Instance-Based Skew Estimation of Document Images by a Combination of Variant and Invariant

Seiichi Uchida
Kyushu Univ., Japan
uchida@is.kyushu-u.ac.jp

Megumi Sakai
Kyushu Univ., Japan
sakai@human.is.kyushu-u.ac.jp

Masakazu Iwamura
Osaka Pref. Univ., Japan
masa@cs.osakafu-u.ac.jp

Shinichiro Omachi
Tohoku Univ., Japan
machi@aso.ecei.tohoku.ac.jp

Koichi Kise
Osaka Pref. Univ., Japan
kise@cs.osakafu-u.ac.jp

Abstract

A novel technique for estimating geometric deformations is proposed and applied to document skew (i.e., rotation) estimation. The proposed method possesses two novel properties. First, the proposed method estimates the skew angles at individual connected components. Those skew angles are then voted to determine the skew angle of the entire document. Second, the proposed method is based on instance-based learning. Specifically, a rotation variant and a rotation invariant are learned, i.e., stored as instances for each character category, and referred for estimating the skew angle very efficiently. The result of a skew estimation experiment on 55 document images has shown that the skew angles of 54 document images were successfully estimated with errors smaller than 2.0 degree. The extension for estimating perspective deformation is also discussed for the application to camera-based OCR.

1. Introduction

For document image processing, the estimation of geometric deformations is an important problem. For example, many researchers have widely investigated the estimation of skew deformation, which severely degrades performance of OCR. Recently, the estimation of perspective deformation also becomes important problem toward the realization of camera-based OCR [1].

In this paper, a novel deformation estimation method is proposed and its performance is evaluated qualitatively and quantitatively via several experiments. In principle, the proposed method can estimate various geometric deformations such as perspective deformation and affine deformation. In this paper, we will focus on the document skew (i.e., rotation) estimation problem, which will be a reasonable problem to observe the basic performance of the proposed method.

The proposed method possesses two novel properties. First, the proposed method estimates the skew angles at individual connected components. (Note that each connected component may be a character). Those skew angles are then voted to determine the skew angle of the entire document. This fact implies that the proposed method does not rely on the straightness of text lines, whereas the conventional skew estimation methods totally rely on the straightness. Thus, the proposed method can be applied to documents whose component characters are laid out irregularly. This property is very favorable for camera-based OCR whose targets will often be short words and/or characters laid out freely.

Second, the proposed method employs instance-based learning for estimating the skew angle by referring stored instances. The simplest realization of instance-based skew estimation will be done by using font images as instances; the skew angle of each connected component is estimated by rotating and matching the font image to the connected component. The rotation angle giving the best match is the estimated skew angle. This simple realization, however, is very naive and requires huge computations. Specifically, it requires $O(N \cdot C \cdot K)$ image matchings, where N is the number of connected components in the target document, C is the number of instances (i.e., the number of assumed character categories), and K is the number of quantized angles (e.g., 360 for the estimation of 1° resolution). The proposed method avoids this problem by using a *rotation invariant* and a *rotation variant* as the instances and therefore requires only $O(N)$ (or less) computations.

The rest of this paper is organized as follows. After a brief review of the conventional methods in Section 2, the proposed method is described in Section 3. The role of the variant and the invariant is also detailed. Through these de-

scriptions, it will be clarified that the proposed method does not rely on the straightness of text lines. In Section 4, the performance of the proposed method is observed via a skew estimation experiment of several document images. After remarking the extensibility of the proposed method for estimating deformations other than rotation in Section 5, a conclusion is drawn in Section 6 with a list of future work.

2. Related Work

Thee skew estimation strategies of the conventional methods are classified into two types, global estimation and local estimation. The global estimation strategy utilizes global features such as projection histogram, whereas the local estimation strategy utilizes local features such as the principal axis of adjacent connected components. In the latter strategy, local skew angles are estimated first by the local features and then combined to determine the skew angle of the entire document. Although the local estimation strategy is minority, it possesses several good properties. Especially, its robustness to irregular layouts (such as short or scattered text lines, figures, mathematical notations, and multi-column layouts). It also has the extensibility to non-uniform skew estimation problems such as document image dewarping and multi-skew detection [2].

