비젼 기반의 포인팅 기기를 위한 퍼지 스크린 검출기

논 문
58P-3-10

Fuzzy Screen Detector for a Vision Based Pointing Device

고 재 원†
(Jae-Won Kho)

Abstract - In this paper, we propose advanced screen detector as a tool for selecting the object for tracking and estimating its distance from a screen using fuzzy logic in vision based pointing device. Our system classifies the line component of the input image into horizontal and vertical lines and applies the fuzzy rule to obtain the best line pair which constitute peripheral framework of the screen. The proposed system improves the detection ratio for detecting the screen in relative to the detector used in the previous works for hand-held type vision based pointing device. Also it allows to detect the screen even though a small part of it may be hidden behind other object.

Key Words : Fuzzy Logic, Screen Detector, Computer Vision, Pointing Device

1. Introduction

As interactive multimedia have come into wide use, user interface such as remote controllers or classical computer mice have several limitations that cause inconvenience. Classical computer mice are available only on the supporting surfaces and can perceive only two-dimensional movement. For this reason, various type of pointing devices which can detect 6 DOF-movement in free space, have been developed[1-8]. Examples include hand gesture recognition systems[1-3], tilt-sensing tools[4], magnetic tracker[5], joystick-based device, gyro mouse[6], soap[7] and cubic mouse[8]. With a remote pointing device users can control the movement of a mouse pointer on the screen at an appreciable distance in mid-air. However, devices for remote pointing have performed poorly, inconvenient to use, cost too high, and show low performance or counter-intuitive manipulation in relative to existing computer mice. And there is no general model yet like existing optical or ball mouse. Therefore, our previous study have focused on obtaining vision based hand-held typed pointing device with higher DOF motion sensing suitable for more general purpose while also being less expensive than other devices based on computer vision. Vision-based pointing devices can be divided into two types: a fix positioned camera type and a hand-held type. The former devices track human motion or object manipulation in front of a fixed camera and the results are used to guide the mouse pointer's motion. In contrast, the latter devices have one or more cameras in the handheld device to detect the device motion itself like a traditional mouse. A camera imbedded hand-held type of pointing devices have its own convenience in relative to the fix positioned camera type of pointing devices which recognize human gestures. However, few studies have been reported concerning those type of devices on account of difficulties about object tracking and pose recognition. Accurate recognition of an object for tracking is critical component affecting their performance. In the previous study, we developed a vision based pointing device using a small camera and proposed a method to detect a screen and to track its corner points to estimate the movement of this device and the user's distance from the screen, which enable the control of a cursor on the computer display[9]. But, the previous method showed the limitation in the detection of the screen when a part of it might be hidden behind other object such as decoration on a monitor and a small cup in front of the screen, because the system detects the overall frame shape of the screen.

For practical use of the handheld pointing device based on screen detection, it should be more flexible in the various situations include that a color of the monitor frame is similar to the background or screen image's color, and as well as that, monitor frame is hidden as mentioned above. To cope with these situations and to increase detection ratio of the screen, we extracted the

† 교신저자 정회원 : 유한대학 컴퓨터제어과 교수·공박
E-mail : jwkho@yuhan.ac.kr
접수일자 : 2009년 7월 8일
최종완료 : 2009년 8월 17일

line component from input images using hough transform[10] and then the fuzzy inference process[11] was provided to select lines that constitute the frame of the screen from the horizontal and vertical straight lines. The use of fuzzy logic has benefits concerning the flexible detection of the screen. The linguistic parameters considered as characteristic of the screen's edge line can make available the disturbed information resources by assignment of weight values of each parameters.

While the boolean method, at times, misdetection of small part of screen by only small disturbance by other object or noise might result in misdetection for a whole screen image.

The remaining part of the paper is organized as follows: Section 2 describes the process to get vertical-horizontal pairs of line components, which are used as candidates for the frame of screen in an input image. Section 3 presents the fuzzy model to estimate the similarity to predefined characteristics of the screen and also to estimate the suitability for tracking. Section 4 shows the experimental results using proposed fuzzy detector algorithm applied to sample screen image. Concluding remarks are included in section 5.

2. Candidate Line Pairs for the Screen

In this section, we describe how to get line components in an input image and to pair the two vertical lines with the two appropriate horizontal lines to meet the needs of minimum requirement to form a rectangular shape.

