PAPER

Skew Estimation by Parts

Soma SHIRAISHI[†a)], *Student Member*, Yaokai FENG[†], *Nonmember, and* Seiichi UCHIDA[†], *Senior Member*

SUMMARY   This paper proposes a new part-based approach for skew estimation of document images. The proposed method first estimates skew angles on rather small areas, which are the local parts of characters, and subsequently determines the global skew angle by aggregating those local estimations. A local skew estimation on a part of a skewed character is performed by finding an identical part from prepared upright character images and calculating the angular difference. Specifically, a keypoint detector (e.g. SURF) is used to determine the local parts of characters, and once the parts are described as feature vectors, a nearest neighbor search is conducted in the instance database to identify the parts. Finally, a local skew estimation is acquired by calculating the difference of the dominant angles of brightness gradient of the parts. After the local skew estimation, the global skew angle is estimated by the majority voting of those local estimations, disregarding some noisy estimations. Our experiments have shown that the proposed method is more robust to short and sparse text lines and non-text backgrounds in document images compared to conventional methods.

key words: skew estimation, part based, keypoint detector

1. Introduction

Skew estimation is an important preprocessing step when digitizing scanned document images [1]. Document skew occurs when a document is not set properly on a scanner, and it can be a problem for the following processes such as OCR and document analysis. Many skew estimation methods have been proposed [2], such as the projection profile method [3], the Hough transform method [4], the nearest neighbor method [5]. Conventional methods generally have two requirements. First, text lines have to be straight and long enough. Second, text regions of a scanned document image have to be found in advance with a certain character extraction process. Recently the targets of scanners are not only regular business documents but also ones with short and sparse text lines on various backgrounds [6]. This fact requires us to develop a more robust skew estimation method to deal with various types of documents.

In this paper, we propose a novel skew estimation method to deal with character skew on various types of documents. The method is a part-based method where local parts of alphabet character images are the unit of estimating local skew angles. Since the proposed method utilizes local parts of characters, and does not rely on straightness of text lines, it can estimate the skew angle of a document image

consisting of short and sparse text lines. Furthermore, by utilizing the majority voting of local estimations, it is possible to remove the bad effects of backgrounds in a scanned document image to skew estimation. The proposed method can be directly applied to grayscale images without any connected component analysis on a binarized image for character extraction. In addition to a brief overview of initial trials proposed in [7], this paper provides several totally new experimental results for further evaluation.

The rest of the paper is organized as follows. In Sect. 2, conventional methods are first reviewed briefly. In Sect. 3, the principle of the proposed method is described. The experimental results are shown in Sect. 4. Finally Sect. 5 concludes the paper.

2. Related Work

2.1 Projection Profile Methods

Projection profile methods can be applied to binarized document images. On a document image, a histogram called a "projection profile" is acquired by accumulating the number of black pixels along parallel sample lines. In the case of a horizontal text, a projection profile has the narrowest peaks when taken horizontally along rows. The simplest method [3] takes projection profiles along expected orientations, and the one with the narrowest peaks gives the skew angle.

In a modified version proposed by Akiyama et al. [8], an input document image is divided into several strips first. A projection profile is, then, calculated on each of the strips at an angle. On each strip, a projection profile appears with its peaks gradually shifted from the adjacent strip in accordance with the skew angle. Therefore the skew angle is calculated from the strip width and the shift length. Sadri et al. [9] proposed a projection profile method based on particle swarm optimization to deal with larger skew angles.

2.2 Hough Transform Methods

The Hough transform can be applied to skew estimation (e.g., [4]). The lines passing through the centroids of characters in a document image can be found with the Hough transform. The slope of the lines corresponds to the document skew [10].

Amin and Fischer [11] proposed a modified version where neighboring connected components are grouped to

handle more complex layouts. To each of the groups of connected components, the Hough transform is applied to determine the document skew angle within the group. After the skew angle has been estimated for each group, the groups with similar skew angles are combined. Subsequently the group with the most angles is determined. The final skew estimation is done by averaging the skew angles within that group. Singh et al. [12] proposed a method based on a block adjacency graph algorithm to reduce the number of image pixels and speed up skew estimation.

2.3 Nearest Neighbor Methods

Nearest neighbor methods have an advantage in handling larger skew angles than the above-mentioned two types of methods. In the method proposed by Hashizume [5], connected components representing characters in a document are found first. Subsequently the nearest neighbor of each component is found, and angles of the lines connecting the centers of gravity of the nearest neighbor components are calculated. Since the spaces between characters on a line is generally smaller than the spaces between lines, it is resonable to assume that both the components in a nearest neighbor pair belong to the same text line. The calculated angles are accumulated in a histogram, and the peak of the histogram indicates the skew angle of the document.

A modified version proposed by O'Gorman [13] connects not only the nearest neighbors but also multiple neighbors to make the method more robust to noise. In this way, it eliminates less reliable estimations including those made by the connection across the lines.

The above three groups of conventional methods have a possibility of failing to estimate the correct skew angle on document images where characters are sparsely laid out, or backgrounds exist in images. This is because they utilize straightness of text lines in an image. Specifically, sparsely laid out characters possibly make a false text line, and backgrounds are sometimes falsely detected as characters.

2.4 Other Existing Approaches

There also exist several methods other than the above three groups. We note that they also have the same problem since they rely on straight text lines. Bar-Yosef et al. [14] proposed a skew estimation method based on a distance transform. In this method, an input image is converted to a distance transform image. Using gradient directions between text lines in the distance transform image, the skew angle can be estimated. Another approach was proposed in [15], where a parallelograms covering algorithm was used to find the background area.

Local estimation approaches have been used to deal with backgrounds in images [16]–[20]. In these methods, local skew estimations on rather small areas are first conducted, and then those estimations are combined into a reliable global skew estimation. Those methods, however, still rely on straightness of short text lines.

Takeda et al. [21] proposed an image retrieval method where they used rotation-invariant local features extracted from an input image to find the exact document from a large-scale database even under perspective distortion. The skew estimation method proposed in this paper works in a manner similar to this method. However, in comparison to Takeda's method, the proposed method does not need the exact document image in a database.

