Grayscale Feature Combination in Recognition based Segmentation for Degraded Text String Recognition

Jun Sun

Fujitsu R&D Center
Eagle Plaza B10th floor, Xiaoyun Rd. No.26
Chaoyang Dist.Beijing,100016, P.R.China
sunjun@frdc.fujitsu.com

Yoshinobu Hotta, Katsuhito Fujimoto
Yutaka Katsuyama, Satoshi Naoi
Fujitsu Laboratories Ltd
Kawasaki, Japan
y.hotta, fujimoto.kat@jp.fujitsu.com
katsuyama, naoi.satoshi@jp.fujitsu.com

Abstract

Grayscale feature is very effective for degraded character recognition. While many papers focus on different feature extraction algorithms on single character recognition, few deals with the impact of the selected feature on segmentation. For recognition-based segmentation, a good recognition performance on single character may not always have good performance on segmentation. In this paper, two types of grayscale feature, the R-Feature and the S-Feature, are proposed based on dual-eigenspace decomposition. The R-Feature is suitable for single character recognition. The S-Feature is suitable for text string segmentation. These two feature are combined to further improve the performance for degraded Japanese text string recognition.

1. Introduction

Degraded character recognition is a very important topic in digital camera based document image processing. For degraded character image, as the degradation level increases, the performance of binarization drops dramatically, which causes big problem for binary-image-based feature extraction. In such case, feature directly extracted from grayscale image has more advantages in keep the shape and structure information of the character.

While many grayscale feature extraction methods have been proposed in recent years[1][2][3][4], they all concentrated on single character recognition. Few paper deals with the performance of the feature on character segmentation. For real text string recognition, recognition-based segmentation is a commonly used strategy[7]. As shown in Figure 1,the first row is the grayscale image of the degraded text string. The second row is the binarization result. Based on connected component analysis on the binary image and the estimated average character width, the dissection step produces two kind of segments: the basic segment (the third row) and the synthesized segment (the fourth row). The basic segment is the un-separable basic unit in image based segmentation. One character image might contains one or more basic segments (for example, left and right structure character image or broken character image). The synthesized segment includes the combination of two or more basic segments under the constraint of the aspect ratio of the segment. During dissection(image based segmentation), no recognition is involved. Therefore one segment might contain one or more characters(touching cases). Also, one character might contain one or more basic and/or synthesized segments. After character dissection, every segment is recognized as some character category. Finally, DP searching is performed to search the segment combination with the minimum total cost. The advantage of recognition-based segmentation is that there is no strict requirements on the dissection step. The recognition is embedded inside the segmentation.

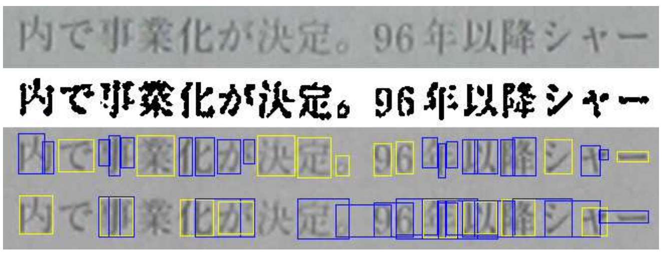

Figure 1: Segmentation of degraded text string.

In the case of degraded text string recognition, more difficulties come out for recognition based segmentation:

1) Broken characters increase the number of segments: a broken character image will be dissected into two or more basic segments. As a result, the number of basic and synthesized segments increases dramatically (the 3rd and 4th row in Figure 1). The correct segments (yellow rectangles) are merged into the incorrect segments (blue rectangles).

2) Difference between character image and non-

character image is diminished by heavy degradation from blurring and low resolution.

It should be noticed that the "recognition" in segmentation is not the same as the "recognition" in single character recognition. The former focuses on how to separate character from non-character. The latter always assumes the input is an character image with unknown category. This situation is similar with the difference between face detection and face recognition.

In this paper, two types of grayscale features derived from dual-eigenspace decomposition are compared. The R-Feature is good at single character recognition. The S-Feature is good at separating character image from non-character image. A combination of the two features can further improve the performance of text string recognition.

2. Grayscale feature extraction and recognition

Many camera based document images contain heavy degradation caused by low resolution, blurring and distortion, which will bring great trouble to binarization. One solution to overcome this problem is to extract the feature directly from the grayscale character image. The grayscale features proposed in this paper are based on eigenspace decomposition [5][6].

