

A Learning Pseudo Bayes Discriminant Method Based on Difference Distribution of Feature Vectors

Hiroaki Takebe, Koji Kurokawa, Yutaka Katsuyama, and Satoshi Naoi

FUJITSU LABORATORIES LTD.

4-1-1, Kamikodanaka Nakahara-ku, Kawasaki, 211-8588, Japan

{takebe.hiroaki, cross, katsuyama,
naoi.satoshi}@jp.fujitsu.com

Abstract. We developed a learning pseudo Bayes discriminant method, that dynamically adapts a pseudo Bayes discriminant function to a font and image degradation condition present in a text. In this method, the characteristics of character pattern deformations are expressed as a statistic of a difference distribution, and information represented by the difference distribution is integrated into the pseudo Bayes discriminant function. The formulation of integrating the difference distribution into the pseudo Bayes discriminant function results in that a covariance matrix of each category is adjusted based on the difference distribution. We evaluated the proposed method on multi-font texts and degraded texts such as compressed color images and faxed copies. We found that the recognition accuracy of our method for the evaluated texts was much higher than that of conventional methods.

1 Introduction

It is often difficult to recognize characters in deformed fonts compared with Mincho and Gothic, which are the most popular fonts for Japanese characters, and in texts degraded significantly depending on the input/output conditions. There have been a number of attempts to solve these problems. Nagy and Xu, for example, proposed a method in which character prototypes are extracted directly from a given text and recognition is performed based on template matching by using the prototypes [2]. Moreover, Ho proposed an adaptive text recognition method that does not depend on ground truth; it is used to first recognize the most frequently occurring words, and then obtain character prototypes from these words [3]. In these methods, the classifier can be “trained” only for the font and degradation condition of a given text, so that it could more accurately recognize the text. However, it is difficult to use these methods for Japanese OCR, because there are more than 3000 categories in Japanese. Thus, to solve the above problems with Japanese OCR, a discriminant function needs to be adapted dynamically to a given font and degradation condition. Omachi proposed methods that detect blurred regions in a character image, and modify a discriminant function based on the detected regions [4][5]. In these methods, eigenvalues and eigenvectors of the covariance matrix of each category are transformed into new values and vectors based on the blurred regions. Blurred regions are detected by

repeatedly “eroding” the character image. The detected blurred regions determine the vector operation in the feature vector space, which is performed for eigenvalues and eigenvectors of each category. However, this method is limited to blurred text images and cannot be used with font variations and the other image degradations. In this paper, we propose a method, that dynamically adapts pseudo Bayes discriminant function to any font and any image degradation condition. In this method, the characteristics of character pattern deformations due to font variations and image degradation are expressed as a statistic of a difference distribution, and information represented by the difference distribution is integrated into the pseudo Bayes discriminant function. The difference distribution represents autocorrelation of character pattern deformations, and the formulation of integration results in adjusting a covariance matrix of each category based on the difference distribution.

The difference distribution is defined and its integration into the pseudo Bayes discriminant function is formulated in Chapter 2. Chapter 2 also presents our experimental results of character recognition, which demonstrate the effectiveness of our method. In Chapter 3, we describe how this method can be used for text recognition and explain how the process of calculation can be made efficient. We also present our experimental results of text recognition, which demonstrate the effectiveness of the method.

2 Learning Pseudo Bayes Discriminant Method

2.1 Difference Distribution

The shapes of different character patterns in one font have common properties. For example, their strokes may be thin/thick or rounded/angular, and so on. The shapes of character patterns input/output by using the same process also have common properties. For example, their strokes in a particular direction may be blurred, and so on. These properties are reflected in the statistic of the difference distribution generated by difference vectors of the character patterns. A difference vector can be defined as follows. Let a feature vector of a character pattern of category C be \mathbf{x} and the average vector of category C be \mathbf{m}_C . Then a difference vector is defined as

