Design and Implementation of a Real-time Region Pointing System using Arm-Pointing Gesture Interface in a 3D Environment

Yun-Sang Han, Yung-Ho Seo, Kyoung-Soo Doo and *Jong-Soo Choi*

Department of Digital Imaging, Graduate School of Advanced Imaging Science,
Multimedia and Film, Chung-Ang University,
221 Huksuk-Dong, Dongjak-Ku, Seoul 156-756, Korea
E-mail : deathknight@imagelab.cau.ac.kr, warmlove@imagelab.cau.ac.kr,
dooks@cau.ac.kr, jschoi@cau.ac.kr

ABSTRACT

In this paper, we propose a method to estimate pointing region in real-world from images of cameras. In general, arm-pointing gesture encodes a direction which extends from user's fingertip to target point. In the proposed work, we assume that the pointing ray can be approximated to a straight line which passes through user's face and fingertip. Therefore, the proposed method extracts two end points for the estimation of pointing direction; one from the user's face and another from the user's fingertip region. Then, the pointing direction and its target region are estimated based on the 2D-3D projective mapping between camera images and real-world scene. In order to demonstrate an application of the proposed method, we constructed an ICGS (interactive cinema guiding system) which employs two CCD cameras and a monitor. The accuracy and robustness of the proposed method are also verified on the experimental results of several real video sequences.

Keywords : Pointing region estimation, Motion analysis, Camera calibration, 3D reconstruction.

1. Introduction

As the ubiquitous technology has advanced, the researches on the technologies to provide more convenient services using computer are actively conducted. In general, people mark their interest region to be provided the service and show others their intention with the pointing gesture using hand movements. That is, if the users point a specific product or information among the many products or information displayed in the shop to show their active interest in them, they can receive only the specific information they want from the service provider in an easier way [1]. Therefore, the estimate of the pointed direction and the detection of the pointing region is a very important matter to the service provider. However, in order for the computer to perform the estimate of the pointed direction as playing the role of service provider, it is required to sense the real-time movements by human face and hands and to analyze them to respond to the requests [2,3].

The estimate of the pointed direction using the analysis of the hand movements is applied to and studied in the field of gesture recognition and robot control. Watanabe estimated the front face using the linear discriminant analysis for the 4-direction characteristics of the face region from each direction with the eight cameras installed in a 45°-interval radial shape. In this method, the two closest cameras to the direction of the estimated front face are selected and the pointed direction is estimated by extracting the eyes and the fingertip [4]. This method has the advantage to minimize the concealed region of the fingertip according to the camera directions. However, it is an experiment in a restricted condition and thus has a disadvantage to use many cameras. Yamaguchi proposed a method to deliver the commands to a robot using the user's hand movements or to control the robot's moving direction through the pointing movement [5]. The estimate of the pointed direction is made in a way that the markers are attached to the parts of human body to obtain the RGB information which is used to estimate the pointed direction and to move the robot to the estimated point.

We propose a system to estimate the real-world pointed region using the face and the fingertip of the user by two cameras. In addition, it composes the "Interactive Cinema Information Guiding System" consisting of two cameras and a beam projector for the actual application of the proposed system. This system is not limited to a one-way unilateral information supply system but can be used as a system to provide various cinema information on user's demand.

First, using the planer pattern images captured by the two cameras, we calculate the projection matrix for each camera. Next, after removing the first background by the subtraction of the background image from the current image, it removes the soft shadow using the hue difference in the HIS color space, which results in the interest region. After finding the head through the template matching process and the location of the fingertip by the hue value of the HSV color space in the interest region, it estimates the 3-dimensional coordinates of the two points using the projection matrix obtained through the above process and provides the cinema information in the pointed region in the real world.

2. Proposed Method

2.1 Camera Calibration

Camera calibration is essential to calculate the 3D information of important features from human.

When a point on 3D real world is supposed as X_i, there are corresponding points x_L and x_R acquired from two cameras and we can calculate P in the condition of $X_iP_R = x_{iR}$ and $X_iP_L = x_{iL}$. Camera calibration is to estimate P matrix the can be decomposed in terms of the internal and external parameters as

$$P = K\,[\,R \mid T\,]$$

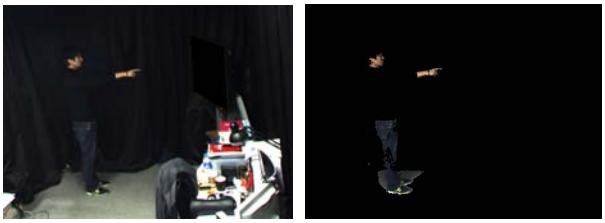
$$K = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix},\ R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix},\ T = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} \quad (1)$$

where the 3×3 symmetric matrix K is internal camera parameters matrix, R is the rotation matrix, and T is the translation vector. 3D coordinate of the real world can be estimated from a camera coordinate x_i by using $x_{iR} = X_i P_R$ and $x_{iL} = X_i P_L$ which are relations between x_i and projection matrix P.

