International Journal on Document Analysis and Recognition (IJDAR) (2018) 21:1–20
https://doi.org/10.1007/s10032-018-0296-z

ORIGINAL PAPER

Text and non-text separation in offline document images: a survey

Showmik Bhowmik[1] · Ram Sarkar[1] · Mita Nasipuri[1] · David Doermann[2]

Received: 7 February 2017 / Revised: 9 November 2017 / Accepted: 27 February 2018 / Published online: 8 March 2018

Abstract

Separation of text and non-text is an essential processing step for any document analysis system. Therefore, it is important to have a clear understanding of the state-of-the-art of text/non-text separation in order to facilitate the development of efficient document processing systems. This paper first summarizes the technical challenges of performing text/non-text separation. It then categorizes offline document images into different classes according to the nature of the challenges one faces, in an attempt to provide insight into various techniques presented in the literature. The pros and cons of various techniques are explained wherever possible. Along with the evaluation protocols, benchmark databases, this paper also presents a performance comparison of different methods. Finally, this article highlights the future research challenges and directions in this domain.

Keywords Text/non-text separation · Segmentation · Offline document images · Engineering drawing · Map · Unconstrained handwritten document · Newspaper · Journal · Magazine · Check · Form · Survey

1 Introduction

Recent advancements in computer vision, document processing and the ability to capture document's content using mobile devices have amplified research in the automatic acquisition of textual and graphical content from offline documents, scene images and video [1–3]. These sources contain what can broadly be categorized as text and non-text (graphics, halftone, drawings, etc.) which require different processing to convert them into a machine readable form. Therefore, the task of identifying these two classes in the document images is required. In the literature, we find that this task is referred to in many different ways including: *Text and Non-text Separation* [4], *Text localization* [5], *Text detection and extraction* [6] and *Suppression of non-text components* [7] among others. In this paper, we survey the state-of-the-art in *Text/Non-text separation* and use this phrase to refer to the general task. The problem of text and non-text separation may formally be defined as:

If I is a non-empty set of elements which represents an image and ε is a non-empty set of output classes such that $\varepsilon = \{C_T, C_{NT}\}$, where CT represents the text class and CNT represents the non-text class, then the problem of text and non-text separation can be defined as finding a function "F" such that $F : I \rightarrow \varepsilon$.

In a wide range of applications, text/non-text separation plays a vital role in early stages of a document analysis system. Since early 1980s a significant number of published articles have appeared dealing with the challenges of developing effective methods which are robust to varying document styles ([4,8–13]). A survey on page segmentation and zone classification was published in 1999 [14], but no articles have been found in the literature which cover newer methods. Surveys have, however, been published on the specific challenges of applications including map processing [15], engineering drawing understanding [16] and text detection and recognition in imagery [17]. In these surveys, text and non-text separation was not the main focus. Despite its importance, there is no high-quality survey where text and non-text separation is the prime focus.

The paper is organized as follows. Section 2 presents taxonomy of text/non-text image classes and the challenges and applications associated with each. Section 3 describes state-of-the-art of methods for general offline document images. Section 4 gives a brief description of standard databases along with the evaluation protocols used for performance assessment of the methods reported in the literature. This section also presents a detail performance comparison among the methods reported for different classes of documents. Sec-

✉ Ram Sarkar
raamsarkar@gmail.com

1 Jadavpur University, Kolkata, India

2 Institute for Advanced Computer Studies, University of Maryland, College Park, USA

tion 5 provides a general summary of the task along with directions for future research.

2 A taxonomy of images with related challenges and applications

It was mentioned earlier that text and non-text separation is an essential part of a document image analysis system and depending on the application and image content, images may need to be processed vary differently. The most common classes of images on which the text/non-text separation task is performed include:

– Online document images [18]
– Scene images and Video frames [17]
– Web images [19]
– General offline document images [14]

Among them, first three are only peripheral to our main interests, so we will provide pointers to the literature but not address them in detail. Our discussion will mainly be focused on general offline document images.

