2000년도 대한전기학회 하계학술대회 논문집 2000.7.17-20

컴퓨터 인터페이스를 위한 Hand Gesture 인식에 관한 연구

장호정, 백한욱, 정진현
광운대학교 제어계측공학과

A Study of Hand Gesture Recognition for Human Computer Interface

(Ho Jung Chang , Han Wook Baek , Chin Hyun Chung)
Dept. of Control & Instrumentation Engineering, Kwangwoon Univ.

Abstract - GUI(graphical user interface) has been the dominant platform for HCI(human computer interaction). The GUI-based style of interaction has made computers simpler and easier to use. However GUI will not easily support the range of interaction necessary to meet users' needs that are natural, intuitive, and adaptive. In this paper we study an approach to track a hand in an image sequence and recognize it, in each video frame for replacing the mouse as a pointing device to virtual reality. An algorithm for real time processing is proposed by estimating of the position of the hand and segmentation, considering the orientation of motion and color distribution of hand region.

1. Introduction

Hand gesture recognition from video images is one of considerable interest as a means of providing simple and intuitive human-computer interfaces. Recently as the new wave of portable computing, the era of pen computer without keyboard and the core technology of pen computing consists of hand-written character and gesture recognition. However, most of previous gesture-based applications need special hardwares such as a mouse, stylus pen, or data glove in order to input a gesture. Although current gesture input devices are well improved, it is more or less unnatural to make gestures using special hardwares. Especially for novice users who are afraid to use computer device. so In this paper a method is described to recognize gestures by extracting hand regions and to control window manager from these recognition results. Color image processing using a skin color model and labeling techniques are used to extract hand image. The trajectory of the hand is obtained by connecting hand region extracted from each frame an input image sequence using the modified BMA(Block Matching Algorithm). This trajectory is analyzed to recognized a gesture according to the predefined window management rule, the window manager uses the gesture recognition technique to control the display, tracking and execution of a menu.

2. Hand Region Segmentation Using C Information

The hand region segmentation using color information consists of the construction of model-database and the extraction process of hand region segmentation.

Since Gray coordinate include only the normal information of each pixel, the hand region included in some kind of textural image can be hardly extracted. Thus, in this paper, the hand region is extracted by the color coordinate, which is used YCbCr that can separate luminance and chrominance. YCbCr is needed less coding-bit than RGB with omitting the luminance of each pixel color and is less sensitive to the circumstance of lighting.

2.1.1 Color Analysis

Most color models in use today are oriented either toward hardware (such as for color monitors and printers) or toward applications where color manipulation is a goal (such as in the creation of color graphics for animation). The hardware-oriented models most commonly used in practice are the RGB (red, green, blue) model for color monitors and a broad class of color video cameras; the CMY (cyan, magenta, yellow) model for color printers; and the YIQ model, which is the standard for color TV broadcast. In the third model the Y corresponds to luminance, and I and Q are two chromatic components called inpahse, and quadrature, respectively. Among the models frequently used for color image manipulation are the HSI (hue, saturation, intensity) model and the HSV (hue, saturation, value) model.

The YCbCr color space used in this paper is international standard for digital video signals. This color space decomposes of luminance and chrominance information from a real world image. When we do region segmentation with the color information, the characteristics of this color space can do simply region segmentation and remove luminance factors of surroundings and some other components which disturb the segmentation. So, the usage of YCbCr shows the decrease of error factors and the available preprocessing of the efficient segmentation.

The standard luminance used by the television industry is computed by

$$Y_l = 0.299R + 0.587G + 0.114B$$

The human visual system has less sensitivity toward variations of color than toward variations of luminance. A simple way to obtain a chrominance-luminance decomposition is to calculate the luminance Y, and make this one of the components of the system. Next, one subtracts Y from the other color components obtaining the combinations R-Y, G-Y, and B-Y, which do not carry luminance information

The YCbCr system is the international standard for digital video signals.
It is obtained from the Y, R-Y, B-Y system by means of the following transformation

$$Y = 16 + 235\,Y_l$$

$$C_b = 128 + 112\left(\frac{0.5}{1-0.114}(B - Y_l)\right)$$

$$C_r = 128 + 112\left(\frac{0.5}{1-0.299}(R - Y_l)\right)$$

2.2 Hand Region Segmentation

The hand region segmentation algorithm involves the use of color information in a fast, low-level region segmentation process. The aim is to classify pixels of the input image into skin color and non-skin color. To do so, we have devised a skin-color reference map in YCbCr color space.

