

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Usefulness of Only One User's Handwritten Character on Offline Personal Character Recognition

Shinji Tsuruoka, Masahiro Hattori, Yasuji Miyake,
Haruhiko Takase and Hiroharu Kawanaka

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/53272>

1. Introduction

The variety of individual person should be respected in present-day life, and in character recognition the usage of characters written by the individual person is one of important problems. Handwritten character recognition is strongly required as a means of input to personal terminal machines such as smartphone, tablet PC and so on. One of problems on handwritten character recognition is low accuracy and the correct rate of the character recognition is not enough for user's request. To improve the accuracy, the characters written by one writer, who is called "*a specific writer*", are effective for simple characters such as alphabet, numerals and symbols in online system [1-5]. The specific writer's characters employed for on-line character recognition system. However, specific writer's characters are not employed on most offline commercial system.

We are considering the usage of character forms written by a specific writer to improve the recognition rate. The variety of character forms by five writers is shown in Fig. 1. The problem of the grouping the variety of character forms is that the distribution of characters for one category is wide, and that the boundary of the category would be not appropriate for character recognition. We think that the one specific writer would write the similar character forms, and that the distribution of the character by the specific writer is narrower than many writers.

We proposed some personal recognition dictionaries (a pure personal dictionary and three adaptive dictionaries) generated from many characters written by one specific writer [6, 7]. The problem of personal recognition dictionary is the writing cost of the characters written by one specific writer, and personal dictionary has not used on offline OCR system up to the

present. In this chapter, we discuss two approaches for generating personal adaptive dictionary in offline character recognition.

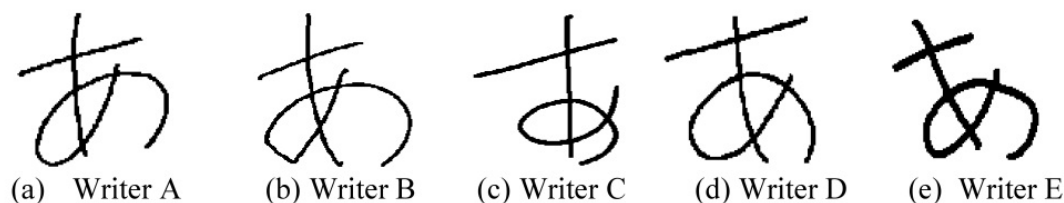


Figure 1. Variety of character forms written by different writers in Japanese HIRAGANA 'a'.

The first approach employs many characters written by a specific writer and many writers to generate a personal adaptive dictionary. In first approach, we proposed three types, that is, "Renewal type dictionary", "Modification type dictionary", "Mixture type dictionary" [6, 7] made by the compound of many characters by the specific writer and many writers. We evaluated the usefulness such as recognition accuracy, storage size of the three types for Japanese "Hiragana" character at offline. The experimental result shows that the personal dictionary is effective for recognition accuracy in comparison with the general dictionary generated by the characters written by many writers, and that the accuracy improved from 97% to 99%. However, the problem of personal dictionary is a large writing cost for each specific writer.

The second approach employs only one character written by the specific writer for all categories, and the only one character written by the specific writer selects one similar writer registered in recognition system. Some writers would write the similar character forms such as Fig. 2. The personal adaptive dictionary is generates using the characters written by the similar writer and many writers. We proposed two types, that is, "Similar mean dictionary", "Similar feature space dictionary" [9]. We compared two proposed types for Japanese character "Hiragana" at offline. The experimental results show that only one character for all categories is very effective for the improvement of recognition accuracy, and the character recognition rate is improved from 82% for the general dictionary to 91% by the proposed adaptive dictionary.



Figure 2. Character forms in the same category Japanese HIRAGANA 'a' by five writers.

Section 2 gives the properties for personal offline character recognition system and the outline of our character recognition system "Weighted Direction Index Histogram Method (WDIHM)" [10,11] which include the feature extraction for histogram of the direction and the modified quadratic discriminant function (MQDF)"[10,11]. Section 3 describes the generating methods of personal adaptive dictionary combined by the characters of specific

writer and many writers. Section 4 presents the usage of the characters written by the similar writer of a specific writer, which is low writing cost and that the accuracy of recognition is higher than the general dictionary. We think that the usage of the similar writer is useful for generating the adaptive dictionary.

2. Personal offline character recognition system

2.1. Usage of characters by a specific writer

The character forms written by individual writers have the variety shown in Fig. 1. Most writers do not write the standard character form, and they have some writing habits. Many researches and developers are interesting in the usage of the characters written by a specific writer. In on-line character recognition, the usage of the characters written by a specific writer is researched widely [1-5]. In offline character recognition, however, the personal recognition dictionary is not used generally at the present, as the writing cost of the characters is large for a specified writer.

We investigated the character forms written by some same writers. The specific writer has a writing habit, and the character forms written by the specific writer are similar each other. We guessed that the personal common feature such as writing habit for each writer is stable shown in Fig.3, and that the personal common feature for each individual writer is useful for personal character recognition. We are considering the extraction method and the usage of the personal common feature for character recognition system. We appear two generating methods of personal dictionary as follows.

- 1. The usage of many characters written by a specific writer (adaptive dictionary)
- 2. The usage of only one character written by a specific writer (similar dictionary)


























Writer a					
Writer b					
Writer c					
Writer d					
Writer e					

Figure 3. Variety of character forms by four writers in Japanese Hiragana ‘to’.

2.2. Properties for personal offline character recognition system

Personal offline character recognition system should adapt to a specific writer using the characters written by the specific writer. In the initial stage, as the number of the characters written by the specific writer is small, the dictionary of the recognition system should be the general dictionary, which is made from the characters written by many writers. Because the average of the accuracy for every writer should be high, and most of the writers desires the high recognition accuracy in the initial condition.

2.2.1. Calculation cost of a specific writer

The personal dictionary should be updated to the character form of a specific writer. The calculation cost of the updating the dictionary should be low for each input character. The recognition system with large cost such as neural networks, Dynamic Programming (DP) and Hidden Markov Model (HMM) is not suitable for personal offline character recognition.

2.2.2. Storage size for generating personal dictionary

The storage size for generating personal dictionary depends on the usage of the learning character, and the less storage of the size is desired. We think that the feature vectors to generate the personal dictionary should be in the mobile terminal machine of the user or cloud, and the writer can use the character forms by many writers.

2.3. Outline of character recognition system “Weighted Direction Index Histogram Method (WDIHM)”

We employ “Weighted direction index histogram method (WDIHM) [10, 11]” as the personal character recognition algorithm. The procedure of WDIHM consists of (1) binarization for extracting character area, (2) normalization of position and size (Figure 4(a)), (3) border following and 4-direction index coding (Figure 4(b)), (4) generation of 4-direction index histogram for 7x7 sub-regions (Figure 4(c)), (5) compress of 4x4 (sub-regions) x4 (directions) (= 64 dimension) histograms using Gaussian weighted filter (Figure 4(d),(e)).

This algorithm is popular in off line character recognition for Japanese handwritten characters, and we are developing this algorithm to personal character recognition. The feature vector in the recognition method employs the four direction index histogram, and the dimension of the feature vector is 64 (= 7x7x4).

The mean vector and covariance matrix of feature vector $x = (x_1, x_2, \dots, x_{64})^T$ for a category l are given in Equation (1), (2)

$${}_l\mu = \frac{1}{N} \sum_{i=1}^N {}_l x_i \quad (1)$$

$${}_l\Sigma = \frac{1}{N} \sum_{i=1}^N ({}_l x_i - {}_l\mu)({}_l x_i - {}_l\mu)^T \quad (2)$$

Where, $i = 1, 2, \dots, N$ and N is the number of learning characters.

