Text Input System Using Hand Shape Recognition

Ahn,Yang-Keun
Realistic Media Platform Research Center
Korea Electronics Technology Institute
Seoul, Republic of Korea
e-mail: ykahn@keti.re.kr

Jung,Kwang-Mo
Realistic Media Platform Research Center
Korea Electronics Technology Institute
Seoul, Republic of Korea
e-mail: jungkm@keti.re.kr

***Abstract*— This paper presents a method to recognize Korean language input using hand information extracted from 3D information taken from a single camera in a smart device environment. The presented method uses the shape information of an image of the hand as input for Korean mobile phone keyboards. Depth information is used to extract the region of the hand in real time. Through preprocessing, noise is removed from the extracted region, and the hand image is normalized with respect to its size and movement. The normalized image of the hand is projected using the rotation invariant Zernike basis function to ascertain its moment. These moment values of the hand's shape are compared to values saved in a database, and the character that has the closest moment value is input. The proposed method, which is designed to overcome the obstacle that people have unique hand shapes, makes use of a system that can easily become familiarized with an individual's hand shape information.**

Keywords-Hand Gesture; Gesture Recognition; Text Input System; Hand Shape; Shape Recognition.

I. INTRODUCTION

Electronic devices have become a staple of everyday modern life. While such devices have been growing smaller, their performance, an indicator of the progress made by the electronics industry, has been improving continuously. However, input methods for electronics are still restricted to traditional methods such as the mouse and keyboard. Recent advances have introduced the advent of touch control methods, but it is still difficult to control devices from a distance. Remote controls have offered some solutions to the issue of long-distance control, but even this method has the disadvantage of requiring hardware to be carried by the user.

In efforts to solve these types of issues, human computer interaction (HCI) research has been steadily progressing [1]. Within the field of HCI, hand movements that can simply and conveniently express a wide range of information are most commonly used [2]. Such hand movements are capable of controlling electronic devices, and may be applied to the fields of user interfaces (UI) and user experience (UX).

The Shape Key Input Interface (KII) presented in this paper is a system that recognizes the shape of the hand, associating it to a character in Korean to be used as input. Key factors in object recognition such as scale, translation, and rotation are normalized to achieve permanence in scale and translation.

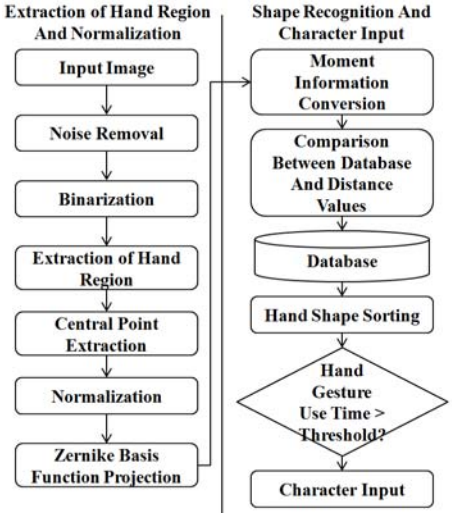

Figure 1. System Control Algorithm

Permanence in rotation is reached through application of a Zernike Shape Descriptor [4]. Rather than touching a designated space within a 3-dimensional space, recognizing the hand shape that corresponds to a keypad button makes input simpler. Additionally, a relatively small number of hand shapes are used, consequently allowing users to provide input without having to look at a keypad. To overcome the issue of different users having unique hand shapes, a system that can easily familiarize itself with users' hands and be customized according to users' tastes was implemented.

The system proposed in this paper is composed largely of two parts, a region extraction and normalization part and a shape recognition and character input part. The algorithm of this system is shown in Figure 1.

II. ZERNIKE MOMENT

Zernike moment is defined as a complex othogonal moment with an absolute value that is rotation invariant. The Zernike moment can be considered as a projection of the basis function, and the basis function of the n-th order at its m-th repetition is defined as follow:

$$V_{nm}(x,y) = V_{nm}(\rho,\theta) = R_{nm}(\rho)\exp(jm\theta) \quad (1)$$

 ISBN: 978-1-61208-468-8

V_{nm} is a set of equations that are normal to a unit circle in polar coordinates. Hence, there is no repetition of information. n is 0 or a positive integer, and m is a number that is even when $n-|m|$, and a non-negative integer when $|m| \leq n$. ρ is the distance from the origin to the point (x,y), and has a domain of $0 \leq \rho \leq 1$. θ is the angle that the point (x,y) creates with the x axis and has a domain of $0 \leq \theta \leq 2\pi$.

