Noname manuscript No.
(will be inserted by the editor)

Interval type-2 fuzzy set and Human vision based Multi-scale Geometric Analysis for text-graphics segmentation

Soumyadip Dhar · **Malay K Kundu**

the date of receipt and acceptance should be inserted later

Abstract This paper presents a novel method for texture-based text-graphic segmentation in a text embedded image. In the method, features are computed applying Multi-scale Geometric Analysis(MGA). The MGA of the image is done by Non-subsampled contourlet transform(NSCT). The NSCT sub-bands help to generate the features which represent textures of the text portions and graphics portions of the image. In a segmentation process, the uncertainties arise mainly for two reasons: one is the ambiguity in gray level and other is the spatial ambiguity. Here the uncertainties are managed by interval type2 fuzzy set (IT2FS). The human vision model called human psychovisual phenomenon (HVS) is incorporated in the process for generating the interval type-2 fuzzy membership functions (IT2FMF). The efficiency of the proposed scheme is measured on the benchmark dataset. The robustness and performance bound of the proposed technique under noise corruption are measured statistically using modified Cramer-Rao bound. We found that effectiveness of the features by NSCT in combination with the IT2FS are quite promising in comparison to the state-of-the-arts methods.

Keywords text-graphics segmentation · Interval type-2 · fuzzy · HVS · segmentation bound

1 Introduction

Text-graphics separation or localization of text in a text embedded image is an important research field. It has diverse applications like helping a visually handicapped person, number plate detection of a vehicle, analyzing the contents of a document

Soumyadip Dhar
RCCIIT, Beliaghata, Kolkata-700010, India
Tel.: +91-9883344995
E-mail: rccsoumya@gmail.com

Malay K Kundu
ISI, B.T.Road, Kolkata-700108, India

image, image search, target detection etc. The separation of text embedded portions from the graphics also helps to store and transmit the document image efficiently using different compression techniques. The lossy and lossless compression techniques are used for storing the non-text/graphics and texts respectively. The prerequisite for the automatic analysis of a document which contains graphics is to separate the text portions from graphics. Thus, the accurate boundary detection between the text and the graphics portions are necessary. To detect the boundaries one should use a proper scheme which can represent the features of the two regions efficiently. Moreover, the process should have the ability to manage the uncertainties of the texture features. Mainly the ambiguity in gray level [1] and ambiguity due to the position of pixels [2] causes the uncertainties in segmentation. The uncertainties are increased with the addition of noise, the rotation of the image in different angles and, also due to the gray value change of pixels dynamically.

2 Related work

The conventional approaches for text-graphics separation can be categorized into two group of methods; one is the top-down method and another is the bottom-up method. The probable text portions are identified first in the top-down method[3–5]. This is followed by division into paragraphs, text-line, and words. The popular top-down technique is X, Y-cut algorithm [6]. The algorithm detects the white spaces by horizontal projection and vertical projection. Grana et al.[7] used the X, Y-cut method for text-graphics separation from document. The bottom-up method first works by clustering the pixels. The clustered pixels are then arranged together to differentiate the text portions and graphics portions[8–10]. The detection of text pixels by the Maximally Stable External Regions (MSERs) for top-down method have been utilized by many researches[11–17]. Stroke features based techniques were also proposed by [18, 19] for text-graphics localization in the bottom-up approach. Zhu et al.[20] used HOG based features for the text-graphics separation. The above-illustrated methods have the constraints that they require prior knowledge about the image for accurate segmentation. That means the efficiency depends upon the resolution, character sizes, distaces between the lines and character orientations.

Some researchers nowadays prefer unsupervised learning or Deep learning for the text-graphics segmentation. Here the Convolutional Neural Network(CNN) was utilized to differentiate text from graphics region[21–24]. The methods show quite an efficiency in performing the localization task. The drawbacks of the methods are that they are computationally costly and they are trained with a huge data. Moreover, the methods are mainly text dependent. The cause for high efficiency of CNN based methods is that people trained the networks with a huge image data and they are not easy to collect. Moreover, the networks mainly detect the texts in horizontal or near horizontal text positions which make them difficult to use practically [25].

A lot of researchers prefer texture-based text-graphics segmentation in an image. The distinguishable texture properties of text and graphics are the main logic behind this approach. Thus, researchers treated the text separation from graphics as a texture segmentation problem. Different types of texture representing features like discrete

cosine transform(DCT)[26], wavelet transform[27,28], Gabor filter[29] are used in text-graphics segmentation.

The wavelet transform(WT) and the alternatives of WT (like M-band wavelet transform, M-band wavelet packet transform etc.) used by the methods in [26–29] can represent the texture of an image efficiently. However, there are limitations of WT based segmentation. The WT features have insufficient rotational invariance and anisotropic properties. So, the methods can still be improved to achieve the higher level of accuracy. The Multi-scale Geometric Analysis (MGA)[30] can overcome the deficiencies of WT-based methods. The MGA can be done using the transformation tools like Ripplet, Curvelet, and Contourlet etc., and they have been used in different problem domains [31]. MGA can detect edges efficiently which are the anisotropic features of an image. This motivate us for using the MGA tool in text-graphics segmentation as the text locations are highly anisotropic in nature. Moreover, a small number of methods described above are found to manage the uncertainties related to the text-graphics segmentation. The uncertainties can arise in feature generation for text embedded image segmentation as the different regions of the image are not crisply defined [32] and owing to the generation of an adequate number features. The uncertainty handling methods for text regions segmentation were proposed by [33–35].

