

TEXT LINE SEGMENTATION IN HANDWRITTEN DOCUMENTS BASED ON DYNAMIC WEIGHTS

Costin-Anton Boiangiu¹
Mihai Cristian Tanase²
Radu Ioanitescu³

ABSTRACT

Identification of text lines in documents, or text line segmentation, represents the first step in the process called ‘Text recognition’, whose purpose is to extract the text and put it in a more understandable format. The paper proposes a seam carving algorithm as an approach to find the text lines. This algorithm uses a new method that allocates dynamic weights for every processed pixel in the original image. With this addition, the resulting lines follow the text more accurately. The downside of this technique is the computational time overhead.

KEYWORDS: OCR, text line segmentation, handwritten documents, dynamic weights

1. INTRODUCTION

The process of extracting lines from a document is used as a basis for document structure extraction, handwriting recognition or text enhancement. There are numerous methods ([13], [9], [5]) that address the printed document line extraction problem which is usually reduced to global skew search (the text lines are parallel with each other, but not necessarily horizontal). On the other hand, when dealing with handwritten document the problem becomes more complex ([14], [12], [11], [10], [8], [7], [6], [4]): the lines are not parallel with each other, same letters do not have same sizes, text lines have letters that extend to other text lines, higher text organization cannot be defined (paragraphs, subsections etc.).

In any image, document or not, with handwritten or printed text, each pixel can be associated with an importance (i.e., how much that pixel influences the overall image). In this paper the importance of the pixels in an image is given by the energy map, which is further used to compute the energy cost map and, finally, the seam carving algorithm is used to detect the text lines.

2. RELATED WORK

There are many techniques that address the problem of text line segmentation. They are generally divided into two categories:

- techniques that work directly on the gray scale image

1 Associate Professor PhD Eng, "Politehnica" University of Bucharest Romania, 060042 Bucharest, costin.boiangiu@cs.pub.ro.

2 Engineer, VirtualMetrix Design Romania, 060104 Bucharest, mihaicristian.tanase@gmail.com

3 Engineer, European Patent Office (EPO) Germany, Bayerstrasse 34, Munich, rioanitescu@epo.org

- techniques that use, as input, the binary representation of the image.

For the gray scale images, there are algorithms that use the, so called, projection profiles representing the sum of all the pixels values in a given direction. This method, applied in a pre-computed direction, which usually uses the Hough transform, can accurately identify text lines in printed documents, but fails to produce even moderate results when applied to handwritten documents.

Different workarounds of this problem were proposed: statistical division of the text page, which tries to closely follow the local skew of the handwritten text; smearing methods, which fills the space in-between text characters; Hough transform which selects all lines that have the value accumulated in Hough space greater than a given threshold; repulsive-attractive method in which the current line is attacked by the pixels from the text and is repelled by previous found text lines. For the binary images, there are various grouping methods that can be used in addition to the above mentioned techniques.

In this paper, the energy map of a gray-scale image is used to compute the energy cost map which, in turn, represents the input for the seam carving algorithm. The result is a combination between the original document and lines that follow the text which represents the segmentation boundaries.