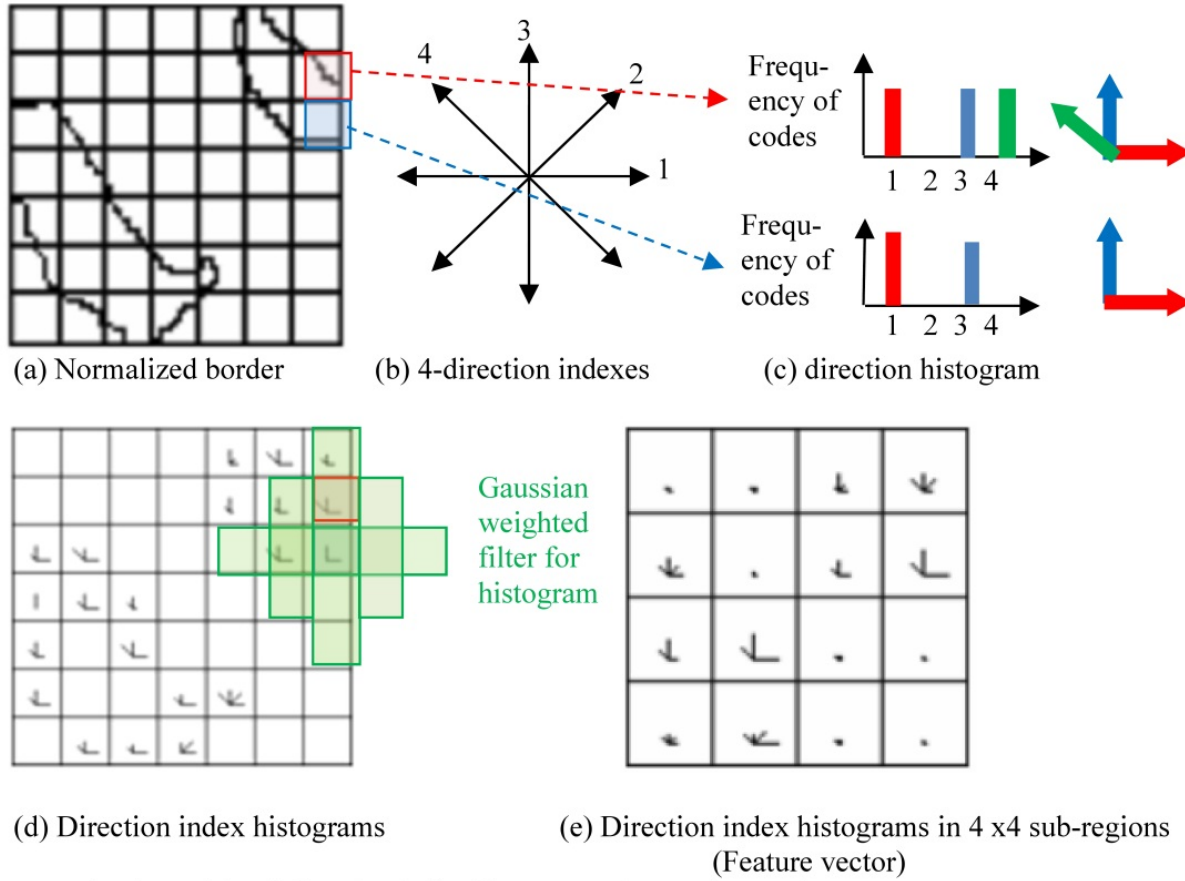


Figure 4. Weighted direction index histograms feature (WDIH).

Quadratic discriminate function (QDF) of an n -dimensional feature vector \mathbf{x} is given Equation (3) for a category l . $P(l)$ is the *a priori* probability for the category l .

$$f_l(\mathbf{x}) = (\mathbf{x} - \mu_l)^T \{\Sigma_l\}^{-1} (\mathbf{x} - \mu_l) + \ln |\Sigma_l| - 2 \log P(l) \quad (3)$$

The QDF becomes optimal in the Bayesian sense for normal distributions with known parameters [11]. On the limited samples, the performance of QDF is degraded because of estimation error, as the parameters become non-optimal. QDF has some problems such as the recognition accuracy, computation time, storage and so on.

We proposed the modified quadratic discriminate function (MQDF) [10, 11] (equation (4)). In our personal character recognition, we employ the modified quadratic discriminate function (MQDF). MQDF for each category is based on the principal component analysis (PCA), and it employs a mean vector, a set of eigenvectors and eigenvalues of a covariance matrix on feature vector for each character category (Fig. 5).

In recognition phase, from the input character the feature vector is extracted, and the MQDF value is calculated for each category. The recognition result, that is the recognized category, is determined by the minimum of the MQDF value for each category.

$${}_l g(x) = \sum_{i=1}^{k-1} \frac{\{ {}_l \varphi_i^T (x - {}_l \mu) \}^2}{{}_l \lambda_i} + \sum_{i=k}^n \frac{\{ {}_l \kappa_i^T (x - {}_l \mu) \}^2}{{}_l \lambda_k} + \ln \left(\prod_{i=1}^{k-1} {}_l \lambda_i \cdot \prod_{i=k}^n {}_l \lambda_k \right) \quad (4)$$

where, x : a feature vector of a character sample

${}_l \mu$: mean vector of character in category l

k : the number of used eigenvalues ($k < n$) determined by the designer

n : the dimension of feature vectors

${}_l \varphi_i$: i -th eigenvector in category l of covariance matrix

${}_l \lambda_i$: i -th eigenvalue in category l of covariance matrix

T : transpose of a vector

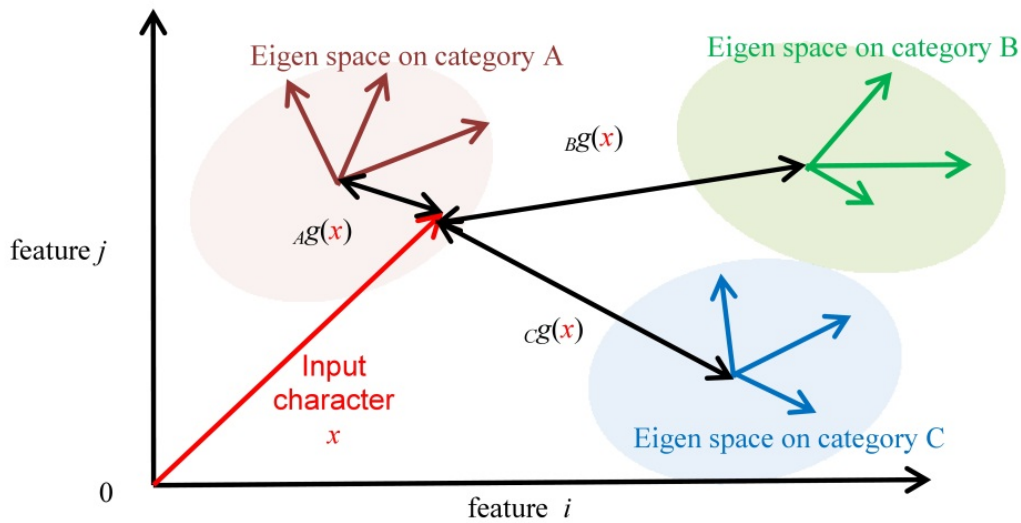


Figure 5. Distance using discriminant function MQDF in feature space

The most conventional handwritten OCRs employ a general dictionary, which is generated by many characters written by many general writers to grasp the variety of character forms. The general dictionary consists of the mean vectors, eigenvalues and eigenvectors for each category. The mean vector is made from the feature vectors of learning characters, and the eigenvalues and eigenvectors are calculated by the covariance matrix on the feature vectors. The general dictionary is generated at software developer usually.

3. Generating methods of adaptive personal dictionary

3.1. Pure personal dictionary and adaptive dictionary

A pure personal dictionary is generated by many characters written by a specific writer, and the dictionary reflects the writing habit of a specific writer. The personal dictionary consists of the personal mean vector and personal covariance matrix for each category (Equation (5), (6)). The recognition accuracy could be the better than the general dictionary by many writers as the distribution of characters written by one specific writer is narrower than the distribution of characters in general dictionary (Fig. 6).

$${}_1\mu_p = \frac{1}{N_p} \sum_{i=1}^{N_p} {}_1x_{pi} \quad (5)$$

$${}_1\Sigma_p = \frac{1}{N_p} \sum_{i=1}^N ({}_1x_i - {}_1\mu_p)({}_1x_i - {}_1\mu_p)^T \quad (6)$$