3 Proposed method

Here we present an MGA based text-graphics region segmentation which can segment the image by managing the uncertainties. Here, The NSCT sub-bands are utilized to represent the texture properties of the text-graphics regions. In the proposed scheme the uncertainties generated in the NSCT feature are reduced by IT2FS[36] which is an efficient tool for uncertainty management. It is a fact that human vision which is modeled as HVS is an efficient tool for separating different objects in an image. Thus, in the proposed scheme we incorporate the visual model for managing the uncertainties. In the proposed scheme, we use the (HVS)[37] for uncertainties handling. In HVS the contrast at a point in an image depends on its surrounding. This information is utilized in the proposed scheme for interval generation of interval type-2 membership functions(IT2FMFs). The uncertainties are managed in the interval type-2 fuzzy domain. The uncertainties in the features are reduced by minimizing the fuzzy entropy in the IT2FS domain. Figure 1 shows the basic steps of our proposed scheme. A statistical measure called modified Cramer-Rao bound[38] is used for performance bound of the proposed scheme.

3.1 Novelty of the proposed scheme

The novelties of the proposed segmentation scheme are

(1) Here we propose a text non-text/graphics region segmentation of color text embedded image with the help of MGA tool NSCT. The NSCT decomposes an im-

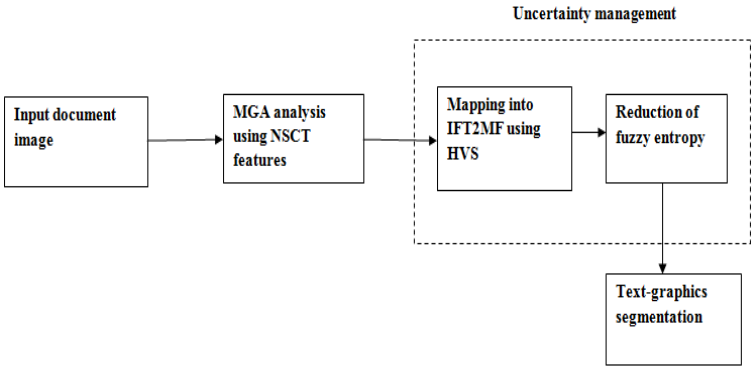

Fig. 1 The basic steps of the proposed scheme

age into sub-bands which represents the image into various scale and orientations. The sub-bands represents the fine details of the text regions efficiently.

(2) In the proposed scheme the uncertainties during the segmentation process are managed by the IT2FS. The IT2FS minimizes the uncertainties in each NSCT sub-band and generate finer features.

(3) For the IT2FS generation the HVS model is incorporated. The interval of the IT2FMF at a pixel is weighted in accordance with the position a pixel in the HVS regions. The HVS model helps to prevent the loss of anisotropic nature of edges.

The organization of the paper is as follows. Section 4 and 5 describes NSCT and IT2FS respectively. The proposed IT2FS generation and weight generation for interval are described in section 6. The algorithm is presented in the Section 7. Section 8 comprises of results discussion, comparison with state-of-the-art methods and performance evaluation using modified Cramer-Rao bound.

4 MGA of an image by Nonsubsampled Contourlet Transform(NSCT)

For MGA of an image, the NSCT is an effective tool. The purpose of NSCT [39] is to transform the text and graphics portions in an image into two different discontinuities. NSCT is a computational framework, which is used to analyze an image in various scales and directions. Different stages of analysis are performed by two sets of filter banks: Non-Subsampled Pyramid (NSP) and Non-Subsampled Directional filter bank (NSDFB) [40].

NSP of NSCT is used for decomposing an image in various scales. It is also shift-invariant as it requires no down-sampling or up-sampling. It is generated iteratively by two channel Non-Subsampled Filter Bank (NSFB). It produces one low-frequency sub-band and one high-frequency sub-band at each decomposition level of NSP. The successive NSP filter bank decomposition stages are utilized to disintegrate iteratively the low-frequency component. The motivation is to represent the irregularities in the image. NSP filter bank produces k +1 sub-bands, made up of one Low-Frequency

Sub-band (LFS) and k High-Frequency Sub-bands (HFSs). The size of each sub-band is same as the original image where k represents the levels of decomposition.

The NSDFB is computed by removing the downsamplers and up-samplers of the DFB with the upsampling the filters subsequently[39]. This produces a tree made up of two-channel NSFB. In each of the stages of the NSP, the NSDFB makes a decomposition of 2^l directions, where l is the level number in the NSDFB. The multi-directional information is acquired by NSCT by the decomposition. The decomposition into NSCT sub-bands also comes with a redundancy of $r = 1 + \sum_{j=1}^{k} 2^{l_j}$, where l_j represents the level number in the NSDFB at the scale j. We refer to [39] for further details about NSCT.

5 Interval type2 fuzzy sets :

In a T2FS, the MF(membership function) is also fuzzy. A type-2 set $\widetilde{A}$ is specified by T2MF function $\mu_{\widetilde{A}}(x,u)$, where $x \in X$,i.e

$$\widetilde{A} = \{((x,u), \mu_{\widetilde{A}}(x,u)) | \forall x \in X, \forall u \in J_x \subseteq [0,1]\} \tag{1}$$

in which $0 \leq \mu_{\widetilde{A}}(x,u) \leq 1$. $\widetilde{A}$ can be formulated as

$$\widetilde{A} = \int_{x \in X} \int_{u \in J_x} \mu_{\widetilde{A}}(x,u)/(x,u), J_x \subseteq [0,1] \tag{2}$$

where $\mu_{\widetilde{A}}(x,u)$ is the secondary membership function and J_x is the primary membership of x which is the secondary membership function domain. Here $\int\int$ denotes union over all permissible x and u. For discrete set $\int$ is given by $\sum$. When all $\mu_{\widetilde{A}}(x,u) = 1$ in the Eq.2 ,then $\widetilde{A}$ represents the interval type-2 fuzzy set(IT2FS) [36].That means in IT2FMF, the secondary grade of the secondary MF is one. Interested readers can go through [36] for details.

