Efficient Three Dimensional Rotation Estimation for Camera-based OCR

Kanta KURAMOTO, Wataru OHYAMA,
Tetsushi WAKABAYASHI and Fumitaka KIMURA
Graduate School of Engineering, Mie University
Kurimamachiya-cho, Tsu-shi, Mie Japan
{kanta, ohyama}@hi.info.mie-u.ac.jp

Abstract

Camera-Based Optical Character Recognition (CBOCR) has attracted interests of many researchers in both computer vision and document analysis research fields. A significant challenge in CBOCR is how we handle characters of those appearances are affected by three-dimensional (3D) rotation due to locational relationship between a printing plane and camera. Proper handling of these 3D rotated characters is expected to improve the performance of both detection and recognition of camera-captured characters. In this paper, we propose an efficient implementation of 3D rotation estimation for camera-captured characters. The proposed implementation requires small memory load and short computational time. We employ Linear Discriminant Function (LDF) instead of Modified Quadratic Discriminant Function (MQDF) for further memory reduction. The results of experimental evaluation using a large-scale alphanumeric character dataset showed that small number of dimensionality of original feature vector is sufficient for keeping accuracy of 3D rotation estimation and total amount of memory required for 3D rotation estimation is reduced from 141.0 MB to 6.6 MB.

1 Introduction

Rapid improvement of performance and availability of consumer-level digital imaging devices made significant demands where these devices were used for capturing documents and characters to be coded in natural scenes. These demands have attracted interests of researchers in both computer vision and document analysis research fields to establish Camera-Based Document Analysis and Recognition (CBDAR) or Camera-Based Optical Character Recognition (CBOCR) techniques. In the last two decades, researchers have continued expanding traditional scanner-based document analysis and recognition techniques to camera-based situation and developing new techniques based on observation from machine vision researches [1, 2].

CBDAR and CBOCR suffer from difficulties which are not involved in scanner-based scenarios. The first difficulty appears in recognition stage. Distorted appearance of characters due to perspective projection and three dimensional (3D) rotation easily corrupts recognition performance. A memory-based recognition technique[3] and a rectification-based approach[4] have been proposed to overcome this distortion problem. The second difficulty is that detection of character is affected by complex natural scenes. As seen in the result of ICDAR 2013 Robust Reading Competition[5], text detection accuracy does not exceed 80%. For text detection in natural scene image, an edge-based method has been proposed in [6]. Other approach using scene context for character detection in natural scene image has been proposed in [7]. This method tried to use information other than character shape to improve character detection performance.

A Connected-Component (CC)-based approach for text detection was proposed by Kuramoto[8]. This method discriminates extracted each CC into character or non-character using estimated 3D rotation angle of recognized character. This method achieved accurate character/non-character discrimination and character recognition and it is confirmed that rotation angle of characters is useful information for character detection.

A major drawback of the method in [8] is that the method requires a large amount of memory for estimating character rotation angle. In this paper, we proposed an efficient implementation of 3D rotation angle estimation which keeps estimation accuracy using low dimensional feature vectors and the small number of eigenvectors in classifiers. Moreover, we reduced required memory amount by employing Linear Discriminant Function (LDF) for rotation angle estimation.

2 Rotation-free character recognition and 3D rotation angle estimation

The proposed efficient 3D rotation angle estimation of camera-captured character is based on the CBOCR method proposed by Narita [9] and Kuramoto [8]. In this section, we review their CBOCR and rotation estimation methods.

Fig. 1 and Fig. 2 show the process flow of training and recognition stages in their method, respectively.

In the training stage, the method translates the center of the enclosing rectangular of a non-rotated character to the origin of 3D left handed Cartesian coordinate system and generates training samples of 3D rotated characters by virtual (on memory) 3D rotation process. After generating 3D rotated characters, feature vectors are extracted from these rotated characters and mean vectors, eigenvalues and eigenvectors of covariance matrices are calculated to construct a classification dictionary which consists of learning model for each character class and also for each rotation angle class.

The Gray-scale Gradient Feature[10] is employed to describe appearance of input character image. The gradient feature is extracted by the following procedure:

1. As pre-processing, the position and size of an input image is normalized.

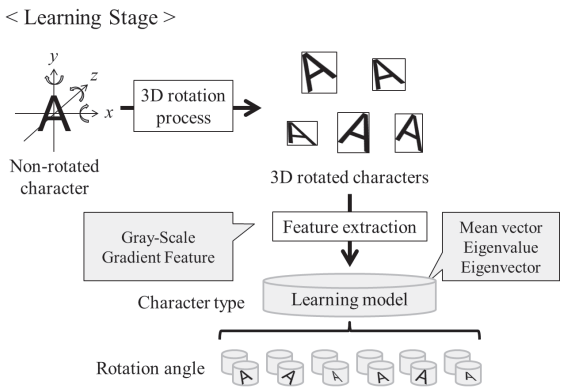

Figure 1. The process flow of Kuramoto's CBOCR method (training stage), cited from [8].