2.5 The Conventional Part-based Method

A solution to the problems caused by sparse text lines and backgrounds has been proposed in [22]. Similarly to the method proposed in this paper, it is also a part-based skew estimation method. In that method, the skew angle is estimated based on a skew estimation of each connected component (∼ a character). Specifically each connected component in an input image is identified by so-called rotation invariants. Once the connected components are identified, a skew angle of each of the connected components can be estimated. Finally the most frequent angle is chosen as the global skew angle. A drawback of this method is that it greatly relies on the accuracy of the connected component extraction.

3. Part-based Skew Estimation Method

3.1 The Key Idea

The proposed method first estimates the local skew angle for each of the local parts of characters in an input document image. The term "local part" in this paper refers to a part of a character such as the parts shown in the bottom of Fig. 1. We can estimate the skew of this local part by comparing its appearance to the identical part without skew. For instance, we can estimate the skew angle, observing the tilt of the corners or the edges.

The local parts should be selected around corners or crossing points of characters for distinctiveness. Accordingly "keypoint detectors" such as SURF and SIFT are appropriate for this purpose. Figure 1 shows an example of local parts detected by SURF [23]. In this example, the local parts are detected around the corner and large curvature areas, and simple straight line segments are not detected as local parts.

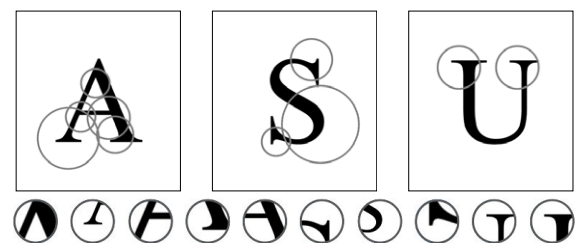When utilizing local parts of a document image, it can-

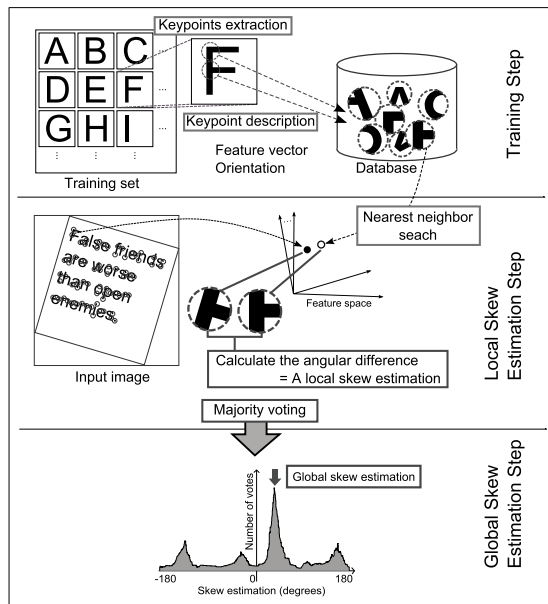**Fig. 1** Local parts detected by the SURF detector.

Fig. 2 The principle of the proposed method.

not be guaranteed that all the selected parts are distinctive enough. This is because characters often consist of similar and common parts. However, this is not a serious problem; it is possible to efficiently disregard the erroneous estimations acquired from those indistinctive parts by aggregating a sufficient number of the local skew estimations. The aggregating method is described in greater detail later in 3.4.

One of the benefits of using keypoint detectors is that preprocessing is not necessary. For example, any binarization process, which is required for connected component analysis, is not applied to an input image. The above keypoint detectors are directly applicable to grayscale images.

As shown in Fig. 2, the proposed method is comprised of three steps: a training step, a local skew estimation step, and a global skew estimation step. Their objectives are, respectively, to prepare a database of upright character parts, to estimate the local skew angles of the detected parts, and to aggregate the local skew estimation for a reliable global skew angle of the whole image.

3.2 Training Step

In the training step, reference local parts are stored into a database. First, upright character images (i.e., font images) are prepared as a training set. If necessary, multiple fonts of every character are prepared. Then local parts on each character image are detected automatically with a keypoint detector (SURF is used in this paper). Figure 1 is an example of the local parts detected with SURF. Each local part is described as a 128-dimensional feature vector and its dominant brightness gradient direction is also calculated. There are three important properties of SURF which are essential to estimate the local skew in the next step.

- First, the same local parts should be detected even if

the scale and/or the skew of the image has changed.
- Second, it can determine the "dominant orientation" at each local part based on brightness gradient of the part. In this paper, we employed the same calculation method used in SURF to calculate the orientation of keypoints, where the sum of the Haar wavelet responses around the point were utilized [23]. If the target image undergoes a skew (i.e., a rotation) of θ, the dominant orientation of a local part is also changed by θ from the original orientation.
- Third, SURF feature vector is skew (and scale) invariant. SURF adaptively changes the orientation and scale of its local part and describes the part as a vector. Consequently, the resulting vector keep unchanged regardless of the scale and skew of characters in the target image.

As shown in Fig. 2, all of the SURF feature vectors of the training images and their dominant orientations are paired and stored in a database. Each paired entry is considered as an instance and used in the local skew estimation step.

3.3 Local Skew Estimation

On an input image, local parts are detected in the same manner as the training step. The skew angle of each local part of the input image is estimated by referring to the database. Specifically, as shown in Fig. 2, the nearest neighbor search (measured by Euclidean distance in the feature vector space) is done for each of the detected local parts in the input image.

Because of the above-mentioned invariance of the SURF feature vectors on the skew and scale of the image, we can expect that the detected local part and its nearest neighbor are the identical local part of a certain character. Of course this is not the case when an image contains components that are not in the database such as graphs and pictures. This case is discussed in the next step. Recalling the second property of SURF explained in 3.2, the skew angle of each of the detected local parts in the input image is estimated just by calculating the difference of the dominant orientations of the detected local part and its nearest neighbor from the database.

3.4 Global Skew Estimation

The global skew angle is finally estimated by aggregating the estimation results of all the local parts of the input image in the previous step. Some estimations on the local parts deviate considerably from the true global skew angle. This deviation comes from the fact that not all of the detected keypoints are distinctive enough to find the correct match in the database. This is due to the similarities shared by different parts of characters (e.g. the top part of "p" and the bottom part of "d"). The incorrect local estimations are also acquired from backgrounds, including components such as

graphs and pictures, since it is impossible to know the skew angle of them without any stored instance for reference.

