Skew Estimation of Document Images Using Bagging

Gaofeng Meng, Chunhong Pan, Nanning Zheng, *Fellow, IEEE*, and Chen Sun

***Abstract*—This paper proposes a general-purpose method for estimating the skew angles of document images. Rather than to derive a skew angle merely from text lines, the proposed method exploits various types of visual cues of image skew available in local image regions. The visual cues are extracted by Radon transform and then outliers of them are iteratively rejected through a floating cascade. A bagging (*bootstrap aggregating*) estimator is finally employed to combine the estimations on the local image blocks. Our experimental results show significant improvements against the state-of-the-art methods, in terms of execution speed and estimation accuracy, as well as the robustness to short and sparse text lines, multiple different skews and the presence of non-textual objects of various types and quantities.**

***Index Terms*—Bagging estimator, document image analysis, floating cascade, Radon transform, skew estimation.**

I. Introduction

NOWADAYS, the scanner is becoming a common device in modern office for converting paper document into its digital format. Document skew commonly occurs during the scanning process when paper documents are not fed properly into the scanner. The existence of such artifacts will cause some significant problems in the subsequent steps, such as page layout analysis and character recognition. Skew estimation and correction, thus, becomes an important issue in the field of document image analysis and understanding [1], [2].

Many methods have been developed to address the problem in the past decades [3], [4]. The basic principle is to derive a skew angle by exploring certain visual cues. Among these visual cues, straight text lines are extensively studied. Many methods are, thus, proposed — including the projection profiles based techniques [5]–[9], the transition counts based techniques [10], [11], the cross-correlation based techniques [12], [13], the Hough transform based techniques [14]–[18] and the nearest neighbor clustering based techniques [19]–[22].

The projection profiles based method [5]–[9] is one of the oldest but most effective approaches. Typically, the method first calculates the projection profiles of a document image at various angles. Each projection profile is then evaluated by a criterion function. The function should reach its maximum when the projection profile is calculated at the correct skew angle. Therefore, one can immediately determine the skew angle by examining the maximum of the function. The projection profiles based method works efficiently for various document images. However, its performance often deteriorates with short and sparse text lines, the presence of nontextual objects and image noises caused by imperfect scanning.

Instead of accumulating the pixel values in an 1-D array, the transition counts based method [10], [11] counts the numbers of transitions between foreground and background pixels and accumulates them in a histogram. Maximum variance of transition counts is often taken as the criterion function for determining the skew angle [11].

The basic idea of cross-correlation based method comes from an observation that the correlation between any pair of vertical lines at a fixed distance is generally maximized, if the shift distance along the vertical direction meets the intercept incurred by the skewed horizontal text lines [12], [13]. The method has a low computational complexity, but is sensitive to document layouts and the presence of nontextual objects.

Hough transform is another widely used technique for skew detection [14]–[16], [18]. It first transforms an image into its parameter space and then searches for the local maxima of the space to estimate the skew angle. Hough transform can not be directly implemented in the real-scanned document images due to its great memory demanding and high time complexity. Many efforts are, thus, made to reduce the computations by minimizing the number of transformations performed [16], or reducing the amount of input data by using a simplified representation of image, such as block adjacent graph (BAG) [15], [17], [18].

The nearest neighbors clustering based method [19]–[22] first extracts connected components from a document image. For each component, the direction of its nearest neighbor is estimated and accumulated in a histogram. The angle that corresponds to the maximum peak of the histogram is taken as the dominant image skew [19]. The nearest neighbors clustering based method is generally slow since the components labeling procedure has a quadratic time complexity $O(n^2)$, where n is the number of components in a document image. Furthermore, to eliminate the false maxima in the histogram, small or large nontextual components are necessary to be filtered in a preprocessing step [20], [21]. This may also increase the computational cost.

Manuscript received August 30, 2009; revised February 09, 2010. First published March 15, 2010; current version published June 16, 2010. This work was supported in part by the National Science Foundation of China under grant No. 60675012, 60635050, and in part by the National "863" High-Tech Program of China under grant No. 2009AA012104. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Maya R. Gupta.

G. Meng is with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, and also with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: gfmeng@nlpr.ia.ac.cn).

C. Pan is with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190 China (e-mail: chpan@nlpr.ia.ac.cn).

N. Zheng and C. Sun are with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049 China (e-mail: nnzheng@mail.xjtu.edu.cn; sunnie.sunc@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2010.2045677

In general, the straight text lines based methods work well for most document images. However, these methods are often severely affected by the presence of scan-introduced distortions and large areas of nontextual objects. Recently, many other image visual cues, such as straight lines of tables, border lines of textual blocks, edges of inserted graphics and straight strokes of large characters, have been exploited as well in the literature. Liu *et al.* [23] proposed a method for skew estimation by using the border lines extracted from text or nontext regions. Chou *et al.* [24] proposed a method based upon the piecewise covering of objects by parallelograms. Yuan *et al.* [25] made use of the border lines of convex hull of image components and their groups for skew estimation. These methods are capable of tolerating the presence of nontextual objects in the image, like forms, tables, pictures and scanning noises.

Although skew estimation has been studied for decades, finding faster and more robust algorithms remains an interesting topic. The difficulties are multifold. A robust method has to be capable of coping with document images with complicated layouts, low image resolutions, significant amounts of scanning artifacts, sparse and short text line, multiple skews due to page distortion, and nontextual objects of various types and quantities. Additionally, some practical document images may consist of complex backgrounds or inaccurate visual cues of skew, e.g., handwritten document images.

To address these difficulties, one way is to use multiple visual cues of image skew. This is often beneficial especially when one kind of visual cues is absent in the image and the others are plentiful. Following this idea, we propose a general-purpose method in this paper for skew estimation by integrating multiple local visual cues of image skew using a bagging estimator.

The proposed method is full-ranged ($\pm 90^\circ$), extremely robust to excessive noises and the presence of large areas of nontextual objects, and highly competitive in estimation accuracy and execution speed. Moreover, our method does not require additional preprocessing steps to remove image noises, to extract textual regions or to segment characters. These remain potential problems in the field of document image analysis. The proposed method is applicable to a wide variety of document images, including textual documents, line-drawing documents, handwritten documents and document images with sparse and short text lines, scan-introduced distortions and large areas of nontextual objects of various types and quantities.

II. Basic Principle

Let x be an $L \times L$ image block that is randomly drawn from a skewed document image D. Denote θ_t the skew angle of D and $\phi(x)$ a skew estimator on the image block x. The skew estimation on image blocks using bagging (***b**ootstrap **agg**regat **ing***) is the average over x of $\phi(x)$ [26], i.e.,

$$\theta_A = E_x\phi(x) = \sum_x \phi(x) Pr(x|D) \tag{1}$$

where $Pr(x|D)$ is the underlying sampling probability function of x on D.

