

Text Separation from Graphics by Analyzing Stroke Width Variety in Persian City Maps

Ali Ghafari-Beranghar

Department of Computer Engineering, Science and Research
Branch, Islamic Azad University,
Tehran, Iran

Ehsanollah Kabir

Department of Electrical and Computer Engineering, Tarbiat
Modares University,
Tehran, Iran

Kaveh Kangarloo

Department of Electrical Engineering, Islamic Azad University,
Central Branch, Tehran, Iran

Abstract—Text segmentation is a live research field with vast new areas to be explored. Separating text layer from graphics is a fundamental step to exploit text and graphics information. The language used in the map is a challenging issue in text layer separation problem. All current methods are proposed for non-Persian language maps. In Persian, text strings are composed of one or more subwords. Each subword is also composed of one to several letters connected together. Therefore, the components of the text strings in Persian are more diverse in terms of size and geometric form than in English. Thus, the overlapping of the Persian text and the lines usually produces a complex structure that the existing methods cannot handle with the necessary efficiency. For this purpose, the stroke width variety of the input map is calculated, and then the average line width of graphics is estimated by analyzing the content of stroke width. After finding the average width of graphical lines, we classify the complex structure into text and graphics in pixel level. We evaluate our method on some variety of full crossing text and graphics in Persian maps and show that some promising results in terms of precision and recall (above 80% and 90%, respectively) are obtained.

Keywords—Document image analysis; text/graphics separation; stroke width; raster map; Farsi; Persian; text segmentation; text label

I. INTRODUCTION

Text extraction is a fundamental task in graphical document image analysis. This problem frequently occurs in many applications like the map, form processing and engineering drawing interpretation where text and graphics are processed in mainly different ways [1]-[20]. Text and graphics are usually separated for later analysis and recognition; indeed, text recognition is completely different from graphics recognition in general. Current OCR systems cannot recognize text labels in complex mixed text and graphics. Moreover, both government and business organizations must frequently

convert existing paper maps of raster maps into a machine-readable form that can be interfaced with the current geographical information systems (GIS) or optical character recognition (OCR).

Despite the many studies that have been reported on the text layer extraction from the map, there does not exist any study on the effect of Persian language in the map processing research area [15]. Recent research suggests that language is one of the influential factors in the process of extracting text layers from a map [18]. In Persian, unlike English, one or more characters can be connected to each other and create a subword. In fact, the word (text string) consists of one to several subwords. Therefore, the components of text strings in Farsi have variable dimensions. On the other hand, the overlapping of subwords and lines creates a large and complex structure as shown in Fig. 1.

On the other hand, there are very rich and diverse sources of maps in which the information they contain is valuable. Information from past land and geographic areas that remain untouched without study. Therefore, in order to extract the Persian text layer from the map of the study, it is necessary to design a new method to solve these complex overlapping text and graphical lines. This work will open up a way for further study on the researchers in this area.

In this paper, we provide stroke width-based approach, which is a local descriptor of text, in urban maps that contain a wealth of text labels. Due to the complexity of the text and lines in these types of maps, macro-micro features are combined to separate the subwords and lines.

The rest of this paper is organized as follows: In Section 2, the proposed method is described in detail. And the experimental results and analysis of the performance are given in Section 3. Finally, the concluding remarks are given in Section 4.