In Ishitani [3], a local skew angle is estimated within a circular window on a document. This local skew estimation is done by searching for the direction which gives the most intensity transitions. Among the estimated skew angles at different windows, the most reliable one is chosen as the global skew angle. Jiang et al. [4] have employed a least mean square fitting for estimating local skew angles. They choose the global skew angle by voting those local skew angles. Lu and Tan [5] have determined a group of connected components (called a nearest-neighbor chain) which comprise a word (or a sub-word) by a region growing technique and then its skew angle is estimated. The global skew angle is chosen as the medium or the mean of the local skew angles. Lu and Tan [6] have proposed an interesting method which utilizes the straight strokes of individual characters for estimating local skew angles.

All of those conventional methods rely on the local straightness of the text lines and/or character strokes. The proposed method does not assume any straightness and thus possesses far more robustness to irregular layouts than the conventional methods. As noted in Section 1, this property is favorable for camera-based OCR.

3. Instance-Based Skew Estimation

3.1. Learning instance

The proposed method estimates the skew angle of each connected component in the target document by referring stored instances. The detail of the estimation will be discussed in Section 3.2. This section describes how to learn the instances, i.e., how to prepare the instances.

The instances are comprised of a rotation variant $p_c(\theta)$ and a rotation invariant q_c where $c \in [1, \ldots, C]$ denotes the character category[1] and θ denotes the skew angle. They are prepared according to the following steps:

1. Define the C character categories which will be included in target documents.
2. For each category c,
   (a) prepare the font image $\boldsymbol{R}_c$,
   (b) measure the value of the rotation invariant q_c, and
   (c) measure the value of the rotation variant $p_c(\theta)$ by rotating $\boldsymbol{R}_c$ by θ.

While any rotation variant and invariant can be used, the following simple variant and invariant are used as $p_c(\theta)$ and q_c in this paper:

$$p_c(\theta) = \frac{\text{area of bounding box of } \boldsymbol{R}_c \text{ at } \theta}{\text{area of black pixels of } \boldsymbol{R}_c}, \quad (1)$$

$$q_c = \frac{\text{area of convex hull of } \boldsymbol{R}_c}{\text{area of black pixels of } \boldsymbol{R}_c}. \quad (2)$$

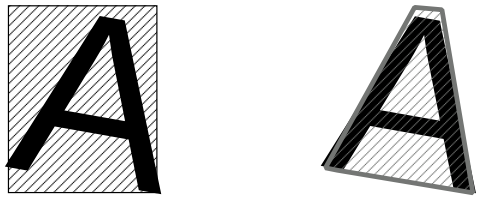
Fig. 1 shows the bounding box and the convex hull of a character. The area of the bounding box depends on the rotation angle θ, whereas the area of the convex hull does not. Fig. 2 shows the variant $p_c(\theta)$ of "y"(Times-Roman) as a function of θ. This function is stored as an instance together with q_c.

The rotation variant of (1) becomes a periodic function of $[-45°, 45°]$. Thus, the variant cannot distinguish, for example, $30°$ and $120°$. If necessary, it is possible to avoid this periodic property by using a variant other than (1).

Note that both $p_c(\theta)$ and q_c are scale and shift invariants. (Thus, q_c is an invariant to similarity transformation.) This scale and shift invariance implies that the proposed method can estimate the correct skew angle regardless of the character size and position.

Although we should define the categories at the learning step, this definition need not to be strict; that is, the proposed method can estimate correct skew angle even if the

[1]Different fonts and styles belong to different categories. Thus, $C = 52 \times 3 \times 2 = 312$, when we assume three styles (e.g., "upright", "italic", and "bold") and two fonts (e.g., "Times-Roman" and "Sans Serif") for 52 categories of "A"∼"Z" and "a"∼"z."



(a) (b)

Figure 1. (a) Bounding box and (b) convex hull of $c =$"A."

