Int. J. Advance Soft Compu. Appl, Vol. 11, No. 3, November 2019
ISSN 2074-8523

Thai Finger-Spelling Sign Language Recognition Employing PHOG and Local Features with KNN

Thongpan Pariwat[1], Pusadee Seresangtakul[2*]

NLSP Laboratory, Computer Science Department
Khon Kaen University, Khon Kaen, Thailand 40002
[1] e-mail: thongpan.par@kkumail.com
[2] e-mail: pusadee@kku.ac.th

Abstract

Sign Language recognition is an important tool for the hearing-impaired in order to communicate with both hearing-impaired and hearing individuals. Similarities in finger-spelling sign language are one of the main factors or problems influencing the accuracy of sign language recognition. This research focuses on one-stroke, Thai finger-spelling sign language (TFSL) in methods of feature extraction with the pyramid histogram of oriented gradients (PHOG) and local features, as well as the application of K-Nearest Neighbors (KNN) recognition. We present, herein, a Thai finger-spelling sign language recognition system (TFSLR) used to classify alphabets shown in similar gestures. Five signers postured fifteen Thai alphabet characters, in which images were taken in five repetitions, totaling 375 finger-spelling images. Our experiment utilizes five-fold cross-validation in order to evaluate the projected system effectiveness. Additionally, we compared the results of each experiment involving PHOG with the amalgamation of PHOG and the local features. The results showed that such amalgamation was capable of handling similarly signed characters with an average accuracy of 97.6%.

Keywords: *Finger-spelling sign language, histogram of oriented gradients, k-nearest neighbors, PHOG.*

1 Introduction

Sign language is the language used for communication by the deaf or hearing impaired, as a significant communicative barrier exists between the hearing impaired and the non-hearing impaired. Incorporating hand, nose, and mouth movements to communicate; very few people are capable of learning such a difficult form of language. For this reason, it is necessary to improve upon the existing sign language recognition systems as a communication tool for all users.

Research in sign language recognition has enjoyed universal appeal. However, there are no standards in existence, as each country develops its own methods; including American Sign Language (ASL) [1], Indian Sign Language (ISL) [2], Chinese Sign Language (CSL) [3], and numerous others. In the case of Thai sign language, the sign language recognition systems in place include both Thai Sign Language (TSL) [4] and Thai Finger-spelling Sign Language (TFSL) [5]. Deaf schools in Thailand incorporate Thai finger-spelling as the sign language model used in their curriculum. Thai finger-spelling signs are categorized into three groups, determined by the number of finger gesture strokes: one-stroke with 15 characters, two-stroke with 24 characters, and three-stroke with three characters. In the use of sign language in according with the declaration of the Thailand's Royal Institute Dictionary, 'ฅ' (/k^h/) and 'ฃ' (/k^h/) cannot be found, and are therefore no longer used [6]. Researchers have primarily focused on the one-stroke signs, as this group is the starting point for both the two-stroke and three-stroke groups.

In previous research, we proposed a recognition system of Thai finger-spelling sign language with global and local features, utilizing Support Vector Machines (SVM) [5]. Global features describe the entire hand image as the area of the convex hull, area of hand density, differences between the bounding box and hand area, differences between the convex hull and hand area, and a group of histograms, such as the vertical and horizontal histogram, and the vertical histogram (starting at the centroid). Local features describe the inner hand detail, consisting of the density and Hu moment. Five right-handed signers were employed in our experiment. They wore dark long-sleeved shirts, and stood in front of a blue background. Our research focused on the 15 characters of the one-stroke gestures of the TFSL. Experiments were user-independent, with the cooperation of three features: 1) the amalgamation global and local features; 2) global features; and 3) local features. The effectiveness of this research was examined through five-fold cross validation. Recognition was conducted via SVM with four kernels; Linear, Polynomial, RBF, and Sigmoid. The highest accuracy (91.20%) was achieved from the global and local features combination via the RBF kernel. However, similarly signed characters resulted in much lower accuracy. The differences in these alphabetical characters can be discerned from the positioning of the thumb; such as 'ต' (/t/), 'น' (/n/), 'ม' (/m/), 'ส' (/s/), and 'อ'

(/ʔ/), as show in Fig. 1 (a); and two similar signed characters, 'ด' (/d/) and 'ร' (/r/), as show in Fig.1 (b).

