Three Dimensional Rotation–Free Recognition of Characters

Ryo Narita, Wataru Ohyama, Tetsushi Wakabayashi and Fumitaka Kimura
Graduate School of Engineering
Mie University
Tsu-shi, Mie, Japan
Email: {narita, ohyama, waka, kimura}@hi.info.mie-u.ac.jp

***Abstract*—In this paper, we propose a new method for three dimensional rotation-free recognition of characters in scene. In the proposed method, we employ the Modified Quadratic Discriminant Function (MQDF) classifier trained with samples generated by three-dimensional rotation process in a computer. We assume that when recognizing individual characters, considering three-dimensional rotation can approximately handle the recognition of perspectively distorted characters. The results of the evaluation experiments using printed alphanumeric characters as an evaluation data set, consisting of approximately 600 samples/class for 62 character classes, show that the recognition rate is 99.34% for rotated characters while it is 99.59% for non rotated characters. We have empirically confirmed that the rotated characters given as the training data set do not negatively affect significantly to recognition of non rotated characters. Moreover, 437 characters extracted from 50 camera-captured scenes were correctly recognized and the feasibility of real world application of our method has been confirmed. Finally we describe on three dimensional rotation angle estimation of characters for detecting local normal of the surface on which the characters are printed aiming to scene analysis by shape from characters.**

***Keywords*-rotation-free character recognition; camera-based character recognition; character-based scene analysis; shape form characters;**

I. Introduction

Recent improvement of size and quality of images taken by commercial digital camera devices raises the demand of using them as an input device for ubiquitous character recognition. If such devices capable of recognizing characters in real three–dimensional (3D) scene were available, they are expected having a variety of applications. For instance, more suitable car navigation system can be realized by integration with traffic sign recognition. Translation of words or sentences on signboards can be done in the same way.

When we capture characters in real 3D space using cameras, the characters on captured image are usually rotated in 3D and recieve perspective distortion. The 3D rotation and perspective distortion easily deteriorate the performance of character recognition.

Many researches have been done in order to recognize camera–captured characters [1]. One of the conventional methods is to extract the text line from an image and recognize the characters after skew correction of the text line [2]. This approach does not work if the characters are not printed along straight line nor text line is extracted correctly.

Figure 1. A standard character.

Figure 2. The characters rotated around x–axis, y–axis and z–axis by 45 degree.

In this paper, we propose a new method of 3D rotation–free recognition of characters in scene. In this method, we employ the Modified Quadratic Discriminant Function (MQDF) classifier trained with samples generated by 3D rotation process in a computer. Hence there is no need to correct rotation of input characters and it is possible to recognize rotated characters without increasing processing time. We assume that when recognizing individual characters, considering 3D rotation can approximately handle the recognition of perspectively distorted characters.

On the contrary, affine invariant character recognition which is robust for rotatory or projective distortion and real-time layout-free character recognition [3] have been proposed. In the similar fashion of our method, Hase et al. applied parametric eigen-space method [4] to recognize rotated characters [5]. However, their method is only focusing on two-dimensional rotation, so that it is quite different from our method that considers 3D rotation and can handle projective distortion. We also study on 3D rotation angle estimation of characters for detecting local normal of the surface on which the characters are printed aiming to scene analysis by shape from characters. We will call a character with no inclination or rotation as a standard character (Figure.1) and that has rotation as a rotated character (Figure.2).

II. Proposed method

The processing flow of proposed method is shown in Figure 3. Corresponding the center of the enclosing rectangular of a standard character to the origin of 3D left

1520-5363/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDAR.2011.169

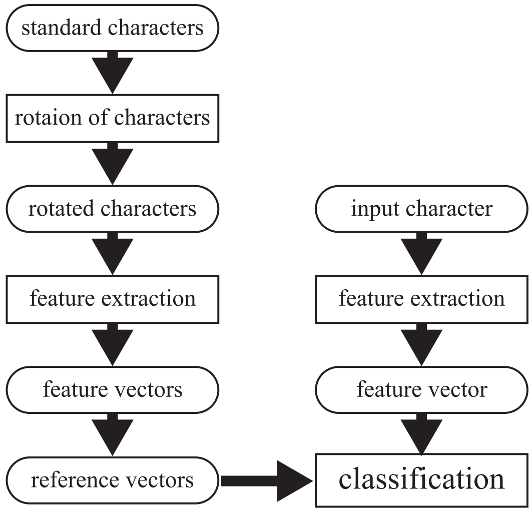

Figure 3. The processing flow of proposed method.

handed Cartesian coordinate system, the rotation processing is performed on a computer to generate 3D rotated character images.