In IT2FS, Footprint of uncertainty (FOU) expresses the uncertainty related with the IT2FS. In the FOU, the lower MF $\underline{\mu}_{\widetilde{A}}$ and upper MF $\overline{\mu}_{\widetilde{A}}(x)$ are utilized for uncertainty measure in IT2FS. Kacpryzk and Smidtz [41] in their paper proposed the measure of uncertainty for the IT2FS. We used the same measure for our method. It is defined as

$$\xi_k(\widetilde{A}) = \frac{1}{N} \sum_{i=1}^{N} \frac{1 - max(1 - \overline{\mu}_{\widetilde{A}}(x), \underline{\mu}_{\widetilde{A}}(x)))}{1 - min(1 - \overline{\mu}_{\widetilde{A}}(x), \underline{\mu}_{\widetilde{A}}(x)))} \tag{3}$$

Where cardinality of the fuzzy set is given by N.

6 Proposed mapping of image into IT2FS and uncertainty handling

6.1 Generation of IT2MF

Let $I(i,j)$ be the gray value of the (i,j)th pixel of a $P \times Q$ dimensional L level image I. In the proposed scheme the locations within the image I where the pixel values are almost same i.e homogeneous with respect the surrounding pixels, are made more

homogeneous. The other pixels remain unchanged. These operations increase the homogeneity and contrast within the image. So, the ambiguities of an image decrease and thus uncertainties get reduced.

To perform the above-mentioned operations the image is transformed into IT2FS I_{IT2FS} by taking the $(2m+1)\times(2m+1)$ overlapping window W where m is an integer and $m>0$. The window contains $V=(2m+1)^2$ pixels in it. Out of V elements, a combination of r elements is considered. For each combination, the type-1 membership value is computed by restricted equivalent function(REF)[42] as follows.

$$REF(x_1,x_2)=1-|x_1-x_2|^{0.5} \tag{4}$$

So the membership value of $\mu_{C_i}(x)$, where $x=I(m,m)$ and C_i represents the set ith combination of W taking r at a time is given by

$$\mu_{C_i}(x)=1-\frac{|x-m_o|^{0.5}}{L} \tag{5}$$

where $m_o=\frac{\Sigma_{y\in C_i}}{r}$ and L is the maximum pixel value. So, here we get $\binom{n}{r}$ such combinations of $\mu_{C_i}(x)$, where $i=1,2...,\binom{n}{r}$. The IT2FMFs are generated by combining the $\binom{n}{r}$ type-1 MF. In the $I_{IT2FS}(x)=[\underline{\mu}(x)\ \overline{\mu}(x)]$ lower membership $\underline{\mu}(x)$ and upper membership $\overline{\mu}(x)$ of $I(m,m)$ are given by

$$\begin{aligned}\underline{\mu}(x)&=w\times t-norm(\mu_{C_1}(x),\mu_{C_2}(x),.....,\mu_{C_{\binom{n}{r}}}(x))\\ \overline{\mu}(x)&=w\times t-conorm(\mu_{C_1}(x),\mu_{C_2}(x),.....,\mu_{C_{\binom{n}{r}}}(x))\end{aligned} \tag{6}$$

where w represents the weight of the center pixel $I(m,m)$ with respect to its background in the $W\times W$ overlapping window. The interval at each point $I_{IT2FS}(m,m)$ represents the uncertainties at that point. More the intervals more will be the uncertainties. So, for the uncertainties minimization, the following operation is done.

$$\bar{I}(m,m)=\begin{cases} I(m,m) & \text{if } I_{IT2FS}(m,m)<\alpha \\ \bar{I}_\alpha(m,m) & \text{if } I_{IT2FS}(m,m)\geq\alpha \end{cases} \tag{7}$$

where $\bar{I}_\alpha(m,m)$ is the mean gray value of the set of pixels within a local overlapping window of size $(2m+1)\times(2m+1)$ around $I(m,m)$ and its position is at the center of the local window. Here α lies between [0 1].

Though the interval I_{IT2FS} can represent uncertainties, there is a possibility that the operation in 7 dilutes the edges in an image. This causes the loss of anisotropic features in the image. To minimize the effect we assign weights to the intervals depending on the background. The choice of the weight w depends upon the regions in which $I(i,j)$ lies with respect to its background in the HVS. In the next subsection, we discuss it.

6.2 Choice of the weights for IT2FMFs by HVS

In HVS the minimum difference between a point and its immediate background, so that the point can be detected is called incremental threshold [37]. The HVS divides an image pixel into three regions; De Vries-Rose region(DVR), Weber region(WR) and Saturation regions(SATR) with respect to its background. The discriminating power in DVR is more than the other two regions. In the DVR, WR and SATR the incremental threshold varies as $\triangle B_T \propto \sqrt{(}B)$,$\triangle B_T \propto B$ and $\triangle B_T \propto B^2$ respectively where $\triangle B_T = |B-I|$. Here, $B(i,j)$ is the background intensity of $I(i,j)$. The intensity is calculated by taking the neighborhood pixels of $W \times W$ window where $I(i,j)$ is the center pixel. An image pixel can reside in three zones: (1) Low uncertain zones as the pixel resides in a homogeneous region, (2) Low uncertain zone as the pixel can be differentiated from other pixels i.e high discriminative zone and (3) High uncertain zone due to the low discriminative power of the zone. The low uncertain and high uncertain zones in DVR, WR, and SATR are shown in Figure 2. So, we can say if a pixel resides in one of the three regions DVR, WR and SATR and the corresponding conditions for the incremental threshold are satisfied, the uncertainties will be reduced. So it is logical and convenient to multiply low weight low uncertain zones and high weights for high uncertain zones. Again, since DVR has higher discriminative power, we consider the lowest weight for that region. As a result, the intervals of IT2FMF will be small and the uncertainties will be low. So, the uncertainties will be low as the pixel $I(i,j)$ is in De Vries-Rose i.e $\alpha_1 B_1 \leq B < \alpha_2 B_1$ and

$$I \geq (B + K_3\sqrt{B}) \quad \text{or} \quad I \leq (B - K_3\sqrt{B}) \tag{8}$$

