Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 389547, 10 pages
http://dx.doi.org/10.1155/2014/389547

Research Article

Fuzzy-Based Segmentation for Variable Font-Sized Text Extraction from Images/Videos

Samabia Tehsin,[1] Asif Masood,[1] Sumaira Kausar,[2] and Fahim Arif[1]

[1] *MCS, National University of Science & Technology (NUST), Islamabad, Pakistan*
[2] *College of E & ME, National University of Science & Technology (NUST), Islamabad, Pakistan*

Correspondence should be addressed to Samabia Tehsin; tsamabia@yahoo.com

Received 28 October 2013; Revised 23 December 2013; Accepted 14 January 2014; Published 19 March 2014

Academic Editor: Yi-Hung Liu

Textual information embedded in multimedia can provide a vital tool for indexing and retrieval. A lot of work is done in the field of text localization and detection because of its very fundamental importance. One of the biggest challenges of text detection is to deal with variation in font sizes and image resolution. This problem gets elevated due to the undersegmentation or oversegmentation of the regions in an image. The paper addresses this problem by proposing a solution using novel fuzzy-based method. This paper advocates postprocessing segmentation method that can solve the problem of variation in text sizes and image resolution. The methodology is tested on ICDAR 2011 Robust Reading Challenge dataset which amply proves the strength of the recommended method.

1. Introduction

Recently there has been a rapid surge in multimedia reservoirs that raised the need of efficient retrieval, indexing, and browsing of multimedia information. Several methodologies are presented in the literature to retrieve image and video data, which exploit color, texture, shape, and relation between objects, and so forth. However, embedded text in images can be extraordinarily instrumental for data retrieval as visual texts in multimedia communicate information regarding news headlines, title of movie, trade-name of products, summaries of sports contest, date and time of events, and so forth. Such information can be influential for the understanding and retrieval of images or videos.

Text implanted in images may be categorized in two classes, namely, caption text and scene text. Caption text is imposed over the image in the editing process for example news headings and match summary/score. It is also referred to as artificial text or superimposed text, whereas scene text is an actual part of the scene, that is, brand name of the product during commercial break, text on sign-board, name plate and text visible on dresses or product, and so forth.

One of the key challenges posed to the text detection process is to deal with text size variations. The text variation may be classified in two types: firstly, the variation of spatial resolution of images and secondly the variation of font sizes within an image. This paper focuses on the above mentioned problem in text detection and provides viable solutions for both categories of the problem.

The rest of the paper is ordered as follows. Section 2 highlights some related work of the field. Section 3 introduces the proposed method to segment text in images. Section 4 presents the dataset used and results of text segmentation algorithm. Section 5 provides some concluding remarks.

2. Literature Review

A variety of techniques for text extraction have appeared in the recent past [1–6]. Comprehensive surveys can be traced explicitly in [7–9]. These techniques can be categorized into two types mainly with reference to the utilized text features, that is, region-based and texture-based methods [10]. Texture-based methods pertain to textural properties of the text, distinguishing it from the background. These techniques mostly use Gabor filters, Wavelet, Fast fourier transform, Spatial variance, and so forth. This approach further uses machine learning methods such as support

vector machine (SVM), multilayer perceptron (MLP), and adaBoost [11–15]. Region-based methods use distinct region features to extort text content. This methodology deals with the color dissimilarity of the text and its surrounding pixels. Procedures based on color, edge, and connected components are frequently exercised in this category [16–19]. These techniques typically work in the bottom up fashion by initially segmenting the small regions and lately grouping the potential text regions. Region-based methods are generally composed of three modules: (1) segmenting the image into small regions which aims at segregating the character regions from its background, (2) merging and grouping of small regions to form words and sentences, and (3) differentiating between text and nontext objects.

Segmentation identifies the occurrence of different regions in the image but does not recognize the relation between these regions. It is substantial to merge the characters of a word to form a text object, because most of the text detection techniques work on group of characters and it is very difficult to detect the isolated character [20, 21]. This grouping can utilize the pixel level features or can exploit the high level features.

Presently, few pixel level merging methods are introduced in the literature pertaining to text detection. Dilation is the most commonly used merging technique [22–26], wherein the dimensions of the morphological operator intrinsically characterize the range of the homogeneous segmented regions. Consequently, hefty text blocks are tending to oversegmentation, whereas diminutive text areas are possibly skipped. Fixed size of the structuring element can only materialize for limited spatial resolution and small range of font sizes. Besides, size of the structuring element should be dependent upon the size of the text but usually has the fixed value which cannot deal with the variation in resolution of image and size of text. Some methodologies in literature utilize pyramid approach to solve this problem and extend the range of text sizes for detection [23, 27, 28]. This highly increases the computational requirements or demands for parallel processing mechanisms.

Object level merging is more close to human vision and deals with the objects and regions instead of pixels. It connects the potential character objects to form the text strings. Hence, the grouping and merging are dependent upon some high level features which gives better performance.

Wolf and Jolion [29] used disparity in heights and positions of the connected component to merge the characters. Minetto et al. [27] developed a grouping step, based on the space between the two text areas relative to their height. Pan et al. [30] built component relation using minimum spanning tree. This text detection method merges the characters into words using shape and spatial difference. Gonzalez and Bergasa [31] suggested that characters of the same word should have several similar characteristics, for instance, stroke size, altitude, position, adjacency, and constant interletter and interword spacing.

