

Gesture Recognition Algorithm for Understanding Man-Machine Interaction and Its Adaptation for New ASIMO

Nobuo HIGAKI*

Taro YOKOYAMA*

Kikuo FUJIMURA**

ABSTRACT

An image-based gesture recognition module has been developed as a part of the biped walking humanoid robot ASIMO's man-machine interaction function in order to enhance its autonomy. ASIMO recognizes human gestures based on statistical analysis, using spatio-temporal variance of the three-dimensional (3D) positions of the face and hand. It is able to distinguish 10 kinds of gestures and postures on both the left and right side. It is also able to recognize a position pointed to by an operator from the relative position of the operator's head and hand and from the direction of the arm. This is calculated by PCA (Principal Component Analysis) on a 3D-point set observed in the proximity of his arm. These components are implemented in a layered structure to reduce complexity, and in an optimally designed communication structure to minimize processing resources. This effort has resulted in a comprehensive image processing system in ASIMO's onboard computer.

1. Introduction

At the end of October 2000, Honda introduced an upgraded version of ASIMO, a small humanoid robot that could adapt to the human living environment⁽¹⁾. The objective was to develop a robot that would actually serve a useful role in close proximity to human beings. This ASIMO was then improved and a new model equipped with a man-machine interaction function and network connectivity was announced in December 2002 (Fig. 1).

ASIMO has previously served in various ways as an attraction and performed reception duties at a variety of locations, thereby gradually gaining acceptance in human society. The former ASIMO, however, performed such actions as walking and shaking hands through handling by an operator who used a remote terminal or portable controller. The robot was not able to interpret human pointing and perform interactive actions autonomously. ASIMO will need to understand human actions and perform actions in collaboration with human beings if it is to evolve into a robot that can be useful to human beings in human society.

Now the previous ASIMO's systems have been expanded by equipping it with a man-machine interaction function using image processing in order to make it a humanoid robot

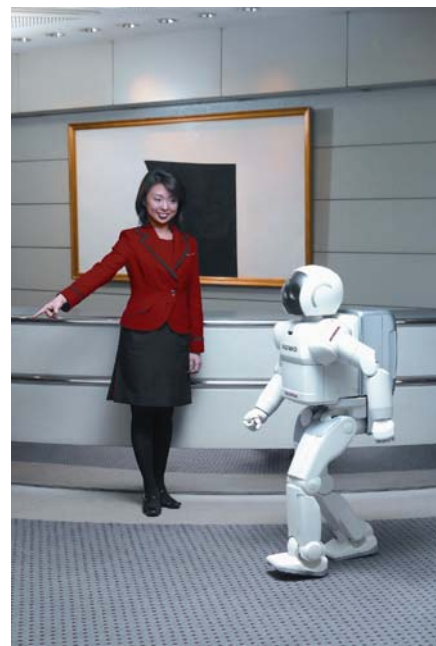


Fig. 1 New ASIMO

* Wako Research Center

** Honda Research Institute USA, Inc

that is endowed with autonomy⁽²⁾.

In general, people find it easier to understand what a robot is doing if it has a humanoid configuration, which also suits the robot for coexistence with human beings in ordinary environments. When giving movement instructions to a humanoid robot by pointing, however, people tend to deal with it in the same way they deal with other people, because it does have a human-like shape. This means that greater value is placed on the capability for communication with human beings than is the case for robots with other configurations.

The means of communication most used among human beings is the voice, whereas communication using gestures fulfills a role of nearly equal importance. Communication by gesture becomes even more important when pointing to a general location, when verbal expression would tend to become wordy, and in cases when the voice would not reach the recipient. Communication using gestures is also effective in conveying nuances that are not easily expressed in words.

Many different proposals have been made for research on gesture recognition. A natural interface with a robot would best be achieved by a means of recognition through non-contact sensors that use image information and so on, since these do not require wearable sensors such as magnetic sensors and data gloves, which impose bodily restraints⁽³⁾. The methods of gesture recognition using image processing that have been proposed so far, however, have involved such issues as fixed cameras⁽⁴⁾, ⁽⁵⁾, numerous limitations with respect to backgrounds and human images⁽⁶⁾, high computation costs and the need for dedicated processing systems⁽⁷⁾, and so on.

The image processing system installed in ASIMO, in particular, demands high-speed techniques that can carry out actions at the same time it handles image processing for face recognition, obstacle recognition, and so on, as well as the gesture recognition that is the subject of the present report. Although this does have the advantage of being able to utilize other processing results, the combination of multiple image processing functions generally makes processing more complex. Even if it were possible to provide greater functional variety and higher reliability, the computation costs become very high, as noted above, and these issues hinder real-time processing. Moreover, even if the different processes could be carried out by single functions, it would be extremely difficult to deal with those functions in an integrated manner.

This paper reports on gesture recognition processing as a real-time and on-board operation that is conducted together with image processing, including face recognition and obstacle recognition.

2. Overview of the Image Processing System

Figure 2 shows an overview of the image processing system installed in the new ASIMO.