Image as an energy map

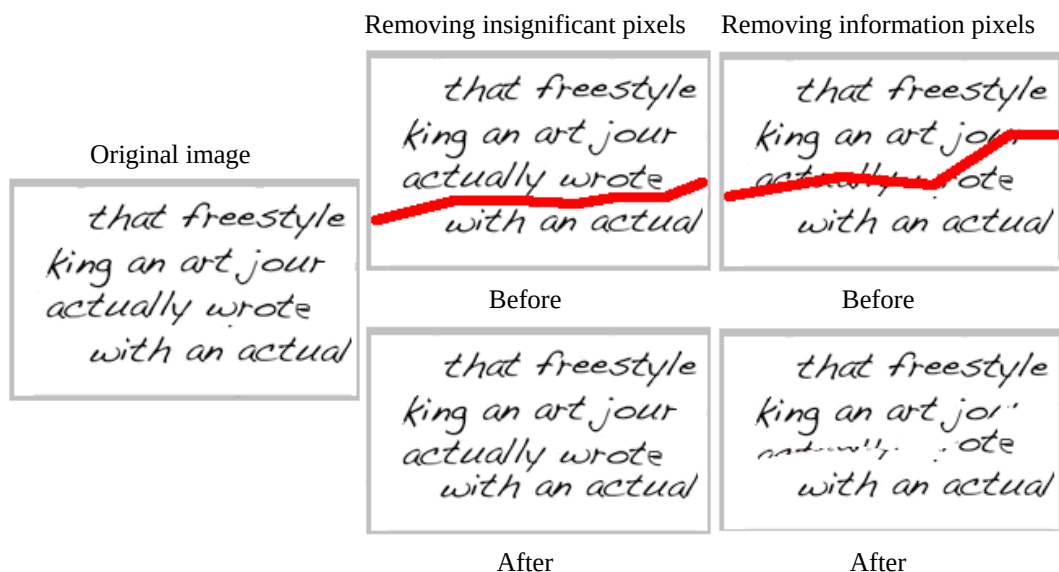


Figure 1 Exemplification of the energy map concept

The energy map of an image represents the information quantity map. Each pixel in the energy map has a value associated with it that represents the amount of information that the given pixel stores in the image. If a high energy pixel is removed from the image, the resulting image has a significant drop in detail, whereas removing a low energy pixel results in a negligible information loss.

Figure 1 illustrates this concept. Removing a set of pixels, each belonging to a homogeneous area will result in almost insignificant information loss compared to extracting the second set of pixels.

From this example, an observation can be made: pixels belonging to homogeneous areas have low energy and pixels belonging to areas with high variations have high energy. This observation leads to the idea of viewing energy map as the derivate of the original image.

In [4], the energy map is computed as the distance transform of the binary image (the objects of interest in the image are the text characters). The output of distance transform represents a map of pixels, each assign with the distance from that pixel to the closest text pixel. For this energy map, high energy represents areas furthest from the text and small energy represents text or areas very close to text.

Three methods are used for calculating the energy map:

$$|\nabla f| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}$$

- a. Magnitude of the gradient, namely: the edge receive higher energies.
- b. Gaussian 1st derivate
- c. Inverse distance transform'

To use distance transform, first, the original image needs to be transformed in a binary image (a binary threshold is applied), than the distance transform algorithm is applied which outputs an image where small values are associated with the text and high values with the background. To have a consistent implementation (i.e., high energy represents high variations), the last image needs to be inverted, which results in an image where the high energy values are associated with text.

3. DYNAMIC WEIGHTS

The next step in the process is to calculate the energy cost map. Similar to [11] and [4], this cost map assigns to each pixel a minimum value calculated with the formula below:

$$M(i, j) = 2 * e(i, j) + \min(w[\text{neighbors_number}/2 + k] * M(i + \text{direction}, j + k))$$

Where $-\text{neighbors_number}/2 < k < \text{neighbors_number}/2$ and direction represents the direction of processing the energy map, which is either left-to-right (+1), or reverse (-1).

To obtain a more accurate energy sum map, both the energy sum maps and linear interpolation the result are computed similar to [11] and the results are displayed in Figure2.

To avoid discontinuous lines, a maximum neighborhood of 3 is used (i.e., neighbor pixels are 8-connected). The weights are constant throughout energy cost calculations. As shown in figures form “Test and Results” section, the results are good to excellent when the documents has close to horizontal lines, but when there lines are skewed, the constant

weight allocation leads to erroneous detection. This is where the idea of dynamic weight allocation comes into play. To accurately follow the text, the direction of the text lines has to be available at each pixel. In other words, for each pixel, the weights are calculated according to local direction of the text line in that pixel.

