ON PERFORMING SKEW DETECTION AND CORRECTION USING MULTIPLE EXPERTS' DECISION

Nicolae TARBĂ[1]
Daniel SCHMIDT[2]
Anda - Elena POPOVICI[3]
Eduard STĂNILOIU[4]
Cristian AVATAVULUI[5]
Marcel PRODAN[6]

Abstract: *One of the main objectives of image processing algorithms is extracting information. An important class of such algorithms is the one responsible for Optical Character Recognition (OCR). A common problem among OCR systems is the need to work with properly oriented images. The process of finding the skew angle and rotating the initial image by its opposite is known as deskew. This paper presents a voting-based approach to solve the document deskew problem, combining results from different algorithms to cover a wide range of document types. The proposed method proved both precise and reliable, suitable to be integrated into mass-production retroconversion systems.*

Keywords: *skew correction, document deskew, image processing, skew detection, layout analysis*

1. Introduction

A. Problem Motivation

The main purpose of OCR algorithms is to convert documents into digital format for easier manipulation and preservation. They achieve this by matching pixels of detected characters to pixels of known characters.

If the documents are to be retrieved correctly, the images processed by the OCR algorithms should be easy to read, and in the above context, this means that they

[1] Engineer, University Politehnica of Bucharest, 060042 Bucharest, Romania, nicolae.tarba@upb.ro
[2] Engineer, University Politehnica of Bucharest, 060042 Bucharest, Romania, daniel.schmidt@stud.fsa.upb.ro
[3] Engineer, University Politehnica of Bucharest, 060042 Bucharest, Romania, ungureanu.anda.elena@gmail.com
[4] Teaching Assistant, PhD Student, University Politehnica of Bucharest, 060042 Bucharest, Romania, eduard.staniloiu@cs.pub.ro
[5] PhD Student, Eng. University Politehnica of Bucharest, 060042 Bucharest, Romania, cristianavatavului@gmail.com
[6] PhD Student, Eng., University Politehnica of Bucharest, 060042 Bucharest, Romania, marcoprod@gmail.com

should be straight and not noisy. Because of this, several preprocessing algorithms, such as denoising, thresholding, de-skewing, etc. should be applied to the image to improve the accuracy of the OCR. This paper will focus on skew correction, in order to ensure proper general image document retroconversion and enable correct text line retrieval [14].

B. Previous Work

Hough Transform [4] takes the image from the Cartesian coordinate system to the polar coordinate system. By doing this, each point/pixel (x, y) will have a corresponding sine wave in polar coordinates. If the Cartesian points are collinear, then their corresponding sine waves will intersect, the point of intersection having as one of the coordinates the angle of the line that goes through the Cartesian points.

Another possibility is to use the Fourier Transform [5]. When applying the Fourier Transform on a skewed image, the resulting image will have the same skew. Because of this, the Hough Transform can be used to de-skew images that do not have any lines.

Projection profiling [6] takes place in the Cartesian coordinate system, where the entire image is projected on one of its axes. If the image is a skewed document, the projection on the vertical axis will resemble a uniform distribution. On the other hand, if the image is an un-skewed document, the result of the projection will be a saw-like distribution, its variance being maximal. The algorithm simply iterates through the angles in an interval, looking for the projection with the highest variance.

Again, this algorithm may be used on the Fourier Transform to de-skew images that are not necessarily image documents.

Another approach is to use Machine Learning techniques for skew detection and deskewing. Kassm and Achkar [8] present a model for License Plate Recognition, which uses cascaded layers of Convolutional Neural Networks for detecting a plate, detecting its skewing angle, deskewing the image and using the deskewed image for character recognition.

2. Proposed Method

This paper proposes a voting system that combines the results from Hough Transform, Projection Profiling, and Frequency Domain Hough Transform into one answer. Because of this, a confidence level was added to the output of each algorithm.

Implementation

The steps taken for the Hough Transform algorithm are:

1. Pre-process the image

   a. Threshold the image [10]
   b. Identify contours (letters/groups of letters) in the thresholded image
   c. Generate a new image by placing points at the same coordinates as the anchors of the contours found in the previous step
2. Look for lines using the OpenCV [7] V4.12 function *HoughLines* (the function returns a list of the most voted lines, in descending order of votes).
3. Compute the confidence
   a. The lines that have fewer votes than the most voted one, divided by a constant, are eliminated (different values were given to the constant in order to normalize the confidence values of the algorithms)
   b. Compute the sum of all votes for the lines that are left
   c. Compute the sum of the votes for the lines that have the same angle as the dominant line
   d. Divide the two sums and obtain a number between 0 and 1

Figure 1. From left to right: a. initial image; b. image with anchor points; c. skew angle detection

The steps for the Projection Profiling algorithm are:

