American Sign Language recognition by using 3D geometric invariant feature and ANN classification

Conference Paper · November 2014

DOI: 10.1109/BMEiCON.2014.7017372

Watcharin Tangsuksant
King Mongkut's University of Technology North Bangkok

American Sign Language Recognition by Using 3D Geometric Invariant Feature and ANN Classification

Watcharin Tangsuksant[1], Suchin Adhan[2], Chuchart Pintavirooj[1]
[1]Department of Biomedical Engineering, Faculty of Engineering
[2]Department of Engineering Education, Faculty of Industrial Education
King Mongkut Institute of Technology Ladkrabang Bangkok, Thailand
w.tangsuksant.m@hotmail.com, kasuchin@kmitl.ac.th, chuchartpintavirooj@gmail.com

Abstract—**Communication between normal and disabled person has been developed in several researches. The hand gesture is one of important communication for the deaf, especially American Sign Language (ASL) which is used in order to represent each alphabet (A-Z). This paper aims to translate ASL from static postures. Besides, this research also designs the glove with six different colored markers and develops algorithm for alphabet classification. Moreover the system is set by two cameras in order to extract 3D coordinate points from each marker. There are three main important processes for algorithm consisting of marker detection by using Circle Hough Transform, computation of all feasible triangle area patches constructed from 3D coordinate triplet that is novel feature, and feature classification using feedforward backpropagation of Artificial Neural Network. The experimental result shows average of accuracy is 95 percent that is high performance and feasibility for proposed method.**

Keywords-component; American Sign Language recognition, Geometric invariant feature, 3D extraction, ANN

I. INTRODUCTION

Several researches focus on sign language translation issue in order to help the deaf easily communicate with normal person. The sign languages are general languages that consist of local and universal languages. There are many active researches on automatic local sign languages translation such as Indian, Malaysian and Thai sign languages [1-3]. Among all, American Sign Language (ASL) is one of the most important and widely used universal languages for deaf. Consequently, there are many researchers emphasizing to develop automatic ASL translation system.

Ensoek Jong *et. al.* [4] presented finger-gesture recognition. Their researches designed glove that use conductive material which was called by “Velostat”. The glove is equipped with microcomputer and Bluetooth module which make the glove inconvenient for user. N. Tanyawiwat and S. Thiemjarus [5] designed a wireless sensor glove for ASL fingerspelling gesture recognition with installed five flex sensors on fingers and 3D accelerometer on the back of the hand. The accuracy of experimental is lower than 80 percent.

In 2012, M.S. Sinith *et. al.* [6] proposed HSL recognition by using Support Vector Machine (SVM) for tracking six sign languages including A, H, I, O, L and W. Although the experimental overall result wass around 90 percent but the performance is poor for distinguishing the W and O alphabet. In addition, M. Ashraful Amin and Hong Yan [7] developed sign language recognition. The Principle Component Analysis (PCA) was used for feature extraction and Fuzzy C-Means (FCM) for classification. Andreas Domingo *et al.* [8] used LAB View for recognizing alphabet A-Z. Their research calculated centroid of each end of fingers for feature extraction. In 2013, Rudy Hartanto and others [9] focused on creating a system to translate A-Z sign language (J and Z is not tested). The hand contours were used to localize hand area and SIFT algorithm was used to compute the key point of each posture. The result was promising.

Additionally, there is a fascinating technique which is 3D technique. 3D technique is applied on many researches for sign language translation [10-14]. Liya Ding and Aleix M. Martinez [10] presented algorithm to recover the 3D shape and motion trajectory of hand from 2D video sequences of ASL words. But their research used manual detection around knuckles of each finger. In 2013, Myoung-Kyu Sohn and others [13] proposed a hand gesture recognition method by using 3D hand motion trajectory technique with depth camera, and then normalized for translation invariant feature extraction and K-NN classification. Our research proposes a 3D novel technique for automatic translation of ASL using geometric invariant feature extraction. Our features are intrinsic, local in nature and affine invariant which is efficient for classifying the ASL languages.