In this paper, by using the robust camera calibration method [6] of Z. Zhang using the 2D planer pattern, we estimate projection matrixes of two cameras.

2.2 Background Image Removal

Fig.1 shows the extracted object from the camera image input. After estimating the primary interest region using the RGB difference between the current image and the background image in order to extract the moving object in the camera images, it minimizes the interest region using the background subtraction to remove the soft shadow by the hue difference in the HSV color space [7].



(a) Input image (b) Object extraction

Fig. 1 : Foreground region after the background subtraction.

2.3 Human Body Detection

To estimate the user's pointed region, the face and the fingertip of the user should be detected. For real-time estimate of the pointed region, only the object is extracted in the previous process. First, the face region is extracted through the template conformation within the region of the extracted object. The equation (2) is to calculate the normalized correlation coefficient (NCC), and the face location is detected through the obtained correlation coefficient.

$$\tilde{R}(x,y) = \frac{\sum_{y'=0}^{h-1}\sum_{x'=0}^{w-1}\tilde{T}(x',y')\tilde{I}(x+x',y+y')}{\sqrt{\sum_{y'=0}^{h-1}\sum_{x'=0}^{w-1}\tilde{T}(x',y')^2\sum_{y'=0}^{h-1}\sum_{x'=0}^{w-1}\tilde{I}(x+x',y+y')^2}} \quad (2)$$

$\tilde{I}$ means the current image obtained from the cameras, $\tilde{T}$ means the template image, and *h* and *w* mean the height and width of the template image.

To detect the fingertip, the hand area is extracted first using the skin color [8,9], and the pixel position farthest distant from the face region is obtained.

2.4 Pointing Region Estimation

Each of the face region and fingertip region are detected for each of the two cameras. The calibration of each camera is finished after calibrating the radiation distortion. Equation (3) is an equation to perform the linear triangulation to estimate the 3-dimensional coordinates from the two cameras [10]. $p^{1T}, p^{2T}, p^{3T}, p'^{1T}, p'^{2T}$ and p'^{3T} are row vectors of the projection matrix for each camera, and (x, y), (x', y') are the coordinates of the camera. The 3-dimensional coordinates in the space can be estimated from the SVD of the matrix A.

$$\mathbf{AX} = \begin{bmatrix} xp^{3T} - p^{1T} \\ yp^{3T} - p^{2T} \\ x'p'^{3T} - p'^{1T} \\ x'p'^{3T} - p'^{2T} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \mathbf{0} \quad (3)$$

From the above equation, the 3-dimensional coordinates for the face region and the fingertip region of the user are obtained. The pointing point is the intersection point where the indication line to connect the coordinates of the user's face to the coordinates of the fingertip and the interest area cross with each other.

2.5 Interactive Cinema Information Guiding System(ICIGS)

To realize the proposed algorithm, the "Interactive Cinema Information Guiding System" is designed which is an applied system to guide the cinema information from the user movement information [11].

Fig.2 shows the drawing of the proposed "Interactive Cinema Information Guiding System" in which two 1394 cameras in the user's lateral side and a monitor in the user's rear side are arranged. At the front side of the user, 9 cinema previews will appear. If the user points one among the cinema posters, the stored preview for the pointed cinema will be played for 5~10 seconds, and after the preview the user can point one of the 9 posters again

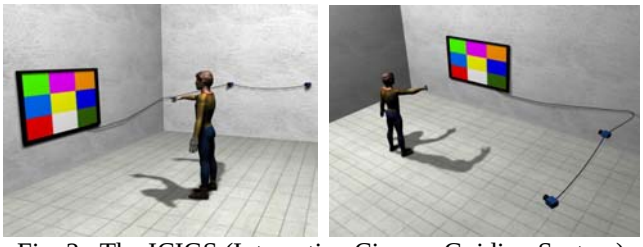


Fig. 2 : The ICIGS (Interactive Cinema Guiding System)

