Layout-Free Dewarping of Planar Document Images

[†]Masakazu Iwamura, [†]Ryo Niwa, [†]Akira Horimatsu, [†]Koichi Kise,
[††]Seiichi Uchida, and [†††]Shinichiro Omachi

[†]Graduate School of Engineering, Osaka Prefecture University
1-1 Gakuencho, Naka, Sakai, 599-8531 Japan

[††]Faculty of Information Science and Electrical Engineering, Kyushu University
744 Motooka, Nishi, Fukuoka, 819-0395 Japan

[†††]Graduate School of Engineering, Tohoku University
6-6-05 Aoba, Aramaki, Aoba, Sendai, 980-8579 Japan

Abstract

For user convenience, processing of document images captured by a digital camera has been attracted much attention. However, most existing processing methods require an upright image such like captured by a scanner. Therefore, we have to cancel perspective distortion of a camera-captured image before processing. Although there are rectification methods of the distortion, most of them work under certain assumptions on the layout; the borders of a document are available, textlines are in parallel, a stereo camera or a video image is required and so on. In this paper, we propose a layout-free rectification method which requires none of the above assumptions. We confirm the effectiveness of the proposed method by experiments.

1. INTRODUCTION

Camera-based document analysis and recognition[1–4] has been an active area in the field of document analysis and recognition. As compared to ordinary document analysis methods with scanners, camera-based methods employ a digital camera as an input method. An advantage of camera-based methods comes from the portability of digital cameras. Capturing images of documents and characters by cameras is extremely easy and fast as compared to scanners. In addition, cameras enable us to capture large posters and signs while they are on display. Thus camera-based methods offer new ways of applying document analysis and charcter recognition for improving our life.

However, it is not simple to realize camera-based methods due to the large difference between images captured by scanners and those by cameras. In other words, camera-captured images are with much more degradation. Most of available techniques of document analysis and character recognition have been developed under the assumption that images are captured by scanners. Hence if one intends to apply these techniques to camera-captured images, it is nesessary to normalize them to be close enough to images captured by scanners. We are concerned in this paper with rectification of camera-captured images with a special focus on perspective distortion.

Most existing rectification methods work under certain assumptions on the layout. For example, some methods rely on the fact that the borders of a document shape rectangle.[5] Some employ the assumption that textlines are typically laid out in parallel.[5,6] Although there are existing methods without assumptions on layout, they require special devices such as stereo cameras[7] or special data such as video images.[8] Therefore, without such a special device nor data, the existing methods cannot rectify a layout-free image such as shown in Fig. 1.

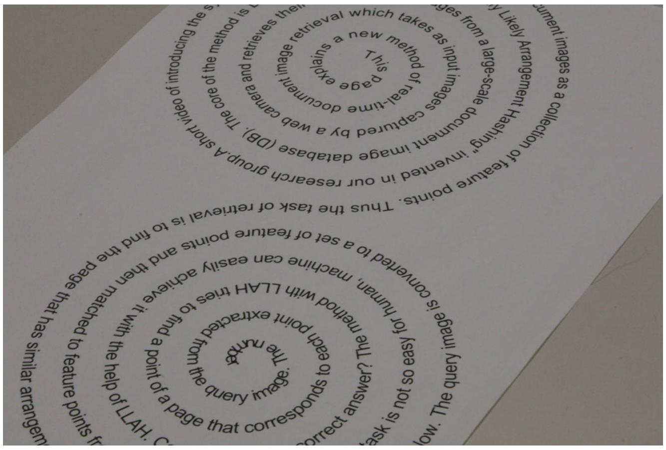


Figure 1. A planar document image difficult to rectify using existing methods.

Although we have proposed a method capable of removing perspective distortion of a layout-free image, [9] the method is limited to obtain an affinely distorted image at best. In this paper, we propose a throughout layout-free dewarping method. The method consists of two steps; the first step obtains an affinely distorted image from the perspectively distorted one, and the second step cancels the affine distortion. The point of the first step is to normalize images based on two values: variants and invariants of perspective transformation. They can be calculated irrelevantly to layout of documents. Due to this property, the method is applicable to documents with various layouts, though most existing methods work under certain assumptions on the layout. The point of the second one is to estimate an affine transformation matrix as a byproduct of rough recognition of affinely distorted characters.