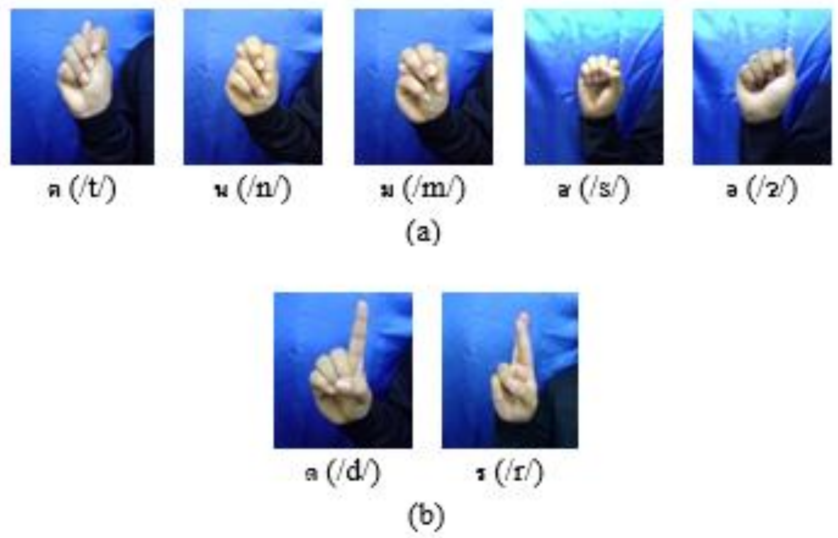


Fig. 1. Signs having similar hand gestures.

A Pyramid Histogram of Oriented Gradients (PHOG) is widely used in the areas of object detection, image retrieval, vehicle classification, and facial recognition [7]. PHOG provides the image features through spatial layout. The local shape is the spatial shape descriptor, which consists of the gradient orientation at each pyramid resolution level [8]. This system offers the use of vision-base techniques as the basis for image processing in the development of new Thai finger-spelling sign language recognition systems. Our research focuses on the one-stroke TFSL, and classifies similar dialects using PHOG and local features with KNN.

2 Related Work

The barriers incurred through the lack of recognition and understanding of finger-spelling sign recognition language have prompted a great deal of research into improving today's existing systems. The Vision based technique is commonly used in the development of Thai finger-spelling sign language recognition systems. Pariwat, et al. [5] proposed a Thai finger-spelling sign language recognition system, which uses global and local features with SVM. Their experiment results revealed average accuracies for RBF (91.2%), linear (86.4%), polynomial (80%), and sigmoid (54.7%) in the amalgamation of global and local features. Chansri, et al. [9] presented the Kinect Sensor process, based on HOG and neural networks to develop a novel TFSL recognition system. The distances tested demonstrated recognition accuracies of 72.92% at 1.2m, 81.25% at 1.0m, and 88.33% at 0.8m. Adhan, et al. [14] presented a recognition system involving the geometric invariant feature, as well as ANN. The purpose of this study was to translate the gestures in TSFLR. Two-dimensional image recognition, based on an evolved geometric invariant feature, was used to translate their 42 Thai sign language letters into the Thai alphabet. This system also presented a two-layer feed-forward neural network. In our experiment, we employed the use of gloves with six different color markers. The results of the TSL 42-alphabet recognition system provided an average accuracy of 96.19. Silanon [15] proposed alternative ways in which to choose the best classifiers, in order to create a strong classifier within the TFSLR system. Its variable was a cascaded classifier, based on HOG features and an adaptive boost (i.e., AdaBoost). Ten hand signers expressed 21 static hand posture images, in which to test and train within our classification tests, achieving an approximate accuracy of 78%. Furthermore, the framework of a finger-spelling sign language for the Arabic alphabet was the study of recognition based on three descriptors: geometric hand features, Hu moment, and discrete orthogonal Tchebichef moment. The recognition rates were determined for both KNN (89.35%) and SVM (86.90%) [13].

