A Novel Method for De-warping in Persian Document Images Captured by Cameras

Hadi Dehbovid hadi.dehbovid@gmail.com
Electrical Engineering Department, Faculty of Engineering,
Islamic Azad University,Science and Researches Branch,
Tehran, Iran.

Farbod Razzazi farbod_razzazi@yahoo.com
Electrical Engineering Department, Faculty of Engineering,
Islamic Azad University, Science and Researches Branch,
Tehran, Iran

Shahpour Alirezaee sh_alirezaee@yahoo.com
Electrical Engineering Department, Faculty of Engineering,
Zanjan University, Zanjan, Iran Zanjan, Iran

Abstract

In this Paper, We proposed a novel algorithm for de-warping of Persian document images captured by the cameras. The aim of de-warping is to remove page distortions and to straighten document images captured by the cameras, so that the documents are readable to the OCR system. Recently, the industrial implementation of the images captured by digital cameras has significantly expanded. Most of the studies carries out so far in this regard have focused on the documents written in Latin and few researches have been conducted regarding Persian documents. The original idea of the proposed algorithm is based on the segmentation of the components of texts. In this algorithm, an effective technique is offered for detection of the upper and lower baselines, which is used in estimation of the slope of the words. Moreover, vertical shift of the warped words is done through fitting a quadratic curve fitted to the centers of the words in a line in relation to the horizontal line. The suggested algorithm is examined by qualitative and quantitative measures and the results of its implementation on various documents indicate a 92% accuracy of the proposed technique in correction of the location and angle of the words.

Keywords: Geometric Distortion, OCR, camera based OCR, Image Archives

1. INTRODUCTION

Recently, the significant role of digital cameras in preparation and analysis of document images is expanding. The reason behind this development is the fact that compared to scanners; such cameras are lighter, more applicable and less expensive. Cameras can be employed in cases where scanners are not of any use. Imaging of thick books and hand-written historical books that have to not be touched are examples of such cases. Such advantages have resulted in applicability of digital cameras in a wide range of applications such as printed documents digital library and natural scenes containing textual contents.

The analysis of the document images captured by these cameras can be divided into different categories based on the type of the document, the technology used in the process of imaging and the desired application.

Generally, images can be captured from any context containing textual contents (such as a paper document, a subtitled video frame, a car plate number, or an open book). In fact, the differences in method of the imaging, will determine the complexities of the textual contents extraction.

The conditions of capturing images with cameras are more complex and different, in comparison to scanners. Therefore, techniques that are more effective and robust have to be applied for the analysis of camera-based document images. Novel algorithmic technologies used for document analysis, generate desirable results with pure, flat text documents. By pure texts, we mean those documents that do not include images or formulas and, and those in which a single language is used. Flat documents refer to documents that are free from geometric distortions. In such techniques, the assumption is a high-quality simple-structured document image (i.e. a straightened black colored text on white background).

Unfortunately, there are no such assumptions for camera-based systems, resulting in new challenges for their applications. Examples of such challenges include low resolution, variable brightness in the page, perspective distortion, non-planer surfaces, complex backgrounds, zooming and focusing on a specific part, movements of the objects while a photo is captured, as well as warping of the content. The significant challenge that we are going to address here is geometric warping of the document images. Warping of the document images is one of the main complexities of the pre-processing stage, and different techniques have been devised so far in order to remove this problem. In fact, the actual aim of de-warping of the pages is to align the document image captured by camera before the OCR stage, so that the document is readable by an OCR system. The most important types of such distortions are those resulting from perspective and paper curvatures.

In the recent years, various techniques are proposed to recover the distortion of the warped document images [1]. These techniques are categorized into hardware versus software-based page reconstruction sets. An example of the first type is the reconstruction of a 3D shape by means of hardware such as stereo cameras [2,3], photometric devices [4], and laser-based scanners [5]. Page reconstruction using a single camera in an uncontrolled environment [6, 7, and 8] can be mentioned as an instance of the second group. In the present article, our assumption is an image taken by a digital camera, without any hardware post-processing. Therefore, the focus of our study is on the second type of algorithms.

Among the second type of techniques mentioned above, three main techniques of restoration of the warped images are mostly focused on. These are representation of a continuous skeletal image for document images de-warping (SKEL)[9], segmentation–based recovery of the warped document images (SEG) [10], and a coordinates-translation model for guiding the documents in one direction and recovery of the distorted document (CTM) [11].