$\mathbf{d} = \mathbf{x} - \mathbf{m}_C$. Next, a difference distribution is defined as follows. Let a character

pattern set be $P = \{\mathbf{x}_i | i = 1, 2, \dots, n\}$ and let the category of each pattern be known. Then we can obtain a set of difference vectors of P and let it be

$\{\mathbf{d}_i | i = 1, 2, \dots, n\}$ and the autocorrelation matrix be $\mathbf{S}_p = \frac{1}{n} \sum_{i=1}^n \mathbf{d}_i \mathbf{d}_i^t$. Then we

define the difference distribution of pattern set P as a normal distribution with autocorrelation matrix \mathbf{S}_p given as a covariance matrix. Autocorrelation matrix \mathbf{S}_p represents the average autocorrelation of pattern deformations of P . If there are a sufficient number of categories in P , \mathbf{S}_p approximates the matrix generated by

averaging the covariance matrices of all categories. This matrix may be used for pre-classification by multiple discriminant analysis (canonical discriminant method) [1].

2.2 Formulation of Integration

The pseudo Bayes discriminant function (a modified quadric discriminant function) can be used to accurately recognize greatly deformed character patterns such as handwritten ones [7][8]. Integrating information represented by the difference distribution into the pseudo Bayes discriminant function will result in more accurate recognition of characters in deformed fonts or in degraded texts. We formulate the integration as follows. First, we integrate the difference distribution into a quadric discriminant function. Let character pattern set P_0 be given and the category of each pattern of $P = \{\mathbf{x}_i \mid i = 1, 2, \dots, n\} \subset P_0$ be known. Let the difference distribution of P be D_p . Then we define a new discriminant function for each category as follows. A new discriminant function of a specific category C for a feature vector \mathbf{x} of an unknown pattern is an expected value of the quadric discriminant function of C obtained using difference distribution D_p as a probability density function. The discriminant function defined above is calculated by using the following equation (Appendix A)

$$\Phi_c(\mathbf{x}) = (\mathbf{x} - \mathbf{m}_c)^t (\boldsymbol{\Sigma} + \boldsymbol{\Sigma}_c)^{-1} (\mathbf{x} - \mathbf{m}_c) + \log |\boldsymbol{\Sigma} + \boldsymbol{\Sigma}_c|, \quad (1)$$

where \mathbf{m}_c is an average vector of category C , $\boldsymbol{\Sigma}_c$ is a covariance matrix of category C , and $\boldsymbol{\Sigma}$ is a covariance matrix of difference distribution D_p . This function equals a discriminant function obtained by replacing $\boldsymbol{\Sigma}_c$ with $\boldsymbol{\Sigma} + \boldsymbol{\Sigma}_c$ in the quadric discriminant function. This means that we can adjust the covariance matrix of a category by using the difference distribution. If the unknown pattern has the same deformation characteristics as those of the pattern set with the difference distribution, the recognition accuracy for the unknown pattern can be improved. This is illustrated in Fig.1. Recall that the formula for replacing $\boldsymbol{\Sigma}_c$ with $\boldsymbol{\Sigma} + \boldsymbol{\Sigma}_c$ is in common with the RDA method, which reduces the estimation error of eigenvalues of a covariance matrix generated by small samples [6]. Next, we shift Φ_c to a pseudo Bayes discriminant function. The pseudo Bayes discriminant function is obtained by equalizing small eigenvalues in the quadric discriminant function [7]. The computational and memory costs can be reduced by the approximation. Kimura approximated an optimal discriminant function where unknown parameters of a covariance matrix are estimated by using Bayes approach [8]. Here, in the same way as in the shift from a quadric discriminant function to a pseudo Bayes discriminant function [7], the following new discriminant function is defined by