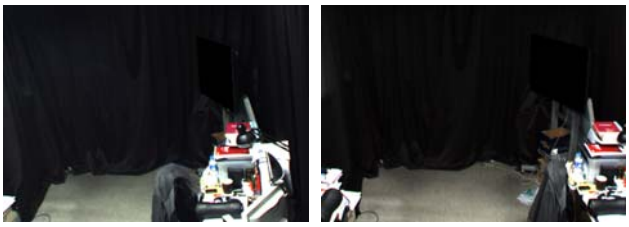


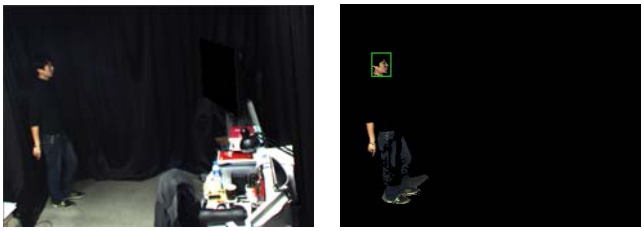
Fig. 3 : The real environment and extraction of interesting region

In addition, the information on the projected side is extracted to compose the plane of the interest area pointed by the user. Fig.3 shows the results of the extraction of the interest area pointed by the user from the two cameras.

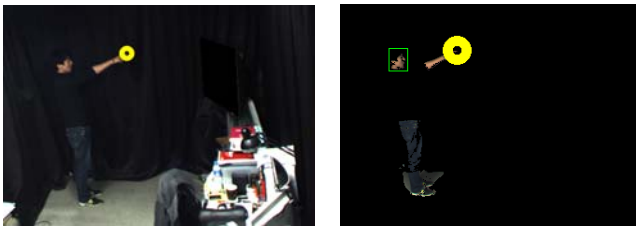
3. Experiment Result

We proposed a system to detect the user's face and hand using the two cameras and to estimate the pointed region by the user in the 3-dimensional space. In addition, it designs an "Interactive Cinema Information Guiding System" to which the proposed algorithm is applied

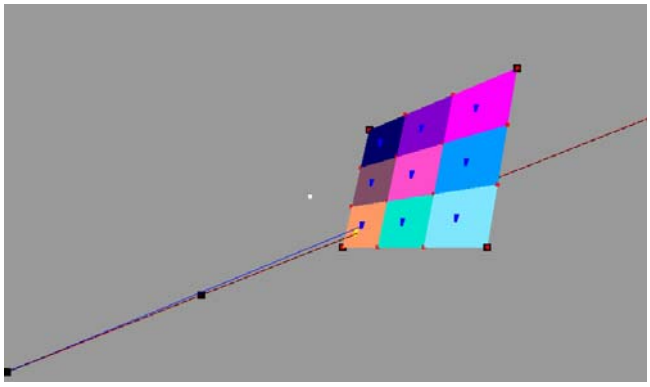
3.1 Pointing Region Estimation



(a) A face detection



(b) A face and a fingertip detection

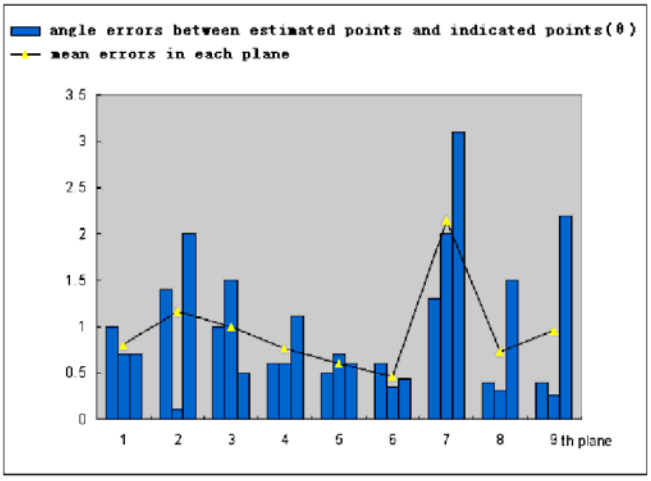


(c) Result of indicated point estimation

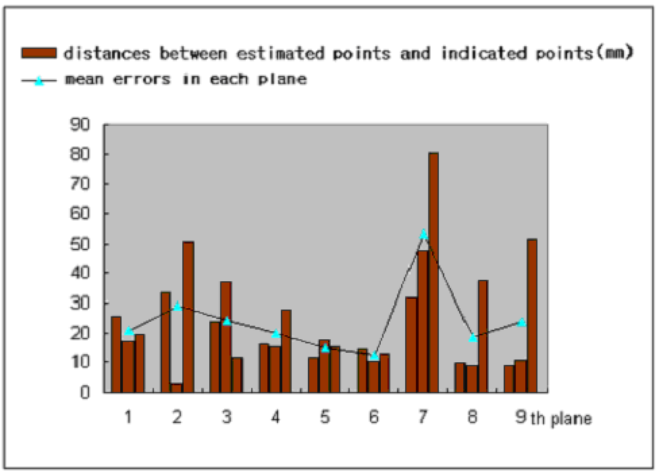
Fig. 4 : Result of pointing region estimation.

