Skew Estimation by Instances

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

This paper proposes a novel skew estimation method by instances. The instances to be learned (i.e., stored) are rotation invariants and a rotation variant for each character category. Using the instances, it is possible to estimate a skew angle of each individual character on a document. This fact implies that the proposed method can estimate the skew angle of a document where characters do not form long straight text lines. Thus, the proposed method will be applicable to various documents such as signboard images captured by a camera. Experimental evaluation using synthetic and real images revealed the expected robustness against various character layouts.

1. Introduction

Skew estimation is one of indispensable preprocessing modules in OCR. Most of conventional skew estimation techniques assume that text lines are long and straight. This assumption is valid for most paper document images acquired by, for example, flat-bed scanners and therefore the conventional techniques have performed well for those document images.

Recently, many researchers have tried to apply their OCR technologies to document images acquired by cameras [1]. One of many difficulties of this new trial is that the assumption of the straightness is not always valid. Text lines in scene (e.g., signboard and poster) are often very short or laid out irregularly and does not have the properties of long text lines.

An idea to relax the assumption for camera-based OCR is to estimate the skew angles at individual characters independently of each other. If the individual skew angle (hereafter called *local skew angle*) can be estimated, the skew angle of the entire document image (hereafter called *global skew angle*) will be estimated by voting those local skew angles and choosing the most frequent one. Clearly, the independent estimation does not require the assumption about long and straight text lines.

How can we estimate the local skew angle of a character? A possible choice is a matching-based method where font images (i.e., reference character images) are matched to the character under various rotation angles. The font image and the rotation angle which give the best match may be the category and the local skew angle of the character, respectively. This simple and naive method has good points; it can estimate the correct skew angles of documents with freely laid-out characters, scattered short text lines, and multi-columns. In addition, it is easy to deal with documents with mathematical notations and figures by rejecting characters without any good match.

This matching-based method can be seen as an instance-based method. That is, in the learning step, rotated font images are "learned" (i.e., stored) as instances. Then, in the estimation step, the local skew angle in the target document is estimated by matching it to the learned instances. Recently, various instance-based methods have been proposed in pattern recognition, computer vision, and image processing areas according to the increase of computational power. They often show better performance than conventional model-based methods, which may fail easily in exceptional cases that the model does not assume.

The above matching-based method, however, has a serious drawback; it requires huge computational complexity. In fact, we must repeat a costly image matching process for every category, every rotation angle, and every character in the document. Consequently, if there are C pre-defined categories and K quantized skew angles, the skew estimation for a document with N characters requires $O(C \cdot K \cdot N)$ image matchings. In addition, the method is excessive for the main aim of the skew estimation since it involves a matching-based character recognition process. That is, a

978-0-7695-3337-7/08 $25.00 © 2008 IEEE
DOI 10.1109/DAS.2008.22

character recognition result is obtained simultaneously with the skew estimation result.

This paper proposes a novel skew estimation method where the local skew angle is estimated in a more efficient way. Specifically, the proposed method uses *rotation invariants* and a *rotation variant* as instances. The local skew angle is estimated by referring those scalar instances instead of high-dimensional instances, i.e., font images. The proposed method requires only $O(N)$ (or less) computations, which is far less than $O(C \cdot K \cdot N)$ of the matching-based implementation. Of course, the proposed method inherits the good points of the matching-based method.

The performance of the proposed method is evaluated qualitatively and quantitatively via several experiments on synthetic document images and real document images. The evaluation will be done by observing the accuracy of the estimated skew and computation time.

2. Related Work

While most of conventional methods estimate the global skew angle of the document in one-step manner, several conventional methods estimate it in two-step manner. The latter methods first estimate the local skew angles and then estimate the global skew angle by choosing the most reliable local skew angle through voting, etc.

The most important merit of the latter methods is 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 deal with nonuniform skew and multiple-skew [2].

Ishitani [3] has estimated a local skew angle by searching for the direction which gives the most intensity (black/white) transitions. Among the estimated local skew angles, 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 neighboring connected components by a region growing technique and then its skew angle is estimated as a local skew angle. 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 still 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.