II. PROPOSED METHOD

In this research, black glove installed with 6 makers is designed and constructed. Five makers are attached on the fingertip leaving one maker on the palm center. Subject wearing the black shirt and the black glove will be captured with two USB cameras installed on the desk. The 640x480 pixel color image acquired from two camera are then used for 3D extraction of maker coordinates using a well-known DLT algorithm . Figure 1 shows our block diagram.

The DLT algorithm requires camera calibration. The camera calibration procedure is performed to obtain intrinsic and extrinsic parameters of camera using a chessboard pattern as used in [16]. The camera calibration matrix is shown in Eq.1 where f_x and f_y are focal length in x-axis and y-axis

978-1-4799-6801-5/14/$31.00 ©2014 IEEE

respectively, s is skew parameter, x_o and y_o are principal point coordinates of the camera. Extrinsic parameters as shown in Eq.2. It consists of rotation matrix and translation vector. The M matrix is a 3-by-4 matrix named projection matrix as shown in Eq.3. The extrinsic and intrinsic matrices of two camera are then used to compute the 3D coordinate of markers.

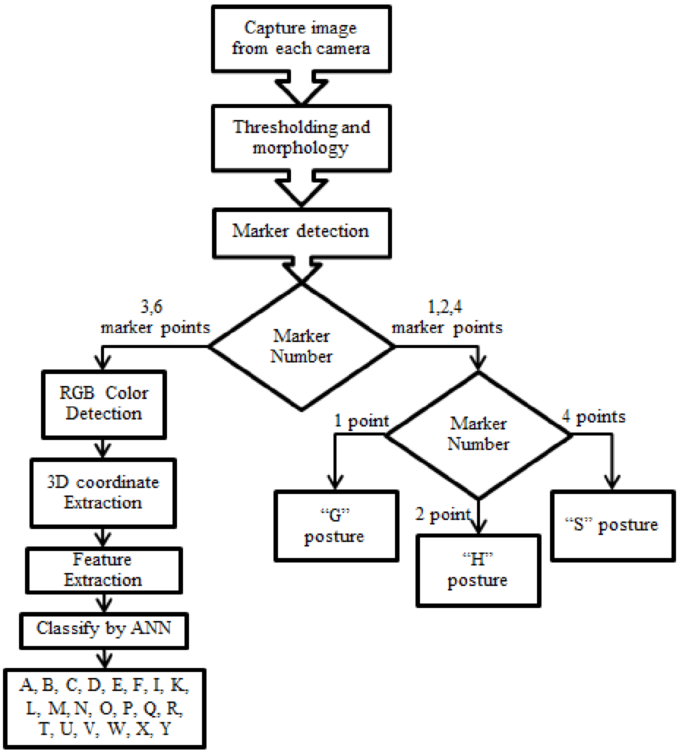

Figure 1. Block diagram of ASL recognition

$$Intrinsic = \begin{bmatrix} -f_x & s & x_o \\ 0 & -f_y & y_o \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

$$Extrinsic = \begin{bmatrix} r_{11} & r_{12} & r_{13} & T_x \\ r_{21} & r_{22} & r_{23} & T_y \\ r_{31} & r_{32} & r_{33} & T_z \end{bmatrix} \quad (2)$$

$$M = Intrinsic \times Extrinsic = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \quad (3)$$

A. Image Acquistion and Pre-processing

This research uses two cameras for capturing the image of each alphabet postures simultaneously. Figure 2 shows image of static alphabet posture of ASL. The original color images size is 640×480 pixels. Then the color image is pre-processed including thresholding and eliminating some small noises using opening and closing method as shown in figure 3.