After generating rotated characters, feature vectors are extracted from those rotated characters and the reference vectors are obtained for each character class. When recognizing input characters, we extract feature vectors without correcting its rotation and recognize it by the MQDF classifier.

We use Weighted Direction code Histogram (WDH) as a feature vector [6] and the MQDF as a classifier [7].

The WDH feature vector of size 392 is extracted by the following procedure.

1) The chain coding is applied to the contour pixels of the normalized character image. Vector sum of adjacent two chain elements is taken to procedure 16 directional code.
2) The normalized character image is divided into 169 (13 horizontal $\times$ 13 vertical) blocks. The number of the contour pixels in each direction is counted in each block to produce 169 local direction code histograms.
3) The spatial resolution is reduced from 13×13 to 7×7 by down sampling every two horizontal and every two vertical blocks with 5×5 Gaussian filter. Similarly the directional resolution is reduced from 16 to 8 by down sampling with a weight vector $[1\ 2\ 1]^{\mathrm{T}}$, to produce a feature vector of size 392 (7 horizontal, 7 vertical, and 8 directional resolution).
4) Variable transformation taking square root of each feature element is applied to make the distribution of the features Gaussian–like.

The 5×5 Gaussian filter and the weight vector $[1\ 2\ 1]^{\mathrm{T}}$ in the step 3) are the high–cut filters to reduce the aliasing due to the down sampling. Their size was empirically determined for this purpose.

The MQDF is defined by

$$g(X) = \frac{1}{\alpha\sigma^2}\Bigg[\|X - M\|^2 - \sum_{i=1}^{k} \frac{(1-\alpha)\lambda_i}{(1-\alpha)\lambda_i + \alpha\sigma^2}\left\{\Phi_i^{\mathrm{T}}(X - M)\right\}^2\Bigg] + \sum_{i=1}^{k} \ln{(1-\alpha)\lambda_i + \alpha\sigma^2} \quad (1)$$

where X and M are feature vector and the mean vector of a class respectively, and λ_i and Φ_i are the i-th eigenvalue and eigenvector of the covariance matrix, respectively, $\sigma^2 I$ and α are an initial estimates of the covariance matrix and a confidence constant, respectively. The class which minimizes $g(X)$ is selected as the recognition result. The required computation time and storage is $\mathrm{O}(kn)$.

III. Evaluation experiment of recognition rate

In order to check the feasibility of our method, we have done the recognition experiment to compare the recognition rate when learning only standard characters and learning rotated characters (including standard characters).

We used printed alphanumeric characters as an evaluation data set consisting of approximately 600 samples/class for 62 character classes.

There are serial numbers for each data and we used even number data as learning data, odd number data as evaluation data. We have changed the rotation angle by intervals of 15 degree for x-axis and y-axis within the range of -45 to 45 degree, z-axis within the range of -30 to 30 degree. When rotating character images, we have done linear interpolation to decrease the aliasing which comes from rotation of images. Moreover, we have done the normalization of character size in three ways, such as (1) normalization maintaining aspect ratio, (2) independent normalization of width and height, (3) independent normalization of width, height and center of gravity, in the preliminary experiment. The best result of the recognition rate was obtained when normalizing in the way (1). Therefore, we normalized learning data before and after rotating its character and normalized evaluation data before feature extraction by maintaining aspect ratio. We normalized character size into 52×52 pixels.

Figure 4 shows the recognition rates. Since some of the characters are so resembling to one another that even humans can not tell the differences, so we treated upper case and lower case, {1, l, I} and {O, o, 0} as single classes respectively in the experiment. When recognizing rotated characters, the recognition rate was 55.20% when learning only standard characters, but it improved to 99.34% when rotated characters were learned. The breakdown of mis-recognized characters is as follows. The most of the failure was mis-recognition of ‘g’ as ‘9’ and then mis-recognition of