A. Related Works

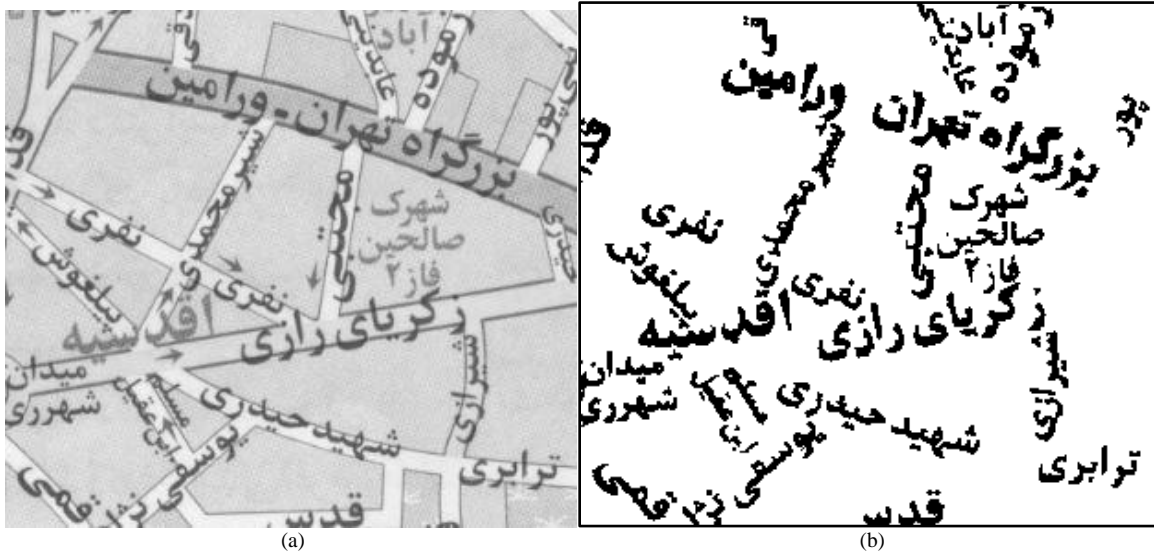


Fig. 1. Part of street map of Tehran, b) Persian Text layer of map.

In this section, we review some related studies in the literature. Chiang has presented an extensive survey in map processing area [15]. Fletcher and Kasturi [1] proposed a method for text layer segmentation. The main assumption is based on not touching text and graphics. By analyzing the size and geometrical features of objects, text components are separated from graphical lines. The approach works well for simple maps in which text and graphics are not touched together. Such simple maps are not so common in varieties of applications like city maps in which text and graphical lines fully overlapped each other. In contrast, we propose the method not need to satisfy these limitations. In complex maps, these assumptions cannot be practical.

Cao and Tan [4] propose a method based on the assumption that simple touch occurs between a limited number of characters of a text string and associated lines. This method uses a thinning process of the input image to detect the region intersection. Then using some heuristic rules, the lines and the text are separated from each other. This method is able to separate a simple overlapping pattern between lines and text. Especially they assume that text characters are separate. In contrast with our method, we do not assume that the characters are necessarily separate from each other. In addition, our method can deal with complex patterns of overlapping text and lines as in Persian text words and lines in high-density city maps.

Tombre et al. [5] proposed a method to consolidation Fletcher and Kasturi's method. They assume that some characters of the text must be non-touched. Based on non-touching characters to lines, it looks for characters associated with graphic lines. Using skeletonization of big line structure, some heuristics rules are used to classify parts of the structure into text stroke and line segments. Tomber's method can be used in simple maps based on finding a major number of characters in a text string, while in Persian language maps, these conditions cannot be met. In the case of Persian text

strings, the characters can be joined to each other, and it can not necessarily be assumed that a significant portion of the characters of the text can be separated from the associating lines, especially in complex maps with a high rate of text and lines overlapping as it is common in Persian city maps.

Cheng and Liu [6] proposed a method based on the assumption that a line as an interferential curve in the text image must be detected and then separated from it. The graph representation of input image is obtained using thinning process. Then, the shortest path algorithm is used to detect interferential curve and removes it from the input image. The text layer obtained from this process is an image thinned in which text quality has greatly reduced. The main limitation of this method is the separation of the lines depending on the thinning of the whole image, so the text image quality from the perspective of text recognition process is controversial. In addition, the method is assumed that the length of the lines should be greater than a predefined threshold and the curvature gradient should be smooth. Therefore, application of the method is not practical, especially in real maps in which these assumptions are not satisfactory. Compared to our proposed method, there are no limiting assumptions on the shape of lines or gradients, as well as line length.

Zhong's method [7] is based on the approximation of text and graphics intersection region. The intersection region is approximated by a polygonal shape. They detect the intersection of text stroke and graphical lines objects using heuristics rules. In complex text and graphics overlapping, the method has complex intersection regions and their heuristics rules cannot be applied efficiently to detect membership of intersection region. In contrast with our method, Zhong's method detects intersection region of text and lines, while our approach focuses on analyzing the complex big line structures that occur usually in complex Persian maps.