This paper largely expands a preliminary trial by the authors [7] at various points for improving the skew estimation performance. For example, new invariants are introduced

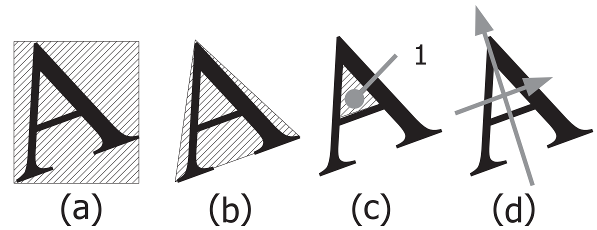


Figure 1. (a) Bounding box, (b) convex hull, (c) hole, and (d) two principal axes.

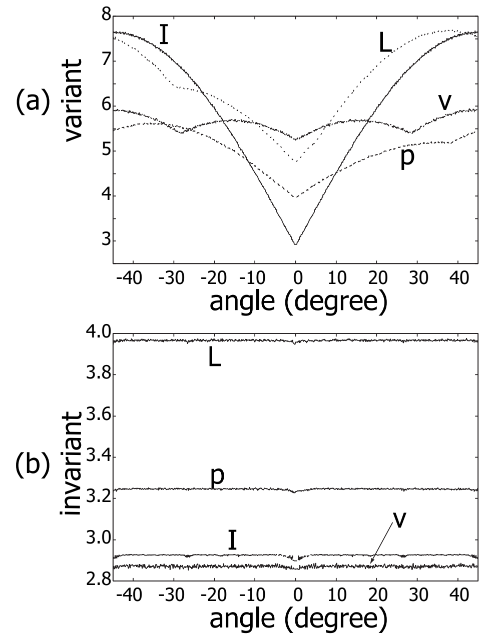


Figure 2. (a) Variant $p_c(\theta)$ and (b) invariant $q_{c,\mathrm{hull}}$ of several categories.

for expecting their complementary characteristics. In addition, experimental evaluation is far more detailed and extended to observe the estimation performance on real document images. (In fact, in [7], only synthetic images were used for the evaluation.)

3. Rotation Variant and Invariants

In advance to the discussion about the proposed skew estimation method, a rotation variant $p_c(\theta)$ and three rotation invariants $q_{c,\mathrm{hull}}$, $q_{c,\mathrm{hole}}$, $q_{c,\mathrm{ratio}}$ are introduced in this section, where $c \in [1, \ldots, C]$ denotes the category of characters and θ denotes the skew angle.

Note that the proposed method can employ other variants and/or invariants. In addition, the proposed method can employ only one or two among the three invariants. We, however, insist on employing the three invariants, since they have complementary characteristics as shown later.

For each category $c \in [1, \ldots, C]$,

1. Prepare the font image $\boldsymbol{R}_c$.
2. Measure the rotation invariants $\boldsymbol{q}_c = (q_{c,\mathrm{hull}}, q_{c,\mathrm{hole}}, q_{c,\mathrm{ratio}})$ using $\boldsymbol{R}_c$.
3. Measure the rotation variant $p_c(\theta)$ at all possible θ while rotating $\boldsymbol{R}_c$ with θ.

Figure 3. Instance learning for skew estimation.

3.1. Rotation variant

The rotation variant $p_c(\theta)$ is defined as

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

Fig. 1 (a) shows the bounding box of Times-Roman "A." Since the area of the bounding box depends on the rotation angle θ, $p_c(\theta)$ is a rotation variant. Fig. 2 (a) shows the variant $p_c(\theta)$ of several categories. The variants of "I" and "L" change drastically according to θ whereas the variant of "v" only changes subtly.

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

3.2. Rotation invariants

Three rotation invariants are defined as

$$q_{c,\mathrm{hull}} = \frac{\text{area of convex hull of } \boldsymbol{R}_c}{\text{area of black pixels of } \boldsymbol{R}_c}, \tag{2}$$

$$q_{c,\mathrm{hole}} = \text{the number of holes in } \boldsymbol{R}_c, \tag{3}$$

$$q_{c,\mathrm{ratio}} = \text{the ratio of two principal axes of } \boldsymbol{R}_c. \tag{4}$$

Figs. 1 (b), (c), and (d) illustrate the convex hull, the hole, and the principal axes of Times-Roman "A." Since those invariants depend on the character shape and not on the character pose, they do not change with the skew angle. Hereafter, $\boldsymbol{q}_c$ denotes the invariant vector $(q_{c,\mathrm{hull}}, q_{c,\mathrm{hole}}, q_{c,\mathrm{ratio}})$. Fig. 2 (b) shows the invariant $q_{c,\mathrm{hull}}$ of several categories. Although the graph shows small perturbation due to the noise at image acquisition, it is almost constant against the change of θ.