Fig.4 shows the experiment to estimate the face and the fingertip of the user. Fig.4 (a) shows the result of face detection by the template matching method mentioned above, Fig.4 (b) shows the result of the fingertip extracted using the skin color, and Fig.4 (c) shows the straight line to point what a user intends and the dotted line estimated after the face and the fingertip of the user are extracted from the images.

To test the accuracy of the pointed region, the centers of the 9 planes were pointed three times dividing the approximately 1020mm × 575mm real-world plane into 3×3. Fig.5 (a) and (b) show the angle errors and distance errors from the indicating points estimated from the centers of the planes. The average distance error from the center of the plane was measured as 24.12mm, and the average angle error is measured as 0.97 ° in the restricted experimental condition. Because the size of a plane is set as approximately 340mm × 191mm, the measurement error is regarded to be within the permissible range.



(a)



(b)

Fig. 5 : Errors between estimated point and center point of the plane (Unit : mm)

3.2 Implementation of Interactive Cinema Information Guiding System

Fig.6 shows the proposed ICIGS (Interactive Cinema Information Guiding System). After pointing region of user is estimated, information of the movie is being provided in the pointing region.

Fig. 6 : ICIGS (Interactive Cinema Information Guiding System)

4. Conclusion

In the experiment, the reliability is verified by pointing the cinema information presented in the 9 planes and by repeating re-pointing at the preview-ending point. In addition, the possibility for practical use of the proposed algorithm is verified by composing “Interactive Cinema Information Guiding System” consisting of two cameras and a monitor for the application of the proposed system. The future study will be extended to the topics of recognition of multiple persons and efficient camera arrangement to make it a more practically usable system.

Acknowledgment

This work was financially supported in part by a grant from Seoul R&BD Program(10570), Seoul R&BD Program(TR080601) and the Ministry of Education and Human Resources Development (MOE) under the second stage of BK21 program.

5. REFERENCES

[1] A. Pentland, "Looking at people: Sensing for Ubiquitous and Wearable Computing", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.22, No.1, pp.107-119, 2000.

[2] R. Cipolla, P. A. Hadfiels and N. J. Hollinghurst, "Uncalibrated stereo vision with pointing for man-machine interface", Proc. of IAPR workshop on Machine Vision Application, pp.163-166, 1994.

[3] N. Jojic, B. Brumitt, B. Meyers, S. Harris, "Detection and Estimation od Pointing Gestures in Dense Disparity Maps", Proc. of International Conference on Automatic Face and Gesture Recognition, pp.468-475, 2000

[4] H. Watanabe, H. Hongo, K. Yamamoto, "Estimation of Omni-Directional Pointing Gestures using Multiple Camera", IEE of Japan, Vol.121-C, No.9, pp.1388-1394, 2001.

[5] T. Yamaguchi, E. Sato, "Humantronics and RT-Middleware", Advances in Information Processing and Protection, Springer, 2007.

[6] Z. Zhang, "A flexible new technique for camera calibration", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 22, No. 11, 1330-1334, 2000.

[7] Sung-Eun Kim, Chang-Joon Park, In-ho Lee, "A Tracking Method of End-effectors in a Vision-based Marker-free Motion Capture System", IEEE Conf. on Cybermatics and Intelligent Systems, Vol. 1, 129-134, 2004.

[8] J. Yang, W. Lu, A. Waibel, "Skin Color Modeling and Adaptation", (CMU-CS-97-146), CS Department, CMU, PA, U.S.A., 1997.

[9] S. Kr. Singh, D. S. Chauhan, M. Vatsa, R. Singh, " “A Robust Skin Color Based Face Detection Algorithm, Tamkang Journal of Science and Engineering", Vol. 6, No. 4, pp. 227-234, 2003.

[10] R. Hartley, A. Zisserman, "Multiple View Geometry", Cambridge Univ. Press, 2000.

[11] Y. S. Han, Y. H. Seo, K. S. Doo, J. S. Choi, "The Real World Pointing Region Estimation System using 3D Geometry Information", in Proc. ITC-CSCC2007, Vol. 3, pp. 1111-1112, 2007.