2.1 Online document images

The use of smart phones, tablet PCs and electronic white boards with pen-based writing interfaces has become widespread. Users use these systems to take notes, draw sketches and produce diagrams. For the recognition of text present in the documents created, the separation of text from the non-text is also very important and has attracted the attention of a large group of researchers around the world. A number of informative articles have recently appeared [20–24]. One of the unique challenges they face is the absence of prior knowledge regarding the content layout and inherent ambiguity present in handwritten content.

2.2 Scene images and video frames

Recognition of text from natural scene images and video frames is one of the most popular areas of research in recent years [1,2,17,25,26]. It has a wide range of applications including multimedia retrieval [27,28], visual input and access [29,30] and industrial automation [31,32]. The primary task in this context is to detect the presence of text first, and if text is found then to localize it [33,34] and recognize it [35,36]. The task has several challenges [17] imposed by the environmental complexity (scene complexity, uneven lighting), the image acquisition process (blurring/defocusing, perspective distortion, etc.) and the complex nature of text (variations in aspect ratio and fonts, multi-oriented/ curved text, multilingual environments, etc.).

2.3 Web images

The extraction and recognition of text from web images has also drawn the attention of researchers due to its need for effective indexing, semantic analysis and ranking of the web pages [19]. Extraction as well as recognition of text is very challenging for several reasons such as low resolution of the images, presence of large number of graphic objects, anti-aliasing properties of these images [37,38] and so on. Finding a uniform representation of the text in the web images is another important goal of this domain, as this uniform representation of the text can be used for numerous applications including visualization, translation, summarization, voice browsing [39], automatic content analysis [40] and web document analysis [41].

2.4 General offline document images

In this domain, our focus is on applications such as document image analysis and recognition [14], map processing system [15], engineering drawing understanding system (EDUS) [4,16,43], information spotting including word spotting, symbol spotting and logo spotting [44]. In the literature, general document images are categorized in different ways by the different researchers [4,14], primarily based on the layout or structure of the documents. In this paper, we define four classes, based on the challenges that one may encounter, and they are:

1. **Class-1:** Newspapers, journals, magazines, books, etc.
2. **Class-2:** Engineering drawings, bank checks, forms, etc.
3. **Class-3:** Advertisements, book/magazine covers, posters, postal envelopes, etc.
4. **Class-4:** Unconstrained handwritten documents images.

It is very hard, however, to establish a clear categorization on the basis of the complexity and these classes do overlap. For example, a journal article may contain an engineering drawing or a magazine may contain a map. Similarly, a map may be hand drawn or an advertisement may contain structured diagrams.

Examples of documents which belong to Class-1 are shown in Fig. 1 and are typically characterized by a clear separation between regions based on their structured layouts. The fact that content is spatially disjoint makes them simpler to process from text and non-text separation point of view. Furthermore in these documents, the adjacent components in the same region[1] that form words, text lines or even

[1] Please note that in the literature, the terms region and zone are often used interchangeably, with the term zone being preferred when a region

