Available online at www.sciencedirect.com
SCIENCE DIRECT®
Pattern Recognition 38 (2005) 2333–2350
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Fiducial line based skew estimation

Bo Yuan*, Chew Lim Tan

Department of Computer Science, School of Computing, National University of Singapore, Singapore 117543, Singapore

Received 5 October 2004; received in revised form 28 January 2005; accepted 7 March 2005

Abstract

Skew estimation for textual document images is a well-researched topic and numerals of methods have been reported in the literature. One of the major challenges is the presence of interfering non-textual objects of various types and quantities in the document images. Many existing methods require proper separation of the textual objects which are well aligned from the non-textual objects which are mostly nonaligned. Some comparative evaluation work on the existing methods chooses only the text zones of the test image database. Therefore, the object filtering or zoning stage is crucial to the skew detection stage. However, it is difficult if not impossible to design general-purpose filters that are able to discriminate noises from textual components. This paper presents a robust, general-purpose skew estimation method that does not need any filtering or zoning preprocessing. In fact, this method does apply filtering, but not on the input components at the beginning of the detection process, rather on the output spectrum at the end of the detection process. Therefore, the problem of finding a textual component filter has been transformed into finding a convolution filter on the output accumulator array. This method consists of three steps: (1) the calculation of the slopes of the virtual lines that pass through the centroids of all the unique pairs of the connected components in an image, and quantizes the arctangents of the slopes into a 1-D accumulator array that covers the range from $-90°$ to $+90°$; (2) a special convolution on the resultant histogram, after which there remain only the prominent peaks that possibly correspond to the skew angles of the image; (3) the verification of the detection result. Its computational complexity and detection precision are uncoupled, unlike those projection-profile-based or Hough-transform-based methods whose speeds drop when higher precision is in demand. Speedup measures on the baseline implementation are also presented. The University of Washington English Document Image Database I (UWDB-I) contains a large number of scanned document images with significant amount of non-textual objects. Therefore, it is a good image database for evaluating the proposed method.

Keywords: Skew estimation; Centroids; Component pairs; Fiducial lines; Noise immunity; UWDB-I

1. Introduction

Text lines in documents are horizontally or vertically oriented by human reading customs. Due to misalignment in the digitization process, the text lines in the resultant images usually deviate from their original orientation by certain amount, which is referred to as skew angle. Skew angle detection is one of the important processing steps in document image understanding. It has drawn extensive studies and a large array of techniques has been developed [1–10]. There are also in-depth reviews [11,12] and comparative evaluations [13] available.

The general principle of skew detection for textual document images is to find a proper representation of text lines, and develop a method to draw the correspondence

* Corresponding author: Tel.: +65 6874 6396; fax: +65 6775 7717.
E-mail address: yuanbo@nus.edu.sg (B. Yuan).

0031-3203/$30.00 © 2005 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2005.03.023

between certain property of the chosen representation and the actual orientation of the images. Popular classes of skew detection methods in the literature include those based on projection-profiles [1–4], Hough-transforms [5,6], nearest-neighborhood [7,8], and morphological operations [9,10]. Different approaches compete on the ground of detection accuracy, time and space efficiencies, abilities to detect the existence of multiple skews in the same image, and robustness in noisy environments (such as graphical objects or their residues after filtering) and scan-introduced distortions (such as object-touching due to insufficient resolution, or warping due to non-flat scanned surfaces).

The skew detection process of the projection-profile-based methods starts by marking a set of points (fiducial points) to represent the objects in an image. Then, it makes parallel projection of these points onto an accumulator array, and a chosen premium function is evaluated on the accumulator array. If the accumulator array is successively rotated by a fixed interval in a range, a series of premiums are obtained. The premium function reaches extremes when the projection is along the text line orientations. Usually, detection is carried out in multiple rounds to achieve good speed and accuracy, either from coarse to fine accumulator rotation interval or from sub-sampled to full resolution image.

The idea of using Hough-transform to detect skew angles of document images is that if the objects of a text line can be represented by a set of points (fiducial points) $\{x, y\}$, each of these points can be mapped with normal parameterization $\rho = x \cos\theta + y \sin\theta$ to a sinusoidal curve in the quantized $\rho - \theta$ parameter space by scanning the whole range of parameter θ in certain intervals. Therefore, if the mapped curves are accumulated in the parameter space, the global maxima $\{\rho_{max}, \theta_{max}\}$ correspond to the prominent text line orientations of the image. All the skew detection methods deploying Hough transforms must make aggressive computation reduction in order to achieve acceptable performance for real world documents, usually with compromises in accuracy and other properties.

Nearest-neighborhood (k-NN) provides a spatial clue for local grouping of objects that belong to a text line. The positions of grouped objects are then used to approximate the orientation of the text line. Although the nearest-neighbor searching process is global, the skew angle detection process is local. Therefore, its skew detection precision is usually not as high as that of the other methods that detect skew on global scale.

Noises, the residual components from photographic objects or scan artifacts after binarization, in textual document images affect all skew detection methods in different ways. Projection-profile-based methods in principle are able to deal with noises only when the noises are uniformly distributed on the whole page (rare in real-world documents) so that the peak evaluation on the projection-profiles is not compromised. Hough-transform-based methods are also able to deal with moderate amount of noises of irregular distribution. However, the presence of excessive noises may produce false maxima which are the results of noise points that happen to be collinear. This severely undermines peak-searching for the already fussy and weak signals in the 2-D accumulator array, as well as drastically slows down the detection speed. The nearest-neighborhood-based methods' ability to deal with excessive noises is generally weak, as noises may easily corrupt the buildup of neighborhoods. The same can be said about the methods using morphological operations, which can be thought of as alternative ways of object grouping at the pixel level.

Noise filtering is commonly used as a preprocessing measure in all classes of the skew detection methods. This is more for the efficiency than for the robustness of the detection algorithms. These filters used are either based on the prior knowledge of the samples to be processed or common sense in order not to be too aggressive in component elimination. It is difficult if not impossible to design general-purpose filters that are able to discriminate noises from textual components.

This paper proposes a skew estimation method that is robust to process document images that contain a large amount of non-textual components or noises. Its noise immunity comes from its uses of component-pairs rather than individual components like most of other methods do. This exploration of the inter-component correlation enables higher linear alignment identification, which leads to easier design of a filter for identifying the peaks that correspond to prominent skew angles in the resultant 1-D accumulator array. It is a general-purpose, full-range ($\pm 90°$) skew estimation method that is straightforward in principle, simple in implementation, parameter-free, and highly competitive in detection accuracy and execution speed.

2. The proposed method

This method works on the extracted components from a binary image. It uses the centroids of the components as the representation the components (*fiducial points*), as centroids are rotation-invariant, thus a proper choice for skew detection in the range of $\pm 90°$. It traverses all the unique pairs of components to calculate the slopes of the virtual lines passing through their centroids (*fiducial lines)*, as illustrated in Fig. 1. The arctangent values of the slopes of the fiducial lines are computed and quantized into an accumulator array to form an angle histogram, as shown in Fig. 3. The prominent peaks in the histogram are the candidates for the detected skew angles. Fig. 2 superposes the fiducial lines drawn along the angle at the peak position of the histogram in Fig. 3 on the input image in Fig. 1.

As Fig. 4 shows, the fiducial lines highly concentrate along the direction of the text lines and widely spread over other angles, the contribution from the components in the same text line overwhelm that from different lines or regions. This is the reason why this method is robust in the existence