2. OBTAINING AFFINELY DISTORTED IMAGES

The first step of the proposed method is to obtain an affinely distorted image (Fig. 2(b)) from the original image (Fig. 2(a)). The procedure presented in this section is based on Ohta's method. [10]

2.1 Estimation of vanishing line

Imagine a document image suffering from perspective distortion. There usually exists the same characters, say 'a,' printed in the same size. For simplicity, we assume such a special document here. Due to the perspective distortion, the observed sizes of the characters vary by their positions; a character near the camera has a large area S_1 and that far from the camera has a small area S_2. As shown in Fig. 2(c), we can estimate a vanishing point from the two areas S_1 and S_2 with the following equation derived by Ohta et al.[10]

$$\frac{f_2}{f_1} = \frac{S_2^{1/3}}{S_1^{1/3}}, \tag{1}$$

where f_1 and f_2 are the distances of them from the vanishing point.

The next step is to estimate a vanishing line from the vanishing points. The process is described in Sec. 2.3.

Once a vanishing line is obtained, two vanishing points on the vanishing line are arbitrarily selected, as shown in Fig. 2(d). Since two parallel lines intersect at a vanishing point, the quadrilateral formed by the four lines is inherently a parallelogram. Thus, the affinely distorted image as shown in Fig. 2(b) is obtained by applying the linear transformation which transforms the quadrilateral (Fig. 2(d)) into a parallelogram (Fig. 2(e)).

Note that the linear relationship of the areas, shown in Eq. (1), can estimate only one vanishing line. Therefore, it cannot cancel perspective distortion. This also comes from the fact that we cannot distinguish from the appearance between an affinely distorted image of an upright object and an upright image where an affinely distorted object is printed.

