JOURNAL OF KOREA MULTIMEDIA SOCIETY, VOL. 8, NO. 6, JUNE 2005(pp. 768-775)

Vision-Based Finger Spelling Recognition for Korean Sign Language

Jun Park†, Dae-hyun Lee††

ABSTRACT

For sign languages are main communication means among hearing-impaired people, there are communication difficulties between speaking-oriented people and sign-language-oriented people. Automated sign-language recognition may resolve these communication problems. In sign languages, finger spelling is used to spell names and words that are not listed in the dictionary. There have been research activities for gesture and posture recognition using glove-based devices. However, these devices are often expensive, cumbersome, and inadequate for recognizing elaborate finger spelling. Use of colored patches or gloves also cause uneasiness. In this paper, a vision-based finger spelling recognition system is introduced. In our method, captured hand region images were separated from the background using a skin detection algorithm assuming that there are no skin-colored objects in the background. Then, hand postures were recognized using a two-dimensional grid analysis method. Our recognition system is not sensitive to the size or the rotation of the input posture images. By optimizing the weights of the posture features using a genetic algorithm, our system achieved high accuracy that matches other systems using devices or colored gloves. We applied our posture recognition system for detecting Korean Sign Language, achieving better than 93% accuracy.

Keywords: Image Processing, Gesture Recognition, Posture Recognition, Finger Spelling, Sign Language, Human Computer Interaction

1. INTRODUCTION

Allowing for natural interactions between computers and humans, hand posture (or gesture) recognition has been one of the active research topics in Human Computer Interaction (HCI) areas. The applications include Virtual Environments (VE), Robotics, tele-presence, and sign languages. In sign languages, hand-postures (hand shape and the orientation) are commonly used to spell names and words that are not defined in the dictionary.

※ Corresponding Author : Jun Park, Address : (121-791) 72-1, Sangsu-Dong, Mapo-Gu, Seoul, Korea, TEL : +82-2-320-3032, FAX : +82-2-332-1653
E-mail : jpark@ cs.hongik.ac.kr
Receipt date : Sep. 8, 2004, Approval date : April 27, 2005
† Computer Science Department, Hong-ik University.
†† Computer Science Department, Hong-ik University.
(E-mail : egosus430@empal.com)
※ This work was supported by 2003 Hongik University Research Fund.

For most applications, a set of distinctive hand postures can be defined to easily detect postures and differentiate one posture from another. However, in a given sign language, a set of postures is predefined usually mimicking the alphabet shapes. Therefore finger spelling postures are often difficult to recognize using computing systems. Some postures are similar or even identical in a sign language: for example, postures for 'ㄷ'([d]) and 'ㅌ' ([t]) are similar and postures for vowel 'ㅏ'([a:]) and digit '1' are identical in Korean Sign Language (KSL). For these reasons, hand posture recognition for sign languages is more challenging than for other applications.

Finger spelling recognition researches can be categorized into two major groups: instrument-based and vision-based. Instrument-based approach often employs expensive devices such as data gloves. These devices are also cumbersome and inadequate for recognizing elaborate finger spelling. Vision-based approach is less expensive. However, many of the vision-based systems employ colored patches or colored gloves to help posture recognition.

In this paper, a vision-based finger spelling recognition system is introduced. Instead of using cumbersome devices or colored gloves, our method detected user's bare hand assuming that there are no skin-colored objects in the background. In our method, captured hand region images were separated from the background using a skin detection algorithm, and hand postures were recognized using a two-dimensional grid analysis method. Our method is not subject to the size of the input posture image. The rotation of the posture image was also corrected using the covariance matrix of the hand region. By optimizing the weights of the posture features using a genetic algorithm, our system achieved high accuracy that matches other systems using devices or colored gloves. We applied our posture recognition system for detecting KSL, achieving better than 93% accuracy.