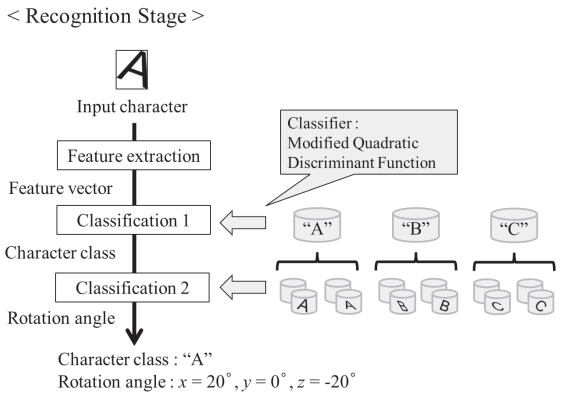

Figure 2. The process flow of Kuramoto's CBOCR method (recognition and rotation estimation stages), cited from [8].

2. The input binary image is converted into a gray-scale image by repeatedly applying a 2×2 mean filter 4 times.

3. The gray-scale image is normalized so that the mean gray scale becomes zero with the standard deviation value equals 1.

4. The direction and strength of gray-scale gradient on each pixel is obtained by applying a Roberts filter to the normalized image.

5. The obtained direction of the gradient is quantized in 32 orientations with $\pi/16$ intervals.

6. The normalized character image is divided into $(2n-1)^2$ ($(2n-1)$ horizontal $\times(2n-1)$vertical) blocks. The strength of the gradient in each direction is accumulated in each block to produce $(2n-1)^2$ local orientation histograms.

7. The directional resolution is reduced from 32 to 8 by down sampling with weight vectors [1 4 6 4 1] and [1 2 1]. Also, the spatial resolution is reduced from $(2n-1)\times(2n-1)$ to $n\times n$ by down sampling every two horizontal and every two vertical blocks with 5×5 Gaussian filter, to produce a feature vector of which dimensionality is $8n^2$ (n horizontal, n vertical, and 8 directional resolution).

The 5×5 Gaussian filter, the weight vector [1 4 6 4 1] and [1 2 1] in the step (7) are the high-cut filters to reduce the aliasing due to the down sampling. Their size was empirically determined for this purpose.

In the recognition and estimation stage, we extract a feature vector of the input character by the same procedure in the learning stage. Then the method classifies the extracted feature vector using learning model of each character class. Modified Quadratic Discriminant Function (MQDF)[11] is employed as classifier. MQDF is defined by

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

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 for character recognition is $\mathrm{O}(kn)$.

After identifying its character class of input image, the method estimates the rotation angle of input image using trained model of that character class to identify its rotation. The original method also used MQDF classifier for this process.

Since this approach does not requires rotation normalization of input character, it can recognize rotated characters in the same computational time as non-rotated characters. Also, since this approach recognizes without word/line segmentation characters individual, there is no restriction of text layout.

3 Efficient implementation of 3D rotation estimation

3.1 Strategies of the proposed implementation

A major drawback of the method described in 2 is that the method requires large memory load for character rotation angle estimation. The main contribution in this paper is proposing efficient implementation which significantly reduces this memory load while keeping the estimation accuracy.

The dimensionality of feature vectors and the number of eigenvectors in MQDF classifier are critical factors to control the memory requirement while the number of rotation angle classes should not be reduced to keep capability for angle estimation. The method handles estimation of the character rotation angle as a classification where an input character image is classified into one of 847 three-dimensional rotation angle classes. The classifier has one dictionary consisting of reference vector and 20 eigenvectors for each angle class. When the method employes feature vectors of 392 dimensionality, the angle estimation requires 1.7GB memory capacity for all character classes. This

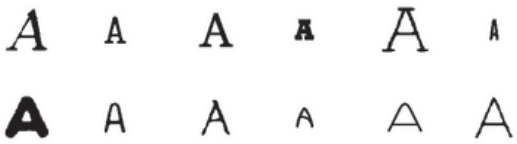

Figure 3. Examples of alphabet 'A' used as data set.

Table 1. Base-line performance of 3D rotation estimation for various fonts: the numbers represent estimation error in degree.

	Helvetica	Century	All-font
x,y	12.25	10.40	11.66
z	2.70	2.49	2.75

requirement of memory capacity is quite larger than that of 9.6MB for character class recognition.

The strategies for memory reduction in the proposed implementation consists of main three approaches:

1. dimension reduction of original feature vectors,
2. reduction of the number of eigenvectors in classifiers,
3. replacing MQDF classifiers by LDF.

In this section, we describe each approach and evaluate its effectiveness by experiments with a large-scale dataset.