Bagging is an effective way to improve the accuracy of estimation [27]. Breiman pointed out that the instability of the estimation method is the vital element of bagging [26], i.e., if perturbing x can cause significant changes in $\phi(x)$, then bagging can improve accuracy.

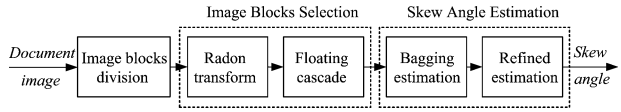


Fig. 1. Overview of the proposed method. It consists of three modules: image blocks division, image blocks selection, and skew angle estimation.

Considering the mean-squared estimation error of $\phi(x)$ on x, we have

$$E_x(\theta_t - \phi(x))^2 = \theta_t^2 - 2\theta_t E_x\phi(x) + E_x\phi(x)^2. \tag{2}$$

Using $E_x\phi(x) = \theta_A$ and applying the inequality $EZ^2 \geq (EZ)^2$ to the third term in (2) gives

$$E_x(\theta_t - \phi(x))^2 \geq (\theta_t - \theta_A)^2. \tag{3}$$

One can see from the above inequality that the squared error of bagging estimation is always lower than the mean-squared error averaged over x of $\phi(x)$.

Generally speaking, the accuracy of skew estimation on an image block depends upon two factors, namely, the block skew estimator $\phi(\cdot)$ and the image block x. Employing a robust skew estimator is essential for an accurate skew estimation. On the other hand, taking an informative image block with sufficient visual cues of image skew, say, a textual block, is also critical for a highly accurate skew estimation.

Notice that the inequality in (3) always holds. This motivates us to further enhance the bagging estimation in an indirect way. Specifically, rather than to directly decrease $(\theta_t - \theta_A)^2$, we endeavor to reduce its upper bound, i.e.,

$$E_x(\theta_t - \phi(x))^2 = \sum_x (\theta_t - \phi(x))^2 Pr(x|D). \tag{4}$$

Intuitively, by minimizing its direct upper bound, the error of bagging estimation can be depressed accordingly, thus, giving an improvement on the estimation.

Basically, if an image block contains sufficient informative cues and the block skew estimator $\phi(\cdot)$ is accurate, the squared estimation error on x, i.e., $(\theta_t - \phi(x))^2$, will be quite small. As a result, to minimize (4), we have at least two tricks: to employ a robust block skew estimator $\phi(\cdot)$ and/or to take an optimal block sampling function $Pr(x|D)$, which draws the informative image blocks from a document image D with high probabilities.

In the proposed method, Radon transform based skew estimator is used. The estimator is robust and adaptive to various types of image blocks, including textual blocks, graphic blocks and their mixture. To make an optimal block sampling function $Pr(x|D)$, we first generate a group of local image blocks, and then a resample procedure is implemented through a floating cascade to reject outliers, which contain few or erroneous cues of image skew. The resample procedure can be considered as an iterative approximation to the optimal sampling function $Pr(x|D)$.

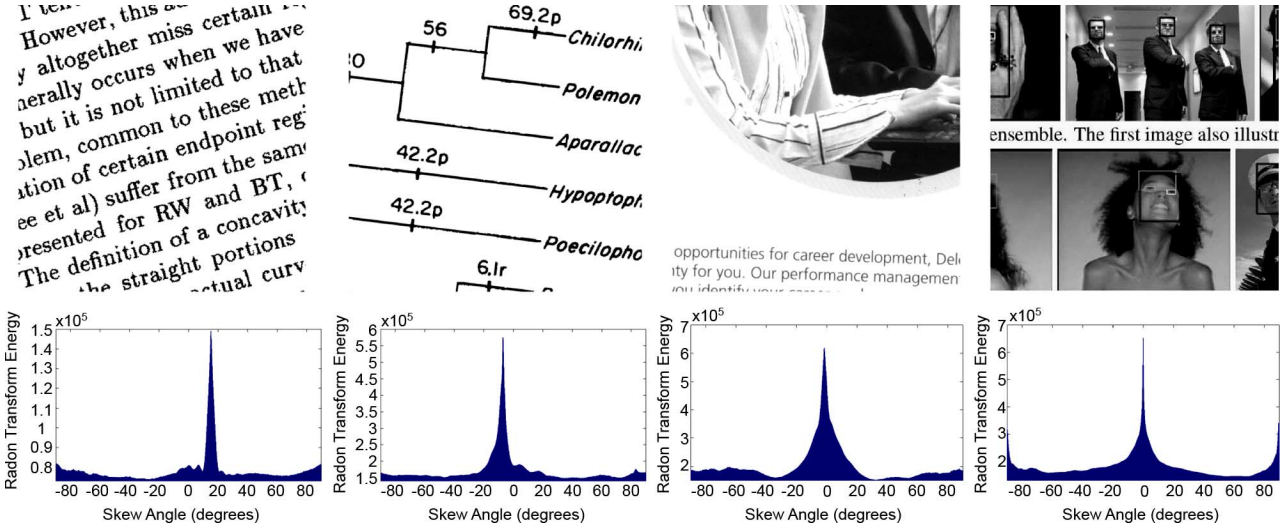


Fig. 2. Energy functions of Radon transform. (Top) various types of local image blocks, (bottom) the corresponding energy functions of Radon transform.

III. Skew Estimation Using Bagging

An overview of the proposed method is illustrated in Fig. 1. An input image is first divided into squared blocks. Then visual cues of image skew within each local image block are extracted using Radon transform. The outliers of them are iteratively rejected through a floating cascade. The skew angle of image is finally achieved by combining estimation results on the selected image blocks.

A. Image Blocks Division

Many methods have been proposed to draw local blocks from a document image. Avanindra *et al.* [28] presented a method based upon Monte Carlo sampling. The basic idea is that the skew angle can be correctly estimated provided that more than half of the sampled blocks are textual. This method works well for the document images that are dominant in text lines. However, if text lines in an image are sparse, the performance will be badly degraded.

In our method, a document image is divided into nonoverlapped squared blocks with equal size $L \times L$. The block size L is set proportionally to the image width or height to make it adaptive to different image resolutions. Every image block is also verified and the blank ones that contain few edge points are dropped.

B. Radon Transform

The Radon transform of an image block is an integral transform consisting of the integral of the image function over straight lines [29]. Given the straight line defined parametrically by

$$\begin{cases} x(t) = t\sin\theta + s\cos\theta \\ y(t) = -t\cos\theta + s\sin\theta \end{cases} \tag{5}$$

where s is the distance from the origin and θ is the angle from the x axis, the Radon transform of an image function $f(x, y)$ is defined by

$$\Re[f](\theta, s) = \int_{-\infty}^{+\infty} f(x(t), y(t))dt. \tag{6}$$

For a fixed angle θ, $\Re[f](\theta, s)$ can be viewed geometrically as the projections of image data at θ.