2. PREVIOUS RESEARCH

There are two major approaches for detecting hand postures or gestures: instrument-based and vision-based. In instrument-based approach, glove-based devices (e.g., Data-gloves) have been used to capture hand gestures by directly measuring the joint angles and hand pose (position and orientation) [1-3]. These devices produced high recognition accuracy, and were successfully used in virtual reality and sign language applications. For example, Hernandez-Rebollar et al. invented a glove-like device using MEMS (Micro Electro Mechanical System) and accelerometers to recognize finger spelling of American Sign Language[1]. However, these devices are mostly expensive, cumbersome, and inadequate especially for recognizing elaborated gestures such as finger spelling[4].

Alternatively, vision-based approach is inexpensive, often efficient, and hence widely used for hand gesture recognitions[6-13]. However, general approach of vision-based hand gesture recognition is a challenging task because the hand-finger combination involves in motions of high DOF (degrees of freedom)[5]: there are 17 active joints above the wrist in human hands[1]. Consequently, the dimension of the problem space would disenable the application of naïve approaches in estimating the hand-finger joint angles. The noise and self-occlusion may also increase complexity and inhibit successful recognitions.

In order to minimize the computational overhead in hand gesture recognition, artificially designed markers can be attached to human hands or gloves. For example, multiple color patches were used for Japanese finger alphabet recognition. Various features (color combination of visible features, their dispersion, and the hand direction) were used for robust detection[12]. Colored glove was used to recognize finger spelling in German sign language. Characteristics such as contact of two areas and angle between two areas were used to distinguish hand-shapes[7]. These methods seemed to work properly. However, using patches requires laborious preparation, and wearing colored gloves might still cause uneasiness and inconvenience to the users.

Vision-based gesture recognition of bare hands is a more difficult task than using patches or colored gloves. Generally, two approaches have been employed for vision-based hand-articulation recognition[6,13]. One is model-based approach, where the input image is compared with the projections of a 3D hand model[6,7]. Fingertips[8] or line features[9] were used to enhance the robustness and accuracy. Although arbitrary hand postures can be recognized, this approach is not appropriate for finger spelling recognition. It is because finger spelling recognition involves in a set of limited number of postures, not arbitrary postures. Finger spelling recognition also requires interactive and real-time systems, while model-based approach is involved with search in high dimensional space, requiring high computation.

The other major vision-based recognition approach of bare hand is appearance-based. The sample gesture images are used during the learning stage to build relationship between the image features and the gesture configuration[10,11]. Terrillon et al. separated hand regions from face and background regions, and applied phase only correlation filter to recognize hand postures[10]. However, their approach was sensitive to the orientation of the input posture image. Shimada et al. used a neural network pattern recognition model to recognize finger spelling of Japanese Sign Language[11]. Their method was sensitive to the orientation, and the recognition accuracy was also low (about 85%).

Our approach is vision-based and appearance-based with bare hands without wearing any expensive or cumbersome devices or gloves. By employing many robust techniques such as robust skin detection, grid analysis, size normalization, rotation correction, and genetic algorithm, our system achieved high accuracy that matches other systems using devices or colored gloves.

3. FINGER SPELLING RECOGNITION SYSTEM

Overview of our system is as shown in Fig. 1. Video images were sampled periodically to obtain posture images, and the skin region was detected based on RGB values. From skin color regions, hand region was detected and separated from the background. Because the feature data extracted from the grid are influenced by the rotation of the posture, the grid was rotated to match with the input posture image.

Then posture features (described in section 3.3) were extracted from the grid and the hand region. A distance function was used to compute the similarity measure of the input posture features with

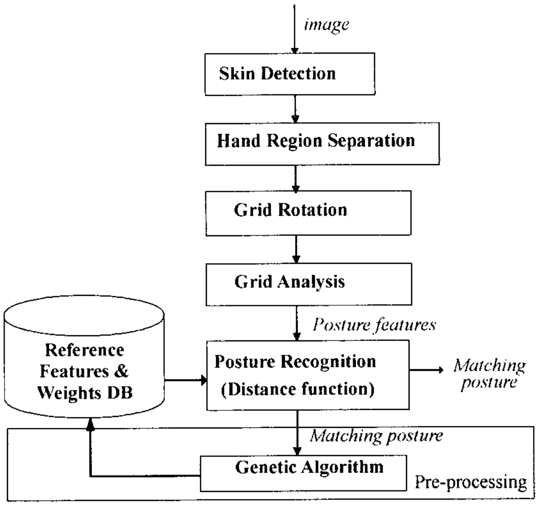


Fig. 1. System Overview.

the reference posture features. Because the weights of the features greatly affect on the performance of the distance function, we used a genetic algorithm to find optimal weight values.