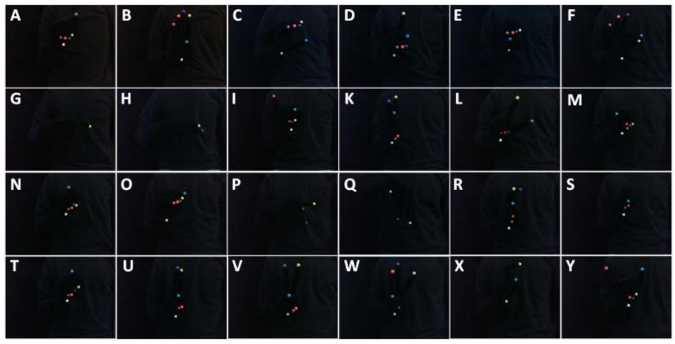

Figure 2. ASL static posture

B. Marker Detection

After pre-processing, next step is marker detection using CHT. The basic of Circle Hough Transform is an algorithm that calculates the coordinates for center of circle which is defined by maximum accummulator array M[a,b] as shown in Eq.4. The calculation relies on radius r, moreover x and y are coordinate of edge object.

$$0 < \theta < 360 \begin{cases} a = x - r\cos\theta \\ b = y - r\cos\theta \end{cases} \quad (4)$$

In this paper, we exploit CHT to detect the circular marker as shown in figure 3. The radius of marker is defined around 5 to 20 pixels due to inconstant distance of each subject hand that effects to radius. However, the number of marker detected is then used as the pre-classification step to distinguish the G, H and S posture since there are one, two and four markers for G, H and S respectively. The detected marker for all other hand sign symbols is then used in the feature extraction and classification process.

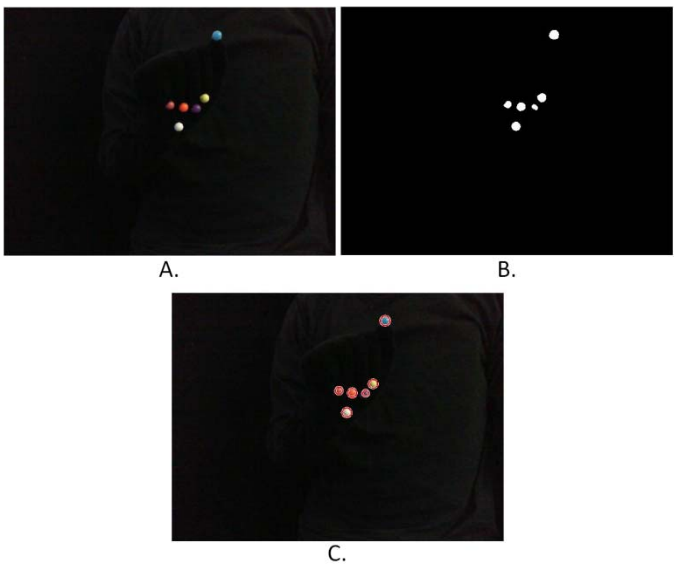

Figure 3. A. Original image B. Pre-processing process
C. Marker detection by using CHT

C. *Color Detection*

This step proposes the color detection in order to match each marker centers from two images. We create window 5×5 pixels around each center then calculate mean of Red, Green and Blue of these window after that calculate sum of mean difference of Red, Green and Blue as Eq.5,

$$sum\ of\ difference = (\bar{R} - \bar{R}') + (\bar{G} - \bar{G}') + (\bar{B} - \bar{B}') \qquad (5)$$

where $\bar{R}, \bar{G}, \bar{B}$ and $\bar{R}', \bar{G}', \bar{B}'$ are mean of Red, Green, Blue from first and second camera respectively. The least value is coordinate center.

D. *Three-dimensional Coordinate Extraction*

The DLT algorithm is used in order to transform 2D image coordinates into 3D object space coordinates. This algorithm required at least two images. The projection matrix for each camera is denoted M' and M'', respectively. The projection matrix of first camera is divided into three row vectors $m_1'^T$, $m_2'^T$ and $m_3'^T$. Likewise $m_1''^T, m_2''^T$ and $m_3''^T$ are three rows of projection matrix of second camera. The point of each image from first and second cameras are denoted (u', v', w') and (u'', v'', w''), respectively (normalize w' and w'' into 1) .