The central idea of the skeletal technique is based on the extraction of the surface skeleton of the document in which connected branches estimated by means of Bezier curves, determine the internal space of the text lines. CTM is a model based on the translation of the cylindrical coordinates and through biasing the document, restores the warped book. Besides, the basis of SEG technique lies upon the segmentation of the words and correction of the skew of these words, using the baselines estimated for them, and finally the vertical translation of these warped words. The three models are used in de-warping Latin document images; however, none of the previously proposed methods can remove the distortion of such documents completely. The SKEL method is able to remove the distortions resulting from page curvatures, but fails to do so for perspective distortions, nor can it restore the distortions occurred to formulas and some components of the documents belonging to the adjacent page. CTM can recover formulas to some extent and also works for the components of the document which are apparent from the neighboring pages, however, it fails to remove the perspective and page curvature distortions. SEG model also is applicable for the parts appearing from the adjacent pages as well as distortions of formulas and page curvatures, yet unable to work for perspective distortions. Among these three models, the SEG method is of more applicability with regard to de-warping documents[13]. Therefore, in the present article, this model is considered as the base model and its improved version is proposed as a technique for restoration of the distortions occurred to documents written in Persian or Arabic. It has to be pointed that since the SEG model is based on the features of Latin orthography, it is not directly applicable to Persian texts. The experimental results presented in this study also confirm such a fact.

The remaining of present paper is structured as follows: in section 2, the related literature is reviewed and. In the 3rd section, the proposed algorithm is represented. Section 4 includes the experimental results and discussions on the recovered images quality and finally the article will be summarized in section 5.

2. SEGMENTATION-BASED MODEL FOR DE-WARPING

2.1 SEG Model

As was mentioned before, SEG model is considered as the base model of this study. Such a technique improves the quality of the images captured by digital cameras through the following four stages:

In the first stage, noisy areas and the black margins as well as the noisy components of the text belonging to adjacent pages are detected and reduced. The original idea of black margins discrimination is based on vertical as well as horizontal projections. First, the image is smoothed, and then the initial and final offsets from the corners and textual areas are calculated with regard to vertical and horizontal projections.

Finally, the black margins are removed through the analysis of the connected segments of the image. In the next step, the words are reconstructed through smoothing.

During the second stage, words and baselines of the text are detected through implementing the segmentation model for the documents for which noise reduction has been carried out. The adjacent words are linked to each other consecutively in order to define the text lines.

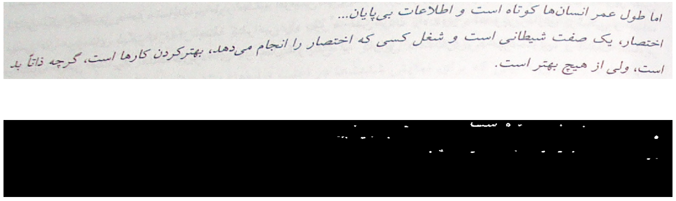
This is performed by the sequential extraction of the words on the right and left sides of the first detected word after a stage by stage top-down scan of the document. For each determined word, the upper and lower baselines are calculated [12].

In the third stage, the initial binary warped image is restored through de-warping and translation of the words with regard to the upper and lower baselines. For each word, the skew is calculated using the slopes of the corresponding upper and lower baselines. Then, all the detected words are de-warped and transformed, so that a primary estimation of the recovered binary image is at hand. Ultimately, it is in the fourth stage that the smoothed image is reconstructed.

2.2 Limitations of the original model with regard to Persian documents

SEG technique have a number of limitations when used for Latin texts and even more problems emerge when this model is used for documents written in Persian. Regarding English texts, there is no appropriate filter for accurate detection of noises and removing them. Besides, in such texts, the estimtion of the upper and lower baselines is not always error free.