Several researches have therefore developed approaches using vision-based sign language recognition systems. Auephanwiriyakul, et al. [4] proposed a unique system of Thai sign language translation with the employing both SIFT an HMM. The signer-dependent tests achieved an average accuracy of 86-95%, while the signer-semi-independent tests achieved 79.75%, and the signer-independent tests scored 76.56%. Lim, et al. [10] proposed a feature covariance matrix with a practical filter for isolated sign language recognition. The experiments yielded an 87.33% accuracy rate in ASL. Yang, et al. [11] proposed a unique form of continuous sign language recognition, in which gestures were stated in terms of sequence segmentation and recognition, comprising two major components of level building, based on a fast-hidden Markov model. Over 100 sentences exist in their Chinese sign language dataset by KINECT, and their library consisted of five signs each. Their experiment results expressed unmatched recognition effectiveness with low computation times. They determined that the seconds of runtime per sentence was 8.98, with an error rate of 12.20%; which were compared with several existing techniques.

3 The Proposed Method

The method developed in this study consisted of both a training phase and a testing phase. The training process is comprised of hand segmentation, the creation of feature vectors using PHOG and local feature extraction, and the creating of a KNN model. Both the testing and training processes begin with hand segmentation and the creation of feature vectors based on KNN recognition. Within the training phase, feature vectors eventually generated the KNN classification model. The KNN model was used to identify characters for text display. The process of each phase is presented in Fig. 2.

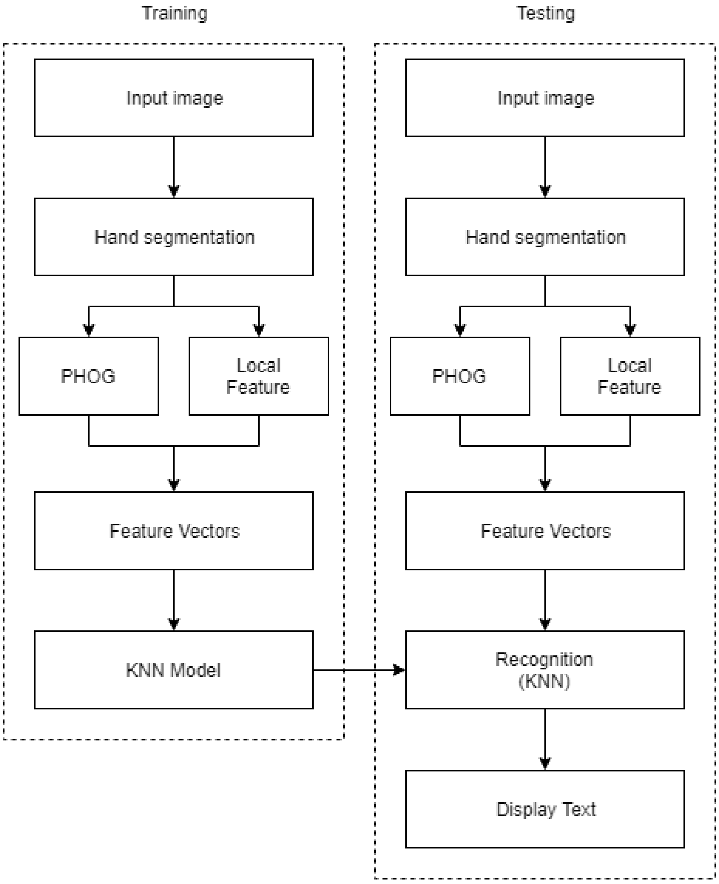


Fig. 2. Thai finger-spelling sign language recognition system framework.