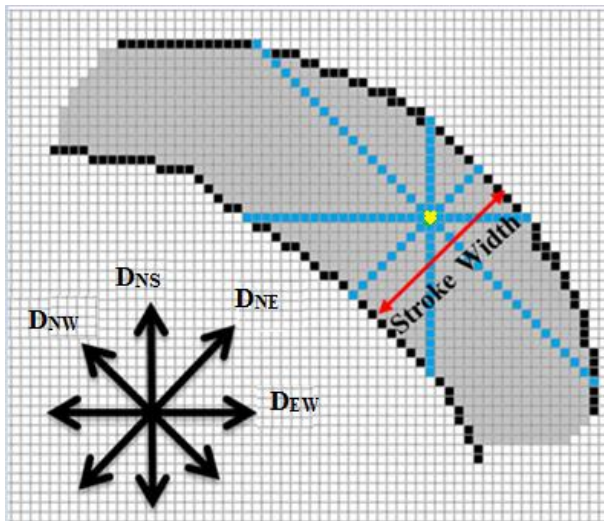


Fig. 2. Stroke width in the yellow pixel has been shown by the red double arrow as the smallest distance between the four directions passing through in the yellow pixel: DEW as east-west, DNS as north-east, DNE as north-east direction, and DNW as north-west direction.

Luo and Kasturi [8] designed some directional morphological operator to extract linear features from maps. Their method can separate lines touched to text. However, the process of morphological is based on manual iteration design. So the method is the difficult approach practically to apply and is dependent on expert knowledge of maps processed. In contrast, our method is not dependent on operator's effort to the analysis of maps.

Li et al. [9] proposed an OCR-based method by training prototypes of characters separated by an operator as training data using template matching approach. They assume some characters of text layer in the map are separable before completely text layer separation. Their approach is language and font dependent and requires user involved in text layer separation. In contrast, our method is not dependent on a specific font.

Tofani and Kasturi [10] proposed a method based on a priori knowledge about the text color of maps. So it is the map dependent method and it cannot apply to some variety of non-color maps. They assume text layer has two major colors, Black and Pink. So they experimentally by color thresholding separate text layer for color maps, then text separated from the graphical line. They find text region and then using line tracking, based on line thickness all pixels of line tracked are removed from the image. They assume that line width is constant and color of text layer is known. However, in contrast, we do not assume that the color of text layer is known.

Chiang and Knoblock [11] propose a user-centric approach to separate text layer of color map by some example as a sample text region or non-text sample to recognize text and non-text color. So based on these text colors found, text layer is extracted from map. In contrast, we propose a new method to extract text layer automatically, to generalize text layer separation from color map, complex text and graphics

overlapping is consider using a micro level feature. In contrast, our proposed method is automatic.

Roy et al. proposed a method [12] and a system [13] based on color in raster maps. In [4], the color segmentation is applied to large graphics components to separate into smaller parts that differ in color. This method can be used for maps that have both high-level quality print and the color of text and graphics distinct from one another while color feature cannot be used in many maps, especially the historical maps. In Addition, the color of text and graphics cannot be distinguished even in color maps. In contrast, our proposed method is independent of the colors of the map. In [5], they perform the color segmentation to get different color layers. They assume that the dominant colors represent the layers of the map. This method works well for maps in which color of text layer is completely different from other layers of the map. However, this assumption is not common in all raster maps.

Levachkine et al. [14] proposed a global dynamic thresholding method to convert the RGB color map into the binary image to detect foreground layer of the map. The text layer is separated from the obtained binary image using connected component analysis. In fact, they use color as a global feature to segment foreground layer, they assume that binarization map is the suitable approach for foreground layer separation. This method can work on some maps, but in maps where the color contrast between the text layer and the rest of the layers is low, it causes loss of parts of the text.

Velázquez [15] proposed global thresholding method on the linear combination of RGB color space components (R, G, B) to obtain a binarized image map. Then V_line technique is used to detect a bounding box on the text string to feed to OCR that is trained for synthetic characters with help of Gazetteer (place names of geographical names list). They assume that some major parts of text strings are not touching to graphics. So the method can separate text string that only a limited number of characters are in touch with the associated line. In addition, the method is language dependent and it requires user's effort for the training of synthetic characters.

In contrast with above approaches [10], [12]-[15], our method is not dependent on the color of maps. It can handle the complex overlapping text and graphics situation. In addition, we do not assume about the shape or line with or line pattern to be separated from text objects as well as the isolation of characters, which is a hypothesized in the above approaches.