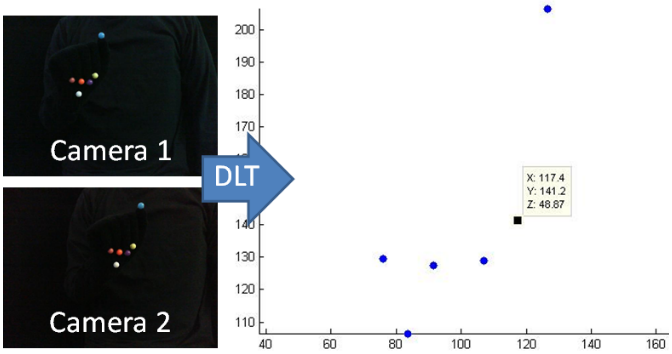

Figure 4. 3D coordinates extraction by DLT

The 2D coordinate of the corresponding marker is acquired and then computed for 3D coordinate computation The equation is shown in Eq.6 and the result as shown in figure 4.

$$\begin{bmatrix} u'm_3'^T - w'm_1'^T \\ v'm_3'^T - w'm_2'^T \\ u''m_3''^T - w''m_1''^T \\ v''m_3''^T - w''m_2''^T \end{bmatrix} \tilde{P} = A\tilde{P} \qquad (6)$$

Where P is the2D coordinate of the corresponding marker and $\tilde{P}$ is the extracted 3D coordinate. Size of matrix A is 4-by-4. Vector $\tilde{P}=[X\ Y\ Z\ 1]^T$ is a homogenous form of point P(X, Y, Z). We are interested in a nontrivial solution (P≠0) of Eq.6, *i.e.* considering the case where $\det(A) = 0$. Singular Value Decomposition (SVD) had been performed to solve P.

D. *Feature Extraction*

This paper proposes a new feature that computes all feasible triangle area patches constructed from 3D coordinates triplet. For n detected marker, there are number of area patches as shown in Eq. 7 and then sort each area from smallest to biggest.

$$C_{n,3} = \frac{n!}{(n-3)!3!} \qquad (7)$$

After that compute the each triangle area by using Heron's formula as shown in Eq.8. Length of each side is represented A, B and C, S is shown in Eq.9.

$$Triangle\ Area = \sqrt{S \times (S-A) \times (S-B) \times (S-C)} \qquad (8)$$

$$S = \frac{A+B+C}{2} \qquad (9)$$

In case of P, Q and X posture, these postures have only three detected markers. There are only one triangle area of P, Q and X. To discriminate among these three symbols, another angle feature is used which is three vertex angle of triangle patch. The detected feature (area and angle) is then sorted in a decreasing order. To derive the scale invariant feature, the feature sequence is normalized by the maximum value prior to the applying as the input of the Artificial Neural Network (ANN) classifier.

E. *Classification by Using Artificial Neural Network*

The structure of biological neural consists of dendrite, cell body and axon for collecting, processing and output signal respectively. Likewise, the basic of ANN consists of input layer (X), hidden layer (Z) and output layer (Y). Each input has difference weight values, after that multiply input with weight value. When sum of these values more than the threshold, output to the next neuron. Figure 5 shows the basic structure of ANN. In our research input X is the sorted area and angle sequence.

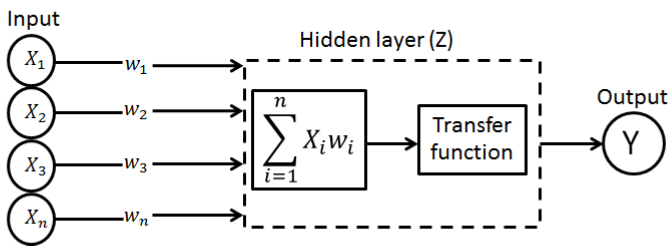

Figure 5. Basic structure of ANN

Feedforward backpropagation training neural network is used in this research; there are two processes which are training process and application process. For the training process in this research we uses 2,100 images training data (100 images per a static posture). The input layer contains feature of each posture and hidden layer is determined empirically to be 21 neurons and 21 output layer neurons. During training the weight and biases are iteratively adjust to minimize the sum square error. We use Log-sigmoid transfer function so as to fix rang of output value 0 to 1. For application process, we test the accuracy of algorithm with 480 images for static posture.

