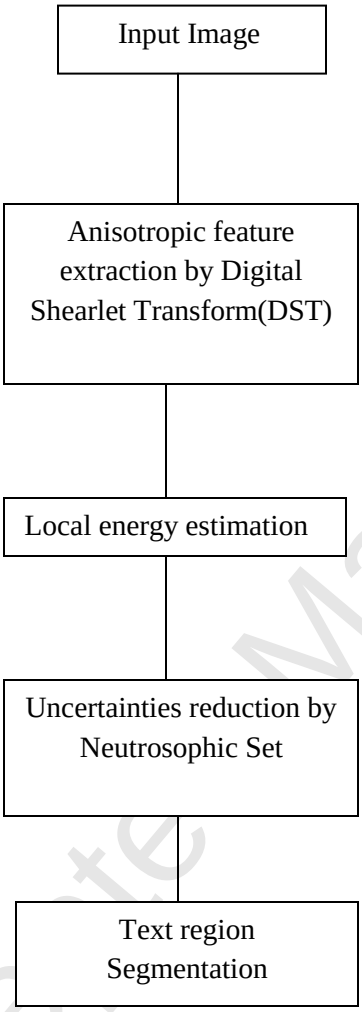
Highlights

- In this paper, a novel text region segmentation method based on Digital Shearlet Transform(DST) is proposed.
- The uncertainties in the DST features are handled by Neutrosophic set theoretic approach.
- The proposed method shows robustness under different perturbations.

Graphical Abstract



Accurate segmentation of complex document image using Digital Shearlet Transform with Neutrosophic Set as uncertainty handling tool

Soumyadip Dhar[a], Malay K. Kundu[b]

[a]*RCC Institute of Information Technology ,Kolkata*-700015,*India, rccsoumya@gmail.com*
[b]*Indian Statistical Institute,Kolkata*-700108,*India, malay@isical.ac.in*

Abstract

In any image segmentation problem, there exist uncertainties. These uncertainties occur from gray level and spatial ambiguities in an image. As a result, accurate segmentation of text regions from non-text regions (graphics/images) in mixed and complex documents is a fairly difficult problem. In this paper, we propose a novel text region segmentation method based on Digital Shearlet transform(DST). The method is capable of handling the uncertainties arising in the segmentation process. To capture the anisotropic features of the text regions, the proposed method uses the DST coefficients as input features to a segmentation process block. This block is designed using the Neutrosophic set(NS) for management of the uncertainty in the process. The proposed method is experimentally verified extensively and the performance is compared with that of some state-of-the-art techniques both quantitatively and qualitatively using benchmark dataset.

Keywords: *Shearlet* , *Digital Shearlet transform(DST)*, *neutrosophic set*, *uncertainty handling*, *document image*, *segmentation*

1. Introduction

Due to the rapid development of digital technology over the last few decades, usage of digital documents becomes a common practice. The practice is due to low cost for storage, easy storage, and transportability of the digital documents. Automation of digital document analysis involves the segmentation of text regions from non-text regions/graphics. Proper text and non-text region segmentation help to store and retrieve the text documents

efficiently. In the area of text and graphics region segmentation, a lot of research work has already been reported so far. These works include segmentation of structurally complex (color and gray level) document and the camera captured natural scene images.

The conventional approaches to image segmentation consist of hard partitioning the image space into meaningful regions by extracting its features. But, the regions are not crispy defined due to incomplete or imprecise input information, the ambiguity/vagueness in an input image, ill-defined and /or overlapping boundaries among the regions, and the indefiniteness in defining/extracting features. An image possesses ambiguity within each pixel because of the possible multi-valued levels of brightness. The uncertainty in an image pattern may be explained in terms of gray level [1] or spatial ambiguity [2] or both. Gray level ambiguity means " indefiniteness" in defining a pixel as white or black(i.e object or background) and it considers global information. Spatial ambiguity refers to "indefiniteness" in shape and geometry(e.g defining centroid, sharp edge etc.) within the image[3]and it takes care of local information. The uncertainties may increase in images due to noise, rotation and dynamic gray level change. Due to the uncertainties, the proper feature generation for segmentation process is difficult. In order to locate the segmentation boundary between the regions accurately, it is necessary that the segmentation process should be capable of handling the uncertainties in an effective manner. The ultimate output of the process will be associated with least uncertainty and the output retains as much of the 'information content' of the image as possible. Thus, it is natural and convenient to avoid committing ourselves to a specific(hard) decision about the segments. The segments should be represented by the subsets which are characterized by the degree to which each pixel belongs to them.

1.1. Related work

Conventional methods for text region segmentation include top-down and bottom-up approaches. In the top-down approach, at first, the probable text regions are identified. Then the regions are split into paragraphs, text-line, and words. X, Y-cut algorithm [4], which works by detecting white space using horizontal and vertical projection, is one of the most popular top-down approaches. Chen et al. [5] proposed a top-down approach using canny

filter and a vertical, horizontal edge map to detect the text regions. Alex Chen et al. [6] in his top-down approach identified the text regions of an image by their statistical properties. On the contrary, bottom-up approach attempts to cluster the pixels and then group together to obtain the text and non-text regions. A bottom-up approach by a block-wise text pixel segmentation method was proposed by Haneda et al. [7]. Here the initial segmentation was refined by a connected component classification (CCC). Yi et al. [8] developed a bottom-up algorithm for segmentation of text in natural scenes based on adjacent character grouping method and text line grouping method. Their method suffered from difficulties like the segmentation failure for a ligature, multicolored text regions, and text sets comprise of less than three characters. Currently, the bottom-up approach, based on Maximally Stable External Regions (MSERs) [9] has gained popularity. The MSERs are based on the idea of taking regions which stay nearly same through a wide range of thresholds. Gomez et al. [10] proposed an algorithm for text extraction based on MSERs and human perception of text in natural scenes. MSERs based text region segmentation was also used by [11, 12, 13, 14]. Another bottom-up approach based on seed generation was proposed by Bai et al. [15].Cho et al.[16] proposed a text region segmentation based on MSERs and canny edge detector. However, the methods described above, are sensitive to character size, scanning resolution, inter-line, and inter-character spacing. They suffer from low accuracy rate when prior knowledge about the image content is not available.

Apart from the above mentioned methods, some researchers used deep learning based methods for the text region detection where Convolutional Neural Network(CNN) was used for text region classification. He et al. [17] used CNN for text component filtering and incorporated Contrast-Enhanced MSERs(CE-MSERs) for text region detection. Huang et al. [18] proposed a deep learning based method, which integrated the MSERs and CNN. In the method, MSERs was used for text region detection and CNN based classifier was utilized for true identification of text regions. However, the method and other deep learning based methods were computationally costly, as they required tremendous data for training and the methods were text dependent. The performance boost of deep learning methods may be due to the training with huge amount of data which may not be publicly available [19].

On the other hand they could only detect horizontal or near horizontal text regions. These two limitations may restrict the practical application of deep learning based methods [19].

Another approach is used by the researchers where it is assumed that text regions have distinctive texture properties. The text region texture properties are quite different from the background texture properties. So segmentation of text from non-text regions is treated as a texture segmentation problem. Zhou et al. [20] combined the three different texture features, such as oriented gradient (HOG), mean of gradients (MG) and local binary patterns(LBP) for text region segmentation. Some of the techniques of this approach use Gabor filter, Wavelet transforms etc., as a representation tool for feature extraction. The proposed method uses a similar methodology for text and non-text region segmentation. The dyadic wavelet based text region segmentation was reported in [21, 22]. But the dyadic wavelet coefficients do not yield rotation invariant features. The limitation can be overcome using Gabor filter. Chan et al. [23] proposed a method for text region segmentation that involves computation of local energy for texture using a bank of orthogonal pairs of Gabor filter. Gabor filter bank was also used for generating features by Nirmal et al. [24] to detect the text regions. The limitation of the methods is that Gabor filter banks come with the high computational costs. To overcome the limitation of both the dyadic wavelet transform and Gabor filter, M-band wavelet transform and M-band wavelet packet transform are preferred by many researchers. Acharyya et al. [25] used M-band wavelet transform to segment the text regions from gray images. Kumar et al. [26] designed matched wavelet and Markov Random Field model for document image segmentation. But, the Wavelet transform can only handle the point singularities and cannot capture the curve like features properly.

The methods described above and most of the methods reported in the literature do not have the capacity to capture the anisotropic features (edges, curves) of the text regions properly. Moreover, they cannot handle the uncertainties inherent in an image and the uncertainties that can arise in feature generation [27] . Maji. et al. [28] used a rough-fuzzy clustering technique on M-band packet wavelet features to handle the uncertainties. Kundu et al. [29] proposed a text region segmentation method combining M-band packet wavelets and fuzzy c-means algorithm. But their uncertainty handling models were used during the

segmentation only and had no effect on feature generation. As a result, the performance of the methods did not achieve the expected level of accuracy. So, that is the major motivation of the current investigation for a better solution.

1.2. Proposed Method

We propose an accurate text region segmentation method which can capture the curve and edge like feature of text regions in an image and also handle the uncertainties in the image during feature generation. The features help to represent the two different textures of text and non-text regions properly. The method is based on Neutrosophic Set(NS) and multi-resolution analysis of the image using Digital Shearlet Transform(DST) transform [30]. The motivation for using DST is that, it can handle the anisotropic features better than the wavelet transform. Moreover, it has better multi-directional, shift invariant and excellent multiscale image decomposition property than the curvelet transform and contourlet transform [31, 30].

Neutrosophic Set was proposed by Florentin Smarandache as a new branch of philosophy dealing with the origin, nature and scope of neutralities [32]. It has a powerful capacity to deal with the uncertainty and is better than other uncertainty handling model [33]. The concept was successfully used in image thresholding [34], image denoising [35], image segmentation [33] and color texture image segmentation [36]. In NS theory, every event has not only a certain degree of truth but also a falsity degree and an indeterminacy degree that have to be considered independently from each other [32] and represented as true, false and indeterminate set respectively.