B. Local Stroke Width Feature

Stroke width of the text is defined as the distance between two parallel edges of a stroke. Fig. 2 shows part of a stroke in an image and the stroke width associated with it by a red double arrow in the yellow pixel of stroke [16].

II. PROPOSED METHOD

The block diagram of the proposed method has been shown in Fig. 3. Details of each step have been explained in the following sections.

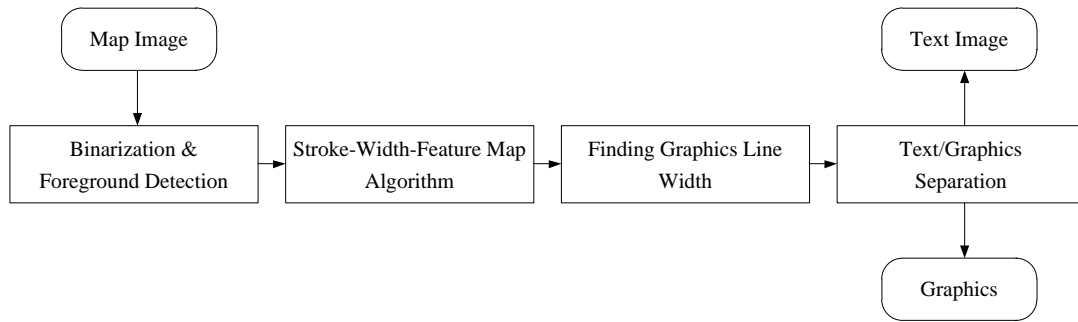


Fig. 3. The block diagram of the proposed method.

A. Binarization and Foreground Detection

Text and graphical lines are in the foreground layer of the map. So, it is used the best known local adaptive binarization method proposed by Sauvola [17] to detect foreground pixels. The Sauvola's method accurately keeps text labels and lines as foreground pixels while removing the background pixels. So, empirically, by setting binarization parameters the foreground of the input map image is detected properly to be used for further analysis. Without loss of generality, it is assumed that the foreground and background pixels have been assigned by '1' and '0' respectively.

B. Stroke-width-feature Map Algorithm

In this section, we present an algorithm for generating stroke width feature map. The Epshtein's method [18] was simplified to calculate local stroke width for each foreground pixel [16]. In Fig. 2 the local stroke width of the foreground yellow pixel has been shown by a red double arrow. So we will obtain a stroke-width feature map as an image of the equal size in which each pixel content has a stroke width. The stroke-width feature map is obtained using algorithm 1 as follows: for each foreground pixel, four distances are calculated. Each distance is the length of the line segment passing through the pixel in four directions of north-south, east-west, north-east and north-west. The smallest distance is obtained as the local stroke width of this foreground pixel.

ALGORITHM 1. Create SW-feature map

Input: BW, binary image of input map image

Output: SW image with the same size as of BW

$SW = \text{Zeroes (size of BW)}$;

For each foreground pixel of BW, p_i in position of (x, y)

Calculate Four Distances as Follows:

$D_{EW}^i =$ number of adjacent connected pixels p_i on the east-west direction until reaching to a background pixel

$D_{NS}^i =$ number of adjacent neighbor pixels p_i in north-south direction until reaching to a background pixel

$D_{NE}^i =$ number of adjacent neighbor pixels p_i in north-east direction until reaching to a background pixel

$D_{NW}^i =$ number of adjacent neighbor pixels p_i in north-west direction until reaching to a background pixel

$$SW(x, y) = \min(D_{EW}^i, D_{NS}^i, D_{NE}^i, D_{NW}^i);$$

End

C. Finding Graphic Line Width

After obtaining a stroke-width feature map, the histogram of this feature map is calculated as follows: for each stroke width value in the feature map, we count the number of pixels which have the same stroke width. So, stroke width distribution is found as a histogram. Fig. 4(c) shows the stroke width histogram of the overlapped text and graphics shown in Fig. 4(a). The histogram shows the stroke width variety which is available in the mixed text/graphics input image. Obviously, there is two major stroke widths in this text image: the first stroke width belongs to graphical line overlapped with the text, and the second one belongs to the text. Therefore, the histogram shows the stroke width content of the input image.