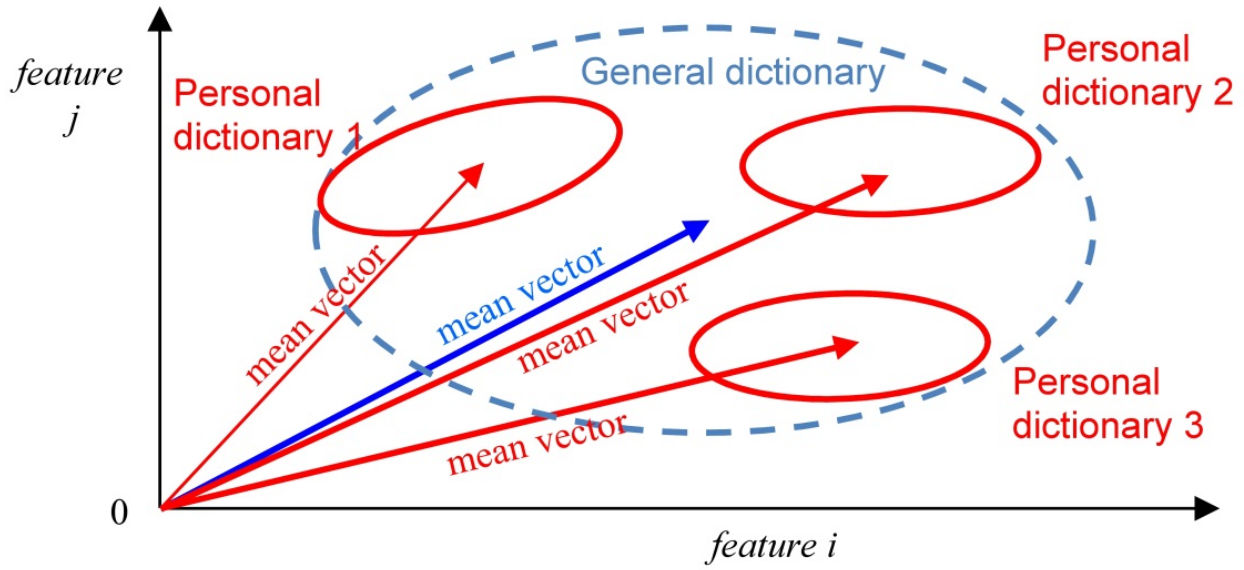


Figure 6. Distribution of personal dictionary and general dictionary

We prepared the set of characters written by five writers using mechanical pencil and one character is written for each frame. The set consists of 10 characters per category and the character sets are 46 categories without a voiced consonant mark “ ” and a P-sound mark “ ” in Japanese “Hiragana” characters shown in Table 1. We employed it to generate personal dictionary.

We examined the comparison between the personal and general dictionary for Japanese Hiragana characters (46 categories), and the recognition rates are shown in Fig. 7 when the number of learning characters is ten characters / category [6]. The mean recognition rate of personal dictionary (99.0%) is 2.2 better than the general dictionary (96.8%). The incorrect category of the recognition result is limited at some categories, and the character form of the category is different from the general form. The recognition rates depend on the number of learning characters, and the lack of learning character is one of the important problems. The problem of personal dictionary is the writing cost of a specific writer.

We proposed new three types of adaptive personal dictionary to reduce the writing cost [6, 7]. The adaptive dictionary is made from the characters written by a specific writer and many general writers. The recognition rates of the following three adaptive dictionaries are higher than the pure personal dictionary.

あ	か	が	さ	ざ	た	だ	な	は	ば	ぱ	ま	や	ら	わ
い	き	ぎ	し	じ	ち	ぢ	に	ひ	び	ぴ	み		り	ん
う	く	ぐ	す	ず	つ	づ	ぬ	ふ	ぶ	ぷ	む	ゆ	る	
え	け	げ	せ	ぜ	て	で	ね	へ	べ	ぺ	め		れ	
お	こ	ご	そ	ぞ	と	ど	の	ほ	ぼ	ぽ	も	よ	ろ	を

Table 1. 46 pure sound categories and 25 categories with the voiced consonant mark and the P-sound mark in Japanese HIRAGANA

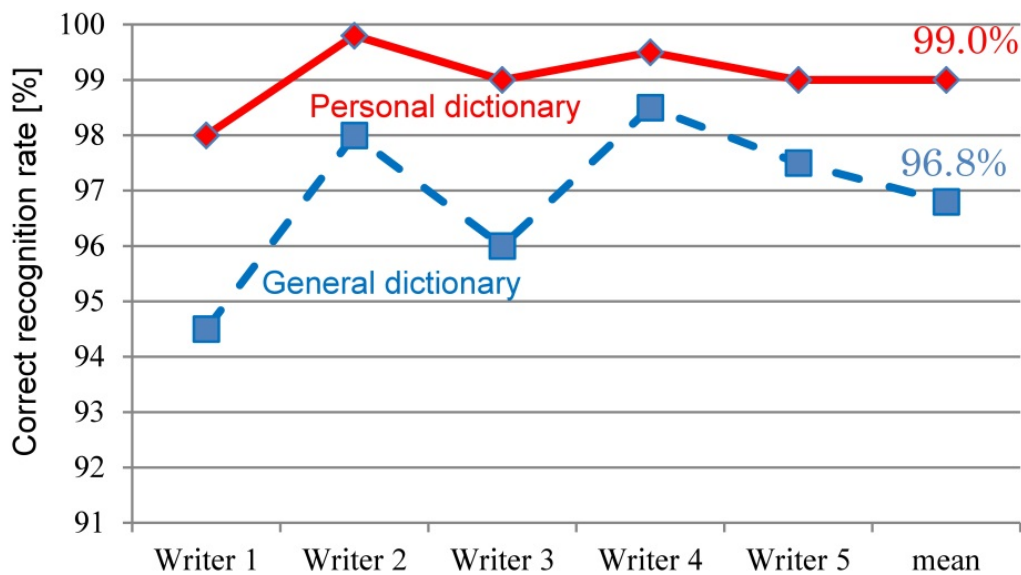


Figure 7. Recognition rates of personal dictionary and general dictionary for 46 categories

3.2. Renewal type dictionary

The mean vector, eigenvalues and eigenvectors in the renewal type are generated by many characters written by the specific writer and many general writers, when the number of written characters by a specific writer increased (Equation (7) and (8)). The weights of the specific writer and general writer are equal to generate the dictionary. Yoshimura et al presented useful for Japanese character recognition using pattern matching method [8]. The recognition rate of the personal dictionary is better than the general dictionary using WDIHM. The problem is the writing cost by a specific writer, and the number of writing characters is more than 5 characters per category initially to become the recognition rate which is better than the general dictionary.

$${}_1\mu_{pr} = \frac{1}{N_p + N} \left(\sum_{i=1}^{N_p} {}_1x_{pi} + \sum_{i=1}^N {}_1x_i \right) \quad (7)$$

$${}_1\Sigma_{pr} = \frac{1}{N_p + N} \left\{ \sum_{i=1}^{N_p} ({}_1x_i - {}_1\mu_{pr})({}_1x_i - {}_1\mu_{pr})^T + \sum_{i=1}^N ({}_1x_i - {}_1\mu)({}_1x_i - {}_1\mu)^T \right\} \quad (8)$$

3.3. Modification type dictionary

In modification type dictionary, mean vectors only is updated when the number of written characters by the specific writer increased, and it is the same to pure personal dictionary by Equation (5). The eigenvalues and eigenvectors are the same as the general dictionary from Equation (2), and they are not updated as the stability of eigenvalues and eigenvectors is low when the number of character written by a specific writer is little. The problem of the recognition rate is unstable for some writers when the number of leaning characters is less than 5 characters per categories.

3.4. Mixture type dictionary

In mixture type dictionary, the mean vector employs the combination of the general mean vectors and the specific writer's mean vector given by the following equation, where the number of characters by a specific writer p is N_p . The eigenvalues and eigenvectors are the same as the general dictionary (Equation (2)).

$${}_l\mu_{pm} = \frac{1}{1 + N_p}({}_l\mu + \sum_{i=1}^{N_p} {}_l x_{pi}) \quad (9)$$

where, ${}_l\mu$ mean vector of category l in general dictionary

N_p : the number of characters written by specific writer p

To understand the distributions of three adaptive dictionaries Fig. 8 shows the mean vectors and the existence space of most samples on general dictionary and three type personal dictionaries in feature space, where the mean vector and the existence space are illustrated as an arrow and an ellipse, respectively. The existence space on mixture type and modification type are the same as the general dictionary, and the existence space of the renewal dictionary is narrower than the other dictionaries.