In the proposed feature extraction method we use the DST to transform the image into different shearlets translates(sub-bands)which have different scales and orientations. For each shearlet coefficient in a sub-band, the energy is computed from the transform coefficients over a overlapping window (size adaptively varied) around each pixel. The transform coefficients are itself rotational invariant. The anisotropic features of the text regions are captured by the DST efficiently. The anisotropic features represent the text regions irrespec-

tive of the character size, inter-line, and inter-character spacing. Thus, the DST captures the texture property of the text regions to differentiate it from the non-text/graphics regions. The texture represents the assembly of text fonts, unlike the conventional methods which used textual and graphical attributes like font size, text line orientation etc for text region segmentation. So the proposed method can work in a generic environment. But, DST may introduce uncertainties in features due to the presence of redundant feature (multiple numbers of shearlet sub-bands) information. This is in addition to conventional type uncertainties present due to the discrete gray level ambiguity, the spatial ambiguity which may increase due to different perturbations and low scanning resolutions etc. These are major reasons for inaccuracy for any segmentation process in locating true segmentation boundaries. So in order to tackle this problem effectively, we have to use the NS as an uncertainty handling model, which can reduce error due to uncertainties for achieving the better segmentation.

In the proposed uncertainty handling scheme, at the beginning, the features(local energy of shearlet coefficients) are mapped into NS domain in order to classify them into three different sub sets, the true, indeterminate and false. In this stage, the uncertainties are reduced iteratively with the increase of values of entries in true and false subsets along with the reduction values of entries in indeterminate sub sets. After this reduction of the uncertainties, the feature selection is done on the true subsets by an unsupervised feature selection algorithm to generate finer features. This step helps to reduce the uncertainties due to the presence of a large number of redundant features. With this re-categorized feature subset, the final segmentation is done using NS based clustering. This is why in the proposed method the two stage uncertainty handling capabilities is expected to give better results than the other existing uncertainty handling models having no such provision of successive uncertainty reduction mechanism.

1.3. Novelty of the proposed method

The novelties or the major contributions of the of the paper are

(1) : We propose an efficient texture based method by the DST for text region segmentation. The highly efficient rotation and translation invariant anisotropic features generated by the shearlets in DST help to capture the different textures of text and non-text regions in a complex background. Thus, it greatly improves the performance of text region segmentation over the state-of-the-art methods.

(2) :In the proposed method, the uncertainties in the sub-bands of the DST are handled by the neutrosophic set. The uncertainties are due to spatial and gray level ambiguity in an image. Moreover, additional uncertainties are introduced due to redundant information generation of the shearlet sub-bands in different scales. The uncertainties in the NS domain are handled in two steps. In the first step, the uncertainties in a feature itself are reduced and in the second step uncertainties during the segmentation are handled. For this in the first step, iteratively the uncertainties in each feature are reduced in NS domain. On the top of that to generate the finer features, features selection in the NS domain based on Maximal information compression index(MCI) is done. In the second step, uncertainty during the segmentation is reduced by clustering in the NS domain. This two-step uncertainty reduction process in text/non-text segmentation is more powerful than the conventional methods which handle the uncertainties by fuzzy-c-means or rough-fuzzy c-means with no such provision of two-step uncertainty reduction. They handle the uncertainties only during the segmentation process. Such process has no effect on feature generations.

(3) : The proposed method is tested under different perturbations i.e noise corruption, rotation, and dynamic gray level changes. Compared to the state-of-the-art methods, our method shows satisfactory robustness under the different perturbations.

The paper is organized as follows. Section 2 describes DST, representation of NS components and unsupervised feature selection. The proposed method and the algorithm are presented in the Section 3. The Section 4 comprises of results discussion, comparison with other methods and performance evaluation.

2. Theoretical Preliminaries

2.1. Shearlet system

Shearlet systems are designed to efficiently encode anisotropic features such as singularities concentrated on lower dimensional embedded manifolds. To achieve optimal sparsity, shearlets are scaled according to a parabolic scaling law encoded in the parabolic scaling matrix A_a, $a > 0$ and exhibit directionality by parameterizing slope encoded in the shear matrix S_s, $s \in \mathbb{R}$, defined by

$$\mathrm{A}_a = \begin{bmatrix} a & 0 \\ 0 & \sqrt{a} \end{bmatrix} \text{ and } \mathrm{S}_s = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}$$

For appropriate choices of the shearlet $\psi \in \mathrm{L}^2(\mathbb{R}^2)$, the Continuous Shearlet Transform

$$\mathcal{SH}_\psi : \mathrm{f} \to \mathcal{SH}_\psi \mathrm{f}(a, s, t) = \langle \mathrm{f}, \psi_{ast} \rangle \tag{1}$$

is a linear isometry from $\mathrm{L}^2(\mathbb{R}^2)$ to $\mathrm{L}^2(\mathbb{S})$. Hence, shearlet systems are based on three parameters: $a > 0$ being the scale parameter measuring the resolution level, $s \in \mathbb{R}$ being the shear parameter measuring the directionality, and $t \in \mathbb{R}^2$ being the translation parameter measuring the position. When $s < |1|$, this produces the cone adapted Continuous Shearlet Transform. It allows an equal treatment of all directions in contrast to a slightly biased treatment by the Continuous Shearlet Transform.

A discrete shearlet transform for $\psi \in \mathrm{L}^2(\mathbb{R}^2)$, is a collection of functions of the form

$$\psi_{j,k,m} = 2^{3j/4} \phi(\mathrm{S}_k \mathrm{A}_{2^j} \cdot - m) : j \in \mathbb{Z}, k \in K \subset \mathbb{Z}, m \in \mathbb{Z}^2 \tag{2}$$

where K is a carefully chosen indexing set of shears. Note that the shearing matrix S_k maps the digital grid $\mathbb{Z}^2$ onto itself, which is the key idea for deriving a unified treatment of the continuum and digital setting.The discrete shearlet system defines a collection of waveforms at various scales j, orientations controlled by k, and locations dependent on m. To avoid the biased treatment of directions which the discrete system inherit, the cone adapted shearlet system is defined as

$$\mathcal{SH}(\phi, \psi, \widetilde{\psi}) = \{\phi(\cdot - m) : m \in \mathbb{Z}^2\} \cup \{\psi_{j,k,m}, \widetilde{\psi}_{j,k,m} : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\} \tag{3}$$

where $\widetilde{\psi}_{j,k,m}$ is generated from $\psi_{j,k,m}$ by interchanging both variables, and $\psi_{j,k,m}$, $\widetilde{\psi}_{j,k,m}$ and ϕ are L^2 functions.

2.2. Digital Shearlet Transform

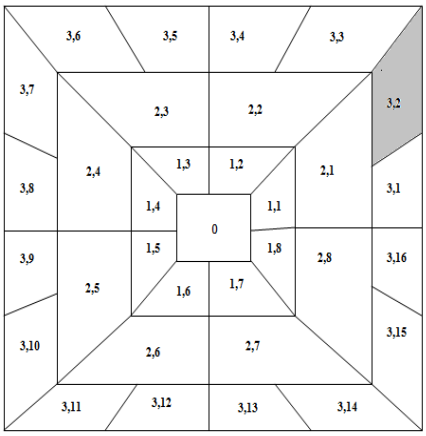


Figure 1: Digital Shearlet transform of an image into 4 levels. The 0th level is the low frequency shearlet. Each index represents the level and orientation of the shearlet in that level e.g the shearlet (3,2) represents the level 3 with orientation 2.

In the proposed method the Digital Shearlet transform [37] is based on cone adapted discrete shearlet system with compactly supported shearlets. In compactly supported shearlet system generator, it is conjectured that no tight shearlet frame exist. The DST is constructed in three steps: (1)A non-separable structure of shearlet generation, (2)digitization of shearlet operators and (3)generation of digital shearlet filter.

A non-separable structure of shearlet generator $\hat{\psi}$ is defined as

$$\hat{\psi}(\xi) = \mathrm{P}(\frac{\xi_1}{2}, \xi_2)\psi_1 \otimes \phi_1(\xi) \tag{4}$$

where P is a 2d directional filter, ϕ_1 is 1D scaling function associated with wavelet MRA and ψ_1 is the corresponding 1D wavelet function and $\xi = (\xi_1, \xi_2) \in \hat{\mathbb{R}^2}$, $\hat{\mathbb{R}^2}$ being the plane in the Fourier domain. Because of $\psi_{j,k,m} = \psi_{j,0,m}(S_{k/2^{j/2}})$, two ingredients are required for digitization of shearlets, Digital shearlet filter $\psi^d_{j,0}$ and digital shearlet operator $\mathrm{S}^{\mathrm{d}}_{k/2^{j/2}}$. For any dicreate 2D signal the shearlet operator is defined as $\mathrm{S}^{\mathrm{d}}_{k/2^{j/2}}(x) = ((x_{\uparrow 2^{\mathrm{d}\alpha_j}} *_1 h_{j/2})(\mathrm{S}_k \cdot) *_1$

$\overline{h_{j/2}})_{\downarrow 2^{\mathrm{d}\alpha_j}}$, where $x \in \mathrm{L}^2(\mathbb{Z}^2)$,$\uparrow 2^{\mathrm{d}\alpha_j}$, $\downarrow 2^{\mathrm{d}\alpha_j}$ and $*_1$ are upsampling, downsampling and convolution operator along X axis and $\mathrm{d}_{\alpha_j} = \lceil j(2-\alpha_j)/2 \rceil$. Here the $\alpha_j(=1)$ measures the degree of anisotropy for each scale $j \geq 0$. Then the digital shearlet filter is given by

$$\psi^{\mathrm{d}}_{j,k} = \mathrm{S}^{\mathrm{d}}_{k/2^{j/2}}(x) * \mathrm{p}_j * (\mathrm{g}_{J-j} \otimes \mathrm{h}_{J-j/2}) \tag{5}$$

Where p_js are the Fourier coefficients of $\mathrm{P}(2^{\mathrm{J}-j-1}\xi_1, 2^{\mathrm{J}-\frac{j}{2}}\xi_2)$,$\mathrm{h}_{\mathrm{J}-j/2}$ is the low pass filter associated with the scaling function ϕ_1,$g_{\mathrm{J}-j}$ is the corresponding high pass filter, associated with the wavelet function ψ_1 and $\mathrm{J} \in \mathbb{N}$ is the highest scale to be considered(i.e $j < \mathrm{J}$ for all shearlets $\psi_{j,k,m}$). Now the digital shearlet transform of a digital signal $\mathrm{f} \in \mathrm{L}^2(\mathbb{Z}^2)$ is given by

$$\mathrm{DST}^{2D}_{j,k,m} = \mathcal{SH}^{\mathrm{d}}_{j,k,m} = (\overline{\psi^{\mathrm{d}}_{j,k}} * f) \tag{6}$$

for $j \in \{0, \mathrm{J}-1\}$ and $|k| < \lceil 2^{j/2} \rceil$

The shearlet systems use regular translations on the integer lattice with the shearing operations. The system provides the structure of the integer grid with directionality. Obviously, these two properties lead to an implementation of the digital shearlet transform exploiting discrete convolutions. This allows a shift-invariant transform by simply skipping the anisotropic downsampling. As a result, digital shearlet transform is highly redundant. The shearlet sub-bands generated by the DST have the same size of the input. Mainly, in two ways the uncertainties(Figure 2) are to be handled in the features generated by the DST. They are

1 : Each shearlet coefficient (feature) of a shearlet sub-band represents one pixel in one scale and orientation. Thus, the contrast between the coefficients in each shearlet sub-band should be increased to reduce the uncertainty due to gray and spatial level ambiguity in an image.