By analyzing the stroke-width histogram we can find dominate stroke width of graphic lines and text labels on the map. In city maps, it is observed that the graphic lines on the map are finer than the text associated with it. So, it can be found the most frequent small stroke width as an estimation of the average width of graphic lines. For example, in Fig. 4(c) the smallest dominant stroke width appears at 2. So, it can be concluded that this stroke width is the average graphic line width of the input image.

D. Text/Graphics Separation

By analyzing the histogram of stroke width, the average width of graphic lines can be found. Since in city maps the widths of graphic lines are nearly fixed, so this estimation is used as the threshold of the text/graphics. In the stroke-width feature map, each pixel has the local stroke width associated with that point, so this threshold can be used to classify any point into two classes: graphics and texts. In addition, we observe that in maps usually, the widths of graphic lines are finer than the text stroke width. Therefore, we can classify each pixel of the stroke-width feature map using this threshold. For each point in stroke-width feature map, if pixel value, i.e., local stroke width, is less than the threshold, then it is classified as graphics; otherwise, it is recognized as text. So, two images will be obtained as follow: one for text and the other for graphics. For each pixel that is classified as text, it is assigned a '1' in the corresponding position in a new image (whose size equals the feature map image), and for each pixel classified as graphics, it is assigned '0' (in the same new image).

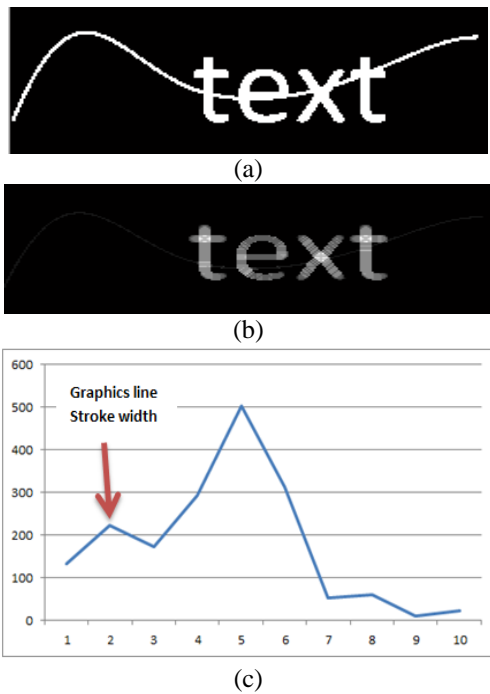


Fig. 4. a) text and overlapping line image, b) corresponding SW image, c) Stroke width variety histogram.

After classification of pixels into text and graphics, in the text layer, some points appear as noise. Size filter is applied to connected components of text layer extracted to remove them.

III. EXPERIMENTAL RESULTS

A. Data Set

To the best of our knowledge, there exists no standard published data set on maps to evaluate our method. So, for our experiments, we gathered 5 real Persian city map images scanned at 300 dpi from sources like major map publishers, Sahab Geographic and Drafting Institute [19] and National Cartographic Center (NCC) [20]. The most important and dominant characteristics of these collected maps are full crossing where text and graphics are overlapping each other. In some collected city map images, multiple graphic lines have overlapped the text labels, as shown in Fig. 8(a). Graphic line patterns differ due to map publisher styles in map production processes. In some maps, lines are continuous (see Fig. 7(a)) and in some others are dotted or line-segment-dotted, as shown in see Fig. 6(a).

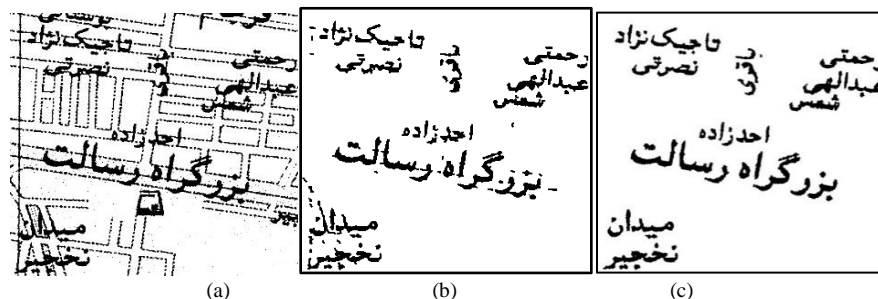


Fig. 5. a) Part of map from data set, b) Text layer extracted by proposed method, c) Corresponding grand truth; the fragmentary text labels in borders have been removed.

