regular intervals means that the light source is either visible or obscured depending on the position, resulting in the projection of a striped texture. When seen from position 1, the grids appear to overlap and the light source is seen through slit-like gaps. Changing the position of viewing reduces the amount of the light source that can be seen, and from position 3 the slits obscure the view, and the light source cannot be seen. If a subject for imaging was present, it would be imaged as bright at position 1 and as dark at position 3. This means that when a white plate is positioned at the viewing positions, a vertical striped texture is projected (Fig. 10).

3. Difference Camera System

This section will provide an overview of the elements composing the difference camera system and the basic processing employed.

3.1. System Configuration

To create the system, one or two projectors were fitted in two difference cameras (Fig. 11). The image sensors had 640×480 effective pixels, with each pixel being $12 \mu\text{m}$ square. The baseline length was 170 mm, and the focal length of the lenses was 8.5 mm.

The difference cameras consisted of two camera heads, a camera control unit, and a camera interface unit (Fig. 12). The image signals output from the image sensors were sent to a PC via the camera control unit and camera interface unit. The projector emitted light from a laser diode (LD) in response to instructions from the camera control unit.

After the integrated images were output from the cameras and linear correction had been applied to them by the PC, the differences between the images were identified and distance information was obtained by stereo vision processing.

3.2. Pixel Circuits

The circuits within a pixel are divided into a photoelectric conversion section, transfer and drain gates, a charge accumulation section, and reading circuits (Fig. 13). The photoelectric conversion section converts light into electrons. The transfer gates connect to the charge accumulation section, where the electrons produced are accumulated (integrated).

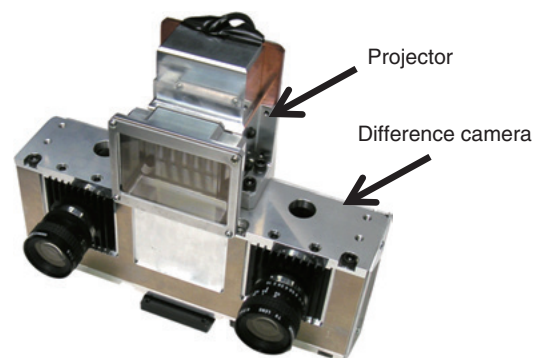


Fig. 11 Difference camera system

The drain gate is connected to the power source, where the electrons produced are extinguished. The electrons accumulated in the charge accumulation section are converted to magnitudes of current by the reading circuit and read out to the exterior of the pixel. The sensors developed for this project have four charge accumulation sections, and two of these are used when the sensor is operating as a difference camera.

The signal read out to the exterior of the pixel is converted from analogue to digital in the sensor, and output to the exterior as a 14-bit digital signal.

3.3. Pixel Configuration

Each pixel is a square with $12\ \mu\text{m}$ sides. The photoelectric conversion section is positioned at the focal point of a micro lens. The use of a pinned photodiode structure for the photoelectric conversion section reduces noise production. The electrons produced in the photoelectric conversion section are sent to any of the four charge accumulation sections via the transfer gates or to a drain via the drain gate.