2.1 Dual eigenspace decomposition

Dual eigenspace decomposition includes two kind of eigenspace: the unitary eigenspace and the individual eigenspace. The unitary eigenspace serves as the 1st level feature extraction and coarse classification. The individual eigenspace is built on the feature extracted from the unitary eigenspace and is used for the 2nd level feature extraction and fine classification. This coarse to fine recognition structure efficiently improves the processing speed, which makes real time recognition possible.

The unitary eigenspace is built from character images of all categories. suppose a character image with size of $ndim = w * h$ is represented by a vector $x = [x_1, x_2, \ldots, x_{ndim}]^T$ using the raster scanning order. The unitary eigenspace is constructed by Principal Component Analysis (PCA) on the covariance matrix of character template images of all categories:

$$COV = \frac{1}{P_c} \sum_{i=1}^{P} \sum_{j=1}^{N_c} (m_{ij} - m)(m_{ij} - m)^T, \qquad (1)$$

where P is the number of the character categories. N_c is the number of templates for every category. $P_c = P * N_c$ is the number of total character template images in all categories. m is the mean vector for all character template images. m_{ij} is the jth character template image in the ith category. The first n eigenvectors of matrix corresponding to the first largest n eigenvalues spans the unitary eigenspace, $U = [u_1, u_2, \ldots, u_n]^T$.

Since the unitary eigenspace is constructed on the samples of all categories, the discrimination power is not strong enough. In order to further improve the recognition performance, an individual eigenspace is built for every character category using the PCA feature from the unitary eigenspace. The covariance matrix for the ith category is obtained as:

$$\begin{aligned} COV_i &= \frac{1}{M_i} \sum_{k=1}^{M_i} (y_i^{(k)} - c_i)(y_i^{(k)} - c_i)^T, \\ y_i^{(k)} &= U^T (x_i^{(k)} - m) \\ c_i &= \frac{1}{N_c} \sum_{j=1}^{N_c} c_{ij} \qquad i = 1, 2, \ldots, P \end{aligned} \qquad (2)$$

where $y_i^{(k)}$ is the PCA feature of the kth training sample in the ith category. c_{ij} is the PCA feature of the jth character template image in the ith category, m_{ij}. c_i is the mean feature of the ith category. M_i is the number of training samples for the ith category. The first n_i eigenvectors of COV_i corresponding to the first n_i largest eigenvalues, $\tilde{U}_i = [u_1^i, u_2^i, \ldots, u_{n_i}^i]^T$, spans the individual eigenspace for the ith category.

2.2 Character image normalization

In order to remove the influence of degradation, precise image registration is first performed on every segment image[4]. The registered segment image has uniform size and the brightness of the image pixel value is compensated.

Figure 2 shows the result of the registration of some segments in Figure 1. The top row is the images of the correct segments. The second row shows the images of the basic segments. The third and the fourth row are the images of synthesized segments.

In order to further improve the recognition performance against degradation, definite canonicalization [8] is used to filter the mean value of the image:

$$c = (1/\sqrt{n}, ..., 1/\sqrt{n}), \qquad (3)$$

$$x' = x - (c \cdot x)c \qquad (4)$$

Finally, the energy of normalized vector, x', is regulated into unit length. Intuitively, the normalization step transfer the character image from inside a hypercube with lattice length of 255 into the surface of a hypersphere.

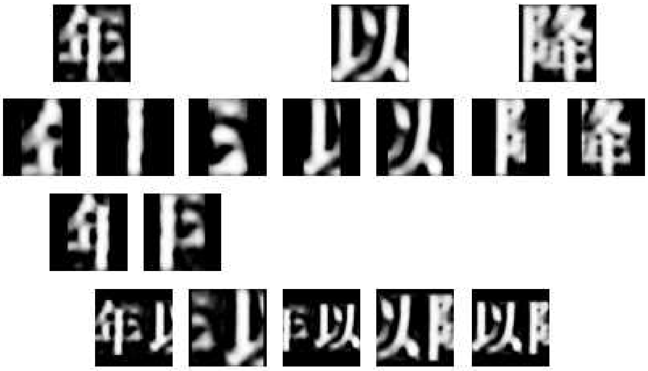

Figure 2: Segment image registration.

2.3 R-Feature extraction and recognition

The unitary eigenspace can be regarded as a transformation from the normalized image domain to the frequence domain. The individual subspace is built upon the frequence space. The R-Feature is extracted in the frequence domain.

