A New Approach for Skew Correction of Documents Based on Particle Swarm Optimization

Javad Sadri
School of Computer Science,
McGill University, 3480 University Street,
McGill Center for Bioinformatics,
Montreal, Quebec, Canada, H3A 2A7,
Email: javad.sadri@mcgill.ca

Mohamed Cheriet
Synchromedia Laboratory,
Ecole de Technologie Superieure (ETS),
1100 Notre-Dame Street West,
Montreal, Quebec, Canada, H3C 1K3,
Email: mohamed.cheriet@etsmtl.ca

Abstract

This paper presents a novel approach for skew correction of documents. Skew correction is modelled as an optimization problem, and for the first time, Particle Swarm Optimization (PSO) is used to solve skew optimization. A new objective function based on local minima and maxima of projection profiles is defined, and PSO is utilized to find the best angle that maximizes differences between values of local minima and maxima. In our approach, local minima and maxima converge to the locations of lines and spaces between lines. Results of our skew correction algorithm are shown on documents written in different scripts such as Latin and Arabic related scripts (e.g. Arabic, Farsi, Urdu,...). Experiments show that our algorithm can handle a wide range of skew angles, also it is robust to gray level and binary images of different scripts.

1. Introduction

Optical Character Recognition (OCR) has many important applications such as: automatic bank cheque processing, mail address recognition, and historical document recognition. In most of these applications, because of inappropriate document position or feeder failure in scanners, copiers, or fax machines, skew is introduced into the document images. Technically, skew is defined as the deviation of the base lines of the text from the horizontal direction [1]. An example of skewed document can be found in Figure 1. Skew angles of as little as 1 degrees can be apparent to human viewer, and most of document analysis and recognition algorithms require the images to be deskewed if the skew angle is greater than a few degrees [2, 3]. Therefore, skew correction is an important pre-processing task in every automatic document processing system.

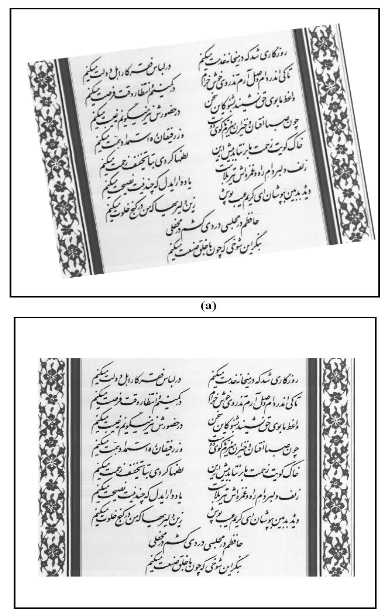

Figure 1. (a) A Persian (Farsi) document with skew about -8.5 degrees, (b) The same document after it has been deskewed.

Several methods have been proposed for skew detection and correction of document images. Many methods have used projection profiles for skew detection [4]. In general, projection profile methods are limited to estimate skew angle within ± 10 degrees [2,5]. The second category of methods are based on Hough Transform [6, 7]. Hough Transform is applied to find the best direction of the straight lines in documents. Another group of skew correction methods, uses k-nearest neighbor clustering on connected com-

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.268

IEEE computer society

ponents, and they try to find the best overall direction of neighboring components [2,8]. See [9] for a survey on skew correction methods. Although many methods have been proposed for skew detection and correction, many of them can only be applied to binary images or only to machine printed documents. Some methods also entail high computational cost such as hough transform methods. Some other methods only deal with small skew angles. For example, although projection profiles are easily can be constructed, due to high computational cost of exhaustive search, the range of search for skew angles is usually restricted to ± 10 degrees.

In this paper, we propose two novel modifications in order to improve projection profile methods. First, we introduce a new method for measuring variations in projection profiles. Second, in order to avoid exhaustive search for finding the best skew angle, we propose utilizing the Particle Swarm Optimization (PSO). PSO algorithms are search and optimization algorithms that in the recent years have received a great deal of attention in optimization [10,11]. Our skew correction method is able to correct the skew of both handwritten and typewritten texts in gray level and binary document images with wide range of image qualities. It is also able to handle skew angles greater than ± 10 degrees.