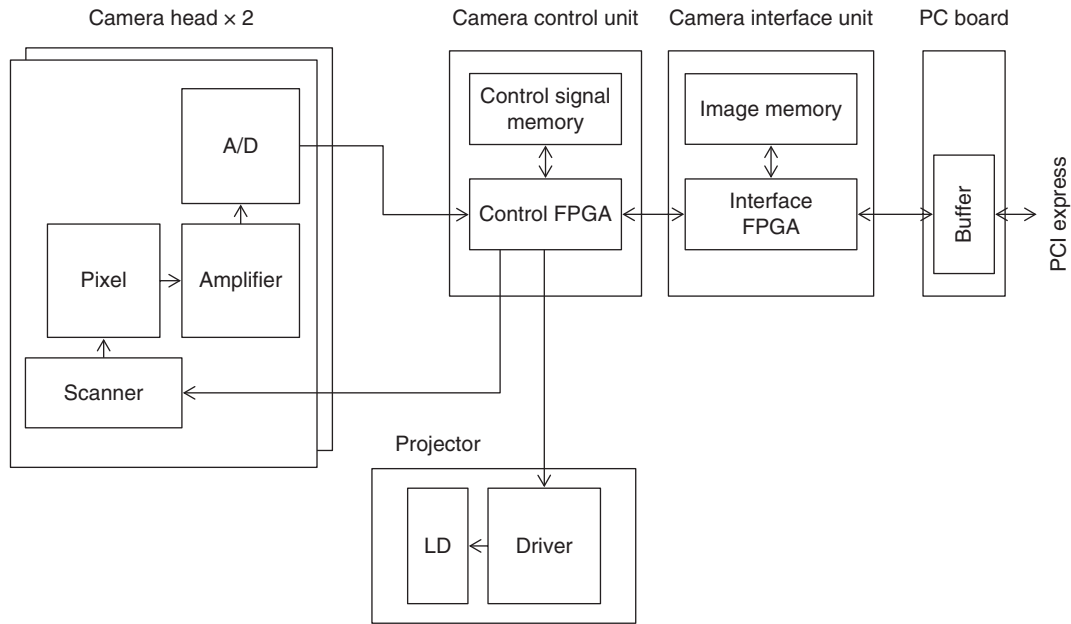


Fig. 12 Difference camera block diagram

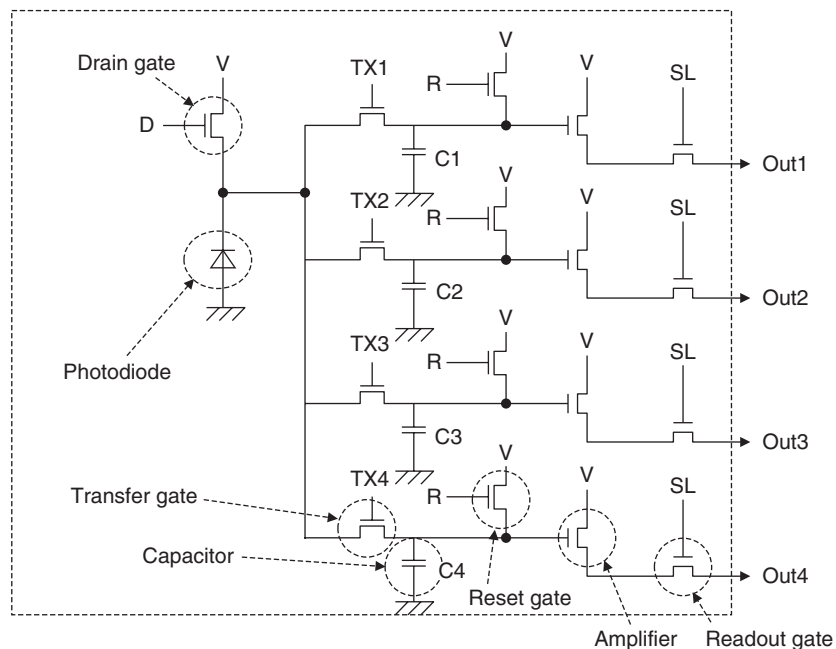


Fig. 13 Pixel circuit

3.4. Projector

In the projector, a diffuser is positioned in front of a semiconductor laser, and two grids are positioned beyond the diffuser. A semiconductor laser with a peak output of 300 W, a wavelength of 800 nm, and a minimum light emission interval of 100 μ s is used as the light source for the projector. The grids used to project a striped texture are vertical stripes with line widths of 0.25 mm and intervals of 0.25 mm transferred onto glass plates. There is a 20 mm gap between the grid plates.

For binocular stereo vision, two types of projector (projecting vertical stripes and diagonal stripes) are combined to help prevent mistaken correlation. Using vertical stripes only, identical textures appear repeatedly in the search range when a left-right correlation is attempted, and a single distance cannot be determined. The addition of diagonal stripes restricts the texture to a single example, making it possible to determine the distance. When the distances are roughly known in advance, the search range can be narrowed and vertical stripes alone are sufficient for the determination of distance.

3.5. Shutter Control

One exposure period is 100 μ s, and there are 50 exposure periods (Fig. 14). The time from commencement to completion of exposure (the exposure period) is 5 ms. The transfer gates are opened for a period that allows for the transfer of a sufficient number of the electrons produced by the received light pulse, and for the remaining time the electrons are discarded via the drain gate.

3.6. Camera Image Signal Processing

To help ensure that input of the same light intensity is output at the same light intensity, each charge accumulation section is corrected via broken-line approximation. By means of this correction, the variation in output values between pixels

becomes equivalent to or less than the shot noise generated by photon number fluctuation, which cannot be removed.

Differences are identified for linearized signals. Negative results are sometimes obtained when differences are identified, and an offset is applied to results for difference to help prevent this.

3.7. Stereo Vision Processing

Following distortion correction, conversion to match epipolar lines, and brightness correction, correlation processing is applied. The sum of absolute difference (SAD) is employed in order to identify correlations. The domain for identification of correlations is 13 horizontal pixels by 11 vertical pixels.