The image processing system is made up of two CCD cameras installed in ASIMO's head and a frame grabber and PC board that performs image processing that are installed in ASIMO's body. The angles of each of ASIMO's joints and other such posture information required to compute camera positions and angles and other such camera parameters used during image processing are held in memory on an electronic control unit (ECU) board.

The computer used for image processing is a commercially available PC board with a CompactPCI bus.

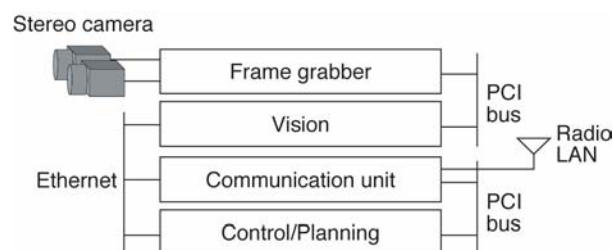


Fig. 2 Image processing system

The computer CPU is a 1.2-GHz Intel Pentium III-M.

A pair of CCD cameras are used to capture images in order to obtain depth information by stereo vision.

A frame grabber is used to capture the image signals output by the cameras. When using stereo vision to identify the location of an object in real space, the frame grabber must acquire the camera parameters matched to each image. For this reason, the frame grabber was designed with the capability for a synchronization signal to synchronize each image with its camera parameters. This makes it possible to perform accurate image processing of time-varying images captured by moving cameras without a need for complex processing.

3. Image Processing Algorithm

3.1. Overview

Figure 3 shows the software configuration of the image processing unit.

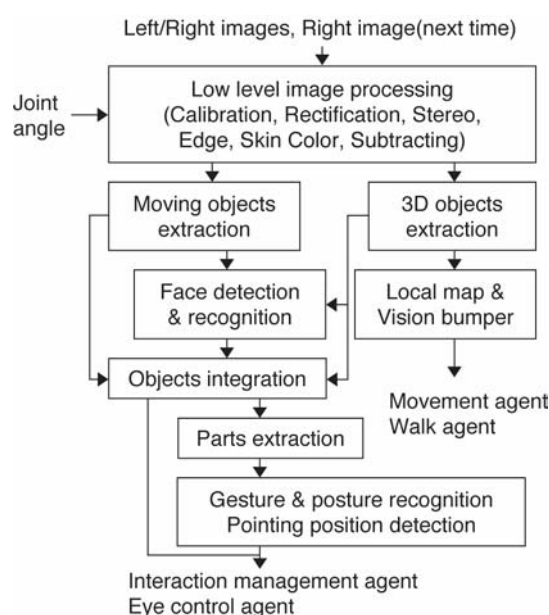


Fig. 3 Vision system software configuration

The images input from the cameras are processed by a low-level image processing unit to correct distortions (upper left row in Fig. 4). This is then followed by computation of depth information (right middle row in Fig. 4), edge information (left lower row in Fig. 4), skin color region information (right lower row in Fig. 4), and movement region information (left middle row in Fig. 4).

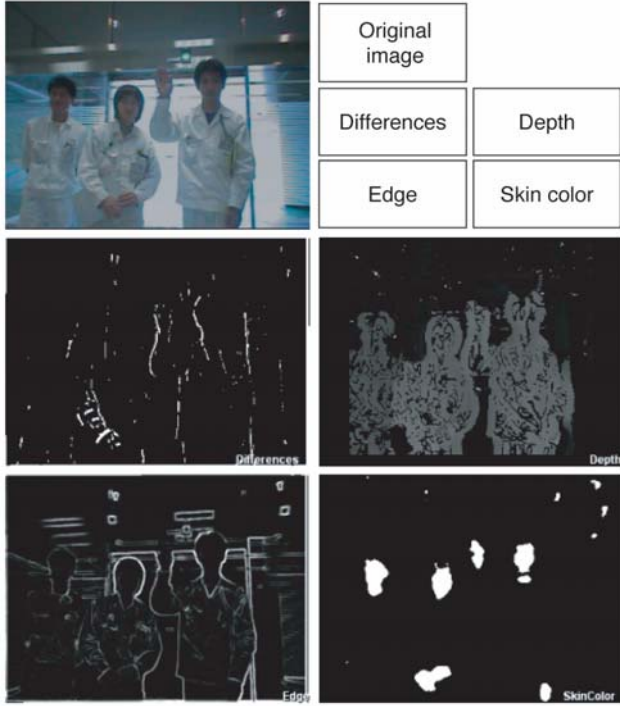


Fig. 4 Preprocessing images

Depth information is obtained by stereo processing using block correlation. Edge extraction is carried out using a Sobel operator. The skin color region is extracted by threshold value operation over the HLS (hue, lightness, saturation) space. Movement region information is extracted by a finite difference calculation after background movement has been estimated from the variance between the image and the camera parameters and the image has been aligned with the background.

These items of information provide the basis for moving object extraction, 3D object extraction, and face detection object extraction processing. At that point integrated processing of the object is carried out in order to integrate the objects within each frame and to process the objects across frames accordingly.