In order to suppress the effect of the erroneous local estimations, we use a simple majority voting scheme as shown in Fig. 2. Specifically, each local skew angle is voted into its corresponding bin of an angle histogram. The width of each bin is predetermined according to the skew sensitivity of the succeeding character recognition and also to the resolution of the selected keypoint detector. In this paper, the bin width is set to 0.5 degree and thus the histogram has 720 bins in total. The global skew angle is estimated as the angle of the bin with the maximum votes.

4. Experiments

We conducted several experiments to evaluate the performance of the proposed method. We used several different datasets for each experiment. In the following experiments the geometry of the images in the training set is common. All the training images were in grayscale. Each of the training images contains the full shape of one character. The character was printed in 120 pt, approximately 200 × 200 pixels, on each image of approximately 500 × 500 pixels with the white background. The size was selected experimentally as above for SURF detector to be able to stably detect keypoints and extract features. The character size dependency of the proposed method is mentioned in 4.1, and its experimental result is shown in Fig. 7. The size of the voting bins was set 0.5 degree in all of the experiments.

4.1 Document Images in a Single Font

We evaluated the basic performance of the proposed method in this experiment. Using SURF detector and descriptor, we first created a database of local parts of characters (including the upper and lower case alphabet characters and the numbers, all in Times New Roman). Several test images in the test set are shown at the top of Fig. 3. This test set consisted of 7200 text-only images (20 kinds of images each of which was rotated in one degree increments to from −180 degrees to 179 degrees). The characters in the test set were also in Times New Roman. The size of the test images was 1280 ×
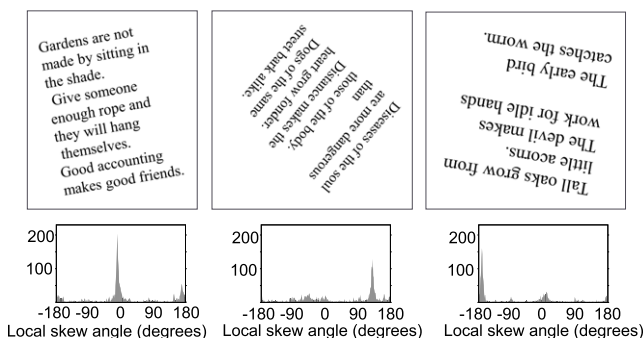
1881 pixels when set upright, and the size of the characters was approximately 100 × 100 pixels.

The voting results are shown under the corresponding test images at the bottom of Fig. 3. In the voting results, large peaks are acquired on the estimated angles whereas other incorrect bins have smaller numbers of votes. Figure 4 shows the actual local estimations on the test image. Both the arrows in dark colors indicating the accurate local estimation and the pale colors indicating the inaccurate local estimations can be seen in the image. However, the accurate local estimation was dominant, and it was selected as the global estimation by majority voting.

The result of 7200 images shown in Fig. 5 indicates that the proposed method is capable of estimating text skew efficiently. The absolute error of 99.6% of the test set was less than or equal to 2.0 degree. The average absolute error was 0.4 degree.

In addition, we have conducted a comparison experiment between the proposed method and three other conventional methods: the projection method [8], the Hough transform method [11], and the nearest neighbor method [5]. Due to the range limitation of the conventional methods, skew angles of test images were limited to [−10, 10]. The 20 images were rotated to from −10 degrees to 10 degrees in one degree increments.

As shown in Fig. 6, the performance of all the methods were similar except the nearest neighbor method, which had relatively larger errors. An advantage of the proposed method is that it does not need an explicit character ex-
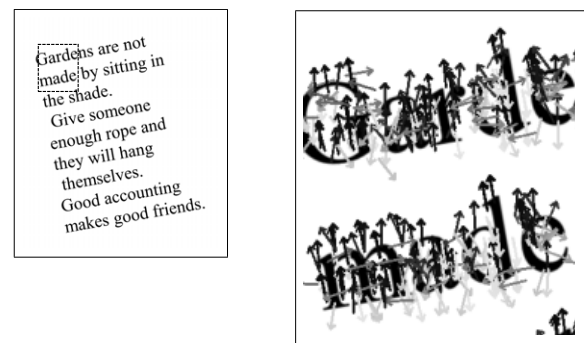

Fig. 4 An original input image (left), and local estimations on the detected keypoints (right). The estimations are shown by arrows. The darker color indicates a smaller estimation error.

Fig. 3 Tex-only document images and corresponding local skew estimation voting results.
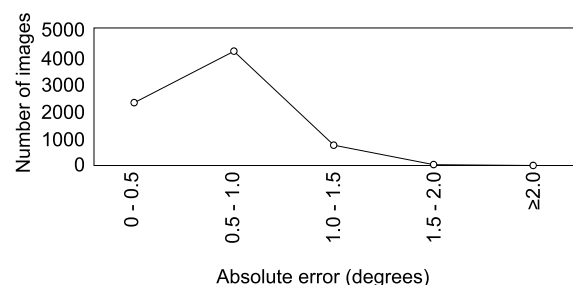
Fig. 5 The skew estimation error on the text-only document image test set.
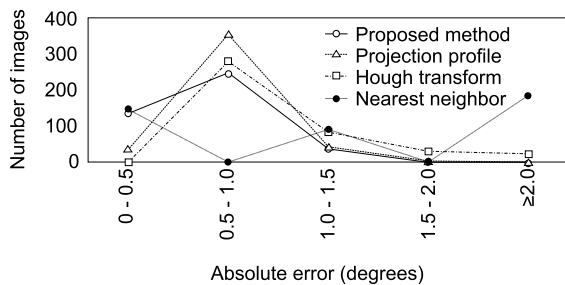
Fig. 6 The comparison result of the proposed method and the three conventional methods on the text-only document image test set.
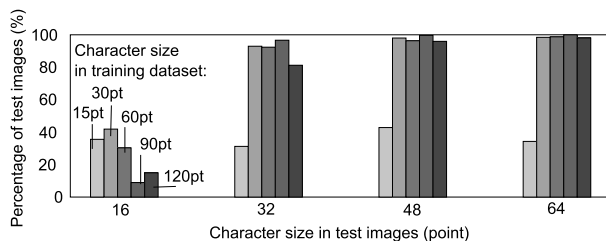
Fig. 7 The character size dependency of the proposed method. The y-axis shows the percentage of the test images whose absolute error did not exceed 2 degrees.