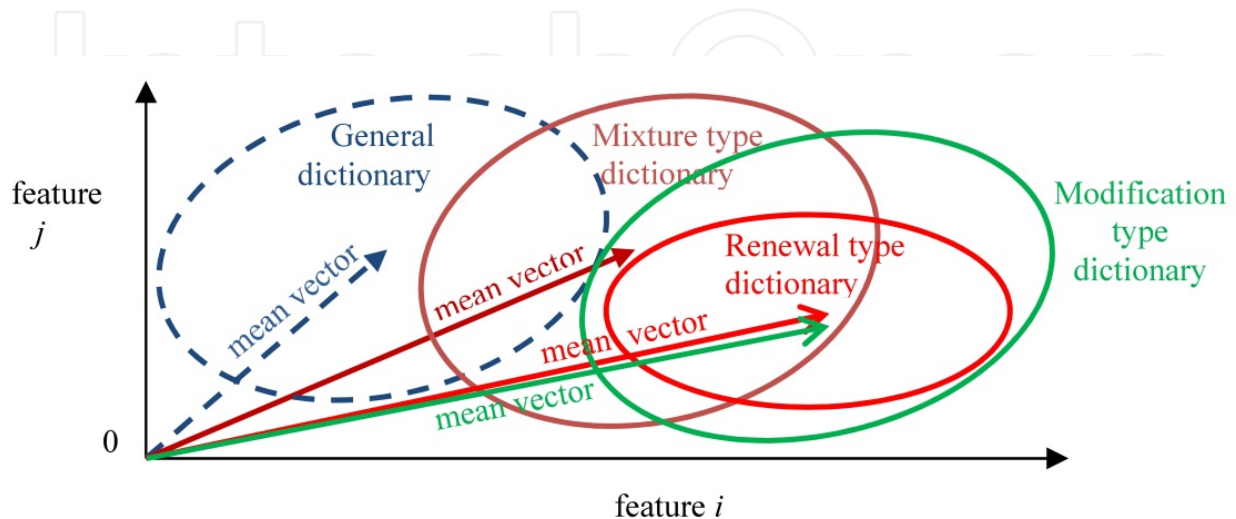


Figure 8. Comparison of general dictionary and personal dictionary in feature space.

3.5. Comparison of four personal dictionaries

Fig. 9 shows the correct recognition rates on the number of characters written by a specific writer on personal dictionary and three adaptive dictionary types (renewal, modification, mixture) for 46 categories without the voiced consonant mark and the P-sound mark. The general dictionary is made from the characters written by 200 writers per category in the character data base ETL9B by the Electro-Technical Laboratory (ETL) of Japan [at present, the National Institute of Advanced Industrial Science and Technology (AIST) of Japan].

The recognition rate of the modification type at the end of left (the number of characters 0) is the recognition rate of the general dictionary. The recognition rates of modification type and mixture type are better than the general dictionary. The recognition rate of mixture type is better than the other types from 2 learning characters to 8 learning characters. The recognition rates of three adaptive dictionaries are the better than the personal dictionary and the best recognition rate is mixture type dictionary. Table 2 shows the properties of personal dictionary and the adaptive dictionaries, and the recalculation costs of modification and mixture dictionary are less than the personal dictionary and the renewal type as the modification and mixture type dictionary recalculate only the mean vector.

The mixture type dictionary would be the best solution as the personal dictionary from above mentioned experiments. However, the problem of the mixture type dictionary needs at least one character per category, and the specific writer must write the characters for the number of categories. The writing cost of a specific writer is very large when the number of categories is large. For example, the number of Japanese Kanji characters is more than 6000 categories.

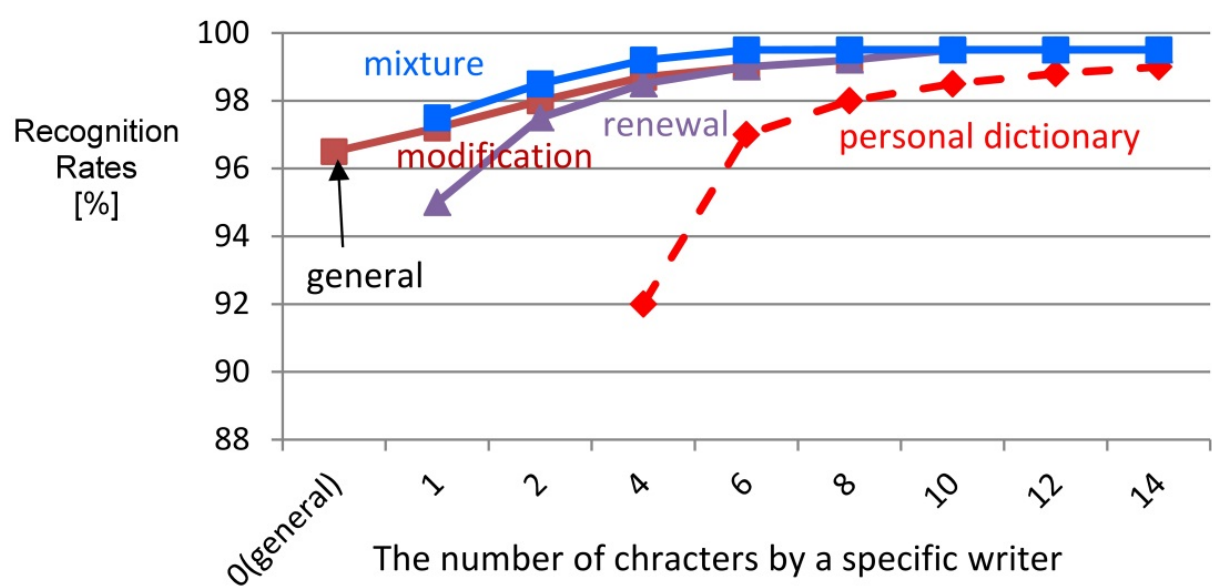


Figure 9. The correct recognition rates on the number of learning characters by five writers for 46 categories

Dictionary type			Personal	Renewal	Modification	Mixture
Components	Mean		personal	personal + general	personal	personal + general
	Eigenvalues and Eigenvectors		personal	personal + general	general	general
Recognition rates	The number of characters by a specific writer	less than 6 character	< general	< general in 1, > general in more than 2	mixture > modification \geq renewal > personal	
		more than 6 characters		mixture \geq modification \geq renewal > personal > general		
	Recognition rate [%] (10 character / category)		99.0	99.3	99.5	99.5
Re-calculation cost			1400	1400	1	
Storage size			65	65	1	

Table 2. Properties of personal dictionary and three adaptive type dictionaries.

3.6. Effect of one character per category

We prepared 20 character images per category for ten writers by pen tablet for 71 “HIRAGANA” character categories, and the character categories include 46 pure sound categories, 20 sound categories with a voiced consonant mark ‘ ゛ ’ and 5 sound categories with a P-sound mark ‘ ゜ ’ shown in Table 1. The resolution of the character is 100 x 100 pixels. We didn’t use the time information of tablet and used the image information only. As the feature vector we used the histograms of 4x4 sub-regions, 4 directions, that is, 64 dimensional feature vector.

Fig. 10 shows the correct recognition rates on the number of characters written by a specific writer for all Hiragana 71 categories in mixture type dictionary. The recognition rates of mixture dictionary (93.7% in mixture (10) and 90.8% in mixture (1)) are better than the general dictionary (82.4%). Only one character such as mixture (1) in Fig. 10 is very effective to improve the recognition rate, and the ten characters such as mixture (10) in Fig. 10 can saturate the recognition rates.

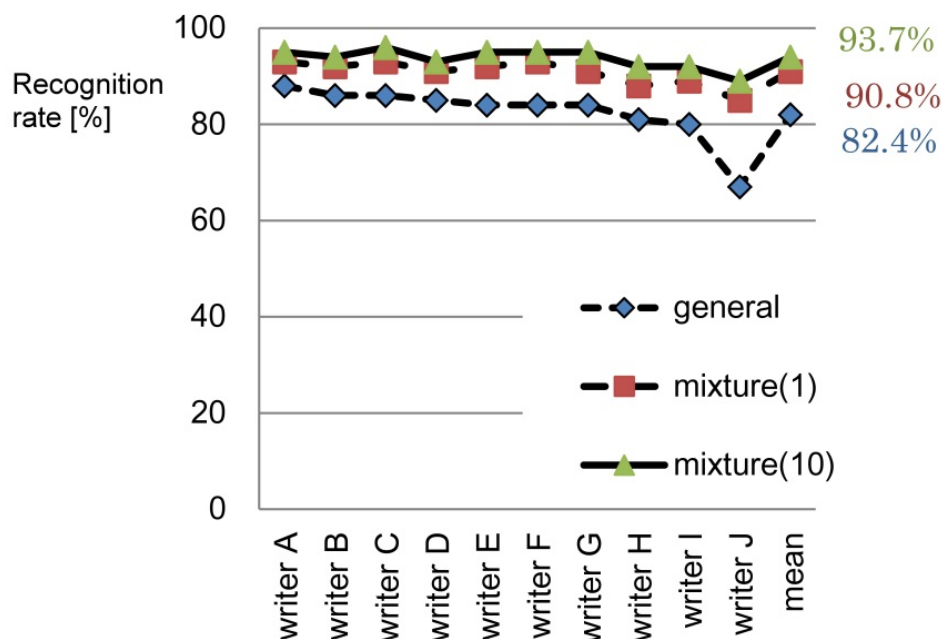


Figure 10. Effect of writer's characters in mixture type dictionary for 71 categories

Fig. 11 shows the relation of four mean vectors of personal, mixture (10), mixture (1) and general dictionaries. The mixture (1) approaches personal mean using only one character per category, and it is effective for the improvement of recognition rate.