$R_{nm}(\rho)$ is a radial polynomial of the Zernike moment, and is expressed as given in equation (2).

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n-|m|}{2}-s\right)!} \rho^{n-2s} \quad (2)$$

A Zernike moment of order n and repetition m Z_{nm} is expressed as in equation (3).

$$Z_{nm} = \frac{n+1}{\pi} \iint_{x^2+y^2 \leq 1} f(x,y) V_{nm}^*(x,y) dxdy \quad (3)$$

where V* represents the complex conjugate. To solve the Zernike moment for a discrete image, (3) can be approximated as shown in (4).

$$Z_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x,y) V^*(x,y), x^2+y^2 \leq 1 \quad (4)$$

A. *Rotation invariant properties of Zernike moment*

The rotation invariant properties of the Zernike moment can be derived as follows. When an image $f(x,y)$ is converted into polar coordinates $f(\rho,\theta)$, an image rotated by an amount α is defined as given in equation (5).

$$f^r(\rho,\theta) = f(\rho,\theta+\alpha) \quad (5)$$

Applying equation (5) to equation (3) results in equation (6).

$$Z_{nm}^r = Z_{nm} \exp(jm\alpha) \quad (6)$$

Equation (6), shown above, is valid, and as such, it can be deduced that a rotated image affects only the topological values. Expressing this results in equation (7).

$$|Z_{nm}^r| = |Z_{nm}| \quad (7)$$

Hence, using the Zernike moment's absolute value as a feature value reveals its rotation invariant nature [5].

III. HAND SHAPE RECOGNITION INPUT METHODOLOGY

This section deals with using the Zernike moment to carry out hand shape recognition, and then translating information of the hand shapes to character input.

A. *Hand separation method*

As shown in Figure 2(a), the closest objects in the depth image are selected as potential candidates that may be the hand. This step resolves the issue of blindly selecting the closest object as the hand. Once the candidates are selected, the image is binarized (see Figure 2(b), and then labeling is performed to detect blobs, as seen in Figure 2(c). Once the blobs are found, the largest blob is selected as the hand Figure 2(d). Figure 2 illustrates the above process.

After hand area recognition, in order to perform Zernike basis function projection, the size and movement must be normalized. To normalize the hand region into even sizes, the hand's center of mass must be determined. To determine the center of mass, a distance transform is used. This changes the image such that the brightest pixel becomes the center of mass Figure 3(a). Once the center of mass is found, using it as a center, an inscribed circle is drawn. The inscribed circle can be found by taking the image of the hand region and transforming it into a contour image Figure 3(b), after which the minimum and maximum distances from the contours to the center of mass are calculated. The minimum distance is used to find the inscribed circle, and all pixel information below the edge of the circle is deleted as a means of removing the wrist area, leaving only an image of the hand Figure 3(c). Conversely, using the maximum distance, a circumscribed circle around the hand can be drawn, and then the maximum distance value is normalized to a predefined size in order to ensure invariance in size and movement Figure 3(d).

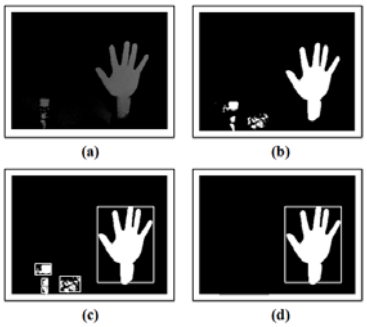

Figure 2. Hand separation process (a) Depth image (b) Binarization (c) Labeling (d) Largest blob detection