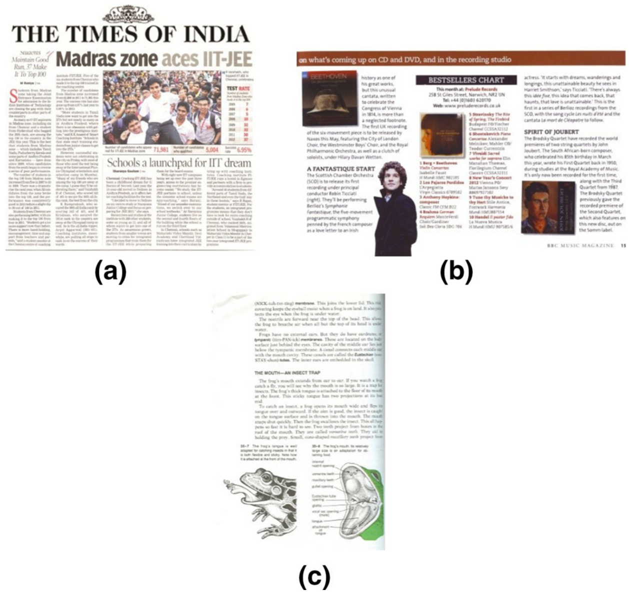


Fig. 1 Examples of documents belong to Class-1: **a** newspaper, **b** magazine, **b** book

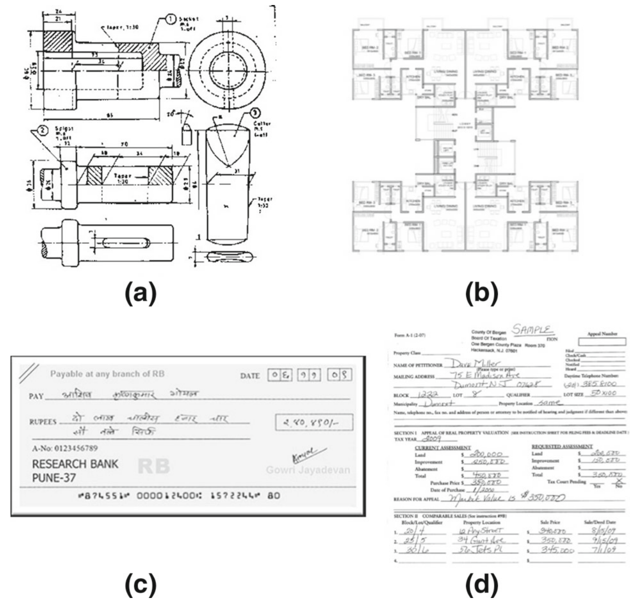


Fig. 2 Examples of documents belong to Class-2: **a** Engineering drawing **b** Architectural drawing **c** Filled-in Indian check [45] **d** Filled-in form

paragraphs typically have the same font and/or style. Like other classes however, the separation task may become more difficult with increased noise or background content.

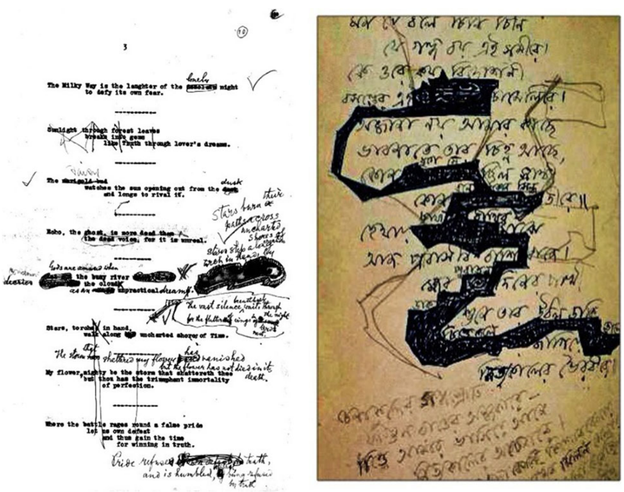
Examples of engineering drawings, bank check and form belonging to Class-2 type documents are shown in Fig. 2. This category of document is more complex than the documents that fall under Class-1 due to the presence of comparably higher levels of non-uniformity in size, shape and orientation of text regions, and the possible existence of touching components. Furthermore, important engineering drawings are often stored for archival purposes in printed form and used in ways that can introduce a significant amount of noise similar to the content. The overall quality of these documents may be affected by aging [16], for example. But the documents of this category generally contain specific types of non-text elements like, checks and forms mostly contain straight lines, dash lines, tables and rectangular (and/or circular) boxes as non-text. Most of the text components (except signatures and other few components) present in these documents generally follow horizontal orientation. Although, the components present in various engineering drawings have multiple orientations, but these, however, typically follow a specific drawing standard such as ANSI and ISO and tend to have some consistency. These drawing standards specify conventions regarding the style, type, size and other characteristics of the content. These tighter geometrical constraints make the text/non-text separation easier than with documents in Class-3 and Class-4.