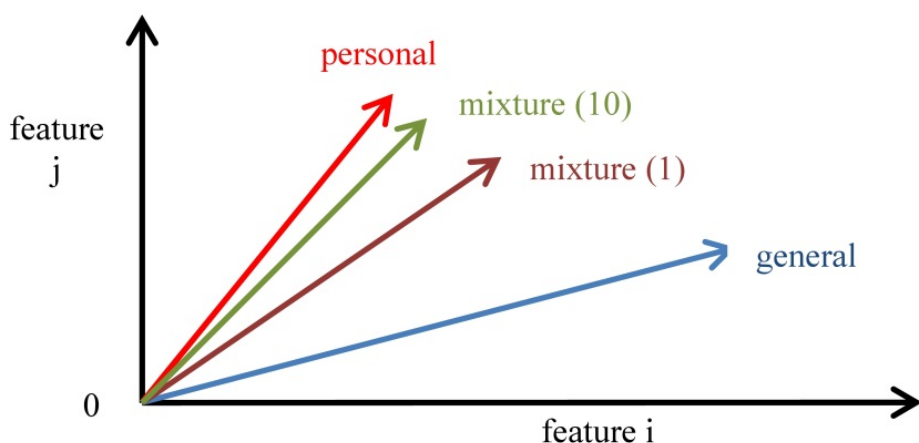


Figure 11. Mean vectors of personal, mixture (10), mixture (1) and general dictionaries in feature space

4. Usage of one character by a specific writer and similar writers

4.1. Outline of usage of one character by a specific writer and similar writers

To resolve the writing cost of a specific writer in the above mentioned discussion, we proposed two new generating methods of adaptive dictionary, especially mixture type dictionary, using only one character for all categories [9]. The key idea is the usage of characters written by the similar writer registered in advance. We assume that the writer's writing feature of one category is much alike to the writer's writing feature of the other

category shown in Fig. 12, and that character form of the similar writer selected by one category and one character is similar to the character form by a specific writer in every category, as some writer verification researches appear that the writing feature of one category is similar to every category [12, 13]. Fig. 12 shows that the curvature of arc and the direction of character lines are similar for each writer.

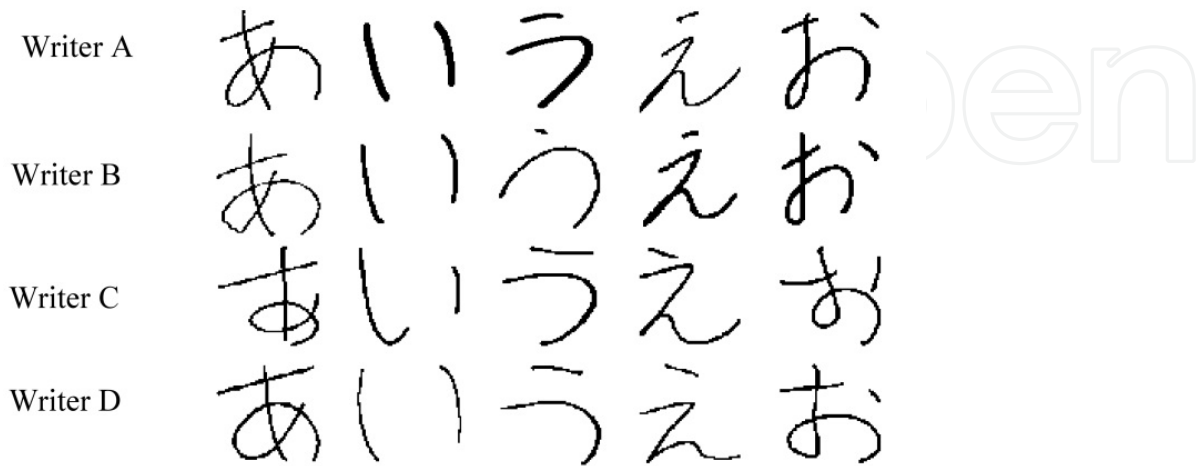


Figure 12. Variation of characters written by four writers.

The outline of our proposed method can be explained as the following procedure.

1. In preparing process, some writers write the set of handwritten characters for all categories to generate an adaptive dictionary for each writer such as "Writer A" and "Writer B" in Fig. 12. The feature vector of the character is extracted by WDIHM mentioned in 2.3.
2. An adaptive dictionary, which consists of the mean vector, the eigenvalues and the eigenvectors of the feature vector for each category, is generated from the set of handwritten characters by only one writer. We prepare the adaptive dictionary for each writer, and call the writer "similar writer" in this chapter. The number of similar writers is limited at the initial operation phase of the character recognition system.
3. In learning process, one character written by one specific writer selects the most similar writer by the minimum value of MQDF among the registered similar writers in Fig. 13. The specific writer would be the specific user of a personal terminal machine. In recognition process, the recognition system employs the recognition dictionary of the similar writer for every category. Fig. 14 shows that the similarity on writing habit for two categories, and that the relative position of writers is similar between category A and B. The recognition process using adaptive dictionaries for each similar writer is shown in Fig. 15.
4. The selected adaptive dictionary is updated by the character written by the specific writer to adapt the character form written by the specific writer. Two new adaptive methods are proposed in the following two sections.

When the user employs mobile terminal machines such as smartphone and tablet personal computer (tablet PC), a new user uses the adaptive dictionary of the similar writers in file saver on the Internet shown in Fig. 16. As the adaptive dictionary of the new writer would be updated and be stored in the Internet file saver, the number increases according to the number of users of the proposed system.

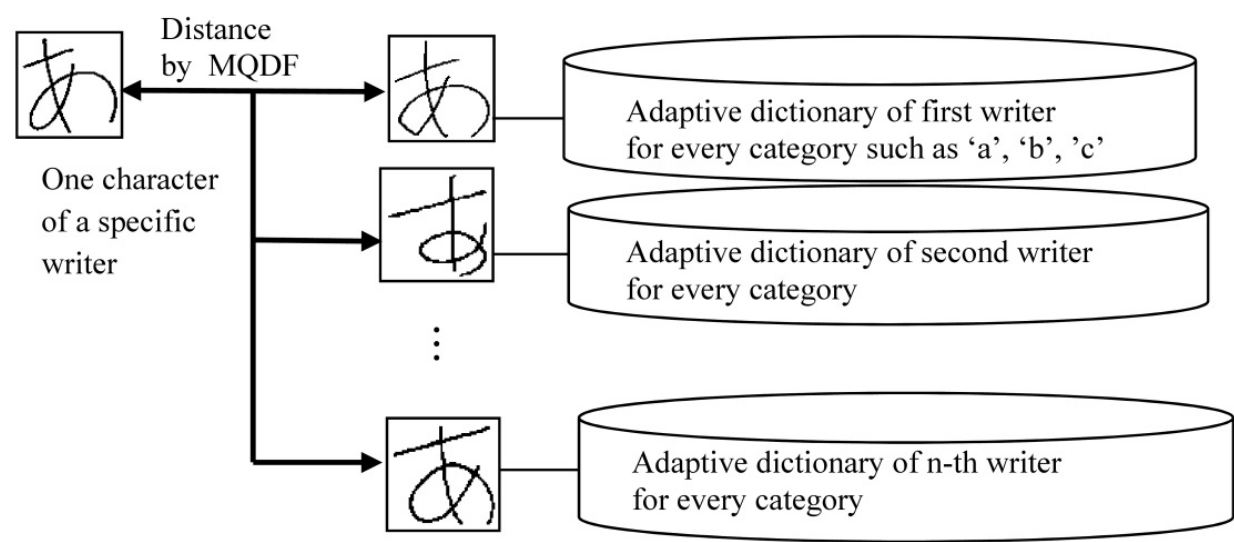


Figure 13. Selection of the most similar writer by one character of the specific writer in learning process

