Multiple Filter Mask Learning of Feature Extraction in Character Recognition

Yoshiaki Kurosawa
Toshiba Solutions Co.

Abstract

An incremental learning for the feature extraction of OCR was applied to the learning of multiple filter mask in the feature extraction. The multiple filter mask means plural masks which are used for one local region of an input pattern. The output of the filter is provided to a recognition unit as a feature vector. In the recognition unit, distance is calculated by a reference vector and the feature vector. Both the reference vectors and the feature extraction filter masks are gradually changed by recurrence equations.

The purpose of this learning was similar character discrimination. The effectiveness of the learned masks was checked as an accuracy improvement. The learned masks were visually displayed and they were checked by human observation in the experiment. It was found that the masks had good performance in recognition accuracy, and that the masks were grown at the important area of a pattern for discrimination through the learning.

Keywords: OCR, Feature, Filter, Learning

1. Introduction

Feature extraction is one of the primal factors of pattern recognition systems, despite that discrimination processes tend to be emphasized in comparison. The problem of the feature extraction is that it needs an adhoc technique for each application, then it requires extensive human resources for development. One of the solutions to this problem is to introduce a new learning technique for creating an adequate feature extraction method automatically for each application.

This report describes a method to fix the filter mask values in the feature extraction of OCR as an example of such a learning technique. For this purpose, an evaluation function and steepest descent method were adopted as the learning method. This evaluation function is the same one which was introduced in the Probabilistic Descent Method (PD) [1] by Amari. This PD was based on a loss function and was improved in the speech recognition field as MCE/GPD (Minimum Classification Error Criterion / Generalized Probabilistic Descent) [2]. The PD is a concept by which Learning Vector Quantization (LVQ) is obtained. Therefore the method treated in this report is deeply related to LVQ. In other words, it is the extended version of LVQ including the feature extraction filter mask learning.

One of the characteristics of this learning system is that it has two feedback loops; the reference update and the feature extraction filter update in contrast to that LVQ has only the reference update. This structure with MCE/GPD was called as Discriminative Feature Extraction (DFE) [3].

There are several approaches in the feature extraction filter learning: Filter Parameter Learning [3] [4] [5], Filter Selector Learning [6] [7], Filter Mask Learning [8][9]. Among these approaches, the Filter Mask Learning (FML) was examined in this research because the filtering is the most popular technique for feature extraction in the OCR field and FML is used for the direct learning of the filter mask values.

The learning system adopted here resembles 3 layer neural nets [10]. The intermediate layer of those NN corresponds to the filter masks in the system adopted here. Also, Cognitron [11] is one of the related systems. There were other learning systems for feature extraction [12] [13] and they showed good performance against complex and difficult problems.

Through these previous works, the filter masks were not observed visually in many cases, and it was too complex to analyze the inner workings in some cases. The purpose of this research was to investigate the learning process of feature extraction filter masks, and the human visual observation of these generated masks.

This report describes Multiple Filter Mask Learning (MFML) which is the introduction of multiple masks into FML. This was successive research from the preliminary investigation [8][9] on the subject of FML. The concept of multiple masks itself is not new, but the effectiveness of multiple masks in FML had not been tested yet, and was the main purpose of this research.
2 Overview of the Learning System

2.1 Basic FML System

The basic recognition system used in this research is shown in Fig. 1. Local regions are assigned in an input image region and local filter masks are defined and correspond to each local region. They are arranged as a filter mask array shown in Fig. 1. Inner products are calculated as the elements of an output feature vector by the local region vectors which consist of pixel values and the local masks. A nonlinear conversion is executed after this inner products calculation. The output feature vector is provided to an Euclidean distance unit and a distance is calculated with a reference vector.

The learning process of FML consists of a reference vector update and a filter mask update. Both the reference vectors and the filter masks are gradually changed through learning by recurrence equations. The actual recurrence equations are shown below.

2.2 Recurrence Equations

Let $i$ be the number of the local region, $f^{(i)}$ be a vector as the local filter mask corresponds to the local region $i$, $z^{(i)}$ be an input vector, and $\hat{x}^i$ be a filter’s output value, then the filtering transformation is described by the inner product of $\hat{x}^i = f^{(i)} z^{(i)}$.