3.1 Hand Region Separation

To detect user's hand from a real environment image, the skin regions need to be identified from the background. To be less affected by illumination conditions, transformation of color space from RGB to HSV was often regarded helpful. However, according to Shin et al., color space transformation doesn't seem to make any difference in skin detection[14]. Comparisons in RGB values look simple but are no worse than other methods in skin detection[15]. Wark et al. exploited that simple thresholding method with R/G ratio detected skin-like colors effectively[16]. We tested previous methods of skin region detection to find a method with desired performance. In our implementation, we employed a method similar to that of Peer et al.[17], which is summarized in the following.

$R>95$ and $G>40$ and $B>20$ and
$R>G+15$ and $R>B+15$

In some situations, detected skin region boundaries were too rough to be processed for posture recognition. To smooth boundary regions, we performed edge-detection (horizontal / vertical / diagonal) and boundary smoothing operations. This smoothing process may be switched off to reduce computation sacrificing image quality and recognition performance.

$$\begin{bmatrix} 1 & 3 & 1 \\ 0 & 0 & 0 \\ -1 & -3 & -1 \end{bmatrix} \quad \begin{bmatrix} 1 & 2 & 0 \\ 2 & 0 & -2 \\ 0 & -2 & -1 \end{bmatrix} \quad \begin{bmatrix} 0.025 & 0.1 & 0.025 \\ 0.1 & 0.5 & 0.1 \\ 0.025 & 0.1 & 0.025 \end{bmatrix}$$

Horizontal edge R-L diagonal edge Smoothing

3.2 Size Normalization and Rotation Correction

By extracting the bounding box of the hand region, the size of the input posture image was normalized before grid analysis. The feature data extracted from the grid are sensitive to the rotation of the posture images. As shown in the images of Fig. 2, small rotation of the hand results in a different look, which may lead to failure in recognition.

To solve this rotation sensitivity problem, orientation of the hand posture was calculated so that the grid can be rotated to match with the input posture image. First, the covariance matrix of the hand region was calculated based on the hand region and the center of gravity. From the covariance matrix, two perpendicular eigen-vectors were derived, one of which lies along the longest span of the hand region. The angular value between the input posture image and the reference posture image is obtained. The grid was, then, rotated using the computed angular value to extract the feature values of the input posture image with the corrected orientation, i.e. in the same orientation as the reference features were extracted.

3.3 Posture Recognition

After hand region separation, size normalization, and rotation correction, hand posture features were extracted from the image. The features were as follows.

- The ratio of width and height of the hand region
- The orientation of the hand region
- The center of gravity of the hand region
- Normalized positions of start and end grid points overlapped with the hand regions (horizontal / vertical)

Because the skin detection algorithm we used was somewhat affected by self-shadow and illumination condition, there were situations when noise in the images affects on the recognition performance. To reduce the effect of noise, small fragments were eliminated heuristically.

Then we used a distance function to calculate the closest posture. Because the significance of the individual features varies (some features affect more on the distant function than others), the performance of the distance function depends on the weight of each feature parameter. However, it is very difficult to identify which feature is more significant than others by how much.