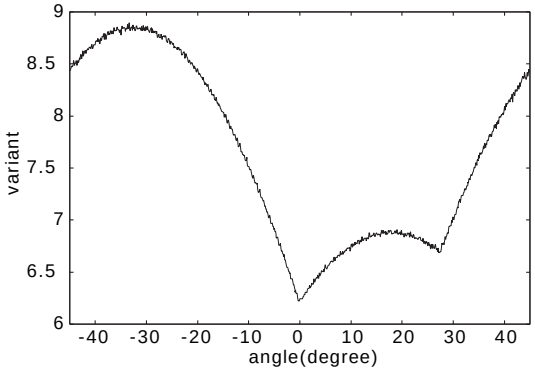


Figure 2. The variant $p_c(\theta)$ of "y."

target document includes the characters whose instances are not learned. As discussed later, this robustness comes from the voting strategy for determining the skew angle of the entire document. In Section 4, the robustness will be experimentally shown through the skew estimation result of a mathematical document which includes several undefined characters (i.e., mathematical symbols).

3.2. Skew estimation by instances

The estimation of the skew angle of a binarized document image is done by a three-step manner: (i) the estimation of the category of each connected component by the invariant q_c (→3.2.1), (ii) the estimation of the skew angle of the connected component by the estimated category c and the variant $p_c(\theta)$ (→3.2.2), and (iii) the estimation of the skew angle of the entire document by voting (→3.2.3).

3.2.1 Category estimation by invariant

Let $\boldsymbol{X}$ denote a connected component ($\simeq$ a character) of the target document image. The category of $\boldsymbol{X}$ can be estimated by comparing its invariant value q_x to the stored instances $\{q_c|c = 1, \ldots, C\}$. If $q_c = q_x$, c is the estimated category of $\boldsymbol{X}$. Since q_x and q_c are rotation invariants, this estimated category will not change under any skew angle.

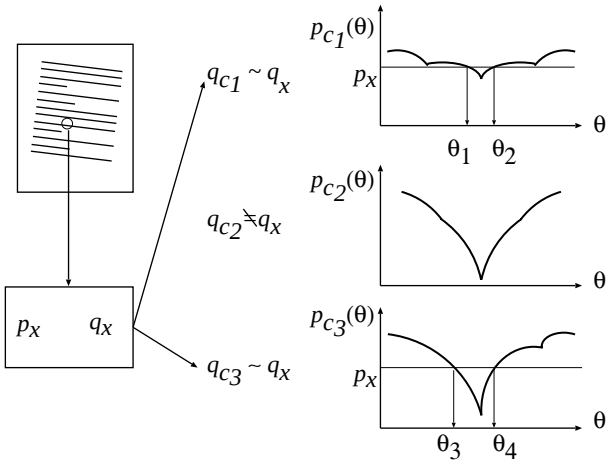


Figure 3. Estimation of skew angle by variant and invariant.

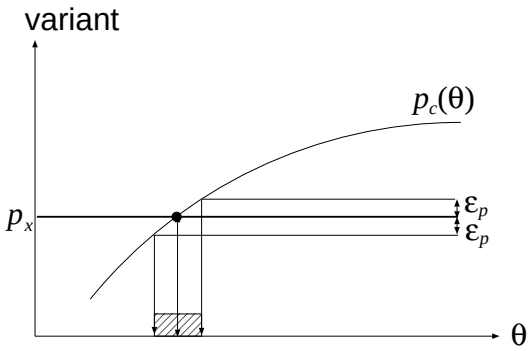


Figure 4. The range of the skew angle for variant p_x.

In practice, the connected component $\boldsymbol{X}$ may be contaminated by noises at image acquisition and thus q_x will be changed from its original value. Thus, all the categories satisfying $|q_c - q_x| \leq \epsilon_q$ are considered as the estimated categories, where ϵ_q is a non-negative constant. In the example of Fig. 3, we will obtain two estimated categories c_1 and c_3 because $q_x \sim q_{c_1} \sim q_{c_3}$, whereas $q_x \not\sim q_{c_2}$.

The category candidate c which satisfies $|q_c - q_x| \leq \epsilon_q$ can be found very efficiently by a look-up table which is indexed by the invariant value q_x. This efficiency is especially useful when a large number of categories are defined for dealing with various fonts, mathematical symbols, multilingual documents, and so on.