Based upon $\Re[f](\theta, s)$, we define an energy function of Radon transform of $f(x, y)$ with respect to θ by

$$energy(\theta|f) = \int_{-\infty}^{+\infty} \Re[f](\theta, s)^2 ds. \tag{7}$$

For a fixed θ, $energy(\theta|f)$ is actually the squared deviation of image projections from zeros. Generally, $energy(\theta|f)$ is highly peaked for an informative image block. Its maximum corresponds to the skew angle of the block. As a consequence, one can estimate the skew angle of $f(x, y)$ by

$$\hat{\theta} = \arg\max_{\theta} energy(\theta|f). \tag{8}$$

The above criterion for skew estimation may fail for image blocks with few informative visual cues of image skew. This occurs frequently when there are much graphics in the document. However, if a local image block contains sufficient cues of image skew, the estimator can often produce the correct results even though the block includes some nontextual objects. Shown in Fig. 2 are several examples of energy functions of Radon transform on various types of local image blocks.

C. Image Blocks Selection

We commonly observe that some image blocks are noninformative or even erroneous, especially when few visual cues of image skew are available. Eliminating such outlier blocks from the bagging estimation will help to improve its accuracy.

Blocks selection can be efficiently implemented through a cascade (see Fig. 3). Let n be the total number of stages in the cascade, and denote $\delta_k (1 \le k \le n)$ and $\hat{\theta}^{(k)}$ the angular resolution and the estimated skew angle of image at the kth stage, respectively. Without loss of generality, we initially set $\delta_0 = 90°$ and $\hat{\theta}^{(0)} = 0°$ for skew estimation in $(-90°, 90°]$.

For each image block $f_i (1 \le i \le N)$, its energy function of Radon transform is calculated at a sequence of angles $\theta_j (1 \le j \le m)$, which are quantized by δ_k from a detection range

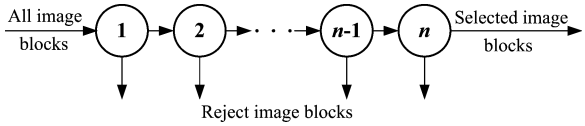


Fig. 3. Schematic diagram of blocks selection cascade. Image blocks are passed through each stage of cascade to iteratively reject noninformative ones.

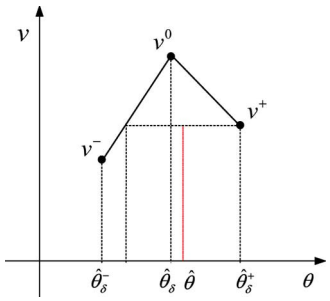


Fig. 4. Refinement of skew estimation by using the symmetry of peak.

$[\hat{\theta}^{(k-1)} - \delta_{k-1}, \hat{\theta}^{(k-1)} + \delta_{k-1}]$. The approximate skew angle of f_i at the kth stage can, thus, be estimated by

$$\hat{\theta}_i^{(k)} = \underset{\theta_j, 1\le j\le m}{\arg\max}\, energy(\theta_j|f_i, \delta_k). \tag{9}$$

Here, $energy(\theta_j|f_i, \delta_k)$ is the discrete version of $energy(\theta|f_i)$ with respect to an angular resolution δ_k. The skew angle of image at the kth stage can be estimated by summing up the energy functions of each image block, i.e.,

$$\hat{\theta}^{(k)} = \underset{\theta_j, 1\le j\le m}{\arg\max} \sum_i energy(\theta_j|f_i, \delta_k). \tag{10}$$

Finally, we reject the image blocks that satisfy $|\hat{\theta}_i^{(k)} - \hat{\theta}^{(k)}| > \delta_k$. The rest image blocks are then passed to the next stage for further selection at a refined angular resolution. The procedure above is repeated until termination conditions for blocks selection are satisfied. More details will be discussed in Section IV.

D. Skew Angle Estimation

Denote

$$v(\theta_j) = \sum_i energy(\theta_j|f_i, \delta), (1 \le j \le m) \tag{11}$$

and $\hat{\theta}_\delta$ the skew angle of image estimated by (10), where δ is the angular resolution of the last stage of cascade.

We observe that a roughly symmetric peak always appears around the maximum of $v(\theta_j)(1 \le j \le m)$. As stated by Baird in [5], this symmetry of the maximum peak can be exploited to improve the estimation of $\hat{\theta}_\delta$. Here, we take three points of $v(\theta_j)(1 \le j \le m)$ around its maximum peak and calculate their symmetric center to refine the estimation, as illustrated in Fig. 4, i.e.,

$$\hat{\theta} = \hat{\theta}_\delta + \frac{\delta}{2}\left(\frac{v^+ - v^-}{v^0 - \min(v^-, v^+)}\right) \tag{12}$$

where $v^0 = v(\hat{\theta}_\delta)$, $v^+ = v(\hat{\theta}_\delta + \delta)$ and $v^- = v(\hat{\theta}_\delta - \delta)$.

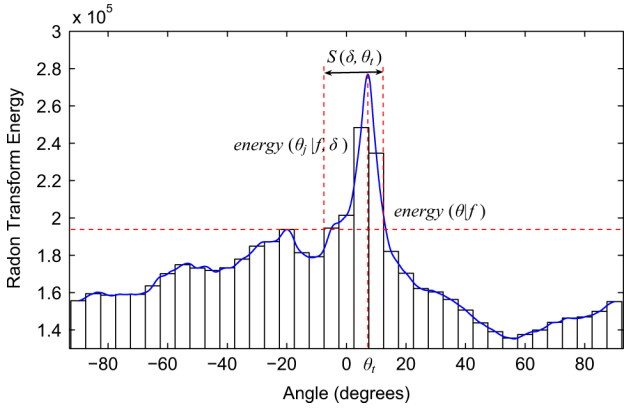


Fig. 5. $S(\delta,\ \theta_t)$ of an informative image block ($\delta = 5°,\ \theta_t = 7.2°$).

IV. Cascade Design

In this section, we will discuss several issues about the design of cascade for blocks selection, including angular resolutions determination, optimal cascade design and termination conditions for a floating cascade.

A. Angular Resolution

For an informative image block f, we assume in this section that its skew angle θ_t can be correctly estimated by the Radon transform based skew estimator, i.e.,

$$\theta_t = \underset{-90°<\theta\le 90°}{\arg\max}\, energy(\theta|f).$$

The largest angular resolution that can be used for the cascaded skew estimation of f is closely related to the major peak of $energy(\theta|f)$.

To exploit this property, we introduce a continuous angle set S, satisfying:

1) $\theta_t \in S$;
2) For all $\theta_j \in S$, $energy(\theta_j|f, \delta)$ increases monotonically if $\theta_j \le \theta_t$ and decreases monotonically if $\theta_j \ge \theta_t$;
3) $energy(\theta_j|f, \delta) > energy(\theta_k|f, \delta), \forall \theta_j \in S,\ \forall \theta_k \notin S$;
4) $S' \subseteq S, \forall S'$ that satisfies the above three conditions.