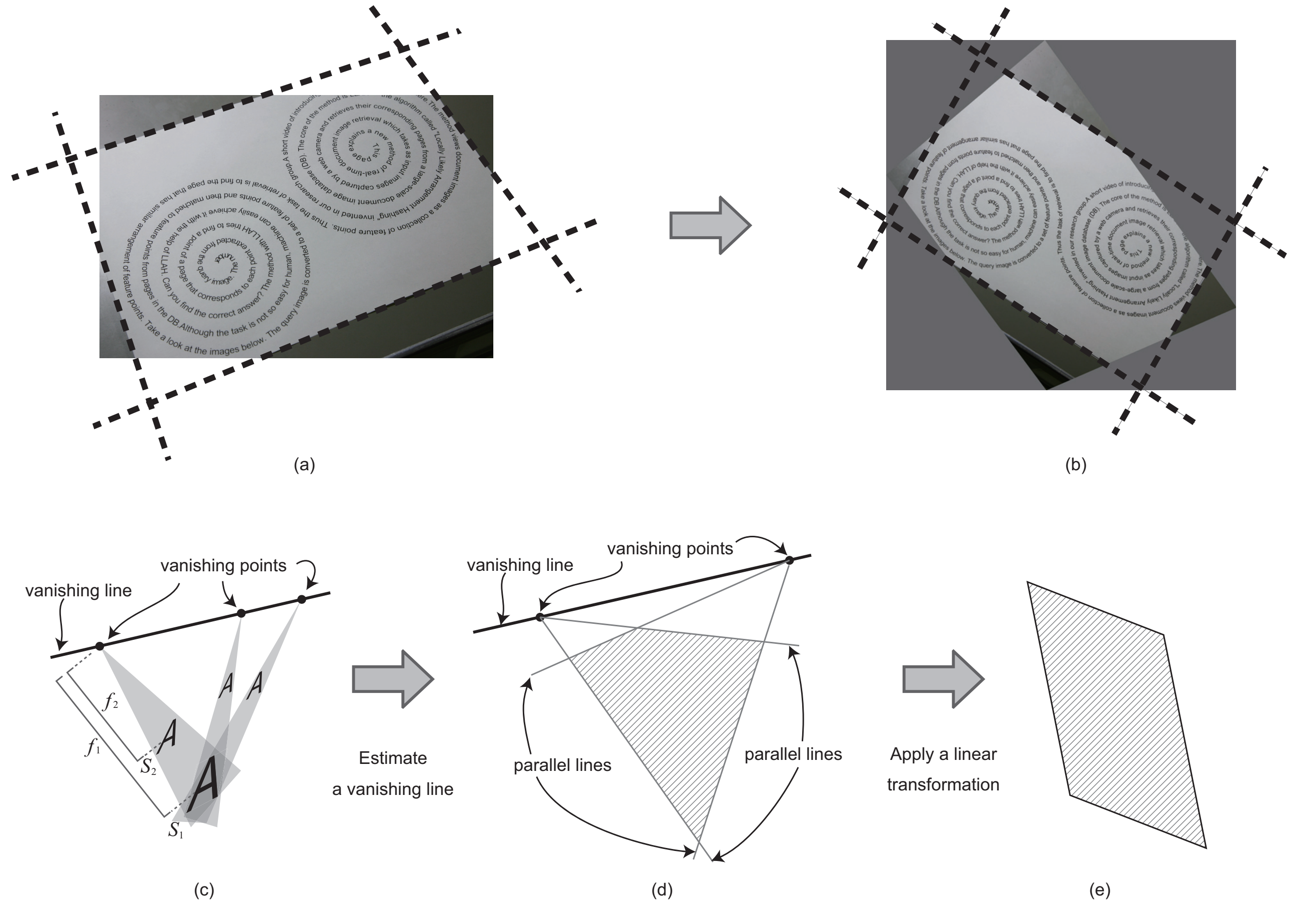


Figure 2. An overview of obtaining an affinely distorted image. (a) The original image. (b) The affinely distorted image. (c) Principal to obtain a vanishing line. (d) Two pairs of parallel lines intersecting at vanishing points. (e) Obtain the affinely distorted image by applying a perspective transformation.

2.2 Clustering using area ratios

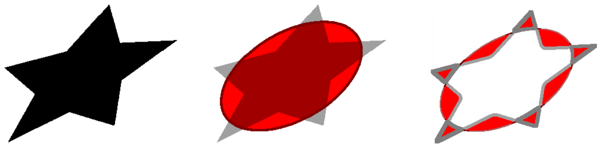
The method of estimating the vanishing line presented in Sec. 2.1 works only if the document contains characters of only one category. However, such a situation is not practical. Thus, we have to distinguish characters into categories in advance.

The most easily conceived method might be character recognition. However, recognizing perspectively distorted characters is not an easy task. In addition, recognition is excessive and classification is enough, because we do not need characters in the same category but characters with the same area for the task. Thus, the characters are classified by the k-means clustering algorithm using affine invariants. Although an affine invariant is not an invariant against perspective distortion, an affine invariant in a small region can be approximated as a projective invariant.

In practice, some clusters should not be used because of a lot of noises. In the experiment of the paper, we selected 20 largest clusters from 50 clusters.

2.3 Robust estimation of vanishing line

After applying the clustering method presented in Sec. 2.2, a reliable vanishing line has to be estimated from the clusters. In order to avoid bad effects of outliers, we estimate a vanishing line candidate for each cluster by the RANSAC algorithm.[11]



(a) Symmetric figure

(b) A-ellipse

(c) Disagreement part with A-ellipse

Figure 3. Regions for calculating the DE-parameter.

The algorithm examines 50 lines each of which is calculated from two randomly selected points. Then, vanishing line candidate which has the largest number of vanishing points within 100 pixel distance is selected.

A candidate of the vanishing line is estimated for each cluster. Then, the most reliable vanishing line is estimated from the vanishing line candidates by robust estimation of mean of directional data.[12] Let $x_1, \ldots, x_n$ be n angles of the vanishing line candidates. The mean of them, say μ, is determined to mimize

$$\sum_{i=1}^{n} \rho(t(x_i^T \mu; \kappa)), \tag{2}$$

where

$$t(u;k) = \text{sign}(u) 2\kappa (1-u)^{1/2} \tag{3}$$

$$\rho(t) = \begin{cases} t^2/2, & |t| \leq c \\ c|t| - c^2/2, & |t| \geq c \end{cases}. \tag{4}$$

We used $\kappa = 20$ and $c = 2$ for the experiment.

3. AFFINE INVARIANTS FOR CLASSIFYING AFFINELY DISTORTED CHARACTERS

In this section, we discuss the classification ability of affine invariants used in Sec. 2.2; DE-parameter proposed by Ohta et al.[10] is introduced and then a set of better affine invariants for classifying affinely distorted characters is proposed.

3.1 DE-parameter

A-ellipse of a given figure shown in Fig. 3 is defined by Ohta et al.[10] The A-ellipse has the same covariance matrix to the figure. For the figure of Fig. 3(a), the A-ellipse is shown in Fig. 3(b). The DE-parameter is the ratio A/B of the areas, where A is the area of the disagreement part between the given figure and the A-ellipse as shown in Fig. 3(c), and B is the area of the given figure.

3.2 Classification of characters by DE-parameter

In the paper of Ohta et al.,[10] five simple figures such as a circle and a triangle were classified by the DE-parameter. However, more complex figures such as alphabets and numerals were not. Thus, we investigate it by an experiment.