B. Evaluation Methodology

In this section, we evaluate our method on a collected data set. The results of the proposed method are shown on some varieties of collected city maps. To show quantitative performance evaluation, the common standard metrics like precision, recall, and f-measure have been measured using the corresponding map ground truth illustrated in the following section:

$$Precision = \frac{T_p}{T_p + F_p} \quad (1)$$

$$Recall = \frac{T_p}{T_p + F_n} \quad (2)$$

$$F_{measure} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (3)$$

Here T_p is the true positive result, F_p is the false positive and F_n is the false negative result. To measure these factors, text ground truth of map was used as a true reference for the result and the parameters are defined as follow:

T_p : The set of text image pixels that is confirmed by the corresponding text ground truth

F_p : The set of text image that is not confirmed by the text ground truth

F_n : The set of text ground truth that is lost

$F_{measure}$: The harmonic mean of precision and recall

C. Ground Truth of Text Layer

We manually provided text ground truth for some of the maps of our data set to evaluate quantitatively efficiency of our proposed method. At first, the map images were converted to binary images using Sauvola's. Then, we manually removed all the non-text pixels from the map like symbols, graphic lines and background texture that occur in binarization process. In addition, the fragmentary text labels in the map borders are neither true text nor graphic so these objects were also cleared. Fragmentary texts have been removed in the corresponding original map, as shown in Fig. 5(c).

Based on the results, the proposed method can extract the Persian text from a big line structure without dependence on the type of lines or the pattern shape of the lines.

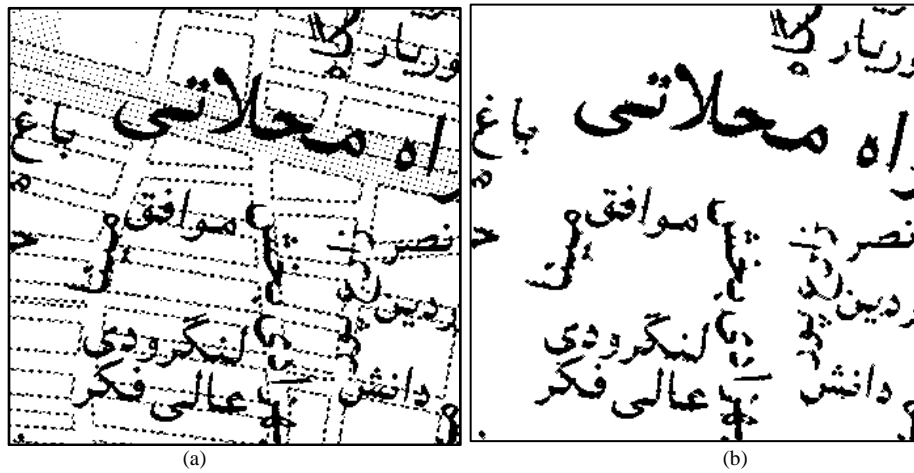


Fig. 6. a) Part of Tehran map, b) Text layer extracted by proposed method.

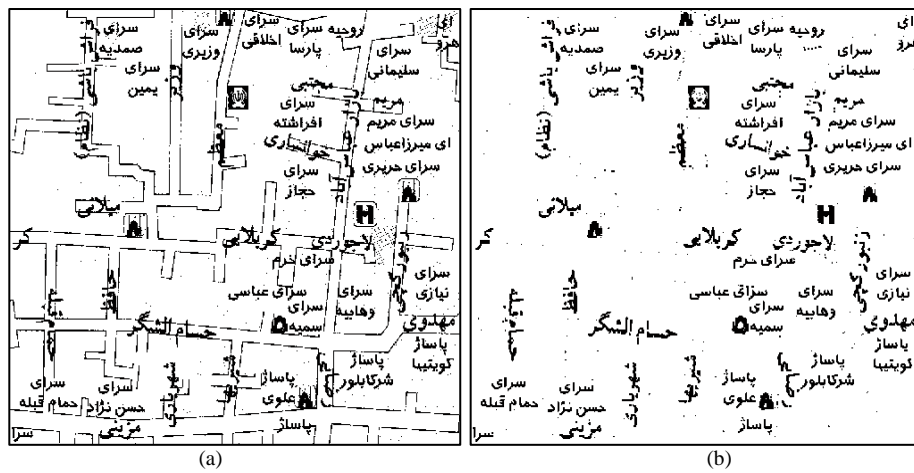


Fig. 7. a) Original map, b) Text layer extracted by proposed method.