Shi et al. [32] used the graph model to merge the neighboring regions to form text strings. The adjoining nodes for each node are those ones that persuade the certain conditions based upon difference in color, position, width ratio, and height ratio. Character candidates are linked into pairs in Yao et al. [33] method. If two regions have similar stroke widths (ratio between the mean stroke widths is fewer than 2.0), matching sizes (ratio between their characteristic scales does not surpass 2.5), and similar colors and are closely placed (distance between them is less than two times the sum of their characteristic scales), they are tagged as a couple. Subsequently, a greedy hierarchical agglomerative clustering approach is exercised to combine the pairs into candidate chains.

Though these features are defined by strict boundaries in the existing techniques, the relation between the neighboring characters is not crisp. It is principally inequitable to declare a character as a neighbor if its distance to height ratio is 1.50 or less, whereas the same verdict gets void, if the ratio turns to even 1.51. The parameters to define the proximity of potential character should have been diffused instead of crisp logic. Thus, there is a need to architect a merging process in which the rules of inference are formulated in a general way, utilizing diffused categories. There is a requirement to frame a system which gives some weight to each of the features used for measuring the degree of neighborhood. Moreover, the similarity obtained by the currently reported features mostly does not correspond to human perception. Human perception of propinquity, similar heights, and similar color cannot be fully expressed using discrete and rigid boundaries or thresholds. These linguistic variables can be better defined by the fuzzy logic.

3. Methodology

Component extraction or segmentation is the procedure of dividing a digital image into multiple fragments, called superpixels [34, 35]. The objective of segmentation is to reduce the computational complexity of the under process image and make its representation easier to analyze. Image segmentation is classically used to trace objects and boundaries in images. In particular, image segmentation is the process to label the pixels of image, where the pixels with same labels share some common characteristics such as color, intensity, and texture and; moreover, edge detection is a basic instrument used in most image processing applications to obtain sharp alteration in intensity of the region boundaries.

Proposed segmentation method consists of two processes: splitting and merging. Splitting is performed by the traditional region-based segmentation techniques, whereas merging is based on the novel fuzzy-based method. Figure 1 provides the architecture of the proposed work.

3.1. Splitting. There exists sharp transition between the text and its background. Edge detection is the budding segmentation tool for text images because sharp intensity transition is the common feature in all the text objects. Exploiting this feature, edge detection along with the connected component labeling is used for segmentation in the proposed methodology, where Sobel edge detection technique is used for edge detection, and image dilation is applied to connect the broken edges.

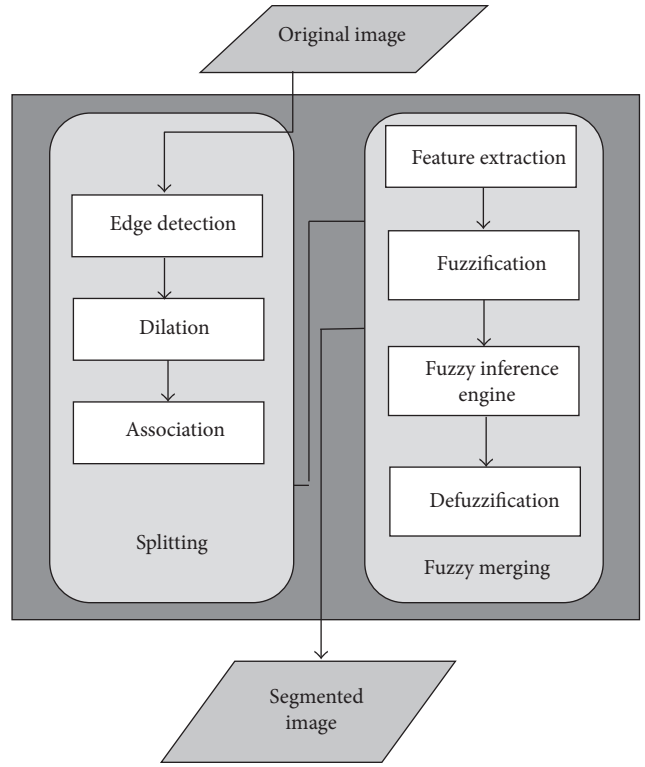

Figure 1: Architecture of the proposed method.

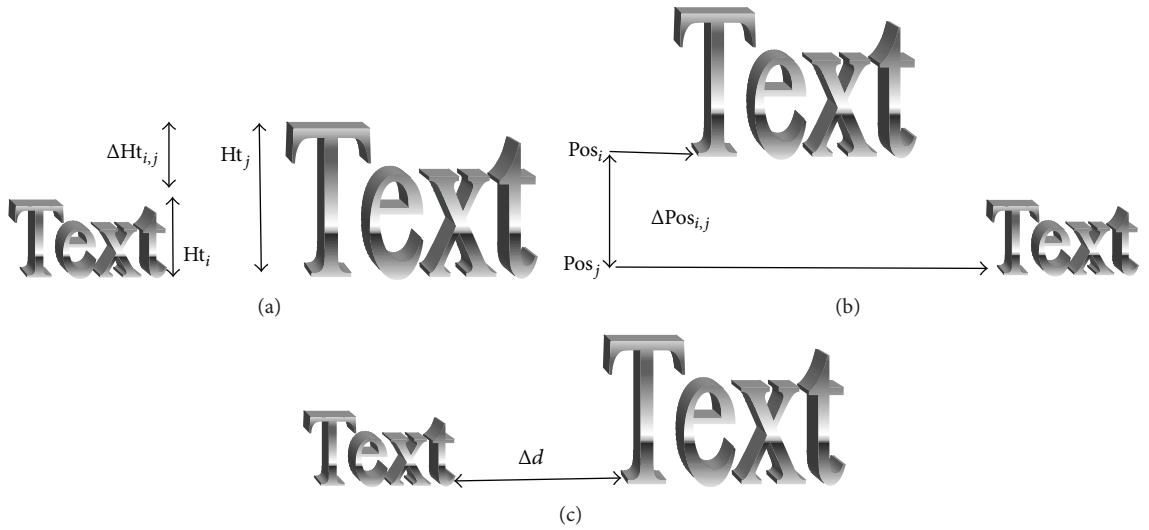

Figure 2: Factors fed into the fuzzy system: (a) height, (b) position, and (c) distance.