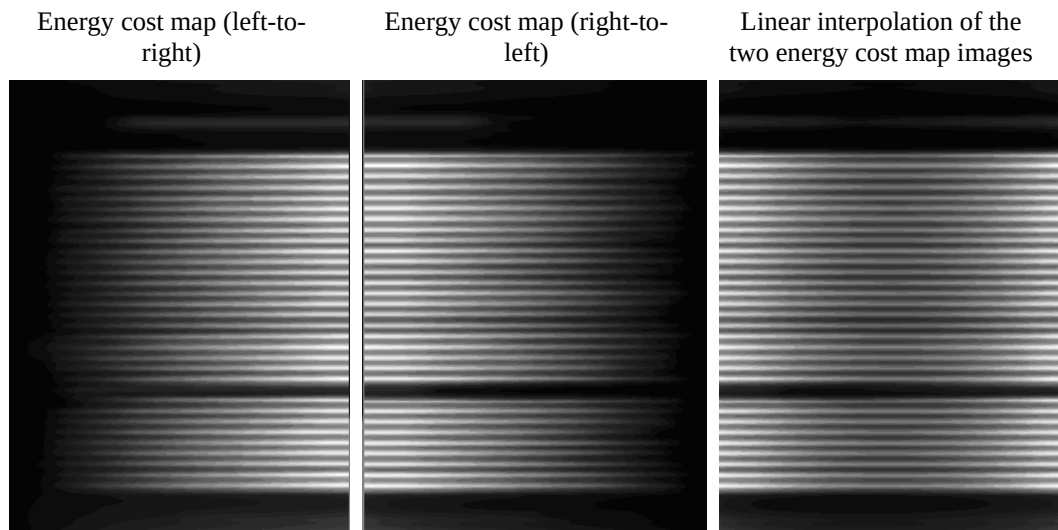


Figure 2. Energy sum map results

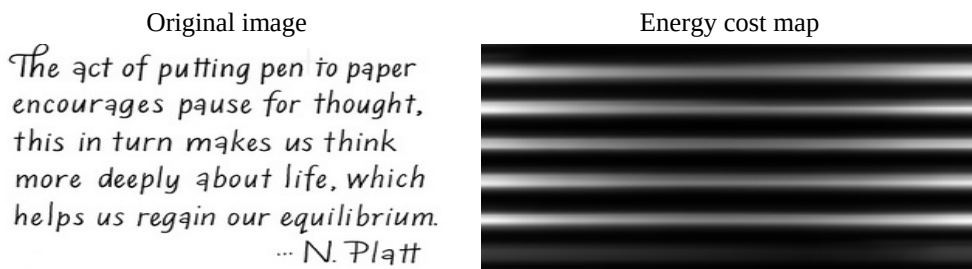


Figure 3. Energy cost map of an image

The algorithm of dynamic weights allocation is presented below.

- Input: original image, window width, window height
- For every pixel:
 - For every variation of the window (by skewing)
 - Sum up the pixels in the new skewed window
 - If the sum is less than the current minimum then
 - Update the current minimum
 - Save the window variation
 - Rescale the saved variation to $[-\text{neighbors_number} / 2, +\text{neighbors_number} / 2]$
 - Calculate the weights giving the rescaled saved window variation

Experimentally, the best results were obtained when the width and height of the scanned window are set to at least the average character width and height values. The minimum and maximum variation angles are obtained empirically: the documents verified showed a maximum skew of approximately 20 degrees. Using this maximum angle, the minimum variation is set to -20 degrees and maximum to 20, leading to a 40 degrees interval, which is more than enough for most documents.

To calculate the weights distance function is used.

$$f(x, x_0, y, y_0) = \sqrt{(x - x_0)^2 + (y - y_0)^2}$$

Because of the way energy cost function is calculated, the formula becomes:

$$f(y, y_0) = \sqrt{1 + (y - y_0)^2}$$

Which is equivalent to:

$$f(x) = \sqrt{1 + x^2}$$

To use the rescaled saved window variation, the formula becomes:

$$f(x, \text{var}) = \sqrt{1 + (x - \text{var})^2}$$

which is equivalent to the geometric translation of variable “x” with “-var” units.

The number of operations necessary is:

$$\text{operations} = w * h * \text{nr_variations} * \text{window_width} * \text{window_height}$$

The last step is to identify the text lines, known as the “seam carving” process. In this step the energy cost map is scanned, searching for minimal values. The algorithm is:

Input: energy cost map having width w and height h, neighbors_number

Output: h lines

For each line (1..h):

For each column (1..w):

P = pixel on right that has minimum cost in the energy cost map and belongs to the neighborhood defined by neighbors_number

Add P to the current carved seam, identified by the line index

If P belongs to a previous found seam, then

Stop this loop

Continue to the next line

4. TESTS AND RESULTS

To test the dynamic weights method, a number of images of handwritten documents have been used. There are three types of tests, each calculating the energy map differently, as described in Section 3.