In our method, to find optimal weight values, a genetic algorithm was used. We used a test set of 1550 images of hand postures and evaluated with 1000 population. According to the experiment, use

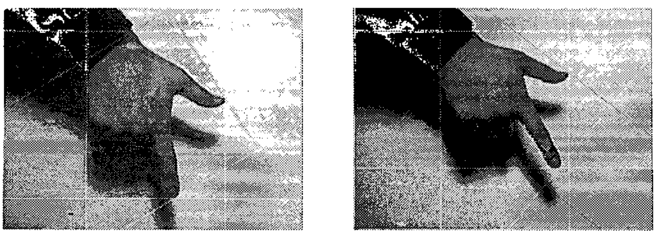


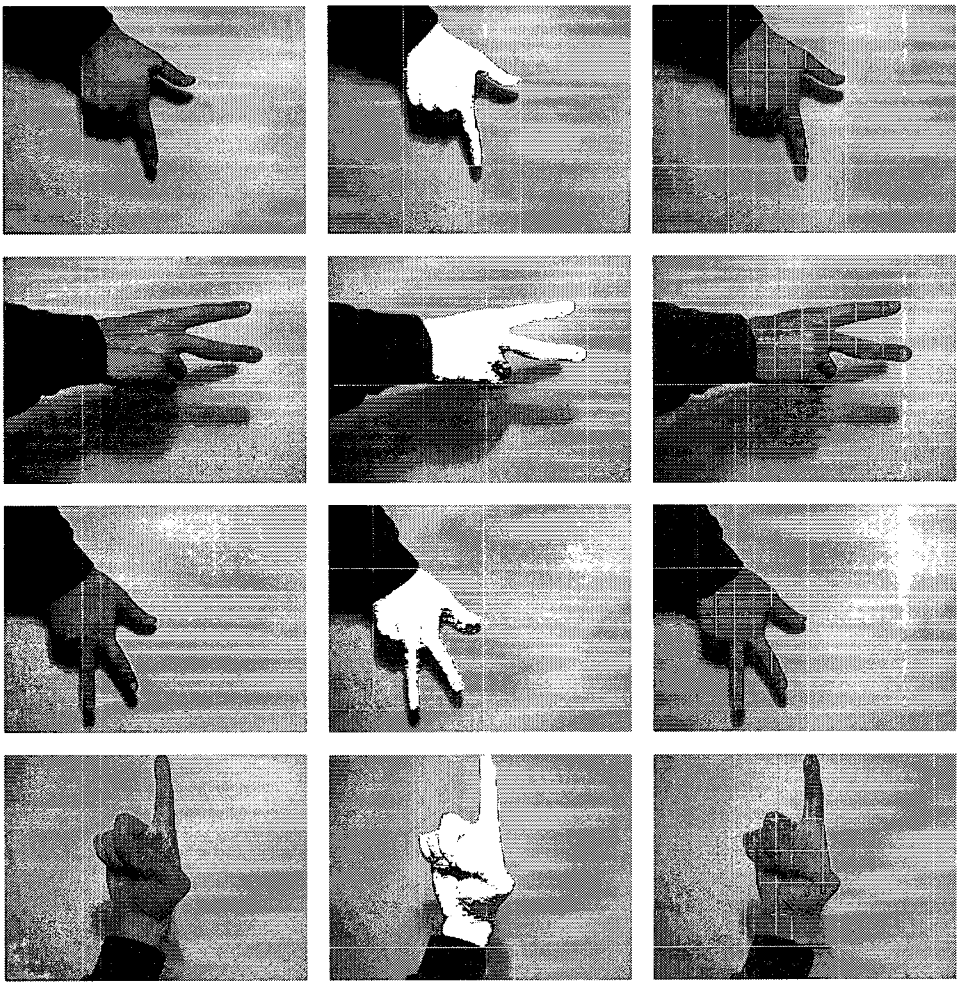
Fig. 2. Rotation Sensitivity of Posture Recognition.

of computed optimal weights increased the recognition accuracy up to 93%. Comparing with below 70% without using weights, the performance was greatly enhanced thanks to optimal weights.