In case the original model is implemented with Persian texts, none of its stages works appropriately and the only output will be some noisy pixels. Thus, the original model is merely considered as a basis for offering a novel algorithm. The original technique bears many problems if used for restoration of the Persian document images. It faces lots of complexities in case of detection and elimination of noisy areas. Moreover, due to the differences in the writing structure of English and Persian, segmentation of the words in Persian texts also will be problematic, since unlike English texts, in Persian documents the height and slopes of the letters of the words do not follow a specific order. Therefore, again the original SEG model will face difficulties in determining the upper and lower baselines. In other words, in Persian, the slope detection and transformation of the words in a geometrically warped text have to be done in a structured manner relying on the whole page. Whereas, in English, these are conducted for each word separately, independent from other words on the page. Another critical limitation of the original model is that the translation of words in each line is based on the highest point in the written form of the first word of the same line. This is less problematic for English texts, since their writing structure is more organized; however, with regard to Persian texts, difficulties are abundant due to the fact that in Persian the height of the text line does not correspond to the height of the words. In figure 1, implementation of the base model on a Persian imaged document is discussed. As we can observe, the original method has no application for Persian texts and the distortions resulting from warping of the image cannot be restored. This limitation made us to propose a new technique in order to remove the above mentioned problems. In our suggested model, robust solutions are presented for the stated limitations.



اما طول عمر انسان‌ها کوتاه است و اطلاعات بی‌پایان...
اختصار، یک صفت شیطانی است و شغل کسی که اختصار را انجام می‌دهد، بهترکردن کارها است، گرچه ذاتاً بد
است، ولی از هیچ بهتر است.

FIGURE 1: Implementing the original model on a Persian sample, from top to down, a. the original document image, b. the output image

3. THE PROPOSED ALGORITHM

The flow chart of the proposed algorithm is demonstrated in figure 2.

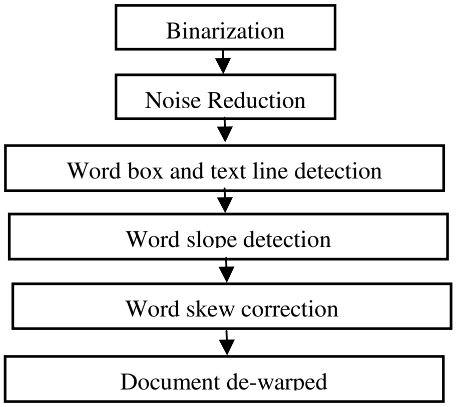


FIGURE 2: Flowchart of the proposed algorithm

As the chart indicates, in this algorithm, first, the image is binarized (black as background and white as foreground) by means of the threshold extracted from the Otsu algorithm, so that the text components of the image are indicated. Then, the bounding boxes of the words are extracted. In the next step, based on the height of the bounding boxes of all words in the text, a histogram is calculated. The maximum value of the histogram corresponds to the average character height *M*. Those histograms whose heights are more than a.M or less than M/b are removed. In addition, those with a width less than M/c are also removed, so that in the next stage of the algorithm, we do not face a difficulty. In fact, at this stage, small letters and segments are eliminated. Empirically, a, b, and c are selected in a way that undesired text components and noises are removed and only the main words remain.

Further in the procedure, two low-pass filters are used in order to horizontally smooth the words within the bounding boxes. The thresholds of these filters are determined empirically. If the thresholds are considered lower than the determined desirable values, the noisy areas and the redundant letters within the bounding boxes that have a small width and length, will not be removed. In addition, if the determined thresholds are increased, then some of the words that are, in fact, the main components of the texts, will be missed. After the horizontal smoothing, once more, the noisy areas are removed. Then, the remained words are detected. In this stage, through a top-down scanning, the word in the highest point of the page is detected and is labeled as K, to which, this to the first text line *L* is assigned. Next, the words are scanned from right to left and the first word, the word that neighbors at a small distance in the left side of word *K*, is detected. Furthermore, it is assumed that these two words which are closest to each other horizontally, should have something in common vertically too, therefore the word which is detected as the closest one is not selected from the following line. This closest word is labeled as K_f. Continuing our scanning, we got to the last word on the left. The same procedure is followed for all the words on the right side of the first word and label all words found in a left to right scanning order and assign them to the first text line. Furthermore, a flag is enabled for these components to indicate that they will not participate to further calculations and labeling. Subsequently, all of the above procedures were repeated for the remained words till we got to the last word of the image. In this way, all of the words of the image were detected and for each word, the corresponding lines were also determined, which, in turn, are used for detection and correction of the skew of the words, as is explained bellow.