Next, extraction processing is performed on each object or on each object that is subject to tracking, and finally recognition processing is performed for gesture and posture recognition and pointing position detection. After this, the processing results are sent to the human-machine interaction processor and fed back into action by the robot.

Thus the processes in this system are layered through the object extractor, parts extractor and recognition unit, and the process results are put together at each layer in order to keep the processing from becoming complex. At the same time, information from other processes is used effectively so that the search area for each process is kept limited. Means such as these were used to reduce the computation costs and achieve operation at 100 – 200 msec/frame with on-board equipment.

The various processes from object extraction to gesture and posture recognition and pointing position detection are described below.

3.2. Moving Object Extraction

This process performs extraction of the person (moving object) who is the object of gesture recognition. First, the depth, edge, and movement information obtained by low-level image processing are used to establish an approximate rectangular solid region that covers the moving object. Then more detailed shape and position information are extracted from the image information within that region.

In the depth direction, the depth d is extracted from the depth information and movement information taken from the camera that has the largest number of edge points in motion. That value is used as a reference in setting the object depth range for processing.

Next, the pixels within this object depth range (or the pixels for which the edge has been extracted (the pixels shown in blue in Fig. 5)) are extracted and the number of pixels extracted from the image along the vertical axis is counted. The line for which that number is largest is taken as the centerline of the processing region (Fig. 5). A distance corresponding roughly to human shoulders is marked off to the left and right of this centerline, according to the depth d , and the left and right processing regions are established accordingly.

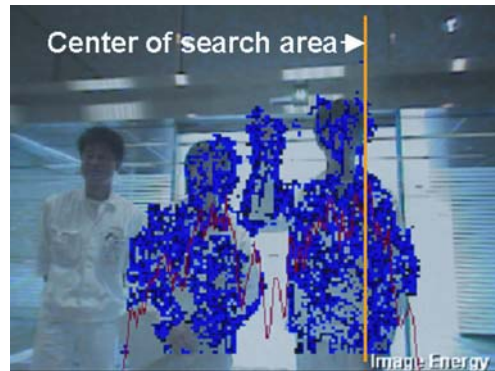


Fig. 5 Center of search area (left-right)

Finally, the vertical processing region is established from the processing depth d and the camera parameter values as the range from the floor to a height above the floor that can fully cover a human body.

The more detailed shape and position of the moving object are extracted from the image information inside the search areas established by the above processes.

Methods for extracting target objects from image information include the segmentation method, which is based on clustering of characteristic features, the contour extraction method⁽⁸⁾, which links together the detected edges, and the dynamic contour model (Snakes⁽⁹⁾), which deforms closed curves so as to minimize a predefined energy.

This system must be able to extract detailed portions of target objects such as faces and hands from complex scenes. Here, therefore, extraction of moving bodies is done using Snakes, which is tolerant of noise and can extract the contour shapes of target objects.

The Snake is a contour model based on the principle of minimized energy. It changes the contour expressed as $v(s) = (x(s), y(s))$ so that the energy calculated in Eq. (1) will be a minimum value.

$$E_{snake} = \int_0^1 E_{int}(v(s)) + E_{image}(v(s)) + E_{con}(v(s)) ds \quad (1)$$

The calculation here of the image energy $E_{image}(v(s))$ uses the intensity of the edges in the search areas described above.

The position of the center of gravity (x_i, y_i) of the moving object is calculated from the positions of the points on or inside of the contour of the extracted moving object i and the angle θ_i of the moving object is determined. The distance d_i to the moving object is calculated from the depth information in the pixels for the points inside the contour, and the position of the moving object in real space is determined. The depth information in the pixels inside the contour of the moving object, which was already extracted, is then updated (the parallax is set to zero), and the process is repeated according to the set processing range in order to extract multiple people.

Figure 6 shows the results for multiple moving objects extracted using Snakes. The rectangles represent the search areas for moving objects within the image, and the points inside the contours represent the centers of gravity of each moving object in the image. The figures to the side express the distance and angle to each moving object.

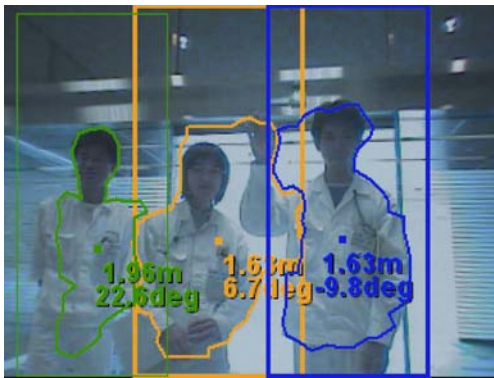


Fig. 6 Extraction of humans

In this way, this method establishes a general rectangular solid region that covers the moving object, and then extracts its more detailed shape and position. Consequently, the method is capable of extracting individual moving objects at high speed even in a scene that has more than one person in a row.