The rest of this paper is organized as follows. In Section 2, the technical details of our method are explained. In Section 3, we show some experimental results. Finally in Section 4, we draw our conclusion.

2. Methodology

In this section, we explain the details of our skew detection and correction algorithm. In the first subsection, we explain our modified projection profile and objective function. Then in the next subsections, we explain using PSO for finding the best skew angle, and also skew correction.

2.1. Modified Projection Profiles

Projection profiles (or image histograms) are calculated by adding gray levels (or binary values) of all the pixels along all the rows in the image. For example, consider an image $I(x, y)$ where, x and y represent rows and columns, respectively. The projection profile $PP(x)$ of I is calculated according to Equation 1. The normalized version of $PP(x)$ is denoted by $P(x)$, and it is calculated based on Equation 2, as follows:

$$PP(x) = \sum_{y} I(x, y) \quad (1)$$

$$P(x) = c \cdot \frac{PP(x)}{\max_x(PP(x))} \quad (2)$$

Here, c is a constant, and it shows the height of the normalized histogram. In this paper, for our experiments, we choose $c = 10$. As an example, projection profile $P(x)$ for the document image in Figure 1-a is obtained, and it is shown in Figure 2-a. As can be seen in this figure, normally document histograms are very noisy functions, and they have many spurious local maxima and minima. In order to simplify measuring the variations in these histograms, we use spline curves to smooth their noises. Splines are piecewise polynomial functions which have a very good capacity to approximate complex functions. They can be easily constructed, and they mainly used in applications such as interpolation and smoothing of one-dimensional or multidimensional data (for more details on splines and their applications refer to [12]). We used one-dimensional cubic spline curves for smoothing of projection profiles of document images. As shown in Equation 3, smoothed version of $P(x)$ by cubic spline is shown by $\widehat{P(x)}$, where cubic spline smoothing operator is denoted by $\Psi(.)$.

$$\widehat{P(x)} = \Psi(P(x)) \quad (3)$$

An example of application of cubic splines for smoothing of $P(x)$ is shown in Figure 2-b. As seen in this figure, $\widehat{P(x)}$ is much smoother than $P(x)$, and it has much less number of local maxima/minima. In Figure 2-b, we have found all local maxima and minima for $\widehat{P(x)}$. In this figure, the heights of local minima and maxima have been shown by solid and dashed bars, respectively. Here, local minima are correspondong to those rows of the image that have lower gray values (or more black pixels), and local maxima are corresponding to those rows of the image that have higher gray values (or more white pixels). If the original skewed document in Figure 1-a to be rotated in different angles, the height of these local minima and maxima in Figure 2-b will be changed. It can easily be verified that if the original skewed document to be rotated in the correct direction of deskewing, the heights of all (or most of) the local minima (here solid bars) will decrease, and the heights of all (or most of) the local maxima (here dashed bars) will increase. Comparison of Figure 2-b and Figure 3-b shows this effect. Now based on this and other similar observations, we define our objective function $F(\theta)$ for all skew angles θ as in Equation 4.

$$F(\theta) = \sum_{i} \widehat{P}(x_{max_i}) - \sum_{j} \widehat{P}(x_{min_j}) \quad (4)$$

Here, $\{x_{max_i}\}$ and $\{x_{min_j}\}$ are the sets of all local maxima and minima of $\widehat{P(x)}$ (computed at skew angle θ). In fact, based on our definition in Equation 4 and based on our previous observation, $F(\theta)$ measures the variations of