Footnote 1 continued
is rectangular. For simplicity however, we will use the term region to mean either zone or polygonal region in this document.

Fig. 3 Examples of Class-3 documents: **a** Book cover **b** Magazine cover **c** Poster image **d** Postal Envelop and **e** Map

Documents categorized as Class-3 (Fig. 3) typically have a complex layout, overlapping text and non-text elements, a large variability in the size of the text and possibly a complex colored background with different illumination and texture. These factors make them more challenging to process than the previous two classes [14,46].

Documents belonging to Class-4 (Fig. 4) are the most complex ones in terms of text and non-text separation because here we consider purely handwritten documents. In general, handwritten documents have large variations as well as ambi-



(a) (b)

Fig. 4 Examples of Class-4 documents: handwritten manuscripts [47]

guity in the size, shape and orientation of their components [7]. Even the adjacent components that form a word often have different size and/or orientation (i.e., skew and slant). Additionally in such documents, a single component often does not represent a complete character, rather it can be a part of a character. This non-uniformity is a function of the writing style and varies with the language, script and writer. Component overlap is also a frequent issue in this category of documents. Handwritten documents have additional challenges like arbitrary layout, non-uniformly skewed text lines, variation on the overall quality, and they can be written on a variety of paper and can be old.

3 Features and techniques for separation

We begin our survey with the features and techniques cited the most in the literature. Since features tend to be used widely across the field, we have not categorized those within our taxonomy. We simply list several classes in Sect. 3.1.

Techniques, however, tend to be more restrictive so we organize those by class although again, some can be applied to multiple classes. Classification of text and non-text is generally done at any of three levels: the pixel level, the component level or the region level. Although initial processing may depend on document complexity, the final representation is at one of these levels. Many recent techniques for text/non-text separation for general offline document images are available in the literature. In this section, those techniques are grouped according to the class of documents they will most likely be applied to and are briefly discussed in the following subsections. It should be noted that depending on document complexity, some algorithms can be applied to multiple classes, but that decision should ultimately be made by those, implementing a solution.

3.1 Features commonly cited

For the classification of text and non-text components/regions present in an offline document image, three types of features are typically used by researchers: Texture-based features [48–50], shape-based features [4,13,51] and color information-based features [52–54].

3.1.1 Texture-based features

Generally, texture features are very useful when dealing with dense documents. Most of the region classification-based methods apply texture features for text/non-text discrimination. Examples are first-order statistical features (e.g., median and modal pixel value) [5], quantitative characteristics of texture (e.g., mean brightness, mean-square deviation, mean difference of mean brightness of the blocks) [56], gray level co-occurrence matrix (GLCM)-based texture features [55,57], run length-based features [7,58,59], white tiles-based texture feature [60], local binary pattern (LBP)-based features [61], transitional features [62] and spatial features [63].

3.1.2 Shape-based features

For documents having text scattered through the image like Class-2 or some specific cases of Class-4, the texture features are not very useful. Here, researchers prefer various shape information of the connected components (CCs) for their classification as text/non-text. Examples of such features are height, width, aspect ratio and thickness of CCs [64–66].

3.1.3 Color-based features

Some methods also use the color information-based features [52–54]. But the use of such methods is limited to Class-3 documents and some color documents belonging to Class-4 category. Color features are very sensitive to multi-colored text, illumination, etc. Generally, pixel classification-based methods use these types of features.