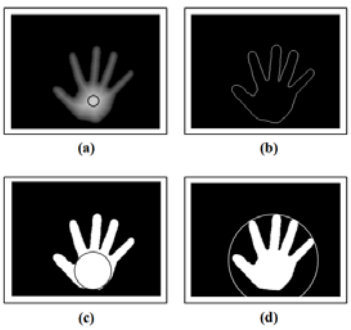

Figure 3. Normalization (a) Distance conversion (b) Contour image (c) Inscribed circle (d) Circumscribed circle

B. Hand image feature value extraction

The absolute value of the Zernike shape descriptor is rotation invariant, making strong recognition of the size, movement, and rotation of the hand images Figure 4(a) possible. In the previous section, the Zernike basis function, combined with the normalized image Figure 4(b) was projected to find the moment value Figure 4(c), leading to the determination of the image's feature values. To further describe the process, Figure 4 is provided below.

C. Recognition enhancement through labeling

Using only the numerical moment values calculated in the previous section, it is possible to proceed with matching and still retain a high recognition rate. However, during the comparison stage when the distance values are compared, outliers may lead to similar or completely different images being detected. This is especially true when there are numerous class variables, which is also when sorting performance is at its lowest. To offset such issues, a labeling technique is used to detect the number of fingers in the image. This particular metric is chosen as the number of fingers on a hand excellently reflects the geometric properties of the hand. However, the discernible types of shape information are limited, and only a maximum of five may be detected.

Comparable classes are largely divided into 5 groups and matched in order to minimize the possibility of error. The use of labeling to detect the number of fingers is shown below in Figure 5. Using labeling to find the number of fingers is simple. All pixel data within a circle with a certain radius extending from the center of mass are deleted. Figure 5(a) is the original image, and Figure 5(b) is the image of the circle used for blob detection. The number of blobs indicate the number of fingers, which in Figure 5(c) is 5.

Images with only one finger detected can the compared with other images with the same number of fingers to improve recognition rates and lower the number of match attempts required to sort images.

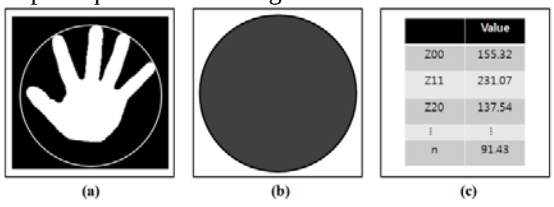

Figure 4. Feature value extraction process (a) Normalized image (b) Zernike basis function (c) Moment values

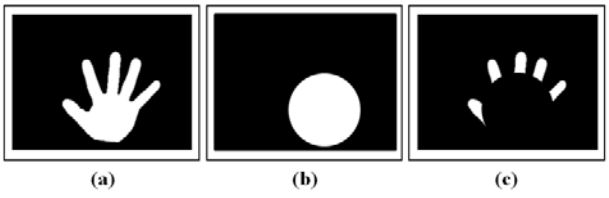

Figure 5. Labeling process (a) Original image (b) Circle image (c) Resultant image

D. User database creation

Identical hand shapes may actually turn out to have differing shape images or different Zernike moment values. Hence, this paper suggests the creation of a user independent and user independent database.

The user independent database is filled with hand shape information of numerous different people. The user dependent database saves all the moment values of each user's hand shapes.

The reason for creating two separate databases is to ensure a high rate of correct hand recognition even for first time users, and to lay groundwork for creating a personalized database to reach near-100% recognition rates.

E. Character input using hand recognition

To perform character input using images of hand shapes, a keypad, as shown in Figure 6, is used. The keypad differs from normal keypads in that each key corresponds to a certain hand shape. Users may imitate the shapes using their hands to type Korean characters. Figure 7 lists the 15 hand shapes and the numerical values they are linked to.