Geometrically, S is the span range of the major peak of $energy(\theta_j|f, \delta)$ that corresponds to θ_t. Notice that S varies with δ and θ_t, so we can rewrite it as $S(\delta, \theta_t)$. An example of the angle set $S(\delta, \theta_t)$ of an image block is illustrated in Fig. 5.

To reveal the role of $S(\delta, \theta_t)$ in the cascaded block skew estimation, we consider its essential size defined by

$$|S(\delta)| = \min_{-90°<\theta_t\le 90°} |S(\delta, \theta_t)| \tag{13}$$

where $|\cdot|$ is the length of an interval.

A correct cascaded skew estimation is closely related with $|S(\delta)|$. We have the following proposition.

Proposition 1: For an informative image block f_i with a skew angle θ_t, the correctness of its approximate skew estimation at the kth stage can always be ensured, i.e., $|\hat{\theta}_i^{(k)} - \theta_t| \le \delta_k, \forall \theta_t \in (-90°, 90°]$, if $\exists \delta < \delta_k$, such that $|S(\delta)| \ge \delta_k$.

The proof is straightforward. One has to notice that the approximate skew angle of an informative image block can always

Fig. 6. Some examples of 500 local image blocks manually selected for calculating the distribution of $|S(\delta)|$.

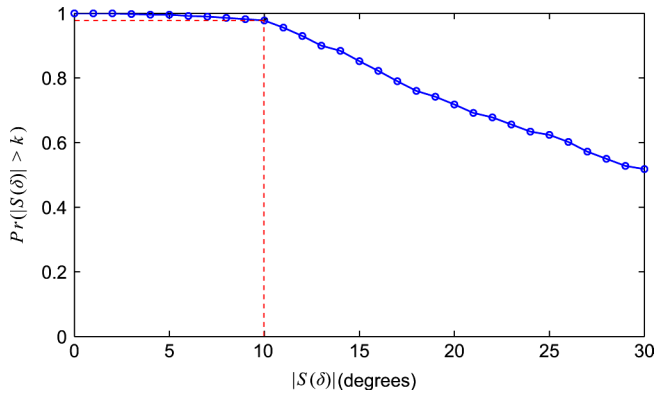


Fig. 7. Cumulative probability distribution of $|S(\delta)|(\delta = 0.5°)$ on the 500 manually selected informative image blocks.

be correctly estimated by (9) if at least one of its quantized angles at the kth stage falls into S. Obviously, this always holds for $|S(\delta)| \geq \delta_k$.

To explore the statistic property of $|S(\delta)|$, we manually selected 500 informative image blocks of various types from a wide variety of document images. An image block is selected based upon our visual inspection if we think some visual cues of image skew are available in the block. Though the selection of informative image blocks is somewhat subjective, it does reflect our common sense in the skew estimation of document images. Some examples of the selected informative image blocks are illustrated in Fig. 6.

We calculate $|S(\delta)|$ for the selected 500 image blocks at $\delta = 0.5°$. The cumulative probability distribution is illustrated in Fig. 7. From the result one can observe that the $|S(\delta)|$ of more than 95% of image blocks is larger than 10°. The percentage begins to drop rapidly as $|S(\delta)|$ is beyond 10°. This implies that more erroneous estimations are likely to happen if a larger angular resolution is adopted. As a result, we empirically limit the largest angular resolution of each stage to 10°.

B. Optimal Cascade Design

The computational cost for selecting an informative image block with $L \times L$ size is $O(L^2 \cdot \Psi)$, where Ψ is the total number of angles used for quantizing the skew detection range of each stage. Following from the procedure of blocks selection, we have

$$\Psi = \sum_{k=1}^{n} \left(\frac{2\delta_{k-1}}{\delta_k} + 1 \right) \tag{14}$$

where $\delta_k > 0 (1 \leq k \leq n)$ is the angular resolution of the kth stage.

Normally, it is not necessary to use a large block size L for an approximate block skew estimation. As a result, we can reduce the computational cost by decreasing the block size using image down-sampling techniques. Alternatively, we can choose the stage number n and the angular resolution $\delta_k (1 \leq k \leq n)$ of each stage to construct an optimal cascade.

Taking the derivative of (14) with respect to each $\delta_k (1 \leq k \leq n-1)$ and letting it be zero, we have

$$\frac{\partial \Psi}{\partial \delta_k} = \frac{2}{\delta_{k+1}} - \frac{2\delta_{k-1}}{\delta_k^2} = 0, \ (1 \leq k \leq n-1) \tag{15}$$

or equivalently

$$\frac{\delta_0}{\delta_1} = \frac{\delta_1}{\delta_2} = \cdots = \frac{\delta_{n-1}}{\delta_n} \triangleq C = \left(\frac{\delta_0}{\delta_n} \right)^{1/n}. \tag{16}$$

Substituting (16) into (14) gives

$$\Psi = 2nC + n = 2n \left(\frac{\delta_0}{\delta_n} \right)^{1/n} + n. \tag{17}$$

The optimal n and $\delta_k (1 \leq k \leq n)$ that minimize Ψ can be derived from (15) and (17), if the minimum angular resolution δ_n is selected. Recall that we limit the largest angular resolution of each stage to 10°, i.e., $\delta_1 \leq 10°$. Therefore, if we select the minimum angular resolution $\delta_n = 0.01°$, an optimal cascade with $n = 6$ stages can be derived from (17). Accordingly, the angular resolution of each stage can, thus, be determined as $\delta_1 = 10°$, $\delta_2 = 2.5°$, $\delta_3 = 0.5°$, $\delta_4 = 0.1°$, $\delta_5 = 0.04°$ and $\delta_6 = 0.01°$, respectively. By using an optimal cascade, the total computational cost for selecting an informative image block can be decreased to $O(64L^2)$, while the original computational cost is $O(18000L^2)$.

C. Floating Cascade

Practically, some document images, such as handwritten pages and scanned documents with curved text lines, consist of inaccurate cues of image skew. The skew angles of such images are often vaguely defined. As a result, it is impossible to make a highly accurate skew estimation for these images. In this case, the process of block selection is not necessary to be continued at stages with smaller angular resolutions. Using a floating cascade, which has a varying number of stages that is adaptive to the accuracy of image skew estimation is, thus, very advantageous.