The second stage of the proposed algorithm, includes the determination of the upper and lower baselines for each word which delimit the main body of the words (see Fig. 3), by which, we are able to find the slope of each word. For each word, the upper and lower baseline is defined via a simple linear equation as is shown in equation (1), (2).

$$y = a_{ij}x + b_{ij} \tag{1}$$

$$y = a'_{ij}\, x + b'_{ij} \tag{2}$$

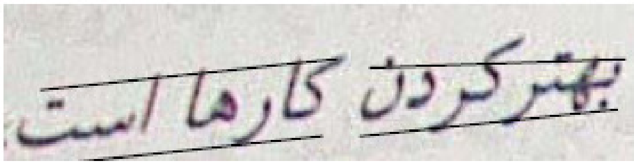


FIGURE 3: Example of upper and lower baseline estimation.

In these equations, the a_{ij}, b_{ij}, a'_{ij}, b'_{ij} coefficients are obtained based on the characteristics of the word. Due to the atypical written forms of the Persian words, and because of the fact that the height of the words does not follow a specific order, it is not possible to estimate the upper and lower baselines for each word correctly, thus it is necessary to consider a number of successive words as one phrase and find the upper and lower baselines for that phrase. This is done through transforming the bounding box of a single word to a bigger one, which encloses a sequence of words. Experimentally, we came up with the conclusion that if we consider each word together with its two preceding words and three following words as a phrase, and regard them in a common bounding box, the baselines will be estimated with better accuracy.

In the next stage, the skew of the words was estimated and also rotated. Here, the slope of each word is derived from the corresponding baseline slopes. Upper and lower baseline slopes of word *Kij* are denoted as:

$$\theta_{ij}^{u} = \arctan(a_{ij}), \theta_{ij}^{l} = \arctan(a'_{ij}) \qquad (3)$$

Using the estimated slope for each word, the angle of that word in relation to the horizontal line was also determined through equation 4. This angle, of course, derived from the comparison we drew between the estimated angles of the upper and lower baselines. This procedure was carried out through equation 4, for the first word of each line, and equation 5, for the other words.

$$\theta_{io} = \begin{cases} \theta_{io}^{ll}, if\ |\theta_{io}^{u}| < |\theta_{io}^{l}| \\ \theta_{io}^{l}, otherwise \end{cases} \qquad (4)$$

$$\theta_{ij} = \begin{cases} \theta_{ij}^{ll}, if\ |\theta_{ij}^{ll} - \theta_{ij-1}| < |\theta_{ij}^{l} - \theta_{ij-1}| \\ \theta_{ij}^{l}, otherwise \end{cases} \qquad (5)$$

Then we re-formed the bounding box of each word and in each bounding box, we transformed those pixels that were translated in relation to the horizontal line, using the following formula (6).

$$\begin{cases} x_{\min} = x \\ y_R = \max(1, round((x - x_{\min}) * \sin(-\theta_{ij}) + y * \cos(\theta_K))) \end{cases} \qquad (6)$$

Estimation and correction of the slope of each word, was carried out based on the slope of previous word in the same line, and when the words in a line finished, this procedure also is ended, and for the new line, the correction of the skews was performed based on the words of the same line and their estimated slopes.

The final stage of the proposed algorithm, consists of transformation of the warped words which was corrected in terms of their slopes. This stage is of paramount importance and, in fact, the main part of the proposed algorithm is related to the transformation of the words with a corrected angle, and sequencing the words in a line, so that all the words in a line are in the same direction and on a horizontal line.

In this part, the criterion is not the maximum value of the bounding box of each word. In fact, here, first the geometric center of all bounding boxes of the words are found, then a polynomial curve of degree two or three is fitted to these centers, using an MMSE scale. In the present study, experiments indicated that the most desirable results will be gained if a quadratic curve is used. As figure 4 indicates, after fitting the curve, the difference between the centers of the bounding boxes with the curve, symbolized as *d* (see Fig. 4), should be estimated and recovery should be carried out with regard to the horizontal line, through the following formula(7).

$$\begin{cases} y_{Rs} = \max(1, (round(y + d))) \\ x_{Rs} = x \end{cases} \qquad (7)$$

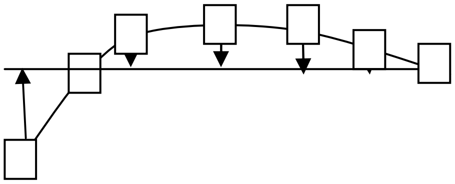


FIGURE 4: Words alignment model.