4. Skew Estimation by Instances

4.1. Learning Step

The proposed method utilizes the rotation variant and the invariants as instances for efficient local skew estimation. The procedure of the instance learning is illustrated in Fig. 3. Here, the characters printed by different fonts or styles belong to different categories. For example, 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", the number of categories is $C = 52 \times 3 \times 2 = 312$.

On the measurement of the invariant (Step 2 of Fig. 3), its value was first measured while rotating $\boldsymbol{R}_c$ from $-45°$ to $45°$ and then determined as the average value. This averaging is necessary because the invariant will fluctuate due to noise at image acquisition as shown in Fig. 2 (b) and therefore needs to be averaged to have its unique and representative value.

On the measurement of the variant (Step 3 of Fig. 3), the skew angle θ must be discretized in practice. Discretization interval (equivalently, the number of discretization levels) is crucial and defines the accuracy of the skew angle In this paper, the angle θ is discretized with $0.1°$ interval. Consequently, the variant $p_c(\theta)$ will be measured at 900 angles within $[-45°, 45°]$.

In practice, it is not necessary to prepare instances for all possible categories. That is, even if the target document includes several rare characters whose instances are not learned, the proposed method can estimate correct skew angle. As discussed later, this robustness comes from the voting strategy for determining the skew angle of the entire document. In Section 5, the robustness will be experimentally shown through the skew estimation result of a mathematical document which includes several undefined characters (i.e., mathematical symbols).

4.2. Estimation step

Fig. 4 illustrates the skew estimation step of the proposed method. There are three main modules in the estimation step, that is, (i) category estimation of each connected component, (ii) estimation of the local skew angle, and (iii) estimation of the global skew angle. Note that since the skew estimation will be done at the preprocessing stage, the target document image is not separated into characters yet. Thus, those modules are performed at each connected component instead of each character.

4.2.1 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

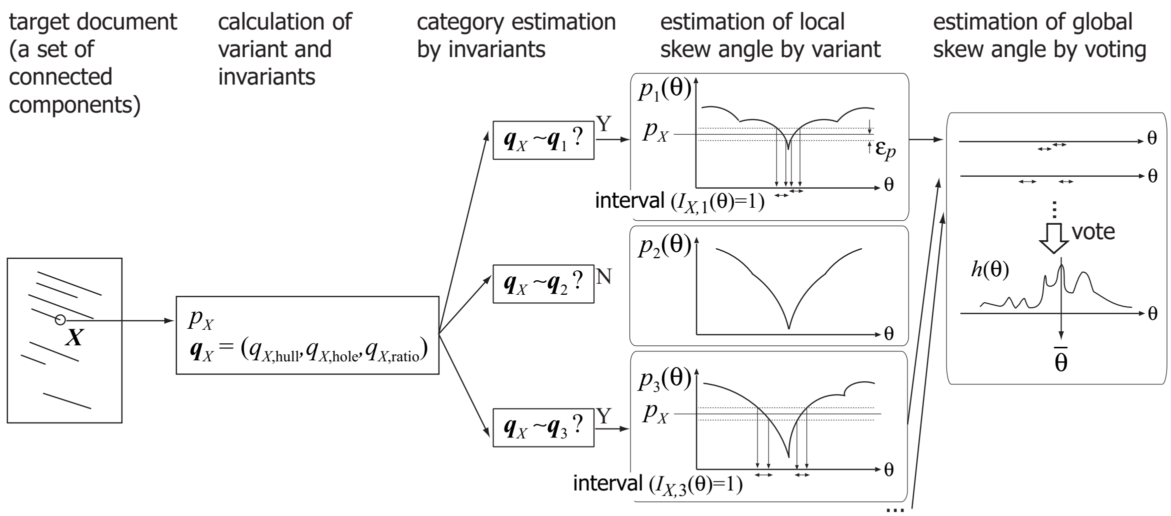


Figure 4. Overview of the proposed skew estimation method.

by comparing its invariant vector $\boldsymbol{q}_X$ to the stored instances $\{\boldsymbol{q}_c | c = 1, \ldots, C\}$. If $\boldsymbol{q}_X \sim \boldsymbol{q}_c$, c is a category candidate of $\boldsymbol{X}$. Specifically, if $\boldsymbol{q}_X$ satisfies the following conditions, c is a category candidate:

$$\left.\begin{array}{l} |q_{c,\text{hull}} - q_{X,\text{hull}}| \le \epsilon_{q_\text{hull}} \\ q_{c,\text{hole}} = q_{X,\text{hole}} \\ |q_{c,\text{ratio}} - q_{X,\text{ratio}}| \le \epsilon_{q_\text{ratio}} \end{array}\right\}, \tag{5}$$

where ϵ_{q_hull} and $\epsilon_{q_\text{ratio}}$ are thresholds determined experimentally.