Or the pixel $I(i,j)$ is in the Weber region i.e $\alpha_2 B_1 \leq B < \alpha_3 B_1$ and

$$I \geq (B + K_1 B) \quad \text{or} \quad I \leq (B - K_1 B) \tag{9}$$

Or, the pixel $I(i,j)$ is the saturation region $\alpha_3 B_1 \leq B$ and

$$I \geq (B + K_2 B^2) \quad \text{or} \quad I \leq (B - K_2 B^2) \tag{10}$$

Here B_1 is the maximum background intensity. Depending on the DVR, WR or SATR the weight w is taken as 0.1, 0.5 and 0.7 respectively. For homogeneous regions the weights are assigned as 0.12. When all the conditions mentioned above are not satisfied and the pixel is in uncertain zone $w = 1$. K_1, K_2 and K_3 are the proportionality constants and we follow the same strategy as [37] to calculate them. The α_1, α_2 and α_3 are taken as 0.4, 0.2 and 0.1 respectively.

7 Proposed methodology

In a text document, the text and non-text/graphics regions have distinctly different texture characteristics. Thus, we extract the texture representing features by decomposing it into sub-bands using the MGA tool NSCT. Before applying the NSCT on the color document image, the CIE Lab color planes are generated. In this plane the

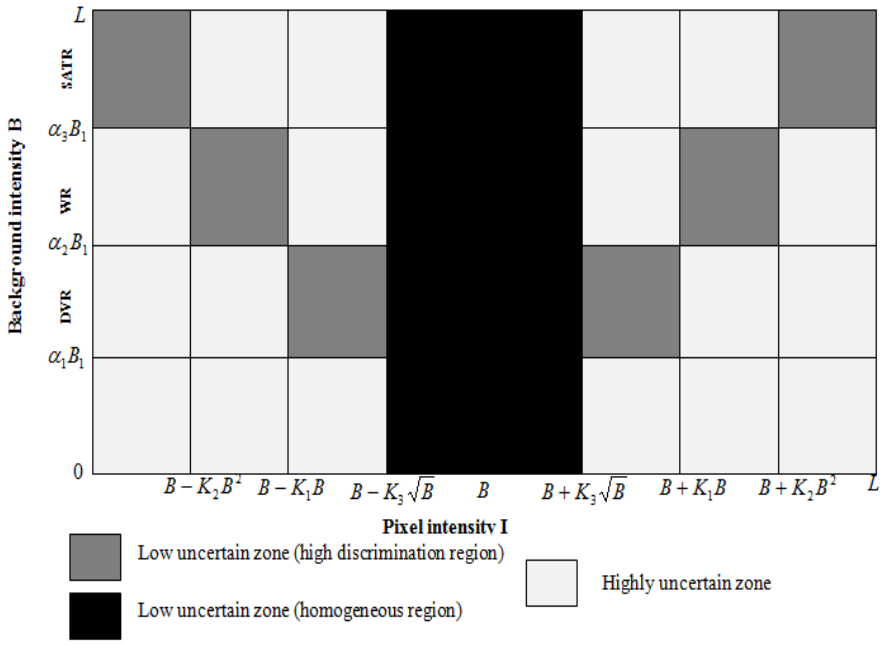

Fig. 2 The zones with low and high uncertainties in the DVR, WR and SATR. The weights are taken small in low uncertain zones compared to high uncertain zones

color metric matches with the human visual perception[43]. The energy of each contourlet coefficient is computed locally using a small overlapping window around the coefficient. The window size is chosen adaptively. The sub-band uncertainties are reduced by mapping the sub-bands into the IT2FS domain. The HVS is incorporated to imitate the efficiency of the human visual capacity and it helps to generate the appropriate membership values. Based on the intervals the homogeneity in the sub-bands are changed and it reduces the feature uncertainties. The uncertainty-reduced sub-bands are used as features for segmentation. This is followed by unsupervised feature selection by Maximal information compression(MIC) index [44]. This is done to minimize the uncertainties further due to redundant information. Finally, internal type-2 fuzzy c-means are used for text and graphics portion segmentation. The steps are shown below:

Step 1 : NSCT transform of the image - In this step, CIE Lab color planes are generated from the input image. A set of sub-bands from each plane are produced by the NSCT transform. From the sub-bands, various scales and directional information can be gained.

Step 2 : Measuring local energy with smoothing:- In each of the sub-band, the local energy of a contourlet coefficient is computed by a nonlinear operation. It is computed over a small adaptive overlapping window having a size w around the

coefficient [45]. Then a filtering using a Gaussian low-pass (smoothing) filter [28] is done to remove the very low energy contained in the sub-bands.

Step 3 : uncertainty reduction in IT2FS:-. To reduce the uncertainties in the sub-bands, the IT2FS using HVS(section6) is generated for each sub-bands. The process is repeated up to ith iteration until $|\xi_k(I_{IT2FS})_i - \xi_k(I_{IT2FS})_{i+1}| \leq \gamma$, where γ is a small positive value.

Step 4 : Unsupervised feature selection:- Here, the discriminative features are chosen by MCI. This can manage the uncertainties that occur due to redundant features.

Step 5 :Separation of text and graphics regions:-Finally, the IT2FCM clustering is applied to the feature vector for segmentation in two classes; text and non-text/graphics regions.

8 Experimental results and discussion.

The proposed scheme joins the benefits of MGA tool NSCT and uncertainty representation using the IT2FS. The IT2FS is generated using the HVS. Here, all the experiments were conducted without having a prior idea about the input images. The empirical performance of the proposed scheme was compared with some state-of-the-arts methods. In our method, the size of the window is taken as 5×5. To demonstrate the capability of our method, in our experiment, we took the text document images from the ICDAR2015 [46], ICDAR 2011, KAIST [47]. The data was also taken from the scanned newspaper images, the magazine, and the publicly available advertisement from the websites. The data also include the camera captured images. The image size varied from 120×120 to 700×700.