First, the PCA feauture is extracted from the normalized character image as in Equation (5) by the unitary eigenspace, U:

$$y = U^T(x' - m). \tag{5}$$

Then, coarse classification is performed by matching the feature, y, with all template feature, c_{ij}. Assuming the first N_{cand} candidate character categories are selected by minimum Euclidean distance, the R-Feature is the reconstructed feature for every category as in Equation (6) by the individual eigenspace.

$$\begin{aligned} \eta_k &= \tilde{U}_k^T(y - c_k), \\ \hat{y}_k &= \tilde{U}_k^T \eta_k + c_k \end{aligned} \tag{6}$$

During the recognition phase, the recognition distance of the R-Feature is taken as the norm of the difference of the PCA feature and its reconstruction:

$$d_k^R = \| y - \hat{y}_k \| \tag{7}$$

The classification is accomplished by sorting the N_{cand} character categories according to the minimization of their recognition distance.

2.4 S-Feature extraction and recognition

Different with the R-Feature, the S-Feature is extracted from the original image domain. First, the R-Feature is recovered back into the normalized image domain:

$$\hat{x}_k = U \cdot \hat{y}_k + m \tag{8}$$

Then the recovered vector is scaled into the original image space:

$$\begin{aligned} \ddot{x}_k(i) &= 255 * (\hat{x}_k(i) - m_1)/(m_2 - m_1) \\ m_1 &= \min\{\hat{x}_k(i)\}, \qquad i = 1, 2, \ldots, n \\ m_2 &= \max\{\hat{x}_k(i)\}, \qquad i = 1, 2, \ldots, n \end{aligned} \tag{9}$$

The physical meaning of the S-Feature, $\ddot{x}_k$, is very clear: it is the restored version of the input segment image by the two-fold eigenspace.

Equation (9) is the reverse transformation of the normalization step in section 2.2, which leverages the basic nature of the registered character image: dark background with pixel value near 0, bright character stroke with pixel value near 255.

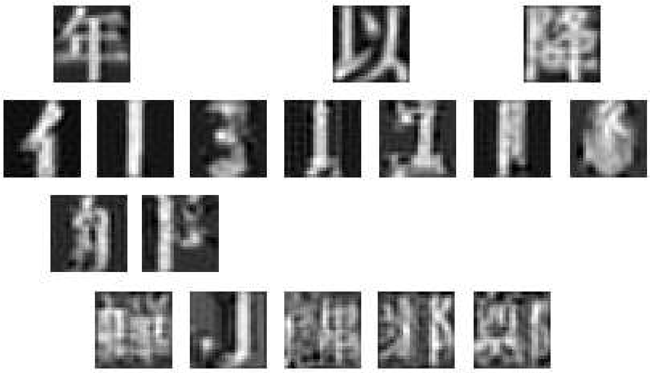

Figure 3: S-Feature of Figure 2

Figure 3 shows the corresponding S-Feature image of the segments displayed in Figure 2. By comparing Figure 2 with Figure 3, we can see that for character image, the restored image can well preserve the structure information. However, for non-character image like the segments in the 2nd, 3rd, and 4th row in Figure 2, much more noise are introduced by the reverse transformation.

This phenomenon is because the unitary eigenspace and the individual eigenspace are all built using character images only. The noise in the reconstruction caused by a degraded version of a character image is much smaller than the noise caused by a non-character image. If the input is a non-character image, the individual eigenspace will choose a most similar character category and calculate the reconstructed R-Feature. Then, the noise introduced by the R-Feature is transferred by the unitary eigenspace back into the image space. The bigger the noise is, the more uneven happens for the pixel value among the background pixels and the pixel value among the stroke pixels. Finally, these unevenness is enhanced by the image reconstruction operation in Equation (9). This special property of the S-Feature makes it very suitable for the segmentation task.

In order to better calculate the difference between two character structure, angle based distance is used instead of Euclidean distance in calculating the recognition distance of the S-Feature:

$$d_k^S = 1 - \frac{x \bullet \ddot{x}_k}{\| x \| \| \ddot{x}_k \|} \quad (10)$$

3. Performance evaluation

For the R-Feature and the S-Feature defined in the previous section, the performance of recognition and segmentation are evaluated separately.

Two testing sets are used here. Testing set 1 is used to evaluate the performance of single recognition, which includes Level-1 Japanese Kanji characters with 19 fonts. These Kanji characters are first typed into slides and the slides are projected onto a screen. Then digital camera is used to capture the character image in three different distances, S1, S2 and S3. Total number of Kanji characters in every distance is 56,335.