We have found that a skin-color region can be identified by the presence of a certain set of chrominance (i.e., Cb and Cr) values narrowly and consistently distributed in the YCbCr color space. With skin-color reference map, the color segmentation can now begin. Since we are utilizing only the color information, the segmentation requires only the chrominance component of the input image.

Consider an input image of $M \times N$ pixels, for which the dimension of Cb and Cr is $\frac{M}{2} \times \frac{N}{2}$. The output of the color segmentation, and hence stage one of the algorithm, is a bitmap of $\frac{M}{2} \times \frac{N}{2}$ size, described as

$$Out = \begin{cases} 1 & \text{if } [\,C_r(x,y) \in [80\ 130]] \cap [\,C_b(x,y) \in [125\ 175]\,] \\ 0 & \text{otherwise} \end{cases}$$

where $x = 0, \ldots \frac{M}{2} - 1$ and $y = 0, \ldots, \frac{N}{2} - 1$.

The output pixel at point (x, y) is classified as skin color and set to one if both the Cb=[80 130] and Cr=[125 175] values at that point fall inside their respective ranges Cb and Cr. Otherwise, the pixel is classified as non-skin color and set to zero.

3. Motion Estimation

The BMA is a popular correlation-based approach to motion estimation and tracking. The modified BMA is used to extract the motion vector wth the previously extracted hand region. Direction and trajectory of hand gesture is derived upon the distribution and direction of the extracted motion vectors. Hence, the motion vector's distribution is statistically processed and applied to the hand gesture recognition.

3.1 Block-Matching Algorithm

Block matching can be considered as the most popular method for practical motion estimation In block matching, the best motion vector estimate is found by a pixel-domain search procedure.

In general, displacements are chosen that either maximize correlation or minimize error between a macroblock and a corresponding array of pel values in the reference picture.

Where the displacement for a pixel (n_1, n_2) in frame k (the present frame) is determined by considering an $N_1 \times N_2$ block centered about (n_1, n_2), and searching frame $k+1$ (the search frame) for the location of the best-matching block of the same size. The search is usually limited to an $(N_1 + 2M_1) \times (N_2 + 2M_2)$ region called the search window for computational reasons.

The matching of the blocks can be quantified according to various criteria including the maximum cross-correlation, the MSE (minimum mean square error) and the MAD (minimum mean absolute distortion) In the minimum MSE criterion, we evaluate the MSE, defined as

Mean Square Error

$$MSE(d_1, d_2) = \frac{1}{N_1 N_2} \sum_{(n_1, n_2) \in B} [\,F_k(n_1, n_2) - F_{k-1}(n_1 + d_1, n_2 + d_2)]^2$$

where B denotes an $N_1 \times N_2$ block, for a set of candidate motion vectors (d_1, d_2). The estimate of the motion vectors is taken to be the value of (d_1, d_2) which minimizes the MSE.

However, the minimum MSE criterion is not commonly used in hardware implementations because it is difficult to realize the square operation in hardware.

Instead, the minimum MAD criterion, defined as

Mean Absolute Distortion

$$MAD(d_1, d_2) = \frac{1}{N_1 N_2} \sum_{(n_1, n_2) \in B} |\,F_k(n_1, n_2) - F_{k-1}(n_1 + d_1, n_2 + d_2)|$$

3.2 Modified BMA

Since the conventional BMA assumes that all pixel's movement are same in block, the assumption is a cause of the blocking effect and the measurement error of movement. Thus, in this paper, in order to compensate the blocking effect and the prediction error of motion vector, the block size is assigned variably. If a block is measured to non-motion or same motion vector, this is assigned big-size region.