III. EXPERIMENTAL RESULT

100 images per a posture are trained by ANN. Figure 6 illustrates graph of feature mean of each posture which are represented by different color lines. Clearly, these lines on the graph are noticeable difference and then can be served as a signature of each posture

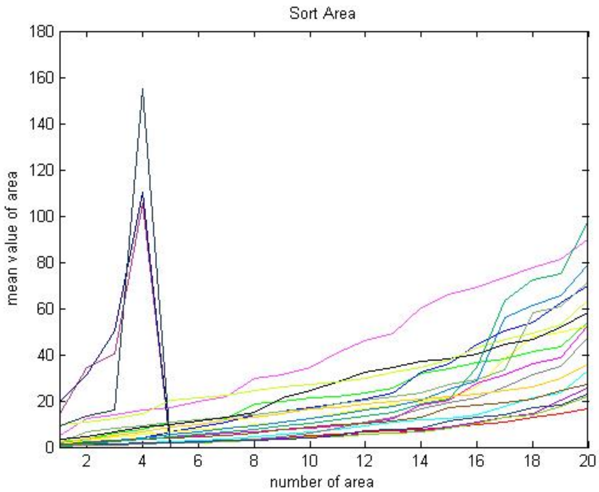

Figure 6. Mean feature of each posture

The experimental result is shown in the Table 1. We test the accuracy of the algorithm with 480 images (20 images per a posture). Performance of this algorithm for static postures Translation of ASL shows that average of accuracy of 95 % computing from Eq.10.

$$Accuracy\ (\%) = \frac{correct\ output}{tatal\ number\ of\ input} \times 100 \qquad (10)$$

IV. DISCUSSION

Based on the experimental result the average of accuracy is 95 percent which is very promising results. In the table 1, there are many postures have 100 percent and some postures is less than. Thus, we will explain advantages and mistakes of this algorithm.

A. Advantage

This research proposes novel featured extraction which is applied by 3D technique. The crucial featured extraction is geometric invariant feature; therefore, it is local and intrinsic and hence is independent to the camera orientation.

B. Error

Although the present algorithm is high performance, there still error is some points. This may be due to the non-circular shape of the marker after the thresholding process. Moreover, light reflection is one of effect that affect to marker detection. This research has to control light environment .

Table 1. Accuracy of algorithm test

Alphabets	Correct (times)	Wrong (times)	Accuracy (%)
A	20	0	100
B	20	0	100
C	20	0	100
D	18	2	90
E	17	3	85
F	19	1	95
G	20	0	100
H	20	0	100
I	20	0	100
K	18	2	90
L	20	0	100
M	20	0	100
N	18	2	90
O	17	3	85
P	20	0	100
Q	20	0	100
R	17	3	85
S	20	0	100
T	17	3	85
U	18	2	90
V	19	1	95
W	19	1	95
X	20	0	100
Y	19	1	95
Average of accuracy (%)			**95.00**

V. CONCLUSION

This paper proposes the feasible method for American Sign Language recognition. We design the glove with 6 different color markers. The markers are automatic detected using Circle Hough Transform. The 2D coordinate extracted from markers captured from two cameras are then used to extract 3D coordinate of marker using DLT. All possible triangle patches constructed from markers triplet are then computed and sorted in an orderly fashion. The areas sequences are then used as input of feedforward backpropagation of Artificial Neural Network for feature classification. The experimental result illustrates the average of accuracy is 95 percent. For the future work, we will be improved in case of dynamic posture and translate in the real-time system.