The performance of the method was measured by Intersection-over-Union(IOU).The threshold is taken as 50%, considering the standard convention in object recognition [48]. The measures are defined in Eq 11, Eq 12, and Eq 13 respectively [25].

$$R = \frac{|n1|}{|n2|} \tag{11}$$

$$P = \frac{|n1|}{|n3|} \tag{12}$$

$$F = \frac{2 \times P \times R}{(P+R)} \tag{13}$$

Where $n1$ is the set of true positive detections,$n2$ is set of the ground truth rectangles and $n3$ is the set of estimated rectangles. R, P, and F are the recall, the precession, and the f-measure respectively.

Figure 3 shows the qualitative results by our proposed method. The performance was compared with state-of-the-art methods Huang [18], Liu [11], Bai [10], Gomez [12], Zhu [20], Cho [16], Kim [26] and Dhar [33]. The quantitative performances are shown in the table 1. From the table, it is observed that average performance of our proposed method is superior to the other methods compared here. Kim et al. [26] used DCT transform based methods for text-graphics localization. The method was not capable of managing uncertainty in the image. The methods in [11,12] used MSERs

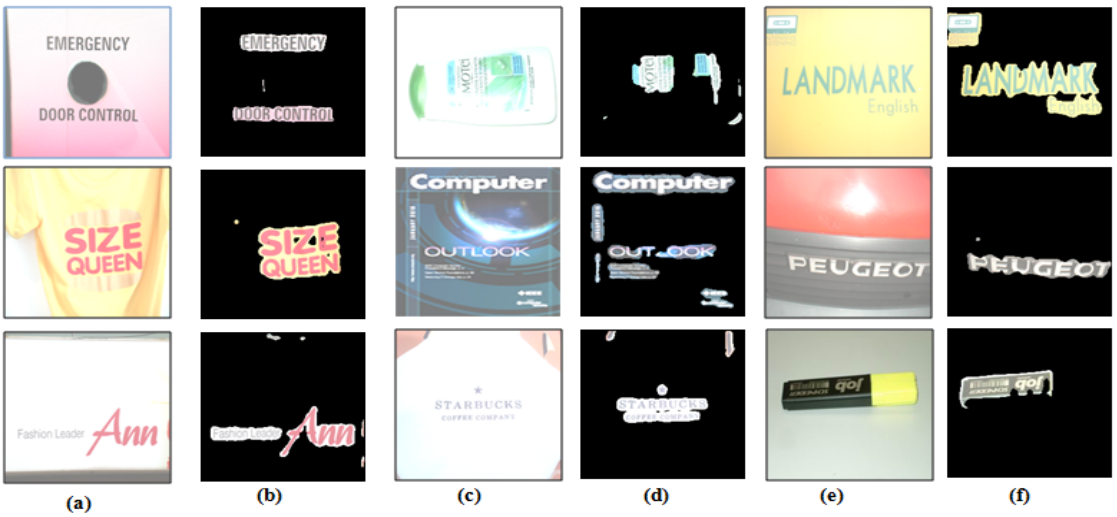

Fig. 3 Column wise(a)(c) and(e) Original color test images. (b),(d) and (f) are the corresponding results by the proposed scheme

based features to differentiate the text-graphics region. From the results, it is evident the sub-bands generated from the MGA tool NSCT and uncertainties management by IT2FS approach are more appropriate than the conventional MSERs [11, 12] based methods. The reason is that the edges i.e the anisotropic features in text portions are captured more accurately by the MGA than the other feature generating tools. The neutrosophic set was proposed for uncertainty reduction in the text document by Dhar et al.[33]. The neutrosophic domain is effective for uncertainty handling. But, three subsets true, false and indeterminate are to be maintained for uncertainties representation in neutrosophic logic. This makes the representation complex. Moreover, in the method [33] the uncertainties were reduced without taking into consideration the loss of edges. In our method, the uncertainties are handled in the IT2FS domain with the help of HVS. Thus, the contrasts of the coefficients in sub-bands were increased and made them suitable for segmentation. The HVS reduced the uncertainties adaptively in different regions and thus took into consideration the minimum loss of anisotropic features. So, the uncertainties reduction technique in our proposed scheme was more efficient than the method in [33]. the proposed scheme also performed better than that of [10]. The reason may be that the method in [10] utilized the Canny edge detector and generated the seed. The detector may fail in the complex background and it could not reduce the uncertainties. Again the method in [16] combined the MSERs and Canny edge detector to detect the text regions, which were less efficient than the proposed MGA features of the text captured by the NSCT. The method in [18] used stroke feature transform. The transform may produce uncertainties because of irregular gradient orientations. The irregularity occurs as the orientations are not perpendicular to the correct stroke edge directions. But, no techniques for uncertainty management was used to reduce the uncertainty.

Table 1 Average performance comparison of text region segmentation results obtained by different methods

Methods	Dataset	*R*	*P*	*F*
Huang [18]	ICDAR2011	0.75	0.82	0.73
Kim [26]		0.76	0.89	0.83
Proposed		0.84	0.89	0.86
Bai [10]	KAIST	0.89	0.83	0.86
Gomez [12]		0.78	0.66	0.71
Proposed		0.89	0.88	0.88
Zhu [20]	ICDAR2015(born digital)	0.80	0.82	0.81
Cho[16]		0.77	0.83	0.79
Liu [11]		0.79	0.86	0.80
Dhar [33]		0.83	0.84	0.83
Proposed		0.85	0.88	0.86