Theoretically, the invariant $\boldsymbol{q}_c$ of the correct category satisfies (5) regardless of skew of $\boldsymbol{X}$, since $\boldsymbol{q}_c$ is the rotation invariant. In addition, $\boldsymbol{q}_c$ of the correct category will satisfy (5) regardless of the character size and position, since $\boldsymbol{q}_c$ is also a scale and shift invariant.

In practice, it is probable that multiple categories satisfy the condition (5) for a single connected component $\boldsymbol{X}$. This is because the invariants are not discriminative enough to determine the category of $\boldsymbol{X}$ uniquely. In other words, it is hard to "recognize" $\boldsymbol{X}$ by only using the feature $\boldsymbol{q}_x$. **Fig. 4** illustrates the selection of two categories c_1 and c_3 as category candidates under the condition that $\boldsymbol{q}_X \sim \boldsymbol{q}_1$, $\boldsymbol{q}_X \not\sim \boldsymbol{q}_2$, and $\boldsymbol{q}_X \sim \boldsymbol{q}_3$.

In practice, it is also probable that no category satisfies the condition (5) when $\boldsymbol{X}$ is distorted by some noise or $\boldsymbol{X}$ belongs to an undefined category. In this case, the connected component $\boldsymbol{X}$ is rejected at this step. (That is, $\boldsymbol{X}$ is not used for estimating local and global skew angle.)

4.2.2 Local skew estimation by variant

The local skew angle of $\boldsymbol{X}$ is estimated by using the stored instance $p_c(\theta)$ of each category candidate c. The variant value p_X is first calculated from $\boldsymbol{X}$ and then compared with the instance $p_c(\theta)$. If $p_c(\theta) \sim p_X$ at an angle θ, θ is a local skew angle candidate of $\boldsymbol{X}$. Specifically, θ is a candidate if it satisfies the condition:

$$|p_X - p_c(\theta)| \le \epsilon_p, \tag{6}$$

where ϵ_p is a threshold.

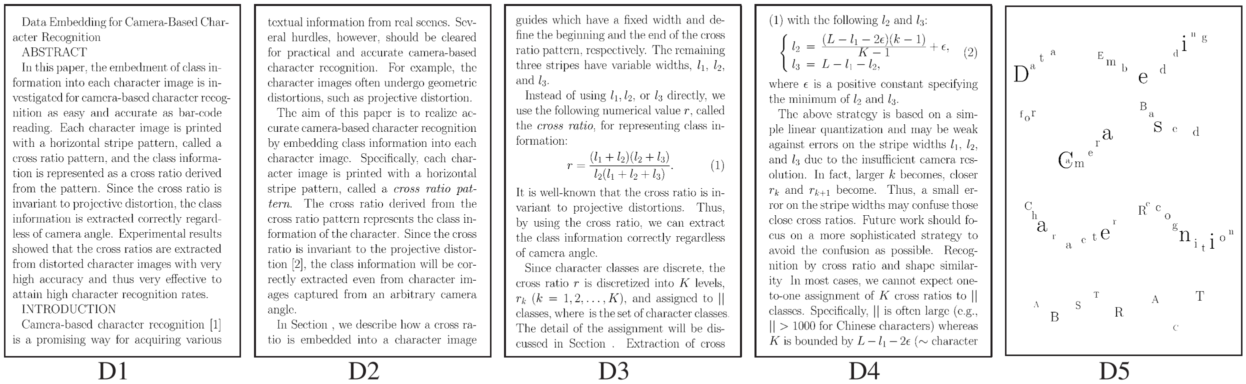
Since the variant $p_c(\theta)$ changes smoothly with θ, the set of θs which satisfy (6) form one or more "intervals" as illustrated in Fig. 4. Each interval is represented as a continuous range of θ where the following function becomes 1:

$$I_{X,c}(\theta) = \begin{cases} 1, & \text{if } |p_X - p_c(\theta)| \le \epsilon_p, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

It is interesting to note that each skew angle candidate has a different reliability. Consider an "I"-shaped character. Its variant $p_c(\theta)$ will change drastically with θ (as shown in Fig. 2 (a)) and therefore the interval becomes short. Consequently, the skew angle candidate from the character is less ambiguous and has a high reliability. In contrast, consider an "o"-shaped character. Its variant will change subtly (like the variant of "v" in **Fig. 2** (a)) and therefore the interval becomes longer. Consequently, the skew angle candidate is very ambiguous and has a low reliability.