Adaptive size of the dilation operator is calculated in consonance with the resolution of the image, which ranges between 3 and 5% of the width of the image. Dilation is performed prior to the fuzzy merging just to minimize the computational efforts. Proposed fuzzy merging method can work without this morphological operation.

3.2. Fuzzy Merging. Succeeding section explains the fuzzy merging process.

Let I be the input image and R the set of all the regions of I, extracted by the above mentioned method. Let $I = \{R_1, R_2, R_3, \ldots, R_m\}$, and m is the total number of regions in the image I.

The problem of merging process can be defined using the graph theory. Let G denote the undirected graph and $V(G) = R$ represent the vertices of the graph G. Edges of the graph G are $E(G) = \{(R_i, R_j) \in RxR \mid i \neq j\}$. These edges show the probability of joining two vertices. Initially, $\forall e \in E(G)$ are set

to null. This probability $p_{i.j}$ can be calculated by fuzzy logic and based upon the four factors. Four factors considered are explained later in the paper.

Fuzzy-based methods assign gradual membership value to the objects, to join with other text instances, which are measured as degrees in [0, 1]. This gives the flexibility to connect the object based on more than one feature, depending upon the different membership values of all the parameters.

3.2.1. Feature Extraction. The merging of character candidates relies on number of factors. Four features are extracted for the decision of joining objects as words or sentences. These features are color, height, position, and distance.

Color. Color is taken as the parameter to join the two text objects. Color of the characters of a single word or sentence is mostly the same. If the colors of two text objects are similar, then these objects can be the candidates to merge. In order to get the degree of similarity, difference between the two colors is calculated.

Lab color coding is used to describe the color of the object. Unlike RGB and CMYK, Lab color coding approximates the human vision system. ΔE is described in the $L^*C^*h^*$ color space with differences in lightness, chroma, and hue calculated from $L^*a^*b^*$ coordinates. Difference of two colors having coordinates (L_1^*, a_1^*, b_1^*) and (L_2^*, a_2^*, b_2^*) can be defined as

$$\Delta E = \sqrt{\left(\frac{\Delta L^*}{K_L S_L}\right)^2 + \left(\frac{\Delta C_{ab}^*}{K_c S_c}\right)^2 + \left(\frac{\Delta H_{ab}^*}{K_H S_H}\right)^2}. \quad (1)$$

Here,

$$\begin{aligned}
\Delta L^* &= L_1^* - L_2^*,\\
C_1^* &= \sqrt{a_1^{*^2} + b_1^{*^2}},\\
C_2^* &= \sqrt{a_2^{*^2} + b_2^{*^2}},\\
\Delta C_{ab}^* &= C_1^* - C_2^*,\\
\Delta H_{ab}^* &= \sqrt{\Delta E_{ab}^{*^2} - \Delta L^{*^2} - \Delta C_{ab}^{*^2}}\\
&= \sqrt{\Delta a^{*^2} + \Delta b^{*^2} + \Delta C_{ab}^{*^2}},\\
\Delta a^* &= a_1^* - a_2^*,\\
\Delta b^* &= b_1^* - b_2^*,\\
S_L &= 1,\\
S_C &= 1 + K_1 C_1^*,\\
S_H &= 1 + K_2 C_1^*,\\
K_L &= 1,\\
K_1 &= 0.045,\\
K_2 &= 0.015.
\end{aligned} \quad (2)$$

Geometrically, the amount ΔH_{ab}^* presents the arithmetic mean of the chord lengths of the equal chroma circles of the two colors.

Height. Difference of heights is the second input parameter for fuzzy system. Only objects with similar heights should be merged because characters of the same word or sentence have the same or similar heights. Difference of heights of two objects is measured as follows:

$$\Delta \mathrm{Ht}_{i,j} = \frac{\left|\mathrm{Ht}_i - \mathrm{Ht}_j\right|}{\mathrm{Ht}_i}, \quad (3)$$

where Ht_i and Ht_j are the heights of ith and jth objects, respectively.

Position. Position of the two objects should be the same for merger. This merging process is proposed for horizontal text only as most of the caption text is horizontally aligned. This can be expanded to other directions by considering position at different angles. Consider

$$\Delta \mathrm{Pos}_{i,j} = \frac{\left|\mathrm{Pos}_i - \mathrm{Pos}_j\right|}{\mathrm{Ht}_i}, \quad (4)$$

where Pos_i and Pos_j are the bottom coordinates of bounding boxes of ith and jth objects, respectively.

Distance. Characters of the same word or sentence are placed closely. The distance between characters varies with the variation in font size and is highly dependent upon the heights of the characters. Distance (Δd) between two regions in an image is calculated by

$$\Delta d = \frac{\min\left(\left|x_i(1) - x_j(2)\right|, \left|x_j(1) - x_i(2)\right|\right)}{\mathrm{Ht}_i}, \quad (5)$$

where $x_n(1)$ and $x_n(2)$ are the left and right coordinates of bounding box of nth object. Figure 2 explains the height, position, and distance phenomena pictorially.

3.2.2. Fuzzification. This step gets the inputs and decides the degree to which suitable fuzzy sets belong by means of membership functions. The input has to be a crisp numerical value bounded to the universe of discourse of the input variable and the output is a fuzzy degree of membership in the qualifying linguistic set. Fuzzification of the input refers to either a table lookup or function estimation.