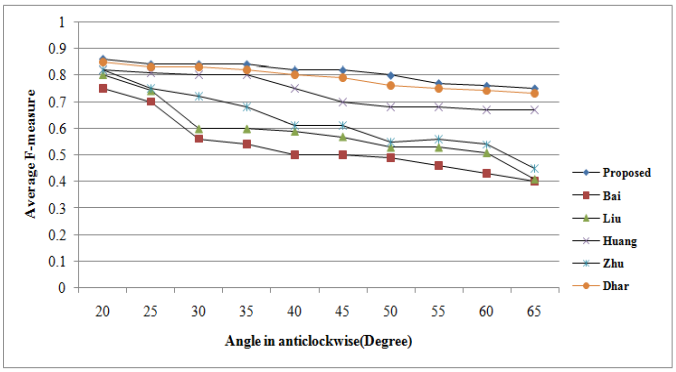

Fig. 4 Average F-measures at different angles of ICDAR2015 dataset(Born digital)

The performances of the proposed scheme under the rotation at different angles and dynamic gray level change are shown in Figure 4 and Figure 5 respectively. Form the results it is evident that our technique is robust against these two perturbations than that of the other methods. The reason is that the NSCT generates the directionally invariant features. Additionally, the IT2FS manages the spatial ambiguity increase in the features which occurs an effect of the rotation. The modification in the pixel values in an image changes the image statistics. Since in the proposed scheme the HVS was incorporated to manage the uncertainties, the IT2FMF changed accordingly with the statistics of the image.

8.1 Statistical validation by the modified Cramer-Rao bound for mean square error (MSE) on noisy images

To check how much robust the method was, we measured its performance under different noise corruptions. For this, we measured the lower modified Cramer-Rao bound[49] for MSE on the ICDAR2015(born-digital) dataset. The measure was used to find out the statistical performance bound of segmentation methods related to a par-

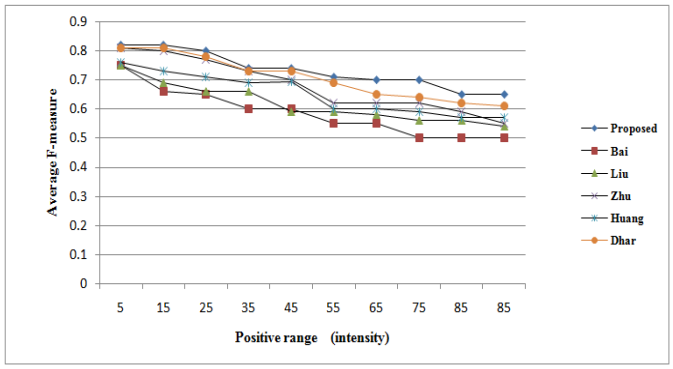

Fig. 5 Average F-measure at different dynamic gray level change ICDAR2015 dataset(Born digital)

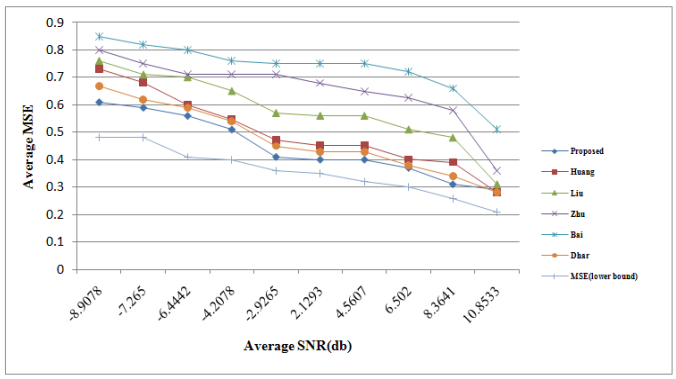

Fig. 6 Average MSE at different SNR with the modified Cramer-Rao bound for minimum MSE on ICDAR2015 dataset(Born digital)

ticular document image content. The bound exhibits the maximum performance that could be achieved by a segmentation algorithm on the particular image. Here the main assumption was that each image included two nonoverlapping basic regions; one was text and another was graphics regions. The ground truth for the image regions given in the dataset was used for the two regions. Each image was corrupted with the Gaussian noise which had a mean zero and variance σ_2. This was executed for all the images (born digital) in the dataset which contains 141 data. The lower bound(average) of MSE and individual MSE(average) for various SNRs starting from lower to higher, are shown in the figure 6. From the graph, it is evident that the proposed scheme produced better results as it has smaller MSE. It also shows good performance in higher SNRs also. The noise addition increases the uncertainty in segmentation and it also changes the image statistics. But in our method, the uncertainties were reduced based on the background intensity of a pixel. So, it is robust against the noise corruption. Also, the performance bound of the proposed scheme and the improvement space to reach the minimum MSE in each SNR can be realized from the graph.

8.2 Proposed method with unsupervised learning

We used our method for unsupervised learning and compared our method with other methods that used the learning technique. For the unsupervised learning based method, we followed the CNN structure for text region segmentation proposed by [50]. We trained the network by the features after the uncertainty management by IT2FS. The CNN network was trained with 229 images and 258 images from the ICDAR2011 ICDAR2005 dataset respectively. The text and graphics patches were created synthetically from the document images. The proposed scheme was tested on the dataset of ICDAR2013 and ICDAR2015(born digital) for comparison. They contain 255 and 141 test images respectively. In the method in [23] method, the features generated by MSERs were entered into CNN for classification. The quantitative results are shown in Table 2. From the result, it is clear the performance of the proposed scheme is superior to [23]. The reason is that the features generated from MSERs and used by He were less robust under the than that of the NSCT features used in the proposed scheme. The method also performed better than the method in [21]. The reason may be the absence of the uncertainty handling method for Aggregated Inception Feature(AIF) used by He. In our method, we reduced the uncertainties in the NSCT features in the IT2FS domain before entering into CNN, where no such techniques were available in their methods.