4.2.3 Global skew estimation by voting

A voting strategy is employed for estimating the global skew angle. 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 is min-



Data Embedding for Camera-Based Character Recognition
ABSTRACT
In this paper, the embedment of class information into each character image is investigated for camera-based character recognition as easy and accurate as bar-code reading. Each character image is printed with a horizontal stripe pattern, called a cross ratio pattern, and the class information is represented as a cross ratio derived from the pattern. Since the cross ratio is invariant to projective distortion, the class information is extracted correctly regardless of camera angle. Experimental results showed that the cross ratios are extracted from distorted character images with very high accuracy and thus very effective to attain high character recognition rates.
INTRODUCTION
Camera-based character recognition [1] is a promising way for acquiring various

textual information from real scenes. Several hurdles, however, should be cleared for practical and accurate camera-based character recognition. For example, the character images often undergo geometric distortions, such as projective distortion.
The aim of this paper is to realize accurate camera-based character recognition by embedding class information into each character image. Specifically, each character image is printed with a horizontal stripe pattern, called a *cross ratio pattern*. The cross ratio derived from the cross ratio pattern represents the class information of the character. Since the cross ratio is invariant to the projective distortion [2], the class information will be correctly extracted even from character images captured from an arbitrary camera angle.
In Section , we describe how a cross ratio is embedded into a character image

guides which have a fixed width and define the beginning and the end of the cross ratio pattern, respectively. The remaining three stripes have variable widths, l_1, l_2, and l_3.
Instead of using l_1, l_2, or l_3 directly, we use the following numerical value r, called the *cross ratio*, for representing class information:
$$r = \frac{(l_1+l_2)(l_2+l_3)}{l_2(l_1+l_2+l_3)}. \quad (1)$$
It is well-known that the cross ratio is invariant to projective distortions. Thus, by using the cross ratio, we can extract the class information correctly regardless of camera angle.
Since character classes are discrete, the cross ratio r is discretized into K levels, r_k ($k = 1, 2, \ldots, K$), and assigned to || classes, where is the set of character classes. The detail of the assignment will be discussed in Section . Extraction of cross

(1) with the following l_2 and l_3:
$$\begin{cases} l_2 = \dfrac{(L - l_1 - 2\epsilon)(k-1)}{K-1} + \epsilon, \\ l_3 = L - l_1 - l_2, \end{cases} \quad (2)$$
where ϵ is a positive constant specifying the minimum of l_2 and l_3.
The above strategy is based on a simple linear quantization and may be weak against errors on the stripe widths l_1, l_2, and l_3 due to the insufficient camera resolution. In fact, larger k becomes, closer r_k and r_{k+1} become. Thus, a small error on the stripe widths may confuse those close cross ratios. Future work should focus on a more sophisticated strategy to avoid the confusion as possible. Recognition by cross ratio and shape similarity In most cases, we cannot expect one-to-one assignment of K cross ratios to || classes. Specifically, || is often large (e.g., || > 1000 for Chinese characters) whereas K is bounded by $L - l_1 - 2\epsilon$ (∼ character

Figure 5. Five original document images for the experiment. D3 and D4 include mathematical expressions.

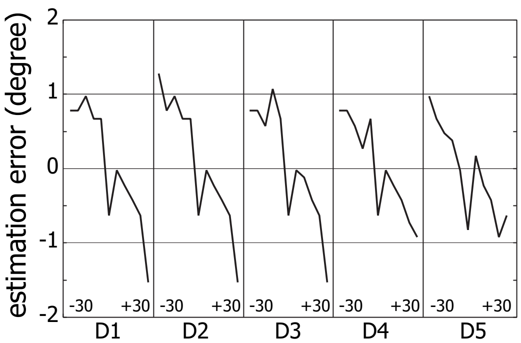


Figure 6. Estimation errors.

imized by voting far more candidates representing the correct skew angle.

The global skew angle $\bar{\theta}$ is finally estimated as the peak of the histogram $h(\theta)$ which is created by voting the skew angle candidates. Using $I_{X,c}(\theta)$, the histogram $h(\theta)$ is defined as

$$h(\theta) = \sum_X \sum_c I_{X,c}(\theta), \quad (8)$$

and $\bar{\theta} = \mathrm{argmax}_\theta h(\theta)$.