3.3. Parts Extraction

Next the system extracts the parts of the people (positions of faces and hands) to be used in gesture recognition.

3.3.1. Face position extraction

Face position extraction extracts the head from the contour information for the extracted moving object. Using that head as a reference point, it establishes a search area and performs the extraction. The position of the head that serves as a reference is defined as the point located at the highest part of a contour within an area expanded laterally for about the width of a human head to the left and right of the center of gravity of the moving object. This point is used as a reference to establish a search area for the face. The size of the search area on the image varies with the distance d_i to the moving object. The search area in the depth direction is

defined as the approximate area that would cover a human face with its center at the moving object distance d_i .

The face position (x_f, y_f) in the image is defined as the center of gravity of the skin color region in the above search area. The face position (X_f, Y_f, Z_f) in actual space can be calculated using (x_f, y_f) , the moving object distance d_i and the camera parameters.

For the purposes of later hand position extraction, the face position following face position extraction is used to further establish an area where the face is estimated to have moved, and skin color information inside that area is deleted.

3.3.2. Hand center position extraction

The hand center position (x_h, y_h) is also defined, in the same way as the face, as the center of gravity of the skin color region within the search area.

The hand search area is defined as the area where the hands can reach, taking the face position (x_f, y_f) as a reference. In some cases there may be another moving object close to the target moving object. In such cases, the center of gravity of the adjacent moving object is set as the boundary of the search area in order to prevent false detection of the hands or face of a moving object that is not a target.

The hand center position (X_h, Y_h, Z_h) in real space is calculated in the same way as the face position.

3.3.3. Tip of hand position extraction

This method for extracting the hand center position (x_h, y_h) derives only the approximate hand position. Therefore, this position is next used to estimate the tip of hand position (x_{ht}, y_{ht}) , which is required for gesture recognition.

First, the hand center position (x_h, y_h) determined in section 3.3.2. is set as the center for defining a tip of hand search area. This search area is defined to be large enough to fully cover a human hand. The size of the region in the image changes according to the distance moved, as in the preceding processes.

Next the system extracts those points within this region that are the top, bottom, left, and right-most points with skin color information in the image. The hand is considered to be extended in the direction of the longer of the distances between the top and bottom-most points and the left and right-most points. The tip of hand is presumed to be the endpoint, either the top or bottom (or left or right).

Finally, it is decided which is the tip of hand—top or bottom (or left or right)—from the relationship between the hand center position (X_f, Y_f, Z_f) and the face position (X_h, Y_h, Z_h) . Specifically, when the hand center position is distant from the face, the hand is considered to be extended, and the end farther from the face is considered to be the tip of hand position. When the hand center position is close to the body, the elbow is considered to be bent, and the end closer to the face is considered to be the tip of hand position. The tip of hand position (X_{ht}, Y_{ht}, Z_{ht}) in real space can be calculated from the coordinates (x_{ht}, y_{ht}) of the point in the image extracted as the tip of hand position and the depth of that point (the parallax).

Figure 7 shows the hand and face search areas and the results of extraction. The red outlines show the respective search areas. The search areas vary in size on the left and right because when the hand is extracted in one frame, that hand position is used as the center of the search area in the following frame, and this narrows the search area.

The use in this way of the extraction results from the previous frame or the moving object extraction results for the search area makes it possible to extract the face and hands with accuracy and high speed.



Fig. 7 Extraction of face & hands position

3.4. Gesture and Posture Recognition

The face position and tip of hand position in real space extracted in section 3.3 are used for recognition of gesture and posture by observation of the temporal variance in the relationships of those positions.

Gesture recognition methods include use of the hidden Markov model (HMM)⁽⁴⁾, but this involves various problems. It requires large amounts of learning data⁽³⁾ to improve the recognition rate, and real-time processing is difficult when the data are extended to 3D⁽¹⁰⁾. The present system, therefore, extrapolates gestures using the Bayesian method⁽¹¹⁾, a method of statistical analysis that allows high-speed processing and statistically optimum judgement, so that it tends not to result in misjudgment. The vector of characteristics x used to characterize a gesture is made up of six parameters, the averages ($\bar{X}$, $\bar{Y}$, $\bar{Z}$) across several frames of the tip of hand position (X_{ht} , Y_{ht} , Z_{ht}) = ($X_{ht} - X_f$, $Y_{ht} - Y_f$, $Z_{ht} - Z_f$), which has the face position in real space as its origin, and the standard deviations (σ_x , σ_y , σ_z).

When the vector of characteristics x is given, the probability density $p(\omega_i|x)$ that it presents is a gesture i can be found through Bayes' theorem using Eq. (2).

$$p(\omega_i|x) = \frac{p(x|\omega_i)p(\omega_i)}{p(x)} \quad (2)$$

Here $p(x|\omega_i)$ is the conditional probability density that when gesture i is given, the image will include the vector of characteristics x . It is expressed in Eq. (3).

$$p(x|\omega_i) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{(x-\bar{x})^T \Sigma^{-1} (x-\bar{x})}{2}} \quad (3)$$

The vector of characteristics x is defined as having covariance matrix Σ and to follow the normal distribution of the expectation value $\bar{x}$.