3.1 Input image

This research examined input images (280 x 288 pixels) in RGB format, captured with a digital camera. Each signer was standing in front of a blue background, wearing a long-sleeved black shirt.

3.2 Hand segmentation

Hand segmentation is the process of distinguishing hands from the background. Following hand segmentation, the process of cutting a hand region of interest (ROI) was undertaken, through the following steps (Fig.3).

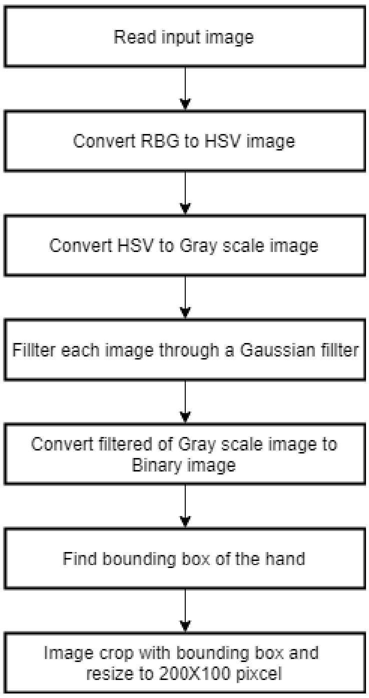

Read input image
Convert RBG to HSV image
Convert HSV to Gray scale image
Fillter each image through a Gaussian fillter
Convert filtered of Gray scale image to Binary image
Find bounding box of the hand
Image crop with bounding box and resize to 200X100 pixcel

Fig. 3. The steps of hand segmentation.

3.3 Feature extraction

This research focuses on the similar characters of the one-stroke, Thai finger-spelling sign language. It was therefore necessary to use feature extraction in both PHOG and local features, allowing us to characterize and differentiate characters having similar features.

3.3.1 Pyramid histogram of oriented gradients (PHOG) and the histogram of oriented gradients (HOG)

Anna Bosch developed PHOG feature extraction in 2007 [7], as an improvement of HOG, performed by the following steps, illustrated in Fig. 4.

Step 1: A hand region of interest (ROI) is created using canny edge detectors for extracting edge contours.

Step 2: The segmentation of images of the edge contours are transferred into cells on the k-level of the pyramid, written as $C_k = 2^k$.

Step 3: Each level of the pyramid is calculated via HOG to obtain the local shape features.

Step 4: The HOG feature vectors of each pyramid level are generated form the PHOG features.

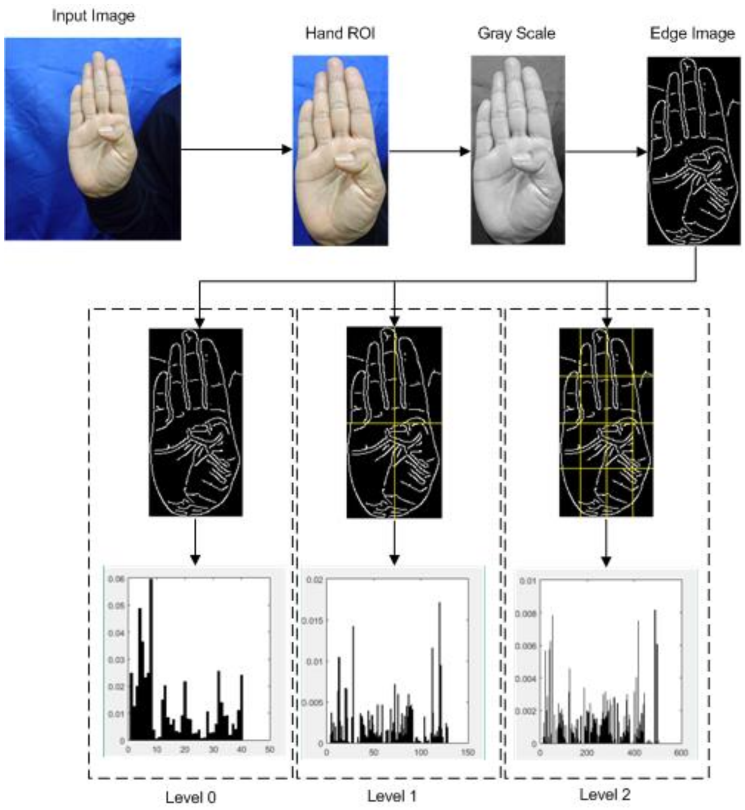


Fig. 4. PHOG feature descriptors of the finger-spelling sign language.

