Enhancing Hand Gesture Recognition using Fuzzy Clustering-based Mixture-of-Experts Model

Jong-Won Yoon  
Yonsei University  
134 Shinchon-dong, Seodaemoon-gu, Seoul, Korea  
+82-2-2123-4803  
jwyoon@sclab.yonsei.ac.kr

Jun-Ki Min  
Yonsei University  
134 Shinchon-dong, Seodaemoon-gu, Seoul, Korea  
+82-2-2123-4803  
loomlike@sclab.yonsei.ac.kr

Sung-Bae Cho  
Yonsei University  
134 Shinchon-dong, Seodaemoon-gu, Seoul, Korea  
+82-2-2123-2720  
sbcho@cs.yonsei.ac.kr

ABSTRACT
Hand gestures have been widely applied to interface as the way of interaction between human and computers. Since a human hand can express various shapes of gestures, previous models for recognizing them cannot distinguish them accurately since they use only single model for recognition. For efficient hand gesture recognition with its enhanced performance, we propose the fuzzy c-means clustering based mixture-of-experts (FME). The proposed method uses multiple local experts obtained via fuzzy c-means clustering and decisions from them are combined with the gating network. To evaluate the performance of the proposed method, we conduct experiments including comparisons with alternative models for hand gesture recognition. As the result of experiments, the proposed model shows improved gesture recognition performance, especially performance on similar hand gesture recognition.

Categories and Subject Descriptors
I.5.2 [Design Methodology]: Classifier design and evaluation

General Terms
Algorithms, Performance, Experimentation

Keywords
Hand gesture recognition, mixtures-of-experts, fuzzy-c-means clustering

1. INTRODUCTION
Hand gestures are one of the most common forms of expressive gestures, and they have great potential to act as a computer interface through the development of the data glove [1-3]. To implement hand gesture-based interfaces, it is required to recognize the shape of human hand from sensory inputs accurately. This can be regarded as one of pattern recognition problems, and various techniques have been adopted. Hand gesture recognition has a major issue, the complexity problem. Since a human hand consists of number of joints, it can express various possible shapes which may include similar ones. It is obvious that the hand gesture-based interface system should recognize various gestures as accurately as possible. However, due to the high degrees of freedom of human hand and diversity of possible gestures, it is difficult to distinguish them precisely [4].

Many researchers have paid their attention to the complexity issue of hand gesture recognition, and tried to implement recognition models to distinguish various gestures accurately with diverse machine learning techniques. However, because they mainly have used only a single model, it is still difficult to recognize various gestures accurately and more difficult to distinguish similar ones.

In order to improve the performance of hand gesture recognition, we focus on the two difficulties, diversity of possible hand gestures and similarity in them. Both of them occur because of the high-complexity of human hands, and the most efficient way to solve the difficulties is to reduce the complexity of hand gesture recognition problem.

One of the ways to reduce the complexity of problems is to use several local experts which are specialized in different parts of the entire problem. This approach is useful for hand gesture recognition as well. Because each local expert is specialized in only a part of all the possible gestures, the variety allocated to each expert is less than that of entire gestures.

In this paper, we propose to use the mixtures-of-experts (ME) model to recognize various hand gestures efficiently. The proposed method separates whole hand gestures into several groups, and generates experts for each group which classify specific ones. Decisions from experts are combined by using a gating network to make final decision of recognitions. Fuzzy c-means clustering (FCM) is used to group whole gestures, and each expert and the gating network are implemented by using multi-layered perceptrons (MLPs).

2. BACKGROUNDS
2.1 Hand Gesture Recognition
Hand gestures have been widely used as a way of interaction for interfaces. As it is not easy to directly determine the shape of a hand based solely on the signals, machine learning techniques for hand gesture recognition have been required [5]. Many researchers have tried to recognize various hand gestures accurately, and lots of traditional pattern recognition techniques have been applied to this domain. Table 1 shows machine learning techniques and applications of previous hand gesture-based interfaces.