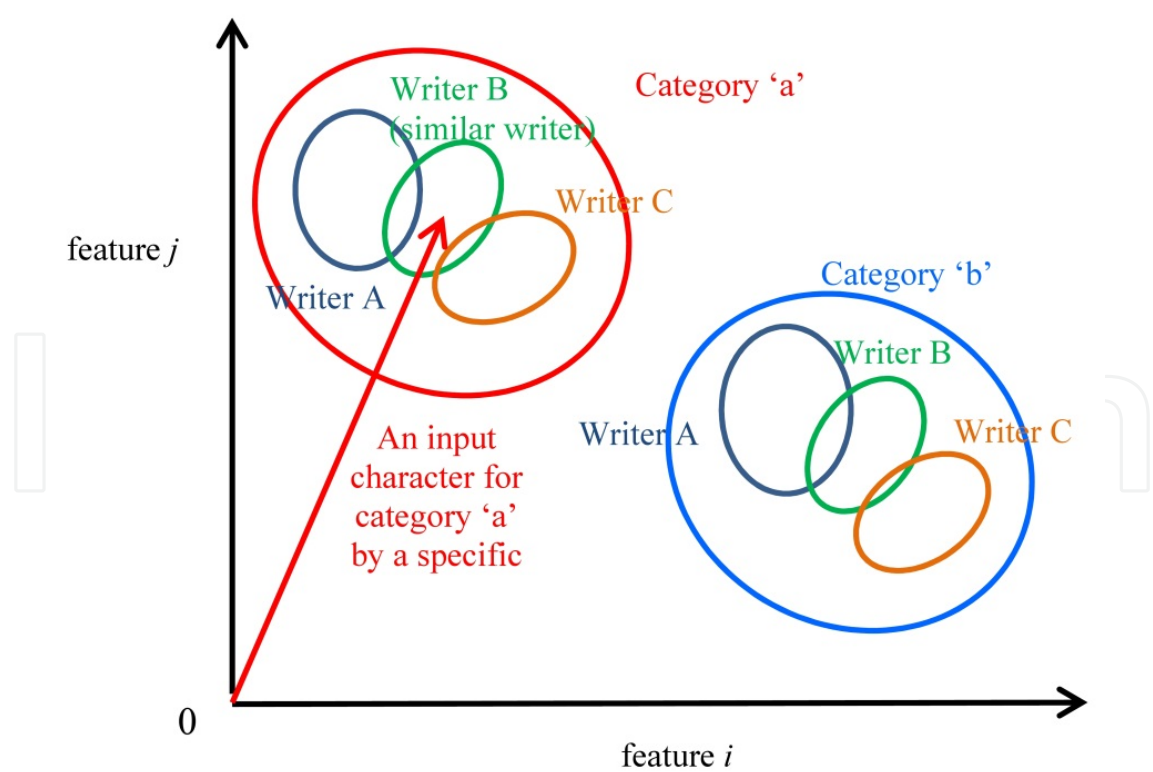


Figure 14. Similarity on writing habit for two categories

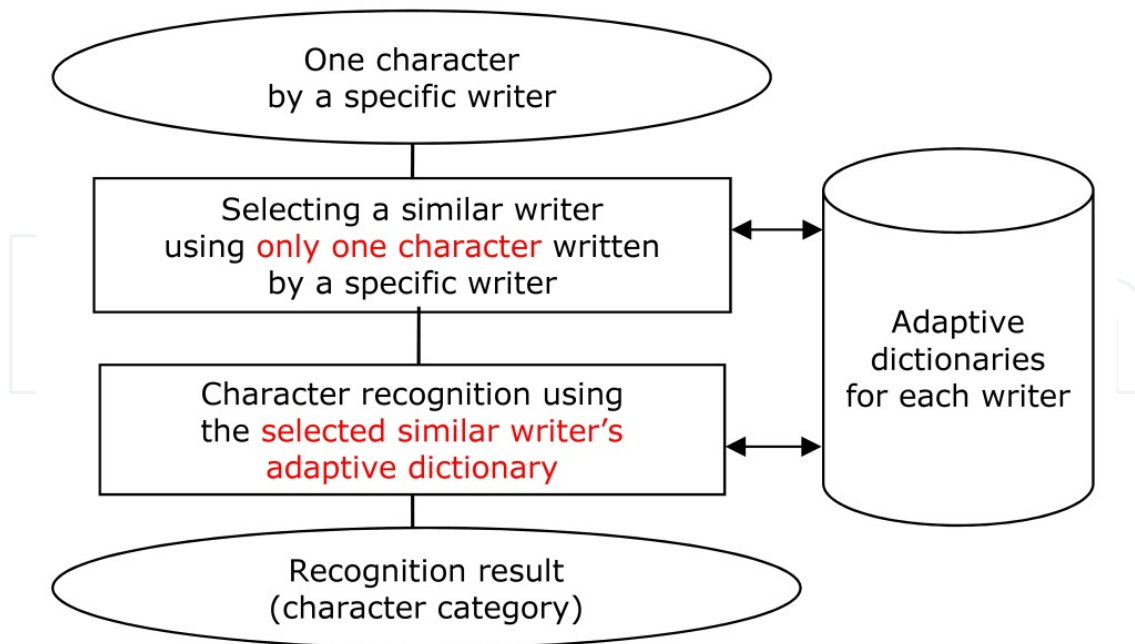


Figure 15. Recognition process using adaptive dictionaries for each similar writer

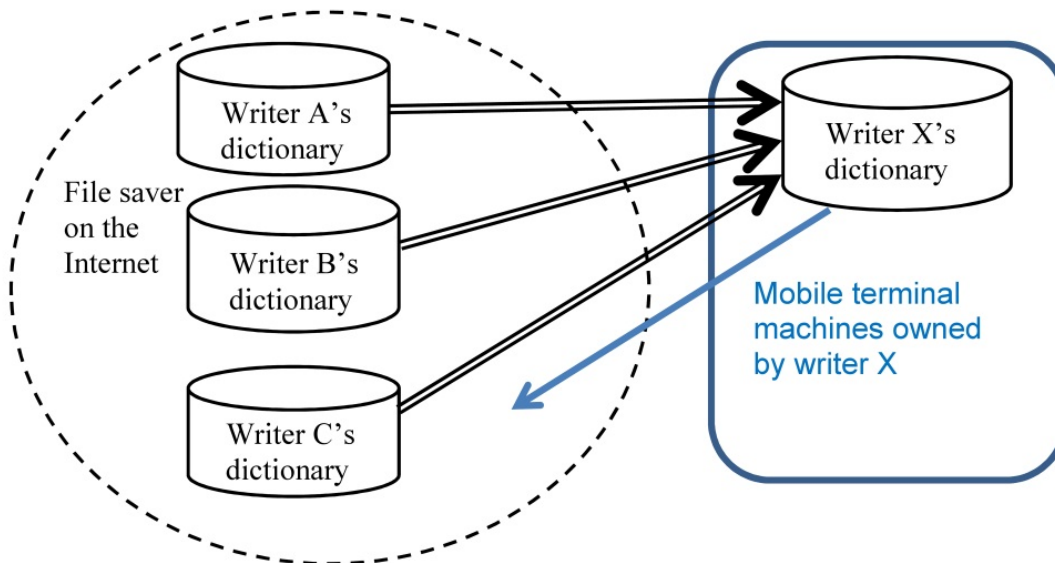


Figure 16. Dictionary generating process using character recognition dictionary on the Internet

4.2. Similar mean dictionary

The similar mean dictionary consists of the mean vector and the set of the eigenvalues and the eigenvectors for each category. In learning phase, the mean vector only is updated in the learning phase, and the set of the eigenvalues and the eigenvectors is the same as the general dictionary.

In the initial phase, the mean vector is the combination of the general mean and the mean vector of the similar writer for each category by equation (10).

$${}_1\mu_s = \frac{1}{1+N_s}({}_1\mu + \sum_{i=1}^{N_s} {}_1x_{s,i}) \quad (10)$$

where, N_s : the number of characters written by a similar writer.

In the leaning phase, the mean vector is updated by the character written by the specific writer (user of personal machines) using the following equations.

$${}_1\mu_p = \frac{1}{N_p} \sum_{i=1}^{N_p} {}_1x_{p,i} \quad (11)$$

$${}_1\mu_{psm} = \frac{N_p}{N_s + N_p + 1} {}_1\mu_p + \frac{N_s + 1}{N_s + N_p + 1} {}_1\mu_s \quad (12)$$

where, N_p : the number of characters written by the specific writer.

In the well learned phase, the number of learning characters written by the specific writer becomes large, and the mean vector closes to the mean vector of the specific writer. The set of the eigenvalues and the eigenvectors is the same as the general dictionary.