3.2 Methods for Class-1 documents

In this subsection, two very commonly used classes of methods for Class-1 type documents are described: *region classification-based* and *CC classification-based*. A region classification-based method makes the decision regarding text or non-text over segmented regions (Fig. 5a), whereas CC classification-based methods perform similar operation over the CCs (Fig. 5b).

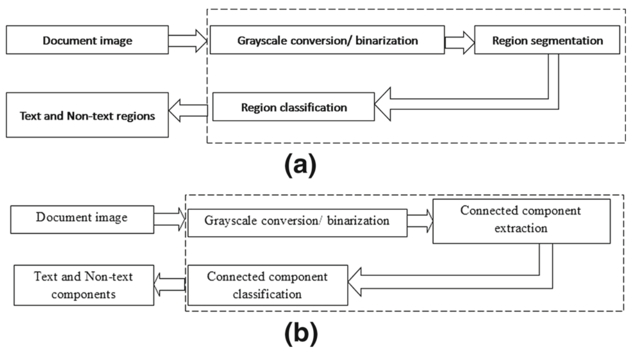


Fig. 5 Schematic diagram of **a** a typical region classification-based method and **b** a typical CC classification-based method

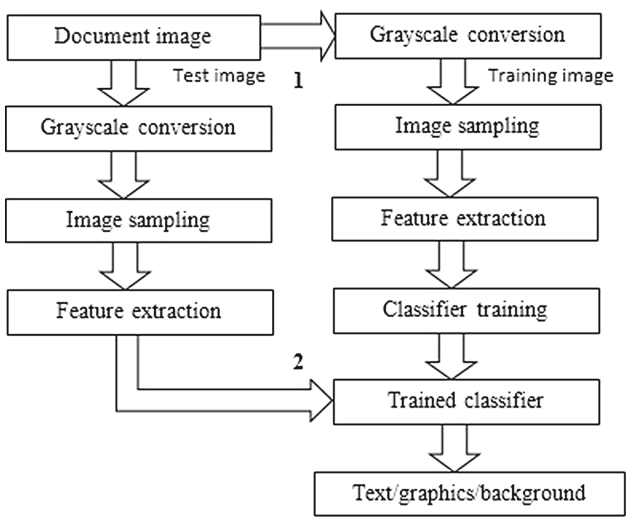


Fig. 6 Several steps performed in a region classification-based method [55]

3.2.1 Region classification-based methods

First, region segmentation is generally performed using a top-down [67–69], bottom-up [70–72] or hybrid [73–75] approach. For top-down approaches, segmentation goes from a coarse level to finer level where large homogeneous regions are first extracted and then refinement is performed at the subsequent levels. In bottom-up approaches, processing is performed in the reverse direction, i.e., from the pixel level to the region level. It starts with merging the pixels to form characters, characters to words, words to text lines and then text lines to regions. Hybrid methods use features of both these approaches for region segmentation and typically is the most widely reported.

Oyedotun and Khashman [55] recently published a region classification-based method for this category of document, and the steps of this method are shown in Fig. 6. They have performed top-down-based region segmentation by shifting a mask of size $m \times n$ over a grayscale input image. From each segmented region, first-order statistical features are extracted including median and modal pixel values along with the GLCM-based second-order statistical features, such as entropy, homogeneity, energy and contrast. A feed-forward neural network-based classifier is trained to classify those regions as text or non-text area. Although the method has achieved reasonable accuracy, authors have not specified how the described method handles a situation where the underlying region of the scanning window includes both text and graphics.

Shih and Chen in [58] and Fan and Wang in [76] use a run length smoothing algorithm (RLSA)-based bottom-up method for initial region segmentation of binarized input images. In [57], for region classification, several shape-based features and black-to-white pixel transition features are extracted, whereas in [76], density features are used. Both these methods have followed a region segmentation policy which may not be useful for document images with non-rectangular shaped regions or document images with complex layout. Thus, they may generate segmented regions comprising both texts and non-texts.