Testing set 2 is used to evaluate the performance on segmentation , which includes degraded text string images captured from "Nikkei Business Magazine" with a Canon PowerShot A80 digital camera under maximum 4 MegaPixels resolution. Total number of text string is 292. Total number of characters is 4491. The correct segmentation result (groundtruth) is labelled manually for every image in testing set 2.

3.1 Performance evaluation for single character recognition

For comparison, PCA feature extracted from unitary eigenspace is also evaluated for single character recognition along with the R-Feature and the S-Feature. Table 1 lists the recognition rate of the R-Feature, S-Feauture and PCA feature on testing set 1.

Table 1: Single character recognition rate of R-Feature, S-Feature and PCA feature(%)

Dataset	S1	S2	S3
R-Feature	**92.80**	**97.46**	**98.51**
R-Feature*	91.47	96.26	97.09
R-Feature**	92.63	97.39	98.40
S-Feature	90.88	96.31	97.82
PCA feature	83.06	89.77	91.97

"R-Feature*" is the result using dual-eigenspace trained by registered character images. "R-Feature**" is the result using dual-eigenspace trained by definite canonicalized registered character images. "R-Feature" is the result using dual-eigenspace trained by the images after normalization operation defined in section 2.2.

The comparison between R-Feature, R-Feature*, and R-Feature** shows the effectiveness of different normalization methods. What's more, we can see from Table 1 that for all 3 distance, $S1$, $S2$, and $S3$, the Dual-eigenspace based features are better than PCA feature. That is because the unitary eigenspace is based on all character categories, the discrimination power is not strong.

Finally, the performance of the R-Feature is better than that of the S-Feature in single character recognition. That means for character image, recovery from the common unitary eigenspace "blurs" the difference between similar characters. Hence, the S-Feature is not a suitable feature for single character recognition.

3.2 Performance evaluation for segmentation

3.2.1 Segmentation error rate

The performance of segmentation is represented by the segmentation error rate. For every feature, recognition based segmentation is evaluated on testing set 2. The segmentation result is compared with the groundtruth. The segmentation error rate is defined as:

$$Err_{seg} = \frac{num.\ of\ mismatched\ segments * 100}{num.\ of\ total\ groundtruth\ segments} \quad (11)$$

If the region of a segment cannot overlap completely with any of the groundtruth, we call that segment a "mismatched segment". Therefore, the segmentation error rate is the complement of the recall rate of the segmentation. Table 2 lists the segmentation error rate for the R-Feature, S-Feature and PCA feature. The number of groundtruth segments is 4523.

Table 2: Segmentation performance of R-Feature, S-Feature and PCA feature.

	$mismatch\ num.$	Err_{seg}
$PCAfeature*$	2146	47.45%
$PCAfeature$	500	11.05%
$R-Feature$	121	2.68%
$S-Feature*$	130	2.87%
$S-Feature$	**23**	**0.51**%

The "PCA feature*" and "S-Feature*" in Table 2 are the result without the image reconstruction step defined in Equation 9. Two conclusions can be drawn from the segmentation experiments:

(1) The individual eigenspace is helpful in segmentation.

(2) The image reconstruction step is very effective. Simply transferring the R-Feature back to n-dimensional space by the unitary eigenspace cannot necessarily improve the

segmentation performance. The error rate of both PCA feature and the S-Feature all dramatically decreases by the image reconstruction operation.