For each image block f_i, we can estimate its approximate skew angle $\hat{\theta}_i^{(k)}$ by (9) and then accumulate it in an array $H^k(\theta_j)(1 \leq j \leq m)$. The overall skew angle of image $\hat{\theta}^{(k)}$ can

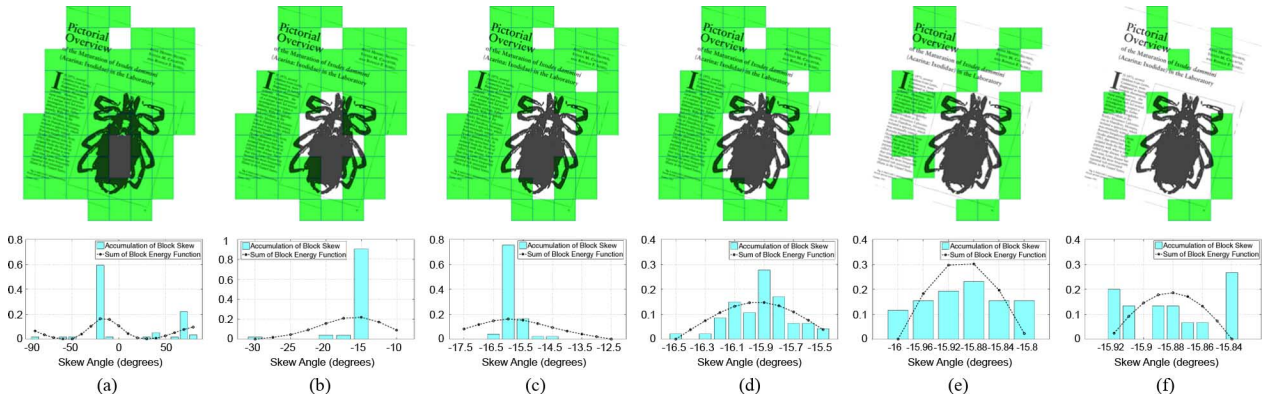


Fig. 8. Process of blocks selection for a document image with text and picture mixed. Totally 6 stages are employed in the floating cascade. (Top) blocks to be selected at each stage (painted in green), (bottom) the accumulation of block skew and the sum of block energy functions. (a) Stage 1, (b) Stage 2, (c) Stage 3, (d) Stage 4, (e) Stage 5, and (f) Stage 6.

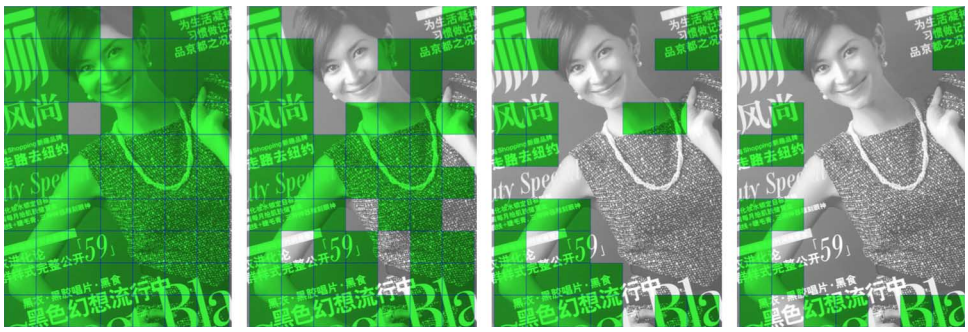


Fig. 9. Blocks selection on document image with complex background. The process of blocks selection stops at the fourth stage.

be estimated by (10). Alternatively, it can be estimated through a majority voting, i.e,

$$\tilde{\theta}^{(k)} = \underset{\theta_j, 1\le j\le m}{\arg\max}\, H^k(\theta_j). \tag{18}$$

The two estimations are generally consistent with each other at each stage of cascade. However, if few accurate visual cues are available in an image, a deviation between the two estimations will appear. This leads us to an empirical termination condition for image blocks selection.

The procedure of blocks selection in the cascade is terminated if one of the following conditions is satisfied:

1) $k \ge n$;
2) $|\hat{\theta}^{(k)} - \tilde{\theta}^{(k)}| > \delta_k$;
3) $|\hat{\theta}^{(k)} - \tilde{\theta}^{(k)}| \le \delta_k$ and $H^k(\hat{\theta}^{(k)}) \le \tau H^k(\tilde{\theta}^{(k)})$, where τ is a prescribed threshold.

The process of blocks selection for a document image with text and picture mixed is illustrated in Fig. 8, in which a cascade with a total of six stages is employed.

Blocks selection using a floating cascade can be adaptive to the estimation accuracy. For images with accurate visual cues of image skew, blocks selection is able to be implemented through more stages, while for images without accurate visual cues, such as handwritten documents, the process of blocks selection is often ceased at the early stage of cascade.

In addition, using a floating cascade for block selection is also beneficial for much complex document images. Shown in Fig. 9 is an example of block selection for image with complex background. In the example, the textual blocks, which are desirable for image skew estimation, are selected. Note that textual region detection is not necessary in the process.

V. Experimental Results

To evaluate the performance of the proposed method, three experiments are implemented on the University of Washington English Document Image Database III (UWDB-III). The first one is on the synthetic images. The skew angles of these images are exactly 0°, so they are quite suitable for testing the correctness and accuracy of the proposed method. The second one is performed on the 1600 images scanned from real printed English journals. The goal of this experiment is to test the robustness of the proposed method to noises, spare and short text lines, and the presence of large areas of nontextual objects etc. The third one is a comparative experiment on UWDB-I, which aims to make a comparison of our method with the state-of-the-art methods that were evaluated on the same dataset.

In the following experiments, we empirically set the block size $L = 1/10 \max(W, H)$ for image blocks division, where W and H are the image width and height, respectively.

A. Synthetic Document Images

In this experiment, the proposed method is first applied to the original synthetic images in UWDB-III. Afterwards, each of the synthetic images is randomly rotated to 10 different angles in the range of $\pm 90°$. The proposed method is then applied to the rotated images.

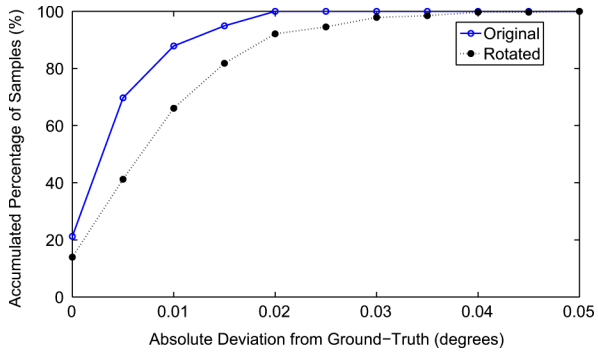


Fig. 10. Skew estimation error rate on the synthetic document images in UWDB-III. Each of synthetic images is randomly rotated to ten different angles in the range of $\pm 90°$.

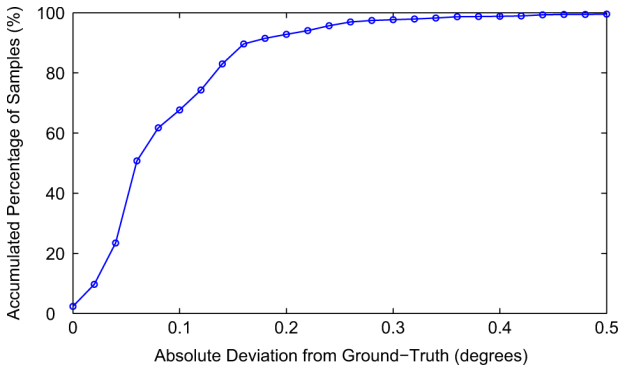


Fig. 11. Skew estimation error rate on the 1600 real-scanned document images in UWDB-III.