Hough transform techniques [10] have been applied conventionally as the way of recognizing the straight lines in the image because of its robustness of handling incomplete image data and high degree of discrimination ability. Surveys of the Hough transform for binary images and applications appear in [12,13]. The basic theory is on a voting process where every edge points of a cluster vote for all possible curves passing through those points.

Points in the $x-y$ plane is mapped to sinusoidal curve in the $\rho-\theta$ plane by the equation (1)

$$\rho = x\cos\theta + y\sin\theta \qquad (1)$$

Then, the line that passes through the two edge points in the $x-y$ plane corresponds to the intersection of two sinusoidal curves in the $\rho-\theta$ plane. An accumulator array is used to count the number of intersection in the $\rho-\theta$ plane. Then it selects the most accumulated point corresponding to the line in the $x-y$ plane. In order to apply the Hough transform, the Canny edge detection algorithm [14] was used to find out the edges of the image.

The extracted segments are classified into vertical line(VL), horizontal line(HL), and diagonal line(DL) using equation (2).

$$\theta_i = \frac{2}{\pi}\tan^{-1}\left(\frac{|y_{i+1}-y_i|}{|x_{i+1}-x_i|}\right),\ \begin{cases} J(\theta)_i = VL & 0.9 \le \theta_i \\ J(\theta)_i = DL & 0.1 < \theta_i \le 0.9 \\ J(\theta)_i = HL & \theta \le 0.1 \end{cases} \qquad (2)$$

The line between i^{th} and $i+1^{th}$ points are categorized based on its angle in respect to the x axis of the image. Figure 1 shows a sample image containing a screen and its classification results. Among the three types of lines, VL and HL groups are selected for the next process.

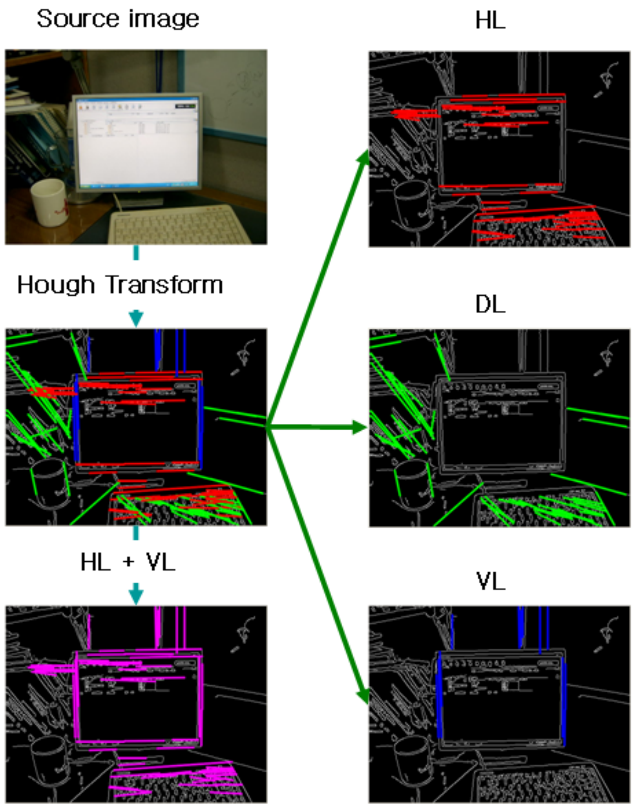

Fig. 1 All the line segments extracted by the hough transform are classified into HL, DL, and VL groups (right side of the figure). The HL and DL groups are selected for the following process steps (left side of the figure).