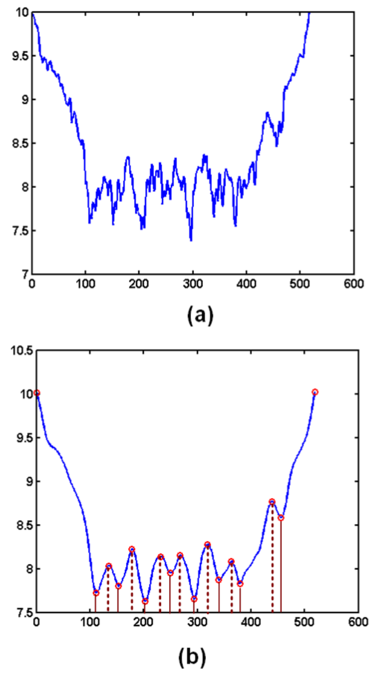

Figure 2. (a) Normalized histogram $P(x)$ for skewed document in Figure 1-a, (b) Using cubic splines, smoothed histogram is obtained and denoted by $\widehat{P(x)}$. In $\widehat{P(x)}$ the heights of local minima and maxima have been shown by solid and dashed bars, respectively.

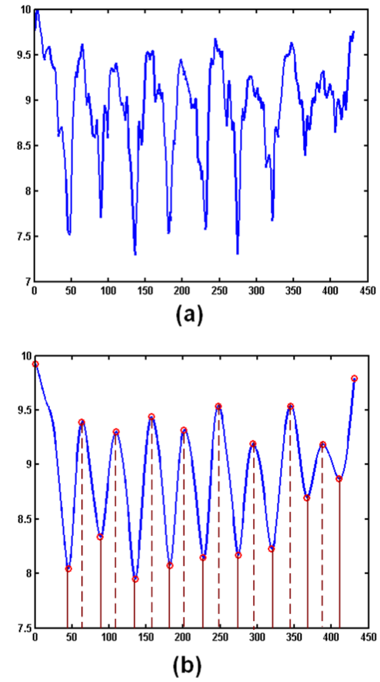

Figure 3. (a) Normalized histogram $P(x)$ for the deskewed document in Figure 1-b, (b) Smoothed histogram $\widehat{P(x)}$ by cubic Splines. In (b) the heights of local minima and maxima have been shown by solid and dashed bars, respectively. For the deskewd document, the location of these bars indicate the location (rows) of the lines and spaces between lines.

document histogram in angle θ. Therefore, in order to find skew angle for a document, we have to maximize its corresponding $F(\theta)$. $F(\theta)$ for the sample document in Figure 1-a has been shown in Figure 4. As seen in this example, $F(\theta)$ has several local maxima and only one global maximum. The global maximum of $F(\theta)$ can be found by an ehuastive search, however, it entails the computation and processing of histograms for every possible angle of θ. In order to expand the range of search for larger skew angels, and also in order to speed up the search and avoid high computational cost of processing of histograms for every angle θ, we propose using PSO. In the next subsection, we will explain how to utilize PSO for the task of global maximization of $F(\theta)$.

2.2. PSO for Skew Detection

PSO is a population-based evolutionary search and optimization algorithm. In PSO each member of the swarm (so-called particle) has a current location, velocity, and direction of the movement. If the search space is d-dimensional, then i^{th} particle of the swarm can be represented by a d-dimensional position vector $\Theta_i = (\theta_i^1, \theta_i^2, \ldots, \theta_i^d)$. The velocity of this particle is represented by another d-dimensional vector $V_i = (v_i^1, v_i^2, \ldots, v_i^d)$. The best previously visited position of the i^{th} particle is denoted as $P_{ibest} = (p_i^1, p_i^2, \ldots, p_i^d)$. If g is the index of the individual in the swarm which currently has the best fitness, its position is denoted by $P_{gbest} = (p_g^1, p_g^2, \ldots, p_g^d)$. In PSO, swarm is initialized with random particles, and then the algorithm is iterated for several generations. In each generation, the particles are manipulated according to the following equations 5 and 6 for their velocities and locations, respectively:

$$v_i^k = v_i^k + c_1\varphi_1(p_i^k - \theta_i^k) + c_2\varphi_2(p_g^k - \theta_i^k) \quad (5)$$

$$\theta_i^k = \theta_i^k + v_i^k \quad (6)$$

Where c_1 and c_2 are positive constants, and φ_1 and φ_2 are random numbers, uniformly distributed between 0, and 1. It has been shown that setting each of c_1 and c_2 close to 2