2 :The total number of shearlet sub-band generated in DST is $\sum_{j=0}^{nscales-1} 2^{\lceil j/2 \rceil + 2}$, where *nscales* represents the number of scales of the DST. The redundant information due to different scales and directions also increases the uncertainty. Proper selection of the sub-bands reduces the uncertainty.

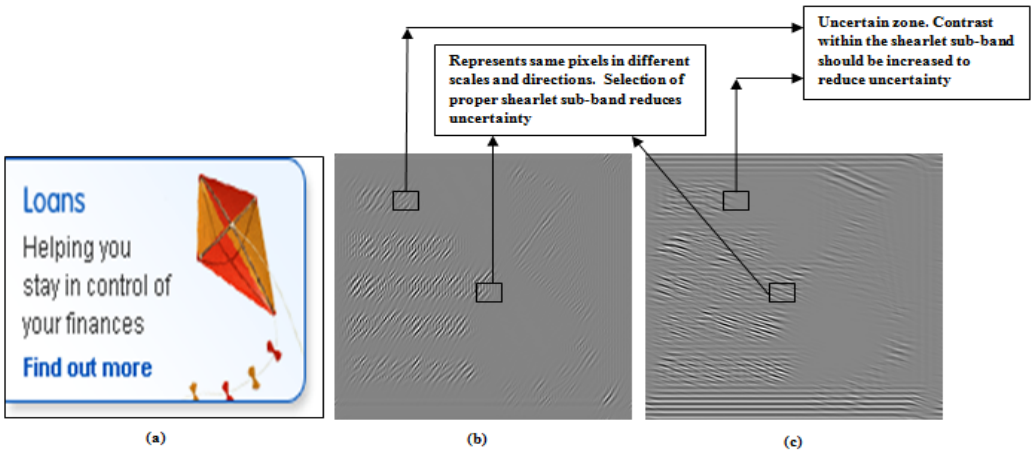


Figure 2: (a) Original image of size 256×256 (b) and (c) shows the two shealets(size 256×256) by the DST of the image in two different resolutions and orientations with uncertain zones.

Apart from the above, uncertainties may occur during segmentation of the features. These uncertainties make it difficult to decide the correct class of a pixel. The uncertainties also increase due to different perturbations. To handle the above uncertainties the Neutrosophic set and Neutrosophic logic are used in the proposed method.

2.3. Neutrosophic set (NS):

Let U be a universe of discourse, and a neutrosophic set $A \subset$ U. An element x from U is noted with respect to A as $x(T, I, F)$ and belongs to A in the following way [32]:
x's degree of belongings in the true subset T is $t\%$, in the indeterminate subset I is $i\%$ and in the false subset F is $f\%$. T,I and F are real standard or non-standard subsets in the open interval of $]^{-}0, 1^{+}[$
with $supT = t\_sup, infT = t\_inf$, $supI = i\_sup, infI = i\_inf$, $supF = f\_sup, infF = f\_inf$ and $n\_sup = t\_sup + i\_sup + f\_sup$, $n\_inf = t\_inf + i\_inf + f\_inf$. In the Neutrosophic Logic [32], the truth (T) and the falsity (F) and the indeterminacy (I) can be any numbers in $[0, 1]$, then $0 \leq T + I + F \leq 3$. For most real world applications, $T, I, F \subset [0, 1]; t + f = 1$ and $i \in [0, 1]$ [33]. In the next section, we will discuss how neutrosophic set can be used as an uncertainty handling tool in the context of image segmentation problem.

2.4. Neurotrosophic image representation for handling uncertainty :

Let r(i,j) be the gray level of the (i,j)th pixel of a P $\times$ Q dimensional L level image B = [r(i,j)],$i = 1,2,..,$P,$j = 1,2,....$Q. When the image is mapped into the neutrosophic domain, the corresponding image is called a neutrosophic image $\mathrm{B_{NS}}$ [36]. It is characterized by three subsets T,I and F. A pixel in an NS domain can be represented as (t,i,f) where the pixel is t true, i indeterminate and f false, and t,i,f belongs to the true subset T, the indeterminate subset I and the false subset F respectively. A pixel B(i,j) in the image domain is mapped into NS domain as

$$\mathrm{B_{NS}}(i,j) = \{T(i,j), I(i,j), F(i,j)\}$$

where $T(i,j)$, $I(i,j)$, and $F(i,j)$ are the membership values t,i,f belonging to true subset T, indeterminate subset I and false subset F respectively. They are represented by the following equations.

$$T(i,j) = 1 - \frac{\bar{r}_{max} - \bar{r}(i,j)}{\bar{r}_{max} - \bar{r}_{min}} \tag{7}$$

$$I(i,j) = 1 - \frac{d_{max} - d(i,j)}{d_{max} - d_{min}} \tag{8}$$

$$d(i,j) = \mathrm{abs}(r(i,j) - \bar{r}(i,j)) \tag{9}$$

$$d_{\max} = \max\{d(i,j) | i \in \mathrm{P}, j \in \mathrm{Q}\} \tag{10}$$

$$d_{\min} = \min\{d(i,j) | i \in \mathrm{P}, j \in \mathrm{Q}\} \tag{11}$$

$$F(i,j) = 1 - T(i,j) \tag{12}$$

where $\bar{r}(i,j)$ is the mean value of the pixels within a local overlapping window of size $w \times w$ around $r(i,j)$ and its position is at the center of the local window. $\bar{r}_{max}$ and $\bar{r}_{min}$ are the maximum and minimum values of $\bar{r}(i,j)$ $\forall i \in \mathrm{P}, \forall j \in \mathrm{Q}$. $d(i,j)$ is the absolute value of the difference between element $r(i,j)$ and its local mean value $\bar{r}(i,j)$. The value of $I(i,j)$ is employed to measure the indeterminacy degree of a pixel $\mathrm{B_{NS}}(i,j)$. Here $I(i,j)$ gives the degree of uncertainties of deciding brightness of the pixel. The indeterminacy of the NS image, which is measured by the entropy of the indeterminate subset, represents the uncertainties present in the gray level image. The changes in T and F influence the

distribution of element in I and the entropy of I. The NS domain representation of a gray-scale image can be found in Appendix A.

In NS domain, two operations, $\alpha - mean$ and $\beta - enhancement$, are used to reduce the indeterminacy of the neutrosophic image. After these two operations, the image becomes more uniform and homogeneous, and more suitable for segmentation. In the next two subsections, we will discuss the $\alpha - mean$ and $\beta - enhancement$ operations mathematically.

2.5. The $\alpha - mean$ operation :

The $\alpha - mean$ operation [33] transforms neutrosophic image pixel $\mathrm{B_{NS}}(i,j)$ to $\mathrm{B_{NS}}\alpha(i,j) = \{\bar{T}(i,j), \bar{I}(i,j), \bar{F}(i,j)\}$ where $\bar{T}$,$\bar{I}$ and $\bar{F}$ are the true, indeterminate and false subsets of $\mathrm{B_{NS}}\alpha$. In this operation, it is checked weather the indeterminate membership value of a pixel is higher than a predefined value α where $0 \leq \alpha \leq 1$. If so, the value is reduced by making the pixel homogeneous with neighboring pixels. Thus, the $\alpha - mean$ operation reduces the uncertainty due to the spatial ambiguity. It is represented as

$$\bar{T}(i,j) = \begin{cases} T(i,j) & \text{if } I(i,j) < \alpha \\ \bar{T}_\alpha(i,j) & \text{if } I(i,j) \geq \alpha \end{cases} \tag{13}$$

where $\bar{T}_\alpha(i,j)$ is the mean value of the pixels within a local overlapping window of size $w \times w$ around $T(i,j)$ and its position is at the center of the local window. Similar operations are done for $\bar{F}(i,j)$. After the operation on the subset T, the indeterminate subset becomes

$$\bar{I}(i,j) = 1 - \frac{\bar{d}_{Tmax} - \bar{d_T}(i,j)}{\bar{d}_{Tmax} - \bar{d}_{Tmin}} \tag{14}$$

$$\bar{d_T}(i,j) = \mathrm{abs}(\bar{T}(i,j) - \bar{\bar{T}}(i,j)) \tag{15}$$

$$\bar{d}_{Tmin} = \min\{\bar{d_T}(i,j) | i \in P, j \in Q\} \tag{16}$$

$$\bar{d}_{Tmax} = \max\{\bar{d_T}(i,j) | i \in P, j \in Q\} \tag{17}$$

where $\bar{d_T}(i,j)$ is the absolute value of the difference between the $\bar{T}(i,j)$ and its local mean value $\bar{\bar{T}}(i,j)$. $\bar{\bar{T}}(i,j)$ is calculated over a local overlapping window of size $w \times w$ around each $\bar{T}(i,j)$ after the $\alpha - mean$ operation on NS image.