4.2.4 Computational feasibility

The proposed method has a strong computational feasibility. In addition to the use of compact scalar instances (i.e., the invariant $\boldsymbol{q}_c$ and the variant $p_c(\theta)$), this strength comes from by the use of look-up tables on referring to the instances. That is, the category candidate c which satisfies the conditions (5) can be found efficiently by a look-up table which works as an inverse mapping from the invariant value to categories. Similarly, the interval of skew angle candidates which satisfy $I_{X,c}(\theta) = 1$ can be found efficiently by a look-up table which works as an inverse mapping from the variant to the interval.

Since the computational cost required by those look-up operations is $O(1)$, each local skew angle is estimated with $O(1)$ computation. Consequently, the global skew angle is estimated with $O(1) \times N = O(N)$ computations.

The computational feasibility of the proposed method may be further improved by limiting the number of connected components to be voted. In fact, it is not necessary to use all N connected components in the document as experimentally shown in 5.4. This is because all the connected components, in principle, will have the same skew angle candidate and thus the voting result will show the peak at the correct skew angle even with a small number of votes.

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.

5. Experimental Results

5.1. Document image samples

Several experiments for evaluating the proposed method were conducted with five documents, D1∼D5. All the documents were created by LaTeX with a single font and style (Times-Roman, upright). Fig. 5 shows those documents. Two documents D3 and D4 include mathematical expressions and thus include italic fonts and mathematical symbols. The document D5, where characters were freely laid

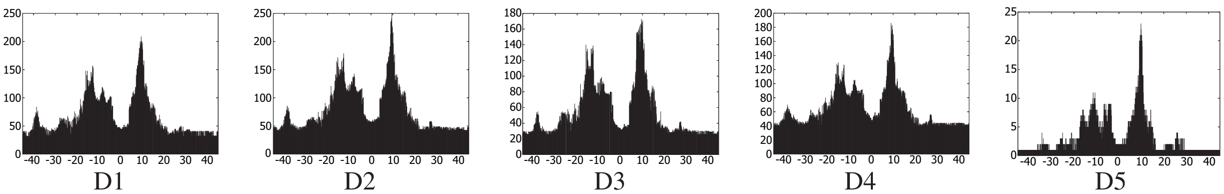


Figure 7. Histograms $h(\theta)$ for D1-D5 with 10° skew.

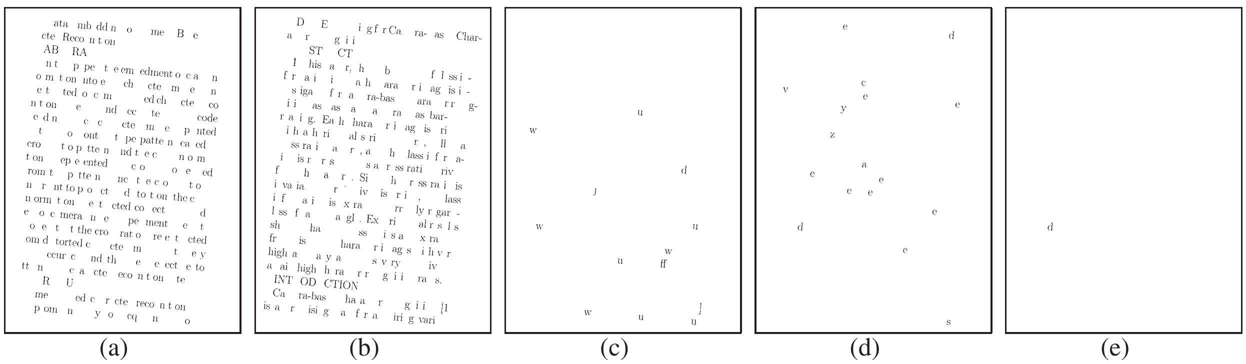


Figure 8. Local skew estimation result of D1 (-5°) with three invariants. See text for details.

out, was prepared to emphasize the robustness of the proposed method to irregular layouts.

Two skew document image datasets were made from those documents.

Dataset1 was prepared for evaluating the accuracy and computation time quantitatively. It contains synthetic skew document images which were created in a PC by applying 11 artificial skews to each of the five original digital document image. The angles of the rotations were $\pm 30^\circ, \pm 20^\circ, \pm 10^\circ, \pm 5^\circ, \pm 2^\circ, 0^\circ$. Consequently, 55 images (600 dpi) were prepared in total.