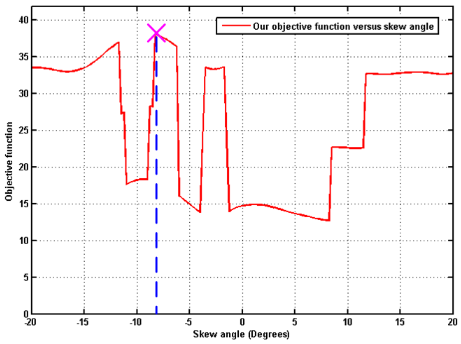

Figure 4. (a) Our objective function $F(\theta)$ for document shown in Figure 1-a for all skew angles θ between ± 20. The global maximum θ_{max} happens at -8.5 degrees (denoted by dashed bar). This value of θ_{max} is the output of our algorithm as the detected skew angle for the document in 1-a. θ_{max} is also the true skew angle for this document.

gets the best overall search performance (for more details on PSO algorithm refer to [10, 11]). In this paper, PSO is utilized in order to find the global maximum of our skew objective function $F(\theta)$ (an example was shown in Figure 4). As seen in this example, maximization of $F(\theta)$ is a one dimensional optimization problem ($d = 1$).

In our PSO implementation, we chose our search space for skew angle θ_i to be in the range of $[-\Theta_{max}, +\Theta_{max}]$. Here, Θ_{max} is the maximum assumed skew angle for the input document. In our experiments, we set Θ_{max} equal to 45 degrees, and we considered V_{max} (maximum speed or displacement in each stage for particles) equal to ± 1 in order to make the search for new angels very smooth. After, some trail experiments, we also set the swarm size equal to 10, and the maximum number of generations equal to 20. Although the maximum number of generations is set to 20, our algorithm will stop earlier if the best particle (θ_{max}) does not move during the last three consecutive iterations (early stopping). Now, by applying these settings and by updating the particles based on the Equations 5 and 6, PSO algorithm can easily find the global maximum θ_{max} for the objective function $F(\theta)$. In the next section, we briefly explain skew correction.

2.3. Skew Correction

After applying PSO algorithm, the best skew angle which is globally maximizes function $F(\theta)$ is found, and it is denoted by θ_{max}. Here, θ_{max} is called detected skew angle, and it is used to perform a rotation on the skewed document in the opposite direction to remove the skew. An example of skew corrected document based on our method was shown in Figure 1-b.

3. Experimental Evaluation

For the experimental evaluation of our method, we collected 100 images of documents of different qualities, containing handwritten, typewritten, also historical documents in both gray level and binary formats. Our data set also includes documents from different scripts such as English and Arabic related scripts (e.g. Arabic, Farsi, Urdu,...). First, we artificially skewed all of these documents with random skew angles between ± 45 degrees (by randomly rotating them), and we recorded their corresponding skew angles, as ground truth information. Then, we applied our skew correction algorithm in order to detect and measure the skew of those documents, and to correct their skews. Experiments showed that, in 96.34% of those skewed documents, our algorithm was able to detect skew angle, and recover the original document image within a maximum error range of ± 1 degrees (skew angles less than ± 1 degrees can be barely seen by eyes). However, in the rest of the documents (3.66%), the algorithm was not able to find the correct skew angle. Investigating of these unsuccessful cases showed that in these cases the PSO algorithm was stuck in one of the local maxima of $F(\theta)$ (skew objective function), or due to early stopping it could not reach to the global maximum (the true skew angle of the document). Also, the average number of iterations of the algorithm (after applying condition for early stopping) was 4. Considering the swarm size of 10, this means that the average number of evaluations of $F(\theta)$ per each document was about 40. This is much lower than the number of evaluations by the exhaustive search for the same documents in the skew range of ± 45 degrees (in fact, atleast 90 evaluations of $F(\theta)$ is required by the exhaustive search). Therefore, the algorithm is much faster than the exhaustive search for skew correction. Figures 5, 6, 7 show some examples of the results of our algorithm.