2.6. The $\beta -$ enhancement operation

In NS domain,the $\beta-$*enhancement* operation [33] is used to enhance the true membership value $T(i,j)$. The value is enhanced if its corresponding $I(i,j)$ value is greater than a predefined value β where $0 \leq \beta \leq 1$. The operation reduces the uncertainty due to gray level ambiguity by lowering the indeterminacy in the neutrosophic image. The β-enhanced image $\mathrm{B}_{\mathrm{NS}\beta}$ is defined as

$$\mathrm{B}_{\mathrm{NS}\beta}(i,j) = (T'(i,j), I'(i,j), F'(i,j)) \tag{18}$$

$$T'(i,j) = \begin{cases} T(i,j) & \text{if } I(i,j) < \beta \\ T'_{\beta}(i,j) & \text{if } I(i,j) \geq \beta \end{cases} \tag{19}$$

$$T'_{\beta}(i,j) = \begin{cases} 2T^2(i,j) & \text{if } T(i,j) < 0.5 \\ 1 - 2(1 - T(i,j))^2 & \text{if } T(i,j) \geq 0.5 \end{cases} \tag{20}$$

$$I'(i,j) = 1 - \frac{d'_{Tmax} - d'_T(i,j)}{d'_{Tmax} - d'_{Tmin}} \tag{21}$$

$$d'_{Tmin} = \min\{d'_T(i,j) | i \in P, j \in Q\} \tag{22}$$

$$d'_{Tmax} = \max\{d'_T(i,j) | i \in P, j \in Q\} \tag{23}$$

$$d'_T(i,j) = \mathrm{abs}(T'(i,j) - \bar{T}'(i,j)) \tag{24}$$

Where $d'_T(i,j)$ is the absolute value of difference between the points $T'(i,j)$ and its local mean value $\bar{T}'(i,j)$ computed over a local overlapping window of size $w \times w$ around the $T'(i,j)$ after the $\beta -$ *enhancement* operation. Now, the membership values in the set T become more distinct and have high contrast.

2.7. Adaptive α and β selection

The α and β are the two important parameters for indeterminacy reduction in a neutrosophic set. These two parameters can affect the segmentation results. So to determine the parameters adaptively depending on the characteristics of individual image, we follow

the same strategy as [36]. The parameters α and β are computed based on entropy EnI of subset I as:

$$EnI = -\sum_{i=1}^{P}\sum_{j=1}^{Q} \mathrm{pb}(i,j)\log_2 \mathrm{pb}(i,j) \tag{25}$$

where $\mathrm{pb}(i,j)$ is the probability of an pixel value at (i,j) in the subset I.

$$En_{max} = \log_2 PQ \tag{26}$$

$$\alpha = \alpha_{min} + \frac{(\alpha_{max} - \alpha_{min})(EnI - En_{min})}{(En_{max} - En_{min})} \tag{27}$$

$$\beta = 1 - \alpha \tag{28}$$

Where $P \times Q$ is the dimension of the image. En_{max} and En_{min} are the maximum and minimum entropy value of I. The α varies in the range $[\alpha_{min}\ \alpha_{max}]$. The true membership values of the neutrosophic image after $\alpha - mean$ and $\beta - enhancement$ operations become the features for segmentation.

2.8. Unsupervised feature selection:

The unsupervised feature selection is used to select the compact set of significant features which has minimum correlation between them. Partitioning of the features is done using the feature similarity Maximal information compression index [38].

The Maximal Information compression index(λ_2) for two features x and y is given by

$$2\lambda_2(x,y) = var(x) + var(y) - \sqrt{(var(x) + var(y))^2 - 4var(x)var(y)(1 - \rho(x,y))^2} \tag{29}$$

Where $var()$ and $\rho()$ denote the variance of a variable and correlation coefficient between two variables respectively.The (λ_2) is the eigenvalue for the direction normal to the principal component direction of feature pair (x,y). The features are partitioned based on k-NN principle using the feature similarity. The compact set of features are chosen based on λ_2 which are the representative of k-neighbouring features. The value of (λ_2) is zero when the features are linearly dependent and increases as the amount of dependency decreases.

2.9. $\gamma - k -$ means clustering algorithm for NS domain:

We use $\gamma - k -$ means clustering [36] for segmentation in NS domain. The clustering algorithm is applied on the true subset of NS image after the $\alpha - mean$ and the $\beta - enhancement$ operations. Let the true subset and the indeterminate subset of NS image after these two operations become $T_{\alpha\beta}$ and $I_{\alpha\beta}$ respectively. Considering the effect of indeterminacy, the true subset $T_{\alpha\beta}$ is transformed into a new subset X for clustering as follows.

$$X(i,j) = \begin{cases} T_{\alpha\beta}(i,j) & \text{if } I_{\alpha\beta}(i,j) \leq \gamma \\ \bar{T}_{\gamma}(i,j) & \text{if } I_{\alpha\beta\gamma}(i,j) > \gamma \end{cases} \tag{30}$$

where $\bar{T}_{\gamma}(i,j)$ is calculated over a local overlapping window of size $w \times w$ around each $T_{\alpha\beta}(i,j)$. The objective function for the clustering is defined by

$$\mathrm{J}_{TC} = \sum_{l=1}^{k} \sum_{i=1}^{H} \sum_{j=1}^{W} \|X(i,j) - Z_l\|^2 \tag{31}$$

where k is the number of clusters and $H \times W$ represents the dimension of X. Since, in the proposed method the number of segments is two, one for text region and other non-text region, here $k = 2$.

$$Z_l = \frac{1}{n_l} \sum_{X(i,j) \in C_l} X(i,j) \tag{32}$$

where J_{TC} is a compactness measure, n is the number of data to be clustered, and C_l is the lth cluster.

3. Proposed Methodology

It is well known that in a mixed text document, the text regions and non-text/graphics regions have distinctly different texture characteristics. With this idea in mind, we extract the texture features by decomposing it into shearlet sub-bands using DST. The local energy of each shearlet coefficient is computed using a small overlapping window of adaptive size around the each coefficient [39]. In order to reduce the uncertainty present in the sub-bands, each of them is then mapped into NS domain. This is followed by feature selection and feature dimensionality reduction. The final feature set thus computed is used for feature

vector generation and segmentation. The algorithm is illustrated in the Figure 3 and the steps are explained below. In the figure

- B_1,B_2..B_n are the shearlet sub-band in DST.
- T_i,I_i,F_i are the true, indeterminate and false value of B_i where $i = 1, 2, ..n$.
- T_i' is the updated true value[Section2.6] of shearlet B_i after the iterative $\alpha - mean$ and $\beta - enhancement$ operations.
- Fe_1,Fe_2..Fe_m are the features after unsupervised feature selection.

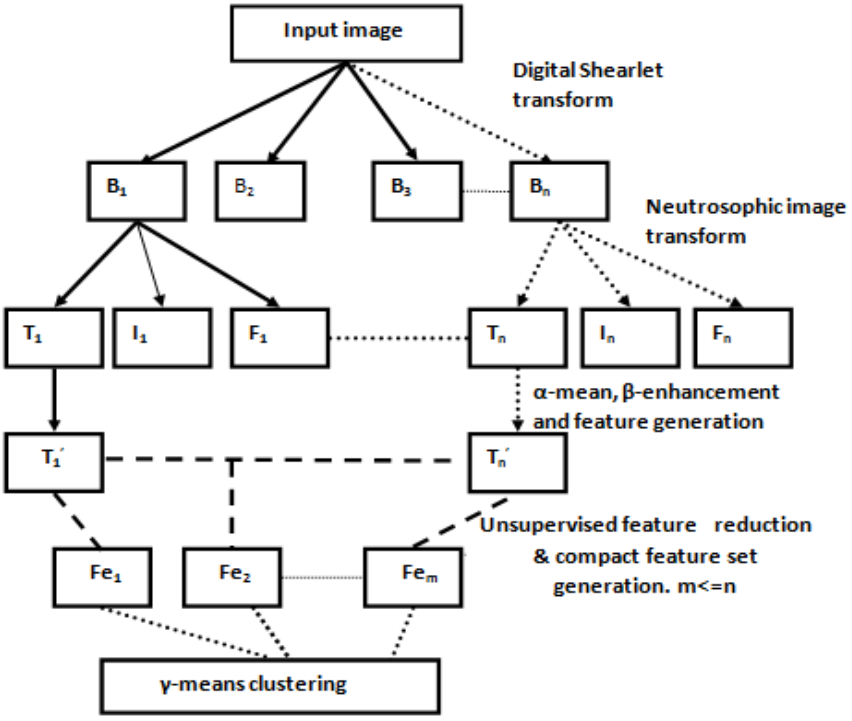


Figure 3: The schematic diagram of the proposed method.

Step 1 : DST transform of the image - In this step, the input image is transformed into DST shearlet sub-bands. Each sub-band contains the different scale and directional information of the image.

Step 2 : Local energy estimation and smoothing of the sub-bands:- In each sub-band, the local energy of a shearlet coefficient is computed by a nonlinear operation, calculated over a small adaptive overlapping window of size w around the coefficient. Then, the very low energy contained in a shearlet sub-band is removed by a Gaussian low-pass (smoothing) filters.

Step 3 : Mapping of the shearlet sub-bands to neutrosophic image:-. To handle the uncertainties in the sub-bands, each of them is mapped into the neutrosophic image.

Step 4 :The $\alpha-$ mean and the $\beta-$ enhancement operation on neutrosophic images:- In this step, to reduce the uncertainties in the neutrosophic images, each of them is subjected to the α-mean followed by the β-enhancement operation. These two operations are done repeatedly until the indeterminate subset entropy of the NS image in Eq 25 becomes unchanged. That means these two operations are done up to ith iteration until $|EnI(i) - EnI(i+1)| \leq \xi$, where ξ is a small positive quantity. The values of α and β are chosen adaptively(as described in sec 2.7). The true membership value T(i,j) of each neutrosophic image represents the textures at different scales and forms the feature for segmentation. The uncertainties within a sub-band is minimized in this step.

Step 5 : Unsupervised feature reduction and feature vector generation :- As the number of features is comparatively large, compaction of the feature set is necessary in order to achieve higher recognition accuracy and better computation efficiency. To get a compact set of features, the features are reduced and selected on the basis of unsupervised feature similarity(described in the section 2.8). The selected features thus used for generating the multidimensional feature vector. The uncertainties due to redundant information generation is reduced in this step.

Step 6 :The $\gamma-$ k $-$ means clustering for segmentation:- Finally the $\gamma - k - means$ clustering (described in section 2.9) is used for segmenting the feature vectors in two classes, comprising of text and non-text regions. The clustering reduces the uncertainties during segmentation.