HOG is a feature descriptor used in computer vision and image processing, which describes the appearance and shape of a local object within an image, through the distribution of intensity gradients or edge directions. These descriptions are derived by dividing images into small pieces, or named cells, which integrate a histogram of gradient directions for pixels within the cells [7]. HOG is performed through the following procedures:

Step 1: Calculate the horizontal and vertical gradients via Sobel, through Equations 1 and 2.

$$G_x = M_x * I = [-1\ 0\ 1] \tag{1}$$

$$G_y = M_y * I = [-1\ 0\ 1]^T \tag{2}$$

Where G_x and G_y are the first derivatives of the image, horizontally and vertically, respectively

Step 2: Find the magnitude and direction of each gradient, through the following formula:

$$|G(x,y)| = \sqrt{G_x\ (x,y)^2 + G_y(x,y)^2} \tag{3}$$

Calculate the direction of the gradient in each image cell at positon *(x,y)*, in the angle range 0-180°, referred to as unsigned gradients; and 0-360°, referred to as signed gradients, as follows.

$$\theta(x,y) = arctan\left(\frac{G_y}{G_x}\right) \tag{4}$$

Where |*G(x,y)*| is the magnitude of the gradient, and *θ (x,y)* is the direction of the gradient.

Step 3: Calculate the histogram of gradients in 8x8 cells, through the following equation.

$$C_b = \sum_{i=1}^{n} q(x,y)_i = b \tag{5}$$

Where C_b is the summation of the orientation bin, *q(x,y)* is the angle of gradient vectors at *(x,y)*, b is the orientation bin considered, and n is the number of the position *(x,y)* in each cell.

Step 4: Calculate the HOG feature vector, through the following equations.

$$V_k = \sum_{i=1}^{n}(|G(x,y)_k| * C_b) \mid q(x,y) = b \tag{6}$$

$$V_k = \begin{bmatrix} V_{k=1} \\ V_{k=2} \\ \vdots \\ V_{k=b} \end{bmatrix} \tag{7}$$

Where V_k is the HOG feature vector of image, and *k* is the number of blocks.

The feature vector of each hand varies, depending on the arrangement and direction of the finger.

3.3.2 Local Features

This research applies local features for describing the description of the inner hand, and utilizes the density along with Hu Moment feature vectors [5]; as explained in the following process.

1) The density, the numerical value of ‘1’ pixel in every single box, was considered from the box size in 4x4 inch segments of each edge image. Fig. 5 outlines the algorithm of the density determination process. Fig. 6 depicts the edge image split to 4x4, as shown in the following example.