Antoncopoulos and Ritchings [77] and Pavlidis and Zhou [78] apply hybrid approaches which use background information to segment input image into regions. After segmentation, authors in [77] employ a white tile-based texture features to develop rules to classify the binarized segmented regions as text or non-text. In [78], a signal cross-correlation function has been used to compute several statistical features like average cross-correlation between adjacent scanlines, average correlation between two scan lines with four intermediate scanlines in between and average number of black runs to separate the text regions from non-text ones. Although both these methods use background information for region segmentation, the method described in [77] is much more flexible than that reported in [78], as the second method has lot of assumptions regarding the spacing of content.

A major problem with the region classification-based methods is that the performance of these methods is often very sensitive to initial region segmentation result. If the initial segmentation process erroneously generates a region having both text and non-text elements, then a region classification-based method classifies the entire region either as text or non-text. In addition, this type of methods often results in regions of a document containing pictures with text embedded. Hence, these methods may fail, as they make the classification decision for an entire region.

3.2.2 Connected component (CC) classification-based methods

In these approaches, CCs are extracted from the input document image and then each CC is classified as text or non-text.

Le et al. recently presented a CC-based method [44] for Class-1 documents. The authors compute a set of features

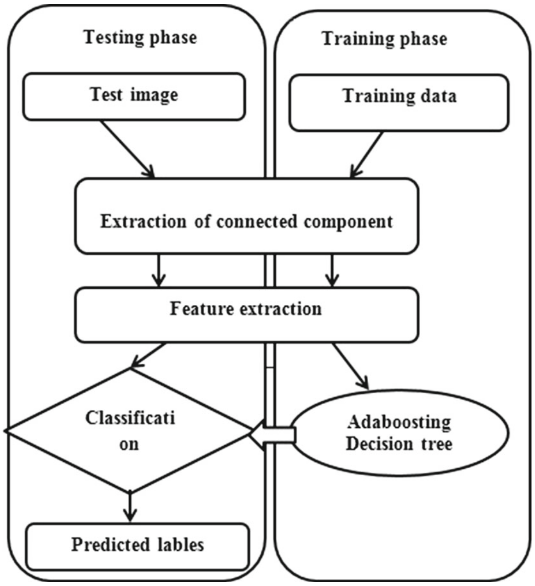


Fig. 7 Stages of a CC classification-based method [44]

using size and shape information, stroke width and the position of CCs. Finally, an Adaboosting decision tree is used to classify the extracted component as text or non-text. Figure 7 presents the steps of the reported method. Though the proposed method has achieved satisfactory results, the authors have not clearly described how they have filtered out the noisy components. The removal of noisy components is an important issue for these categories of methods as the presence of these noisy components can cause serious degradation in recognition result.

In [4], Kim et al. have reported a two-stage CC classification-based method. Initially, extracted components are filtered using a heuristic filter to suppress the non-text. The heuristic filter uses shape, size, aspect ratio, position of the component, pixel density and the number of components contained in the bounding box of a CC for an initial classification. At the second stage, a recursive filter is used which performs homogeneous region detection and white space analysis on the basis of statistical features including median, variance and contextual information to separate the remaining components. On the other hand, Drivas and Aims [79] have used frequency of the components based on height, ratio of the total area covered by the components and the total area covered by white pixels along with the geometrical appearance of the components to devise a set of rules, which are then used to classify the extracted components. The methods described in these papers take a rule-based separation approach, but it is very hard to make generalized rules to classify the components for a large range of documents.

Bukhari et al. report a method [13] where shape and context information are computed from the components to train a multilayer perceptron (MLP) classifier for text and non-text separation. The authors here have assumed that the context of a text component is always structured, but we find the context of a text does not always have this property. For example, the text close to non-text can also have context which appears random. In that case, misclassifications may occur.