We employed 60 characters from numerals and alphabets; 10 figures, 24 lowercase alphabets except 'i' and 'j' which consists of two connected components, and 26 capital alphabets. As shown in Fig. 4, we prepared images whose sizes were

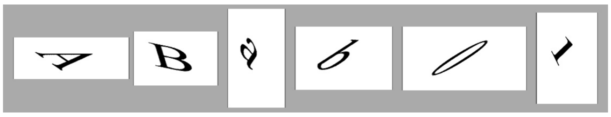


Figure 4. Parts of character images used for the investigation of the classification ability, picked from 1260 affinely distorted images per character.

96×96 pixels and applied various affine transformations. For the sake of that, the affine transformation matrix $\boldsymbol{T} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ was decomposed into

$$\begin{aligned}\boldsymbol{T} &= L(\beta)R(\theta)S(\varphi)A(\alpha) \\ &= \begin{pmatrix} \beta & 0 \\ 0 & \beta \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \\ &\qquad \begin{pmatrix} 1 & \tan\varphi \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha & 0 \\ 0 & 1/\alpha \end{pmatrix},\end{aligned} \tag{5}$$

where

$$\alpha = \pm\sqrt{\frac{a^2+c^2}{ad-bc}}, \tag{6}$$

$$\varphi = \tan^{-1}\frac{ab+cd}{ad-bc}, \tag{7}$$

$$\theta = \cos^{-1}\frac{\pm a}{\sqrt{a^2+c^2}}, \tag{8}$$

$$\beta = \pm\sqrt{ad-bc}. \tag{9}$$

Since β is the scale parameter, we changed the remaining three parameters α, φ and θ in the following ranges: $\alpha = \{1, 2, 3, 4\}$, $\varphi = \{-1.0, -0.9, \cdots, 1.0\}$, and $\theta = \{0.0, 0.1, \cdots, 1.4\}$. Thus by combining them, we applied 1260 affine transformations for each character (1260 comes from $4 \times 21 \times 15$).

The distributions of the DE-parameter are shown in Fig. 5. The figure shows that the DE-parameter is stable to the affine transformations. However, many characters take similar values of the DE-parameter. Thus, it does not have enough ability to classification characters.

3.3 Improvement of the DE-parameter

As described above, a problem of the DE-parameter is its lack of discrimination power for classifying characters. This problem is solved by the TW-value which is an improvement of DE-parameter proposed in.[13] The TW-value is calculated using two ellipses that are similar to the A-ellipse. Some examples are shown in Fig. 7, where m and n in TW(m,n) indicate scale factors of the A-ellipse in percent. The TW-value is calculated as the area ratio defined by:

$$\mathrm{TW}(m,n) = \frac{C}{D}$$

where C is the area of overlap between the given figure (the letter 'A' in Fig. 7) and the region circumscribed by the two ellipses, and D is the area of the given figure, respectively.

In the same manner of the DE-parameter, we investigated if the TW-value has the ability to classify characters. The distributions of the TW-value are shown in Fig. 6. The figure shows that (i) by comparing Fig. 6(a), (b) and (c), the distributions differs by m and n, and (ii) the distributions of the TW-value vary depending on characters. Especially for (ii), characters which have similar values in a TW-value do not have similar values in another TW-value. Therefore, we confirmed that the TW-value has the discrimination ability enough to roughly classify characters.

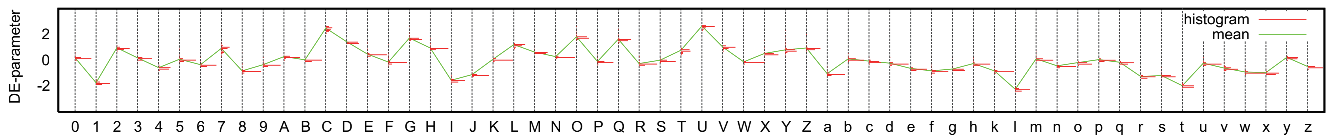


Figure 5. Distribution of DE-parameter. The histogram of the distribution for each character is plotted in red; horizontal axis represents frequency. A green line connects the mean values of the distributions. The distributions are normalized so that the mean is zero and the standard variance is one.