3.1. Computational complexity of the proposed method :

The analysis of computational complexity (worst case) involves the following steps (1) Computation of R number of $P \times Q$ DST shearlet sub-bands upto scale j where $R = \sum_{j=0}^{nscales-1} 2^{\lceil j/2 \rceil + 2}$ (2) NS image generation and $\alpha - mean$, $\beta - enhancement$ operation (3) Unsupervised feature selection (4) $\gamma - K - means$ clustering. The overall complexity of the proposed method is $O(R \times P^2 Q^2)$, where the image size is $P \times Q$.

4. Experimental results and discussion.

4.1. Experimental setup and performance measure :

The proposed text region segmentation method combines the merits of multi-resolution analysis of DST and uncertainty representation using the NS. Experiments were done rigorously, and extensively to judge the ability of the proposed method. The importance of the parameters α and β for text region segmentation was also studied. It is to be noted that all the experiments were carried out without a prior knowledge about the input images. The performance of the proposed method was extensively compared qualitatively and quantitatively with some existing text region segmentation methods. The proposed method was performed using MATLAB2013 with Pentium IV processor. For the proposed method, the values of different parameters are given in the table 1.

Table 1: Values of different parameters used in the proposed method

$Parameter$	$Value$
w	5
En_{min}	0
α_{min}	0.01
α_{max}	0.1
ξ	0.001
γ	0.5

To judge the quantitative performance,four standard measures such as Recall R,Precision P and F-measure f were used [19]. The performance evaluation was based on Intersection-over-Union(IOU), with a threshold of 50%,following the standard practice in object recognition [40]. The measures are defined in Eq 33, Eq 34, and Eq 35 respectively [19].

$$R = \frac{|n1|}{|n2|} \tag{33}$$

$$P = \frac{|n1|}{|n3|} \tag{34}$$

$$f = \frac{2 \times P \times R}{(P + R)} \tag{35}$$

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.

4.2. Datasets:

To verify the efficiency of the proposed method, in the present experiment, we used the document images from the ICDAR2015 [41], ICDAR 2003 [42],ICDAR2011, ICDAR2013, KAIST [43] and also the scanned images from newspaper, magazine, advertisement found on the publicly available websites. The size of the images varied from 120×120 to 700×700.

For quantitative performance measure, different methods used different datasets and some of them are not publicly available. Also, the protocols for measuring the performances were different. So we have to apply them on the same dataset and use the same protocol for performance measure to make the comparisons fair. For quantitative performance measure, we applied all of them to the ICDAR2015 born digital dataset which contains 141 images and used the protocol in [40] for performance measure. For the text region segmentation task the ground truth data is provided in the dataset in terms of word bounding boxes. The reason for applying on this dataset is that born-digital images are inherently low resolution [44]. So ambiguities in the images are more and they are difficult for segmentation. This is the motivation for using this dataset and we used the updated version of it. For measuring the performance under different perturbations we used the same dataset. We also applied our method on KAIST dataset for the quantitative performance measure. It

Table 2: Average performance comparison of segmentation results obtained by different methods using gray level test images from ICDAR2015(Born digital)

Methods	*R*	*P*	*f*
Acharyya [25]	0.58	0.60	0.59
Kumar [26]	0.61	0.63	0.62
Gomez [10]	0.72	0.60	0.66
Maji [28]	0.80	0.82	0.81
Proposed	0.83	0.85	0.84

contains 395 test images for text region segmentation.

4.3. Experiments on gray document images :

To show the effectiveness of the proposed method on gray document images, apart from the standard dataset described earlier, the method was applied to different types of gray images which include images with non-overlapping text, overlapping text and graphics, the text of different sizes and orientations.

We applied the DST on gray level image. The image was decomposed into 4 level with one low and three high-frequency shearlets by DST. The three high frequency shearlets were decomposed into $8(= 2^{\lceil 1/2 \rceil + 2})$,$8(= 2^{\lceil 2/2 \rceil + 2})$ and $16(= 2^{\lceil 3/2 \rceil + 2})$ shearlet sub-bands by the DST. So total $33(= 1 + 8 + 8 + 16)$ sub-bands of the same size of the input image were generated. The high-frequency shearlets contained the local information of text and non-text regions in 8,8 and 16 orientations i.e at each 45°, 45° and 22.5° anti-clockwise directions for the levels 1,2 and 3. This was followed by local energy estimation of DST coefficients. To reduce the uncertainty present in the sub-bands and due to redundant information generation, they were mapped into NS domain for segmentation as described in Section 3.

We compared the performance of the proposed method with four published methods. They are Acharyya [25], Kumar [26], Maji [28] and Gomez [10]. The qualitative results of the text region segmentation by five different methods (Acharyya [25], Kumar [26], Maji [28],

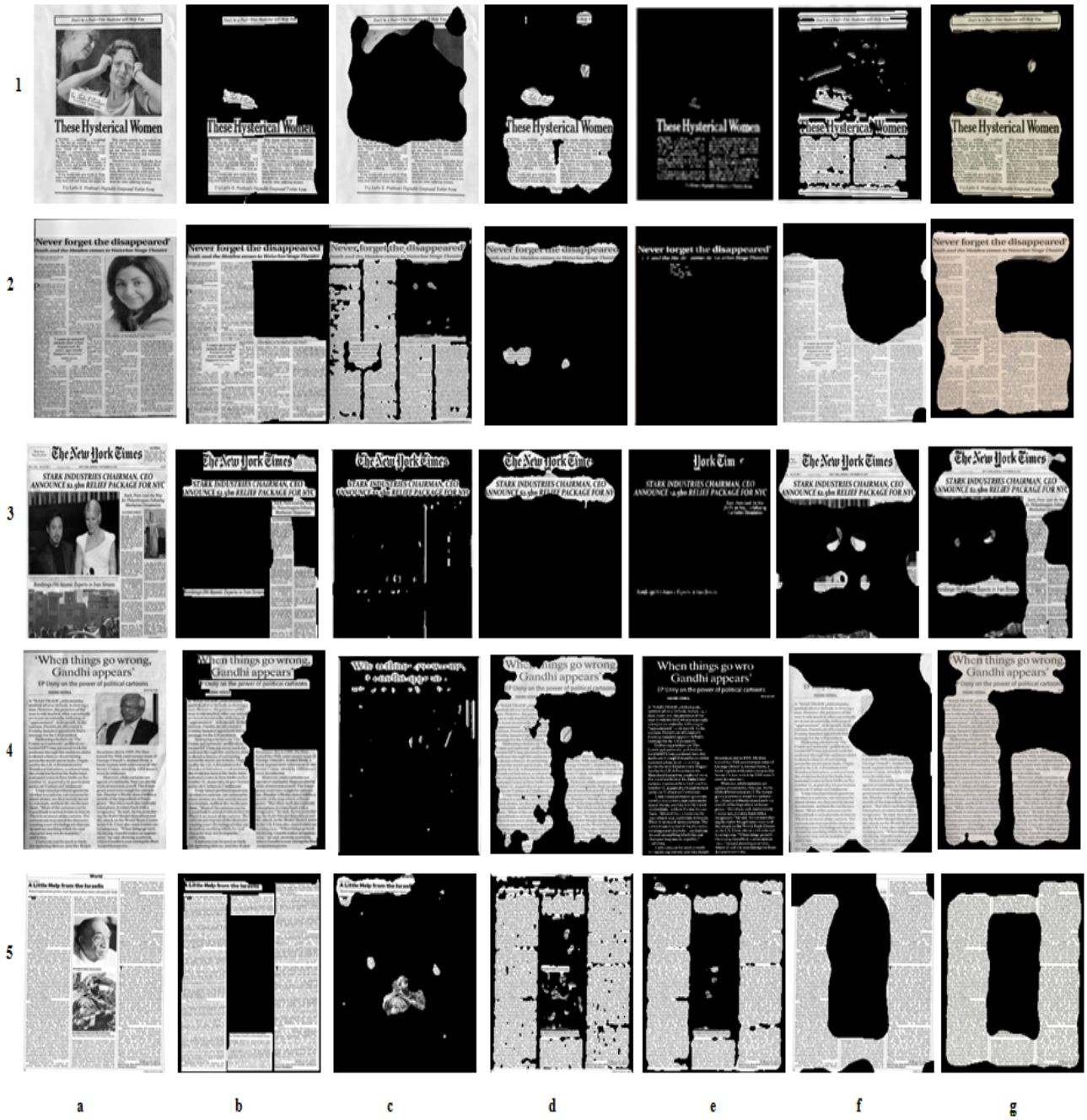


Figure 4: Column wise (a) Original test images (b) Ground truths for segmentation result (c) Results using Acharyya [25] (d) Results using Kumar [26] (e) Results using Gomez [10] (f) Results using Maji [28] (g) Results using proposed method

Gomez [10]and the proposed) are shown in the Figure 4. We applied the proposed method, and all the other methods considered here on the gray version of the ICDAR2015 dataset of Born Digital images. The average quantitative performances of the five methods are shown in the table 2.

It is observed from both the qualitative and quantitative measures that on an average the performance and accuracy of the proposed method are much better than that of Acharyya [25],Kumar [26],Gomez [10] and Maji [28]. The DST used in the proposed method leads to a unified treatment of the continuum as well as the digital realm, while still providing optimally sparse approximations of anisotropic features. Thus, the generated features from the multi-resolution and multi-directional DST shearlet sub-bands and neutrosophic approach were more effective than the conventional MSERs [10] based method. Moreover, the features were quite useful in finding the accurate segmentation boundary than the methods in [25] and [26] which used M-band wavelet and matched wavelet respectively. The M-band wavelet and matched wavelet have less capacity to capture the edges and curvatures of text regions than the DST sub-bands. Again, the method in [28] used the multi-resolution analysis with rough fuzzy clustering for handling the uncertainty during segmentation. However, we reduced the uncertainties in DST feature in two stages. In the first stage, the uncertainties in the shearlets were reduced using iterative $\alpha - mean$ and $\beta - enhancement$ operation during feature generation. That means, the contrasts of the coefficients in the shearlets were increased and that made them suitable for segmentation. This was followed by feature set reduction to handle the uncertainties due to redundancy. In the second stage, the uncertainties during clustering were handled by $\gamma - k - means$ in the NS domain. These two steps reduction of the uncertainties were more effective than uncertainty reduction in [28].

The unsupervised feature selection algorithm in the proposed method takes into account the uncertainties due to feature redundancy and yields a more compact subset of features. We tested the effect of unsupervised feature selection in NS domain and the effect of parameter k (described in the section 2.8) on the performance. The analytical result is shown in Appendix A.1.