4.3. Similar feature space dictionary

The similar feature space dictionary consists of the mean vector and the set of the eigenvalues and the eigenvectors.

In the initial phase, the mean vector and the set of the eigenvalues and the eigenvectors are generated by the combination of similar writer and general writer. The mean vector in the similar feature space dictionary is the same as the similar mean dictionary mentioned in 4.2. The set of eigenvalues and eigenvectors is not the same with the similar mean dictionary, and it is calculated by the covariance matrix on the feature vectors of characters written by the similar writer and the general writers shown in equation (13).

$${}_1\Sigma_{psf} = \frac{1}{N_s + 1} {}_1\Sigma_g + \frac{N_s}{N_s + 1} {}_1\Sigma_s \quad (13)$$

In the learning phase, the mean vector only is updated by the character written by the specific writer (user of personal machines) using equation (11) and (12). The set of the eigenvalues and the eigenvectors are not updated.

4.4. Comparison of four dictionaries

Two new proposed methods in this paper employ one character written by a specific writer (a new user), and the effort of the user is the minimum to reflect the handwritten feature of a specific writer. The comparison of four dictionaries in initial phase is showed in Table 3.

The similar mean dictionary employs the combined mean vector of characters written by the similar writer and the general writers, and it employs the set of eigenvalues and

eigenvectors of the characters written by the general writers. The similar feature space dictionary employs the mean vector and the set of eigenvalues and eigenvectors of characters written by the similar writer and the general writers. The difference of these four dictionaries illustrates in Fig. 17.

Type of recognition dictionary	The number of characters by a specific writer for 71 categories	The writers of mean vector	The writers of Eigenvalues and Eigenvectors
General	0	general	general
Mixture type	71 (minimum)	general+specific	general
Similar mean	1	general+similar	general
Similar feature space	1	general+similar	general+similar

Table 3. Comparison of writing costs for 71 categories and the components of dictionary (mean vector, eigenvalues and eigenvectors) in initial phase

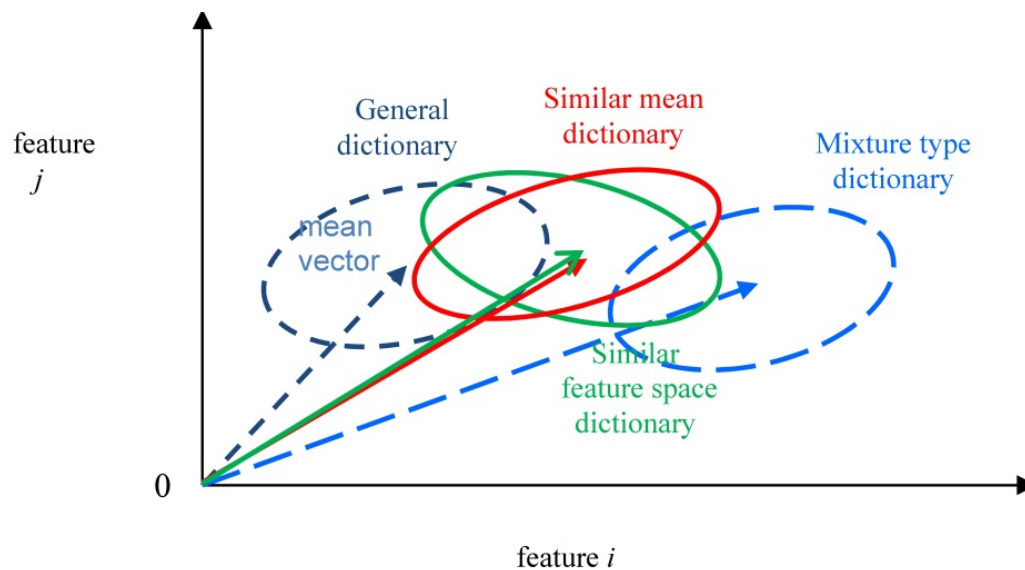


Figure 17. Difference of four dictionaries in feature space

4.5. Experimental results

4.5.1. Similar writer selection by one character in one category

We examined the recognition rates of HIRAGANA categories to select the category written by the specific writer. The average recognition rate for ten writers is obtained in 20 learning characters per category, and the every combination of 71 categories and ten writers is calculated. The category of the maximum recognition rate among 71 categories is HIRAGANA category 'po', and we select the 'po' category that should be written by the specific writer.

The images in Table 4 and Table 5 show the character forms of HIRAGAN category “e” and category “pa”, respectively. The character images in tables show a typical example for each writer. The character forms show the large variety of the writing habit. In Table 4, the MQDF value for one character in category ‘po’ written by “Writer D” is calculated for nine registered writers using the mixture type dictionary, and “Writer C” is selected by the minimum of MQDF value as the similar writer. MQDF values in category ‘e’ are shown in Table 4. The MQDF value of the similar writer is the minimum value among the registered nine writers without the specific writer (Writer D), and the character form of the similar writer (Writer C) is similar to the character form of the specific writer (Writer D). The selection procedure of similar writer would be appropriate in this category.






	Specific writer (Writer D)	Similar Writer (Writer C)	Writer J	Writer A	Writer F
MQDF value	149	158	165	167	169
Character form					

Table 4. Correct result of character ‘e’ using similar mean dictionary and similar space dictionary of the similar writer selected by one character ‘po’ written by a specific writer

Table 5 shows the case of the critical MQDF value of the similar writer. Writer J is selected by the character in category ‘po’ written by “Writer B”. The MQDF value of the similar writer “Writer J” is different from the character form by the specific writer (Writer B), and it is close to the MQDF values of the other writer. The similar writer depends on the written category, and the future problem is the selection of the category for the selection of the similar writer.

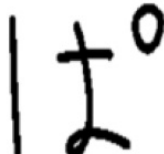
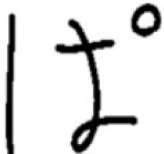
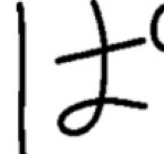

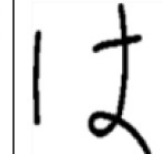
	Specific writer (Writer B)	Similar Writer (Writer J)	Writer E	Writer F	Writer H
MQDF value	138	147	149	154	155
Character form					

Table 5. Correct result of character ‘pa’ using only similar space dictionary

4.5.2. The comparison of four dictionaries

We compared three personal dictionaries (mixture type dictionary, similar mean dictionary and similar feature space dictionary) and the general dictionary. Fig. 18 shows the

comparison of the four dictionaries in the correct recognition rates for ten writers, and the order of writers is sorted by the recognition rates using general dictionary. The rates of mixture type dictionary (90.8% in mean) and the similar feature space dictionary (91.0% in mean) are nearly equal, and these rates are better clearly than the rates of the general dictionary (82.4% in mean) and the similar mean dictionary (84.7% in mean) for all writers. The rates of the similar mean dictionary for 7 writers are better than the general dictionary, and the mean rate for ten writers is better than general dictionary. It is more effective for writer with strong writing habit such as Writer J, and the effect of these dictionaries would increase when the number of similar writes would become large. The recognition rate of similar mean dictionary becomes near the general dictionary as the problem of similar mean dictionary would be the mismatch between the mean vector and the set of eigenvalues and eigenvectors. The number of learning character per category to generate similar mean dictionary and similar feature space dictionary is 10 for every category.