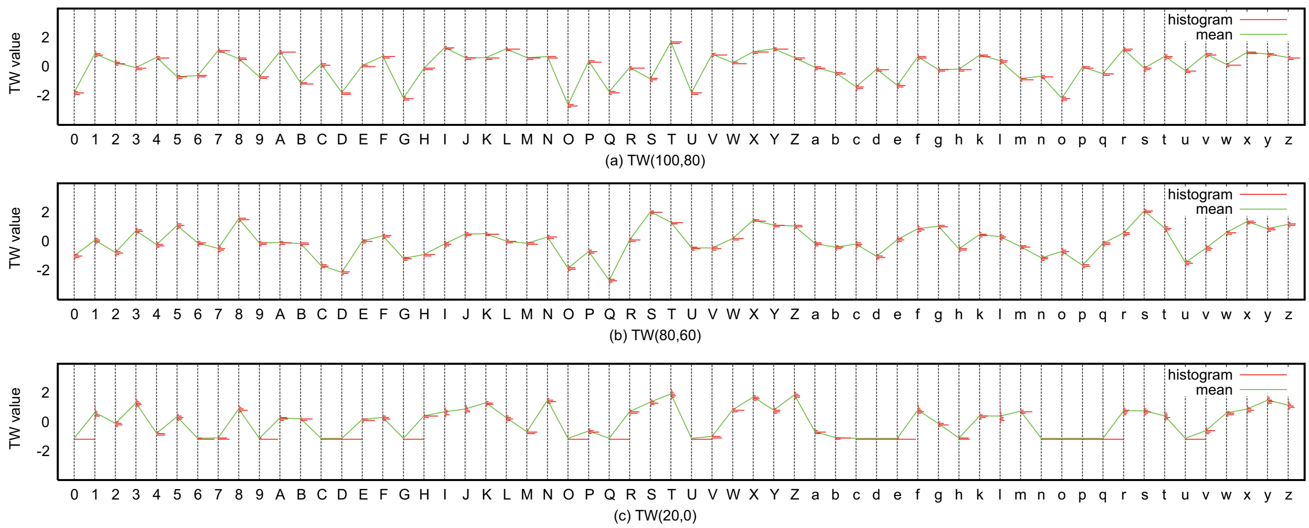
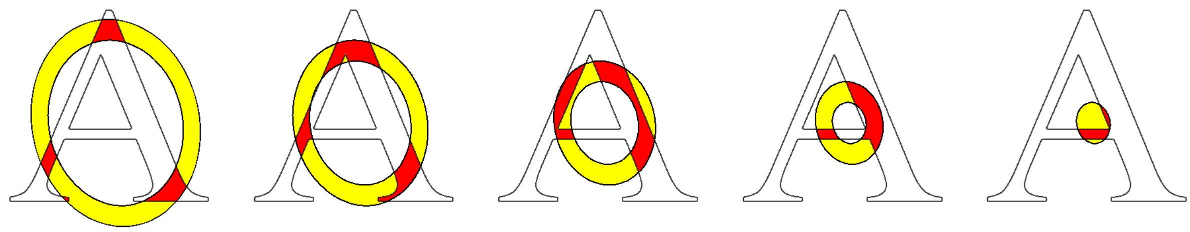


Figure 6. Distributions of TW-values in the same manner as Fig. 5.



(a) TW(100,80) (b) TW(80,60) (c) TW(60,40) (d) TW(40,20) (e) TW(20,0)

Figure 7. Regions for calculating TW-values.

4. OBTAINING UPRIGHT IMAGES USING NORMALIZATION

This section presents a method to estimate an upright image. For the sake of that, we estimate an affine transformation matrix by employing a recognition method for affinely distorted characters. [13] The method is based on the TW-values and an affine normalization as detailed later. One may think that if recognition of affinely distorted characters is possible, we do not need further rectification. However, it is not true. The reasons is that as is well known, the recognition performance of invariants (cf. the TW-values and the Zernike moments) is inherently limited. Thus, it is better to obtain an upright image first and then cancel perspective distortion of the image.

The recognition method for affinely distorted characters [13] is detailed. The principal of estimating the affine transformation matrix is illustrated in Fig. 8. As mentioned above, the method is based on an affine normalization. That is, all the reference

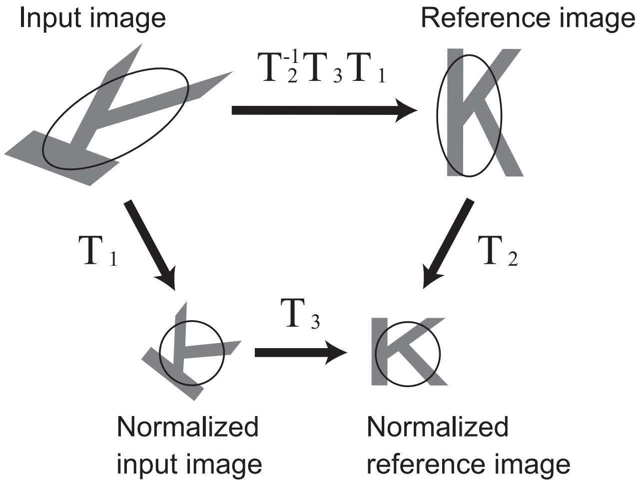


Figure 8. Recognition-based affine normalization. T_1 and T_2 are the affine transformation matrices of the input image and the corresponding reference image for normalization, respectively. T_3 is the rotation matrix between the two normalized images. Finally, the affine transformation matrix from the input image into the reference image given as $T_2^{-1}T_3T_1$ is obtained as a by-product of the recognition result.