We used multi-font character data set exampled in Fig. 3 consisting of approximately 600 samples per one character class. We divided this data set into training and evaluation sets. In the training set, the rotation angle of characters ranges -50 degree to 50 degree around x-axis and y-axis, -30 degree to 30 degree around z-axis in order to avoid similarity problem e.g. 'N' and 'Z'. Concequently, 847 ($11\times11\times7$) samples are generated for training for each character image. Rotation angle of characters in evaluation set ranges -42 degree to 42 degree around x- and y-axes, -30 to 30 degrees around z-axis with the angle interval of 6 degree. 2475 ($15 \times 15 \times 11$) samples are generated as an evaluation data set for each character image.

For evaluation measures, we employ two metrics, i.e. arccosine of inner product between the character normal detected by the method and the actual character normal (rotation angle around x and y-axes) and difference between rotation angles around z-axis.

3.2 Base-line performance for 3D rotation estimation

The base-line performances of 3D rotation estimation for Helvetica font, Century font and All fonts are shown by Table 1. The base-lines are obtained by the condition where 392 dimensional gradient feature vector ($n = 7$) and 20 eigenvectors of covariance matrix for calculation of MQDF classifier. From these results, we observed that (1) estimation performance for rotation around z-axis is higher than that around x,y-axis and (2) the number of fonts contained in evaluation set does not affect estimation accuracy.

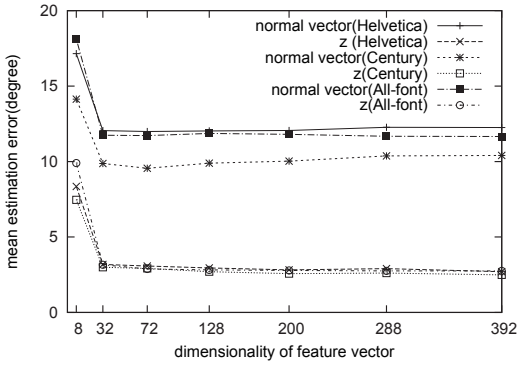

Figure 4. Mean estimation error on angle estimation vs dimensionality of feature vector.

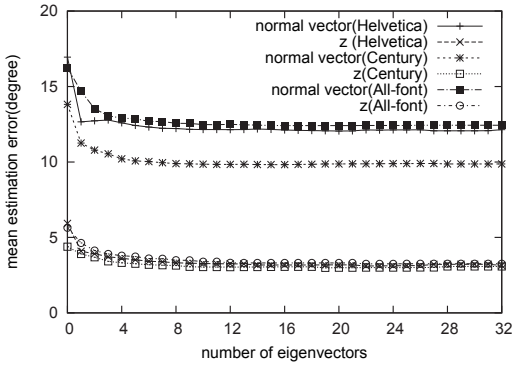

Figure 5. Mean estimation error on angle estimation vs the number of eigenvectors.

3.3 Dimension reduction of original feature vector

Here, we reduced the dimensionality of original feature vector by changing the number of sub-blocks and evaluate the effect of dimension reduction against estimation accuracy. Since the class of input character image is identified in the recognition stage ("Classification 1" in Fig.2), the dimensionality of 392 of original feature vector seems to be larger than necessity for rotation angle estimation. Particularly, the number of sub-blocking (7×7) is considered to be reduced.

The results are shown in Fig. 4. Mean estimation error is stable against dimensionality of feature vectors, except dimensionality of 8 ($n = 1$). These results represent that small size of trained model, which is constructed with small original feature vector, is able to keep estimation accuracy for 3D rotation of characters.

3.4 Reduction of eigenvectors in classifiers

Here, we reduced the number of eigenvectors in MQDF classifiers and evaluated the effects of them against estimation performance. In the calculation of MQDF classifiers, the original estimation method used 20 eigenvectors when dimensionality of original feature vectors are 392. We confirmed that the original dimensionality could be reduced in the above, so, we also expected that the number of eigenvectors in classifiers could be reduced.

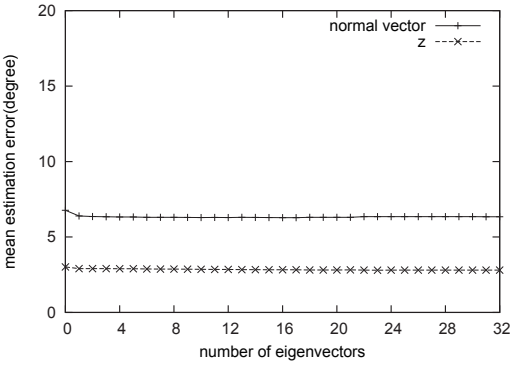

Figure 6. Mean estimation error on angle estimation vs the number of eigenvectors (in the case where the training and the test data consist of only Helvetica font).