Let the inputs to the fuzzy system be represented in the vector notation:

$$\begin{aligned}
x^* &= \left[x_1^* \;\; x_2^* \;\; x_3^* \;\; x_4^*\right]\\
&= \left[\Delta E \;\; \Delta \mathrm{Ht} \;\; \Delta \mathrm{Pos} \;\; \Delta d\right],
\end{aligned} \quad (6)$$

where x^* belonging to R^4 represents real value points. We define symmetric Gaussian function and sigmoid function for the input.

The symmetric Gaussian function is defined by two parameters σ and c:

$$\begin{aligned}
\mu_{A^e}(x_1) &= e^{-(1/2)((x_1-\overline{x}_1{}^{(e)})/\sigma_1^{(e)})^2},\\
\mu_{B^f}(x_2) &= e^{-(1/2)((x_2-\overline{x}_2{}^{(f)})/\sigma_2^{(f)})^2},\\
\mu_{C^g}(x_3) &= e^{-(1/2)((x_3-\overline{x}_3{}^{(g)})/\sigma_3^{(g)})^2},\\
\mu_{D^h}(x_4) &= e^{-(1/2)((x_4-\overline{x}_4{}^{(h)})/\sigma_4^{(h)})^2},
\end{aligned} \tag{7}$$

where $e = 1, 2$; $f = 1, 2$; $g = 1, 2$; and $h = 1, 2$ represent the number of fuzzy sets. $\overline{x}_1{}^{(e)}$, $\overline{x}_2{}^{(f)}$, $\overline{x}_3{}^{(g)}$, and $\overline{x}_4{}^{(h)}$ represent the means of fuzzy sets, where $\sigma_1^{(e)}$, $\sigma_2^{(f)}$, $\sigma_3^{(g)}$, and $\sigma_4^{(h)}$ represent the variances of fuzzy sets.

Third membership function of all the inputs exhibits a progression from miniature start that advances and reached a culmination over time. Sigmoid function is used to express this phenomenon. Consider

$$\begin{aligned}
\mu_{A^s}(x_1) &= \frac{1}{1+e^{-a_s(x_1-c_s)}},\\
\mu_{B^t}(x_2) &= \frac{1}{1+e^{-a_t(x_2-c_t)}},\\
\mu_{C^u}(x_3) &= \frac{1}{1+e^{-a_u(x_3-c_u)}},\\
\mu_{D^v}(x_4) &= \frac{1}{1+e^{-a_v(x_4-c_v)}},
\end{aligned} \tag{8}$$

where $s = 3$; $t = 3$; $u = 3$; and $v = 3$ represent the fuzzy set's number. $a_{s\cdots v}$ and $c_{s\cdots v}$ are the model parameters to be fitted.

The following function is used to map x^* belonging to R^4 into fuzzy set $ABCD$:

$$\begin{aligned}
&\mu_{ABCD}(x_1, x_2, x_3, x_4)\\
&\quad = \mu_A(x_1) * \mu_B(x_2) * \mu_C(x_3) * \mu_D(x_4)\\
&\quad = \min\left(\mu_A(x_1), \mu_B(x_2), \mu_C(x_3), \mu_D(x_4)\right).
\end{aligned} \tag{9}$$

Minimum t-norm operator $(*)$ is used for fuzzification.

3.2.3. Product Inference Engine. Multiple inputs and single output fuzzy rule-base is employed for the current merging problem. Product inference engine (PIE) makes use of fuzzy rule base and linguistic rules. PIE encompasses individual rule-based inference with union combination, min implication, min operator for t-norm, and max operator for s-norm:

$$\begin{aligned}
&\text{Ru}^{(1)}: \text{IF } \Delta E \text{ is } A^1 \text{ and } \Delta\text{Ht is } B^1\\
&\qquad \text{and } \Delta\text{Pos is } C^1 \text{ and } \Delta d \text{ is } D^1 \text{ THEN } y \text{ is } Z^2;\\
&\text{Ru}^{(2)}: \text{IF } \Delta E \text{ is } A^2 \text{ and } \Delta\text{Ht is } B^1\\
&\qquad \text{and } \Delta\text{Pos is } C^1 \text{ and } \Delta d \text{ is } D^1 \text{ THEN } y \text{ is } Z^2;\\
&\text{Ru}^{(3)}: \text{IF } \Delta E \text{ is } A^3 \text{ or } \Delta\text{Ht is } B^3\\
&\qquad \text{or } \Delta\text{Pos is } C^3 \text{ or } \Delta d \text{ is } D^3 \text{ THEN } y \text{ is } Z^1;\\
&\text{Ru}^{(4)}: \text{IF } \Delta E \text{ is } A^1 \text{ and } \Delta\text{Ht is } B^2\\
&\qquad \text{and } \Delta\text{Pos is } C^1 \text{ and } \Delta d \text{ is } D^1 \text{ THEN } y \text{ is } Z^2;\\
&\text{Ru}^{(5)}: \text{IF } \Delta E \text{ is } A^1 \text{ and } \Delta\text{Ht is } B^2\\
&\qquad \text{and } \Delta\text{Pos is } C^2 \text{ and } \Delta d \text{ is } D^1 \text{ THEN } y \text{ is } Z^2;\\
&\text{Ru}^{(6)}: \text{IF } \Delta E \text{ is } A^1 \text{ and } \Delta\text{Ht is } B^1\\
&\qquad \text{and } \Delta\text{Pos is } C^2 \text{ and } \Delta d \text{ is } D^1 \text{ THEN } y \text{ is } Z^2;\\
&\text{Ru}^{(7)}: \text{IF } \Delta E \text{ is } A^1 \text{ and } \Delta\text{Ht is } B^1\\
&\qquad \text{and } \Delta\text{Pos is } C^1 \text{ and } \Delta d \text{ is } D^2 \text{ THEN } y \text{ is } Z^2,
\end{aligned} \tag{10}$$