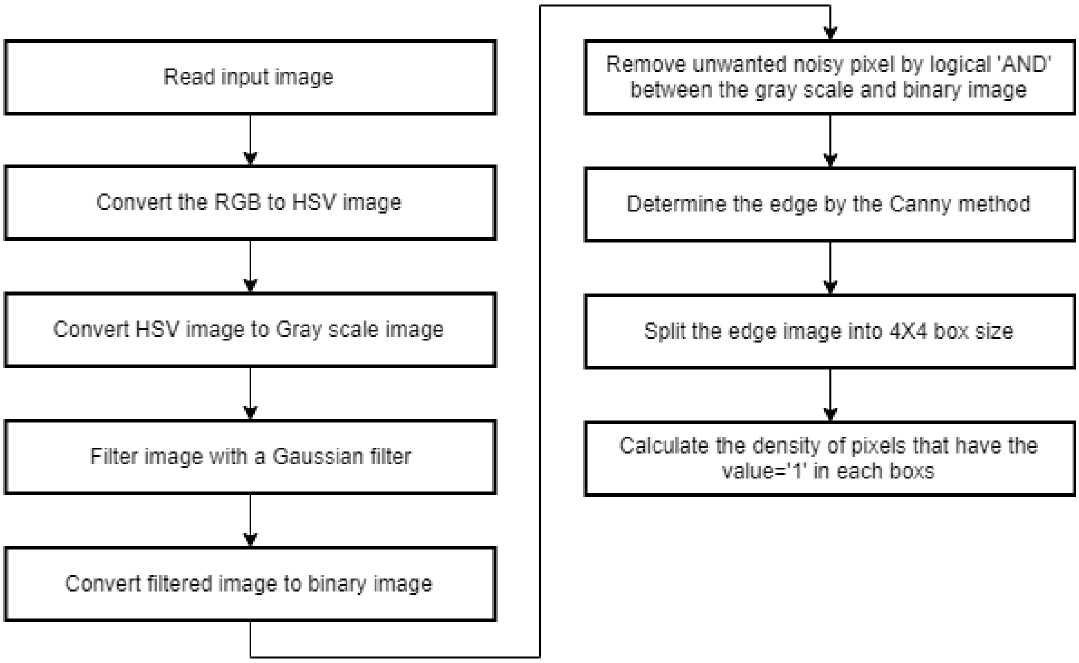
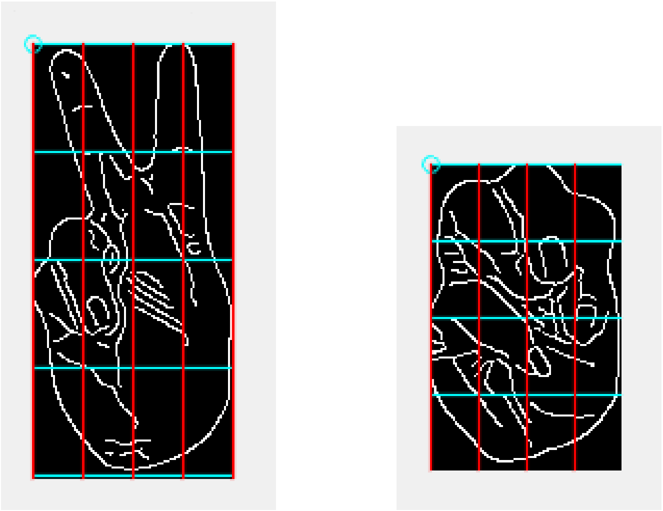

Read input image
Convert the RGB to HSV image
Convert HSV image to Gray scale image
Filter image with a Gaussian filter
Convert filtered image to binary image
Remove unwanted noisy pixel by logical 'AND' between the gray scale and binary image
Determine the edge by the Canny method
Split the edge image into 4X4 box size
Calculate the density of pixels that have the value='1' in each boxs

Fig. 5. The steps of the density feature.



(a) (b)

Fig. 6. Edge image the (a) ‘ก’ (/k/) sign, (b) ‘ม’ (/m/) signs, split to 4x4.

2) The concept of seven moment invariants introduced by Hu, which remains unchanged under image scaling, translation, and rotation [12]. The seven moment invariants are defined in Equation 8, as follows:

$$
\begin{aligned}
\emptyset_1 &= \mu_{20} + \mu_{02} \\
\emptyset_2 &= (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \\
\emptyset_3 &= (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2 \\
\emptyset_4 &= (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \\
\emptyset_5 &= (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})\left[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2\right] \\
&\quad +(3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \\
\emptyset_6 &= (\mu_{20} - \mu_{02})\left[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2\right] \\
&\quad +4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03}) \\
\emptyset_7 &= (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})\left[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2\right] \\
&\quad -(\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2]
\end{aligned}
\tag{8}
$$

Where ϕ_1-ϕ_6 = six absolute orthogonal invariants, ϕ_7 = the shew orthogonal invariant, μ_{pq} = the normalized central moment, and *pq* = the order of central moments up to three.