4. EXPERIMENTAL RESULTS

In this section, the implementation of the proposed algorithm on the Persian document images and the resulting findings are examined. Then a comparison is done between the original algorithm and the one proposed in this paper with regard to their application for the same samples. The samples are captured by a Sony digital camera (S950), with a resolution of 10 megapixels. Totally, a sum of 8 samples were provided and the suggested algorithm is implemented on all of them, the result of which are presented bellow.

Figure 5(a) & 6(a) & 7(a) show three Persian warped samples.

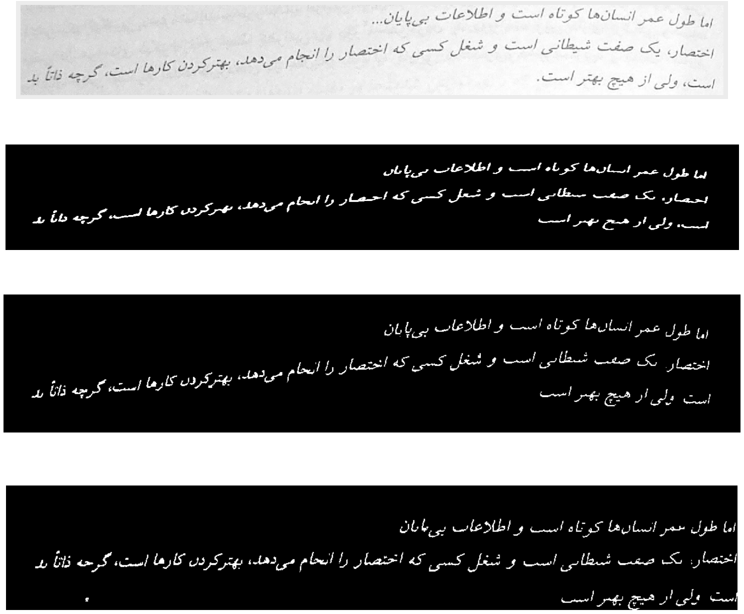
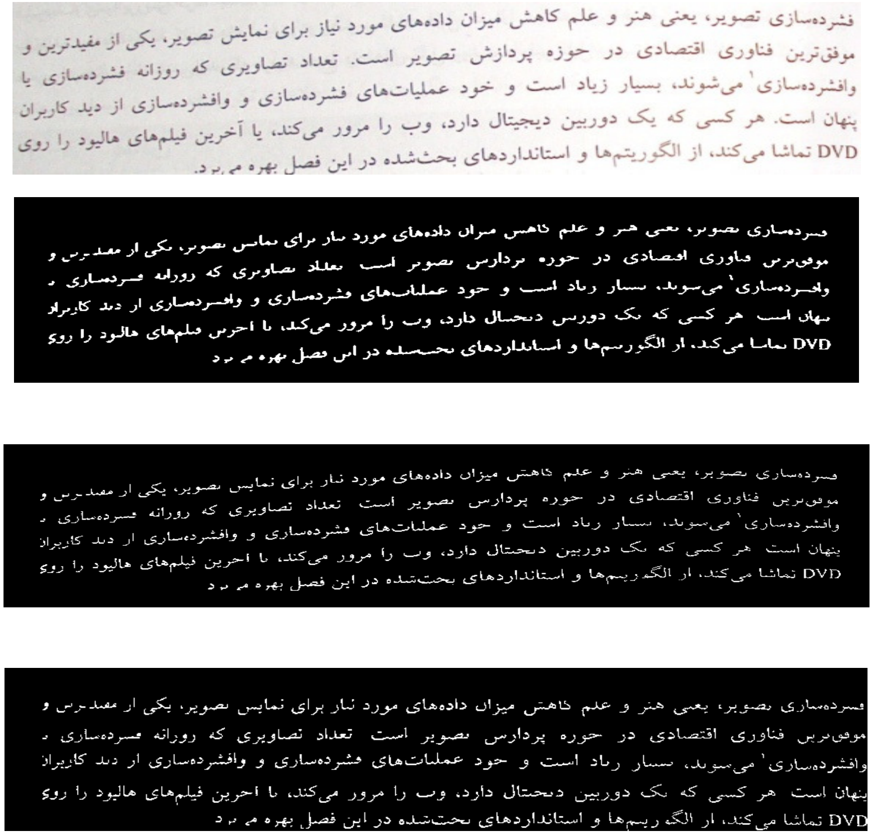


FIGURE 5: from top to down – a. the original image, b. elimination of the noisy areas, determination of the boxes and the text lines, c. estimation and correction of the skews of the boxes, d. transformation of the boxes and the output image of the algorithm.

