Position Tracking of the Mouth Using Image Processing

Hiroki Higa, Kazutoshi Tsuha, Ryota Onaga, and Ikuo Nakamura

Abstract—**In this paper, an image processing method using CMOS camera was applied to an assistive system to care for bedridden people. It was experimentally demonstrated that 2-demensional position tracking of the experimental subject's mouth was performed using image processing. Furthermore, 3-demensional position tracking of the subject's mouth was measured using stereo matching measurement. From the experimental results, it was clear that the 2-demensional postion of his mouth was correctly tracked 90 % in all frames, and that the 3-demensional postion tracking of his mouth was effectively measured.**

I. INTRODUCTION

There are many people who are confined to their beds for a long time by the cause of stroke or traffic accidents. Some of them have strong-minded to be independent of others and to live their own lives. For the purpose of assisting the bedridden people to live, we have developed the assistive robot arm system [1], [2]. Our concept with respect to this system is to mainly carry out nursing care of the bedridden patients by their family and to subsidiarily assist nursing care of them by the assistive robot arm. The system has a drink water command as one of commands. In this case, the end effector (hand) of the robot arm system gets close to the user's face. In order to safely control the assistive robot arm, an image processing method using CMOS cameras was adopted as one of feedbacks, and detection of an experimental subject's mouth was done in real time. Furthermore, a distance between cameras and the user's mouth was measured by using stereo matching measurement.

II. SYSTEM CONFIGURATION AND IMAGE PROCESSING

A. Overview of System

A system configuration of the assistive robot arm is shown in Fig. 1. This is composed of a robot arm, personal computer (PC), control device, monitor display, and CMOS cameras. It has the following functions: (1) to help users to have meals and something to drink, and (2) to inform their personal doctors or their family of what they are doing with the robot arm if necessary, and to e-mail and browse so the users can have many opportunities to communicate with people all over the world. The robot arm has the ability to tilt the screen so as to acquire the user's face from a most advantageous position.

B. Detection of User's Mouth Using Image Processing

To detect the color of the human skin component, the Y- and I-elements of YIQ color system were used. The transform matrix [3] from RGB to YIQ is given by

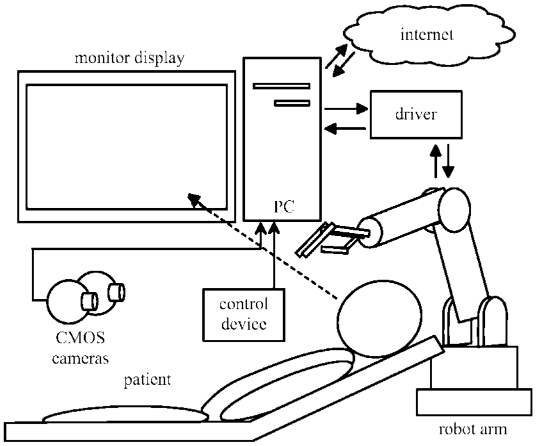


Fig. 1. System configuration of assistive robot arm.

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.522 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}.$$

It is very important to know relative location information among a robot arm, user and some objects such as a bed and table so as to prevent the robot arm from being hit against the user when it is applied to nursing care. Two CMOS cameras were used to obtain a spatial coordinate of an object in the nursing care environment. In this paper, stereo matching measurement [4] was used to calculate the spatial coordinate of the object from a pair of images.

The stereo matching measurement is the method to obtain 3-dimensional coordinate from a pair of images. Fig. 2 shows the principle of the stereo matching measurement. It is defined that the point O_L is the original point in the stereo matching measurement. Two cameras are located at the points O_L and O_R. The coordinates of a point $P(x_p, y_p, z_p)$ to be measured are given by

$$z_p = \frac{d \cdot f}{x_L - x_R},$$

$$x_p = \frac{d \cdot x_L}{x_L - x_R},$$

and

$$y_p = \frac{d \cdot y_L}{x_L - x_R} = \frac{d \cdot y_R}{x_L - x_R},$$

where $P_L(x_L, y_L)$ and $P_R(x_R, y_R)$ are the position coordinates of an object on the images obtained by the left and right

Authors are with the Department of Electrical and Electronics Engineering, University of the Ryukyus, Nishihara, Okinawa 903-0213, Japan

1-4244-0033-3/06/$20.00 ©2006 IEEE.