where (A^1, A^2, A^3), (B^1, B^2, B^3), and (C^1, C^2, C^3) are the input fuzzy membership functions for the same, similar, and different, (D^1, D^2, D^3) are the membership values for minimum, average, and maximum and (Z^1, Z^2) are the output membership functions corresponding to *Not join* and *join*. Triangular curve function is used as output membership function that can be defined by

$$\mu_Z(y_m) = \begin{cases} 0 & y_m \le a_z \text{ OR } c_z \le y_m\\ \dfrac{y_m - a_z}{b_z - a_z} & a_z \le y_m \le b_z\\ \dfrac{c_z - y_m}{c - b_z} & b_z \le y_m \le c_z. \end{cases} \tag{11}$$

More compactly, it can be expressed as

$$\mu_Z(y_m) = \max\left(\min\left(\frac{y_m - a_z}{b_z - a_z}, \frac{c_z - y_m}{c_z - b_z}\right), 0\right). \tag{12}$$

The parameters a_z and c_z define the feet of the triangle and the parameter b_z defines the peak. Figure 3 shows different membership functions used in the system.

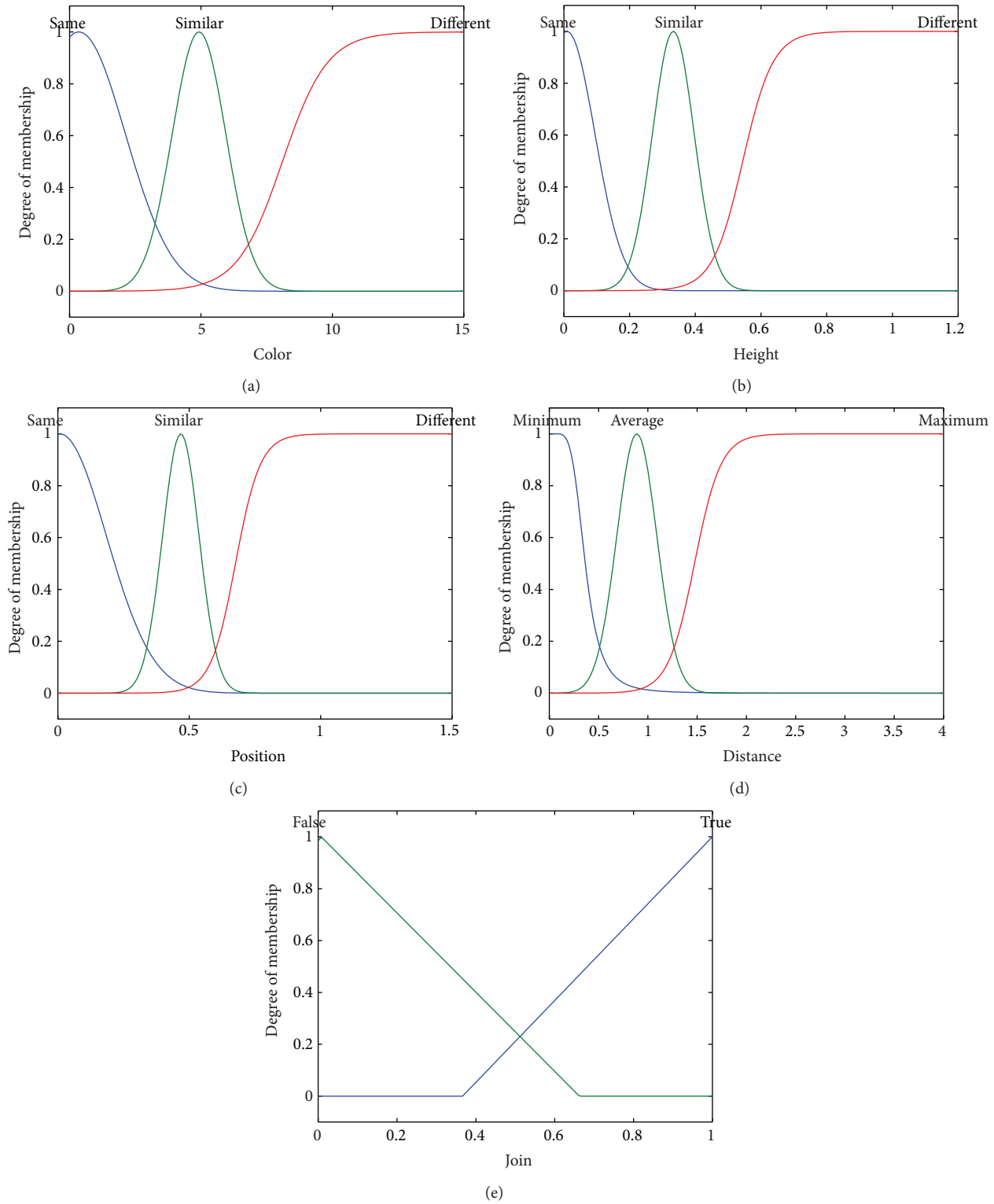

Figure 3: Membership functions: (a) color, (b) height, (c) position, (d) distance, and (e) output.