REFERNCES

[1] J. Rekha, J. Bhattacharya, S. Majumder, “Shape, Texture and Local Movement Hand Gesture Features for Indian Sign Language Recognitino”, Trendz in Information Sciences and Computing (TISC), 2011, 8-9 Dec. 2011.

[2] Akmeliawati R., Ooi M. P., and Ye Chow Kuang, “Real-Time Malaysian Sign Language Translation using Colour Segmentation and Neural Network”, Instrumentation and Measurement Technology Conference Proceedings, 2007. IMTC 2007. IEEE, 1-3 May 2007.

[3] Phonkrit Chanda, Sansanee Auephanwiriyakul, and Nipon Theera-Umpon, “Thai Sign Language Translation System Using Upright Speed-Up Robust Feature and C-Means Clustering”, IEEE World Congress on Computational Intelligence, June 10-15, 2012.

[4] Eunseok Jeong, Jaehong Lee, and DaeEun Kim, “Finger-gesture Recognition Glove using Velostat”, 11th International Conference on Control, Automation and System, Oct. 26-29, 2011.

[5] Netchanok Tanyawiwat, and Surapa Thiemjarus, “Design of an Assistive Communication Glove using Combined Sensory Channels”, Wearable and Implantable Body Sensor Networks (BSN), 2012 Ninth International Conference, 9-12 May 2012.

[6] M.S. Sinith, Soorej G Kamal, Nisha B.,Nayana S, Kiran Surendran, and Jith P S, “Sign Language Recognition Using Support Vector Machine”, 2012 International Conference on Advances in Computing and Communications, 9-11 Aug.,2012.

[7] M. Ashraful Amin, and Hong Yan, “Sign Language Finger Alphabet Recognition From Gabor-PCA Representation of Hand Gestures”, Proceeding of the 6th International Conference on Machine Learning and Cybernetics, 19-22 Aug. 2007.

[8] Andreas Domingo, Rini Akmeliawati, and Kuang Ye Chow, “Pattern Matching for Automatic Sign Language Translation System using LabVIEW”, International Conference on Interlligent and Advanced Systems 2007, 25-28 Nov. 2007.

[9] Rudy Hartanto, Adhi Susanto, and P . Insap Santosa, “Preliminary Design of Static Indonesian Sign Language Recognition System”, Information Technology and Electrical Engineering (ICITEE), 2013 International Conference, 7-8 Oct. 2013.

[10] Liya Ding and Aleix M. Martinez, “Three-Dimensional Shape and Motion Reconstruction for the Analysis of American Sign Language”, Computer Vision and Pattern Recognition Workshop, 17-22 June 2006.

[11] Mahmoud Elmezain, Ayoub Al-Hamadi, Saira Saleem Pathan, and Bernd Michaelis, “Spatio-Temporal Feature Extraction-Based Hand Gesture Recognition for Isolated American Sign Language and Arabic Numbers”, 6th International Symposium on Image and Signal Processing and analysis, 16-18 Sept.2009.

[12] Kin Fun Li, Kylee Lothrop, Ethan Gill, and Stephen Lau, “A Web-Based Sign Language Translator Using 3D Video Processing”, 14th International Conference on Network-Based Information Systems (NBiS), 7-9 Sept. 2011.

[13] Myoung-Kyu Sohn, Sang-Heon Lee, Dong-Ju Kim, Byungmin Kim, Hyunduk Kim, “3D Hand Gesture Recognition from One Example”, 2013 IEEE International Conference on Consumer Electronis (ICCE), 11-14 Jan 2013.

[14] Chenyang Zhang, Xiaodong Yang, and YingLi Tian, “Histogram of 3D Facets: A Characteristic Descriptor for Hand Gesture Recognition”, 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 22-26 April 2013.

[15] Z. Zhang, "A Flexible New Technique for Camera Calibration," IEEE Transactions on Pattern Analysis & Machine Intelligence, Vol. 22, pp. 1330-1334, 2000.