Brightness correction normalizes the luminance of the central pixels based on the average luminance of the domain. By making the intensity of domains with a low reflectivity (dark domains) and domains with a high reflectivity (bright domains) identical, this correction makes it possible to judge the reliability of SAD correlation processing results by means of a fixed threshold value.

Domains that are too dark are not corrected because this would increase noise to too great an extent.

4. Distance Accuracy of Camera System

The distance accuracy of the binocular stereo vision camera system was measured.

The subjects of the measurements were black flat surfaces and white flat surfaces. The reflectivity of the black flat surfaces was 5%, and the reflectivity of the white flat surfaces was 83%. A halogen light was used as ambient light. The light was adjusted to irradiate the subjects at a power corresponding to 100000 lx at close to 800 nm. The light intensity for the projection was 70%, the maximum

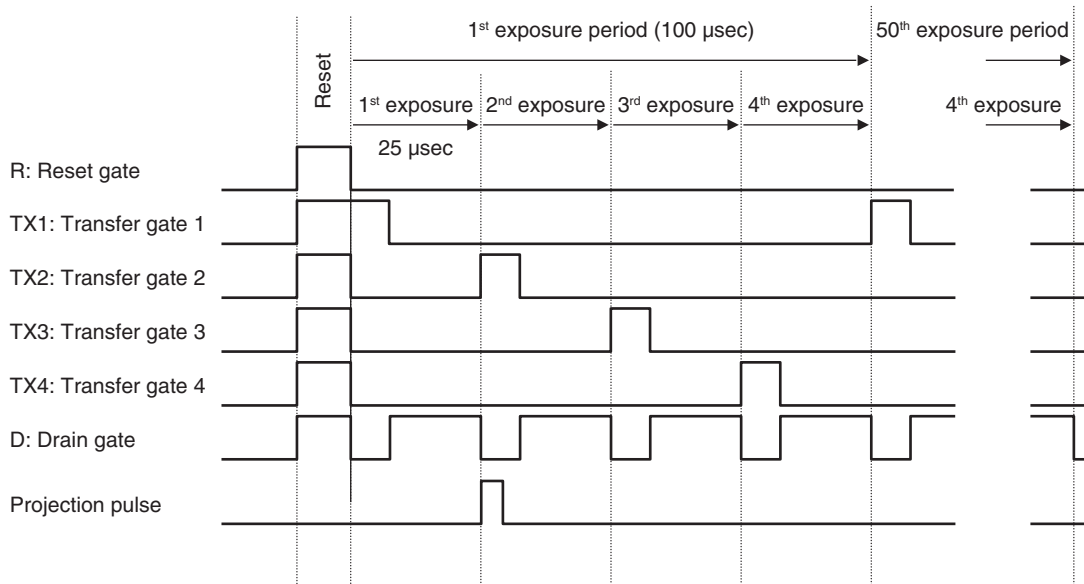


Fig. 14 Shutter timing

intensity allowed for a Class 1 laser.

The distance between the plane and the camera was varied in 2 cm increments using a linear stage. Among the distance data detected by the system, the average value within a domain extending 20 cm in all directions in actual space close to the center of the plane was taken to be the distance to the plane.

Figure 15 shows the images used in measurements and images taken using an ordinary camera with an exposure timing that continued to expose the image after the period of light emission. When the difference camera was used, identical images were captured irrespective of the presence

or absence of ambient light. When the normal camera was used, either the image was not captured due to saturation by ambient light, or, when the sensitivity was reduced to help prevent saturation, the texture became unclear. Because integration processing was conducted for the normal camera for longer than the period of emission of light by the light source, when the brightness of the image was made the same using only projected light, if environmental light was present, the image became brighter. This is to say, the image was significantly affected by environmental light.

Figure 16 shows the relationship between distance and error in measurement results. Error is expressed as