$$\Psi_c(\mathbf{x}) = \frac{1}{\gamma_c^{l+1}} \left\{ \|\mathbf{x} - \mathbf{m}_c\|^2 - \sum_{i=1}^l \left(1 - \frac{\gamma_c^{l+1}}{\gamma_c^i} \right) \left((\mathbf{x} - \mathbf{m}_c) \cdot \mathbf{z}_c^i \right)^2 \right\} + \log \left(\prod_{i=1}^l \gamma_c^i \cdot \prod_{i=l+1}^n \gamma_c^{l+1} \right), \quad (2)$$

where γ_c^i is the i -th eigenvalue of $\Sigma + \Sigma_c$, \mathbf{z}_c^i is an eigenvector corresponding to γ_c^i , and l is an integer smaller than the dimension of the feature vector space n . Let pattern set \mathbf{P}_0 be given and a category of each pattern of $\mathbf{P} = \{\mathbf{x}_i \mid i = 1, 2, \dots, n\} \subset \mathbf{P}_0$ be known. We call it a learning pseudo Bayes discriminant method, to calculate the difference distribution of \mathbf{P} and to recognize unknown pattern of \mathbf{P}_0 based on the value of discriminant function $\Psi_c(\mathbf{x})$.

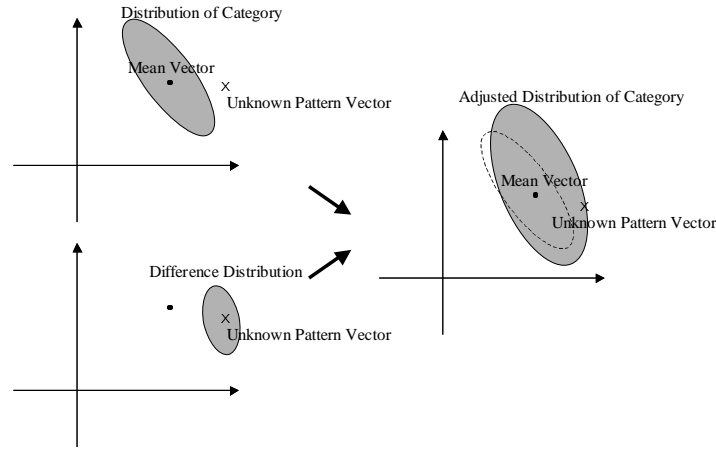


Fig. 1. Adjusting covariance matrix of category by using difference distribution.

2.3 Experimental Results

We used a chain code direction feature. A character image is divided into some sub-windows in the three different directions: horizontally, vertically and diagonally. The total number of sub-windows is 72. So, the dimension of the feature vector is 288. We evaluated the proposed method on about 50 thousand kanji characters in 17 fonts including Mincho, Gothic, Square, Gona, and Nar. For each font, 600 character patterns corresponding to 1/5 of all was supposed to be known, and a difference distribution was generated from the 600 character patterns. We applied the proposed discriminant function to the remaining unknown 2400 character patterns corresponding to 4/5 of all. The recognition accuracy of the proposed method for a total of 40 thousand character patterns was 97.3%, which was 6.8% better than that of the city block distance function and 1.1% better than that of the pseudo Bayes discriminant function.

Table 1. Character recognition accuracy.

City block	Pseudo bayes	Proposed method
90,4%	96,2%	97,3%

**Fig. 2.** Examples of fonts used in experiment.

3 Application for Text Recognition

3.1 Process Outline

We use the learning pseudo Bayes discriminant method for text recognition by dynamically calculating the difference distribution of the input text. First, the conventional recognition process is performed for a text. Then, the characters whose recognition probabilities are high are estimated to be correct, and the difference distribution is calculated from these characters. Next, eigenvalues and eigenvectors are adjusted by using the estimated difference distribution, which is the same as replacing Σ_c with $\Sigma + \Sigma_c$. After that, a second recognition process is performed for the text by using the pseudo Bayes discriminant function with the adjusted eigenvalues and eigenvectors. The estimated difference distribution reflects the information about the font used in the text and the image condition of the text. Thus our method can adaptively recognize characters based on the character pattern deformations in a given text. The recognition process is as follows. Here, we use a conventional pseudo Bayes discriminant function in the first recognition process. We use recognition probability to estimate correctly recognized characters, which is explained in Section 3.2. We use a high-speed approximation method using only principal eigenvalues and eigenvectors to adjust the eigenvalues and eigenvectors of each category, which is explained in Section 3.3.