The database of test input images was consisting of about 500 different types of documents from old letters, library index files, receipts, patents, miscellaneous printed documents (with pronounced skew and complex layout which failed because of that the automatic scanning deskew phase). Despite its relatively small size, the database was considered somehow relevant because of the great variety of input documents types and as a result of the great variety of problems that occurred in the carried automatic batch tests.

Some of the results are discussed below, in order to emphasize the possible solutions, the evolution of the proposed approach, its limitations, strong and weak points and computational costs. One aspect however that cannot be solved using a database of this size is to evaluate the robustness in providing successful handwritten text segmentation in all cases.

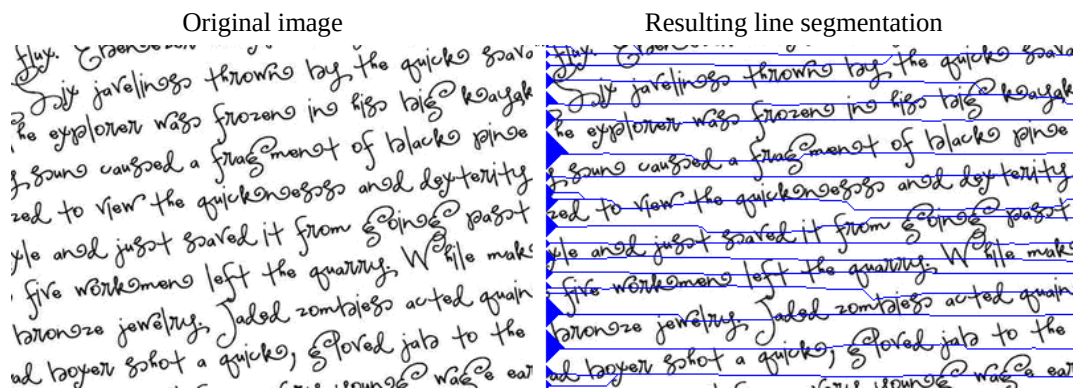


Figure 5. (a) Gaussian 1st derivate using constant weights test 1

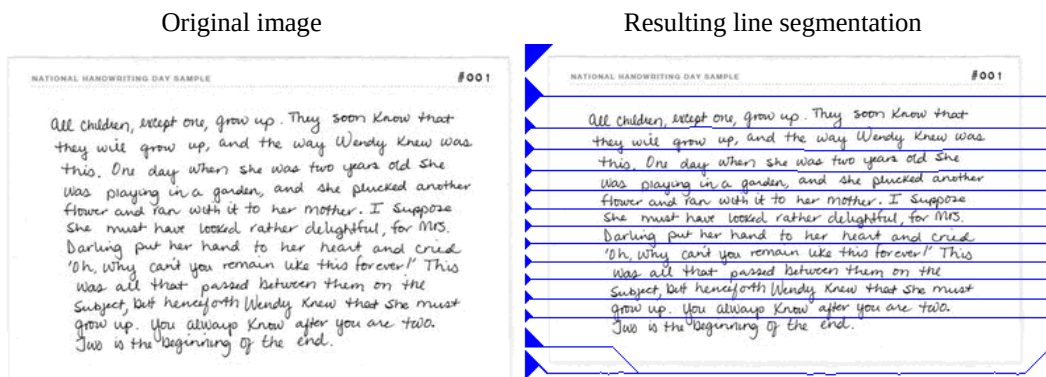


Figure 5. (b) Gaussian 1st derivate using constant weights test 2

Original image

The act of putting pen to paper encourages pause for thought, this in turn makes us think more deeply about life, which helps us regain our equilibrium.
... N. Platt

Resulting line segmentation

The act of putting pen to paper encourages pause for thought, this in turn makes us think more deeply about life, which helps us regain our equilibrium.
... N. Platt

Figure 5. (c) Magnitude of the gradient using constant weights test 1

Original image

flux. Green...
The javelin... thrown by the quick...
the explorer was frozen in his big...
... caused a fragment of black...
... to view the quickness and...
... and just saved it from...
five workmen left the quarry. While...
... acted...
... shot a quick, ...