PIE can be fully defined by

$$\mu_{Z'}(y_m) = \max_{\{e,f,g,h,s,t,u,v,z\}} \left\{ \sup_{x \epsilon X} \min \left(\mu_{ABCD}(x_1, x_2, x_3, x_4), \mu_{A^i}(x_1)\, \mu_{B^i}(x_2), \mu_{C^i}(x_3), \mu_{D^i}(x_4), \mu_Z(y_m) \right) \right\}. \tag{13}$$

3.2.4. Defuzzification. Defuzzification is the mapping of fuzzy values into the real-world values. Center average fuzzifier (CAD) is used as the weighted average of the centers of fuzzy sets as it provides a reasonable approximation:

$$\overline{\omega}_m = y_m^* = \frac{\sum_{c=1}^{n_r} \overline{y}^{(c)} \omega_n^{(c)}}{\sum_{c=1}^{n_r} \omega_n^{(c)}}, \tag{14}$$

where $\overline{y}^{(c)}$ and $\omega_n^{(c)}$ are the center and height of the output fuzzy sets. CAD is chosen because it is computationally less expensive and has more accuracy and continuity when compared to other defuzzifiers [36].

4. Results and Experiments

Dataset of ICDAR 2011 Robust Reading Competition, Challenge 1: "Reading Text in Born-Digital Images (Web and Email)," is applied in this research, wherein the dataset comprises 102 images of test and 420 images of training sets. The above dataset possesses vast variation in font size, resolution, background complexity, and font type. However, ICDAR dataset is recognized as the most widely used benchmark for text detection.

The ranking metric used for the text segmentation task is accurate. Accuracy of segmentation can be defined as

$$\text{Acc} = \sum_{i=1}^{N} \frac{\text{No. of correctly segemented objects}}{\text{Total number of objects}}. \tag{15}$$

In the text detection and localization problem, isolated character is also considered as under segmentation. Proposed method obtained 90.7% accuracy for segmentation of text objects. Comparison of the segmentation results with and without fuzzy merging can be viewed in Figure 4. Segmentation without fuzzy merging is tested for adaptive and fixed size structuring elements. Achieved results show that fuzzy merging has a very effective role in segmentation for text detection.

In order to prove the practicability of the proposed segmentation method, fuzzy merging is added as the post segmentation process in textorter [37], which is the best technique in ICDAR Robust Reading Competition 2011 [38], whereby the results justify a major improvement in the detection rate of textorter. It is also factual that many isolated characters are not detected as text by textorter, as these are not merged as a complete word. The ranking metric used for the text localization task is the harmonic mean which is computed according to the methodology proposed in [39]. It is a combination of two measures: precision and recall. Table 1 shows the comparison of results for different text detection methods.

Table 1: Comparison of proposed work with other techniques.

Method	Recall	Precision	Harmonic mean
Textorter with fuzzy merging	**73.75**	**85.12**	**79.02**
Textorter [37]	69.62	85.83	76.88
TH-TextLoc*	73.08	80.51	76.62

*Stood second in ICDAR Robust Reading Competition 2011 [38].

Figure 5 shows the superiority of the proposed method. Results show that fuzzy merging really enhances the segmentation process for text detection.

Different combinations for input and output membership functions are tested, where the results show that the combination testified in the proposed methodology ensures the best outcome. Gaussian, triangular, sigmoid, trapezoidal, and bell-shaped are commonly used membership functions. These functions are tested for making different combinations of fuzzy inference engine. Gaussian, triangular, and sigmoid functions are defined in Section 3.2.2. Bell-shaped function can be defined as

$$\mu_A(x) = \frac{1}{1 + |(x-c)/a|^{2b}}. \tag{16}$$

The comprehensive bell function can be defined using three parameters a, b, and c, and here the parameter b is mostly positive, whereas parameter c traces the middle of the curve.

The trapezoidal curve is a function of a vector "X" and is dependent upon four scalar parameters a, b, c, and d:

$$\mu_A(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \dfrac{d-x}{d-c}, & c \le x \le d \\ 0, & d \le x \end{cases} \tag{17}$$

or it can be defined compactly as

$$\mu_A(x) = \max\left(\min\left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c} \right), 0 \right). \tag{18}$$

The parameters a and d trace the "feet" of the trapezoid and the parameters b and c set the "shoulders."

Figure 6 shows the comparison of different membership functions regarding four inputs.

5. Conclusion

The paper addresses very crucial problem of text detection, which is variation in font size and resolution. Earlier approaches are primarily dataset specific and unable to deal

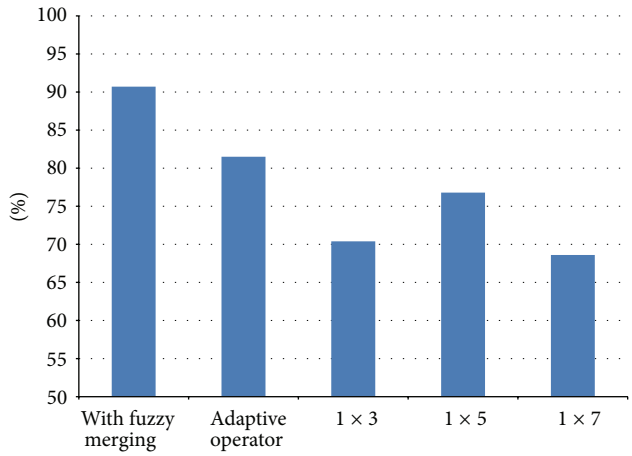

Figure 4: Comparison of proposed methodology with other techniques.

(a) (b) (c)

Figure 5: Results of the proposed method. (a) Original images, (b) after splitting method, and (c) after fuzzy merging.