Fig. 8 Font images in the multi-font training set (from top left to bottom right): Times New Roman, Arial, Impact, Courier New. Bottom: Garamond, Georgia, Verdana.

Fig. 9 Performance on the single-font test set.

traction process. Since the test images used in this experiment did not contain the backgrounds, this advantage did not emerge. As a result, the proposed method and two of the conventional methods showed the similar results. The performance of the proposed method on the test images with the background is examined later in 4.3.
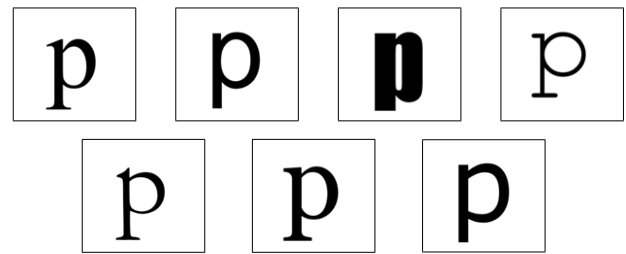
We also tested the character size dependency of the proposed method. The test result is shown in Fig. 7. Although the performance of the proposed method decreased when the character size was 16 points, it was constantly high on larger sizes.

The proposed method performed less accurately on the small characters because the keypoint detection and the feature extraction were not done efficiently on the small characters in the test images. This is the same for the character size of the training dataset. The feature extraction cannot be done efficiently in the training step if the size is too small. In the future, we can tackle this problem by simply magnifying the test images, adjusting the parameters used in SURF, or employing a more appropriate feature extraction method. In fact, we had a better result when we magnified the 16-point test images. We magnified those images four times, and the absolute error of 81% was less than or equal to 1.0 degree.
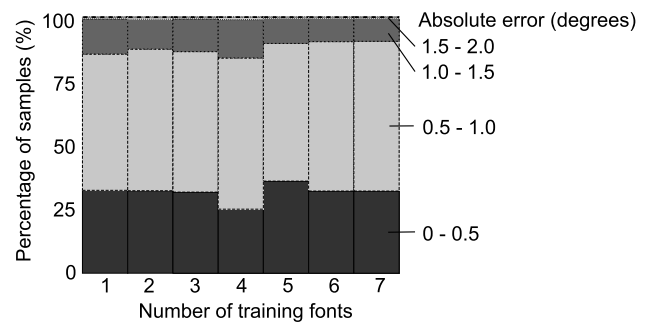
4.2 Document Images in Different Fonts

For the proposed method to be applicable to multiple fonts, more than one kind of font in a training set is necessary. In this experiment we evaluated the performance of the proposed method when there were more than one font in the *training* set.

Ideally, the proposed method should have the follow-

ing two characteristics. First, when a test set has only one kind of font, the accuracy of the skew estimation should not decrease as the number of the fonts in the training set increases. Second, when a test set has multiple fonts, the proposed method should be able to deal with them by having the corresponding fonts in the training set. To check the above characteristics of the proposed method, we used two test sets. One test set, the single-font test set, consisted of only the images containing characters in Times New Roman, which are similar as the ones used in the previous experiment. The other test set, the multi-font test set, consisted of images each of which contained one of the following 7 different fonts, Times New Roman, Arial, Impact, Courier New, Garamond, Georgia, and Verdana, shown in Fig. 8. Note that each document image in this test set contained one of those fonts. We added fonts in the training set one by one in the above order and observed the performance change of the proposed method.

The result of the first test set, the single-font test set, is shown in Fig. 9. The estimation accuracy of the proposed method changed only by a small amount when the other fonts were included in the training set. This is an important characteristic for the proposed method because it implies that we can put many fonts in the training set without seriously decreasing the accuracy of the proposed method.

The accuracy recovered after the number of the fonts in the training set had become five from four. From this fact, it can be said that the local parts detected from the added font, Garamond, was utilized to estimate the skew of Times New Roman. This implies that one font in the training set can be utilized to estimate the skew of other fonts.
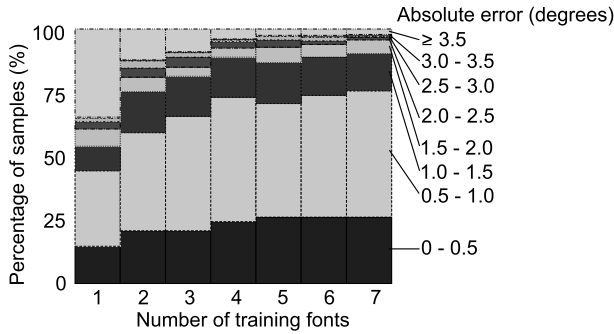
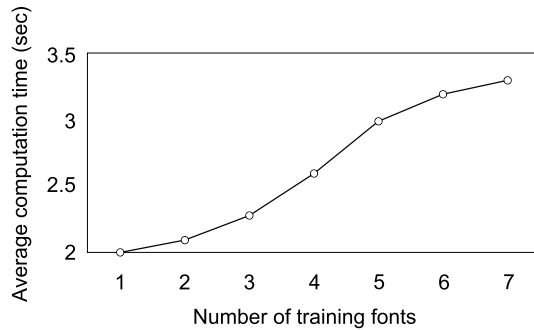**Fig. 10** Performance change on the multi-font test set.

Fig. 11 Average computation time of the proposed method on a test image used in Sect. 4.2.

Figure 10 shows the performance change of the proposed method on the multi-font test set as the number of the fonts in the training set increased one by one. The choices of training fonts and their order to add were the same as the previous experiment, and only the test images were different in this experiment. It can be observed that although the addition of more fonts to the training set increased the performance, the benefit became less as the number of training fonts increased. In other words, it is not necessary to register all the fonts of input images to the database.