4.4. Experiments on noisy, rotated and dynamic gray level changed document images:

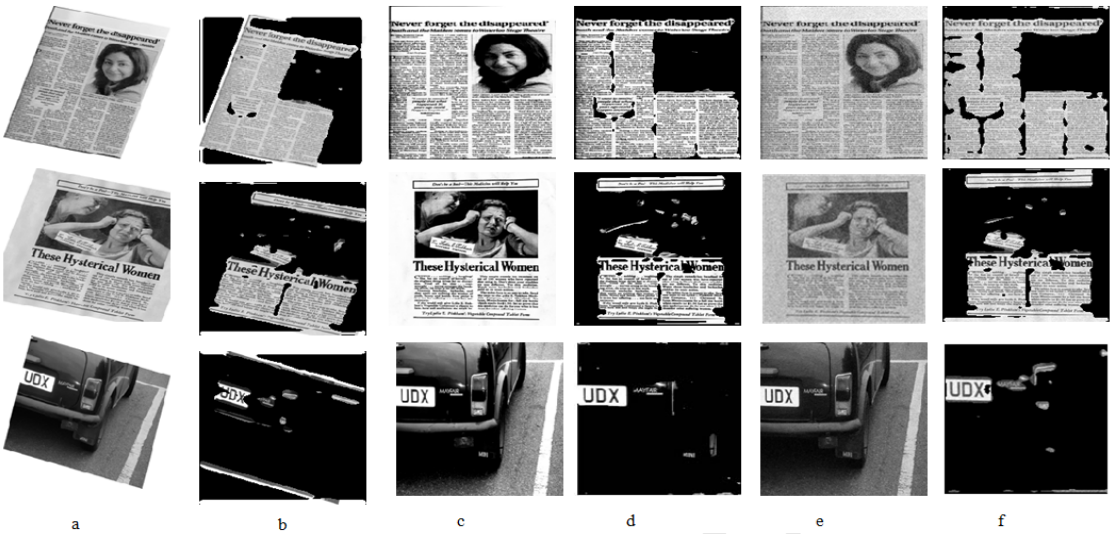


Figure 5: Column wise(a) Rotation of test images by 15.5 degrees clockwise (b) Segmentation results by the proposed method (c) Change by +55 for gray level dynamic range of the test images (d) Segmentation results by the proposed method (e) Test images with Gaussian noise of mean 0 and std 0.16 (f) Segmentation results by the proposed method

To demonstrate the robustness of the proposed method, it was applied to a new set of images with rotation, gray level dynamic range shift, and noise corruption. Some of the test images and the corresponding text region segmentation results are shown in the Figure 5. The test images were generated by moderate rotations, adding Gaussian noise and dynamic gray level intensity change. We compared the quantitative performance of the proposed method and other text region segmentation methods considered here over noisy, rotated and dynamic gray level changed images. ICDAR2015 dataset was used for this purpose.

The uncertainties in the DST shearlet features increase when noise is added to an image. So the noise addition affects the text-region segmentation result. The average F-measures' against average signal to noise ratio (SNR) for all the methods are reported in the graph at Figure 6. The figure illustrates that our method is robust against a moderate amount of noise corruption than that of the other methods compared here. The proposed method is robust against a moderate noise corruption due to the noise handling capacity of shearlets and

due to the adaptive $\alpha - means$ and $\beta - enhancement$ operations in the proposed method. These two operations make the noisy pixels in the neutrosophic image homogeneous with the neighboring pixels and the noise is reduced. Therefore, the uncertainties in the features get minimized.

In the proposed method the DST produces the rotation invariant features due to its excellent directional property. Additionally, the neutrosophic set theoretic approach reduces the spatial ambiguity present in the features due to the rotation. As a result, the method becomes robust against rotation. The average F-measures against the angles of rotation is shown in the graph at Figure 7. From the graph, it can be said that the proposed method can also detect non-horizontal text regions more efficiently than the methods compared here.

The dynamic gray level modification changes the image pixel intensity, and as a result, the gray level ambiguity is increased. The ambiguity is successfully reduced in the NS domain. The reason is that there will be negligible change in $T(i, j)$ membership values due to the changes in r values in the Eq. 7. The average F-measures against the positive and negative gray level shifts are shown in the graphs at Figure 8 and Figure 9 respectively. It is observed from the graphs that the proposed method is fairly tolerant of moderate changes in gray level dynamic range.

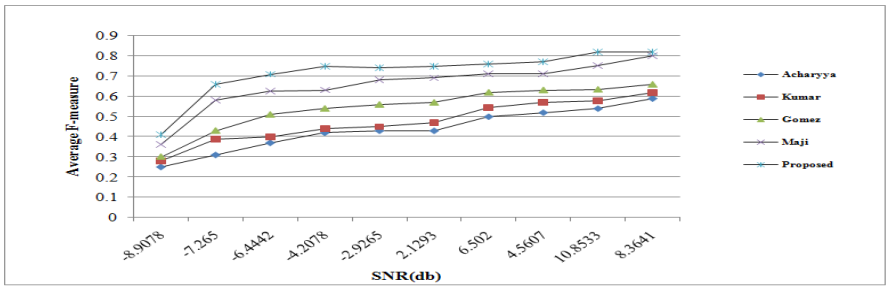


Figure 6: Average F-measure at different SNR on ICDAR2015 dataset(Born digital)

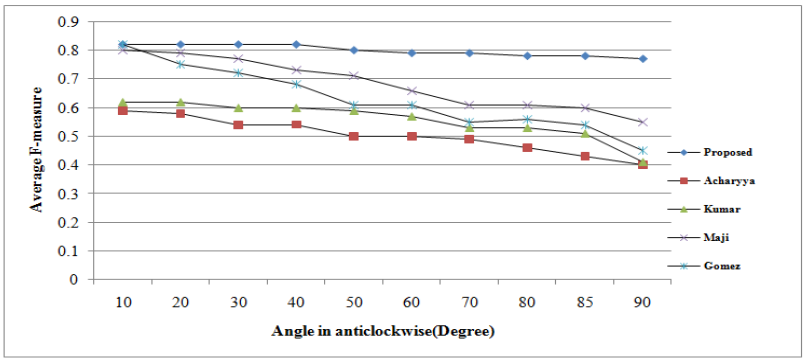


Figure 7: Average F-measure at different orientations on ICDAR2015 dataset(Born digital)

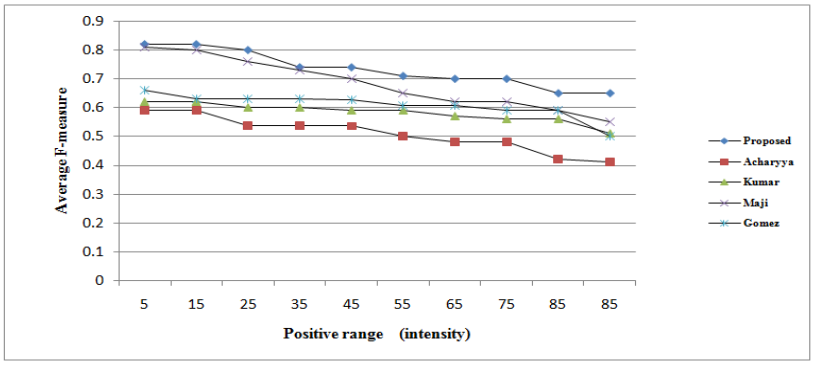


Figure 8: Average F-measure at different positive dynamic range change on ICDAR2015 dataset(Born digital)

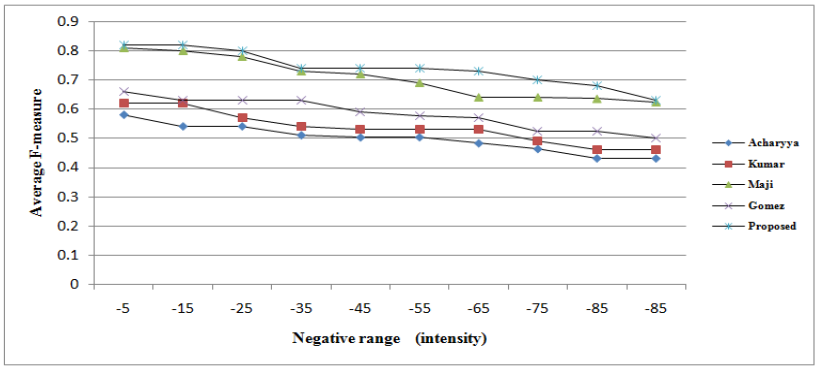


Figure 9: Average F-measure at different negative dynamic range change on ICDAR2015 dataset(Born digital)

4.5. Experiments on colored document images :

To judge the performance of the proposed method on colored text documents, we applied it to the colored image of the scanned newspaper, magazine etc. For this, at first, the colored document image was transformed into the YCbCr plane. The YCbCr basically decouples the intensity and color information, and this representation is very close to the human perceptual model. The human visual system is less sensitive to chrominance than luminance [45]. Accordingly, the weights of features generated from the above three planes were based on the convention perceptual importance, as used in the JPEG 2000 that is Y: Cb: Cr=4:2:1. Each of the Y, Cb, Cr color plane was transformed into a set of frequency shearlets using DST. The total number of shearlet sub-band generated was $99(= 3 \times 33)$. The sub-bands were then mapped to NS domain for uncertainty reduction. This was followed by feature vector generation and segmentation as described in section 3.

We compared the performance of the proposed method with three published methods in [29], [36] and [10] on color documents. The method in [36] used dyadic wavelet transform and NS set theocratic approach for color texture segmentation. As the basic assumption of the proposed method is that the text and non-text regions are two different textures, the proposed method is also capable of segmenting any color texture image which contains two different textures. Here it was assumed that the method in [36] was also capable of differentiating the text and non-text regions.

The quantitative performance of all the three methods and the proposed method were computed using three different metrics described in Eq 33, Eq 34 and Eq 35 respectively. The color text region segmentation results due to four different methods ([29], [36], [10] and the proposed) are shown in the Figure 10.For quantitative performance, we applied all the four methods and the method by Cho [16] on the color images of ICDAR2015. The results are shown in table 3. We also compared the quantitative performance of the proposed method with the methods in [10] and [15] on KAIST dataset. The performances are shown in table 4. It is observed both from the qualitative and quantitative results that the accuracy of the proposed method on average is much better than that of the other methods.