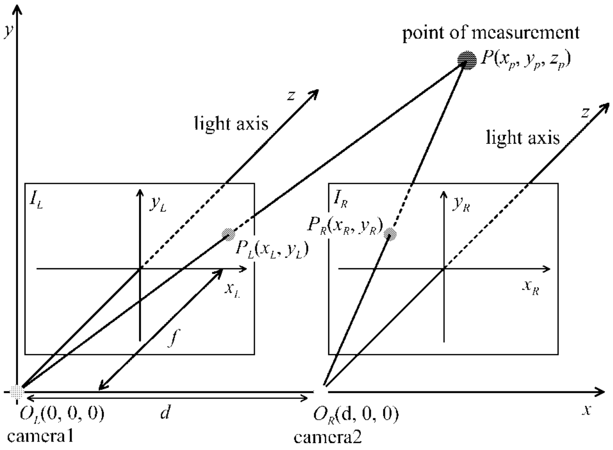


Fig. 2. Principle of stereo matching measurement.

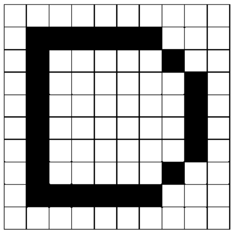


Fig. 3. Example image.

cameras, and d and f are the distance between two cameras and that between the cameras and a frame, respectively.

In general, a camera's lens has a distortion. This leads to cause errors in any analysis performed with video-based motion. Thus, the correction equations of the camera's lens distortion were also used as expressed by eqs. (1) and (2)[5],

$$\begin{aligned}\Delta x &= -x_p + (k_1 r^2 + k_2 r^4 + k_3 r^6)(x - x_p) \\ &+ p_1\{r^2 + 2(x - x_p)^2\} + 2p_2(x - x_p)(y - y_p), \quad (1)\end{aligned}$$

and

$$\begin{aligned}\Delta y &= -y_p + (k_1 r^2 + k_2 r^4 + k_3 r^6)(y - y_p) \\ &+ 2p_1(x - x_p)(y - y_p) + p_2\{r^2 + 2(y - y_p)^2\}, \quad (2)\end{aligned}$$

where $r^2 = (x - x_p)^2 + (y - y_p)^2$, (x, y) and $(x + \Delta x, y + \Delta y)$ are the coordinates before correction and after correction, respectively. (x_p, y_p) is the center coordinate of the image, and parameters k_1, k_2, k_3, p_1, and p_2 are constant.

The Gray Level Run Length method [6] is a way to compress information. For a binary image, there are white pixels having the pixel values of 1 and black pixels having the pixel values of 0. Consecutive pixels of the same value in a given direction consist of a run [7]. Figure 3 shows an example image. Assuming that the first pixel value is known, and that a right edge pixel of the image is connected to the left edge pixel at the next row, the 2-dimensional matrix can be represented by the number of runs of length: 11, 6, 4, 1, 5, 1, 3, 1, 6, 1, 2, 1, 6, 1, 2, 1, 6, 1, 2, 1, 6, 1, 2, 1, 5, 1, 3, 6, 13. This means that the 2-dimensional matrix of 100 pixels is encoded as 29 numbers. In our system, this technique was adopted to detect the experimental subject's mouth in real time.

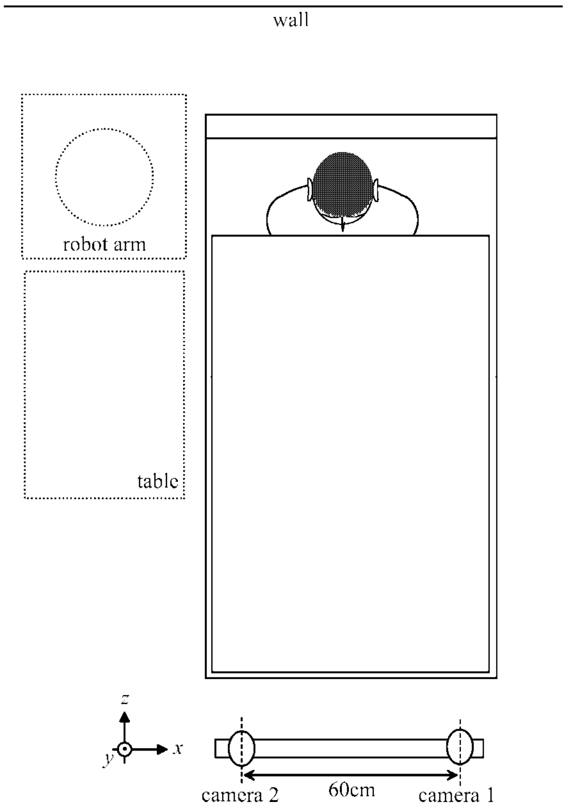


Fig. 4. Environmental setup for nursing care (top view).