Table 2 Comparisons with Deep learning based methods using ICDAR dataset

Methods	Dataset	*R*	*P*	*F*
He [21]	ICDAR2013	0.86	0.88	0.87
He [23]		0.73	0.93	0.82
Proposed		0.85	0.91	0.88
He [21]	ICDAR2015	0.83	0.80	0.82
Proposed		0.80	0.93	0.87

9 Conclusion.

Here we propose an MGA based text-graphics segmentation methodology in the text document images. The method judiciously utilizes the advantages of multi-scale and multi-directional MGA tool NSCT. The uncertainties arise in MGA tool is managed by the IT2FS. For the construction of IT2FS, the human psychovisual phenomenon (HVS) is considered to minimize the loss of anisotropic features. The motivation is to achieve the segmentation result with higher accuracy in comparison to the state-of-the-art methods. It is found that the performance of the NSCT features with IT2FS is the best among the methods compared here. The method also shows robust performance under different perturbations and also under the noise corruption when verified against the modified Cramer-Rao bound. Current research is going on to extend it to a multiclass texture segmentation problem with proper modification.

References

1. C. A. Murthy and S. K. Pal. Histogram thresholding by minimizing gray level fuzziness. *Information Sciences*, 60:107–135, 1992.
2. A. Rosenfield. Fuzzy geometry: An updated overview. *Information Sciences*, 110:127–133, 1998.
3. D. Chen, J-M. Odobez, and H. Bourlard. Text detection and recognition in images and video frames. *Pattern Recognition*, 37:595–608, 2004.
4. S. M. Lucas. ICDAR2005 text locating competition results . *Proceedings of International Conference on Document Analysis and Recognition.*, 1:80–84, 2005.
5. W. Zhu, Q. Chen, C. Wei, and Z. Li. A segmentation algorithm based on image projection for complex text layout. In *AIP Conference Proceedings*, volume 1890, page 030011. AIP Publishing, 2017.
6. G. Nagg, S. Seth, and M. Viswanathan. A protoptype document image analysis system for technical journals. *Computer.*, 25:10–22, 1992.
7. Costantino Grana, Giuseppe Serra, Marco Manfredi, Dalia Coppi, and Rita Cucchiara. Layout analysis and content enrichment of digitized books. *Multimedia Tools and Applications*, 75(7):3879–3900, 2016.
8. Yuanwang Wei, Zhijiang Zhang, Wei Shen, Dan Zeng, Mei Fang, and Shifu Zhou. Text detection in scene images based on exhaustive segmentation. *Signal Processing: Image Communication*, 50:1–8, 2017.
9. C. Yi and Y.L. Tian. Text string detection from natural scenes by structure-based partition and grouping. *IEEE Transactions on Image processing.*, 20(9):2594–2605., 2011.
10. B. Bai, F. Yin, and C. L. Liu. A seed-based segmentation method for scene text extraction. *IAPR International Workshop on Document Analysis Systems*, pages 262–266, 2014.
11. Z. Liu, Y. Li, X. Qi, Y. Yang, M. Nian, H. Zhang, and R. Xiamixiding. Method for unconstrained text detection in natural scene image. *IET Computer Vision*, 2017.
12. L. Gomez and D. Karatzas. Multi-script text extraction from natural scenes. *Proceedings of International Conference on Document Analysis and Recognition.*, pages 467–471, 2013.
13. X.C. Yin, X. Yin, and H.W. Hao K. Hung. Robust text detection in natural scene images. *IEEE Transactions on Pattern Aanlysis and Machine Intelligence.*, 36:970–983., 2014.
14. C. Shi, C. Wang, B. Xiao, and Y. Zhang. Scene text detection using graph model built upon maximally stable extremal regions. *Pattren Recognition letter .*, 34:107–116., 2013.
15. Y. Li and H. Lu. Scene text detection via stroke width. *Proceedings of International Conference on Pattern Recognition*, pages 681–684, 2012.
16. H. Cho, M. Sung, and B. Jun. Canny text detector: Fast and robust scene text localization algorithm. *International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3566–3573, 2016.
17. J. Park, G. Lee, E. Kim, J. Lim, S. Kim, H. Yang, M. Lee, and S. Hwang. Automatic detection and recognition of korean text in outdoor signboard images. *Pattern Recognition Letters*, 31(12):1728–1739, 2010.
18. W. Huang, Z. Lin, J. Yang, and J. Wang. Text localization in natural images using stroke feature transform and text covariance descriptors. *Proceedings of the IEEE International Conference on Computer Vision*, pages 1241–1248, 2013.
19. Haojin Yang, Bernhard Quehl, and Harald Sack. A framework for improved video text detection and recognition. *Multimedia Tools and Applications*, 69(1):217–245, 2014.
20. A. Zhu, G. Wang, and Y. Dong. Detecting natural scenes text via auto image partition, two-stage grouping and two-layer classification. *Pattern Recognition Letters*, 67:153–162, 2015.
21. P. He, W. Huang, T. He, Q. Zhu, Y Qiao, and X. Li. Single shot text detector with regional attention. *Proceedings of International Conference on Computer Vision (ICCV),*, 2017.
22. W. Huang, Yu. Qiao, and X. Tang. Robust scene text detection with convolution neural network induced MSER trees. *Proceedings of European Conference on Computer Vision*, LNCS 8692:497–511, 2014.
23. T. He, W. Huang, Y. Qiao, and J. Yao. Text-attentional convolutional neural network for scene text detection. *IEEE Transactions on Image Processing*, 25:2529–2541, 2016.
24. Z. Tian, W. Huang, T. He, P. He, and Y. Qiao. Detecting text in natural image with connectionist text proposal network. *Proceedings of European Conference on Computer Vision (ECCV)*, 2016.
25. Y. Zhu, C. Yao, and X. Bai. Scene text detection and recognition:recent advances and future trends. *Frontiers of Computer Science*, 10:19–36, 2016.