In the proposed method for a color document image, both the color and texture infor-

mation were combined. The DST shearlet features are able to extract the texture features more accurately than dyadic wavelet transform. Hence, the proposed method provided the better result than the method in [36]. Again, from the performances, it can be said that the feature set and the uncertainty handling technique of the proposed method during feature generation is better than that of [29] during segmentation. The method in [29] used fuzzy c-means for segmentation. The proposed method also performed better than that of the seed based method in [15]. The reason may be that the method in [15] depended on the Canny edge detector for generating the seed. The detector may not work well in the complex background and it can not handle the uncertainties. Again the method in [16] combined the MSERs and Canny edge detector to detect the text regions, which are less efficient than the proposed anisotropic features of the text captured by the DST.

Table 3: Average performance comparison of text region segmentation results obtained by different methods using color test images from ICDAR2015(Born digital)

Methods	*R*	*P*	*f*
Kundu [29]	0.67	0.75	0.71
Sengur [36]	0.74	0.79	0.77
Gomez [10]	0.73	0.76	0.74
Cho [16]	0.77	0.83	0.79
Proposed	0.83	0.84	0.83

Table 4: Average performance comparison of text region segmentation results obtained by different methods using color test images from KAIST

Methods	*R*	*P*	*f*
Gomez [10]	0.78	0.66	0.71
Bai [15]	0.89	0.83	0.86
Proposed	0.89	0.86	0.87

Figure 10: Column wise(a) Original color test image (b) Ground truth for test color image segmentation (c) Results using Kundu. [29](d) Results using Sengur [36] (e) Results using Gomez [10](f) Result using proposed method

4.6. Comparison with deep learning based methods

For the learning based method, we used the CNN with the sheralet features in NS domain for text region detection. We trained the network by using the reduced features in the NS domain. For classification using CNN, we followed the same protocol as used in [46] for text region segmentation using multi-resolution analysis. The CNN text detector had 5 layers. They were 2 convolution layers, 2 average pooling layers, and a fully connected layer. The output layer was consisted of 2 nodes which were text and non-text. To train the CNN we used the 229 training image from ICDAR2011 and 258 images from ICDAR2005 dataset. We synthetically created the text and non-text patches from them. We compared our method with state-of-the-art deep learning based text detection methods. We tested on the same dataset of ICDAR2011 and ICDAR2013 for comparison which contains 255 and 233 test images respectively. In Huang's [18] and He's [17] deep learning method, the MSERs features were used as input to the CNN. The quantitative results are shown in Table 5. From the result, it can be said that the proposed method performed better than that of He and Huang. The reason is that the MSERs features used in the deep learning, are less robust in the rotation than DST shearlet features used in the proposed method. Hence the MSER features become more uncertain than that of the DST features. Moreover, the uncertainties in the DST features were reduced in NS domain before entering into CNN, where no such provisions were available in their methods.

Table 5: Experimental results of Deep learning based methods using ICDAR dataset

Methods	*Dataset*	R	P	f
He [10]	*ICDAR*2011	0.74	0.91	0.82
Huang [15]		0.71	0.88	0.78
Proposed		0.74	0.94	0.84
He [10]	*ICDAR*2013	0.73	0.93	0.82
Proposed		0.75	0.94	0.83

4.7. Additional qualitative results

Some additional results by the proposed method from ICDAR2015(Born digital) dataset and KAIST dataset are shown in Figure11. Further, the method was applied to some more document image in different orientations. In these images, ambiguities occurred due to different orientations. Some test image data from Computer vision Laboratory [47] and corresponding segmentation results are shown in Figure 12. The visual inspection again indicates that the method is fairly tolerant of different orientations .

Figure 11: Row wise (a) Row1 shows images from ICDAR2015 and KAIST. Row2 shows text region segmentation results of corresponding images in the same sequence, using the proposed method.

4.8. Role of α and β parameters :

In this sub-section, we studied the role of the two parameters α and β for text region segmentation. As already stated, the operations are used to reduce indeterminacy in the NS domain. If these two parameters are not appropriate, the proposed method may not generate features properly for text and non-text regions. This may lead to poor segmentation result. In Figure 13, three results are shown for different values of α and β. When the values of α and β were both 0.83 the result is shown in the Figure 13(b). The result in Figure 13(c) was obtained when values of α, β were 0.75 and 0.25 respectively. But when the values were computed adaptively based on the statistics of the I subset, we got the most consistent result as shown in Figure 13(d). The reason is that α is high(low) and β is low(high) when

Figure 12: Column wise(a) Color text image set from Computer vision Laboratory dataset (b) Text region segmentation results of corresponding images in (a) in the same sequence, using the proposed method. (c) Another set of Color text image set corresponding to the same row in (a) with the different orientation (d) Text region segmentation results of corresponding images in (c) in the same sequence, using the proposed method.

601 entropy of I is high(low). High entropy means the image contains more indeterminacy i.e
602 it is difficult to decide whether a pixel is text or non-text pixels. So shearlet sub- band
603 with high indeterminacy i.e $I \geq \alpha$ should be made homogenous by α-means with contrast
604 enhancement by β-enhancement to generate the proper feature vector. In the first two cases,
605 the values of α and β were chosen arbitrarily without considering the document image in
606 hand and hence could not generate the proper feature vector.

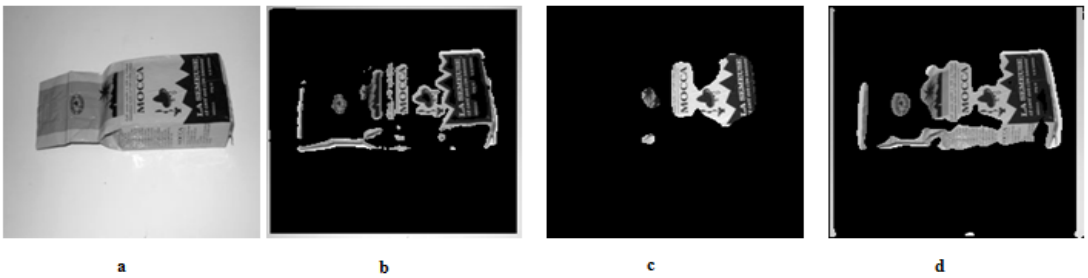


Figure 13: Column wise(a) Original document test image (b)Results using $\alpha = 0.83$ and $\beta = 0.83$ (c) Results using $\alpha = 0.75$ and $\beta = 0.25$(d) Results using adaptive α and β

607

5. Conclusion.

In this paper, we have proposed an DST and NS based text region segmentation methodology in the complex document images. The method judiciously combines the advantages of multi-scale and multi-directional shearlet feature extraction tool like DST together with the uncertainty handling capability of the neutrosophic set. This is done in order to achieve the segmentation result with higher accuracy in comparison to many existing techniques reported in the literature. It is also to be noted that the proposed method performs equally well in another type of segmentation like two class color and gray level texture segmentation. This method also shows robustness in performance under different perturbations like moderate rotation, dynamic range changes, and noise corruption. With suitable modification in the current method, it can be extended for the multiclass problem, which is being currently investigated.

6. Acknowledgement

We would like to thank Dr. Weilin Huang for providing us the code of their paper [17].

7. Reference

[1] C. A. Murthy, S. K. Pal, Histogram thresholding by minimizing gray level fuzziness, Information Sciences 60 (1992.) 107–135.

[2] A. Rosenfield, Fuzzy geometry: An updated overview, Information Sciences 110 (1998.) 127–133.

[3] S. K. Pal, A. Ghosh, M. K. Kundu, Soft Computing for Image Processing, Springer-Verlag Berlin Heidelberg GmbH, 1st edition, 2000.

[4] G. Nagg, S. Seth, M. Viswanathan, A protoptype document image analysis system for technical journals, Computer. 25 (1992.) 10–22.

[5] D. Chen, J.-M. Odobez, H. Bourlard, Text detection and recognition in images and video frames, Pattern Recognition 37 (2004) 595–608.

[6] S. M. Lucas., ICDAR2005 text locating competition results ., Proceedings of International Conference on Document Analysis and Recognition. 1 (2005.) 80–84.

[7] E. Haneda, C. Bouman, Text segmentation for mrc document compression, IEEE Transactions on Image Processing 20 (2011) 1611–1626.

[8] C. Yi, Y. Tian, Text string detection from natural scenes by structure-based partition and grouping., IEEE Transactions on Image processing. 20 (2011.) 2594–2605.

[9] J. Matas, O. Chum, M. Urban, T. Pajdla, Robust wide baseline stereo from maximally stable external regions., Proceedings of British Machine Vision Conference (2002) 384–396.

[10] L. Gomez, D. Karatzas, Multi-script text extraction from natural scenes., Proceedings of International Conference on Document Analysis and Recognition. (2013.) 467–471.

[11] X. Yin, X. Yin, H. H. K. Hung, Robust text detection in natural scene images, IEEE Transactions on Pattern Aanlysis and Machine Intelligence. 36 (2014.) 970–983.

[12] A. Gonzalez, L. Bergasa, J. Yebes, S. Bronte, Text location in complex images., Proceedings of International Conference on Pattern Recognition. (2012.) 617–620.

[13] C. Shi, C. Wang, B. Xiao, Y. Zhang, Scene text detection using graph model built upon maximally stable extremal regions., Pattren Recognition letter . 34 (2013.) 107–116.

[14] Y. Li, H. Lu, Scene text detection via stroke width, Proceedings of International Conference on Pattern Recognition (2012) 681–684.

[15] B. Bai, F. Yin, C. L. Liu, A seed-based segmentation method for scene text extraction., IAPR International Workshop on Document Analysis Systems (2014) 262–266.

[16] H. Cho, M. Sung, B. Jun, Canny text detector: Fast and robust scene text localization algorithm., International Conference on Computer Vision and Pattern Recognition (CVPR) (2016) 3566–3573.

[17] T. He, W. Huang, Y. Qiao, J. Yao, Text-attentional convolutional neural network for scene text detection., IEEE Transactions on Image Processing 25 (2016) 2529–2541.

[18] W. Huang, Y. Qiao, X. Tang, Robust scene text detection with convolution neural network induced MSER trees, Proceedings of European Conference on Computer Vision LNCS 8692 (2014) 497–511.