3.2 Recognition Probability

We use recognition probability proposed in [9]. A recognition probability is obtained as follows. Let the discriminant function value of a character pattern for an i -th candidate character be d_i , and the reciprocal of d_i be $r_i = \frac{1}{d_i}$. If the difference between r_i and another r_j is large, the first candidate is certain to be correct. Then

$r = \frac{r_1}{r_1 + r_2 + r_3 + \dots}$ is defined as the certainty of recognition. We use $r = \frac{r_1}{r_1 + r_2}$ here. Next, the certainty of recognition r is transformed into the recognition probability based on a table $p(r)$, generated as follows. A discriminant function was performed for a large number of character patterns for the evaluation. Let the number of patterns whose certainty of recognition was r be $N(r)$, and let the number of patterns whose certainty of recognition was r and whose 1st candidates were correct be $N_{OK}(r)$. Then the recognition probability for r was defined as $p(r) = \frac{N_{OK}(r)}{N(r)}$.

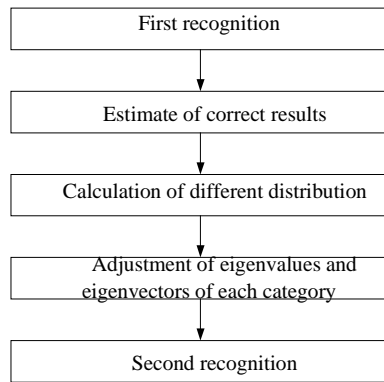


Fig. 3. Recognition process

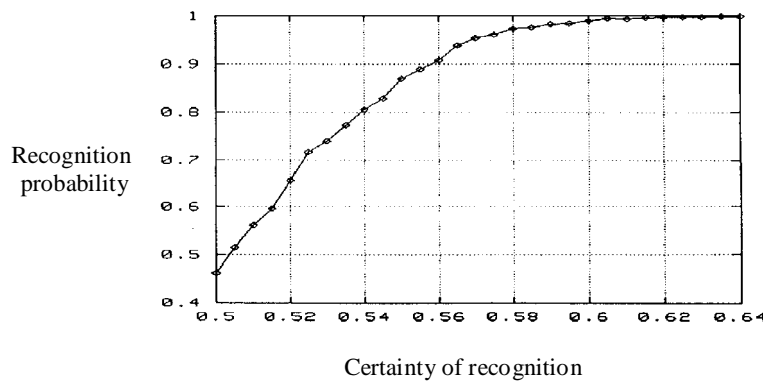


Fig. 4. Relation between certainty of recognition and recognition probability.

3.3 Adjusting Eigenvalues and Eigenvectors

Let the covariance matrix of category C be Σ_c , and the covariance matrix of the difference distribution be Σ . Then we need to calculate the eigenvalues and eigenvectors of $\Sigma + \Sigma_c$. However, calculating the eigenvalues and eigenvectors of all categories will take a very long time. Therefore we calculate the approximate values of the eigenvalues and eigenvectors of $\Sigma + \Sigma_c$ by using the principal eigenvalues and eigenvectors of Σ and Σ_c . The principal eigenvalues of a matrix are those in the higher ranks in the descending order of eigenvalues, and the principal eigenvectors are those corresponding to the principal eigenvalues. Let the principal eigenvalues of Σ be $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_s$, and the principal eigenvectors corresponding to those be $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s$. Let the principal eigenvalues of Σ_c be $\beta_1 \geq \beta_2 \geq \dots \geq \beta_t$, and the principal eigenvectors corresponding to those be $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_t$. Then the approximate matrices of Σ and Σ_c are, respectively, $\tilde{\Sigma}$ and $\tilde{\Sigma}_c$ which are defined below, and we can calculate the eigenvalues and eigenvectors of $\tilde{\Sigma} + \tilde{\Sigma}_c$.