Furthermore, $p(\omega_i)$ expresses the probability density of the prior probability of ω_i . When $p(\omega_i)$ is the normal distribution of the expectation value ω_{i0} and distribution $V[\omega_{i0}]$, then $p(\omega_i)$ can be written as in Eq. (4).

$$p(\omega_i) = \frac{1}{\sqrt{2\pi V[\omega_{i0}]}} e^{-\frac{(\omega_i - \omega_{i0})^2}{2V[\omega_{i0}]}} \quad (4)$$

Since the denominator of Eq. (2) does not depend on ω_i , the probability density of the posterior probability of ω_i , following Eqs. (3) and (4), can be expressed as in Eq. (5).

$$p(\omega_i|x) \propto p(x|\omega_i)p(\omega_i) = \frac{1}{\sqrt{2\pi^3} \sqrt{|\Sigma|V[\omega_{i0}]}} e^{-\frac{(\omega_i - \omega_{i0})^2}{2V[\omega_{i0}]} - \frac{(x - \bar{x})^T \Sigma^{-1} (x - \bar{x})}{2}} \quad (5)$$

Equation (5) is used to determine, for a given extracted vector of characteristics x , the probability relative to all gestures that it is gesture i . The vector with the highest probability is taken as the gesture recognition result for that frame. The final output result is obtained from results across several frames.

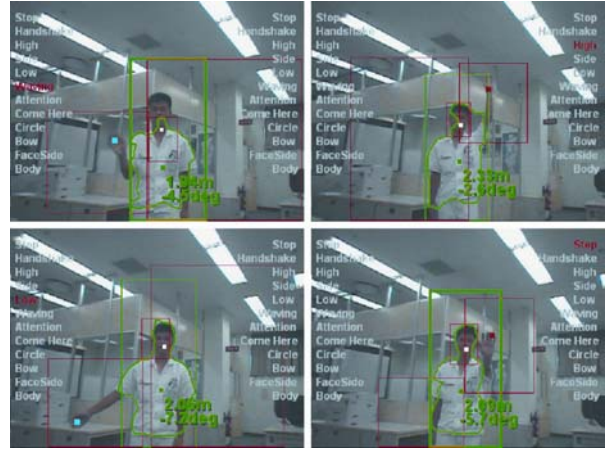


Fig. 8 Gesture & posture recognition results

The system using this method is currently capable of recognizing ten kinds of gesture and posture separately for the left and right hands. These are the handshake, hand circling (moving the hand in a circle beside the body), bye-bye (waving the hand from side to side in front of the body, upper left in Fig. 8), hand swing (waving the hand from side to side above the head), high-hand (raising the hand, upper right in Fig. 8), side-hand (extending the hand out beside the body), low-hand (extending the hand diagonally downward beside the body, lower left in Fig. 8), come here (beckoning by waving the hand up and down with palm down in front of body), stop (holding out the hand palm-forward in front of the body, lower right in Fig. 8), and face-side (raising the hand to the side of the face). The postures and gestures shown in red on the right and left sides of Fig. 8 are recognition results.

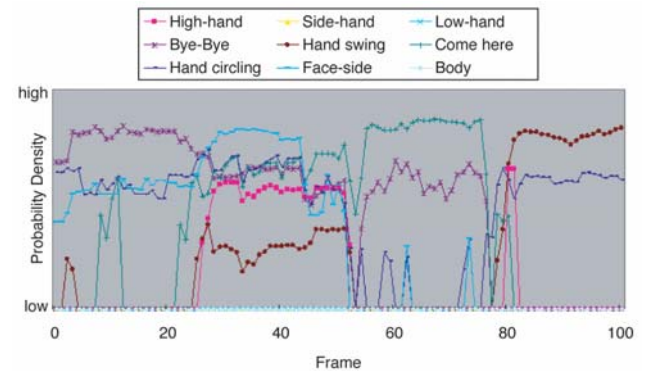


Fig. 9 Probability density

Figure 9 shows the probability densities for the bye-bye, face-side, come here, and hand swing gestures and postures. The graph indicates that the probability densities for gestures and postures performed at that time are clearly higher than the probability densities for other gestures and postures. It is clear that correct recognition is taking place.

3.5. Pointing Position Detection

Pointing position detection processing takes place by the procedure shown in Fig. 10.