Note that the error decreased dramatically after the second font, i.e., Arial was added. This is because these two kinds of fonts were different in terms of the existense of the decorative structures at the end of the strokes. (That is, Times New Roman is a serif font whereas Arial is a sans-serif font). Two fonts, Impact and Courier, have very different shapes compared to the others, so adding them to the training set also improved the performance. Finally adding Garamond (Serif), Georgia (Serif), and Veradana (Sans-serif) did not improve the performance as much because these fonts shared the similar types of shapes to a degree with the already existing fonts, Times New Roman, and Arial.

Figure 11 shows the average calculation times. The experiment was conducted on a computer under 64-bit Windows 7 Professional, with an intel Core i7 CPU 970 @ 3.20 GHz, and 24.0 GB of RAM. The nearest neighbor search for local skew estimation has the time complexity $O(nm)$, where n and m are the numbers of keypoints ex-

tracted from the training set, and an input image respectively. If it is necessary to reduce the computations, we can employ the approximate nearest neighbor search.

4.3 Document Images with Non-character Backgrounds

Scan of documents that contain not only characters, but also non-character backgrounds is recently common as mentioned in Sect. 1. We have conducted two experiments on such documents.

The first experiment was a comparison between the proposed method and three other conventional methods: the projection method [8], the Hough transform method [11], and the nearest neighbor method [5]. As a test set we used synthetic 100 magazine-cover-like images. All the images in the test set are shown in Fig. 12. Since characters in those images were in Times New Roman, we prepared a database of Times New Roman font. Each of the images were rotated to random angles with the range of [−10, 10], ending up with 300 test images in total. The range was decided as above so that all the three conventional methods achieved the highest accuracy. The highest accuracy of the projection profile is achieved within the range of [−10, 10] [10]. The Hough transform method and the nearest neighbor method have their highest accuracy within the range of [−15, 15] and [−90, 90] respectively [10].

Since the conventional methods need character extraction before their skew estimation process (Fig. 14), we first binarized the test images using Otsu's method [24], and found connected components. We removed non-character objects from the images by filtering connected components. For the filtering, we used a method proposed in [25]. In the filtering process, the following three thresholds are necessary. They were the threshold for (1) the size, (2) the aspect ratio of the bounding box, and (3) the ratio between the area of a connected component and its bounding box. Since the first threshold must be determined according to the size of used test images and the accuracy of the filtering is mostly dependent on this threshold, we used two different values. The first one ("Threshold A") was selected so that the characters were rather appropriately extracted on a document, but small remains of the non-character objects were also kept. The second one ("Threshold B") was selected so that the small remains of the non-character objects could be filtered out, but also the small characters were removed. For the second and third thresholds, we used the same fixed number as [25].

The result is shown in Fig. 13. The absolute error of the proposed method was less than or equal to 2.0 degrees for 98.0% of the test set. By the projection profile method, the Hough transform method, and the nearest neighbor method, 34.0%, 68.0%, and 51.7% of the test set had the absolute error being less than or equal to 2.0 degrees respectively.

In this experiment, the proposed method had much more robustness compared to the conventional methods. This is because the proposed method can effectively exploit the local parts of characters in grayscale images. The con-
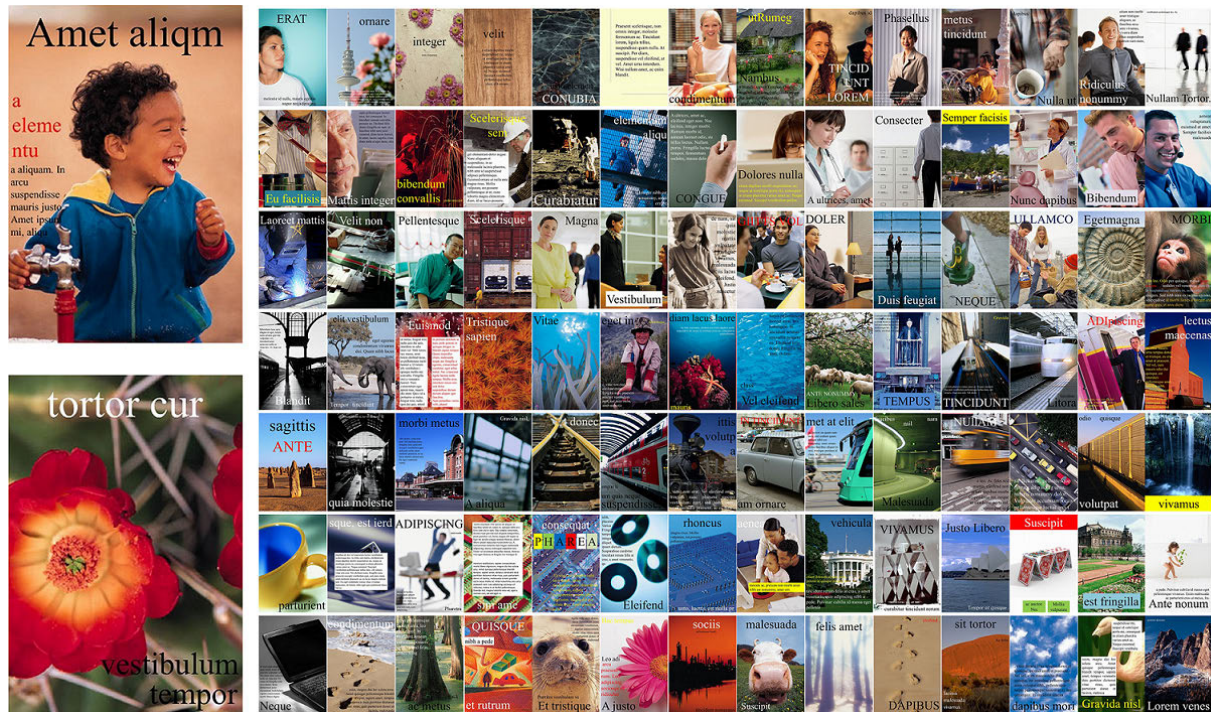
Fig. 12　All of the 100 magazine-cover-like test images.