Let a nonlinear conversion be defined as $x_i = \rho(\hat{x}^i)$ where $x_i$ is the element of an output vector $x$. By this notation, the feature extraction from $z^{(i)}$ is described as $x_i = \rho(f^{(i)^T} z^{(i)})$. If the reference vector is denoted as $\varphi$, the Euclidean distance $S$ is described as

$$S = ||x - \varphi||^2.$$  

The recognition result is determined by the category whose reference has the minimum distance.

For the recurrence equations of $f^{(i)}$ and $\varphi$, $\Delta f^{(i)}$ and $\Delta \varphi$ are calculated from $f^{(i)}$ and $\varphi$. The vector $f^{(i)}$ and $\varphi$ are changed so that the system has better performance through this change. The changed filter masks represented by the vector $f^{(i)}$ are arranged and displayed to be observed visually as shown in Fig. 1.

The recurrence equations are described below. Let $l(d)$ be a loss function and $l'(d)$ represent the differential function. Here, $d$ is defined by $d = S_{ok} - S_{err}$ where $S_{ok}$ is the distance between the correct category’s reference and the input, and $S_{err}$ is the one between the other best category’s reference and the input. The recurrence equations are expressed as follows where $\varphi_i$ is the element of the reference vector $\varphi$.

$$\Delta f^{(i)} = \mp 2\epsilon_f l''(d)\rho'(\hat{x}^i)(x_i - \varphi_i)z^{(i)},$$
$$\Delta \varphi = \pm 2\epsilon_r l''(d)(x - \varphi).$$

Here, $\epsilon_f, \epsilon_r$ are the values which define the strength of the learning. The upper part of $\pm$ or $\mp$ is applied if an input category is the same as the reference’s one and the lower part was applied if they are different.

2.3 Multiple Filter Mask Learning (MFML)

The above mentioned basic learning system which had one filter mask for one local region was applied to a similar character discrimination in which two pairs of similar hand written characters were adopted as a target. A good performance was obtained through experiments of this case [9].

![Figure 1. Outline of recognition system used in FML.](image)

**Figure 1.** Outline of recognition system used in FML.

![Figure 2. The characters “mon-gamae” used in the experiment.](image)

**Figure 2.** The characters “mon-gamae” used in the experiment.

However, it was insufficient in the case of a similar hand written character group which is shown in Fig. 2 as an example. Fig. 2 shows the letters “mon-gamae” which have the same shape at the outside portion and have different shapes at the middle lower area.

To improve the performance for this kind of character group, the system was enhanced to have plural mask arrays as shown in Fig. 3. In this figure, (A) and (B) represent the same system. Just the description is different. In Fig. 3(A), the MFML is described as follows. The plural mask arrays are adopted and the inner products obtained from these mask arrays are provided as the elements of the output feature vector. If the number of mask arrays is $p$ and the number of masks in the array is $q$, the output vector dimension becomes $pq$.

In Fig. 3(B), a filter unit is introduced for one local region. It has plural masks and the output is the plural...
inner product values. The overall filtering system is composed of this filter unit arranged as an array. (not shown in Fig. 3(B)). In this case, the array is called a filter unit array.

3 Environment of Experiments

The input character image size was normalized to be a 40x40 pattern and was provided to the feature extractor. The mask size was 7x7 and the masks were located at intervals of 3 pixels for the horizontal and vertical directions. The output feature vector size was 14x14.

The initial values \( f_{x,y} \) of the filter masks are given by

\[
f_{x,y} = A \cdot \exp\left\{-\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) \right\},
\]

where \( x, y \) is the pixel position and \( A \) is a parameter to determine the strength.

The value \( \hat{x}_i \) is obtained as an inner product of the mask values and the input pattern’s pixel values in the local region. Then it is converted by a nonlinear function. The function adopted here is

\[
\rho(\hat{x}_i) = \frac{1}{1 + e^{(b-a\hat{x}_i)}}.
\]

A sigmoid function is adopted as the loss function and it’s differential forms are

\[
l(d) = \frac{1}{1 + e^{-\delta d}},
\]

\[
l'(d) = \delta \cdot l(d) \cdot (1 - l(d)).
\]

The group of similar characters shown in Fig. 2 were used in the experiments. There were 16000 patterns for learning and 10476 for testing.