$$\Sigma \approx \tilde{\Sigma} = \sum_{i=1}^s \alpha_i \mathbf{v}_i \mathbf{v}_i^t + \alpha \Sigma_{\mathbf{v}^\perp}, \quad \Sigma_c \approx \tilde{\Sigma}_c = \sum_{i=1}^t \beta_i \mathbf{w}_i \mathbf{w}_i^t + \beta \Sigma_{\mathbf{w}^\perp} \quad (3)$$

The calculation is shown in detail in Appendix B. This approximation enables the calculation of eigenvalues and eigenvectors of an n-dim square matrix to be replaced with the calculation of eigenvalues and eigenvectors of, at most, an s+t-dim square matrix.

3.4 Experimental Results

1) Multi-font Texts

We evaluated the proposed method on texts drawn by using three fonts, namely, FC-square, DF-pop, and FC-handwritten, in which characters are often seriously deformed compared with those in the most popular fonts. We prepared 20 texts. They were texts from newspapers, magazines, business documents, and papers. The total number of characters was 22600. The 20×3 texts were drawn using three fonts. The result was as follows.

2) Degraded Texts

Compressed color images are usually degraded and many share the same characteristics. Thus we evaluated the proposed method on compressed color texts. Twenty-two color images of texts from magazines were prepared in the BMP format,

Table 2. Character recognition accuracy for multi-font texts.

	City block	Pseudo bayes	Proposed method
FC-square	65,8%	78,0%	84,3%
DF-pop	84,4%	91,4%	94,2%
FC-handwritten	85,7%	93,9%	95,8%



Fig. 5. Examples used in experiment.

and they were compressed by 75% into the JPEG format. Then, they were transformed into binary images through binarization. We used the proposed method for the binary images. The total number of characters was about 13000. We also evaluated the proposed method on degraded faxed copies. Twenty texts from newspapers, magazines, and papers were prepared, and sent by fax and the faxed copies were then scanned. The total number of characters was about 22000. The result was as follows.

Table 3. Character recognition accuracy for degraded texts.

	City block	Pseudo bayes	Proposed method
Compressed images	93,8%	95,0%	95,9%
FAX images	96,5%	96,7%	97,4%

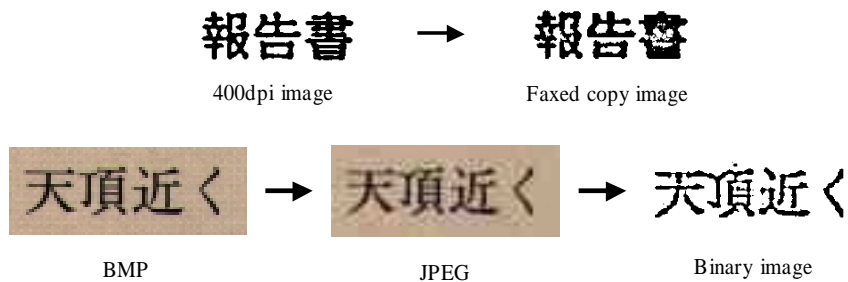


Fig. 6. Examples of texts used in experiment.

3) Processing Time

In the adjustment process of the eigenvalues and eigenvectors, those of $\Sigma + \Sigma_c$ are calculated for each category. The method of calculating directly them of $\Sigma + \Sigma_c$ took about 20 minutes by using a PC with a 1-GHz CPU. However, the method of calculating the approximate values of these eigenvalues and eigenvectors reduced the processing time to about 20 seconds. Moreover dimensionality reduction by feature selection enabled the processing time to be reduced from about 20 to about 4.2 seconds.