(a) The proposed method.

(b) The projection profile method.

(c) The Hough transform method.
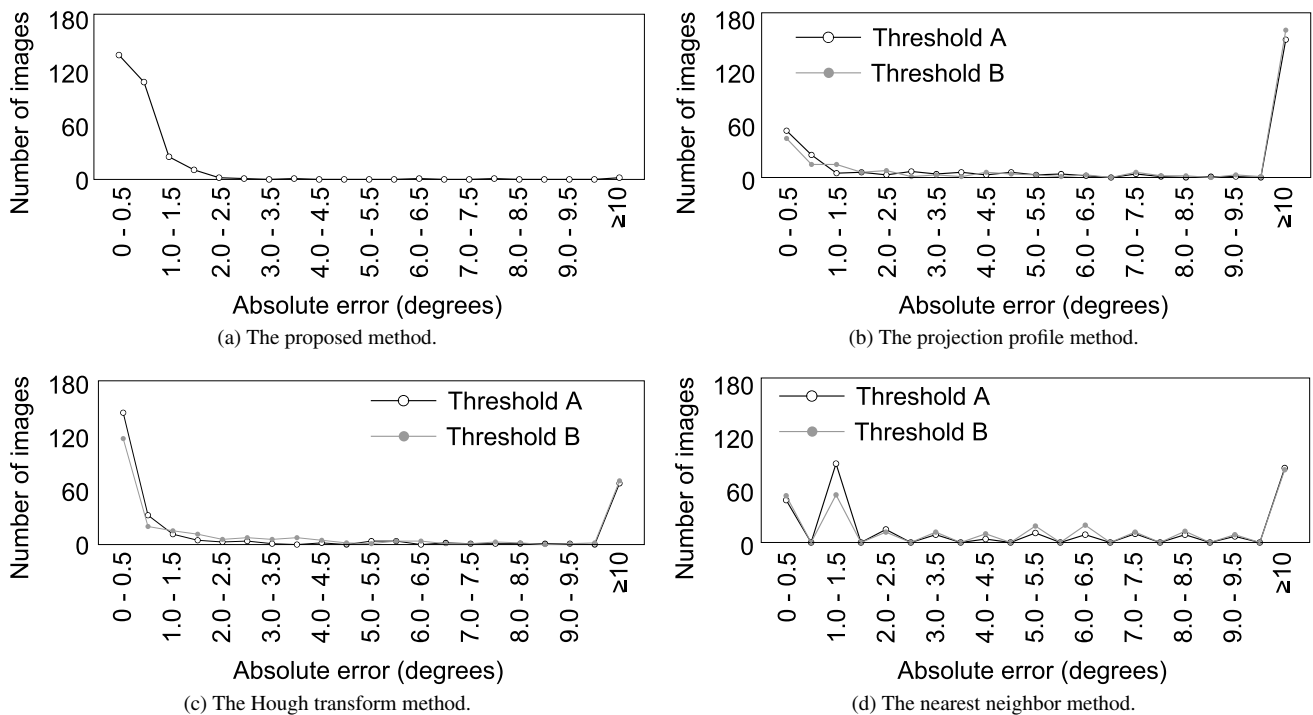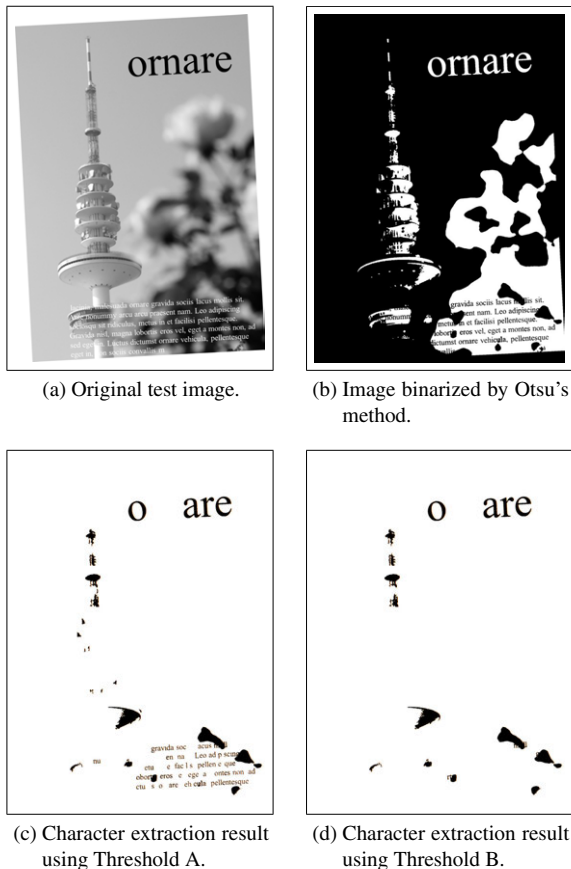
(d) The nearest neighbor method.

Fig. 13　The comparison among the results of the four methods on 300 magazine-cover-like images.

ventional methods, on the other hand, need a certain level of accuracy of character extraction, connected component analysis, and filtering.

　Figure 15 shows one of the test images with the character extraction process for the conventional methods ((a), (b), (c)), and the local estimations made by the proposed method ((d), (e)). Because of the low contrast with the background, characters were not extracted correctly. Accordingly the conventional methods were not able to estimate the skew angle. In contrast, the proposed method does not need the

(a) Original test image.

(b) Image binarized by Otsu's method.

(c) Character extraction result using Threshold A.

(d) Character extraction result using Threshold B.

Fig. 14 The character extraction process for the conventional skew estimation methods.

(a) Input image.

(b) Binarization result.

(c) Character extraction result.