The result is shown in Fig. 5. Each of the mean estimation error were almost the same except that the number of eigenvectors is zero, which is equivalent to using Euclidean distance as classifier. This result shows that small number of eigenvectors are sufficient to estimate rotation angle of multi-font characters.

In order to check the limitation of eigenvector reduction, we have done the experiment for performance evaluation with only Helvetica font for both training and test data sets. The result is shown in Fig. 6. Each of the mean estimation error is not affected by the number of eigenvectors. From this result, we considered that eigenvectors in MQDF classifiers contribute for multi-font rotation angle estimation.

3.5 Replacing classifier

We assume that the distribution of sample due to difference in font shape is the same in all classes in the feature space, and we replace MQDF by LDF. LDF is defined by

$$\begin{aligned} g_j(X) &= W_j^T X + w_{0j}, \qquad (2) \\ W_j^T &= -2M_j^T \Sigma^{-1}, \\ w_{0j} &= M_j^T \Sigma^{-1} M_j, \end{aligned}$$

where, X and M_j are feature vector and the mean vector of class j respectively, and Σ is covariance matrix. If the distributions of samples in every classes are Gaussian and equal to each other, LDF guarantees to be optimum discriminant function. The LDF brings significant reduction of memory and computational time against MQDF.

The result is shown in Table 2. The estimation accuracy of LDF is slightly lower than that of MQDF, but required memory is significantly reduced. In addition, the estimation accuracy of LDF is higher than that of MQDF with reduced the number of eigenvectors.

4 Conclusion

In this paper, we proposed an effective implementation for 3D rotation angle estimation in CBOCR. The results of experiments confirmed that our three strategies; i.e. (1) reduction of original feature dimensionality, (2) reduction of eigenvectors in classifiers, and (3) replacing classifiers from MQDF to LDF, could reduce required memory from 1.7 GB to 6.6MB with keeping estimation performance from the base-line.

Table 2. Average estimation error and memory capacity for various classfiers.

	MQDF ($k = 20$)	MQDF ($k = 1$)	LDF
x,y (degree)	12.25	14.68	13.92
z (degree)	2.70	4.68	3.89
memory capacity (MB)	141.0	19.2	6.6

Future work involves to further improvement of x,y-axis rotation angle estimation by introducing interpolation with classification scores.

References

[1] J.Liang, D.Doermann, H.Li: "Camera-based analysis of text and documents: a survey," International Journal on Document Analysis and Recognition, Vol.7 Issue 2, pp.84–104, 2005

[2] Q.Ye, D.Doermann,:"Text Detection and Recognition in Imagery: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.99, No.PrePrints, p.1, 2015

[3] M.Iwamura, T.Tsuji, K.Kise: "Memory-Based Recognition of Camera-Captured Characters," Proc. 9th IAPR International Workshop on Document Analysis Systems, pp.89–96, 2010

[4] G.K.Myers, R.C.Bolles, Q.T.Luong, J.A.Herson, H.B. Aradhye: "Rectification and recognition of text in 3-d scenes," International Journal on Document Analysis and Recognition, Vol.7 Issue 2–3, pp.147–158, 2005

[5] D.Karatzas, F.Shafait, S.Uchida, M.Iwamura, L.Gomez, S.Robles, J.Mas, D.Fernandez, J.Almazan, L.P. de las Heras: "ICDAR 2013 Robust Reading Competition", Proc. 12th International Conference of Document Analysis and Recognition, pp.1115–1124, 2013

[6] B.Epshtein, E.Ofek, Y.Wexler: "Detecting text in natural scenes with stroke width transform," Proc. CVPR2010, pp.2963–2970, 2010

[7] Y.Kunishige, F.Yaokai, S.Uchida: "Scenery character detection with environmental context," Proc. 11th International Conference of Document Analysis and Recognition, pp.1049–1053, 2011

[8] K.Kuramoto, W.Ohyama, T.Wakabayashi, F.Kimura: "Accuracy Improvement of Viewpoint-Free Scene Character Recognition by Rotation Angle Estimation," Proc. 5th International Workshop on Camera-Based Document Analysis and Recognition, pp. 60–70, 2014

[9] R.Narita, W.Ohyama, T.Wakabayashi, F.Kimura: "A study on three dimensional rotation-free character recognition and rotation angle estimation of characters," Proc. 21st International Conference on Pattern Recognition, pp.677–680, 2012

[10] T.Wakabayashi, S.Tsuruoka, F.Kimura, Y.Miyake: "Increasing the feature size in handwritten numeral recognition to improve accuracy," Systems and Computers in Japan, Vol.26, No.8, pp.35–44, 1995

[11] F.Kimura, K.Takashina, S.Tsuruoka, Y.Miyake: "Modified quadratic discriminant functions and the application to Chinese character recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.9, No.1, pp.149–153, 1987