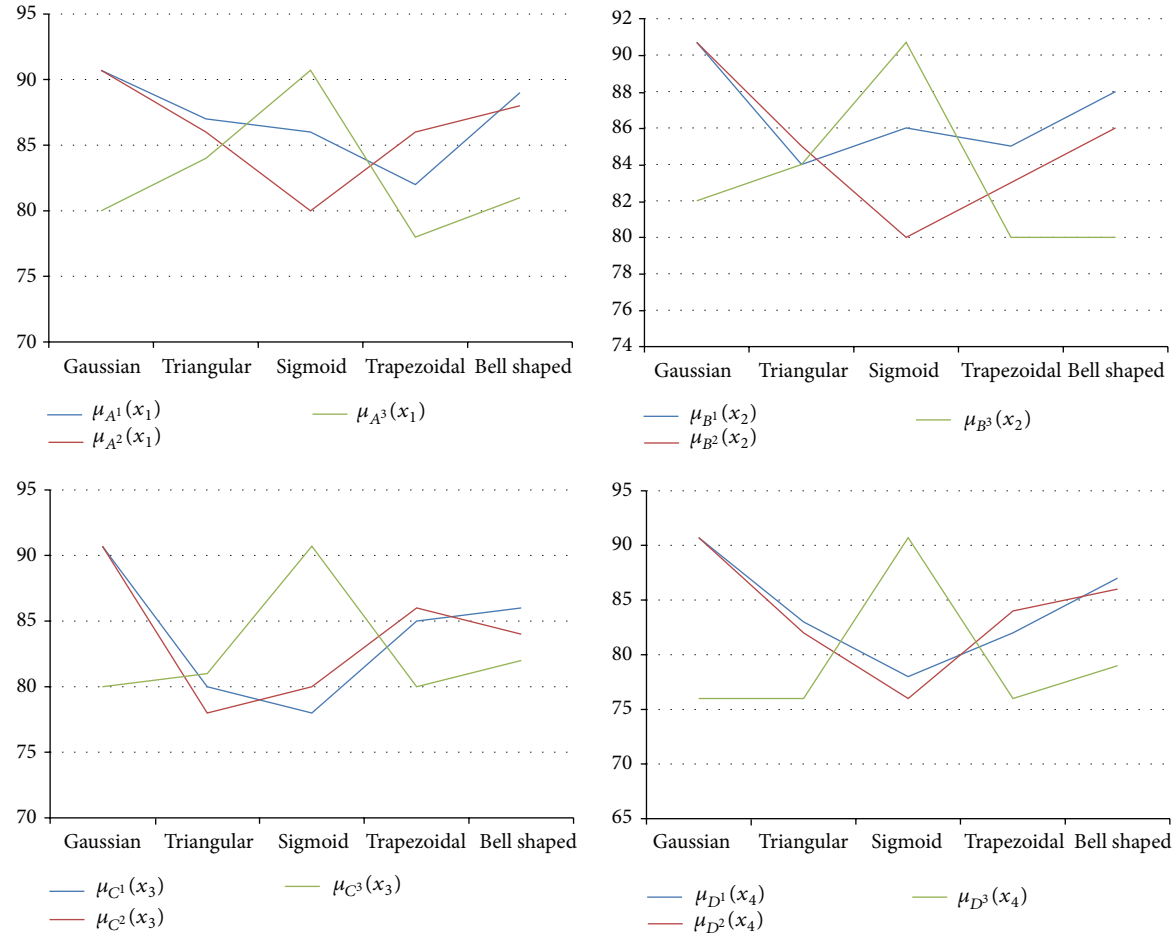

Figure 6: Comparison of different membership functions.

with enormous variation of font sizes. This paper devises a fuzzy-based postprocessing method for segmentation duly operatable with combination of any segmentation method. Four factors are mainly put forth for joining characters into words. These factors are fed into the fuzzy system which gives the verdict of joining or not joining regions. Dataset of ICDAR 2011 Robust Reading Competition, Challenge 1: "Reading Text in Born-Digital Images (Web and Email)," is applied into this research, whereby the results achieved stand out to be productive when pitched against the above referred retrieval problems.

Conflict of Interests

The authors declare that they have no conflict of interests regarding the publication of this paper.

References

[1] H. Li, D. Doermann, and O. Kia, "Automatic text detection and tracking in digital video," *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 147–156, 2000.

[2] K. I. Kim, K. Jung, and J. H. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1631–1639, 2003.

[3] M. Zhao, S. Li, and J. Kwok, "Text detection in images using sparse representation with discriminative dictionaries," *Image and Vision Computing*, vol. 28, no. 12, pp. 1590–1599, 2010.

[4] K. Wang and S. Belongie, "Word spotting in the wild," in *Proceedings of the European Conference on Computer Vision (ECCV '10)*, pp. 591–604, Springer, 2010.

[5] L. Neumann and J. Matas, "A method for text localization and recognition in real-world images," in *Proceedings of the Asian Conference on Computer Vision (ACCV '11)*, pp. 770–783, Springer, 2011.

[6] P. Shivakumara, T. Q. Phan, and C. L. Tan, "A Laplacian approach to multi-oriented text detection in video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 412–419, 2011.

[7] K. Jung, K. I. Kim, and A. K. Jain, "Text information extraction in images and video: a survey," *Pattern Recognition*, vol. 37, no. 5, pp. 977–997, 2004.

[8] J. Liang, D. Doermann, and H. Li, "Camera-based analysis of text and documents: a survey," *International Journal on*

Document Analysis and Recognition, vol. 7, no. 2-3, pp. 84–104, 2005.

[9] C. P. Sumathi, T. Santhanam, and G. Gayathri, "A Survey on various approaches of text extraction in images," *International Journal of Computer Science & Engineering Survey*, vol. 3, no. 4, 2012.

[10] R. Lienhart, *Video OCR: A Survey and Practitioner's Guide*, Video mining, Springer, Burlingame, Calif, USA, 2003.

[11] C. Li, X. G. Ding, and Y. S. Wu, "An algorithm for text location in images based on histogram features and Ada-boost," *Journal of Image and Graphics*, vol. 3, article 003, 2006.