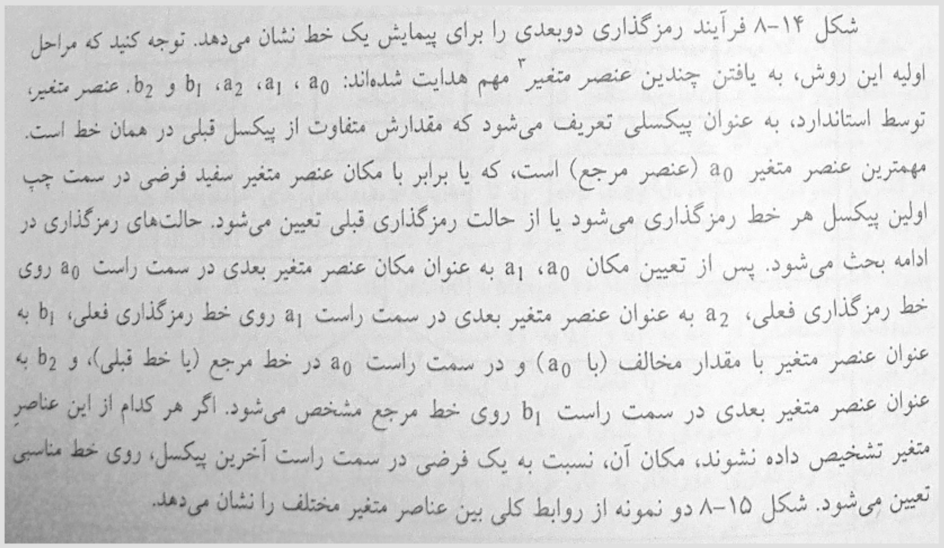
In figure 5, part a, the original document image containing geometric distortion is presented, in b, the noisy areas and the small words are removed, in fact, the points and symbols such as slash, comma, etc. are removed, however, as it can be observed, in the end of second line, a main word is also removed. In part c, for each phrase within a bounding box, its slope is determined and based on the formulas presented in the previous part, the slopes are recovered. In d, the words, for which skew correction was conducted, were transformed. As presented by the figure, this stage was successful for all the three lines, but the dot of the last word of the second line is not places where it should.



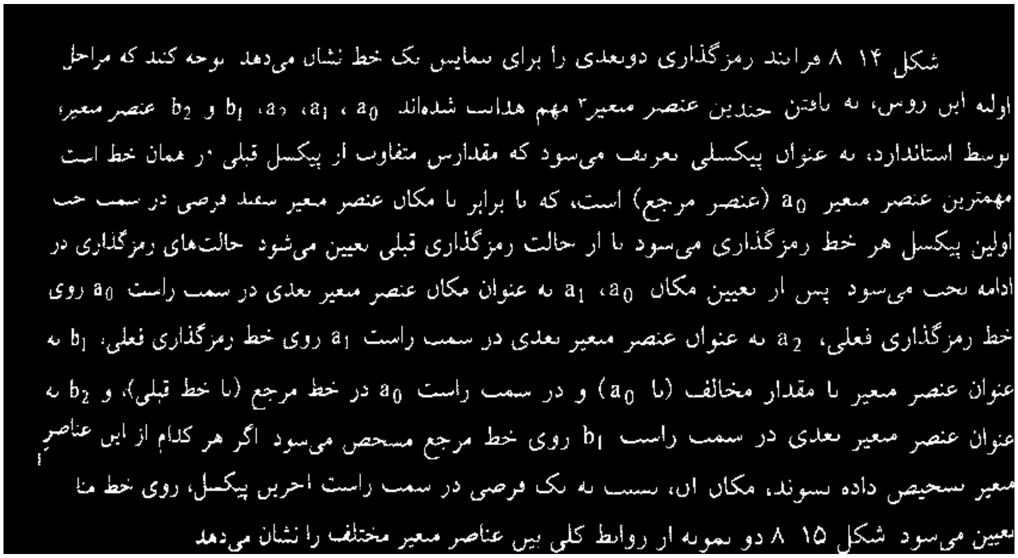
فشرده‌سازی تصویر، یعنی هنر و علم کاهش میزان داده‌های مورد نیاز برای نمایش تصویر، یکی از مفیدترین و
موفق‌ترین فناوری اقتصادی در حوزه پردازش تصویر است. تعداد تصاویری که روزانه فشرده‌سازی یا
وافشرده‌سازی[1] می‌شوند، بسیار زیاد است و خود عملیات‌های فشرده‌سازی و وافشرده‌سازی از دید کاربران
پنهان است. هر کسی که یک دوربین دیجیتال دارد، وب را مرور می‌کند، یا آخرین فیلم‌های هالیود را روی
DVD تماشا می‌کند، از الگوریتم‌ها و استانداردهای بحث‌شده در این فصل بهره می‌برد.