3.2.2 Local skew estimation by variant

For each estimated category c, the skew angle of $\boldsymbol{X}$ can be estimated by comparing its variant value p_x to the stored variance $p_c(\theta)$. The angle θ satisfying $p_c(\theta) = p_x$ is a candidate of the skew angle of $\boldsymbol{X}$ and therefore a candidate of the document skew angle. (Precisely speaking, we will use the relaxed condition $|p_c(\theta) - p_x| \leq \epsilon_p$ instead of $p_c(\theta) = p_x$, where ϵ_p is a non-negative constant. This will be discussed in 3.2.3.)

Consequently, multiple skew angle candidates will be obtained from a single connected component $\boldsymbol{X}$. This is because a single connected component $\boldsymbol{X}$ will have multiple category candidates and, furthermore, each category candidate c will provide multiple skew angle candidates satisfying $p_c(\theta) = p_x$. In the example of Fig. 3, two candidates θ_1, θ_2 are obtained from $p_{c_1}(\theta)$ and two candidates θ_3, θ_4 are from $p_{c_3}(\theta)$. In other words, four candidates are obtained from a single connected component.

Like the above category estimation step, the angle θ which satisfies $p_x = p_c(\theta)$ can be found very efficiently by a look-up table indexed by the variant value. This table is considered as the inverse function $\theta = p_c^{-1}(p_x)$.

3.2.3 Global skew estimation by voting

A voting strategy is employed for estimating the skew angle of the entire document. Roughly speaking, the purpose of the voting is to find the most frequent skew angle among all the candidates obtained by the above (local) skew estimation step. The voting strategy makes the proposed method tolerant to the false category candidates and the false skew angle candidates. Another merit is the tolerance to undefined categories. The bad effect of the undefined categories can be minimized by voting far more candidates representing the correct skew angle.

The skew angle of the entire document is estimated by voting the "range" specified by each skew angle candidate. As shown in Fig. 4, the range is determined as $[p_c^{-1}(p_x - \epsilon_p), p_c^{-1}(p_x + \epsilon_p)]$ by assuming that the true value of the variant p_x lies within $[p_x - \epsilon_p, p_x + \epsilon_p]$. (This range is come from the relaxed condition $|p_c(\theta) - p_x| \leq \epsilon_p$.) The skew angle is finally obtained as the angle where the most ranges are overlapped.

It is noteworthy that the width of the range is negatively proportional to the reliability of the skew angle candidate. This fact can be understood from the following example: Consider an "o"-shaped character. The reliability of the skew angle estimated at the character will be low because the character does not change its shape by any skew. In this case, its skew variant $p_c(\theta)$ will change subtly according to θ and the range determined by the variant becomes wide. In contrast, the variant of an "I"-shaped character, which will provide a highly reliable skew angle, will change drastically according to θ and the range becomes narrow.

3.2.4 Computational feasibility

The proposed method has a strong computational feasibility. This strength is emphasized not only by the use of the invariant and the variant but also by the look-up tables. The proposed method does not perform any try-and-error skew estimation step, unlike the global estimation methods based on the projection histogram and the local estimation methods like [3]. Furthermore, the proposed method requires neither line fitting nor image processing to search neighborhoods of each connected component. The proposed method, of course, does not require any costly image matching procedure, unlike the simple realization of the instance-based skew estimation outlined in Section 1.

The computational feasibility of the proposed method may be further improved by using a limited number of connected components. In fact, it is not necessary to use all the connected components in the document as experimentally shown later. This is because all the connected components, in principle, will show the same skew angle and thus the voting result will show the peak at the correct skew angle even with a limited number of votes.