4 Gaussian Mask Experiments

4.1 The Learning of a Single Mask (FML)

Fig. 4(A) shows the initial masks used in this experiment. The masks are arranged with 14 pieces horizontally and 14 pieces vertically. Each mask has 7x7 pixels. The pixel values are represented by darkness; darker pixels have higher values. The value zero corresponds to darkness observed at the rim of the masks. Here, all masks are the same in Fig. 4(A). Actual values were calculated by Eq. 4 with \( \sigma^2 = 0.8, \sigma^2_y = 0.8 \) and \( A = 0.2 \).

The case of a single mask and the case of four masks in a filter unit were examined as the experiment. Fig. 4(B) shows the learned mask array of the single mask case. As for Fig. 4(B), the contrast was lowered for easy observation. The actual masks had higher contrast compared to Fig. 4(A).

We can see the change of masks inside of the array and the small change in the outside area. The change of the masks appears where the learning patterns are different. This characteristic of the learning method was the same in the case of learning for the similar character pair [9].

The learning cycles were 5,760. Here, as for one learning cycle, all the learning patterns were used and each pattern was used once in one cycle.

4.2 The Learning of 4 masks (MFML)

Four mask arrays were used in the next experiment. If we use the same initial masks for all arrays, it is clear that the learning results of each array were all the same and it cause no improvement or change in the learning comparing to the case of a single mask. It needs to prepare different masks for each array as the initial masks.

A method adopted here to avoid this problem was the “delayed start” process. The first step of this process is to learn the 1st array only. The other arrays are held unchanged until the delayed start point which is predefined learning cycles. Then, in the 2nd step, the 2nd array is
added to the learning target in addition to the 1st array, and the learning continues until the second delayed start point. For the 3rd and the 4th, the same process is repeated.

The mask arrays obtained by this process is shown in Fig. 5. The learning cycles were 8,850 and the delayed start points were 1000, 2000, 3000 and 4000.

![Figure 5](image)

Figure 5. Mask arrays obtained by “delayed start.”

Two or three directions are observed in the masks of columns 4, 5, 7 and 8. It means that plural direction sensitive masks were created in the filter unit. But these masks are uneven and noisy. Sometimes the direction appears and sometimes not. It seems insufficient compared to an artificial filter mask like the direction sensitive filter in the Gabor filter. However, it is important that the method has the ability to make these.

The intensity of the masks are weak for the late start mask array. These 4 arrays are unbalanced in their intensity. It means that the influence of the initial masks remains. This unbalance could be corrected by making the learning cycles uniform. But the important point is the fact that the system tends to be influenced by the initial masks. Therefore, the initial masks are important in the design of this learning system.

4.3 Recognition Accuracy

Table 1. Error rate of MFML with Gaussian initial 4 masks

<table>
<thead>
<tr>
<th>Array</th>
<th>RL</th>
<th>FML</th>
<th>MFML</th>
<th>GC</th>
<th>GB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error</td>
<td>29.5%</td>
<td>18.4%</td>
<td>13.3%</td>
<td>13.9%</td>
<td>10.9%</td>
</tr>
</tbody>
</table>

The recognition accuracy achieved by this learning experiment is shown in Table 1. In the table, “Array” means the number of masks in a filter unit. “Error” is the error rate. There was no rejection in this experiment. “RL” means that the reference was only learned with the initial Gaussian mask. “RL” is equivalent to the learning of LVQ (Learning Vector Quantization) [14] with this Gaussian mask. “FML” and “MFML” mean the error rate which is obtained by the recognition unit in FML and MFML.

“GC” means the result of LVQ with a direction sensitive filter mask which is the same as the mask used in the Gabor filter [15]. This mask was made by the same technique as the Gabor filter’s mask. Only a cosine function was used in this initial mask in contrast to the sine and cosine functions used in the Gabor filter. Also, the non linear function which was used in the Gabor filter as a square root was not used in “GC.”

The equation for calculating mask values is as follows.

\[
f_{x,y} = A \cdot \exp\{-\frac{1}{2}\left(\frac{r_x^2}{\sigma_x^2} + \frac{r_y^2}{\sigma_y^2}\right)\} \cdot \cos(2\pi \frac{r_x}{u}),
\]

\[
r_x = x \cos \theta_t + y \sin \theta_t,
\]

\[
r_y = -x \sin \theta_t + y \cos \theta_t,
\]

where \(x, y\) is the pixel position, \(\theta_t\) defines the angle of
direction and \( A \) is a parameter to determine the strength. This is called the GC mask hereafter.