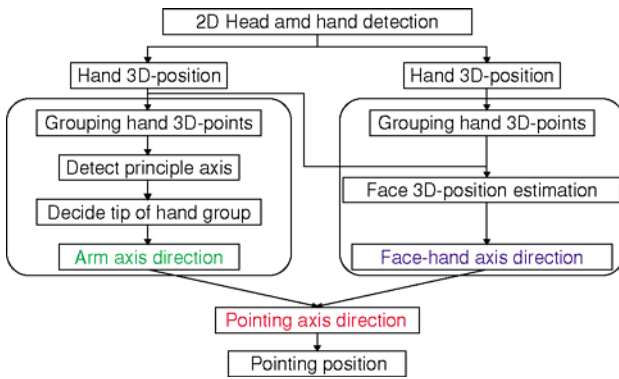


Fig. 10 Flowchart of pointing position detection

First, the tip of hand and head positions obtained by parts extraction from certain image acquisition frames are used as a reference to obtain points corresponding to the head and ends of the arms in a stereo image. Specifically, objects within a certain range by 3D distance from the reference position are taken as the head and ends of the arms. These are called the head group and hand group, respectively.

For the head group, the center of gravity is taken as the head center. For the hand group, the principal axis is determined through principal component analysis of the values of the 3D coordinates. This calculation is performed as follows. There are N points in the group. The average of the coordinate values for these points normalized to 0 is A_n . The normalized group is expressed as a matrix, in the form $A=[A_1, A_2, \dots, A_N]$. The distribution covariance matrix for this matrix is determined, and its eigenvalues and specific vectors are calculated. The specific vector corresponding to the largest eigenvalue is taken as the largest principal axis of the hand group distribution.

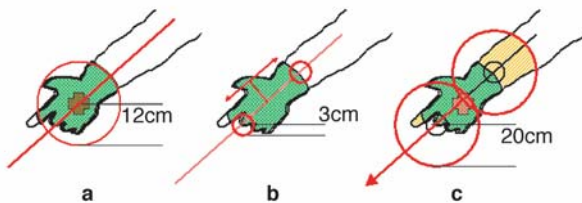


Fig. 11 Arm direction detection

The direction of the end of the arm is sought for the principal axis thus obtained. The method employed is to search for locations in two directions along the principal axis and within a radius of 3 cm from the center of gravity where there are no 3D points (Fig. 11b). The direction in which there are fewer 3D points within 20 cm from the location within 3 cm where there are no 3D points is taken as the end direction (Fig. 11c).

Finally, the relationship of the head group and hand group positions is confirmed, and when the selected direction is not possible in terms of body structure, it is corrected to the reverse direction. If the reverse direction is also not possible, then it is determined that processing could not be accomplished with that image acquisition frame.

The same kind of process is applied to three or more frames, and when the total distribution of the head center position, the arm axis direction and the arm positions lies within certain values, the arm and head posture is considered to have been identified. Arm processing is carried out on both the right and left arms, and the hand on the side that is identified first and that is identified as being farther away is considered to be the pointing hand.

After the arm and head posture has been identified, the face position is estimated on the supposition that the person pointing will be looking at the pointing hand. The head is supposed to be a sphere with a radius estimated from the body height. The surface of the head in the direction from the head center toward the pointing hand is determined. The neck is considered to have rotated perpendicularly to the Z axis, and the value of Z set back to the value of the head center is considered to be the eye position (see Fig. 12 for the coordinate system).

The vector linking the eye position and the hand position is the face-hand vector ($V_{\text{face-hand}}$), and the vector in the arm direction is the arm vector (V_{arm}). The composite of the face-hand vector and arm vector is the pointing vector (V_{pointing}) (Fig. 12). The composite is as shown in Eq. (6).

$$V_{\text{pointing}} = \frac{V_{\text{arm}}}{|V_{\text{arm}}|} + \frac{V_{\text{face-hand}}}{|V_{\text{face-hand}}|} \quad (6)$$

However, if the distance from the head to the pointing hand is 70% or less of the full arm length as estimated from body height, then the elbow is considered to be bent. In this case, the arm vector is taken as the pointing vector.

The location where the line extended from the pointing vector that has the hand position as its origin intersects with the floor surface is considered to be the pointing position (the red line in Fig. 12).

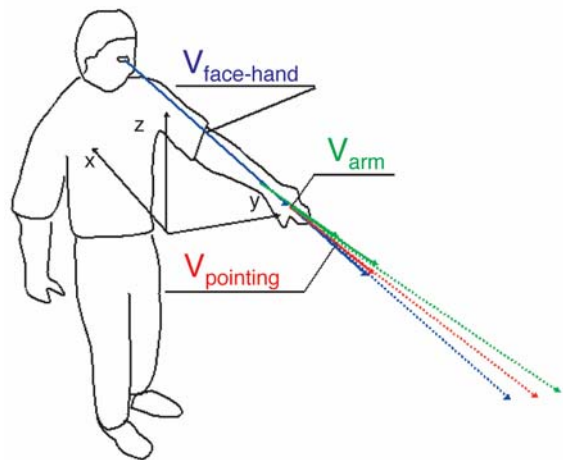


Fig. 12 Pointing gesture overview

The above algorithms were incorporated into the system equipment and the accuracy of this method was verified. The stereo cameras were installed 1.5 m in front of the test subject (ASIMO was stood 1.5 m in front of the test subject).

Images of the subject pointing to seven lateral points at a distance of 1 m in front of the test subject were used. The pointing position was detected using three methods, extrapolation of the pointed floor surface using only the arm vector, using the face-hand vector, and using the pointing vector. Figure 13 shows the detection results.