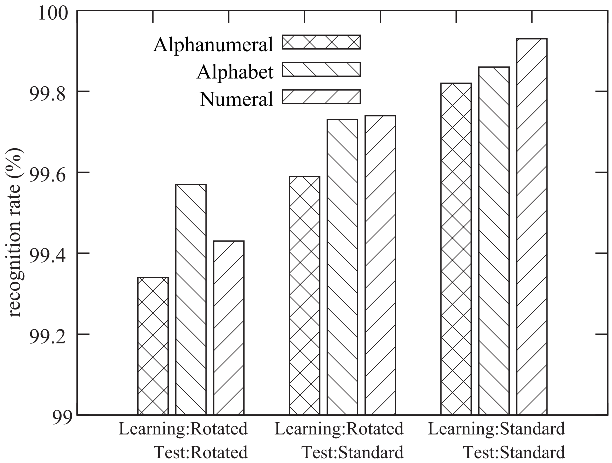

Figure 4. The result of the recognition experiment.

'a' as 'q'. Most of the mis-recognized characters were really resemble to one another that even humans have difficulty to discriminate. When standard characters were given as an input the recognition rate obtained by the method was 99.59%. The rate is only slightly lower than the recognition rate 99.82% obtained when only standard characters were used for both learning and evaluation. We have confirmed that learning rotated characters does not negatively affect significantly when recognizing standard characters. The recognition rates when only alphabet and only numeric character are used both for learning and evaluation are shown together in Figure 4.

IV. Recognition experiment using camera–captured images

We developed the system that automatically detect and recognize the characters in image of scene taken by camera function of cellular phones or digital cameras, and performed the recognition experiment using the rotation–free classifier we described in Section III. We prepared 50 camera-captured images in which there are 437 alphanumeric characters excluding those images with characters that are rotated exceeding the ranges of rotation in learning process. Some examples of the images are shown in Figure 5.

Character segmentation module consists of connected component (CC) detection and character selection. We do not try to detect separating characters 'i' and 'j' in this stage since those characters will be segmented properly after rotation angle estimation and correction more easily.

A. Character segmentation

In the character segmentation module, we extract CCs that are possible characters from the images like Figure 5. The flow of character segmentation is shown in Figure 6.

Figure 5. Some examples of alphanumeric character images which taken by camera function of cellular phones or digital cameras.

The input image is binarized by local threshold method, and the CC analysis and noise removal are performed to the binary image. Then enclosing rectangular of the CC is detected and the corresponding area of the original image is extracted. The extracted image is binarized again with the threshold selected by Otsu ' s method. We leave only the largest CC to get the most character likely candidate. We call the CCs detected by the above method as group A. As in the similar flow, after the binariztion of the input image by local threshold method is performed black and white reverse operation. We call the CCs detected with this reverse operation as group B. Although more CCs than existing characters are included in both groups, we can avoid to fail to extract characters due to mis-judging whether the background is black or white. As the result of character segmentation from 50 camera-captured images, 1026 of CCs that include 437 characters were extracted.

B. Character recognition/rejection

Since there are CCs that are not characters we try to reject those CCs and recognize only those that are characters. The flow of character recognition/rejection is shown in Figure 7.

We use an extended reference vectors (with period '.' added to those we describe in section III) to calculate the MQDF to recognize characters. The value of the MQDF represents the negative log-likelihood of character and represents that the smaller the value is the higher the possibility of the character. If the value is greater than a threshold value, the CC is rejected as it is not a character. We adopted the average of minimum and maximum value of the MQDF, i.e. (minimum value + maximum value)/2, as the threshold value. For example the threshold value is -688.5 for the example in Figure 7.

The reason of adding period to the category set is to deal with small noise components. After the process of rejection, the CC that is recognized as period is also rejected. Hence the dot of the separating characters is also rejected in this stage. Finally, the number of remaining CCs in group A and B are count up and the components in the group that has