character images stored in the system are affinely normalized in advance. The normalization does not rotate images. An input character image is also affinely normalized. Then, the corresponding reference image is found by a matching procedure of rotating the normalized images. As a by-product of the matching procedure, a candidate of the affine transformation matrix (i.e., $\boldsymbol{T}_2^{-1}\boldsymbol{T}_3\boldsymbol{T}_1$ in Fig. 8) is obtained for each character. If all the characters suffer from the same affine transformation, all the candidates of the affine transformation matrix is the same in ideal. However, this restriction is too strong to ensure layout-free. Therefore, the matrix is decomposed into the parameters of the independent scaling α, the shear φ, the skew angle θ and the scaling β, as in Eq.(5). By assuming all the characters have the same independent scaling α and the shear φ, the most feasible values of two parameters are chosen by the k-nearest neighbor density estimator. [14] That is, the point which has the largest distance to the kth nearest neighbor (point) in the 2-dimensional feature space consists of α and φ, is selected.

5. EXPERIMENT

We performed experiments to confirm the effectiveness of the proposed method for two datasets. **Dataset A** contained 40 document images; 10 pages of English journal papers printed by an ink jet printer (EPSON PX-G920) were captured from four different angles by digital cameras (Canon EOS Digital 5D and Fujifilm Finepix F710). **Dataset B** contained five document images; the document image shown in Fig. 9 were printed and captured from five different angles by a digital camera (Canon EOS Digital 5D). The images of the dataset B contained the borders of a document only for evaluation. Experimental results of the proposed method for the datasets are shown in Figs. 10 and 11.

5.1 Restoration of affinely distorted images

We evaluated how much the method presented in Sec. 2 can restore parallel lines with the TW-values. In order to evaluate it, small marks were printed in every 21mm distance on document images in the dataset A. The marks help us evaluate the parallelism. For document images in the dataset B, the borders of the paper were used for evaluation.

The results are shown in Table 1. For all the measures, close to 0 is better. The result shows that the proposed method restored parallel lines for both datasets.

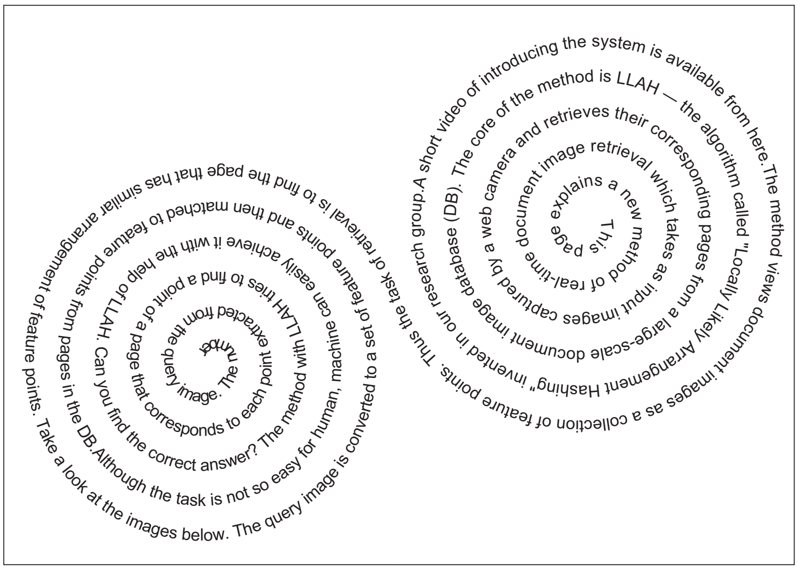


Figure 9. The original image of the spiral text document used as the dataset B.

Table 1. Angles of two parallel lines (degree).