Recognized hand gestures have been used for various target applications, especially sign language recognition which contains a variety of hand shapes. However, even though previous works tried to apply machine learning techniques to the hand gesture recognition, a single model they used is not appropriate to
discriminate various gestures because it is difficult to learn all hand gestures at once due to the complexity problem.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Techniques</th>
<th>Target Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee et al. [7]</td>
<td>ANN</td>
<td>Taiwan sign language recognition</td>
</tr>
<tr>
<td>Fels and Hinton [8]</td>
<td>ANN and RBF</td>
<td>Speech synthesizer</td>
</tr>
<tr>
<td>Ziaie et al. [9]</td>
<td>NB + KNN</td>
<td>Human-robot interaction</td>
</tr>
<tr>
<td>Kamel et al. [10]</td>
<td>SVD</td>
<td>Online signature verification</td>
</tr>
<tr>
<td>Kelly et al. [11]</td>
<td>SVM</td>
<td>Irish sign language recognition</td>
</tr>
</tbody>
</table>

Table 1. Previous hand gesture recognition techniques

To overcome this limitation caused by high-complexity of the hand gesture recognition which has not been solved yet with a single model in previous works, we mainly focus on reducing the complexity of the recognition, and the mixture-of-experts model, especially based on fuzzy c-means clustering, is used to deal with the complexity in this work.

2.2 Mixtures-of-Experts
Mixture-of-experts models [12] consist of a set of experts and a gating network which combines the decisions from experts. One motivation for the mixture-of-experts model is based in the divide-and-conquer principle, which is common to the field of computer science. By this principle, certain complex problems can be decomposed into a set of relatively simple sub-problems. In the mixture-of-experts model, the assumption is that there are separate problems within the larger underlying problem. Modeling these smaller sub-problems is performed by the experts, while the decision that which expert will be used is modeled by the gating network.

According to the divide-and-conquer principle, the mixture-of-experts model should work well for problems that are composed of smaller unconnected ideas. Each expert deals with different features from a different perspective, thereby resolving the small separable problems.

Through the features of the mixture-of-experts model, it has been applied to traditional complicated recognition problems [13,14], and it is also suitable for the hand gesture recognition problem we aim to solve. The complexity of the problem can be reduced by dividing all hand gestures into some subgroups.

One of the considerations when implementing ME models is the way to separate the decision surface and generate experts. In order to divide the decision surface into several subgroups, clustering techniques can be applied. Some hard clustering techniques, e.g., $k$-means clustering, have been widely used. However, it is not easy to define crisp boundaries between several hand gestures to generate experts since there are several similar ones, and some gestures should belong to two or more groups. Thus, hard clustering techniques are not fit to be applied to the domain of hand gesture recognition. Due to this limitation, we propose to use FCM which has soft boundaries to create ME models. FCM was already applied to ME for classification problems [15]. However, the work used FCM only for labeling unsupervised data and the ME model was rarely affected by it. The proposed method can generate ME models more adaptive to the problem since the models are decided by considering the result of FCM.

3. THE PROPOSED METHOD
The main idea of the proposed method, fuzzy c-means clustering based mixtures-of-experts (FME), is to divide whole hand gestures into several groups that each group contains similar ones to train local experts specialized to specific groups.

The proposed method consists of two phases, the training phase and the recognition phase. Figure 1 shows a procedure of the proposed method.

In the training phase, the ME model for hand gesture recognition is generated from training data set. The model consists of $N$ experts and a gating network. Figure 2 shows the model used in the proposed method. The training data set is clustered by using
FCM and used to generate local experts. The gating network is trained with generated experts and all training data.

In the recognition phase, for any input vector \( x \), the model obtains \( N \) local decisions from experts and makes final decision by combining local decisions from experts with an output vector from the gating network. \( M(x) \), the set of local decisions from total \( N \) experts for input \( x \) is defined as below:

\[
M(x) = \{ \mu_1(x), \mu_2(x), \mu_3(x), ..., \mu_N(x) \},
\]

where \( \mu_i(x) \) represents the output of the \( i \)th expert with any given input \( x \). \( G(x) \), the output of the gating network which represents weights of experts for input \( x \) is also defined as follows:

\[
G(x) = \{ g_1, g_2, g_3, ..., g_N \}.
\]

### 3.1 FCM based experts generation

Prior to apply the ME model to hand gesture recognition, local experts must be generated with separate data set. Each expert deals with only a part of data set.