FIGURE 6: from top to down- a, the original image, b. elimination of the noisy areas- determination of the boxes and the text lines, c. estimation and reformation of the skews of the boxes, d. transformation of the boxes and the final output of the algorithm

In figure 6, part b, which is related to elimination of the noisy areas, a limited number of words in the end of the second line as well as those at the end of the last line are missed. In c, the slopes of the phrases are recovered accurately and in d, the transformation process was also successfully conducted.



شکل ۱۴–۸ فرآیند رمزگذاری دوبعدی را برای پیمایش یک خط نشان می‌دهد. توجه کنید که مراحل
اولیه این روش، به یافتن چندین عنصر متغیر[۳] مهم هدایت شده‌اند: a_0 ، a_1، a_2، b_1 و b_2. عنصر متغیر،
توسط استاندارد، به عنوان پیکسلی تعریف می‌شود که مقدارش متفاوت از پیکسل قبلی در همان خط است.
مهمترین عنصر متغیر a_0 (عنصر مرجع) است، که یا برابر با مکان عنصر متغیر سفید فرضی در سمت چپ
اولین پیکسل هر خط رمزگذاری می‌شود یا از حالت رمزگذاری قبلی تعیین می‌شود. حالت‌های رمزگذاری در
ادامه بحث می‌شود. پس از تعیین مکان a_0، a_1 به عنوان مکان عنصر متغیر بعدی در سمت راست a_0 روی
خط رمزگذاری فعلی، a_2 به عنوان عنصر متغیر بعدی در سمت راست a_1 روی خط رمزگذاری فعلی، b_1 به
عنوان عنصر متغیر با مقدار مخالف (با a_0) و در سمت راست a_0 در خط مرجع (یا خط قبلی)، و b_2 به
عنوان عنصر متغیر بعدی در سمت راست b_1 روی خط مرجع مشخص می‌شود. اگر هر کدام از این عناصرِ
متغیر تشخیص داده نشوند، مکان آن، نسبت به یک فرضی در سمت راست آخرین پیکسل، روی خط مناسبی
تعیین می‌شود. شکل ۱۵–۸ دو نمونه از روابط کلی بین عناصر متغیر مختلف را نشان می‌دهد.



شکل ۱۴ ۸ فراىند رمزگذاری دوبعدی را برای سمایس نک خط نشان می‌دهد بوحه کند که مراحل
اوله ابن روس، به ىافتن حندين عنصر معير[۳] مهم هدابت شده‌اند a_0 ، a_1، a_2، b_1 و b_2 عنصر معير؛
بوسط استاندارد، به عنوان پیکسلی بعریف می‌سود که مقدارس متفاوت ار پیکسل قبلی ۰ر همان خط است
مهمترین عنصر معیر a_0 (عنصر مرجع) است، که ىا برابر ىا مکان عنصر معیر سعند فرصی در سمب حب
اولین پیکسل هر خط رمزگذاری می‌سود ىا ار حالت رمزگذاری قبلی بعیین می‌شود حالت‌های رمزگذاری در
ادامه بحب می‌سود پس ار بعیین مکان a_0، a_1 به عنوان مکان عنصر معیر بعدی در سمب راست a_0 روی
خط رمزگذاری فعلی، a_2 به عنوان عنصر معیر بعدی در سمب راست a_1 روی خط رمزگذاری فعلی؛ b_1 به
عنوان عنصر معیر ىا مقدار مخالف (ىا a_0) و در سمب راست a_0 در خط مرجع (ىا خط قبلی)، و b_2 به
عنوان عنصر معیر بعدی در سمب راسب b_1 روی خط مرجع مسحص می‌سود اگر هر کدام از ابن عناصرِ
معیر سحیص داده سوند، مکان ان، سسب به بک فرصی در سمب راست احرىن پیکسل، روی خط منا
بعیین می‌سود شکل ۱۵ ۸ دو بمونه ار روابط کلی بین عناصر معیر مختلف را نشان می‌دهد