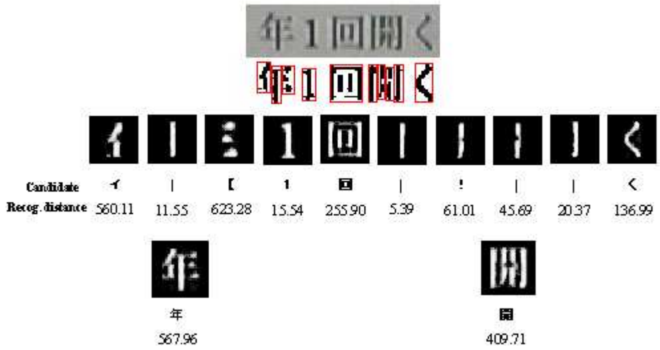

Figure 4: Segmentation result using R-Feature

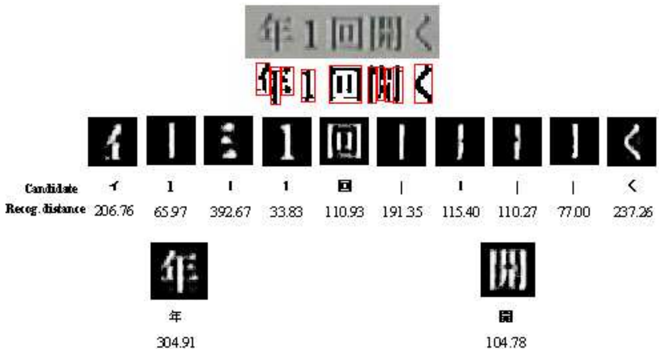

Figure 5: Segmentation result using R-Feature

Figure 4 and Figure 5 show the segmentation results using R-Feature and S-Feature respectively. The first row is the original degraded grayscale text line image. The second row is the binarization result that is used for coarse segmentation. The third row is the normalized image for the basic segments. The fourth row is the normalized image for two synthesized segments. Below the normalized image is the recognition result and the corresponding recognition distance. Notice the recognition result in two figures are not exactly the same because different feature is used. Compared with R-Feature, the recognition distance of S-Feature is more reasonable. The fourth character in Figure 4 will be mis-segmented into 4 components. But in Figure 5, the character can be correctly segmented because the recognition distance is far less than the sum of its four components.

3.2.2 Relative recognition distance

The reason of the success of the S-Feature in segmentation is further analyzed by the distribution of the relative recognition distance (RRD). In order to calculate the RRD for a segment image, the segments generated from a text string are first classified into 3 categories: correct-segment, under-segment, and over-segment.

The correct-segment is a basic segment or synthesized segment that can overlap with a groundtruth segment perfectly. The segment images in the first row of Figure 2 are correct-segments.

The under-segment is a basic segment or synthesized segment that is embedded inside a groundtruth segment. The segment images in the second and the third row in Figure 2 are under-segments.

The over-segment is a synthesiz segment that intersects two or more groundtruth segments. The segments in the fourth row in Figure 2 are over-segments.

The original recognition distance obtained by Equation 7 and Equation 10 is very sensitive to the degradation level of a particular text image. Hence for every segment, the RRD is defined according to the relationship between the position of the segment and the position of the interfered correct-segment:

$$RRD_c = 1 \tag{12}$$

$$RRD_u = \frac{d(SEG_u)}{d(SEG_c)},\ SEG_u \subset SEG_c \tag{13}$$

$$RRD_o = \frac{d(SEG_o)}{\sum_{j=1}^{m} d(SEG_c^j)/m}, \tag{14}$$

$$SEG_o \bigcap SEG_c^j \neq \emptyset.$$

Where SEG is the image region of the segment. $d()$ stands for the recognition distance obtained by the R-Feature or the S-Feature. m is the number of correct-segments that intersect with the over-segment.

Figure 6 shows the histogram of RRD_u of the R-Feature and the S-Feature. If a correct-segment can be decomposed into two or more under-segments, and the summed RRD value of the under-segments is less than 1, segmentation error will occur. We can see from the figure that the number of small-value RRD for the S-Feature is much less than that of the R-Feature. That is very helpful for the segmentation task. Figure 7 shows the histogram of RRD_o of the R-Feature and the S-Feature. Again, the S-Feature has more large value RRD_o than that of the R-Feature.

The fact that the S-Feature has less small-RRD-valued segments than the R-Feature proves that the S-Feature is a more suitable feature for recognition based segmentation.

4. Feature combination for degraded text string recognition

The analysis in the previous section shows that the S-Feature and the R-Feature are complementary: the S-Feature is better in recognition-based segmentation. The

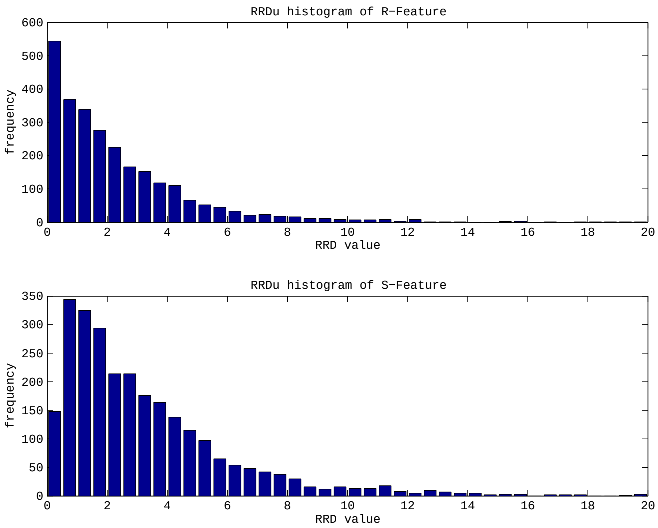

Figure 6: RRD_u histogram of R-Feature (top) and S-Feature (bottom)