When dividing entire hand gestures into several subgroups, each subgroup should guarantee similarities between members in it. This is required to give specialties to local experts in distinguishing some similar gestures even though they are only a part of entire gestures.

Since it is not easy to define crisp boundaries to determine similarities between gestures, the clustering technique which uses soft boundaries is required, such as a fuzzy clustering technique.

A fuzzy clustering approach is less likely to get stuck in the local minimum than a hard clustering approach because it makes soft decisions in iteration through the use of membership values. This approach can be useful for the proposed model.

The most widely-used fuzzy clustering algorithm is the FCM, proposed by Bezdek [16]. It generates a fuzzy partition that provides each piece of data with a degree of membership to a given cluster. The values of the degrees of membership lie between 0 and 1. Values close to 0 indicate the absence of strong association to the corresponding cluster, whereas values close to 1 indicate strong association to the cluster. Figure 3 shows the procedure of the FCM.

For entire training data set \( T = \{ (x_1, y_1), (x_2, y_2), ..., (x_R, y_R) \} \) which contains total \( R \) pairs of an input and a desired output, membership values between training data instances and \( K \) clusters are obtained as the result of FCM. With the membership matrix, the training data set for the \( i \)th expert \( T_i^e \) is defined as below:

\[
T_i^e = \{ (x_j, y_j) \mid u_{ij} > 0, (x_j, y_j) \in T \}.
\]

The \( j \)th training data instance which has the membership value for the \( i \)th cluster that is greater than 0 is assigned to expert \( i \). Even though each instance has the membership value between 0 and 1 for each cluster, generally, it does not have the value greater than 0 for all clusters. Therefore, each expert has only a part of whole training instances. The membership values of the instances in the same cluster may be different, but they are treated equally as one of the training data set for the corresponding expert in experts generation step. The differences between the values are reflected in the gating network after generating experts.

1) Determine the number of clusters \( c \) and the fuzziness parameter \( m \)
2) Initialize the membership matrix \( u_i \) satisfying the condition:

\[
\sum_{i=1}^{c} u_{ij} = 1, \quad 1 \leq j \leq n
\]

3) Compute centroids \( v_i (i = 1,2, ..., c) \):

\[
v_i = \frac{\sum_{j=1}^{n} u_{ij}^m x_j}{\sum_{j=1}^{n} u_{ij}^m}
\]

4) Compute membership values matrix \( U \):

\[
u_i = \left( \frac{1}{d^m(x_i, v_i)} \right)^{\frac{1}{m-1}} \sum_{j=1}^{n} \frac{1}{d^m(x_j, v_i)}
\]

5) Compute the objective function \( J_m \):

\[
J_m(X, U, V) = \sum_{i=1}^{n} \sum_{j=1}^{c} (u_{ij})^m d^2(x_i, v_i)
\]

6) Repeat 3) to 5) until stabilized as:

\[
\left| f_m^{(i)} - f_m^{(i-1)} \right| \leq \varepsilon
\]

### Figure 3. The fuzzy c-means clustering algorithm

MLP is used in this work to implement local experts. Each expert is learned with back-propagation (BP) algorithm as shown below:

\[
\Delta w_{a} = \eta \delta_{a} \mu_{i} (x_j)^T, \quad \delta_{a} = \mu_{i} (x_j) \left( 1 - \mu_{i} (x_j) \right) (y_j - \mu_{i} (x_j)), \quad \delta_{h} = \mu_{ih} (x_j) \left( 1 - \mu_{ih} (x_j) \right) \Delta w_{a}, \quad \Delta w_{h} = \eta \delta_{h} \mu_{i} (x_j)^T, \quad \delta_{h} = \mu_{ih} (x_j) \left( 1 - \mu_{ih} (x_j) \right) \Delta w_{a}, \quad \Delta w_{h} = \eta \delta_{h} \mu_{i} (x_j)^T
\]

where \( w_a \) and \( w_h \) represent the weights of links to output nodes and hidden nodes, respectively, \( \mu_{i} (x) \) is outputs from the hidden nodes of the \( i \)th expert, and \( \eta \) is the learning rate. To train the \( i \)th expert, pair \( (x_j, y_j) \) which belongs to the training data set for the \( i \)th expert \( T_i^e \) is used to decide variations of all weights.