Resulting line segmentation

flux. Green...
The javelin... thrown by the quick...
the explorer was frozen in his big...
... caused a fragment of black...
... to view the quickness and...
... and just saved it from...
five workmen left the quarry. While...
... acted...
... shot a quick, ...

Figure 5. (d) Magnitude of the gradient using constant weights test 2

Original image

The act of putting pen to paper encourages pause for thought, this in turn makes us think more deeply about life, which helps us regain our equilibrium.
... N. Platt

Resulting line segmentation

The act of putting pen to paper encourages pause for thought, this in turn makes us think more deeply about life, which helps us regain our equilibrium.
... N. Platt

Figure 5. (e) Inverse distance transform using constant weights test 1

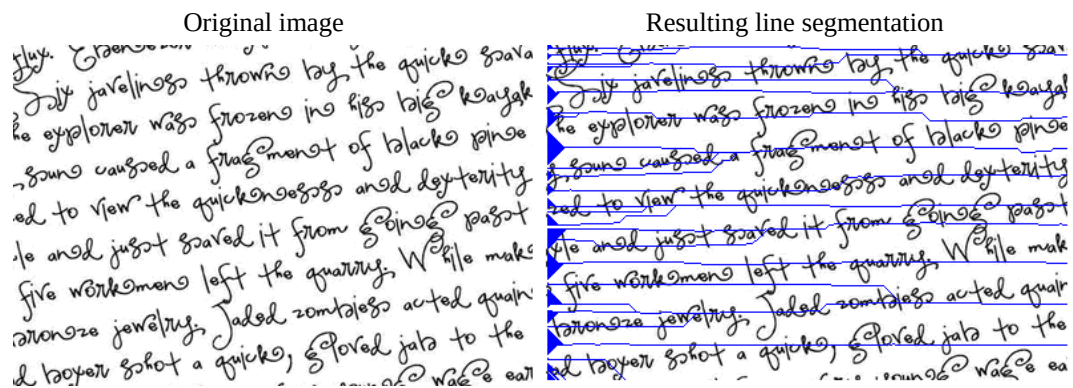


Figure 5. (f) Inverse distance transform using constant weights test 2