4. Conclusions

Skew detection and correction in document images are critical steps before layout analysis, segmentation and recognition. In this paper, we proposed a new method of skew correction for handwritten and typewritten documents in both gray level and binary formats. We modified projection profiles by utilizing spline curves, also we introduced an efficient objective function in order to measure variations in projection profiles in a robust way. Also, in order to avoid exhaustive search for the best skew angle, our method used

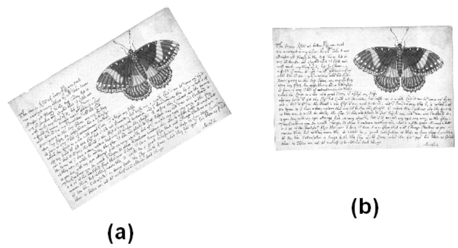

Figure 5. (a) Original skewed image (English), (b) Skew corrected image by our algorithm

Figure 6. (a) Original skewed image (Arabic mixed with Urdu), (b) Skew corrected image by our algorithm

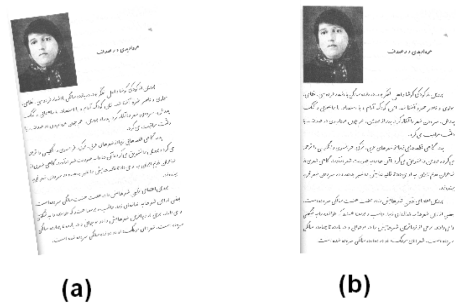

Figure 7. (a) Original skewed image (Farsi), (b) Skew corrected image by our algorithm

Particle Swarm Optimization (PSO) for skew detection. To the best of our knowledge, so far PSO has not been used for skew detection. Utilizing PSO enables our algorithm to search efficiently for the best skew angle in the greater ranges of angles much larger than ± 10. Results of our skew correction algorithm are shown on documents written in different scripts such as Latin and Arabic related scripts (e.g. Arabic, Farsi, Urdu,...). Our experimental results show that our algorithm performs well on gray level and binary document images with different qualities, and it is also robust to different scripts. In the future, we are going to improve our algorithm in order to boost its performance and speed, also, to test its performance on larger collections of documents.

References

[1] P. Slavik, and V. Govindaraju, IEEE Trans. on PAMI, Vol. 23, No. 3, pp. 323-326, March, 2001.

[2] X. Jiang, H. Bunke, and D. Widmer-Kljajo, Proc. of the Fifth International Conference on Document Analysis and Recognition, Bangalore, pp.629-632, 1999.

[3] U. Pal, and B.B Chudhuri, Pattern Recognition Letters 17, Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkatta, India, pp. 899-904, 1996.

[4] A. Bagdanov and J. Kanai, ICDAR, pp. 401-405, 1997.

[5] J. Kumar, T. Kasar, and A. G. Ramakrishnan, IEEE TENCON 2007, pp.1-4, 2007.

[6] S. Srihari and V.Govindaraju, Machine Vision and Applications, 2:141-153, 1989.

[7] C. Singha, N. Bhatiab, and A. Kaurc, Pattern Recognition, Vol. 41, pp. 3528-3546, 2008

[8] Z. Shi and V. Govindaraju, "Skew Detection for Complex Document Images Using Fuzzy Runlength", Proc. of the 7tht ICDAR, 2003.

[9] J. J. Hull, In: J. J. Hull, S. L. Taylor, Document Analysis Systems, vol. II. World Scientific, Singapore. pp. 40-64. Appeared in Document Analysis Systems II, pp. 40-64, 1998.

[10] J. Kennedy and R. Eberhart. Swarm Intelligence. Morgan Kaufmann Publishers, Inc., San Francisco, CA, 2001.

[11] R. Poli, W. B. Langdon, and O. Holland, Proc. of European Conference on Genetic Programming (EuroGP 2005), pp. 291-300, Switzerland, 2005.

[12] J. H. Ahlberg, E. N. Nielson, and J. L. Walsh, The Theory of Splines and Their Applications, Mathematics in Science and Engineering, New York: Academic Press, 1967.