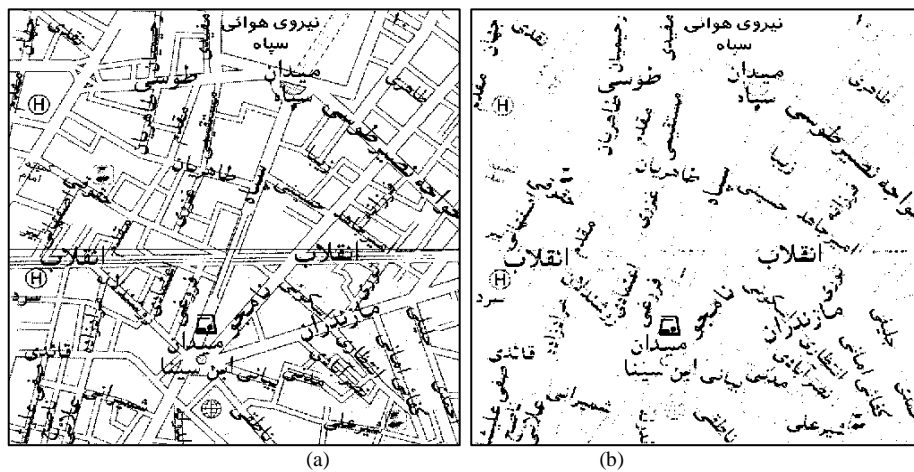


Fig. 8. Original map, b) Separated text layer by proposed method.

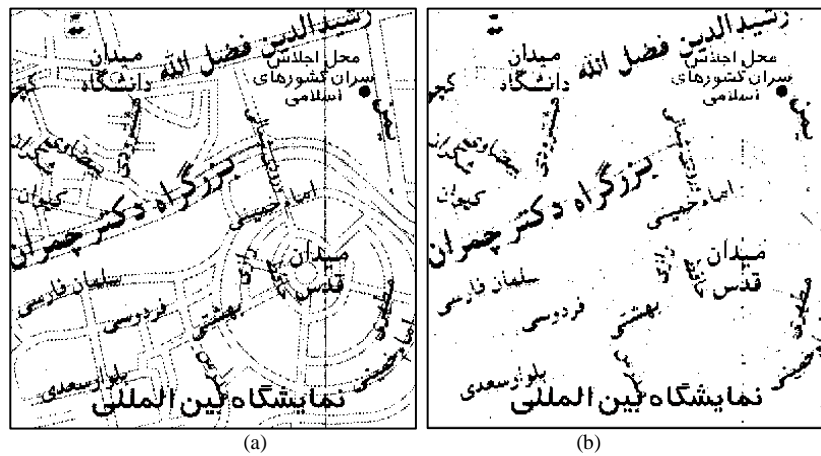


Fig. 9. a) Original street map, b) Text layer extracted by proposed method.

Corresponding text layer grand truth each input image as the evaluation reference was obtained. The quantitative evaluation results are obtained in Table I. The overlapping lines and text have created a whole unit, as shown in Fig. 5(a)-9(a). The proposed method categorizes the foreground pixels into text and graphics according to the average estimated thickness of line structure in the stroke width feature space.

TABLE I. THE PERFORMANCE EVALUATION OF PROPOSED METHOD

Map Image	Precision (%)	Recall (%)	Fmeasure (%)
Fig. 5(a)	79	96	84
Fig. 6(a)	99.6	98.9	99.3
Fig. 7(a)	86.4	91.4	88.8
Fig. 8(a)	86.5	89.6	88
Fig. 9(a)	92.6	92.1	92.3

Based on the results, the proposed method is able to separate the Persian text layer in different situations such as the continuous lines, as shown in Fig. 7(a), the high density of the lines, and even the overlap of several lines from the entire text, as shown in Fig. 8(a), or the curved lines as shown in Fig. 9(a).