		Dataset A		Dataset B	
Before rectification	Mean	5.9	17.1	11.4	2.5
After rectification	Mean	1.7	2.7	2.2	1.2
	Standard deviation	2.0	1.2	1.4	0.4

5.2 Restoration of upright images

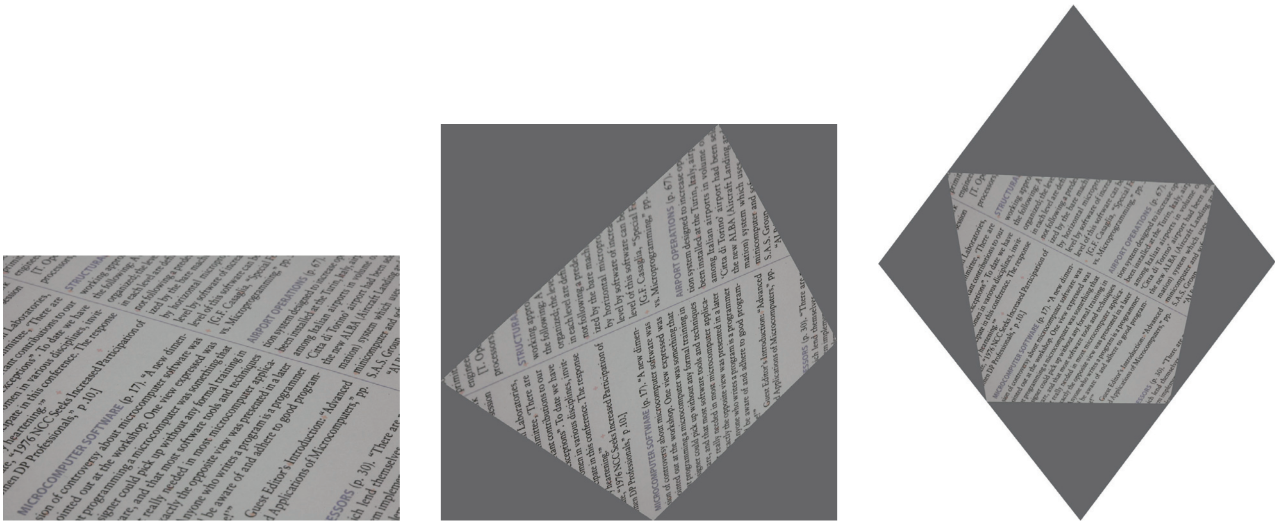
We examined the method of obtaining upright images presented in Sec. 4. For the experiment, $k = 20$ was used for the k-nearest neighbor density estimator. The reference images contained degraded character images of the capital and lowercase alphabets and numerals which were created synthetically by the Gaussian generative learning methods [15] in three different variances. The characters listed in Table 2 were not used for the estimation of the parameters, due to difficulty in character recognition and ambiguity in rotation angles.

We evaluated the results in the same manner in Sec. 5.1. The results are shown in Table 3. "Average largest error" represents the magnitude of residual deformation; the average angle of the most deformed parallel line among parallelograms in a document image was calculated. "Aspect ratio error" means the aspect ratio of restored document image to the original one. For both measures, close to 0 is better. The table shows that the proposed method could restore both typical document such like textlines are laid out in parallel and layout-free document images.

6. CONCLUSION

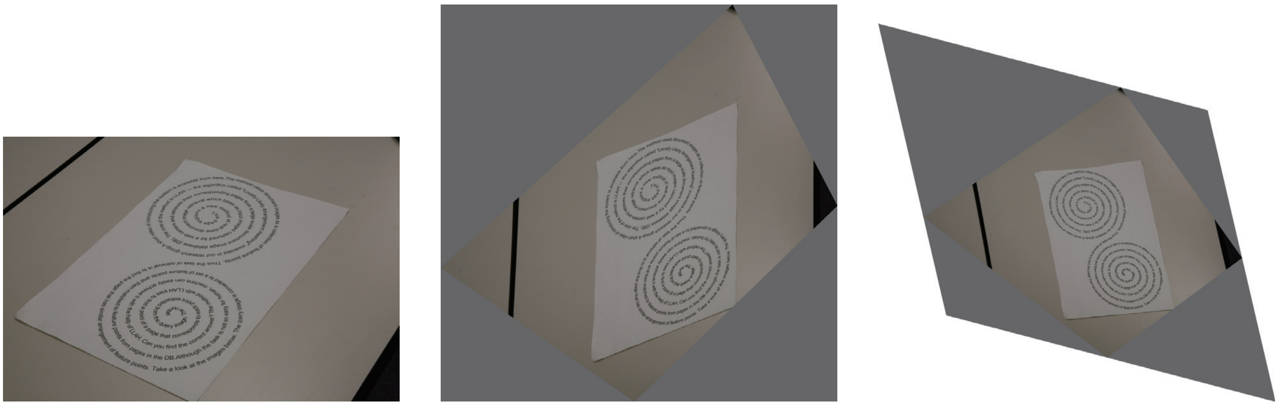
The main contributions of the paper is to propose a throughout layout-free dewarping method. That is, the proposed method does not use the assumptions on the layout where most existing methods use; the borders of a document are available, textlines are in parallel, a stereo camera or a video image is required and so on.