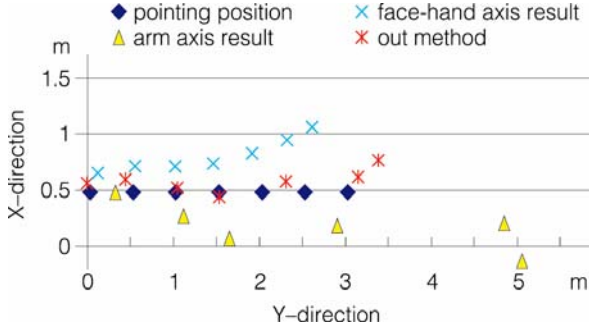


Fig. 13 Result of pointing position detected by each vector

The average error was 0.93 m when the pointed floor surface was extrapolated using only the arm vector, 0.34 cm when using the face-hand vector, and 0.25 cm when using the pointing vector. It is clear that the recognition results are improved by use of the pointing vector, which is a composite of the face-hand vector and arm vector.

During this testing, the test subject simply pointed toward the floor without any deliberate attention to the detection method. By using this method, the system became able to detect a point closer to the intended position indicated by the pointing in a natural manner than just the position derived from the arm vector and face-hand vector.

The accuracy of detection was also determined using six test subjects (S1-S6 in Fig. 14), who each extended an arm to point toward nine points diagonally to the left (1.4 to 4.2 m ahead). They did this three times (Fig. 14).

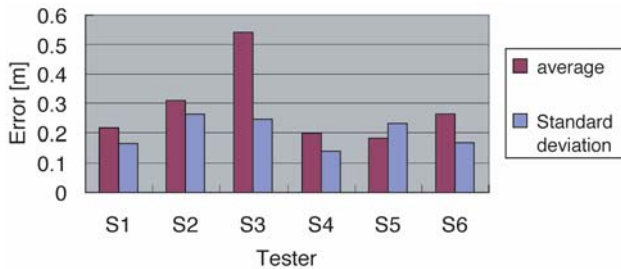


Fig. 14 The detection error for every subject

The average error for all points was 0.28 m (broken down by pointing distance, this was an error of 0.15 m for 1-2 m, 0.25 m for 2-3 m, and 0.45 m for 3-5 m). The range extended from an error of 0.54 m for the test subject with the largest error, to 0.18 to 0.31 m for the other test subjects. There was some individual variation, but this is considered to indicate a largely stable detection capability.

Figure 15 shows an example of pointing position recognition results. The left side of Fig. 15 is an illustration of a stereo image overlaid by recognition results, and the right side is an overhead view of those recognition results.

In this way, the system became capable of extracting a stable, highly accurate pointing position through use of the relationship between the head and arm positions in addition to the direction of the arm.

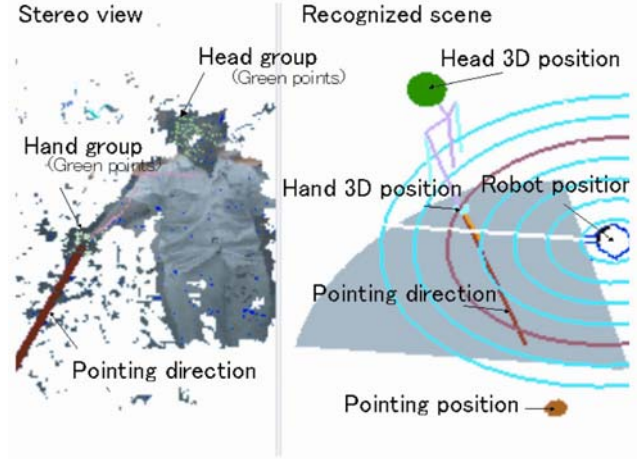


Fig. 15 A Pointing gesture recognition result

4. Test Results

Installation of these gesture recognition functions in ASIMO enabled it to come to where it was called over by a gesture or move to a pointing position.