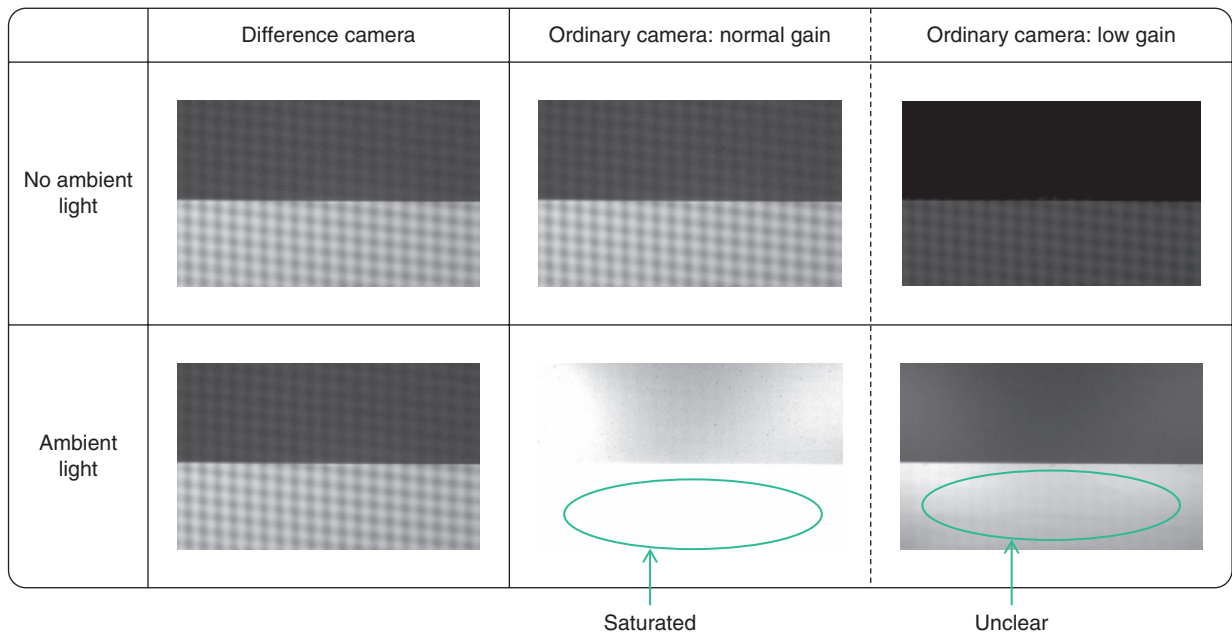


Fig. 15 Difference image and ordinary image

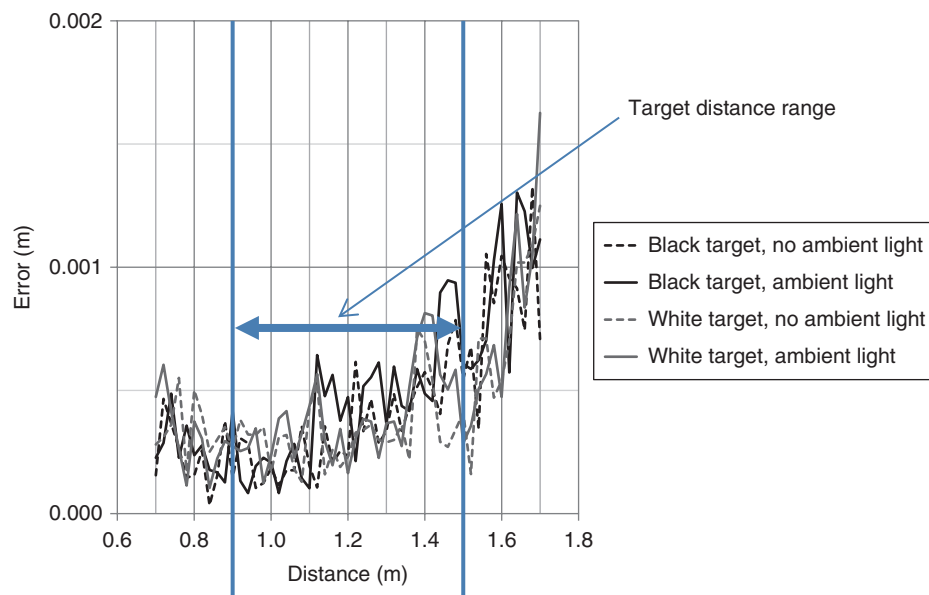


Fig. 16 Measurement error

3σ (three times σ , the standard deviation of distance data measured ten times), added to the absolute value of the difference between the average distance and the actual value of distance. Results for the black flat surface with no environmental light are shown as the black broken line, and results with environmental light as the black solid line; results for the white flat surface with no environmental light are shown as the gray broken line, and results with environmental light are shown as the gray solid line.

The graph indicates that in the range between 0.9 m and 1.5 m, error is 1 mm or less irrespective of reflectivity or the presence of environmental light. This represents sufficient accuracy for walking control.

5. Example of Application of Difference Camera

Because the difference camera makes it possible to remove ambient light and obtain images using only projected light, one application for the camera other than as a distance sensor is in enhancing the resistance of recognition functions to environmental light. The camera would make it possible to obtain performance of the type offered by appropriate image processing under indoor illumination even outdoors.

The grids were removed from the projector to do away with the texture, and images were taken outdoors with the subject in sunlight from different angles. The facial recognition processing used in ASIMO was applied to these images, and facial recognition results for images taken with a normal camera and the difference camera images were compared (Fig. 17).