4. Experimental Results

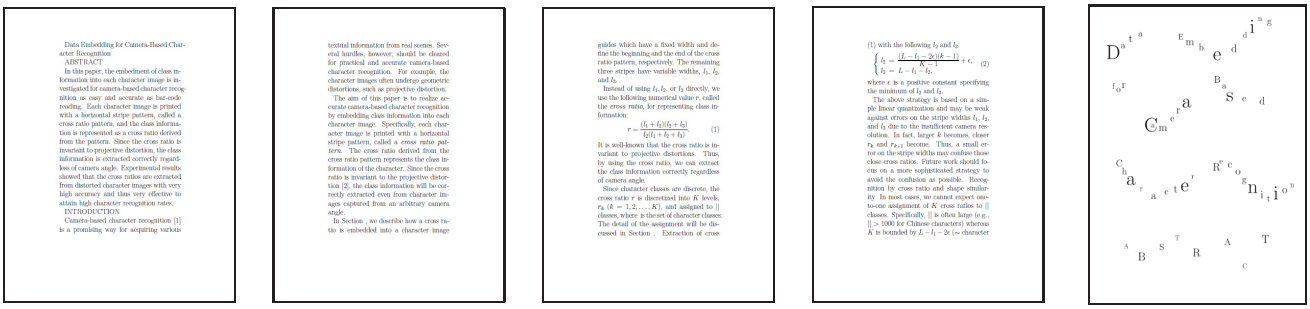
4.1. Document image samples

Five document images were created by LaTeX with a single font and style (Times-Roman, upright) and used for the evaluation of the skew estimation accuracy. Their resolution was 600 dpi. Fig. 5 shows those images, D1∼D5. The number of the defined categories were $C = 52$ ("A"∼ "Z", "a"∼ "z",). It is noteworthy that the two documents D3 and D4 include mathematical expressions and thus include several undefined categories, such as italic fonts and mathematical symbols. The document D5, where characters were freely laid out, was prepared to emphasize the robustness of the proposed method to irregular layouts.

Each document image was rotated $\pm 30°, \pm 20°, \pm 10°, \pm 5°, \pm 2°, 0°$ and thus 55 test images were prepared in total. Fig. 6 shows several rotated images of D1. For every connected component in the document image, its category and skew angle were estimated and voted to determine the skew angle of the entire document image.

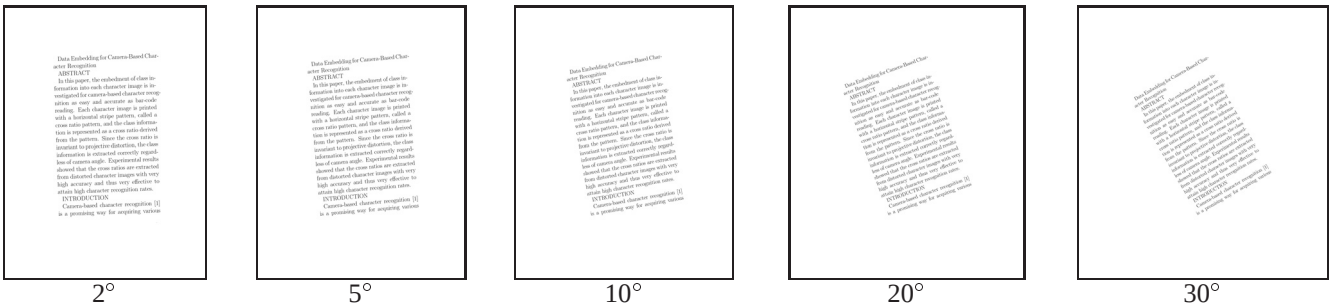
4.2. Preparing instances

For each of the 52 categories, the instances, i.e., the variant $p_c(\theta)$ and the invariant q_c, were measured by using the original font image of Times-Roman as $\boldsymbol{R}_c$ and stored. The resolution of the font image was 1440 dpi. Both the variant



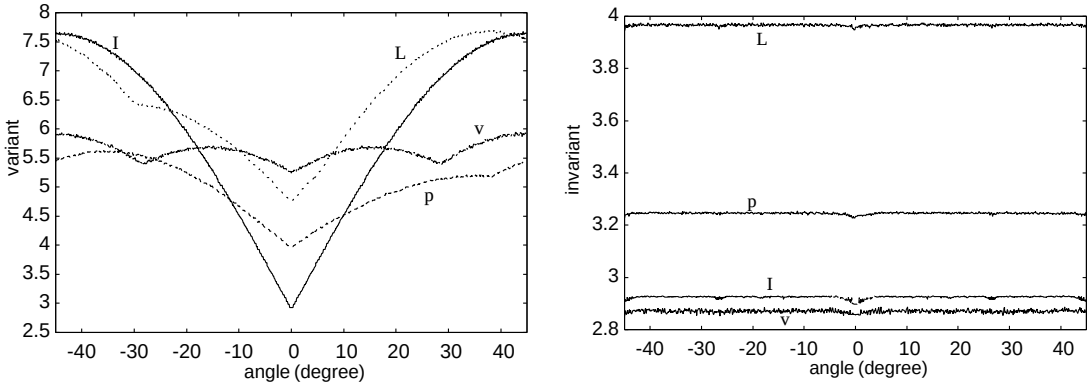
D1 D2 D3 D4 D5

Figure 5. Five document images used in the experiment. D3 and D4 include mathematical expressions.