The procedure for the selecting the candidate pairs of vertical-horizontal lines to be scored by fuzzy logic in the next section is built based on the algorithm described below:

STEP 1 : Select a leftmost vertical line which is designated in Eq.(3) as a first reference line.

$$VL_i = \{(x_1^i, y_1^i), (x_2^i, y_2^i)\} \qquad (3)$$

STEP 2 : Select a uppermost horizontal line HL_j satisfying the following condition.

$$HL_j = \{(x_1^j, y_1^j), (x_2^j, y_2^j) \mid (x_1^j \geq x_1^i) \wedge (y_1^j \geq y_1^i)\} \qquad (4)$$

STEP 3 : Select a next leftmost vertical line VL_{i+1} VLi+1 satisfying the following condition.

$$VL_{i+1} = \{(x_1^{i+1}, y_1^{i+1}), (x_2^{i+1}, y_2^{i+1}) \mid (x_1^{i+1} \geq x_2^j) \wedge (y_1^{i+1} \geq y_2^j)\} \qquad (5)$$

STEP 4 : Select a next uppermost horizontal line HL_{i+1} satisfying the following condition.

$$HL_{i+1} = ((x_1^{j+1}, y_1^{j+1}), (x_2^{j+1}, y_2^{j+1}) \mid (x_1^{j+1} \geq x_2^i) \wedge (y_1^{j+1} \geq y_2^i) \wedge (x_2^{j+1} \geq x_2^{i+1}) \wedge (y_2^{j+1} \geq y_2^{i+1})) \qquad (6)$$

STEP 5 : Add the vector { VL_i, HL_j, VL_{i+1}, HL_{i+1}} to the table of candidate pairs as a row.

STEP 6 : Repeat the STEP 1 to 5 for all the possible combination of vertical and horizontal lines in the image.

The results of line pairs in respect to the sample image are shown in Fig.3. We can verify that some lines that could not pair with other lines are removed in the table of candidate pairs.

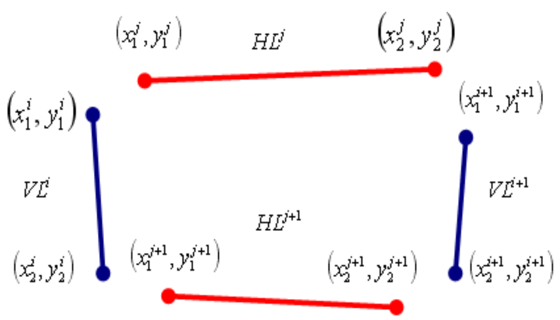

Fig. 2 Vertical and horizontal line pairs

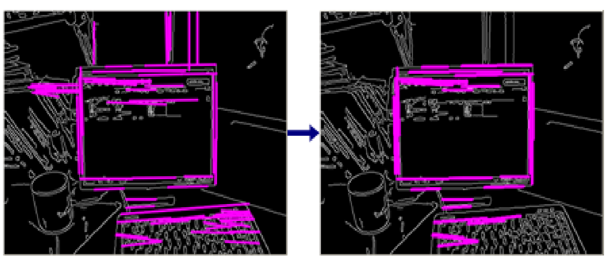

Fig. 3 (Left) All of the vertical and horizontal lines, (Right) Lines in the table of candidate pairs

3. Fuzzy Logic to Evaluate the Matching Degree

As it is shown on the right side of Fig.3, usually the line pairs that constitute the frame of the screen have shorter gap distance between vertical and horizontal lines than the other pairs and have a regular aspect ratio (4:3 or 16:9). The position of the screen in the image is also an important factor, because enough area should be provided for feasible tracking. Therefore, we have defined three parameters to characterize these features.

First, the average of gap distances between neighboring segments in the i^{th} line pair is defined as follows:

$$GDist_i = \frac{\frac{1}{4}\sum_{m=0}^{1}\sum_{n=0}^{1} Line\_dist(VL^{i+m}, HL^{j+n})}{\max\{Line\_dist(VL^{i+m}, HL^{j+n}) \mid \forall i,j \in N, \forall m,n \in [0,1]\}} \qquad (7)$$

Where, N is the total number of line pairs, and $Line\_dist$ function calculates the minimum distance between two lines, for example, $Line\_dist(VL_i, HL_j)$ can be calculated as follows :

$$Line\_dist(VL_i, HL_j) = \sqrt{(x_1^i - x_1^j)^2 + (y_1^i - y_1^j)^2} \qquad (8)$$

$GDist_i$ is normalized to the range [0,1] by dividing its value with the largest gap distance in all line pairs within an image.