The main limitation of the works is the cases where the quality of the map is low or the text with fine fonts, the quality of the extracted text is low. Also, in cases where lines have a high thickness than the average lines width estimated, the proposed method requires redesign

IV. CONCLUSION

All previous methods focus on extraction of English text string from maps. This paper proposes a new method for text layer extraction in Persian language maps. Our method uses local shape feature, stroke width, to separate text from graphics. Therefore, our method does not depend on the geometric shape of graphical lines, i.e. line length, smoothness, and line pattern as these are the important factor in most previous methods. In addition, it suggests the possibility of adding some features like intensity or color features of map objects to improve the performance of the method. This method can handle complex overlapping text and graphics in high-

density maps. However, there is still much work to do. The proposed method cannot deal with poor quality of text in which text stroke width is near the average line width of graphical lines, so new methods should be designed further.

REFERENCES

- [1] Fletcher, L.A. and R. Kasturi, A robust algorithm for text string separation from mixed text/graphics images. *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, 1988. 10(6): p. 910-918.
- [2] Sauvola, J., et al. Adaptive document binarization. in *Document Analysis and Recognition, 1997.*, Proceedings of the Fourth International Conference on. 1997.
- [3] Luo, H. and R. Kasturi, Improved directional morphological operations for separation of characters from maps/graphics. *Lecture Notes in Computer Science*, 1998. 1389: p. 35-47.
- [4] Tofani, P. and R. Kasturi. Segmentation of text from color map images. in *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on. 1998.*
- [5] Li, L., et al., Integrated text and line-art extraction from a topographic map. *International Journal on Document Analysis and Recognition*, 2000. 2(4): p. 177-185.
- [6] Cao, R. and C.L. Tan, Text/graphics separation in maps, in *Graphics Recognition Algorithms and Applications. 2002.* p. 167-177.
- [7] Levachkine, S., et al. Semantic analysis and recognition of raster-scanned color cartographic images. in *Graphics Recognition Algorithms and Applications. 2002.* Springer.
- [8] Tombre, K., et al., Text/graphics separation revisited, in *Document Analysis System V*, E.P. Lopresti, J. Hu, and R. Kashi, Editors. 2002. p. 200-211.
- [9] Zhong, D.X., Extraction of embedded and/or line-touching character-like objects. *Pattern Recognition*, 2002. 35(11): p. 2453-2466.
- [10] Velázquez, A. and S. Levachkine, Text/Graphics Separation and Recognition in Raster-Scanned Color Cartographic Maps, in *Lecture Notes in Computer Science, Graphics Recognition. 2004.* p. 63-74.
- [11] Cheng, Z. and Y. Liu, A Graph-based Method to Remove Interferential Curve From Text Image. *Machine Vision and Applications*, 2006. 17(4): p. 219-228.
- [12] Roy, P.P., J. Lladós, and P. Umapada. Text/Graphics Separation in Color Maps. in *Computing: Theory and Applications, 2007. ICCTA '07. International Conference on. 2007.*
- [13] Roy, P.P., et al. A System to Segment Text and Symbols from Color Maps. in *Graphics Recognition. Recent Advances and New Opportunities, Lecture Notes in Computer Science. 2008.*
- [14] Epshtein, B., E. Ofek, and Y. Wexler. Detecting text in natural scenes with stroke width transform. in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2010. San Francisco, CA: IEEE.*

- [15] Chiang, Y.-Y., S. Leyk, and C.A. Knoblock, A survey of digital map processing techniques. *ACM Computing Surveys (CSUR)*, 2014. 47(1): p. 1.
- [16] Ghafari-Beranghar, A., E. Kabir, and K. Kangarloo, Directional Stroke Width Transform to Separate Text and Graphics in City Maps. *Journal of Computer & Robotics*, 2014. 7(2): p. 1-7.
- [17] Chiang, Y.-Y. and C.A. Knoblock, Recognizing text in raster maps. *Geoinformatica*, 2015. 19(1): p. 1-27.
- [18] Chiang, Y.-Y., et al., Assessing the impact of graphical quality on automatic text recognition in digital maps. *Computers & Geosciences*, 2016. 93: p. 21-35.
- [19] National Cartographic Center. NCC. 2018 [cited 2018 1 Jan]; Available from: www.NCC.org.ir.
- [20] Sahab Geographic and Drafting Institute. 2018 [cited 2018 1 Jan]; Available from: <http://www.sahabmap.com>.