3.4 Recognition

We employed KNN to classify the feature vector data in our research. The KNN is a simple algorithm that maintains all available cases, and distinguishes new variant cases based on similarities. It is useful in both statistical estimation and pattern recognition. Case classification is implemented by a majority vote of its neighbors, which assigns it to the class most common amongst its K-nearest neighbors, if *k* = 1, *k*=3, and *k*=5; the case would be simply assigned to the class of its nearest neighbor, measured by the Euclidean distance function, shown in Equation 9.

$$D = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \tag{9}$$

Where *D* is Euclidean distance, and *(x,y)* is Euclidean vector.

The KNN classifier choice was introduced in the research of Dahmani, et al. [13]; whose sign language finger-spelling recognition system demonstrated that the KNN classifier is usually successful when each class has several possible prototypes, as well as an irregular decision boundary.

4 Experiment and Results

4.1 Data Collection

The five signers employed within this experiment consisted of two professional signers, one hearing impaired signer, and two non-professional signers. The dataset consisted of 15 classes, or 15 hand signs, in which the hand signs were repeated five times, for a total of 375 photos. An example of the 15 hand signs is given in Fig. 7.

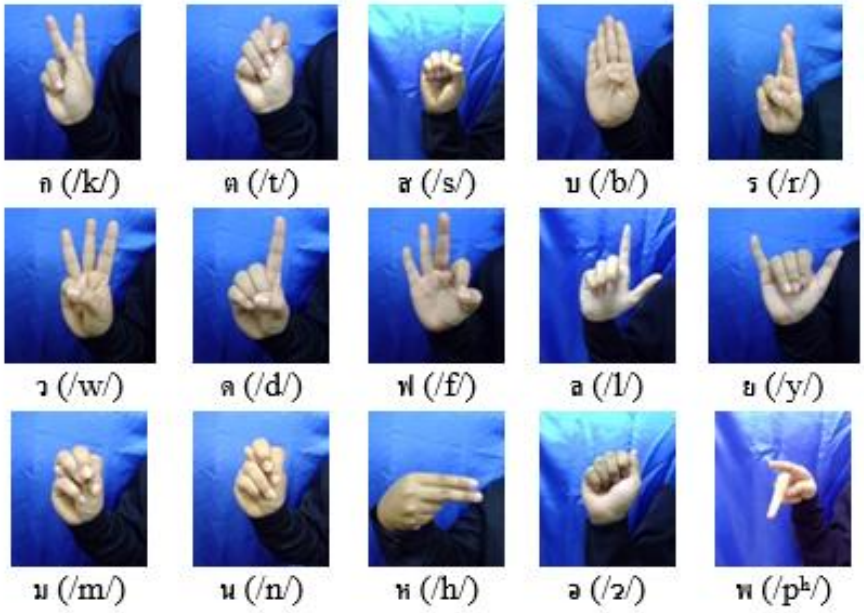


Fig. 7. The fifteen TFSL hand signs.

4.2 Experiment Results

Our experiments focused on user-independent testing, PHOG and local features, and KNN recognition, (k = 1). The PHOG parameters consisted of bin = 5, angle = 360°, and level = 5; and five-fold cross-validation was used to assess effectiveness.

Table 1 shows the experiment results of the five-fold cross validation of Thai finger-spelling sign language recognition, which compared the PHOG and local features with KNN, PHOG features with KNN, and global and local features with the RBF kernel of SVM. The average accuracy in of the PHOG descriptors and local features were combined, totaling 97.60%, the average accuracy of the PHOG descriptors with KNN was 97.33%, and the global and local features with the RBF kernel of SVM was 91.20% [5].

Table 1. Comparisons of the accuracy results of Thai finger-spelling sign language recognition systems.