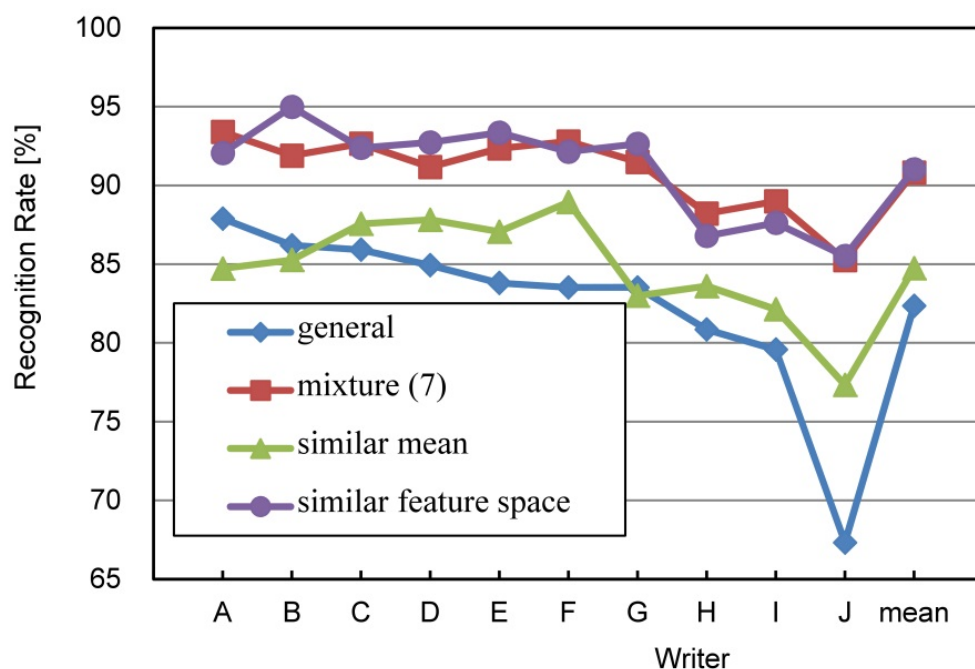


Figure 18. The recognition rates by four dictionaries

One character is the least cost to extract the writing habit of writer in similar mean dictionary and similar feature space dictionary. The writing cost of these dictionaries is $1/\{(the\ number\ of\ category) \times (learning\ characters\ per\ category)\}$ of mixture type dictionary. The writing cost of the specific writer is reduced vastly.

Table 6 shows the comparison of correct recognition rates and the writing cost for general dictionary, mixture type dictionary, similar mean dictionary and similar feature space dictionary. It is confirmed that only one character by a specific writer (user) is very effective for handwritten character recognition.

The character image written by the specific writer in Table 4 is the example of the correct recognition result of character 'e' using similar mean dictionary and similar space dictionary

of the similar writer selected by one character 'po' written by a specific writer. MQDF value (158) by the similar writer in Table 4 is the minimum value for all categories. However, the character image written by the specific writer in Table 5 is the example of the correct result of character 'pa' using only similar space dictionary, and using similar mean dictionary arises an incorrect result as MQDF value for category 'pa' is not the minimum MQDF value for the other categories.

Type of recognition dictionary	Correct recognition rate [%]	The number of characters by a specific writer for 71 categories
General	82.4	0
Mixture type (7)	90.8	$71 \times 7 = 497$
Similar mean	84.7	1
Similar feature space	91.0	1

Table 6. Comparison of correct recognition rates

Table 7 shows the incorrect recognition result of character 'wo' using similar mean and similar feature space dictionaries. The MQDF value (159) of the similar writer for category 'wo' is larger than the category 'chi', and the recognition result becomes the category 'chi'. If the similar writer would be 'Writer D', the input character is recognized correctly. We are considering a new selection method of the similar writer to improve the correct recognition rate.






	Specific writer (Writer F)	Similar Writer (Writer I)	Writer D	Writer A	Writer G
MQDF value	133	159	139	147	166
Character form					

Table 7. Incorrect result of character 'wo' using similar mean and similar feature space dictionaries

The similar mean dictionary and the similar feature dictionary use effectively one character written by the specific writer, and we confirm that the usage of one character will enlarge for personal terminal machines.

5. Conclusions

We explained the usefulness of personal dictionary on offline character recognition using our proposed adaptive dictionary. Three adaptive dictionaries (the renewal type, modification type and mixture type) are introduced by our research group, and the recognition rates of the renewal type, modification type and mixture type are 99.3%, 99.5%, 99.5% for 46 categories, respectively. The recognition rate of mixture type is better than the

other types from 2 learning characters to 8 learning characters. We think that the mixture type dictionary is most useful for personal terminal machines such as smartphone and tablet personal computer (tablet PC). However the problem of the adaptive dictionary is the writing cost, and to resolve this problem we proposed two dictionary generation methods (similar mean dictionary and similar feature space dictionary) using only one character methods by a specific writer.

We examined the recognition rate using handwritten characters of 71 Japanese "HIRAGANA" categories and we obtained the character recognition rate of 91.0 % (the general dictionary made from ETL9B: 82.4 %). The usage of character forms written by a specific user is very effective even if the number of characters by the user is only one, and the character by the user improves the recognition rate of character recognition system vastly.

The future problems are as follows.

5. The selection of the category for the selection of the similar writer
6. The usage of multiple similar writers and multiple categories
7. The application to Chinese characters

Author details

Shinji Tsuruoka*

Graduate School of Regional Innovation Studies, Mie University, Tsu, Mie, Japan

Masahiro Hattori

Previously at Graduate School of Engineering, Mie University, Tsu, Japan

Takuya Kimura

Previously at Graduate School of Regional Innovation Studies, Mie, University, Tsu, Mie, Japan

Yasuji Miyake

Professor Emeritus, Mie University, Tsu, Mie, Japan

Haruhiko Takase and Hiroharu Kawanaka

Graduate School of Engineering, Mie University, Tsu, Mie, Japan

Acknowledgement

We would like to sincerely thank to Prof. Fumitaka Kimura and Associate Prof. Tetsushi Wakabayashi in Mie University, Japan.

6. References

- [1] Tappert C.C (1984) Adaptive on-line handwriting recognition, Seventh International Conference on Pattern Recognition (7th ICPR): 1004-1007

* Corresponding Author

- [2] Connell S.D, Jain A.K (2001) Template-based online character recognition, *Pattern Recognition* 34: 1-14
- [3] Connell S.D, Jain A.K (2002) Writer Adaptation of Online Handwriting Models, *IEEE Trans. PAMI*: 329-346
- [4] LaViola J J, Zeleznik R C (2007) A Practical Approach for Writer-Dependent Symbol Recognition Using a Writer-Independent Symbol Recognizer, *IEEE Trans. PAMI*, 29(11):1917-1926
- [5] Huang Z, Ding K, Jin L (2009) Writer Adaptive Online Handwriting Recognition Using Incremental Linear Discriminant Analysis, *Proc. of International Conference on Document Analysis and Recognition (ICDAR2009)*:91-95.
- [6] Tsuruoka S, Morita H, Kimura F, Miyake Y (1987) Handwritten Character Recognition Adaptable to the Writer. *IEICE Trans. on Information and Systems*, J70-D (10):1953-1960 [in Japanese]
- [7] Tsuruoka S, Morita H, Kimura F, Miyake Y (1988) Handwritten Character Recognition Adaptable to the Writer. *Proc. of IAPR Workshop on Computer Vision*: 179-182
- [8] Yoshimura M, Kimura F, Yoshimura I (1983) On the Effectiveness of Personal Templates in the Character Recognition, *IEICE Trans. on Information and Systems*, J66-D (4):454-455 [in Japanese]
- [9] Tsuruoka S, Hattori M, Kadir M F A, Takano T, Kawanaka H, Takase H, Miyake Y (2010) Personal Dictionaries for Handwritten Character Recognition Using Character Written by a Similar Writer. *Proc. of 12th International Conference on Frontiers in Handwriting Recognition (ICFHR2010)*: 599-604.
- [10] Tsuruoka S, Kurita K, Harada T, Kimura F, Miyake Y (1987) Handwritten "KANJI" and "HIRAGANA" Character Recognition Using Weighted Direction Index Histogram Method. *IEICE Trans. on Information and Systems*, J70-D (7): 1390-1397 [in Japanese]
- [11] Kimura F, Takashina K, Tsuruoka S, Miyake Y (1987) Modified Quadratic Discriminant Functions and the Application to Chinese Character Recognition. *IEEE Trans. Pattern Anal. Mach. Intell. PAMI-9*(1): 149-153
- [12] Yoshimura I, Yoshimura M (1991) Off-Line Writer Verification Using Ordinary Characters as the Object, *Pattern Recognition*, 24(9):909-915
- [13] Yoshimura M, Yoshimura I, Kim H. B (1993) A Text-Independent Off-Line Writer Identification Method for Japanese and Korean Sentences, *IEICE Trans. on Information and Systems*, E76-D (4): 454-461
- [14] Cheriet M, Kharram N, Liu C, Suen C Y (2007) *Character recognition systems*. Wiley & Sons Inc.: 293- 301
- [15] Ding K, Jin L (2010) Incremental MQDF Learning for Writer Adaptive Handwriting Recognition, *12th International Conference on Frontiers in Handwriting Recognition (ICFHR 2010)*: 559-564
- [16] Kawazoe Y, Ohyama W, Wakabayashi T, Kimura F (2010) Incremental MQDF Learning for Writer Adaptive Handwriting Recognition, *12th International Conference on Frontiers in Handwriting Recognition (ICFHR 2010)*: 410-414