Dataset2 was prepared for observing the performance on practical situations. It contains real skew document images acquired by scanning papers where the original document images were printed. When scanning, two rotations (about ± 10) were applied to each document. Consequently, 10 images (600 dpi) were prepared in total.

5.2. Instances

For each pre-defined category, the variant $p_c(\theta)$ and the invariants $\boldsymbol{q}_c$ were measured by using the original image of the Times-Roman font and then stored as instances. The resolution of the font image was 1440 dpi. The number of the pre-defined categories was $C = 52$ (“A”∼ “Z” and “a”∼ “z”,). Both the variant and the invariants were measured by rotating $\boldsymbol{R}_c$ every 0.1° from -45° to 45°. Fig. 2 shows the variants and the invariants of several categories prepared according to this procedure. Note that the italic fonts and the mathematical symbols in D3 and D4 did not have their own instances.

5.3. Accuracy evaluation using synthetic images

Fig. 6 shows the estimation errors by the proposed method for the 55 test document images (5 documents × 11 angles) of Dataset1. For all the test images, the absolute errors were always less than 2.0°. That is, the document image deskewed by the proposed method will have the skew angle less than 2.0°. In addition, except for several images with large skew angles ($\sim \pm 30$), the absolute errors were less than 1.0°. Note that a commercial OCR could recognize the deskewed images of D1 and D2 (i.e., regular documents without mathematical expressions) with 100% accuracy.

As shown in Fig. 6, the skew angles of D5 were also estimated very 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. 7 shows the histogram $h(\theta)$ for D1∼D5 with 10° skew. In every histogram, it is possible to observe a large peak around 10°, i.e., the correct skew angle. It is noteworthy that there is another peak around -10°. This phenomenon is caused by the symmetric property of the variant $p_c(\theta)$. As shown in Fig. 2 (a), the variant tends to be symmetric (i.e., $p_c(\theta) \sim p_c(-\theta)$) because the bounding boxes

Table 1. Distribution of absolute estimation errors.

	Dataset1 55 images				Dataset2 10 images
	<0.5°	<1.0°	<1.5°	<2.0°	success
$q_{c,\text{hull}}$	20	45	49	54	0
$q_{c,\text{hole}}$	7	13	17	25	0
$q_{c,\text{ratio}}$	3	11	30	45	0
$q_{c,\text{hull}}+q_{c,\text{hole}}$	20	40	46	55	0
$q_{c,\text{hull}}+q_{c,\text{ratio}}$	20	48	49	55	7
$q_{c,\text{hole}}+q_{c,\text{ratio}}$	6	11	27	55	0
all	28	50	50	55	9

at θ and $-\theta$ have similar sizes for symmetric characters such as "I" and "v." Thus, if a certain angle θ is voted from a connected component, $-\theta$ is also voted and finally a false peak is formed.

Fig. 8 shows the local skew estimation result for D1 ($-5°$). Each connected component of D1 is plotted in one of Figs. 8 (a), (b),. . .,(e) according to their individual results of local skew angle estimation:

- (a) is the good case where the correct category was selected as a category candidate and then the correct angle was also voted as a skew angle candidate.
- (b) is the rejected case where no category candidate was found.
- (c) is the worst case where the correct category was not selected due to some noise on $\boldsymbol{q}_X$ and therefore the correct angle was also not voted.
- (d) is the unfortunate case where the correct category was selected but the correct angle was not voted due to some noise on the variant p_X.
- (e) is the strange case where the correct category was not selected but the correct angle was voted "by accident" from a wrong category candidate.

As shown in Fig. 8, most connected components were plotted in (a) and (b). They helped to provide the correct skew estimation or did not disturb the estimation. Few connected components were plotted in the wrong cases (c)-(e). Thus, those figures prove that the proposed method performed reasonable.