III. Experimental method

A. 2-Dimensional Position Tracking cf the Experimental Sutject's Mouth

Before the experiments, the correction of lens distortion of cameras was done. The procedure was as follows. A 30 × 28 matrix-like structure, which was drawn with a black square, 3 cm on a side, on a sheet of white paper was fixed on a wall. A distance between two cameras (WebCam 5, Creative Co. Ltd.,) was fixed to 60 cm. The pair of cameras was placed at a distance of 140 cm away from the wall, the matrix-like structure was taken with it, and the pair of cameras was moved back at intervals of 10 cm up to 250 cm from the wall. The resolution of the images was set to 640 × 480 pixels, they were saved in 24-bit Windows bitmap format, and using these data and eqs. (1) and (2), the distortion of the lens was calibrated.

In order to safely perform a drink water command using the assistive robot arm system, 2-dimensional coordinates of the experimental subject's mouth was firstly measured. An experimental setup is shown in Fig. 4. An experimental subject (24 years old) sat on a bed, and camera 1 was used to detect the subject's mouth. The resolution of image was set to 320 × 240 pixels for reducing computation

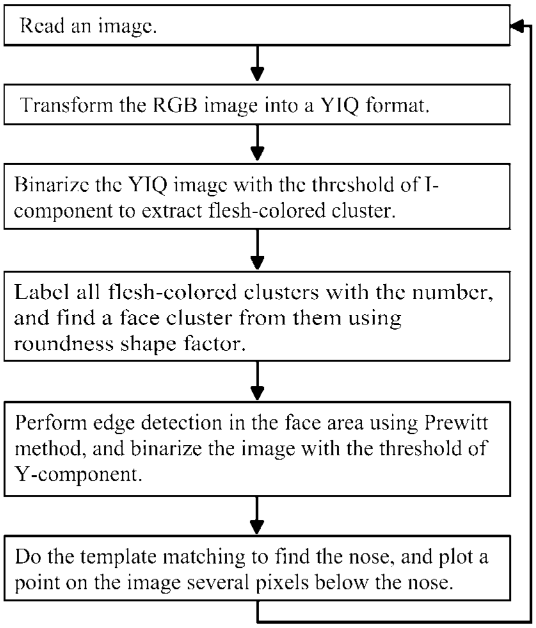

Read an image.
Transform the RGB image into a YIQ format.
Binarize the YIQ image with the threshold of I-component to extract flesh-colored cluster.
Label all flesh-colored clusters with the number, and find a face cluster from them using roundness shape factor.
Perform edge detection in the face area using Prewitt method, and binarize the image with the threshold of Y-component.
Do the template matching to find the nose, and plot a point on the image several pixels below the nose.

Fig. 5. Flowchart of image processing to detect user's mouth.

time, and the image processing program was made using Visual C++. The steps of detecting the experimental subject's mouth are shown in Fig. 5. It was difficult to detect the experimental subject's mouth directly. Thus, detecting the subject's face area first, we obtained his nose position in the face area, and using the nose position, his mouth position was finally detected. A distance between the experimental subject's head and CMOS cameras was about 1.65 m. The sequence of the head movements was as follows: initial position (looking forward), right rotation, initial position, left rotation, initial position, extension, initial position, flexion, and initial position. A white point was plotted at the position of the subject's mouth.

B. 3-Dimensional Position Tracking of the Experimental Subject's Mouth

In the next experiment, 3-dimensional coordinates of the experimental subject's mouth was measured. The resolution of motion-image data was set to 320 × 240 pixels. A distance between the experimental subject's head and CMOS cameras was about 1.65 m. In this case, the cameras 1 and 2 were used, and a red cross-shape mark was attached on the wall as a reference point. The squence of head movements was the same as the previous experiment. A white point was plotted at the position of the subject's mouth.

In addition, the static distance measurements were also performed at the distance of 160 cm between two cameras and a dummy head and that of 170 cm between them, respectively. In this experiment, the dummy head was fixed at the distance of 160 cm away from the two cameras, and the distance was measured using the stereo matching measurement. The measurement in the case of the distance of 170 cm between two cameras and the dummy head was also done.