### 3.2 Decision combination

Each expert covers only a limited part of entire problem space and may give wrong decisions for other parts which were not learned. Therefore, for the input, it is required to decide that which expert is specialized to deal with the given input. This process is called decision combination.

All decisions from experts are combined by using the gating network. The role of the gating network is to decide degrees of reflections of local decisions from experts based on input data.
The gating network is also implemented by using MLP, and must be learned before using it. In order to decide degrees of reflections of local experts, errors of experts, $E_i(x_j)$, should be obtained first. $E_i(x_j)$, error of the $i$th expert for given training data instance $(x_j, y_j)$ is defined as below:

$$E_i(x_j) = |y_j - \mu_i(x_j)|, (x_j, y_j) \in T.$$  \hspace{1cm} (8)

After normalizing $E(x_j)$, we can obtain accuracies of experts by subtracting errors from 1. Accuracies are finally adjusted by multiplying membership values to reflect specialties in the given training data instance:

$$A_i(x_j) = \left(1 - \frac{E_i(x_j) - m_{h_i} \mu_i(x_j)}{m_{h_i} \mu_i(x_j) - m_{h_k} \mu_k(x_j)} \right) u_i, \ x_j \in T.$$  \hspace{1cm} (9)

Obtained $A(x_j)$ is regarded as a target output vector for the $j$th training data instance. As similar to training for experts, BP algorithm is used to train the gating network as below:

$$\delta_{o_g} = G(x_j) \left(1 - G(x_j) \right) \left(\tilde{A}(x_j) - G(x_j) \right), \ (x_j, y_j) \in T.$$  \hspace{1cm} (10)

$$\delta_{h_g} = G_h(x_j) \left(1 - G_h(x_j) \right) w_{o_g} \delta_{o_g}, \ (x_j, y_j) \in T.$$  \hspace{1cm} (11)

$$\Delta w_{o_g} = \eta \delta_{o_g} G_h(x_j)^T, \ x_j \in T.$$  \hspace{1cm} (12)

$$\Delta w_{h_g} = \eta \delta_{h_g} x_j, \ x_j \in T.$$  \hspace{1cm} (13)

where $w_{o_g}$ and $w_{h_g}$ represent the weights of links to output nodes and hidden nodes of the gating network, respectively. $G_h(x)$ is outputs from the hidden nodes of the gating network, and $\eta$ is the learning rate. Contrary to training processes for experts, entire training data set $T$ is used to train the gating network.

Trained gating network is used with local experts to make final decision in the recognition phase. For a given input $x$, the final decision $O(x)$ is obtained by using $M(x)$ and $G(x)$ as follows:

$$O_i(x) = G_i(x) M_i(x) = \sum_{i=1}^{N} \tilde{o}_i \mu_i(x).$$  \hspace{1cm} (14)

The class which has the highest value among the final decision vector $O(x)$ is chosen as the recognition result.

### 4. EXPERIMENTS

#### 4.1 Experimental setup

The experiments were performed to evaluate the performance of the proposed model with the data glove-based input method. The set of hand gestures shown in Figure 5 was used to evaluate the accuracy of the hand gesture recognition since it includes comprehensive hand shapes. The first 24 hand gestures were the American Sign Language (ASL) alphabet and the last hand gesture involved the unfolding of all five fingers to show the palm.

Data was collected from six persons between 23 and 32 years old. Total 4500 gestures were used to train the model, and 2250 gestures were used for testing.