The second parameter is the distance between the center of the i^{th} line pair and center of the image, which is defined as follows:

$$CPose_i = \frac{Point\_dist(CI, CL_i)}{Point\_dist(CI, origin)} \qquad (9)$$

Where, CI is the center point of the image which can be estimated from the resolution of the given image, CL_i is the center of the i^{th} line pair which is consisted with 4 lines, and the function $Point\_dist$ calculates the distance between two points. CL_i can be estimated as follows:

$$CL_i = \left(\frac{x_1^i + x_2^{i+1}}{2}, \frac{y_1^j + y_2^{j+1}}{2}\right) \qquad (10)$$

The parameter $CPose_i$ is also normalized to the range [0,1] by dividing its value with the distance between the center and origin of the image.

Lastly, the third parameter, aspect ratio of the i^{th} line pair is defined by Eq.(11).

$$ARatio_i = \left| \frac{x_2^{i+1} - x_1^i}{y_2^{j+1} - y_1^i} \right| \quad (11)$$

In order to increase the flexibility and reliability of the detection process in this model we used the fuzzy logic. Each parameter described above is too vague and uncertain to directly constitute a clear evidence for detecting process. For example, the shape of the screen in the input image is not uniform caused by perspective deformation which depends on the position of the camera. For this reason, the aspect ratio should be expressed as how close its value is to the format of the normal screen and be considered with other parameters to detect the screen in the image. In our fuzzy logic system, we used the above three parameters as the factors to evaluate the matching degree by fuzzification. We adopted a Mamdani-type model[15] thanks to its simplicity. The linguistic input and output variables and their associated terms used by the fuzzy detector are shown in table 1.

Table 1 Linguistic variables and terms

Variable	Type	Range	Terms
$GDist_i$	Input	[0,1]	Close, Mod, Far
$CPose_i$	Input	[0,1]	RC,NC,NB
$ARatio_i$	Input	[0,1.5]	WW,W,WN,N,NN
$Match_i$	Output	[0,100]	VH,H,M,L,VL

Legend: Mod=Moderate, RC = Right in the Center, NC = Near in the Center, NB = Near in the Boundary, WW = Wider than Wide monitor frame, W = Wide monitor frame, WN = Wider than narrow monitor and Narrower than wide monitor frame, N = Narrow monitor frame, NN = Narrower than Narrow monitor frame, VH = Very High, H = High, M = Medium, L = Low, and VL = Very Low.