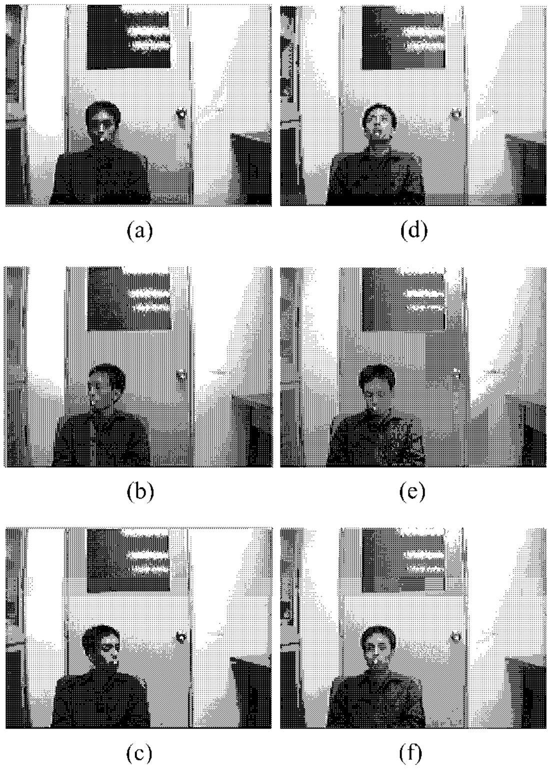

(a) (d)
(b) (e)
(c) (f)

Fig. 6. Experimental results captured from motion-image data. The head movements are (a) initial position (looking forward), (b) the right rotation, (c) the left rotation, (d) the extension, (e) the flexion, and (f) initial position, respectively. Experimental subject's mouth was plotted by white point.

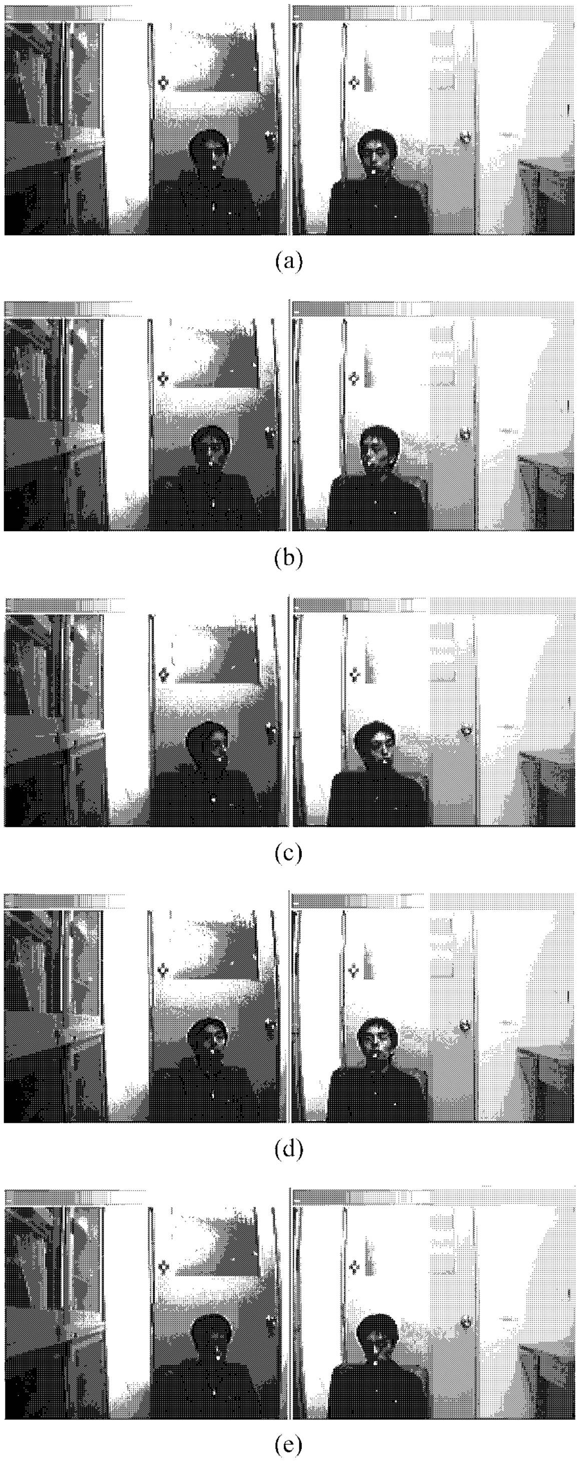
IV. Experimental Results and Discussions