2° 5° 10° 20° 30°

Figure 6. Skewed document images (D1).



(a) Variant (b) Invariant

Figure 7. Examples of variant and invariant.

and the invariant were measured by rotating $\boldsymbol{R}_c$ every $0.1°$ from $-45°$ to $45°$. As noted before, the italic fonts and the mathematical symbols included in D3 and D4 did not have their own instances.

Fig. 7 (a) shows the variants of several categories. The variants $p_c(\theta)$ of "I" and "L" change drastically according to θ whereas the variant of "v" only changes subtly. As noted in 3.2.3, the variants changing drastically are favorable for the reliable skew estimation. In the experiment, however, the variants changing subtly were also used for observing the basic performance of the proposed method.

Fig. 7 (b) shows the invariants of several categories. The invariants, in principle, will not change according to θ. This figure, however, reveals that the invariant fluctuates due to noise at image acquisition. In the experiment, the invariant value was averaged from $-45°$ to $45°$ and then stored as the instance.

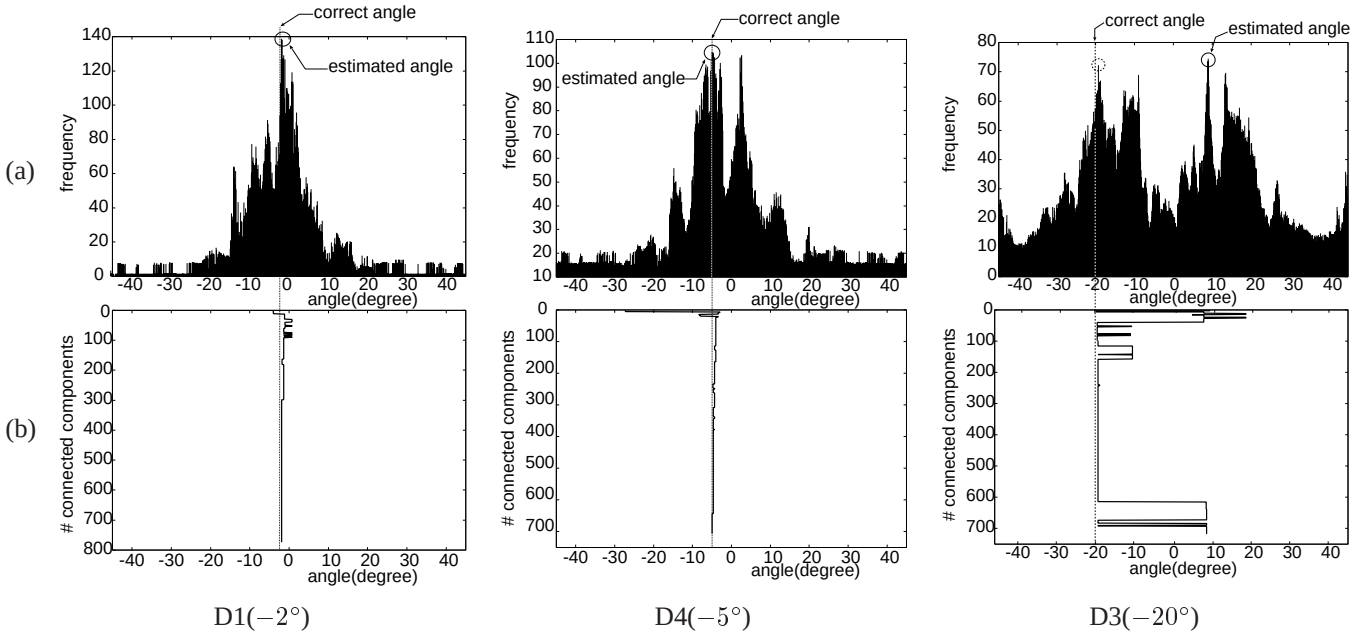


Figure 8. Estimation result. (a) Voting result and estimated skew angle as the peak of the voting result. (b) Change of estimation result according to the number of characters ($\sim$ connected components).