26. S. H. Kim, K. J. An, S. W. Jang, and G. Y. Kim. Texture feature-based text region segmentation in social multimedia data. *Multimedia Tools and Applications*, 75(20):12815–12829, 2016.
27. C.W. Liang and P. Y. Chen. Dwt based text localization. *International Journal of Applied Science Engineering.*, 2:105–116., 2004.
28. M. Acharyya and M.K. Kundu. Document image segmentation using wavelet scale-space features. *IEEE Transactions on circuits and systems on video technology.*, 12(12):1117–1127., 2002.
29. W. Chan and G. Coghill. Text analysis using local energy. *Pattern Recognition* ., 34:2523–2532., 2001.
30. E. L. Pennec and S. Mallat. Image compression with geometrical wavelets. In *Image Processing, 2000. Proceedings. 2000 International Conference on*, volume 1, pages 661–664, 2000.
31. M. N. Do and M. Vetterli. The contourlet transform: an efficient directional multiresolution image representation. *Proceedings of British Machine Vision Conference*, 14(12):20912106., 2005.
32. S. Roy, M.K. Kundu, and G.H. .Granlund. Uncertainty relations and time-frequency distributions for unsharp observables. *Information Sciences*, 89:193–209, 1996.
33. S. Dhar and M.K. Kundu. Accurate segmentation of complex document image using digital shearlet transform with neutrosophic set as uncertainty handling tool. *Applied Soft Computing*, 61:412–426, 2017.
34. M.K. Kundu, S. Dhar, and M. Banerjee. A new approch for segmentation of image and text in natural and commercial text documents. *Proceedings of International Conference on Communications ,Devices and Intelligent system.*, pages 86–88., 2012.
35. P. Maji and S. Roy. Rough -fuzzy clustering and multiresolution image analysis for text-graphics segmentation. *Applied soft computing.*, 30:705–721., 2015.
36. N. N. Karnik and J. M. Mendel. Introduction to type-2 fuzzy logic systems. *Proceeding of International Conference on Fuzzy Systems.*, pages 915–920., 1989.
37. M.K. Kundu and S.K Pal. Thresholding for edge detection using human psychovisual phenomena. *Pattern Recognition Letters*, 4(6):433–441, 1986.
38. R. Peng and P.K. Varshney. On performance limit of image segmentation algorithms. *Computer Vision and Image understanding*, 132:24–38, 2015.
39. A. L. Da. Cunha, J. Zhou, and M. N. Do. The nonsubsampled contourlet transform: Theory, design and applications. *IEEE Transactions on Image Processing*, 15(10):3089–3101, 2006.
40. E. J. Candes and D. L. Donoho. New tight frames of curvelets and optimal representations of objects with singularities. *Communications on Pure and Applied Mathematics*, 57:219–266, 2003.
41. E. Szmidt and J. Kacprzyk. Entropy for intuitionistic fuzzy sets. *Fuzzy sets and systems*, 118(3):467–477, 2001.
42. H. Bustince, E. Barrenechea, and M. Pagola. Restricted equivalence functions. *Fuzzy Sets and Systems*, 157:2333–2346, 2006.
43. F. López, J. Valiente, R. Baldrich, and M. Vanrell. Fast surface grading using color statistics in the cie lab space. In *Iberian Conference on Pattern Recognition and Image Analysis*, pages 666–673, 2005.
44. P. Mitra, C.A. Murty, and S.K. Pal. Unsupervised feature selection using feature similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence* ., 24(3):301–312., 2002.
45. M.K. Kundu and M. Acharyya. M-band wavelets:application to texture segmentation for real life image analysis. *Internation Journal of Wavelets, Multiresolution and Information Processing.*, 1(1):115–119., 2003.
46. ICDAR2015 dataset. `http://rrc.cvc.uab.es/`. 2015.
47. Kaist scene text database. `www.iapr-tc11.org/mediawiki/index.php/KAIST_Scene_Text_Database`. 2011.
48. M. Everingham, L. V. Gool and C. K. I. Williams J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88:303–338, 2010.
49. R. Peng and P.K. Varshney. On performance limit of image segmentation algorithms. *Computer Vision and Image understanding*, 132:24–38, 2015.
50. T. Kobchaisawat and T. H. Chalidabhongse. Thai text localization in natural scene images using convolutional neural network. *Signal and Information Processing Association Annual Summit and Conference(APSIPA)*, 2014.

Soumyadip Dhar received his B.E and M.E degree in Computer Science & Engineering from University of Burdwan and West Bengal University of Technology. He has contributed 2 well known and prestigious journals and several conferences. Currently he is Assistant Professor in department of IT in RCC Institute of Information Technology. His current research interest includes Image/ processing & analysis, soft computing, and machine intelligence.

Malay K. Kundu received his B. Tech., M. Tech. and Ph.D (Tech.) degrees in Radio physics and Electronics all are from the University of Calcutta. Currently he is a full professor in the Machine Intelligence Unit of the Indian Statistical Institute, Kolkata, India. He had been the head of the Machine Intelligence Unit from September 1993 to November 1995 and Professor In-charge (Chairman) of the Computer & Communication Sciences Division of the Institute during 2004 to 2006. He is a Fellow of a The International Association for Pattern Recognition, USA (FIAPR), Indian National Academy of Engineering (FNAE), National Academy of Sciences (FNASc.), India and the Institute of Electronics and Telecommunication Engineers (FIETE), India. A senior member of the IEEE, USA and the founding life member & Vice President of the Indian Unit for Pattern Recognition and Artificial Intelligence (IUPRAI).He was selected as INAE Distinguished Professor in 2013.
His current research interest includes Image/video processing & analysis, soft computing, computer vision, machine intelligence and data security. He has contributed 3 book volumes, about 140 research papers in well known and prestigious archival journals, international refereed conferences and in the edited monograph volumes. He is the holder of nine U.S patents.

Author Photographs

Soumyadip Dhar

Malay K Kundu