(d) Local estimations acuqired from non-character objects

(e) Local estimations acquired from a character region.
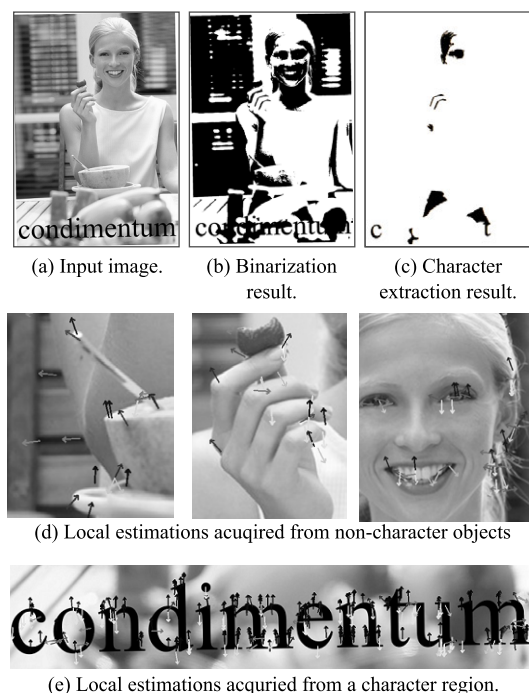
Fig. 15 A magazine-like test image on which the proposed method has an advantage against the conventional methods. ((a), (b), (c): the character extraction process, and (d), (e): local estimations by the proposed method.)
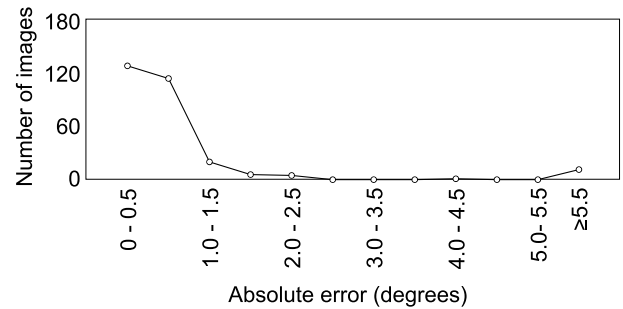
Fig. 16 Skew estimation error on the test set with larger skew angles.

character extraction process; whereas the incorrect local estimations were acquired randomly on the non-character objects (Fig. 15 (d)), many uniform local estimations were acquired from the character region and the estimation became dominant (Fig. 15 (e)).
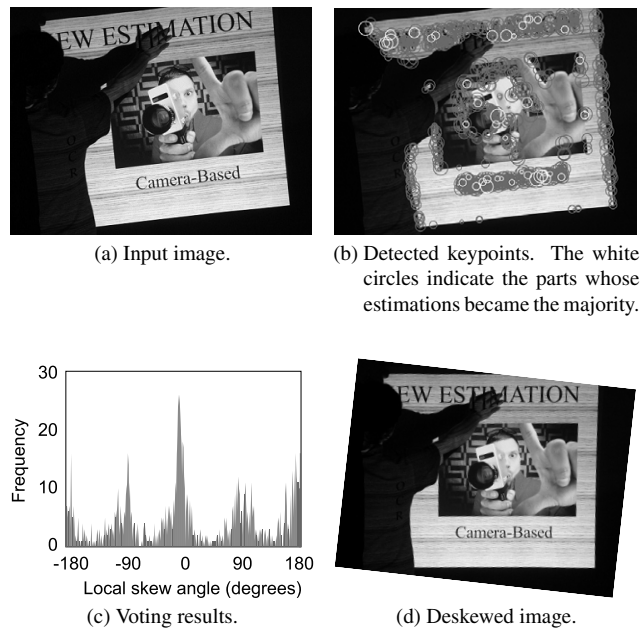
Among the conventional methods, the Hough transform method and the nearest neighbor method had relatively higher accuracy. This is because they utilize majority voting in their methods for less influence of non-character objects. However, their estimations on several documents were still mislead by non-character backgrounds.

The second experiment was conducted to evaluate the accuracy of the proposed method on more largely skewed document images. As mentioned above, the projection profile methods and the Hough transform methods can only deal with the skew angles being less than 10 degrees and 15 degrees respectively, and the nearest neighbor can only deal with the skew angles being less than 90 degrees. One advantage of the proposed method is that it can be used to estimate larger skew angles than the conventional methods. The same 100 magazine-like document images used in the above experiments were skewed more largely and used as a test set for this experiment. The 300 test images were acquired by rotating them to random angles with the range of [−180, 180].

The result is shown in Fig. 16. The proposed method was able to deal with 96% of the test set with an error being less than or equal to 2.0 degrees. The result shows that the proposed method efficiently estimates document skew which is out of the range of application for the conventional methods.

4.4 Camera-captured Image

In actual situations, camera-captured images often contain skewed characters. In order to seek the possibility to apply the proposed method to such images, we conducted an experiment. We applied the proposed method to a camera-captured image shown in Fig. 17 (a). The image was taken from an appropriate position so that the image did not have large perspective skew. Figure 17 (b) shows the detected parts by SURF detector on the test image. By majority voting, the global estimation was made mainly based on the local estimation acquired from the character regions. Fig-

(a) Input image.

(b) Detected keypoints. The white circles indicate the parts whose estimations became the majority.

(c) Voting results.

(d) Deskewed image.

Fig. 17 Skew correction on a camera-captured image by the proposed method.

ure 17 (c) is the voting result, and Fig. 17 (d) shows the deskewed image.

5. Conclusion

In this paper, a new approach for skew estimation was proposed. It is an important characteristics for a skew estimation method to be applicable to documents with sparse text lines and backgrounds. The proposed method is able to deal with those documents by locally estimating the skew angles on parts of characters.

Many comparison experiments conducted on documents containing non-character backgrounds indicate that the proposed method works much more accurately than the conventional methods on those images. This is because of the following reasons. First, since the proposed method is part-based, it can be used for such characters that only parts of them are available. It is difficult for the conventional methods to deal with such cases because, in those methods, it is necessary to binarize an image and extract characters as connected components appropriately so that text lines can be found. Second, the majority voting process enables the proposed method to disregard incorrect local estimations obtained from non-character backgrounds, and pick out only the local estimations calculated from characters.

In the future, database reduction, for instance by clustering, should be taken into consideration for faster computational and more efficiency. Other local feature descriptors that are more accurate than SURF can also be sought for to improve the accuracy of the proposed method. Furthermore we shall work on modifying the proposed method so that it can also be applied to perspective skew, which often occurs in camera-captured images.