FIGURE 7: from top to down- a. the original image, b. final output of the algorithm

Sample	Figure 5	Figure 6	Figure 7	Average of the total samples
Accuracy of the original algorithm	5%	3%	5%	4%
Accuracy of the proposed algorithm	95%	90%	92%	92%

TABLE 1: A comparison between the proposed and the original approaches.

Table 1 indicates the results of applying the suggested algorithm on 8 samples. Greater efficiency of the proposed method is also presented numerically. The results reveal the high applicability of the proposed algorithm presented in this paper.

References should be indicated in the text by consecutive numbers in square brackets, as [1], [2] etc.

5. CONSLUSION & FUTURE WORK

The suggested algorithm is of much efficiency with regard to Persian texts. This algorithm can be implemented on the Persian document images captured by cameras, which have dramatic geometric distortions, as well as photometric distortions. The results of our studies support the effectiveness of this algorithm. This algorithm is generalizable to texts containing formulas and figures as well, about which, researches are being conducted. Besides, reduction of photometric distortions is another area, which can play a complementary role to the practically applications of this algorithm.

6.REFERENCES

[1] J. Liang, D. Doermann, H. Li. *"Camera-based analysis of text and documents: a survey"*. Int. Jour. Of Document Analysis and Recognition, 7(2-3): 84–104, 2005

[2] A. Ulges, C. Lampert, and T. M. Breuel. *"Document capture using stereo vision"*. In Proceedings of the ACM Symposium on Document Engineering, Milwaukee, Wisconsin, USA, 2004

[3] A. Yamashita, A. Kawarago, T. Kaneko and K.T.Miura. *"Shape reconstruction and image restoration for non-flat surfaces of documents with a stereo vision system"*. In Proceedings of 17th International Conference on Pattern Recognition (ICPR) Cambridge UK, 2004

[4] M.S. Brown and W.B. Seales. *"Document restoration using 3d shape: A general deskewing algorithm for arbitrarily warped documents"*. In International Conference on Computer Vision (ICCV), Vancouver, B.C., Canada, 2001

[5] M. Pilu. *"Deskewing perspectively distorted documents: An approach based on perceptual organization"*. In HP Technical Reports, 2001

[6] L. Zhang and C.L. Tan. *"Warped image restoration with applications to digital libraries"*. In Proc. Eighth Int. Conf. on Document Analysis and Recognition, Washington, DC, USA, 2005

[7] A. Ulges, C.H. Lampert and T.M. Breuel. *"Document image dewarping using robust estimation of curled text lines"*. In Proc. Eighth Int. Conf. on Document Analysis and Recognition, Washington, DC, USA, 2005

[8] J. Liang, D.F. DeMenthon, and D. Doermann. *"Flattening curved documents in images"*. In Proc. Computer Vision and Pattern Recognition,San Diego, 2005

[9] A. Masalovitch and L. Mestetskiy. *"Usage of continuous skeletal image representation for document images de-warping"*. In 2nd Int. Workshop on Camera- Based Document Analysis and Recognition, Curitiba, Brazil, 2007

[10] B.Gatos, I. Pratikakis, and K. Ntirogiannis. *"Segmentation based recovery of arbitrarily warped document images"*. In Proc. Int. Conf. on Document Analysis and Recognition, Curitiba, Brazil, 2007

[11] B. Fu, M.Wu, R. Li,W. Li, and Z. Xu. *"A model-based book de-warping method using text line detection"*. In 2nd Int. Workshop on Camera-Based Document Analysis and Recognition, Curitiba, Brazil, 2007

[12] U.V. Marti, H. Bunke. *"Using a statistical language model to improve the performance of an HMMbased cursive handwriting recognition system"*. Int. Jour. of Pattern Recognition and Artifical Intelligence, 15(1): 65–90, 2001

[13] F. Shafait and T. M. Breuel. *"Document Image Dewarping Contest"*. In proc CBDR, 2007