4 Conclusion

We developed a learning pseudo Bayes discriminant method, in which a pseudo Bayes discriminant function is dynamically adapted to the font and degradation condition present in a given text. In this paper, we first defined the difference distribution of a character pattern set, which represents the characteristics of character pattern deformations of the set, and then formulated the integration of the difference distribution into the pseudo Bayes discriminant function. This formulation results in that a covariance matrix of each category is adjusted based on the difference distribution. Next we used this method for text recognition. First, a recognition process is performed for an input text, and correct characters are estimated based on the recognition probability. A difference distribution is generated from the estimated correct characters. Then, the eigenvalues and eigenvectors of each category are dynamically adjusted by the estimated difference distribution, and the recognition process by using pseudo Bayes discriminant function is performed for the text once again. We evaluated the proposed method for multi-font texts and degraded texts such as compressed color images and faxed copies. The recognition accuracy of our method was much higher than that of conventional methods.

References

1. R.O.Duda and P.E.Hart, "Pattern Classification and Scene Analysis", pp.118–121, John Wiley & Sons, Inc., New York, 1973.
2. G.Nagy and Y.Xu, "Automatic Prototype Extraction for Adaptive OCR", Proceedings of Fourth International Conference on Document Analysis and Recognition, Ulm, Germany, August, pp.18-20 1997, 278–282.
3. T.K.Ho, "Bootstrapping Text Recognition from Stop Words", Proceedings of Fourteenth International Conference on Pattern Recognition, Brisbane, Australia, August, pp.17–20, 1998, 605-609.
4. S.Omachi and H.Aso, "A Qualitative Adaptation of Subspace Method for Character Recognition", Trans. of IEICE(D-II), vol. J82-D-II, No.11, pp.1930–1939, Nov., 1999.
5. S.Omachi, F.Sun, and H.Aso, "A Noise-Adaptive Discriminant Function and Its Application to Blurred Machine Printed Kanji Recognition", PAMI-22, 3, pp314–319, March 2000.

6. J.H.Friedman, "Regularized Discriminant Analysis", Journal of American Statistical Association, 84, No.405, pp.165–175, 1989.
7. S.Tsuruoka, M.Kurita, T.Harada, F.Kimura, and K.Miyake, "Handwritten "KANJI" and "HIRAGANA" Character Recognition Using Weighted Direction Index Histogram Method", Trans. of IEICE (D), vol.J70-D, No.7 pp.1390–1397 July, 1987.
8. F.Kimura, K.Takashina, S.Tsuruoka, and Y.Miyake, "Modified quadratic discriminant functions and the application to Chinese character recognition", IEEE Trans. PAMI, vol.9, no.1, pp.149–153, 1987.
9. K.Fujimoto and H.Kamada, "Fast and Precise Character Recognition by Estimating Recognition Probability", Proceeding of the 1996 Information and Systems Society Conference of IEICE, D-361, Sep. 1996.

Appendix A: Formulation of Discriminant Function

Function $q(\mathbf{m}_1, \mathbf{m}_2, \Sigma_1, \Sigma_2)$ obtained by summing up a normal distribution of average \mathbf{m}_1 and covariance matrix Σ_1 weighted with a normal distribution of average \mathbf{m}_2 and covariance matrix Σ_2 , is calculated based on the following equation.

$$\begin{aligned}
 & q(\mathbf{m}_1, \mathbf{m}_2, \Sigma_1, \Sigma_2) \\
 &= \int_{\mathbb{R}^d} \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_1|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m}_1)^t \Sigma_1^{-1} (\mathbf{x}-\mathbf{m}_1)\right\} \cdot \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_2|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m}_2)^t \Sigma_2^{-1} (\mathbf{x}-\mathbf{m}_2)\right\} d\mathbf{x} \\
 &= \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_1 + \Sigma_2|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{m}_1 - \mathbf{m}_2)^t (\Sigma_1 + \Sigma_2)^{-1} (\mathbf{m}_1 - \mathbf{m}_2)\right\}
 \end{aligned}$$