A. 2-Dimensional Position Tracking of the Subject's Mouth

The experimental results of 2-demensional position tracking of the subject's mouth are shown in Fig. 6. The experimental subject's mouth on the processed images were respectively plotted by a white point. It was obvious from the experimental result that the experimental subject's mouth could be correctly detected at the frame rate of 30 frames/sec. We defined the success as the thing that the points obtained by the image processing were in the subject's mouth, and counted how many success points in all frames we obtained in the experiment. It was clear that the success rate of the motion-image processing was 90%.

B. 3-Dimensional Position Tracking of the Subject's Mouth

The experimental result of 3-demensional position tracking of the subject's mouth are shown in Fig. 7. From the experimental result, it was found that the experimental subject's mouth was correctly detected at the frame rate of 25 frames/sec. The rms errors between the distances of the experimental conditions (160 cm and 170 cm) and those obtained by the stereo matching measurement were calculated as an evaluation of the static distance measurement. The result is shown in Fig. 8. It was obvious that the rms errors of distance between two cameras and the dummy head were effectively small.



(a)

(b)

(c)

(d)

(e)

Fig. 7. Experimental results captured from motion-image data in the cases of the head movements of (a) initial position (looking forward), (b) the right rotation, (c) the left rotation, (d) the extension, and (e) the flexion. The both of experimental subject's mouth in right and left processed images were plotted by white points.

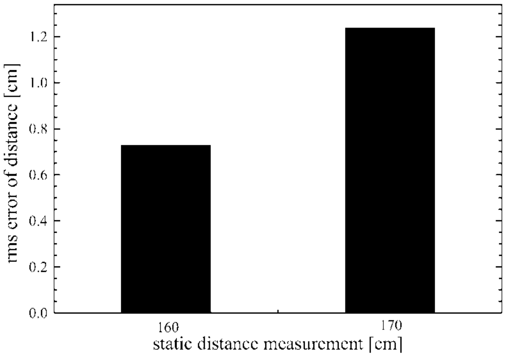


Fig. 8. Rms errors of static distance measurement.

V. Conclusion

In this paper, the motion-image processing program was made for the assistive robot arm system, and the position tracking of the experimental subject's mouth was experimentally verified. The experimental results showed that the 2-dimensional position of his mouth was correctly tracked 90 % in all frames. Further, it was experimentally demonstrated that it was possible to measure 3-dimensional position of the subject's mouth using stereo matching measurement. A further consideration of the performance with the assistive robot arm will be needed for our future work.

Acknowledgments

This work was partly supported by the Tateishi Science and Technology Foundation, and University of the Ryukyus under a grant for scientific research.

References

[1] H. Higa, A. Katayama, K. Tsuha, and I. Nakamura, "A Study on Application of a Robot Arm to Welfare Equipment for Nursing Care -Using Input Interface and Visual Feedback-," Proc. of ITC-CSCC 2004, pp. 7E2L-3-1 - 7E2L-3-4 (CD-ROM), 2004.

[2] H. Higa, K. Tsuha, and I. Nakamura, "A Basic Study on Position Detection of the Lip Using Image Processing," Proc. of ITC-CSCC 2005, pp. 1283-1284, 2005.

[3] H. Nakagawa, "Development of the high-speed image recognition for an autonomous mobile robot," master's thesis of Japan Advanced Institute of Science and Technology, 2002.

[4] T. Agui, and T. Nagao, "Introduction to image processing by C language," Shokodo, pp.127-129, 2001.

[5] H. Inomoto, S. Hukube, K. Akimoto, and Y. Onishi, "The camera calibration by the 2-dimensional target place," Proceedings of Annual Meeting of Japan Society of Photogrammetry 2003, pp.37-40, 2003 (in Japanese).

[6] A. Chu, C. M. Sehgal, and J. F. Greenleaf, "Use of gray value distribution of run lengths for texture analysis," Patt. Recogn. Lett., vol.11, pp.415-420, 1990.

[7] K. Shoji, and T. Miyaji, "High-speed connected element labeling using runs information," IEICE Trans. D-II, vol.121-C, no.2, pp.392-400, 2001 (in Japanese).