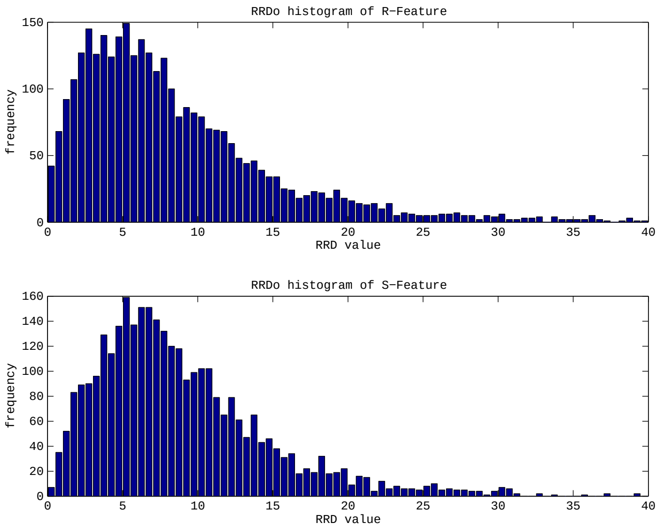

Figure 7: RRD_o histogram of R-Feature (top) and S-Feature (bottom)

R-Feature is better in single character recognition. Thus, a combination strategy is proposed for degraded text string recognition.

(1) Characters in the text string is segmented using the S-Feature by recognition-based segmentation.

(2) The recognition result of every segmented character images is refined by the R-Feature.

Notice that both R-Feature and S-Feature are all derived from the same dual-eigenspace decomposition. Therefore the computation time can be effectively reduced.

Performance evaluation of feature combination is conducted using testing set 2 and testing set 3. The images in testing set 3 include 90 text blocks with variant character size, the minimum size of character is as small as 10*10 pixels. Total number of characters in testing set II is 10253.

Table 3: Recognition of feature combination.(%)

Dataset	R-Feature	Feature combination
testing set 2	93.94	**95.79**
testing set 3	76.82	**80.88**

As shown in Table 3, by combining the S-Feature with the R-Feature, the overall recognition performance can be improved effectively.

5. Conclusions

The impact of grayscale feature on text string segmentation is discussed in this paper. Two kinds of features are proposed for degraded text string recognition: the S-Feature is used for text string segmentation, the R-Feature is used to recognized segmented individual characters. Experiments prove the effectiveness of the feature combination.

References

[1] Wang X. W., Ding X. Q., and Liu C. S. Optimized Gabor filter based feature extraction for character recognition. *Proc. of ICPR*, pp. 223-226, 2002.

[2] Yoshimura, H., Etoh, M., Kondo, K., et al. Gray-scale character recognition by gabor jets projection. *Proc. of ICPR* pp.335-338, 2000

[3] Wang, L., Pavlidis, T. Direct Gray-Scale Extraction of Features for Character Recognition. *IEEE trans. Pattern Analysis and Machine Intelligence* 15(10) pp.1053-1067, 1993

[4] Sun, J., Hotta, Y., Katsuyama, Y., Naoi, S. Low resolution character recognition by dual eigenspace and synthetic degraded patterns. *1st ACM workshop on Hardcopy Document Processing* pp.15-22, 2004.

[5] Duda, R. O., Hart, P. E., Stork, D. G. *Pattern classification, second edition.* A Wiley-Interscience Publication John Wiley & Sons, Inc. pp.568 569, 2001.

[6] Zhang, D., Peng, H., Zhou, J., Sankar, K. P. A novel face recognition system using hybrid neural and dual eigenspace methods. *IEEE trans. System, Man and Cybernetics - part A* 32(6) pp.787-792, 2002

[7] Casey, R. G., Lecolinet, E. A Survey of Methods and Strategies in Character Segmentation. *IEEE trans. Pattern Analysis and Machine Intelligence* 18(7) pp.690-706, 1996

[8] Iijima, T. Theory of pattern recognition. Series of basic information technology 6. *Morikita Publishing Company Ltd.,* 1989