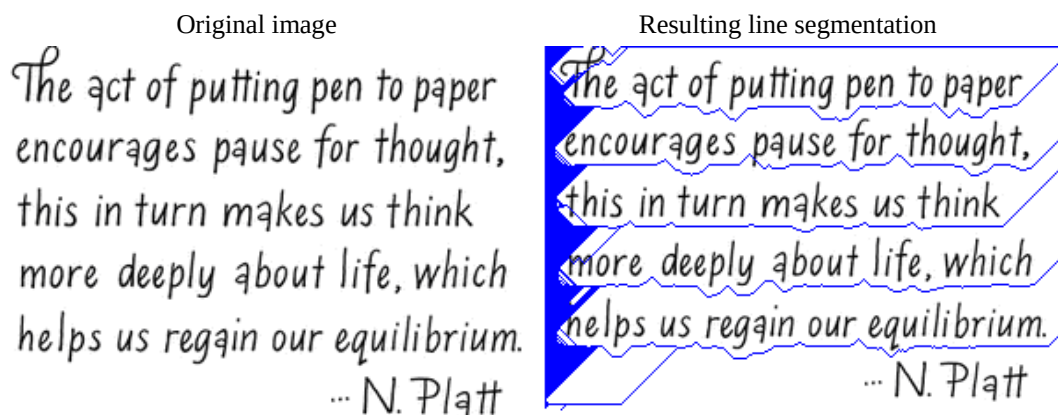


Figure 5. (g) Gaussian 1st derivate dynamic weights test 1

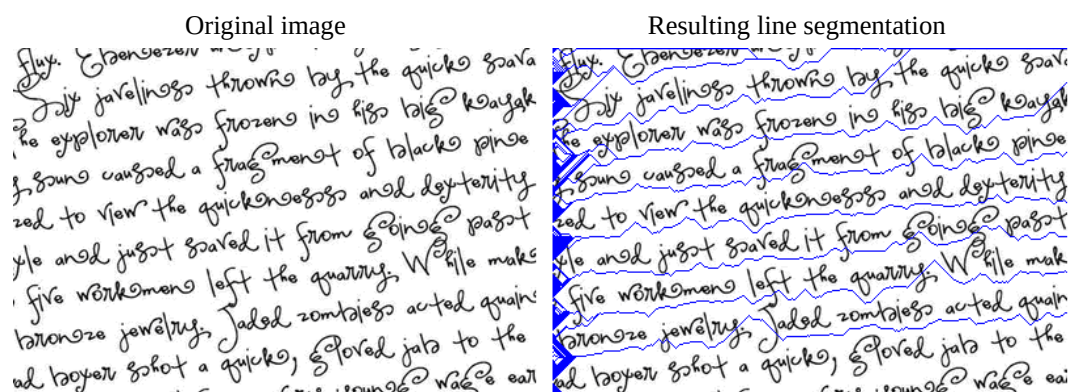


Figure 5. (h) Gaussian 1st derivate dynamic weights test 2

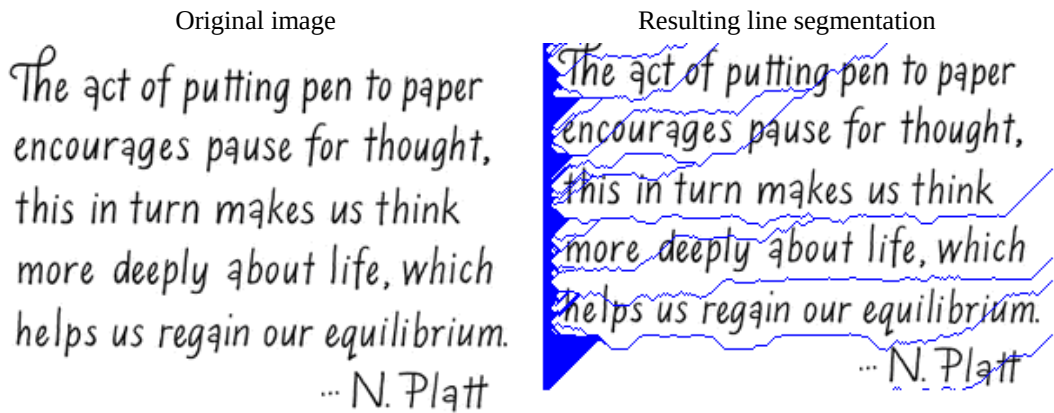


Figure 5. (i) Magnitude of the gradient using dynamic weights

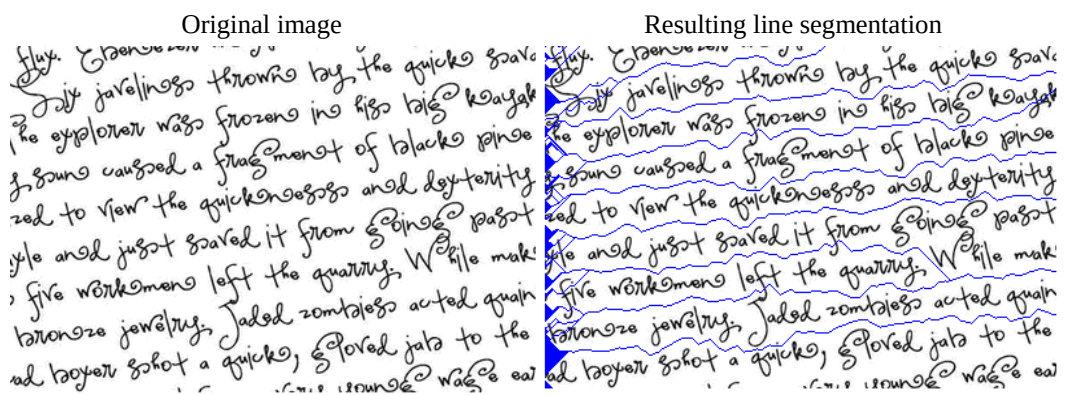


Figure 5. (j) Inverse distance transform using dynamic weights

5. CONCLUSION AND FUTURE WORK

This paper describes an improvement to text line segmentation for handwritten documents based on seam carving using energy map. Experimental results showed that using constant weights when calculating the energy cost map leads to bad results when the text lines are skewed. This is the reason a dynamic allocation of weights based on the local text direction information is proposed. This way, pixels from the same line have higher probability of selection when calculating the minimum values in energy cost map.