The proposed method can be applicable to the spiral text document shown in Fig. 1. One may think that such a document never exists. However, the proposed method is applicable not only for the spiral text document but also the usual layouts. Therefore, the method is widely applicable.



(a) Perspectively distorted image. (b) Affinely distorted image. (c) Upright image.

Figure 10. An experimental result of the proposed method for dataset A.



(a) Perspectively distorted image. (b) Affinely distorted image. (c) Upright image.

Figure 11. An experimental result of the proposed method for dataset B.

Table 2. The characters not used for the estimation of the feasible values of the parameters α and φ, due to their difficulty in character recognition and ambiguity in rotation angles.

2	3	5	7	A	D	E	F	G	K
M	P	R	T	V	W	Y	a	c	e
f	g	h	k	m	r	t	v	w	y

Table 3. Result of restoring upright images (degree).

			Dataset A	Dataset B
Average largest error	Before rectification	Mean	32.7	34.6
	After rectification	Mean	5.4	9.7
		Standard deviation	4.7	6.3
Aspect ratio error	After rectification	Mean	0.24	0.55
		Standard deviation	0.12	0.48

The restriction of the proposed method is that the proposed method requires to store the character images in the fonts which often appear in document images. However, we expect only several fonts are enough to work fine for most documents. The future work includes the investigation of that.

REFERENCES

[1] [*Proc. First Int'l Workshop on Camera-Based Document Analysis and Recognition (CBDAR2005)*] (2005). http://www.m.cs.osakafu-u.ac.jp/cbdar2005/.

[2] [*Proc. Second Int'l Workshop on Camera-Based Document Analysis and Recognition (CBDAR2007)*] (2007). http://www.m.cs.osakafu-u.ac.jp/cbdar2007/.

[3] Doermann, D., Liang, J., and Li, H., "Progress in camera-based document image analysis," in [*Proc. ICDAR2003*], 606–616 (2003).

[4] Liang, J., Doermann, D., and Li, H., "Camera-based analysis of text and documents: a survey," *Int'l Journal of Document Analysis and Recognition (IJDAR)* **7**, 84–104 (2005).

[5] Clark, P. and Mirmehdi, M., "Recognising text in real scenes," *Int'l Journal of Document Analysis and Recognition (IJDAR)* **4**, 243–257 (2002).

[6] Pilu, M., "Extraction of illusory linear clues in perspectively skewed documents," in [*Proc. Computer Vision and Pattern Recognition, 2001 (CVPR '01)*], **1**, 363–368 (2001).

[7] Lampert, C. H., Braun, T., Ulges, A., Keysers, D., and Breuel, T. M., "Oblivious document capture and real-time retrieval," in [*Proc. CBDAR2005*], 79–86 (Aug. 2005).

[8] Sato, T., Ikeda, S., Kanbara, M., Iketani, A., Nakajima, N., Yokoya, N., and Yamada, K., "High-resolution video mosaicing for documents and photos by estimating camera motion," in [*Proc. SPIE Electronic Imaging*], **5299**, 246–253 (Jan. 2004).

[9] Iwamura, M., Niwa, R., Kise, K., Uchida, S., and Omachi, S., "Rectifying perspective distortion into affine distortion using variants and invariants," in [*Proc. Second Int'l Workshop on Camera-Based Document Analysis and Recognition (CBDAR2007)*], 138–145 (Sept. 2007).

[10] Ohta, Y., Maenobu, K., and Sakai, T., "Obtaining surface orientation from texels under perspective projection," in [*Proc. of 7th Int'l Conference on Artificial Intelligence*], 746–751 (1981).

[11] Fischler, M. A. and Bolles, R. C., "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM* **24**(6), 381–395 (1981).

[12] Mardia, K. V. and Jupp, P. E., [*Directional Statistics*], John Wiley & Sons, Inc. (1991).

[13] Horimatsu, A., Niwa, R., Iwamura, M., Kise, K., Uchida, S., and Omachi, S., "Affine invariant recognition of characters by progressive pruning," in [*Proc. the eighth IAPR Workshop on Document Analysis Systems (DAS2008)*], 237–244 (Sept. 2008).

[14] Fukunaga, K., [*Introduction to statistical pattern recognition*], Academic Press, 2nd ed. (1990).

[15] Ishida, H., Yanadume, S., Takahashi, T., Ide, I., Mekada, Y., and Murase, H., "Recognition of low-resolution characters by a generative learning method," in [*Proc. First Int'l Workshop on Camera-Based Document Analysis and Recognition (CBDAR2005)*], 45–51 (2005).