The square green frame shows the domain for facial recognition results. The green or red bars at the left of the images show the degree of match with the pre-recorded subject (that is, the “score”). The longer the bar, the better the match. If the bar is green, the face is judged as being the same person; if the bar is red, the face is judged as being a different person.

The results show that the difference images are not affected by sunlight, and the facial recognition results are similarly unaffected. The ghosting visible in the upper section of the image taken under back-lighting by the normal camera has also been removed in the difference image.

6. Conclusion

A difference camera system able to measure distance to a non-textured surface under sunlight has been developed.

The system achieves a distance accuracy of 1 mm in the range from 0.9 m to 1.5 m, sufficient for use in robot walking control, irrespective of the reflectivity of the subject or the presence of environmental light.

When used with non-textured projected light, the system also enhances the accuracy of facial recognition.

Acknowledgements

The author wishes to express his sincere gratitude to Professor Shoji Kawahito of Shizuoka University for his guidance in the development of the image sensor discussed in this paper.

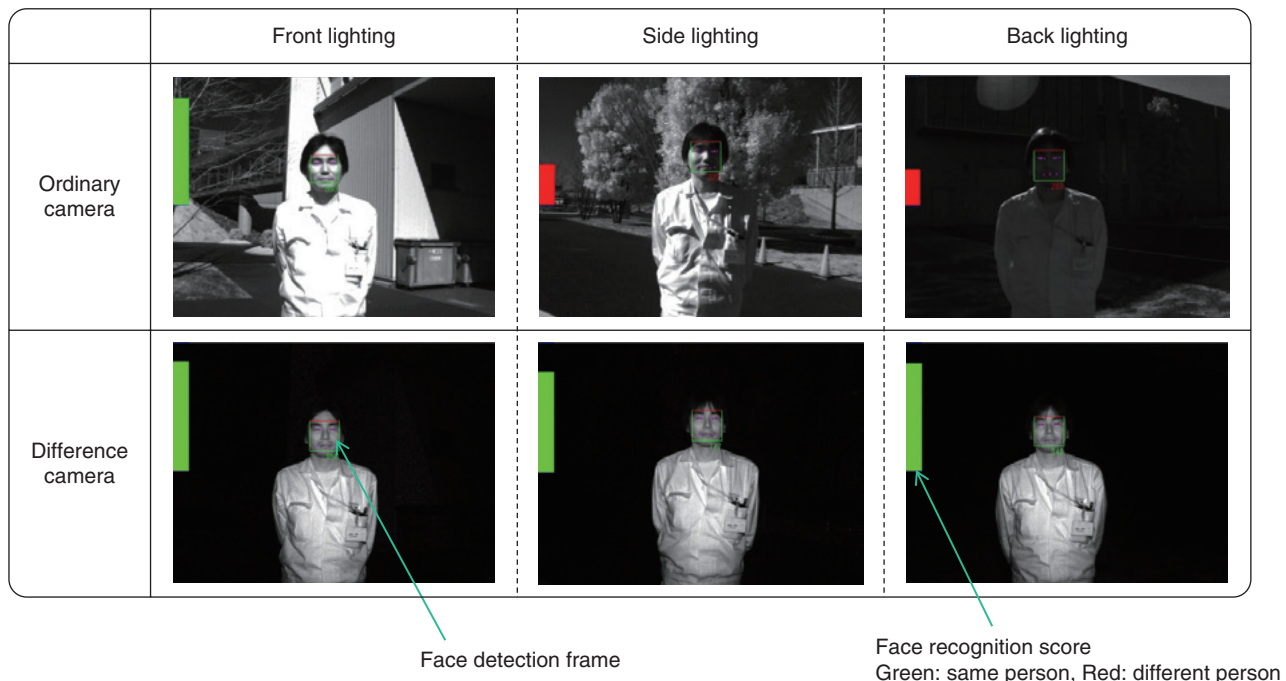


Fig. 17 Face recognition results

References

- (1) Shigemi, S., Kawaguchi, Y., Yoshiike, T., Kawabe, K., Ogawa, N.: Development of New ASIMO, Honda R&D Technical Review, Vol. 18, No. 1, p. 38-44
- (2) Oike, Y., Ikeda, M., Asada, K.: Design and Implementation of Real-Time 3-D Image Sensor With 640 x 480 Pixel Resolution, IEEE Journal of Solid-State Circuits, Vol. 39, No. 4, p. 622-628 (2004)
- (3) Konolige, K.: Projected Texture Stereo, Robotics and Automation (ICRA), p. 148-155 (2010)

■ Author ■



Chiaki AOYAMA