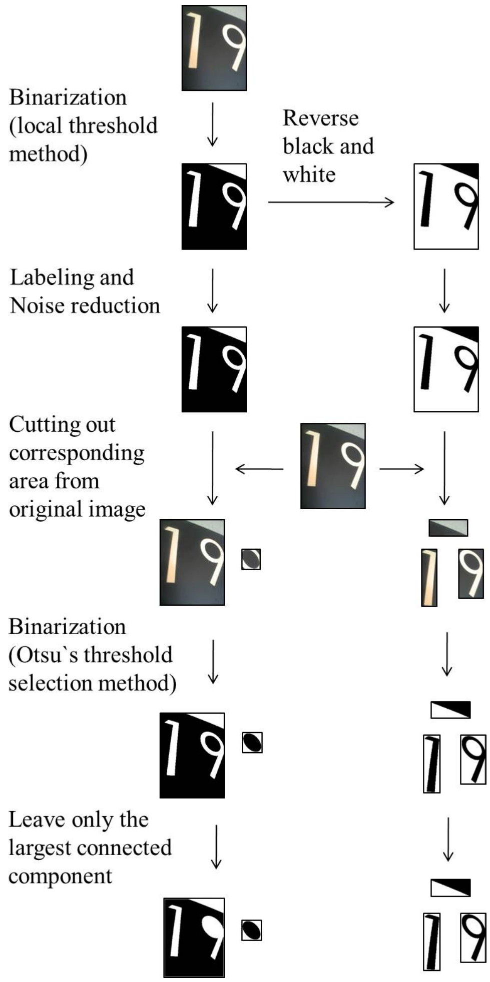

Figure 6. The flow of cutting out characters.

more CCs are selected as final characters.

In the recognition experiment 437 of CCs that should be recognized as a character were left and correctly recognized, while 51 CCs that are not characters were detected as characters. Thus we obtained 100% recall with 89.5% precision. All characters were correctly recognized and the average number of false detection was suppressed to 1.02 (51/50) per image.

As a result we have confirmed that when recognizing individual characters, considering 3D rotation can approximately handle the recognition of perspectively distorted characters. While we excluded separating characters from category set for recognition in the above experiment, we have done the recognition experiment using 11 separating characters manually segmented and obtained 100% recognition rate for those characters.

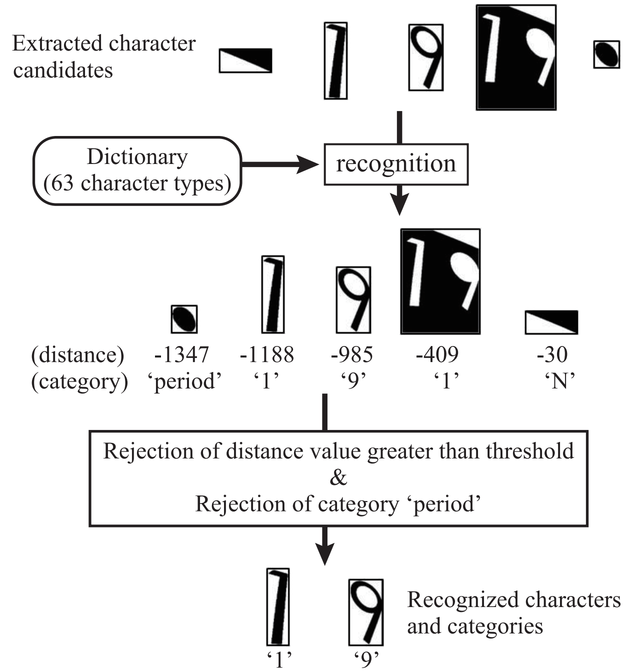

Figure 7. The flow of recognition/rejection character.

V. Application to scene analysis

A. Rotation angle estimation of camera-captured charcters

In this section we describe on three dimensional rotation angle estimation of characters for detecting local normal of the surface on which the characters are printed aiming to scene analysis by shape from characters.

B. Rotation angle estimation method and estimation accuracy

The flow of estimating rotation angle is as follows. After the three dimensional rotation-fee character recognition, the rotation angle is estimated by finding the angle which minimizes the MQDF for the known class of character samples with every rotation angle.

In the experiment we used the same printed alphanumeric characters rotated in the same way as in section III. Therefore 245 ($7\times7\times5$) sets of reference vectors are generated for each class of character. Used feature vector, 392 dimensional WDH, is also the same as in the preceding experiments. To suppress the increase of the processing time according to the increase of the reference vector sets, we used Euclid distance and a linear discriminant function besides the MQDF.

Correct detection rate of rotation angle was 58.01%, 78.90% and 84.82% for Euclid distance, the linear discriminant function and the MQDF respectively. Comparatively high detection rate was obtained by the linear discriminant function and the MQDF. Since the rotation of only around

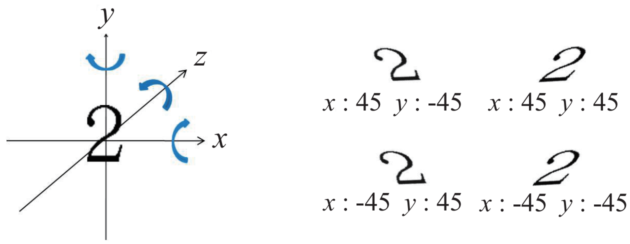

Figure 8. Characters with counter rotation around x-axis and y-axis by 45 degree.