Table 1. Statistics of absolute errors of estimated skew angles.

$\leq 0.5°$	$\leq 1.0°$	$\leq 1.5°$	$\leq 2.0°$
20/55 (36%)	45/55 (82%)	49/55 (89%)	54/55 (98%)

Table 2. Absolute estimation error (degree) at each document image.

skew (°)	D1	D2	D3	D4	D5	average
-30	0.7	0.7	0.7	0.7	0.4	0.6
-20	0.7	0.7	29	0.7	0.4	6.3
-10	0.5	0.9	0.9	0.2	0.3	0.6
-5	0.7	0.7	1.5	0.0	0.2	0.6
-2	0.1	0.9	0.2	0.2	0.2	0.3
0	1.7	1.7	1.7	1.7	0.4	1.4
2	0.6	0.6	0.4	0.3	0.2	0.4
5	0.3	0.3	0.9	0.9	0.3	0.5
10	1.0	1.4	2.0	1.4	0.2	1.2
20	0.7	0.7	0.9	0.7	0.6	0.7
30	0.7	0.7	1.0	1.2	0.4	0.8
average	0.7	0.9	3.6	0.7	0.3	1.2

4.3. Accuracy of estimated skew angles

Table 1 summarizes the absolute errors of the skew angles estimated for the 55 test document images. For 98% (=54/55) of the test images, the absolute error was less than $2.0°$. Table 2 shows the absolute error for each of 55 test images. This table indicates that the estimation accuracy does not depend on the skew angles. The table also indicates that the estimation accuracy is not degraded by the existence of mathematical expressions, that is, the undefined categories. The skew of only one test image ("D3 rotated $-20°$") was poorly estimated. The reason of this failure will be discussed later.

It is noteworthy that the skew angles of D5 were also estimated successfully. This success emphasizes the usefulness of the proposed method since the conventional methods assuming straight text lines will fail to estimate the skew angles of D5.

Fig. 8 (a) shows the histogram of the skew angle candidates, i.e., the voting result. The first and the second histograms have their peaks at the correct skew angles. Consequently, the correct skew angles were estimated. The third histogram is of the failure result ("D3 rotated $-20°$"); the histogram has two rivaling peaks and the false peak won by a slight difference.

Fig. 8 (b) shows the change of the peak according to the increase of connected components. The first and the second

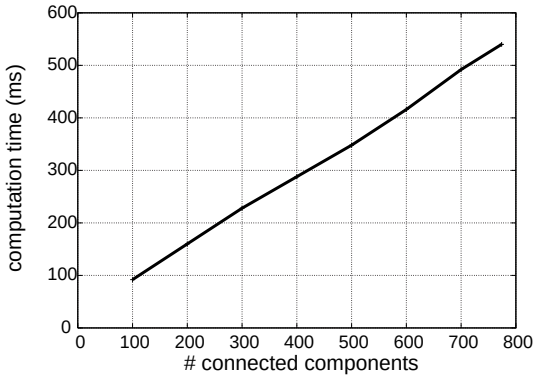


Figure 9. Computation time.

results, i.e., successful results, show quick convergence to the correct peak by $100 \sim 300$ connected components. The third result, i.e., the failure result, also shows the convergence to the correct peak by $200 \sim 600$ connected components; the peak, however, has moved to false one by 600 or more connected components.

4.4. Failure analysis

One of the main reasons of having the false peak in "D3 rotated $-20°$" was the error of the invariant. The invariants of several categories were very sensitive to the noise at the image acquisition. Especially, the invariant sometimes shows a drastic difference from the stored one by the disappearance of thin "serifs" by the low resolution. (As noted above, the resolution of 600 dpi was used on preparing the test images, whereas 1440 dpi was used on preparing the instances.) Since a false invariant always leads to false category estimation, it consequently leads to absolutely false skew estimation.

4.5. Computation time

Fig. 9 shows the computation time (CPU: Intel Pentium D) as a function of the number of connected components used for estimating the skew angle. From this graph, it is shown that the proposed method requires 100$\sim$ 200 ms for each document; this is because the proposed method could reach the correct skew angle with about 100$\sim$300 connected components (Fig. 8(b)).