A careful inspection of Fig. 8 reveals that several categories (e.g., "a") were often rejected (i.e., plotted in (b)). This is because an invariant of some categories is very sensitive to the distortion by rotation. Another inspection reveals that most "o"-shaped characters (e.g., "o", "e"., "c")

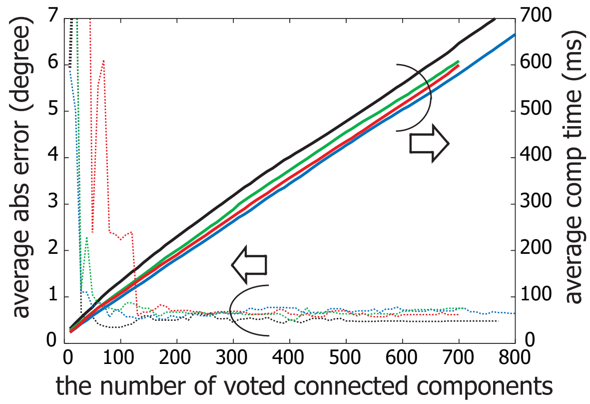


Figure 9. Average computation time and average absolute error for each of D1$\sim$ D4 as functions of the number of voted connected components.

were plotted in (a). This fact does *not* prove that the "o"-shaped characters are useful for the estimation. As noted in 4.2.2, the interval (i.e., the range of θs) voted by the "o"-shaped characters tends to be long because of slow change of $p_c(\theta)$ with θ. Thus the interval may include not only the correct angle but also many incorrect angles and therefore those characters are not very useful.

Table 1 shows the distribution of absolute estimation errors of 55 test images under different usages of the invariants, $q_{c,\text{hull}}$, $q_{c,\text{hole}}$, and $q_{c,\text{ratio}}$. For example, the estimation error was smaller than $1.0°$ at 45 images among 55 test images when the single invariant $q_{c,\text{hull}}$ was used. This table indicates that the invariants must be combined for minimizing estimation error. The reason of this "synergy" will be discussed together with another experimental result in 5.5.

5.4. Computation time

Fig. 9 plots the computation time (solid line) on a PC with a CPU of Intel Pentium D and the absolute estimation error (dashed line) as functions of the number of connected components used. angle. Each line plots the average of 11 images of a document.

This graph clearly shows the computational feasibility of the proposed method. That is, the proposed method could minimize the estimation error by using only $100 \sim 200$ connected components and thus requires 100$\sim$ 200 ms for each document image.

5.5. Accuracy evaluation using real images

Another experiment was performed on Dataset2. Since the original skew angle was unknown for the test images of

Dataset2, the evaluation has been made by visual observation. Among 10 test images, 9 images could have successful results when the three invariants ($q_{c,\mathrm{hull}}$, $q_{c,\mathrm{hole}}$, and $q_{c,\mathrm{ratio}}$) were used. Consequently, it is indicated that the proposed method is applicable to real document images, while there is still large room to improve estimation performance by, for example, employing new invariants and multiple variants.

Table 1 summarizes the results on Dataset2 when one or two or three invariants were used. Surprisingly, there was no test image whose skew was estimated correctly by a single invariant. The difference between the three-invariants case and the single-invariant case clearly reveals the effect of complementary characteristics of the invariants. The invariant $q_{c,\mathrm{hull}}$ is discriminative but unstable against even small distortions. (The shape of convex hull and/or the thickness of character strokes change easily according to image acquisition condition. In contrast, the invariants $q_{c,\mathrm{hole}}$ and $q_{c,\mathrm{ratio}}$ are less discriminative but stable against the distortions. Their combination could reject many less reliable connected components successfully and provide better estimation accuracy.

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 angles of individual connected components very efficiently by using three rotation invariants and a rotation variant. The skew angles are then 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 estimation error was less than $2.0°$ at all the images and $1.0°$ at 50 images. The computational feasibility was also certified experimentally.

Future work will focus on the following points:

- Incorporating other variants and invariants. In order to improve estimation accuracy, it will be effective to use other variants and invariants. Especially, use of multiple variants must be examined to increase the number of reliable votes.

- Document image with nonuniform skew. The proposed method is based on the estimation of local skew angles of a document. This fact implies that the proposed method has a potential to deskew a document with nonuniform skew.

- Removal of less reliable instances. In this paper, the instance of the variant was prepared for every categories. Several instances, however, were less reliable; 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 suppress false category candidates and false skew angle candidates.

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

- Various documents. Since the proposed method is based on voting, it will be robust to interference by figures, tables, and pictorial decorations, which often disturb skew estimation by conventional techniques. This robustness has already been indicated through the skew estimation results of documents with mathematical notations.

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] S. Uchida, M. Sakai, M. Iwamura, S. Omachi, K. Kise, "Instance-based skew estimation of document images by a combination of variant and invariant," Proc. Int. Workshop on Camera-Based Doc. Anal. Recog., pp.53-60, 2007.