Fig. 10 gives the experimental results on the original and rotated images. For the original synthetic images, the average absolute error is 0.007° and the standard deviation is 0.009°. The absolute error is defined as the absolute difference between the estimated skew angle and the given ground-truth. For the rotated images, the average absolute error is 0.011° and the standard deviation is 0.014°. The accuracy is slightly lower than that on the original images. This is mainly due to the image quantization errors introduced by image rotation.

B. Real-Scanned Document Images

Fig. 11 shows the skew estimation error rate on the 1600 real-scanned document images. The average absolute error on the testing dataset is 0.097° and the standard deviation is 0.118°.

In comparison with the results on the synthetic images, the accuracy on the real-scanned images has an obvious decline. Actually, unlike the synthetic images, the ground-truths for the real-scanned document images are unknown and difficult to be precisely determined. This is especially true for document images with inaccurate cues of image skew, such as handwritten documents and scanned documents with multiple different skews. Consequently, it is better in practice to set a reasonable error range for evaluating the estimation results of these images. That is to say, if the estimation error falls into a permissible range, we can accept it. From Fig. 11, one can see that for more than 90% of images, the estimation errors of our method fall within

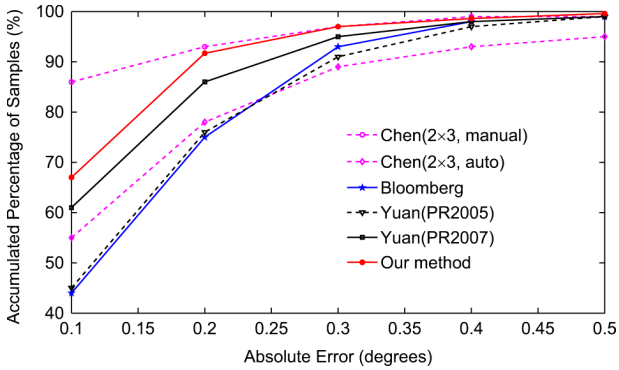


Fig. 12. Performance comparison using the real-scanned document images in UWDB-I. (See Table I for the numerical values).

TABLE I
PERFORMANCE COMPARISON OF FIVE METHODS USING THE REAL-SCANNED DOCUMENT IMAGES IN UWDB-I

	0.0°	0.1°	0.2°	0.3°	0.4°	0.5°
Bloomberg	-	44%	75%	93%	98%	99%
Chen (manual)	-	86%	93%	97%	99%	99%
Chen (auto)	-	55%	78%	89%	93%	95%
Yuan (PR2005)	3%	45%	76%	91%	97%	99%
Yuan (PR2007)	4%	61%	86%	95%	98%	99%
Our Method	2.5%	67.0%	91.7%	97.0%	98.6%	99.6%

a range of $\pm 0.2°$. Such small errors, in fact, give rise to a difference of only few pixels in a text line, which are difficult to be perceived by human eyes. This demonstrates the effectiveness of our method.

The average running time of our method is less than 200 ms for a document image in UWDB-III (2592×3300 pixels). The experiment is implemented by VC6.0 on an Intel Pentium IV PC with 2.8 GHz CPU and 1 GB RAM.

C. Comparisons

In this experiment, we make a comparison of our method with four published works that were evaluated on UWDB-I by Chen *et al.* [30] (the creators of the UW database), Bloomberg *et al.* [31] and Yuan *et al.* [21], [25], respectively.

The accumulated percentages of images versus the absolute estimation errors of the five methods are given in Table I. Their figures are also illustrated in Fig. 12. Since Chen *et al.* [30] and Bloomberg *et al.* [31] did not provide numerical results in their published papers, we adopt the comparative data presented by Yuan *et al.* [25].

As shown in Table I, within 0.1° of absolute error, Chen's method performs the best at 86% with manually fine-tuned parameters. Followed is our method at 67%, then is Yuan(PR2007) at 61% and Chen's method in its automatic mode at 55%. All the methods can estimate the skew angles of about 99% of testing images within 0.5° in their best parameter settings.

If all the methods are set in their automatic mode, our method performs the best. As shown in Table I and Fig. 12, within 0.2° of absolute error, the performance of our method is superior more than 5% to the best one that is ever reported on UWDB-I by Yuan *et al.* [25].

TABLE II
SKEW ESTIMATION ERRORS OF IMAGE L01DSYN WITH DIFFERENT NOISE DENSITY LEVELS

Noise Density	0.0	0.05	0.1	0.2	0.3	0.4
Error (°)	-0.016	-0.008	0.008	0.015	-0.060	-0.420

TABLE III
SKEW ESTIMATION ERRORS OF IMAGE L01DSYN (SKEW ANGLE: 10.420°) WITH DIFFERENT REDUCTION RATIOS (R. R.)

R. R. (%)	100	80	60	40	20	10	5
Error (°)	-0.016	-0.012	0.001	-0.058	0.020	0.080	-0.337

VI. DISCUSSION

In this section, we discuss several performance issues about the proposed method, including its sensitivity to image noises, image sizes, script languages and text font sizes.

A. Sensitivity to Image Noises

Table II gives the skew estimation errors of a randomly selected synthetic image L01DSYN from UWDB-III. The image is rotated counterclockwise 10.420° and then salt & pepper noises with different level densities are added onto the rotated image.

From the table, one can observe that the estimation error increases with the deterioration of image quality. This is consistent with our intuition. As the noise density increases, visual cues of image skew are ruined and difficult to be robustly extracted.

Excessive document noise presents a serious problem for skew estimation. Shivakumara *et al.* [32] proposed an efficient skew estimation method using boundary growing, which can work under noise density up to 0.05 level but not beyond. In comparison, our method can tolerate noise density up to 0.3 level without any additional preprocessing step to eliminate the noises in advance.

B. Sensitivity to Image Sizes

Table III shows the estimation results on the image L01DSYN (image size: 3300×2550, resolution: 300 dpi) with different reduction ratios. The image is reduced to 80%, 60%, 40%, 20%, 10%, 5% of its original size, respectively. One can see from the results that the image size has a minor impact on the estimation results. The estimation accuracy does not drop sharply until the image is reduced to 5% of its original size, i.e., 165×127 (resolution: 15 dpi). Actually, in such a sufficiently reduced image, the quantization errors of the image are overwhelming and very few visual cues are remained for an accurate skew estimation.