1. Pre-process the image
   a. Threshold the image [11]
   b. Identify contours (letters/groups of letters) in the thresholded image
   c. Generate a new image by placing points at the same coordinates as the anchors of the contours found in the previous step
2. Rotate the image and find the angle for which the vertical projection has a maximal variance
3. Compute the confidence
   a. Divide the maximal variance by the sum of all variances (which was found, empirically, to have values between 0 and 0.5)

b. Subtract 0.5

c. Take the absolute value

d. Multiply by a constant (to normalize the confidence values of the algorithms)

e. Subtract the result from 1

The steps for the Frequency Domain Hough Transform algorithm are:

1. Pre-process the image
   a. Pad the image so it becomes square-shaped
2. Compute the Fourier Transform
3. Create the image of amplitudes
4. Threshold the image [12]
5. Look for lines using OpenCV HoughLines
6. Compute the confidence (the steps are the same as the regular Hough Transform)
7. Compute the confidence (the steps are the same as the regular Hough Transform)

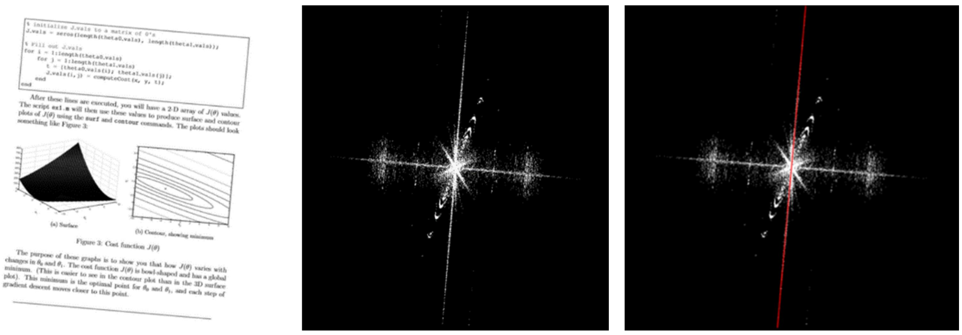
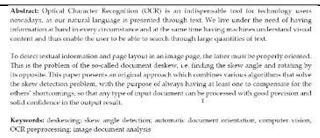
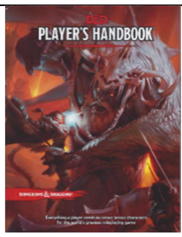
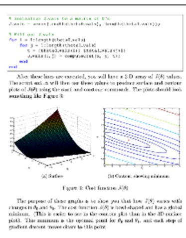


Figure 2. From left to right: a. initial image; b. processed FFT; c. skew angle detection

The voting system combines the angles and confidences resulted from all these algorithms and provides an optimal solution. Three voting mechanisms were approached: best first, unanimous, and weighted voting. The first method picks the result of the algorithm with the greatest confidence, completely neglecting the others. The unanimous voting computes the skew angle as the average of each algorithm's result. The last method, weighted voting works in the same way as unanimous voting, but this time a weighted average is computed. Better results can be obtained if the output with a confidence lower than half of the maximum is discarded.

Evaluation

The test sample consisted of 14 images rotated by angles ranging from -5 degrees to 5 degrees with a 0.1° step. The angle search space was also limited to [-5, 5] to save execution time, but the application can perform on any angle. A dataset of 1414 images might seem small, but this paper is just a proof of concept. Larger tests can be conducted if needed. The dataset needs to be as diverse as possible. It includes several classes of images as can be observed in Table 1.

Class	Sample image	Motivation
Simple blocks of text	 **Abstract:** Optical Character Recognition (OCR) is an indispensable tool for technology users nowadays, as our natural language is presented through text. We live under the need of having information at hand in every circumstance and at the same time having machines understand visual content and thus enable the user to be able to search through large quantities of text. To detect textual information and page layout in an image page, the latter must be properly oriented. This is the problem of the so-called document deskew, i.e. finding the skew angle and rotating by its opposite. This paper presents an original approach which combines various algorithms that solve the skew detection problem, with the purpose of always having at least one to compensate for the others' shortcomings, so that any type of input document can be processed with good precision and solid confidence in the output result. **Keywords:** deskewing; skew angle detection; automatic document orientation, computer vision, OCR preprocessing; image document analysis	The most trivial of cases, good for testing the ability to detect rows
Images with little text		Very little information surrounded by as much noise as possible
Scanned documents with blocks of text on different backgrounds		The most common use case
Text with images		Moderate amount of information and noise

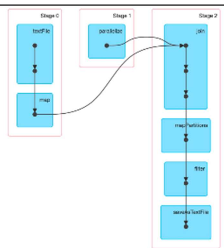
Class	Sample image	Motivation
Handwritten documents		These should be harder for spatial domain algorithms, but the frequency domain shouldn't have problems with them
Diagrams, graphs, etc.		A small amount of information with a moderate amount of noise

Table 1. Classes of images included in the skew detection dataset

3. Performance Measurements

Best First Voting produces the best results as can be observed in Table 2.