References

[1] G. Nagy, "Twenty years of document image analysis in PAMI," IEEE Trans. Pattern Anal. Mach. Intell., vol.22, no.1, pp.38–62, Jan. 2000.

[2] J. Hull, "Document image skew detection: Survey and annotated bibliography," in Docuent Analysis Systems II, ed., J. Hull, pp.40–64, World Scientific, 1998.

[3] W. Postl, "Detection of linear oblique structures and skew scan in digitized documents," Proc. Int. Conf. Document Analysis and Recognition, pp.687–689, 1986.

[4] S.N. Srihari and V. Govindaraju, "Analysis of textual images using the Hough transform," Machine Vision and Applications, vol.2, no.3, pp.141–153, 1989.

[5] A. Hashizume, P.-S. Yeh, and A. Rosenfeld, "A method of detecting the orientation of aligned components," Pattern Recognit. Lett., vol.4, no.2, pp.125–132, April 1986.

[6] G. Meng, C. Pan, N. Zheng, and C. Sun, "Skew estimation of document images using bagging," IEEE Trans. Image Process., vol.19, no.7, pp.1837–1846, 2010.

[7] S. Shiraishi. Y. Feng, and S. Uchida, "A part-based skew estimation method," Proc. Int. Workshop Document Analysis Systems, pp.185–189, 2012.

[8] T. Akiyama and N. Hagita, "Automated entry system for printed documents," Pattern Recognit., vol.23, no.11, pp.1141–1154, 1990.

[9] J. Sadri and M. Cheriet, "A new approach for skew correction of documents based on particle swarm optimization," Proc. Int. Conf. Document Analysis Recognition, pp.1066–1070, July 2009.

[10] L. O'Gorman and R. Kasturi, ed., Document Image Analysis, IEEE CS Press, Los Alamitos, 1997.

[11] A. Amin and S. Fischer, "A document skew detection method using the Hough transform," Pattern Analysis and Application, vol.3, no.3, pp.243–253, 2000.

[12] C. Singh, N. Bhatia, and A. Kaur, "Hough transform based fast skew detection and accurate skew correction methods," Pattern Recognit., vol.41, no.12, pp.3528–3546, Dec. 2008.

[13] L. O'Gorman, "The document spectrum for page layout analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol.15, no.11, pp.1162–1173, Nov. 1993.

[14] I. Bar-Yosef, N. Hagbi, K. Kedem, and I. Disnstein, "Fast and accurate skew estimation based on distance transform," Proc. Int. Workshop Document Analysis Systems, pp.402–407, Sept. 2008.

[15] A. Mascaro, G. Cavalcanti, and C. Mello, "Fast and robust skew estimation of scanned documents through background area information," Pattern Recognit. Lett., vol.31, no.11, pp.1403–1411, Aug. 2010.

[16] U. Pal, M. Mitra, and B.B. Chaudhuri, "Multi-skew detection of indian script documents," Proc. Int. Conf. Document Analysis and Recognition, pp.292–296, Sept. 2001.

[17] Y. Ishitani, "Document skew detection based on local region complexity," Proc. Int. Conf. Document Analysis and Recognition, pp.49–52, 1993.

[18] X. Jiang, H. Bunke, and D. Widmer-Kljajo, "Skew detection of document images by focused nearest-neighbor clustering," Proc. Int. Conf. Document Analysis and Recognition, pp.629–632, 1999.

[19] Y. Lu and C.L. Tan, "Improved nearest neighbor based approach to accurate document skew estimation," Proc. Int. Conf. Document Analysis and Recognition, pp.503–507, 2003.

[20] S. Lu and C.L. Tan, "Camera document restoration for OCR," Proc. Int. Workshop Camera-Based Document Analysis and Recognition, pp.17–24, 2005.

[21] K. Takeda, K. Kise, and M. Iwamura, "Real-time document image retrieval for a 10 million pages database with a memory efficient and stability improved LLAH," Proc. Int. Conf. Document Analysis and Recognition, pp.1054–1058, 2011.

[22] S. Uchida, M. Sakai, M. Iwamura, S. Omachi, and K. Kise, "Skew

estimation by instances," Proc. Int. Workshop Document Analysis Systems, pp.201–208, 2008.

[23] H. Bay, A. Ess, T. Tuytelaars, and L.V. Gool, "SURF: Speeded up robust features," Comput. Vis. Image Understand., vol.110, no.3, pp.346–359, June 2008.

[24] N. Otsu, "A threshold selection method from gray-level histograms," IEEE Trans. Syst., Man Cybern., vol.9, no.1, pp.62–66, Jan. 1979.

[25] P. Saragiotis and N. Papamarkos, "Local skew correction in documents," Int. J. Pattern Recognition and Artificial Intelligence, vol.22, no.4, pp.691–710, 2008.

Soma Shiraishi received B.E. degree from Kyushu University, Japan in 2009. He is currently pursuing his M.E. in Kyushu University. His current research interests include image analysis and sequential pattern recognition.

Yaokai Feng obtained his B.E. and M.E. degrees in Computer Science from Tianjin University, China, in 1986 and 1992, respectively. He obtained the Ph.D. degree in Information Science from Kyushu University, Japan, in 2004. Now, he is an assistant professor in the Graduate School of Information Science and Electrical Engineering, Kyushu University, Japan. His current research interests include pattern recognition, information retrieval and network security.

Seiichi Uchida received B.E., M.E., and Dr. Eng. degrees from Kyushu University in 1990, 1992 and 1999, respectively. From 1992 to 1996, he joined SECOM Co., Ltd., Tokyo, Japan where he worked on speech processing. Currently, he is a professor at Faculty of Information Science and Electrical Engineering, Kyushu University. His research interests include pattern recognition and image processing. He received 2002 IEICE PRMU Research Encouraging Award, MIRU2006 Nagao Award (best paper award), 2007 IAPR/ICDAR Best Paper Award, 2009 IEICE Best Paper Award, and MIRU2011 Excellent Paper Award. Dr. Uchida is a member of IEEE and IPSJ.