Nevertheless, by a simple calculation that $\tan^{-1}(1/165) \cdot 180^\circ/\pi \approx 0.347^\circ$, one can see that the accuracy on the reduced image (5% of the original size, resolution: 15 dpi) is still better than the highest accuracy achieved by linear regression. This demonstrates the superiority of our method over those based upon linear regression analysis.

C. Sensitivity to Script Languages

Since all the samples in UWDB-III are English documents, the experiment on UWDB-III is insufficient to evaluate how the proposed method works on document images in other languages. For this purpose, we make six textual images with A4 size at 200 dpi, consisting of Chinese, Japanese, Korea, French, Indian, and Arabic texts, respectively. In the experiment, the six images are first randomly rotated to 10 different angles within $\pm 90^\circ$, resulting in totally 60 new images. The proposed method is then applied to the 60 rotated images. The average absolute errors and standard deviations on each class of images are listed in Table IV.

TABLE IV
AVERAGE ABSOLUTE ERRORS (AAE) AND STANDARD DEVIATIONS (S. D.) OF IMAGES IN DIFFERENT LANGUAGES

Language	Chinese	Japanese	Korean	French	Indian	Arabic
AAE (°)	0.030	0.050	0.046	0.016	0.005	0.010
S. D. (°)	0.046	0.062	0.058	0.013	0.007	0.011

From Table IV, one can observe that the accuracy on different script languages varies in a narrow range between 0.005° and 0.05°. This reveals that the proposed method is less affected by script languages. However, script languages do affect some methods using base points fitting, since the fiducial points of characters may differ distinctly in different languages and sometimes are difficult to extract due to the adhesion of characters. For example, Indian characters are often connected by a head line (*shirorekha* or *matra*). It may cause serious problems for components analysis based methods [18] due to the difficulty of characters segmentation.

It is also interesting to investigate the small variations of accuracy on different languages. Chinese, Japanese and Korea are orient languages. Characters in these documents are generally mono-spaced and loosely aligned in a text line. The overall accuracy for these images is about 0.05°. For the nonorient language documents, such as French, Indian or Arabic documents, the accuracy is near 0.01°, slightly better than that of orient languages. This is mainly because the nonorient characters are always aligned more compactly than the orient characters and, thus, can provide much accurate visual cues for skew estimation.

For the orient language documents, the accuracy of Chinese document is about 0.03°, higher than that of 0.046° of Korea document and 0.05° of Japanese document. Chinese characters are square character with a uniform size. Consequently, its alignment within a text line is much more regular than those of Korea and Japanese documents. The latter is often a mixture of Chinese words and some other types of characters, such as Japanese katakana and English words.

For the nonorient language documents, Indian document has the highest accuracy at 0.005°. Followed is that of Arabic documents at 0.01° and that of French documents at 0.016°. The short head lines that connect Indian characters may cause problems for components analysis based methods. But for our method, they instead provide much accurate cues for skew estimation.

D. Sensitivity to Font Sizes

Table V gives the experimental results on document images with different font sizes. We make ten English document images using the typeface of *Times New Roman* with font sizes

TABLE V
AVERAGE ABSOLUTE ERRORS (AAE) AND STANDARD DEVIATIONS (S. D.) OF SYNTHETIC IMAGES WITH DIFFERENT FONT SIZES

Font Size(pts)	10	12	14	16	18	20	24	28	36	48
AAE (°)	0.013	0.036	0.017	0.013	0.012	0.014	0.027	0.020	0.013	0.052
S. D. (°)	0.013	0.035	0.016	0.016	0.014	0.015	0.035	0.025	0.020	0.064

ranging from 10pts to 48pts. In the experiment, each image is first randomly rotated to 10 different angles within $\pm 90^\circ$ and the proposed method is applied to the rotated images. From the results, one can see that text font size has a minor impact on the accuracy of the proposed method. The accuracy on images with 36pts font size is not significantly different yet to those on the images with smaller font sizes.

Generally, characters in images with large font sizes are sparsely distributed. As a result, once the image is divided into blocks, the number of characters within one block is often very small. The alignment of characters is, thus, inadequate for the block skew estimation. But on the other hand, the strokes of characters become prominent in this case, the horizontal or vertical directions of which provide reliable indications of the correct image skew. The proposed method is able to take good advantage of these local cues for a robust skew estimation.

VII. CONCLUSION

In this paper, we have proposed a general-purpose and accurate method based upon bagging for skew estimation of document images. The excellent performance of the proposed method benefits much from the combination of a variety of local image visual cues. Such visual cues are proven to be complementary to each other for skew estimation. The visual cues are first extracted by Radon transform and then the outliers of them are iteratively rejected through a floating cascade. A bagging estimator is finally employed to combine the visual cues on local image blocks.

To evaluate the performance of the proposed method, a series of large-scale experiments based upon UWDB-III are conducted. The experimental results show that the proposed method is an accurate and general-purpose method. It is highly competitive in execution speed and estimation accuracy, and extremely robust to document noises, multiple different skews, short and sparse text lines, and the presence of large areas of nontextual objects of various types and quantities.

ACKNOWLEDGMENT

The authors would like to thank the associate editor Prof. M. Gupta and the anonymous reviewers for their remarks and suggestions which improved the manuscript significantly.

REFERENCES

[1] Y. Y. Tang, S. W. Lee, and C. Y. Suen, "Automatic document processing: A survey," *Pattern Recognit.*, vol. 29, no. 12, pp. 1931–1952, 1996.

[2] 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.

[3] A. D. Bagdanov and J. Kanai, "Evaluation of document image skew estimation techniques," in *Proc. SPIE*, San Jose, CA, 1996, vol. 2660, pp. 343–353.

[4] J. J. Hull, "Document image skew detection: Survey and annotated bibliography," in *Document Analysis Systems II*, J. J. Hull and S. L. Taylor, Eds. Singapore: World Scientific, 1998, pp. 40–64.

[5] H. S. Baird, "The skew angle of printed documents," in *Proc. SPSE 40th Ann. Conf. and Symp. on Hybrid Imaging Systems*, Rochester, NY, 1987, pp. 21–24.

[6] E. Kavallieratou, N. Fakotakis, and G. Kokkinakis, "Skew angle estimation for printed and handwritten documents using the wigner-ville distribution," *Image Vis. Comput.*, vol. 20, no. 11, pp. 813–824, 2002.

[7] N. Liolios, N. Fakotakis, and G. Kokkinakis, "On the generalization of the form identification and skew detection problem," *Pattern Recognit.*, vol. 35, no. 1, pp. 253–264, 2002.

[8] R. Kapoor, D. Bagai, and T. S. Kamal, "A new algorithm for skew detection and correction," *Pattern Recognit. Lett.*, vol. 25, no. 11, pp. 1215–1229, 2004.

[9] S. T. Li, Q. H. Shen, and J. Sun, "Skew detection using wavelet decomposition and projection profile analysis," *Pattern Recognit. Lett.*, vol. 28, no. 5, pp. 555–562, 2007.