[19] Y. Zhu, C. Yao, X. Bai, Scene text detection and recognition:recent advances and future trends, Frontiers of Computer Science 10 (2016) 19–36.

[20] G. Zhou, Y. Liu, Q. Meng, Detecting multilingual text in natural scene., Proceedings of International Symposium on Access spaces . (2011.) 116–120.

[21] M. Zhao, S. Li, J. Kwork, Text detection in images using sparse representation with discriminative dictionaries., Image and Vision Computing . 28 (2010.) 1590–1599.

[22] C. Liang, P. Y. Chen, Dwt based text localization., International Journal of Applied Science Engineering. 2 (2004.) 105–116.

[23] W. Chan, G. Coghill, Text analysis using local energy., Pattern Recognition . 34 (2001.) 2523–2532.

[24] S. Nirmal, P. Nagabhushan, Foreground text segmentation in complex color document images., Signal Image and Video Processing . 6 (2012.) 669–678.

[25] M. Acharyya, M. Kundu, Document image segmentation using wavelet scale-space features., IEEE

Transactions on circuits and systems on video technology. 12 (2002.) 1117–1127.

[26] S. Kumar, R. Gupta, N. Khanna, S. Chaudhury, S. Joshi, Text extraction and document image segmentation using matched wavelets and mrf model., IEEE Transactions on Image processing . 16 (2007.) 2117–2128.

[27] S. Roy, M. Kundu, G. .Granlund, Uncertainty relations and time-frequency distributions for unsharp observables, Information Sciences 89 (1996.) 193–209.

[28] P. Maji, S. Roy, Rough -fuzzy clustering and multiresolution image analysis for text-graphics segmentation., Applied soft computing. 30 (2015.) 705–721.

[29] M. Kundu, S. Dhar, 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. (2012.) 86–88.

[30] G. Kutyniok, W. Q. Lim, R. Reisenhofer, Shearlab 3d: Faithful digital shearlet transforms based on compactly supported shearlets., Numerical Analysis (math.NA) arXiv:1402.5670 (2014).

[31] G. Kutyniok, W. Q. Lim, G. Steidl, Shearlets: Theory and applications., GAMM-Mitt 37 (2014) 259–280.

[32] F. Smarandache, A Unifying Field in Logic,Neutrosophy, Neutrosophic Set,Neutrosophic Probability, American Research Press, 3rd edition, 2003.

[33] Y. Guo, H. Cheng, A new neutrosophic approch to image segmentation, Pattren Recognition . 42 (2009.) 587–595.

[34] H. Cheng, Y. Guo, A new neutrosophic approch to image thresholding, New Mathematics and Natural Computation. 4 (2008.) 291–308.

[35] Y. Guo, H. Cheng, A new neutrosophic approch to image denoising, New Mathematics and Natural Computation. 5 (2009.) 653–662.

[36] A. Sengur, Y. Guo, Color texure segmentation based on neutrosophic set and wavelet transform, Computer Vision and Image Understanding. 115 (2011.) 1134–1144.

[37] G. Kutyniok, W. Q. Lim, X. Zhuang, Digital shearlet transforms. In Shearlets: Multiscale analysis for multivariate data,, Spinger, 2012.

[38] P. Mitra, C. Murty, S. Pal, Unsupervised feature selection using feature similarity., IEEE Transactions on Pattern Analysis and Machine Intelligence . 24 (2002.) 301–312.

[39] M. Kundu, M. Acharyya, M-band wavelets:application to texture segmentation for real life image analysis., Internation Journal of Wavelets, Multiresolution and Information Processing. 1 (2003.) 115–119.

[40] M. Everingham, L. V. G. and C. K. I. Williams J. Winn, A. Zisserman, The pascal visual object classes (voc) challenge, International Journal of Computer Vision 88 (2010.) 303–338.

[41] ICDAR2015 dataset, http://rrc.cvc.uab.es/ (2015.).

[42] ICDAR2003, http://algoval.essex.ac.uk/icdar/Datasets.html (2003.).

[43] Kaist scene text database, www.iapr-tc11.org/mediawiki/index.php/KAIST_Scene_Text_Database (2011.).

[44] D. Karatzas, S. R. Mestre, J. Mas, F. Nourbakhsh, P. P. Roy, ICDAR 2011 Robust Reading Competition., International Conference on Document Analysis and Recognition. (2011) 1485–1490.

[45] N. Plataniotis, A. Venetsanopoulos, Color image processing and applications., Spinger Verlag, Heidelberg . (2000).

[46] T. Kobchaisawat, 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).

[47] Computer Vision Laboratory dataset, http://www.vision.ee.ethz.ch/datasets/ (2003.).

Appendix A. NS image representation of gray level values of an image

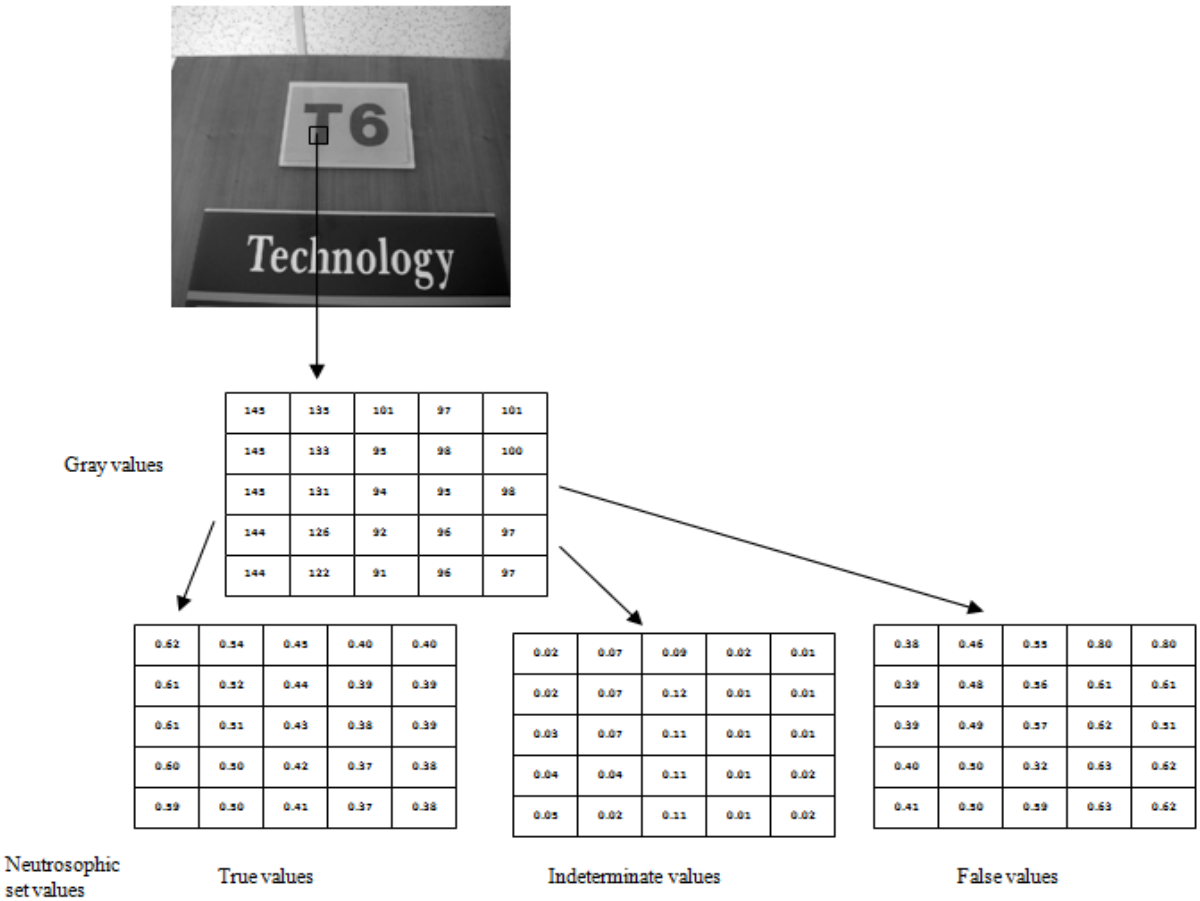


Figure A.14: Neutrosophic set representation of a small portion (marked square) of gray scale image

The image in the Figure A.14 shows the gray values of the original image within the square box and the true,the false and the indeterminate membership values of the corresponding neutrosophic image obtained by the Eq 7-Eq 12. From the third column of the indeterminate subset matrix, it can be said that pixels in that column are the edge pixels, and which is correct.

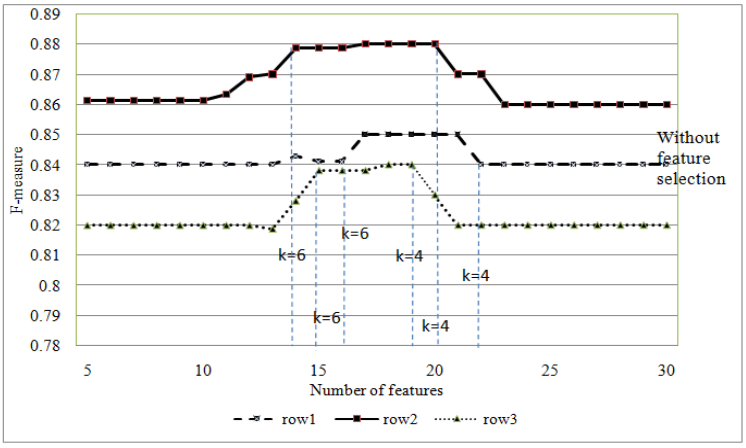


Figure A.15: Variation in f-measure with the size of feature set on three test images. The vertical dotted lines represent the number of features and corresponding f-measure at k=4 or k=6

Appendix A.1. Importance of unsupervised feature selection :

To examine the importance of unsupervised feature selection, we applied the proposed method with and without the feature selection, on the gray level images. The quantitative performance on three images in row 1,row2 and row3 in Figure 4 are illustrated in the graph shown in Figure A.15. In the figure, the f-measures are shown against the number of features. It also illustrates the effect of parameter k in the feature selection algorithm. The figure shows that f-measure becomes low without the feature selection algorithm or when the number of features is very low. It is observed that the number of features generated is lower when the value of k in high. In the figure, the vertical dashed lines point the number of features created when k=6 or k=4 for each sample image.