A new discriminant function of specific category C for feature vector \mathbf{x} of an unknown pattern is defined by $\varphi_c(\mathbf{x}) = -2 \log q(\mathbf{x}, \mathbf{m}_c, \Sigma, \Sigma_c)$, where \mathbf{x} is a n-dimensional feature vector of an input image, \mathbf{m}_c is an average vector of category C, Σ_c is a covariance matrix of category C, and Σ is a covariance matrix of the difference distribution of pattern set P. By removing the constant terms from this equation, we obtain the following discriminant function.

$$\Phi_c(\mathbf{x}) = (\mathbf{x} - \mathbf{m}_c)^t (\Sigma + \Sigma_c)^{-1} (\mathbf{x} - \mathbf{m}_c) + \log |\Sigma + \Sigma_c|$$

Appendix B: Calculation of Eigenvalues and Eigenvectors

We assume that a vector subspace covered by $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s$ and a vector subspace covered by $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_t$ are \mathbf{V} and \mathbf{W} , respectively, and that a minimal vector subspace including \mathbf{V} and \mathbf{W} is \mathbf{U} . Then, both $\mathbf{w}'_1, \mathbf{w}'_2, \dots, \mathbf{w}'_u$ and $\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_t$ that satisfy the following equation are calculated by Gram-Schmidt orthogonalization.

$$\mathbf{U} = \mathbf{V} + \mathbf{W} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s, \mathbf{w}'_1, \mathbf{w}'_2, \dots, \mathbf{w}'_u\} = \{\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_v, \mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_t\}$$

And matrices $\mathbf{V}_W, \mathbf{W}_V, \mathbf{P}, \mathbf{A}$ and \mathbf{B} are defined as

$$\mathbf{V}_W = (\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_s \quad \mathbf{w}'_1 \quad \mathbf{w}'_2 \quad \dots \quad \mathbf{w}'_u),$$

$$\mathbf{W}_V = (\mathbf{w}_1 \quad \mathbf{w}_2 \quad \dots \quad \mathbf{w}_t \quad \mathbf{v}'_1 \quad \mathbf{v}'_2 \quad \dots \quad \mathbf{v}'_v),$$

$$\mathbf{P} = \mathbf{V}_W^t \mathbf{W}_V, \quad \mathbf{A} = \begin{pmatrix} \alpha_1 & & & & & \\ & \ddots & & & & \\ & & \alpha_s & & & \\ & & & \alpha & & \\ & & & & \ddots & \\ & & & & & \alpha \end{pmatrix}, \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} \beta_1 & & & & & \\ & \ddots & & & & \\ & & \beta_t & & & \\ & & & \beta & & \\ & & & & \ddots & \\ & & & & & \beta \end{pmatrix}.$$

Then we can show that

$$\tilde{\Sigma} + \tilde{\Sigma}_c = \mathbf{V}_W (\mathbf{A} + \mathbf{PBP}^t) \mathbf{V}_W^t + (\alpha + \beta) \Sigma_{U^\perp}.$$

Thus matrix $\mathbf{A} + \mathbf{PBP}^t$ is diagonalized as $\mathbf{A} + \mathbf{PBP}^t = \mathbf{X}\Gamma\mathbf{X}^t$,

$$\text{where } \mathbf{X} = (\mathbf{x}_1 \quad \mathbf{x}_2 \quad \dots \quad \mathbf{x}_{s+u}) \text{ and } \Gamma = \begin{pmatrix} \gamma_1 & & & & & \\ & \gamma_2 & & & & \\ & & \ddots & & & \\ & & & & & \\ & & & & & \gamma_{s+u} \end{pmatrix}.$$

Then the obtained diagonal components equal $\gamma_1, \gamma_2, \dots, \gamma_1$, and $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_1$ are calculated according to the following equation $(\mathbf{z}_1 \quad \mathbf{z}_2 \quad \dots \quad \mathbf{z}_{s+u}) = \mathbf{V}_W \mathbf{X}$.