Figure 6. ShapeKII keypad

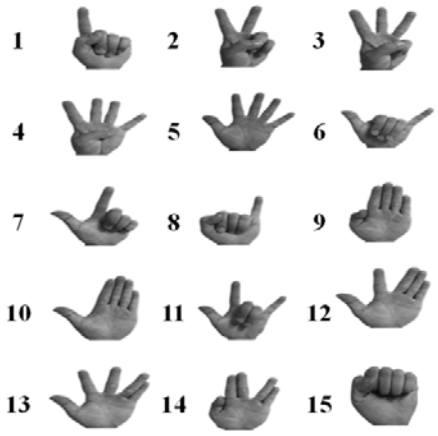

Figure 7. Hand shapes and their associated number values

IV. Experiment Results and Applications

Tests were performed using a computer with an Intel Core i7-2600K 3.4GHz processor and 8GB of memory,

along with a depth camera DS325 [6] linked at 60FPS. Software processing times are listed in [Table 1] below. Most of the time was used in creating the depth image. Considering the time required to compute the Zernike moment and the load required to run the process in real-time above 30FPS, it was decided that a 15-th order Zernike moment would be used. The q-recursive method introduced in [7] was applied to the Zernike moment to reduce the calculation time.

To conduct a validation, for the user independent database, 5 moment class values each of 15 hand shapes were saved – a total of 750 moment values. In the case of the user dependent database, 10 moment values were assigned to a hand shape – a total of 150 moment values. The sentence used to test the system was a Korean pangram,
"키스의고유조건은입술끼리만나야하고특별한기술은필요치않다" that required 125 separate input characters, of which the input speed Figure 8 was measured. Additionally, all hand shapes from number 1 to 15 were repeated 10 times at random to test for accuracy.

The obtained results Figure 8 indicate that the mean time required to type one character is 0.79 seconds. When the user pre-registered his hand shapes, and then attempted to type while looking at the instruction screen, recognition reached over 98%, but the amount of time required between input increased. To increase the typing speed, the user memorized each shape and typed without referring to instructions. Although this method increased typing speed, typo probability increased. Figure 9 and Figure 10 present the recognition rates for 10 users who made each hand shape from number 1 to 15 randomly 10 times each for both the user independent and user dependent cases. The user independent case showed a 96.06% rate, which is acceptable. The user dependent case, on the other hand, although requiring an extra registration process, yielded a higher recognition rate of 98.06%.

TABLE I. PROCESSING TIME

Order(n)	Zernike	Depth Image	Total
0	1 msec	16 msec	17 msec
5	2 msec	16 msec	18 msec
10	7 msec	16 msec	23 msec
15	13 msec	16 msec	29 msec
20	25 msec	16 msec	41 msec