4. Experiment and Conclusion

To obtain image, CCTV camera which has 24bit true color and 256×256 resolution is used. The range that we found to be the most suitable for all the input images that we have tested are Cb=[80 130] Cr=[125 175]. The segment result by using the skin-color reference map can be seen in Fig. 2

A hand region segmentation based on the skin-color reference map can acquire accurate results, if there is a good contrast between skin color and those of the background objects. However, if the color

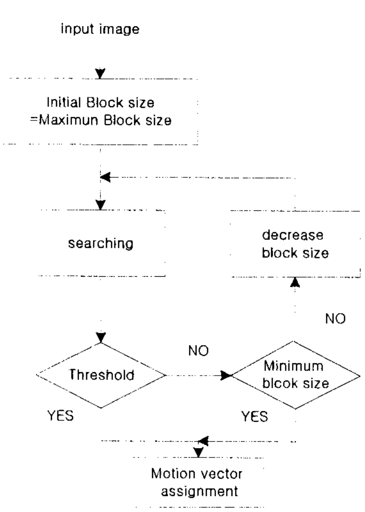

Fig.1. Flowchart of the Modified BMA

characteristic of the backgorund is similar to that of the skin, then the result is of color segmentation is the detection of pixels in a hand region and may also include other areas where the chrominance values coincide with those of the skin color. The noise pixels except valid region can be eliminated by the merging method, which merge each pixel to valid region with a basis of valid object block. No matter what BMA is a popular approach to the motion estimation and tracking, complex or textuaral back-ground disturbs the estimation of motion vector with BMA. Thus, before the BMA, the segmentation based on the color coordinate eliminates invalid region except the hand region. The algorithm proposed in this paper can simplify the color information of image and improves the motion estimation.

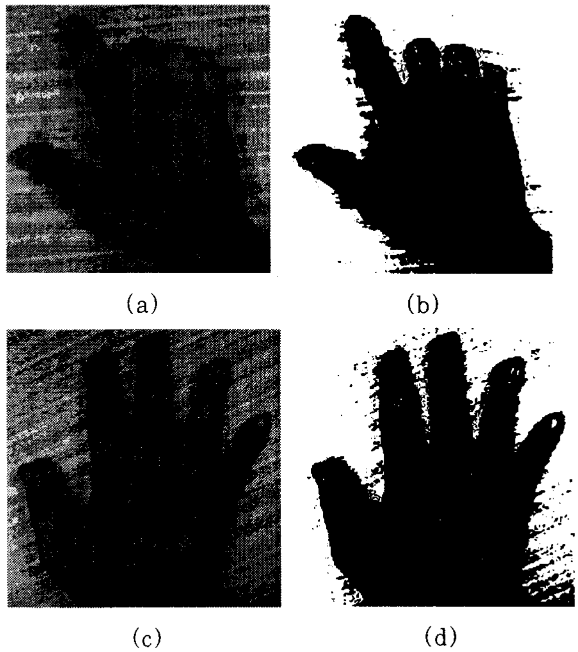

Fig. 2 Results produced by the color-segmentation process which has background compose of texture image.
(a),(c) original image and (b),(d) result of segmentation

[참 고 문 헌]

[1] Rafael C. Gonzalez, “Digital Image Processing”, Addison Wesley Publishing Company, 1993
[2] J. R. Parker "Algorithm for Image Processing and Computer Vision" , Wiley Computer Publishing, 1996
[3] Randy Crane, "A simplified approach to Image Processing" Prentice hall,1997
[4] Jonas Gomes, Luiz Velho, "Image Processing for Computer Graphics", Springer, 1997
[5] Phillip E. Mattison, "Practial Digital VIDEO with Programming Examples In c", Johj Wiley & Sons, Inc. 1994
[6] Won Hee L. "A Study on Segmentation of Video Sequence by Modified BMA" 1997
[7] Jie Yang, Weier Lu, Alex Waibel, "Skin-Color modeling and Adaptation", Pittsburgh, PA 15213,
[8] Richard P. Schumeyer, Kenneth E. Barner, "A Color-Based Classifier for Region Identification in Video", SPIE Vol. 3309
[9] A.M Peacock, S. Matsunaga, D. Renshaw. J. Hannah and A.Murray, "Reference block updating when tracking with block matching algorithm", Electronics Letters 17th February 2000, Vol. 36
[10] Jie-Bin Xu, Lai-Man P, Chok-Kwan Cheung ,"Adaptive Motion Tracking Block Matching Algorithm for Video Coding", IEEE Transactions on Circuit and systems for video Technology, Vol 9