[12] K. I. Kim, K. Jung, and J. H. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1631–1639, 2003.

[13] R. Lienhart and A. Wernicke, "Localizing and segmenting text in images and videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 4, pp. 256–268, 2002.

[14] J. Gllavata, E. Qeli, and B. Freisleben, "Detecting text in videos using fuzzy clustering ensembles," in *Proceedings of the 8th IEEE International Symposium on Multimedia (ISM '06)*, pp. 283–290, IEEE, December 2006.

[15] D. Chen, J.-M. Odobez, and H. Bourlard, "Text detection and recognition in images and video frames," *Pattern Recognition*, vol. 37, no. 3, pp. 595–608, 2004.

[16] J. Fabrizio, M. Cord, and B. Marcotegui, *Text Extraction from Street Level Images*, City Models, Roads and Traffic (CMRT), 3, 2009.

[17] M. León Cristóbal, V. Vilaplana Besler, A. Gasull Llampallas, and F. Marqués Acosta, *Region-Based Caption Text Extraction*, 2012.

[18] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 2963–2970, IEEE, June 2010.

[19] M. Anthimopoulos, B. Gatos, and I. Pratikakis, "A two-stage scheme for text detection in video images," *Image and Vision Computing*, vol. 28, no. 9, pp. 1413–1426, 2010.

[20] L. Neumann and J. Matas, "A method for text localization and recognition in real-world images," in *Proceedings of the Asian Conference on Computer Vision (ACCV '10)*, pp. 770–783, Springer, 2011.

[21] S. T. Deepa and S. P. Victor, "A novel method for text extraction," *International Journal of Engineering Science & Advanced Technology*, no. 4, pp. 961–964, 2013.

[22] R. Farhoodi and S. Kasaei, "Text segmentation from images with textured and colored background," in *Proceedings of 13th Iranian Conference on Electrical Engineering*, Zanjan, Iran, May 2005.

[23] M. S. Das, B. H. Bindhu, and A. Govardhan, "Evaluation of text detection and localization methods in natural images," *International Journal of Emerging Technology and Advanced Engineering*, vol. 2, no. 6, pp. 277–282, 2012.

[24] S. T. Deepa and S. P. Victor, "A novel method for text extraction," *International Journal of Engineering Science Advanced Technology*, vol. 2, no. 4, pp. 961–964, 2013.

[25] S. Li and J. T. Kwok, "Text extraction using edge detection and morphological dilation," in *Proceedings of the International Symposium on Intelligent Multimedia, Video and Speech Processing (ISIMP '04)*, pp. 330–333, IEEE, October 2004.

[26] J. Poignant, L. Besacier, G. Quenot, and F. Thollard, "From text detection in videos to person identification," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '12)*, pp. 854–859, IEEE, July 2012.

[27] R. Minetto, N. Thome, M. Cord, J. Fabrizio, and B. Marcotegui, "Snoopertext: a multiresolution system for text detection in complex visual scenes," in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP '10)*, pp. 3861–3864, IEEE, September 2010.

[28] M. Anthimopoulos, B. Gatos, and I. Pratikakis, "Multiresolution text detection in video frames," in *Proceedings of the 2nd International Conference on Computer Vision Theory and Applications (VISAPP '07)*, pp. 161–166, March 2007.

[29] C. Wolf and J.-M. Jolion, "Extraction and recognition of artificial text in multimedia documents," *Pattern Analysis and Applications*, vol. 6, no. 4, pp. 309–326, 2004.

[30] Y.-F. Pan, X. Hou, and C.-L. Liu, "A hybrid approach to detect and localize texts in natural scene images," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 800–813, 2011.

[31] A. Gonzalez and L. M. Bergasa, "A text reading algorithm for natural images," *Image and Vision Computing*, vol. 31, pp. 255–274, 2013.

[32] C. Shi, C. Wang, B. Xiao, Y. Zhang, and S. Gao, "Scene text detection using graph model built upon maximally stable extremal regions," *Pattern Recognition Letters*, vol. 34, pp. 107–116, 2012.

[33] C. Yao, X. Bai, W. Liu, Y. Ma, and Z. Tu, "Detecting texts of arbitrary orientations in natural images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 1083–1090, IEEE, June 2012.

[34] O. J. Tobias and R. Seara, "Image segmentation by histogram thresholding using fuzzy sets," *IEEE Transactions on Image Processing*, vol. 11, no. 12, pp. 1457–1465, 2002.

[35] N. Senthilkumaran and R. Rajesh, "Edge detection techniques for image segmentation-a survey of soft computing approaches," *International Journal of Recent Trends in Engineering*, vol. 1, no. 2, pp. 250–254, 2009.

[36] L. X. Wang, *A Course in Fuzzy Systems*, Prentice-Hall Press, Upper Saddle River, NJ, USA, 1999.

[37] S. Tehsin, A. Masood, S. Kausar, and Y. Javed, "Text localization and detection method for born-digital images," *IETE Journal of Research*, vol. 59, no. 4, pp. 343–349, 2013.

[38] D. Karatzas, S. R. Mestre, J. Mas, F. Nourbakhsh, and P. P. Roy, "ICDAR 2011 robust reading competition—challenge 1: reading text in born-digital images (web and email)," in *Proceedings of the 11th International Conference on Document Analysis and Recognition (ICDAR '11)*, pp. 1485–1490, IEEE, September 2011.

[39] C. Wolf and J.-M. Jolion, "Object count/area graphs for the evaluation of object detection and segmentation algorithms," *International Journal on Document Analysis and Recognition*, vol. 8, no. 4, pp. 280–296, 2006.