Figure 9. The result of local normal detection for camera-captured scene image.

x-axis (y-axis) just shrinks in height (width) as shown in Figure 2 the accurate detection of these rotation is difficult. Also, the counter rotation around x-axis and y-axis results in the same images as shown in Figure 8. Therefore, we regarded those rotations to be equivalent in calculating the detection rate.

C. Rotation angle estimation of characters in real scene images

We performed the experiment to detect local normal of the surface on which the characters are printed. The used data is camera-captured alphanumeric characters in real scene.

The ambiguity of the sign of rotation described in B. was removed manually in this experiment.

Figure 9 shows the result of normal vector detection. We can see that three dimensional rotation angle detection of characters in scene can be applied to local normal detection.

Although the accuracy of the normal detection is not high so far because we estimated rotation angle by intervals of 15 degree, it will be improved by refinement of the rotation angle and/or parametric interpolation as well as averaging normals of multiple characters. To improve both the accuracy and the processing time, hierarchical and iterative approach with progressive angle refinement will be promising.

VI. Summary

In this paper, we proposed a new method of three dimensional rotation-free recognition of characters in scene. We employed the MQDF classifier trained with samples generated by 3D rotation process in a computer. The results of the evaluation experiments using printed alphanumeric characters as an evaluation data set, consisting of approximately 600 samples/class for 62 character classes, showed that the recognition rate is 99.34% for rotated characters while it is 99.59% for non rotated characters. We have experimentally confirmed that the rotated characters given as the training data set do not negatively affect significantly to recognition of non rotated characters. Moreover, 437 characters extracted from 50 camera-captured scenes were correctly recognized and confirmed the feasibility of real world application of our method.

Finally we studied on the possibility of three dimensional rotation angle estimation of characters for detecting local normal of the surface on which the characters are printed aiming to scene analysis by shape from characters. It was shown experimentally that three dimensional rotation angle detection of characters in scene can be applied to local normal detection.

Following issues are remaining as future research topics:

1) There are some similar methods that recognize characters by absorbing variation of characters. Comparative experiments of those methods and our method will be conducted.
2) The ambiguity of the sign of rotation will be removed using global constraints on scene.
3) The accuracy of the normal detection will be improved by refinement of the rotation angle and/or parametric interpolation as well as averaging normals of multiple characters. To improve both the accuracy and the processing time, hierarchical and iterative approach with progressive angle refinement will be promising.

References

[1] J. Liang, D. Doermann, H. Li, *Camera–based analysis of text and documents: a survey*, IJDAR, vol.7, no.2, pp.84–104, 2005.

[2] G.K. Myers, R.C. Bolles, Q.-T. Luong, J.A. Herson, and H.B. Aradhye, “Rectification and recognition of text in 3-d scenes”, IJDAR, vol.7, no.2-3, pp.147—158, 2005.

[3] M. Iwamura, T. Tsuji, K, Kise, “Memory-Based Recognition of Camera-Captured Characters”, Proc. DAS2010, pp.89-96, 2010.

[4] H. Murase, S.K. Nayar, “3D Object Recognition from Appearance - Parametric Eigenspace Method-”, Trans. of IEICE, Vol.J77-D-II, No.11, pp.2179-2187, 1994.

[5] H. Hase, T. Shinokawa, M. Yoneda, C.Y. Suen, “Recognition of Rotated Characters by Eigen-space”, Proc. ICDAR2003, pp.731-735, 2003.

[6] F. Kimura, T. Wakabayashi, S. Tsuruoka and Y. Miyake, Improvement of Handwritten Japanese Character Recognition Using Weighted Direction Code Histogram, Pattern Recognition, Vol.30, No.8, pp.1329-1337, 1997.

[7] F. Kimura, K. Takashina, S. Tsuruoka and Y. Miyake, Modified quadratic discriminant functions and the application to Chinese character recognition, IEEE Trans. Patter. Anal. Mach. Intell. PAMI-9(1), 149-153, 1987.