In order to recognize hand gestures, we use 14 Ultra data glove from 5DT shown in Figure 4. Used data glove works at 60Hz and sensors measure bending amounts of hand joints with 14 sensors. Each gesture data represents sensory inputs from 14 sensors of the right hand. Figure 6 shows the data glove-based hand gesture interface with the proposed method used in the experiments.

![Figure 4. The data glove used in the experiments](image)

<table>
<thead>
<tr>
<th>Table 2. Spece of experts and the gating network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experts</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>Input nodes</td>
</tr>
<tr>
<td>Hidden nodes</td>
</tr>
<tr>
<td>Output nodes</td>
</tr>
<tr>
<td>Learning rate</td>
</tr>
<tr>
<td>Threshold</td>
</tr>
</tbody>
</table>

The proposed FME for hand gesture recognition was implemented based on the training data. Table 2 shows the description of experts and the gating networks of the implemented model. To create the FME model, the number of experts and the fuzziness parameter were set to 7 and 1.3 respectively. To set the parameters, we tested the model with changing parameter values gradually and chose the parameters with the best performance of the model.

To compare the proposed method with alternative methods for hand gesture recognition, we also used Support Vector Machine (SVM), MLP, Naïve Bayes (NB) which are widely-used classifiers. For SVM, Radial Basis Function (RBF) is used as the kernel function, and the threshold was set to 0.001. For MLP, 50 hidden nodes are used, and the learning rate and the threshold are fixed to 0.3 and 0.005 respectively. Instead of implementing all alternative methods, we used Weka, the well-known library for machine learning techniques, which contains all of the methods for the experiments.

#### 4.2 Experimental Result

Before evaluating the recognition performance, we analyzed the result of expert generation to confirm whether FCM assigns training data into several experts with guaranteeing similarities between gestures in the same expert. Figure 7 shows the distributions of entire training data. The bar chart represents average membership values of each class for each expert.
As shown in the figure, some similar gestures show alike tendency of membership. For example, class 0, class 11, class 12 and class 18 shows similar tendency of membership that the membership value for the expert 6 is a bit high, and we can confirm that they have analogous shapes as shown in Figure 5. Moreover, class 1, class 5, and class 24 show similar tendencies that they mainly belong to expert 4, and their hand shapes are also alike that most of fingers are stretched and put together.

In order to highlight the outstanding performance of the proposed method, we compared the method with alternative recognition methods which use only a single model.

Table 3. Result of comparison experiment (unit : %)

<table>
<thead>
<tr>
<th></th>
<th>MLP</th>
<th>SVM</th>
<th>NB</th>
<th>FME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>92.85</td>
<td>93.31</td>
<td>90.43</td>
<td>96.50</td>
</tr>
<tr>
<td>Error</td>
<td>0.64</td>
<td>-</td>
<td>-</td>
<td>0.28</td>
</tr>
</tbody>
</table>

Figure 7. Average membership values of classes

Figure 8. Error rates of compared methods and FME
For the experiment, the performances of SVM, MLP, and NB were also obtained from the same data set, especially in the case of the proposed method and MLP, the average results of 10 trials for model generation were used. Table 3 shows the result of the experiment.

As shown in Figure 8, FME showed the best performance among various hand gesture recognition methods, and the performance was improved up to 6.13%. Moreover, despite FME uses MLP as a model for its local experts, the performance was 3.71% higher than the method with a single MLP model. This confirmed that using multiple local experts shows better performance than methods which use only a single recognition model.

To evaluate the performance in recognition especially for similar hand gesture sets, the recognition accuracies of all methods for some similar gesture groups were obtained as shown in Table 4.