As mentioned in the beginning of this paper, the dynamic weight calculation has very high computational cost. One method to reduce this complexity is to set smaller parameter values for the number of variations, window width and height. Other methods of optimization can also be considered and applied in conjunction. Also, the seam carving algorithm can be improved by constraining line curvature, which can solve the problem in the picture “Inverted distance transform – Dynamic weights”, in which a seam passes through the space between two words and intersects another seam.

The work presented in this paper is a building block of a much bigger project: a complete, modular, fully automatic content conversion system developed for educational purposes. In

the near future, with the completion of the system and the running in automatic batch processing of large image databases of all kind of skewed documents (containing handwriting or not) the dynamic weights method will be fully evaluated in order to assess its real potential as a preprocessing phase for OCR applied on handwritten documents.

6. REFERENCES

- [1] L. Likforman-Sulem, A. Zahour, B. Taconet "Text Line Segmentation of Historical Documents: a Survey", *International Journal on document analysis and recognition*, vol. 9, no. 2/4, pp 123/138, 2007.
- [2] R. Strand, F. Malmberg, S. Svensson, "Minimal Cost-Path for Path-Based Distances", *Image and Signal Processing and Analysis*, pp. 379-384, 2007.
- [3] S. Avidan, A. Shamir, "Seam Carving for Content-Aware Image Resizing", *ACM Siggraph*, article 10, 2007.
- [4] R. Saabni, J. El-Sana, "Language-Independent Text Lines Extraction Using Seam Carving", *Document Analysis and Recognition (ICDAR)*, pp. 563-568, 2001.
- [5] R. P. dos Santos, G. S. Clemente, T. I. Ren and G. D.C. Calvalcanti , "Text Line Segmentation Based on Morphology and Histogram Projection", *Document Analysis and Recognition (ICDAR)*, pp. 651- 655, 2009.
- [6] V. Papavassiliou, T. Stafylakis, V. Katsouros, G. Carayanni, "Handwritten document image segmentation into text lines and words", *Pattern Recognition*, vol. 43, no 1, pp. 369-377, 2010.
- [7] X. Du, W. Pan, T. D. Bui, "Text Line Segmentation in Handwritten Documents Using Mumford-Shah Model", *Pattern Recognition*, vol. 42, no 12, pp. 3136-3145, 2009.
- [8] N. Tripathy and U. Pal , "Handwriting segmentation of unconstrained Oriya text", *Frontiers in Handwriting Recognition*, pp. 306-311, 2004.
- [9] S. Saha, S. Basu, M. Nasipuri and D. Kr. Basu, "A Hough Transform based Technique for Text Segmentation", *Journal of Computing*, vol. 2, no. 2, 2010.
- [10] D. J. Kennard, W. A. Barrett, "Separating Lines of Text in Free-Form Handwritten Historical Documents", *Document Image Analysis for Libraries*, pp. 12-23, 2006.
- [11] Asi, R. Saabni, J. El-Sana, "Text Line Segmentation for Gray Scale Historical Document Images", *Proceedings of the 2011 Workshop on Historical Document Imaging and Processing*, 120-126, 2011.
- [12] Bar-Yosef, "Input sensitive thresholding for ancient Hebrew manuscript", *Pattern Recognition Letters*, vol. 26, no. 8, pp. 1168-1173, 2005.
- [13] M. Arivazhagan, H. Srinivasan and S. Srihari, "A Statistical approach to line segmentation in handwritten documents", *Proceedings of SPIE*, 2007.
- [14] I, Bar-Yosef, N. Hagbi, K. Kedem, I. Dinstein, "Line segmentation for degraded handwritten historical documents", *Document Analysis and Recognition*, pp. 1161-1165, 2009.