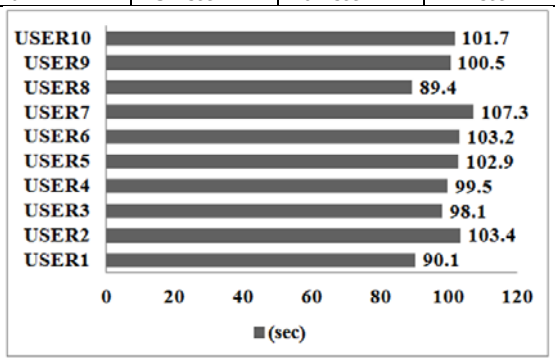

Figure 8. Pangram input time measurement

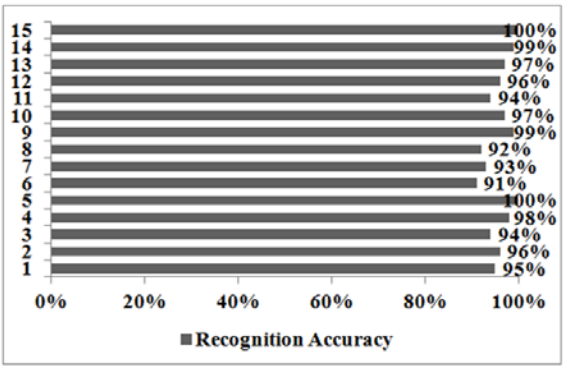

Figure 9. Independent user hand shape recognition accuracy

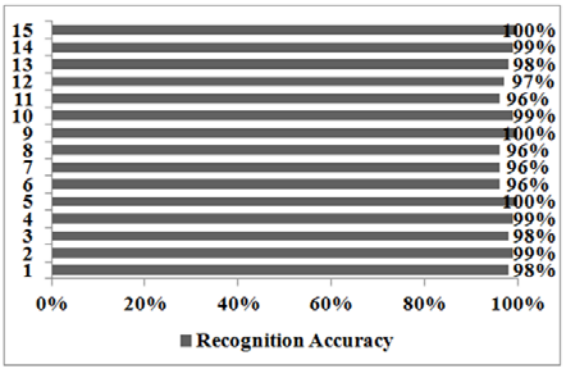

Figure 10. Dependent user hand shape recognition accuracy

A. *Application*

Taking into consideration the findings discussed in this paper, a program named ShapeKII was created. Figure 11 is the hardware setup used. A smart TV using a DS325 camera was installed in the top right. A gesture analysis was performed through a computer connected to the television by HDMI cable. Figure 12 illustrates an example of how ShapeKII is to be used.

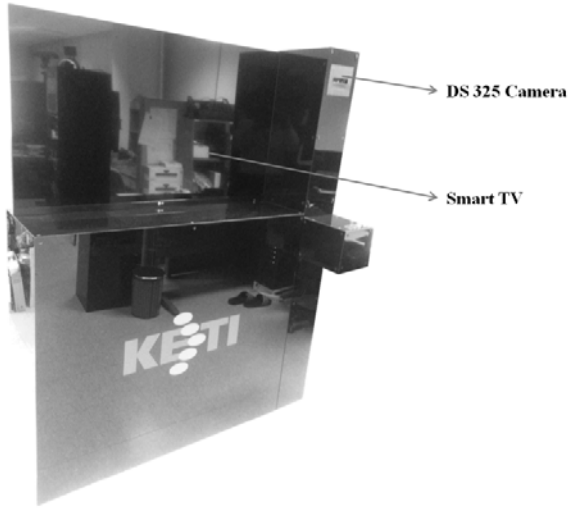

Figure 11. System hardware configuration

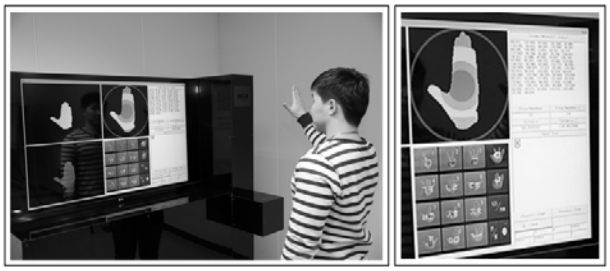

Figure 12. ShapeKII system (a) Character typing using hand shapes (b) Close-up of results screen

V. CONCLUSION

This paper proposed a Korean input system that uses a single depth camera to extract 3D information of a user's hand. The methodology presented worked regardless of the hand's location, using only the information of the hand's shape to type Korean. A user dependent database was employed so users could register their personal hand shapes, thereby increasing user friendliness. The presented method is currently limited to a short-range detection system, but it will be extended to longer-range detection in future work and will also be operable on smart devices.

REFERENCES

[1] R. Watson, "A survey of Gesture Recognition Techniques," Technical Report, TCD-CS-1993-11, 1993, pp. 1-31.

[2] P. Garg, N. Aggarwal, and S. Sofat, "Vision Based Hand Gesture Recognition," Proceedings of World Academy of Science, Engineering and Technology, vol.49, Jan 2009, pp. 972-977.

[3] M. Tosas, and B. Li, "Virtual Touch Screen for Mixed Reality", Proc. European Conference on Computer Vision, Lecture Notes in Computer Science, Prague, Czech Republic, vol 3058, May 2004, pp. 48-59.

[4] C. H. Teh and R T. Chin, "On Image Analysis by the Methods Of Moments" IEEE Transactions on Pattern Analysis and Machine Intelligence, vol, 10, no.4, 1988, pp. 496-513.

[5] A. Khotanzad and Y. H. Hong, "Invariant image recognition by zernike moments," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol, 12, no.5, 1990, pp. 489-497.

[6] http://www.softkinetic.com

[7] C. W. Chong, and P. Raveendran, and R. Mukundan, "A comparative analysis of algorithms for fast computation of Zernike moments," The Journal of the Pattern Recognition Society, 2003, pp.731-742.

 ISBN: 978-1-61208-468-8