4. EXPERIMENTS AND RESULTS

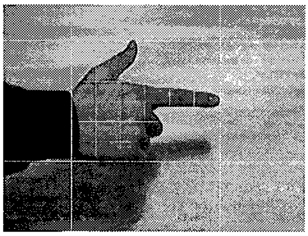
We tested our system for recognizing finger spelling of Korean Sign Language. The feature parameter values of each finger-spelled alphabet were extracted and stored in the database during the pre-processing stage. In Fig. 3, several images of alphabet postures are shown. The original images are shown in the left, the skin-detected images in the middle, and the results of grid analysis in the right column. During the experiment, illumination was controlled so that skin regions get adequate amount of light. Background environment was cleaned up so that there are no skin-colored objects.

Fig. 4 shows an example of word recognition. A sequence of posture images was captured as the user made the alphabet postures for Korean word meaning “tree” (spelled [na:mu:] in Korean). The word is composed of four alphabets, each of which

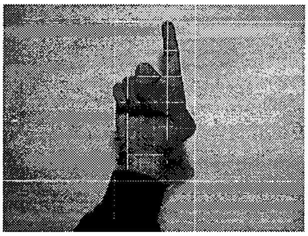


Original images　　Skin region detection　　Grid analysis

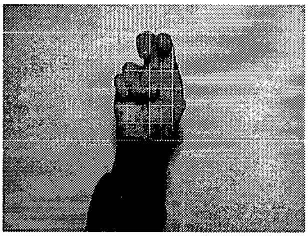
Fig. 3. Steps for Finger Spelling Recognition.



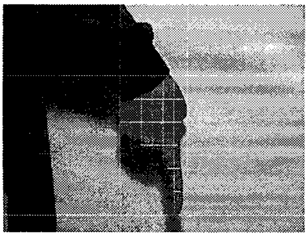
Alphabet for "[n]"



Alphabet for "[a]"



Alphabet for "[m]"



Alphabet for "[u]"

Fig. 4. Recognition of Korean word meaning "tree".

was detected to compose the complete word.

5. CONCLUSIONS

In this paper, we proposed a posture detection system for recognizing finger spelling of a sign language. Our method did not employ any expensive or cumbersome devices such as data gloves. Instead, an inexpensive PC camera was used for vision-based hand posture recognition. A robust skin region detection algorithm was used to separate hand regions from the background. Hand postures were analyzed using a grid analysis method by scanning the hand region images vertically and horizontally. Our method was not subject to the size of the input posture image. The rotation of the posture image was also corrected using the covariance matrix of the hand region. By optimizing the weights of the posture features using a genetic algorithm, our system achieved high accuracy that matches other systems using devices or colored gloves. The method was tested for recognizing Korean alphabets and words to assist communication between hearing impaired and Korean-speaking people. The recognition success rate was above 93%. Combined with other technologies, our posture recognition system is expected to enrich the interactions in other applications including virtual environments.

6. REFERENCES

[1] J.L. Jhernandez-Rebollar, R.W. Lindeman, and N. Kyriakopoulos, "A Multi-Class Pattern Recognition System for Practical Finger Spelling Translation," *Proceedings of 2002 the Fourth International IEEE Conference on Multimodal Interfaces*, pp.185-190, Oct. 14-16, 2002.

[2] R.H. Liang, "Continuous Gesture Recognition System for Taiwanese Sign Language," *PhD Thesis, National Taiwan University*, Taiwan, R.O.C., 1997

[3] Chan-Su Lee, Gyu-tae Park, Jong-Sung Kim, Zeungnam Bien, Won Jang, and Sung-Kwon Kim, "Real-time Recognition System of Korean Sign Language based on Elementary Components," Proceedings of Sixth International

Conference on Fuzzy Systems (IEEE FUZZ'97), pp.1463-1468, 1997, Spain.

[4] Jung-Bae Kim, Kwang-Hyun Park, Won-Chul Bang, and Bien, "Continuous gesture recognition system for Korean sign language based on fuzzy logic and hidden Markov model," *Proceedings of the 2002 IEEE International Conference on Fuzzy Systems, Volume: 2*, pp.1574-1579, 2002.