<table>
<thead>
<tr>
<th>Similar gesture set</th>
<th>MLP</th>
<th>SVM</th>
<th>NB</th>
<th>FME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>78.89</td>
<td>77.22</td>
<td>79.44</td>
<td>90.00</td>
</tr>
<tr>
<td></td>
<td>80.06</td>
<td>80.06</td>
<td>83.33</td>
<td>92.77</td>
</tr>
<tr>
<td></td>
<td>94.44</td>
<td>84.07</td>
<td>91.11</td>
<td>97.22</td>
</tr>
<tr>
<td></td>
<td>94.81</td>
<td>85.92</td>
<td>96.67</td>
<td>96.67</td>
</tr>
</tbody>
</table>

As the result, FME showed the best performance for all similar gesture groups. Alternative methods, which use only single model for hand gesture recognition showed poor performances because of confusions caused when distinguishing similar gestures. FME showed up to 12.71% improved performance in distinguishing similar gestures than other methods since it uses local experts specialized to distinguish similar gestures. This result confirmed that the proposed FME is especially specialized in similar hand gesture recognition.

Even though we used FCM to build ME model, there are also other alternative ways to generate experts. One of the popular ways is a k-means clustering [17]. However, since the k-means clustering based ME (KME) uses crisp boundaries between experts, it is not appropriate to apply to hand gesture recognition problem and we already examined about this limitation at section 2.2. We conducted the comparison experiment with KME and FME to prove this experimentally by showing more superb performance of FME than KME.

The experts of KME were implemented same as FME, and the gating network was trained by using membership of each training data instance as the desired output. In order to compare performances of KME and FME with various number of experts for both methods, we changed the number of experts in both KME and FME from 5 to 10 denoted as KME5, KME6, ..., KME10, and FME5, FME6, ..., FME10.

Table 5 shows the result of comparison with KME6 and FME7, which show the best performance among KME-based methods and FME-based methods respectively. From the experimental result, we can see that FME shows 1.54% improved performance than KME. Additionally, Figure 9 shows error rates of KME and FME methods. We changed the numbers of experts for both methods from 5 to 10. In this figure, FME showed lower error rate than KME for any numbers of experts.

<table>
<thead>
<tr>
<th>KME6 (Best)</th>
<th>FME7 (Best)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>95.18</td>
</tr>
<tr>
<td>Error</td>
<td>0.27</td>
</tr>
</tbody>
</table>

In the proposed method, it is quite important to choose appropriate number of experts and the number of experts should be chosen according to the complexity of domain. Too small number of experts may not reduce the complexity of the problem, and too large number of experts can cause another problem that the local experts may be poorly trained because entire training...
data set is divided into too many groups and each group has only a few of training data instances.

In order to analyze relationships between the number of experts and performance, we observed recognition accuracies by changing the number of experts from 2 to 15. Figure 10 shows the result of the experiment.

With seven experts, the model showed the highest performance, and from 7 to 10 experts, the similar performances were obtained. However, with less than 6 experts, the performance was dropped as the number of experts was decreased. The performance was also dropped when the number of experts was greater than 10. This result shows that the number of experts between 7 and 10 can reduce the complexity of the problem with the hand gesture data set we used efficiently.

5. CONCLUDING REMARKS

In this paper, fuzzy c-means clustering based mixture-of-experts (FME) is proposed to solve the complexity problem in hand gesture recognition and provide enhanced performance. Since the hand gesture recognition is known as a complex problem, the method is aimed to divide the problem into some simple sub-problems, especially with soft boundaries using fuzzy c-means clustering (FCM). It generates experts for each subgroup, and makes decisions by combining outputs from experts and the gating network.

To evaluate the outstanding performance of the proposed method, we conducted experiments with hand gestures data set which contains total 25 gestures including ASL. As the result of the experiments, it was shown that the proposed FME gives higher performance than other alternative methods, particularly, excellent performance in distinguishing similar hand gestures.

In the proposed method, the performance can be changed depending on the gating network even though the same experts. Since we mainly focused on improving the way to generate experts by using FCM, there still remain problems of improvement of the gating network. In the future, various designs for the gating network should be applied in order to enhance the performance of the model. In addition, the methods to find optimal parameters for the model automatically should also be investigated in the future.

6. ACKNOWLEDGEMENTS

This research was supported by the Original Technology Research Program for Brain Science through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0018948), and the MKE (The Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2010-C1090-1021-0008)).

7. REFERENCES