Deskew Method	Accuracy	Average Error	Average Confidence	Execution Time (ms)
Hough Lines	99.150%	0.085°	0.632	4.473
Projection Profiling	99.523%	0.048°	0.545	131.019
Frequency Hough Lines	99.363%	0.064°	0.481	51.512
Best First Voting	**99.568%**	**0.043°**	**0.752**	**187.004**
Unanimous Voting	99.541%	0.046°	0.616	187.004
Weighted Voting	99.553%	0.045°	0.651	187.004

Table 2. Compared performance indicators for the presented method against candidates

The results are also consistent with the ones presented in [1], where similar accuracies were reported for the voting systems.

Comparison with Existing Techniques

The results of the voting systems are slightly better than the individual results of the voters. On a dataset of 1414 images, the accuracy went up by 0.04%, which might seem like a very small increase. However, the starting accuracy was already above 99% so the increase in accuracy is significant.

4. Conclusions

Spatial Hough transform is much faster than the voting systems and its accuracy is only 0.4% lower. It is hard to justify the existence of the voting system without valuing accuracy much higher than computational cost. Spatial Hough transform is a reasonable compromise between accuracy and computational cost, but when compromise is not an option it is great to have better alternatives accuracy-wise. Out of all the alternatives presented in this paper, Best First Voting seems to be the best.

Future Work

Many future improvements will be carried in the near future:

- The dataset will be further diversified
- The confidence computation for each algorithm will be improved
- More voters will be added to the voting system
- Computation-heavy voters may be discarded in favor of lower-cost alternatives
- Redundancy will be minimized by using specialized algorithms for small image classes instead of general algorithms that work well on the same classes as other voters
- Integration into a larger image retroconversion system using several other voting-based techniques [9][11][13] will also be performed.

Acknowledgement

This work was supported by a grant of the Romanian Ministry of Research and Innovation, CCCDI - UEFISCDI, project number PN-III-P1-1.2-PCCDI-2017-0689 / „Lib2Life- Revitalizarea bibliotecilor şi a patrimoniului cultural prin tehnologii avansate” / "Revitalizing Libraries and Cultural Heritage through Advanced Technologies", within PNCDI III

References

[1] C.A. Boiangiu, O.A. Dinu, C. Popescu, N. Constantin, C. Petrescu, *Voting-Based Document Image Skew Detection in Applied Sciences*, Vol. 10, No. 7, 2020, pp. 2236;

[2] I.P. Gurov, A.S. Potapov, O.V. Scherbakov, I.N. Zhdanov *Hough and Fourier Transforms in the Task of Text Lines Detection*, 11th International Conference on Quality Control by Artificial Vision, Fukuoka, Japan, 2013;
[3] Safonov, I. Kurilin *Deskew for Card Image Scanning*, Proceedings of Graphicon, 2011;
[4] A.S. Hassanein, S. Mohammad, M. Sameer, M.E. Ragab *A Survey on Hough Transform, Theory, Techniques and Applications*, JCSI Volume 12, 2015;
[5] N. Beaudoin, S. Beauchemin *An accurate discrete Fourier transform for image processing*, IEEE, 2002;
[6] Papandreou, B. Gatos *A Novel Skew Detection Technique Based on Vertical Projections*, IEEE, 2011;
[7] OpenCV, *Open source Computer Vision library*, URL: https://opencv.org/, Accessed on May 4th, 2020;
[8] G.A. Kassm, R. Achkar, *LPR CNN cascade and adaptive deskewing*. Procedia computer science, Vol. 114, 2017, pp. 296-303;
[9] Costin-Anton Boiangiu, Radu Ioanitescu, Razvan-Costin Dragomir, *Voting-Based OCR System*, The Journal of Information Systems & Operations Management, Vol. 10, No. 2, 2016, pp. 470-486.
[10] Costin-Anton Boiangiu, Ion Bucur, Andrei Tigora, *The Image Binarization Problem Revisited: Perspectives and Approaches*, The Journal of Information Systems & Operations Management, Vol. 6, No. 2, 2012, pp. 419-427.
[11] Costin-Anton Boiangiu, Mihai Simion, Vlad Lionte, Zaharescu Mihai, *Voting Based Image Binarization*, The Journal of Information Systems & Operations Management, Vol. 8, No. 2, 2014, pp. 343-351.
[12] Costin-Anton Boiangiu, Andrei Iulian Dvornic, Dan Cristian Cananau, *Binarization for Digitization Projects Using Hybrid Foreground-Reconstruction*, Proceedings of the 5th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, August 27-29, 2009, pp.141-144.
[13] Costin-Anton Boiangiu, Paul Boglis, Georgiana Simion, Radu Ioanitescu, *Voting-Based Layout Analysis*, The Journal of Information Systems & Operations Management, Vol. 8, No. 1, 2014, pp. 39-47.
[14] Costin Anton Boiangiu, Mihai Cristian Tanase, Radu Ioanitescu, *Handwritten Documents Text Line Segmentation based on Information Energy*, International Journal Of Computers Communications & Control (IJCCC), Vol 9, No 1, 2014, pp. 8-15.