[5] John Lin, Ying Wu, and T.S. Huang, "Modeling the constraints of human hand motion," *Proceedings of 2000 Workshop on Human Motion*, pp.121-126, 2000.

[6] Ying Wu, J.Y. Lin, and T.S. Huang, "Capturing natural hand articulation," *Proceedings of Eighth IEEE International Conference on Computer Vision, 2001 (ICCV 2001)*, Volume: 2, pp.426-432, 2001.

[7] K. Grobel and H. Hienz, "Video-based handshape recognition using a handshape structure model in real time," *Proceedings of the 13th International Conference on Pattern Recognition*, vol.3, pp.446-450, Aug 1996.

[8] Nobutaka Shimada, Yoshiaki Shirai, Yoshinori Kuno, and Jun Miura, "Hand gesture estimation and model-refinement using monocular camera-ambiguity limitation by inequality constraints," *Proceedings of the 3rd conference on Face and Gesture Recognition*, pp. 268-273, 1998.

[9] Jakub Segen and Senthil Kumar, "Shadow Gestures: 3D Hand Pose Estimation using a Single Camera," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Vol.1, pp.479-485, June 1999.

[10] Jean-Christophe Terrillon, Arnaud Piplré, Yoshinori Niwa, and Kazuhiko Yamamoto, "Robust Face Detection and Japanese Sign Language Hand Posture Recognition for Human-Computer Interaction in an 'Intelligent' Room," *The 15th International Conference on Vision Interface*, pp.369-376, Calgary, Canada, May 27-29, 2002.

[11] M. Shimada, S. Iwasaki, and T. Asakura, "Finger spelling recognition using network with pattern recognition model," *SICE 2003 Annual Conference*, Vol. 3, pp. 2458-2463, 4-6 Aug. 2003

[12] Kazuyoshi Yoshino, Toshio Kawashima, and Yoshinao Aoki, "Recognition of Japanese Sign Language from Image Sequence Using Color Combination," *Proceedings of IEEE International Conference on Image Processing*, Vol. III, pp.511-514, September, 1996.

[13] Yasushi Hamada, Nobutaka Shimada, and Yoshiaki Shirai, "Hand Shape Estimation Using Sequence of Multi-Ocular Images Based on Transition Network," *The 15th International Conference on Vision Interface*, pp.362-368, Calgary, Canada, May 27-29, 2002.

[14] M.C. Shin, K.I. Chang, and L.V. Tsap, "Does Colorspace Transformation Make Any Difference on Skin Detection," *IEEE Workshop on Applications of Computer Vision*, pp.275-279, 2002.

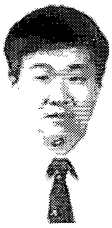
[15] V. Vezhnevets, V. Sazonov, and A. Andreeva, "A Survey on Pixel-Based Skin Color Detection Techniques," *Proc. Graphicon2003*, pp.85-92, Moscow, Russia, 2003.

[16] T. Wark and S. Sridharan, "A Syntactic Approach to Automatic Lip Feature Extraction for Speaker Identification," *Proceedings of ICASSP '98*, pp. 3693-3696, 1998.

[17] P. Peer, J. Kovac, and F. Solina, "Human skin colour clustering for face detection," *EUROCON 2003-International Conference on Computer as a Tool*, 2003.

Jun Park

1993 Computer Science and Statistics, Seoul National University (BS)
1996 Computer Science, University of Southern California (MS)
2000 Computer Science, University of Southern California (Ph.D.)
2001~2002 Post-Doc. Researcher at Rockwell International, Science Center
2002~Present Assistant Professor, Computer Science and Engineering Department at Hongik University
Research Interests : Augmented Reality, Virtual Reality, HCI, Computer Graphics, Interactive Media



Dae-hyun Lee

2003 Computer Science and Engineering, Hongik University (BS)
2005 Computer Science and Engineering, Hongik University (MS)
2005~Present Senior Engineer, Gomid Inc.
Research Interests : Computer Graphics, HCI, Game Development, Real-time 3D Graphics