“GB” means the result of LVQ with the Gabor filter with the same equation as Eq. 7 with the sine function.

The difference between “RL” and “MFML” shows that the learning was effectively done. The result of “MFML” showed good performance compared to the other method, except for LVQ with the Gabor filter (“GB”). It means that MFML could make masks which were equal to direction sensitive GC masks. This proved that MFML with the Gaussian initial mask and a delayed start was effective to some extent. Further improvement could be possible if the uneven characteristics were improved.

5 Direction Sensitive Mask Experiments

5.1 The Learning Results

The next question is, “Is it possible to make a better masks from an even direction sensitive mask using MFML?” The experiment shown here was done for this purpose.

As an initial filter, the direction sensitive GC mask (Eq. 7) was adopted here.

The actual parameters were \( \sigma_x^2 = 5.0, \sigma_y^2 = 20.0, u = 7.0 \) and \( A = 0.3 \). This equation is the same as in the Gabor filter but the size of the mask is not large compared with the cosine’s cycle so that it has only one peak at the center. These masks represent 4 directions and their actual pattern are shown in Fig. 7.

The mask arrays obtained by this process are shown in Fig. 8. The learning cycles were 15,240.

The center and lower area were strongly changed and emphasized from the initial state. In contrast to that, the right, left and top areas remained similar to the initial state. This was all the same for each of the 4 mask arrays.

5.2 Recognition Accuracy

The learning and recognition simulation was done in the specifications of 2 directions, 4 directions and 8 directions. The results are shown in Table 2. In this table, “Array” means the number of mask arrays. “RL” means the reference learning with no mask learning. “GB” means “Gabor+LVQ.”

The comparison between “RL” and “MFML” showed that the learning was effective. The large number of filter masks showed good performance. However, the results from 8 arrays was not good enough for its large number. The results of “MFML” exceeded the results of “GB.” This means that the created masks were almost equal to the Gabor filter if LVQ is adopted as the recognition system. It can be said that the learning was successful in this basic examination.

The one learning cycle needed approximately 20 seconds with 3.4G PenIV PC.

5.3 Recognition Accuracy when a recognition unit was other than LVQ

Finally, the effectiveness of this filter was examined when it was applied to a recognition system other than LVQ. Here, SVM [16] was adopted as the recognition system to examine.

Table 3 shows the results. In this table, “MFML” means the recognition result of SVM with learned masks.
Table 3. Error rate obtained when learned masks are used as feature extraction and SVM for recognition

<table>
<thead>
<tr>
<th>Array</th>
<th>2</th>
<th>4</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFML</td>
<td>10.3%</td>
<td>8.1%</td>
<td>7.5%</td>
</tr>
<tr>
<td>GC</td>
<td>11.6%</td>
<td>9.1%</td>
<td>8.8%</td>
</tr>
<tr>
<td>GB</td>
<td>7.8%</td>
<td>5.0%</td>
<td>4.9%</td>
</tr>
</tbody>
</table>

“GC” means the recognition result of SVM with direction sensitive GC masks. “GB” means the result of SVM with a Gabor filter.

From these results, it can be said that the masks obtained by MFML had better performance than the direction sensitive GC masks, even if the recognition system was SVM.

On the contrary, “GB” was better than “MFML”. Gabor filter has 2 masks, sine and cosine, corresponding to the one mask of MFML, GB also has a square root function in the feature calculation unit, and is one reason for the difference between “GB” and “MFML”.

Basically, it is not appropriate to use a recognition method which is much different from the recognition method used in FML. However, in practical cases, it is possible that we adopt just such a case. The described experiment was done for this situation. In conclusion, several approaches for further improvement are needed and expected in such cases.

6 Conclusion

Through these experiments, MFML was proved to be effective in terms of the system’s recognition accuracy and the generated masks had good performance. The mask patterns prepared for human observation showed that the learning was focused on the important part for discrimination of similar characters.

But it needs more improvement if it is planned to have the learned filter embedded in the recognition system other than LVQ. One approach for improvement is to increase the complexity of the filtering system.

References