[10] Y. Ishitani, "Document skew detection based on local region complexity," in *Proc. 2nd Int. Conf. Document Analysis and Recognition*, Tsukuba, Japan, 1993, pp. 49–52.

[11] Y. K. Chen and J. F. Wang, "Skew detection and reconstruction based on maximization of variance of transition-counts," *Pattern Recognit.*, vol. 33, no. 2, pp. 195–208, 2000.

[12] H. Yan, "Skew correction of document images using interline cross-correlation," *CVGIP, Graph. Models Image Process.*, vol. 55, no. 6, pp. 538–543, 1993.

[13] B. Gatos, N. Papamarkos, and C. Chamzas, "Skew detection and text line position determination in digitized documents," *Pattern Recognit.*, vol. 30, no. 9, pp. 1505–1519, 1997.

[14] D. S. Le, G. R. Thoma, and H. Wechsler, "Automated page orientation and skew angle detection for binary document images," *Pattern Recognit.*, vol. 27, no. 10, pp. 1325–1344, 1994.

[15] B. Yu and A. K. Jain, "A robust and fast skew detection algorithm for generic documents," *Pattern Recognit.*, vol. 29, no. 10, pp. 1599–1629, 1996.

[16] A. Amin and S. Fischer, "A document skew detection method using the hough transform," *Pattern Anal. Appl.*, vol. 3, no. 3, pp. 243–253, 2000.

[17] H. K. Kwag, S. H. Kim, S. H. Jeong, and G. S. Lee, "Efficient skew estimation and correction algorithm for document images," *Image Vis. Comput.*, vol. 20, no. 1, pp. 25–35, 2002.

[18] 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, 2008.

[19] 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.

[20] Y. Lu and C. L. Tan, "A nearest-neighbor chain based approach to skew estimation in document images," *Pattern Recognit. Lett.*, vol. 24, no. 14, pp. 2315–2323, 2003.

[21] B. Yuan and C. L. Tan, "Fiducial line based skew estimation," *Pattern Recognit.*, vol. 38, no. 12, pp. 2333–2350, 2005.

[22] V. Aradhya, G. H. Kumar, and P. Shivakumara, "An accurate and efficient skew estimation technique for south indian documents: A new boundary growing and nearest neighbor clustering based approach," *Int. J. Robot. Autom.*, vol. 22, no. 4, pp. 272–280, 2007.

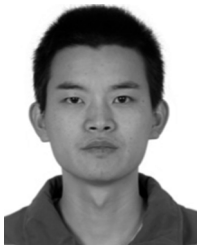
[23] H. Liu, Q. Wu, H. B. Zha, and X. P. Liu, "Skew detection for complex document images using robust borderlines in both text and non-text regions," *Pattern Recognit. Lett.*, vol. 29, no. 13, pp. 1893–1900, 2008.

[24] C. H. Chou, S. Y. Chu, and F. Chang, "Estimation of skew angles for scanned documents based on piecewise covering by parallelograms," *Pattern Recognit.*, vol. 40, no. 2, pp. 443–455, 2007.

[25] B. Yuan and C. L. Tan, "Convex hull based skew estimation," *Pattern Recognit.*, vol. 40, no. 2, pp. 456–475, 2007.

[26] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.

[27] J. V. Hansen, "Combining predictors: Comparison of five meta machine learning methods," *Inf. Sci.*, vol. 119, no. 1–2, pp. 91–105, 1999.
[28] Avanindra and S. Chaudhuri, "Robust detection of skew in document images," *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 344–349, Feb. 1997.
[29] S. R. Deans, *The Radon Transform and Some of Its Applications*. New York: Wiley, 1983.
[30] S. Chen, "Document Layout Analysis Using Recursive Morphological Transforms," Ph.D. dissertation, Univ. Washington, Seattle, WA, 1995.
[31] D. S. Bloomberg, G. E. Kopec, and L. Dasari, "Measuring document image skew and orientation," in *Proc. SPIE, Document Recognition II*, San Jose, CA, 1995, vol. 2422, pp. 302–316.
[32] P. Shivakumara and G. H. Kumar, "A novel boundary growing approach for accurate skew estimation of binary document images," *Pattern Recognit. Lett.*, vol. 27, no. 7, pp. 791–801, 2006.



Gaofeng Meng received the B.S. degree in applied mathematics from Northwestern Polytechnical University, Xi'an, China, in 2002, and the M.S. degree in applied mathematics from Tianjin University, Tianjin, China, in 2005, and the Ph.D. degree in control science and engineering from Xi'an Jiaotong University, Xi'an, Shaanxi, China, in 2009.

In 2009, he jointed the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China as an Assistant Professor. His current research interests include image processing, computer vision, and pattern recognition.

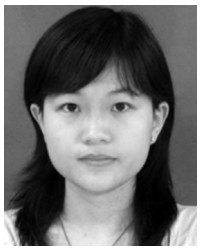
Chunhong Pan received the B.S. degree in automatic control from Tsinghua University, Beijing, China, in 1987, the M.S. degree from Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, China, in 1990, and the Ph.D. degree in pattern recognition and intelligent system from Institute of Automation, Chinese Academy of Sciences, Beijing, in 2000.

He is currently a professor at National Laboratory of Pattern Recognition of Institute of Automation, Chinese Academy of Sciences. His research interests include computer vision, image processing, and remote sensing.

Nanning Zheng (SM'94–F'06) received the B.S. degree from the Department of Electrical Engineering, Xi'an Jiaotong University, Xi'an, China, in 1975, the M.S. degree in information and control engineering from Xi'an Jiaotong University in 1981, and the Ph.D. degree in electrical engineering from Keio University, Yokohama, Japan, in 1985.

In 1975, he joined Xi'an Jiaotong University, where he is currently a Professor and the Director of the Institute of Artificial Intelligence and Robotics. Since August 2003, he has been the President of Xi'an Jiaotong University. His research interests include computer vision, pattern recognition, machine vision and image processing, neural networks, and hardware implementation of intelligent systems.

Dr. Zheng became a member of the Chinese Academy of Engineering in 1999 and has been the Chief Scientist and the Director of the Information Technology Committee of the China National High Technology Research and Development Program since 2001. He was General Chair of the International Symposium on Information Theory and Its Applications and General Co-Chair of the International Symposium on Nonlinear Theory and Its Applications, both in 2002. He is a member of the Board of Governors of the IEEE ITS Society and the Chinese Representative on the Governing Board of the International Association for Pattern Recognition. He also serves as an Executive Deputy Editor of the Chinese Science Bulletin.



Chen Sun received the B.S. degree in information engineering and the M.S. degree in control science and engineering from Xi'an Jiaotong University in 2005 and 2008, respectively, and the general engineering degree from Ecole Centrale Paris in 2006.

In July 2008, she joined Shanghai Baosight Software Co., Ltd. as a software engineer. Her current research interests include image processing and pattern recognition.