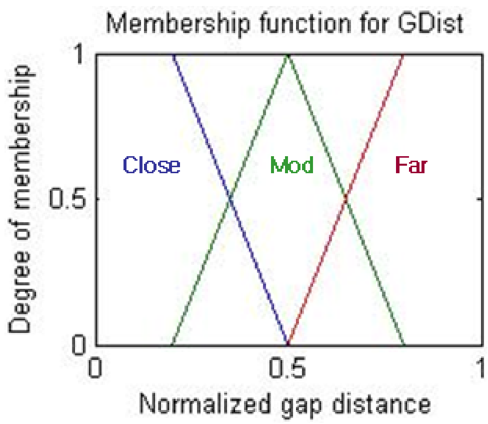

Fig. 4 Fuzzification of the linguistic variable $GDist_i$

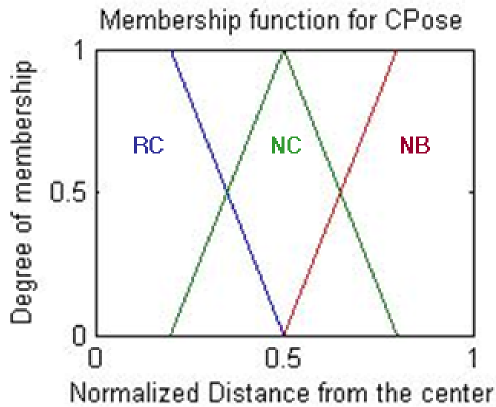

Fig. 5 Fuzzification of the linguistic variable $CPose_i$

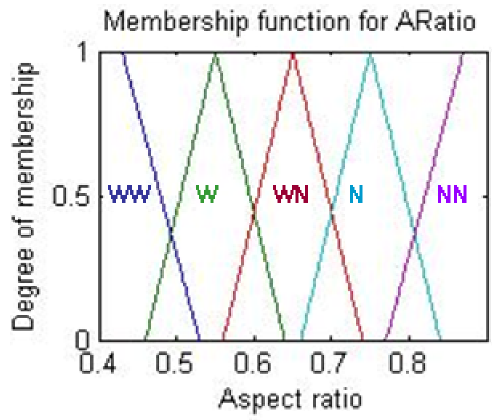

Fig. 6 Fuzzification of the linguistic variable $ARatio_i$

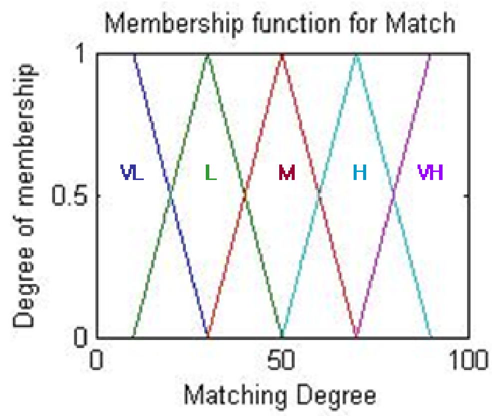

Fig. 7 Fuzzification of the linguistic variable $Match_i$

The values of a fuzzified input execute all the rules in the fuzzy reasoning system, which includes rules like the following:

IF the $GDist_i$ is Close and $CPose_i$ is RC and $ARatio_i$ is WW THEN the $Match_i$ is M.

The rule bases used for our model are shown in Table

2. Fuzzy rules are generated by a heuristic analysis of the relationships of the linguistic variables and desired outcome to represent the matching degree.

For every line pair, each of the fuzzy if-then rules is evaluated. In the deffuzzification procedure, the center of area method is used to convert the output of all rules into a crisp value. Each fuzzy output is multiplied by its corresponding singleton value, then the sum of these products is divided by the sum of all the fuzzy outputs to obtain the final single output variable as is indicated in the following formula:

$$fuzzy\_output = \frac{\sum_{U} \mu_i(w_i) \cdot w_i}{\sum_{U} \mu_i(w_i)} \quad (12)$$

Where, U represents number of all output values which are considered, w_i is the given values of the terms with the corresponding membership function μ_i.

Table 2 Fuzzy rule base

GDist	*CPose*	*ARatio*	*Match*	*GDist*	*CPose*	*ARatio*	*Match*
Close	RC	WW	M	Mod	NC	N	M
Close	RC	W	VH	Mod	NC	NN	L
Close	RC	WN	M	Mod	NB	WW	L
Close	RC	N	VH	Mod	NB	W	L
Close	RC	NN	M	Mod	NB	WN	L
Close	NC	WW	M	Mod	NB	N	L
Close	NC	W	H	Mod	NB	NN	L
Close	NC	WN	M	Far	RC	WW	VL
Close	NC	N	H	Far	RC	W	L
Close	NC	NN	M	Far	RC	WN	L
Close	NB	WW	L	Far	RC	N	L
Close	NB	W	M	Far	RC	NN	VL
Close	NB	WN	M	Far	NC	WW	VL
Close	NB	N	M	Far	NC	W	L
Close	NB	NN	L	Far	NC	WN	L
Mod	RC	WW	M	Far	NC	N	L
Mod	RC	W	H	Far	NC	NN	VL
Mod	RC	WN	M	Far	NB	WW	VL
Mod	RC	N	H	Far	NB	W	VL
Mod	RC	NN	L	Far	NB	WN	VL
Mod	NC	WW	L	Far	NB	N	VL
Mod	NC	W	M	Far	NB	NN	VL
Mod	NC	WN	M				

4. Experimental Results

The defuzzified output value computed by Eq.(12) represent the matching degree of each line pair. Randomly selected eight samples among 505 line pairs and their defuzzified output values are shown in the Fig.8. All the matching degrees of each line pair are compared and then, a line pair which has highest value is selected as the best candidate, which constitutes the frame of a screen. In our experimental results, the pair number 337 (Fig.8-e) can be selected as the best configuration of line pairs.

As shown in Fig.8, we can see that the configuration of a line pair more coincide with a frame of a screen when it has the higher value of matching degree.

(a) pair number 1 (61.811)

(b) pair number 82(23.627)

(c) pair number 103(28.226)

(d) pair number 259(21.448)

(e) pair number 337(68.987)

(f) pair number 343(34.055)

(g) pair number 419(68.273)

(h) pair number 505(16.648)

Fig. 8 Selected sample of line pairs: The values in the parenthesis represent the defuzzified matching degree of each line pair.