5. Estimation of other deformations

Although we have focused on the skew estimation problem in this paper, the proposed method can estimate other geometric deformations by using different invariant and variant. For example, it is possible to estimate the shear deformation of objects which undergo affine transformation. In this case, for estimating the shear η, we will use an affine invariant q_c for estimating the category and an affine variant $p_c(\eta)$ which is a similarity invariant but a shear variant.

The estimation of perspective deformation, which is the most important deformation for camera-based OCR, can also be tackled by the proposed method. This can be realized by the fact that the perspective deformation can be decomposed into affine transformation (6 degrees of freedom) and the perspective component that controls the line at infinity (2 degrees of freedom) [7]. Let ϕ and ψ denote two parameters specifying the perspective component. In this case, we can estimate and compensate the perspective deformation according to the following steps:

1. Estimate the category c of each connected component $\boldsymbol{X}$ by using a perspective invariant q_c.
2. Estimate ϕ and ψ by using a perspective variant $p_c(\phi, \psi)$ which is an affine invariant and a variant to ϕ and ψ. The voting for this estimation is performed on the two-dimensional (ϕ, ψ)-plane.
3. Compensate the perspective component by using the estimated ϕ and ψ. The resulting document image will only undergo an affine transformation.
4. Estimate and compensate the shear η by the procedure described at the beginning of this section.
5. Finally, estimate and compensate the rotation θ by the procedure of Section 3.

Note that if we use another variant $p_c(\eta, \theta)$, the last two steps can be unified into one step.

6. Conclusion and Future Work

A novel technique for estimating document skew (i.e., rotation) has been proposed and its performance has been evaluated quantitatively and qualitatively via several experiments. The proposed method estimates the skew angle of each connected component very efficiently by using a rotation invariant and a rotation variant. The skew angles estimated at individual connected components are subjected to a voting procedure to find the most reliable skew angle as the entire document skew. The experimental result on 55 document images has shown that the skew angles of 54 document images were successfully estimated with errors smaller than 2.0°. The computational feasibility was also certified experimentally.

Future work will focus on the following points:

- Improvement of invariant. As analyzed in 4.4, the proposed method fails mainly due to the error of the invariant. Erroneous invariants always lead to false category candidates and thus lead to absolutely false skew

estimation. A more stable and distinguishable invariant will be necessary. A combinatorial use of multiple invariants is a possible remedy.

- Removal of less reliable instances. In this paper, the instance of the variant was prepared for every categories. Several instances, however, were less reliable as pointed in Section 3.2.3; for example, the variants of "o"-shaped characters do not change drastically by rotation and thus not express the skew angle clearly. In addition, the invariants of several categories are sensitive to noise. Removal of those invariants and variants will exclude false category candidates and skew angle candidates.

- Estimation of deformations other than rotation. As noted in Section 5, the proposed method can be extended to estimate various deformations by using suitable combinations of variants and invariants. Especially, perspective deformation will be the most important one for camera-based OCR.

References

[1] J. Liang, D. Doermann and H. Li: "Camera-based analysis of text and documents: a survey," Int. J. Doc. Anal. Recog., vol. 7, pp. 84–104, 2005.

[2] U. Pal, M. Mitra, B. B. Chaudhuri, "Multi-skew detection of Indian script documents," Proc. Int. Conf. Doc. Anal. Recog., pp. 292–296, 2001.

[3] Y. Ishitani, "Document Skew Detection Based on Local Region Complexity," Proc. Int. Conf. Doc. Anal. Recog., pp. 49–52, 1993.

[4] X. Jiang, H. Bunke, and D. Widmer-Kljajo, "Skew Detection of Document Images by Focused Nearest-Neighbor Clustering," Proc. Int. Conf. Doc. Anal. Recog., pp. 629–632, 1999.

[5] Y. Lu and C. L. Tan, "Improved Nearest Neighbor Based Approach to Accurate Document Skew Estimation," Proc. Int. Conf. Doc. Anal. Recog., pp. 503–507, 2003.

[6] S. Lu and C. L. Tan, "Camera Document Restoration for OCR," Proc. Int. Workshop Camera-Based Doc. Anal. Recog., pp. 17–24, 2005.

[7] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2nd edition, 2004.