Five-fold cross validation	PHOG + Local features with KNN (%)	PHOG with KNN (%)	Global and Local features with (RBF) SVM (%) [5]
1st	100.00	98.67	88.00
2nd	100.00	100.00	82.67
3rd	100.00	100.00	97.33
4th	98.67	98.67	93.33
5th	89.33	89.33	94.67
Average	97.60	97.33	91.20

5 Conclusion

This study examines a newly developed Thai finger-spelling sign language recognition system using PHOG and local features, employing a KNN classifier. We also compared the experiment results between the PHOG and the amalgamation of PHOG with local features. Five-fold cross validation was measured to determine its effectiveness. The amalgamation of PHOG and local features displayed an average accuracy of 97.60%. Moreover, the system can effectively distinguish between the very similar sign gestures within the finger-spelling sign language. In further study, we look forward to developing a system encompassing all 42 alphabet characters of this finger-spelling sign language.

6 Acknowledgement

This research was funded by the Computer Science Department, Faculty of Science, Khon Kaen University, Khon Kaen, Thailand. The authors would like to extend our thanks to Sotsuksa Khon Kaen School for their assistance and support in the development of the finger-spelling sign language data.

7 References

[1] Pansare, J. R., & Ingle, M. (2016, August). Vision-based approach for American sign language recognition using edge orientation histogram. In *2016 International Conference on Image, Vision and Computing (ICIVC)* (pp. 86-90). IEEE.

[2] Tripathi, K., & Nandi, N. B. G. (2015). Continuous Indian sign language gesture recognition and sentence formation. *Procedia Computer Science*, 54, 523-531.

[3] Lei, L., & Dashun, Q. (2015, July). Design of data-glove and Chinese sign language recognition system based on ARM9. In *2015 12th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)* (Vol. 3, pp. 1130-1134). IEEE.

[4] Auephanwiriyakul, S., Phitakwinai, S., Suttapak, W., Chanda, P., & Theera-Umpon, N. (2013). Thai sign language translation using scale invariant feature transform and hidden markov models. *Pattern Recognition Letters*, 34(11), 1291-1298.

[5] Pariwat, T., & Seresangtakul, P. (2017, February). Thai finger-spelling sign language recognition using global and local features with SVM. In *2017 9th international conference on knowledge and smart technology (KST)* (pp. 116-120). IEEE.

[7] Sugiharto, A., & Harjoko, A. (2016, October). Traffic sign detection based on HOG and PHOG using binary SVM and k-NN. In *2016 3rd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)* (pp. 317-321). IEEE.

[8] Roy, P. P., Bhunia, A. K., Das, A., Dey, P., & Pal, U. (2016). HMM-based Indic handwritten word recognition using zone segmentation. *Pattern Recognition*, 60, 1057-1075.

[9] Chansri, C., & Srinonchat, J. (2016, June). Reliability and accuracy of Thai sign language recognition with Kinect sensor. In *2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)* (pp. 1-4). IEEE.

[10] Lim, K. M., Tan, A. W., & Tan, S. C. (2016). A feature covariance matrix with serial particle filter for isolated sign language recognition. *Expert Systems with Applications*, 54, 208-218.

[11] Yang, W., Tao, J., & Ye, Z. (2016). Continuous sign language recognition using level building based on fast hidden Markov model. *Pattern Recognition Letters*, 78, 28-35.

[12] Huang, Z., & Leng, J. (2010, April). Analysis of Hu's moment invariants on image scaling and rotation. In *2010 2nd International Conference on Computer Engineering and Technology* (Vol. 7, pp. V7-476). IEEE.

[13] Dahmani, D., & Larabi, S. (2014). User-independent system for sign language finger spelling recognition. *Journal of Visual Communication and Image Representation*, 25(5), 1240-1250.

[14] Adhan, S., & Pintavirooj, C. (2016, December). Thai sign language recognition by using geometric invariant feature and ANN classification. In *2016 9th biomedical engineering international conference (BMEiCON)* (pp. 1-4). IEEE.

[15] Silanon, K. (2017). Thai finger-spelling recognition using a cascaded classifier based on histogram of orientation gradient features. *Computational intelligence and neuroscience*, 2017.