5. Conclusion

One of the major problem that troubled in the hand-held type of vision based pointing devices is detecting the target used to decide its position and reference for tracking. We improved the performance of these type of pointing device by detecting the screen. To increase the detection ratio of the screen in the variable environmental condition, we extracted the vertical and horizontal line components in the image which can be obtained in the camera imbedded in the pointing device, and then we configured the rectangular formed line pair, which used as candidate for frame of screen. In the process of selecting the best line pair, a fuzzy logic system which has four linguistic variables and 45 fuzzy rules, was used. Experimental results show that proposed fuzzy detector system accurately evaluates the matching degree of each line pair according to the coincidence level.

References

[1] Y.Wu and T.S.Huang, “Hand Modeling Analysis, and Recognition for Vision-Based Human Computer Interaction,” IEEE Signal Processing Magazine, May 2001, pp. 51-60.

[2] A.Wilson and E.Cutrell, “FlowMouse. A Computer Vision-Based Pointing and Gesture Input Device,” Human-Computer Interaction INTERACT '05, (Rome, September2005), Springer(forIFIP), pp.565-578.

[3] P.Nesi and A.D.Bimbo : “A vision-based 3-D mouse”, International Journal of Human-Computer Studies,” Volume 44, Number 1, pp. 73-92, January 1996.

[4] J.ReKimoto, “Tilting Operations for Small Screen Interfaces,” User Interface and Software Technology (UIST'96), pp.167-168, 1996.

[5] V. Kindratenko, and A. Bennett, “Evaluation of Rotation Correction Techniques for Electromagnetic Position Tracking Systems,” Proc. of the 6th Eurographics Workshop on Virtual Environments, pp.13-22, Amsterdam, The Netherlands, June 2000.

[6] I. S. MacKenzie and S. Jusoh, “An Evaluation of Two Input Devices for Remote Pointing,” Proc. of the Eighth IFIP International Conference on EHCI 2001, pp. 235-249, Heidelberg, Germany.

[7] P. Baudisch, M. Sinclair and A. Wilson, “Soap: a Pointing Device that Works in Mid-Air,” UIST'06, October 15 - 18, 2006, Montreux, Switzerland.

[8] B. Fröhlich and J. Plate, “The Cubic Mouse. A New Device for Three-Dimensional Input,” In Proc. of ACM CHI 2000, pp 526-531, April 2000.

[9] S. Jang, M. Cheon and M. Park, “A Vision-Based Approach for Remote Pointing Devices,” ISIS 2007, 8th International Symposium on Advanced Intelligent Systems, Sokcho-City, Korea, September 5 - 8, 2007.

[10] P.V.C. Hough, “Method and means for recognizing complex patterns,” U.S. Patent 306965, 1962.

[11] M.Sugeno, Industrial Applications of Fuzzy Control, Elsevier Science Pub. Co.,1985.

[12] J. Illingworth and J. Kittler, “A survey of the Hough transform,” Computer Vision, Graphics, and Image Processing, 44, pp.87-116, 1988.

[13] V. F. Leavers, “Which Hough transform?,” CVGIP: Image Understanding, vol.58, no.2, pp.250-264, September 1993.

[14] J. Canny, “A computational approach to edge detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.8, no.6, pp. 679-698, Nov. 1986.

[15] E. Mamdani and S. Assilian, “An Experiment in linguistic synthesis with a fuzzy logic controller,” International Journal of Man-machine studies, 7(1), pp.1-13, 1975.

저 자 소 개

고 재 원 (高 在 源)

1957년 2월 21일생. 1981년 연세대 공대 전기공학과 졸업. 1983년 동 대학원 전기공학과 졸업(석사). 1992년 동 대학원 전기공학과 졸업(공박). 1986~1992 삼성전자 자동화연구소 선임연구원. 1993~1995 고등기술연구원 생산기술연구실 책임연구원. 1996~현재 유한대학 컴퓨터제어과 교수

Tel : 02-2610-0833

Fax : 02-2682-6080

E-mail : jwkho@yuhan.ac.kr