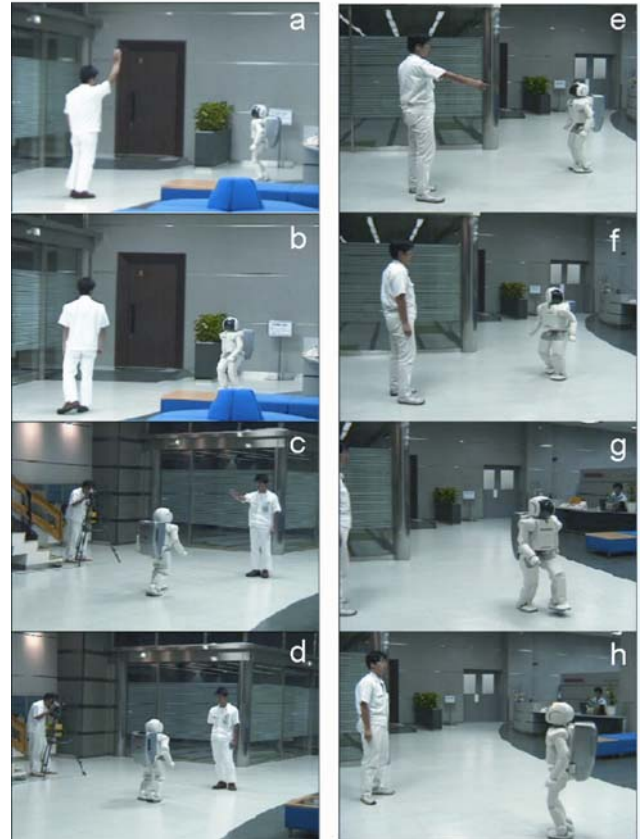


Fig. 16 Demonstration images a: recognize “high-hand” posture, b: start to follow, c: recognize “stop” posture, d: stop walking, e: recognize pointing position, f: start to walk, g: walk to pointing position, h: arrive and stand by

Specifically, ASIMO has become able to obey a person when it sees a person raising a hand high or gesturing to come. In addition, when the person gestures ASIMO to stop, it is able to stop on the spot. (See examples on left in Fig. 16.)

When a person tries to shake hands with ASIMO, it will extend its right hand to shake, and in response to other gestures and postures, it will perform the same gestures and postures as a person.

In addition, when ASIMO receives voice input that requests it to move, such as by saying "go over there," it has been enabled to recognize the position where the person is pointing, and to move to that location. (See examples on right in Fig. 16.)

5. Conclusion

A gesture recognition function using image processing was installed on the biped walking humanoid robot ASIMO in order to enhance ASIMO's autonomy.

- 1) The system extracted people (moving objects) that are the objects of gesture recognition on the basis of movement information, and further used color information to extract the human head and hands. From the temporal variance in the 3D positions of the head and hands, it used statistical analysis to carry out gesture and posture recognition, and has become able to recognize ten types of gesture and posture.
- 2) Further, the direction of the arm was determined by principal component analysis of the 3D point set observed in the vicinity of the hand. Then the direction a person was pointing and the point being indicated were extracted from the relationship between the positions of the head and arm and the direction of the arm.

This made it possible to use gestures to give ASIMO some instructions for actions that up to now had been provided by an operator using a remote terminal to execute operating instructions one by one. This enhanced ASIMO's operability and autonomy.

The present system implements object extractor, parts extractor, and recognition unit functions in a layered structure. The results of the processes are put together at each layer, and the system has thus avoided complexity in the processing. Furthermore, information from other processes is used effectively so that the search area for each process is kept limited. Means such as these were used to reduce the computation costs. As a result, this system realized a comprehensive image processing system that was the first in the world capable of operating at a speed of approximately 100 – 200 msec/frame with on-board equipment.

References

- (1) Hirose, M., Takenaka, N.: Development of Humanoid Robot ASIMO, Honda R&D Technical Review, Vol. 13, No. 1, p. 1-6 (2001)
- (2) Sakagami, Y., Watanabe, R., Aoyama, C., Matsunaga, S., Higaki, N., Fujimura, K.: The intelligent ASIMO : System overview and integration, IROS, p. 2478-2483 (2002)
- (3) Yachida, M., Iwai, Y.: Human Motion Analysis and Recognition for Interaction, Technical Report of IEICE, PRMU99-142 (1999)

- (4) Takahashi, M., Ogura, N., Watanabe, S.: Recognition of mixture motions using Hidden Markov Model, Technical Report of IEICE, PRMU 2000-151 (2001)
- (5) Takahashi, K., Sakaguchi, T., Ohya, J.: Remarks on a Real-Time, Non-contact, Non-wear, 3D Human Body Posture Estimation Method, Trans. IEICE (D-II), Vol. J83-D-II, No. 5, p. 1305-1314 (2000)
- (6) Ishii, H., Mochizuki, K., Kishino, F.: A Motion Recognition Method from Stereo Images for Human Image Synthesis, Trans. IEICE (D-II), Vol. J76-D-II, No. 8, p. 1805-1812 (1993)
- (7) Takahashi, K., Seki, S., Kojima, H., Oka, R.: Spotting Recognition of Human Gestures from Time-Varying Images, Trans. IEICE (D-II), Vol. J77-D-II, No. 8, p. 1552-1561 (1994)
- (8) Ballard, D., Brown, C.: Segmented Images, Computer Vision, p. 147-205 (1987)
- (9) Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active Contour Models, Proc. of 1st ICCV, p. 259-268 (1987)
- (10) Muller, S., Eickeler, S., Rigoll, G.: Crane Gesture Recognition Using Pseudo 3-D Hidden Markov Models, In 4th IEEE International Conference on Automatic Face and Gesture Recognition, p. 398-402 (2000)
- (11) Kanatani, K.: Bayesian estimation, p. 149-150 (1995)

■ Authors ■



Nobuo HIGAKI



Taro YOKOYAMA



Kikuo FUJIMURA