Methods based on CC classification work at lower level than that using region classification so they can overcome the erroneous classification of an entire region. Most of these methods also detect the text appearing within a non-text region. But these methods often suffer from the presence of touching components. For example, if a text component is attached with a non-text component, then in these methods, it may be considered as a single component and can be mistakenly classified as one of the two classes. This issue is less common in Class-1 documents, which is why such methods are often applied.

3.2.3 Hybrid methods

In the literature, some hybrid methods are also reported which combine both region and CC classification-based approaches. A classic method is introduced by Fletcher and Kasturi [10] which extract the CCs from input image and then filter them on the basis of area and aspect ratio to suppress non-text ones. Then, a Hough transform is used to group the components together to form logical strings. Each logical string is examined on the basis of structural features including size, area, variance in aspect ratio, etc., as well as pixel density to separate components at final level. In [80], Chowdhury et al. present another method, in which texture analysis is performed on the grayscale image to separate halftones at first stage and then the resultant images are binarized for further processing. From the binarized image, CCs are extracted and pixel density information along with basic characteristics of English characters is used to classify the extracted components.

In addition to these methods, other methods also exist in literature, which have not received much attention from the researchers. One such example is a multi-resolution morphology-based method [81], introduced by Bloomberg et al. This method performs a series of morphological operations to remove halftones from the document image. But it is not very efficient in separating other types of graphics like drawings and so Bukhari et al. [82] have reported an improved version of it.

3.3 Methods for Class-2 documents

Unlike Class-1 documents, text in engineering drawings is generally scattered throughout the image. Thus, it is not, in general, possible to segment these images into large isolated polygonal regions dedicated to text or non-text [83]. This restricts the use of typical region classification-based techniques and requires a finer segmentation.

Methods in the literature for separation of text and non-text in engineering drawings are either performed on the *raster image* (which is the scanned version of original image) or on *vectorized image* (which is the result of vectorization of raster image) [83].

3.3.1 Raster image analysis-based methods

Raster images are a scanned representation copy of original image. Most of the raster image analysis-based methods follow a typical CC classification-based approach. Tomber et al. [84] propose a raster image analysis-based method where initially non-text components are separated by defining a threshold on the size of the bounding box of the components. At the next level, separation is carried out on remaining components by using the pixel density and elongation thresholds. In [42], Lu et al. examine the extracted CCs using observed characteristics of the elements present in an engineering drawing on the ground of various geometrical features like shape, size, local stroke density and length of the components. Similarly in [85–87], extracted components are analyzed using black pixel count, alignment criteria and a set of heuristic rules (which have been designed using height and width of the components), respectively.

Recently, Do et al. [88] have reported a multi-learned dictionary-based method. In this method, they develop two separate dictionaries, one for text part and another for non-text part. Non-overlapping patches present in those dictionaries are represented using a sparse representation and based on which patches from the input images are classified.

3.3.2 Vectorized image analysis-based methods

Vectorized images are the result of a vectorization process [89,90] performed on the raster image. Methods developed to perform the separation task on this type of images, generally, accept vector-representations (arcs and lines) as input and classify them as text or non-text [83]. For example, in [16], Dori and Wenyin use length and stroke like appearance to get candidate texts. In [91], Dori and Velkovitch have developed a set of rules based on the position, shape and connectivity of the components to eliminate the non-text components.

Most of these methods follow a rule-based strategy to classify the extracted components. Although these rules are formed using heuristics associated with engineering drawing standards, they can still be a bit restrictive as defining an equally effective set of rules for a wide range of documents is very hard.

While developing a complete system for processing of filled-in checks and forms, researchers perform the separation task for various purposes like, filled-in data extraction and effective storage of input images [92]. To perform the separation in these documents, similar to the engineering drawings, they follow CC analysis-based approach.

Casey et al. in [92] attempt to identify the extraneous markings or lines (mostly curved) that may pass through the form data, by analyzing the height of the components. They consider components having height greater than the expected height of the characters, as candidate of extraneous markings. Using a continuity-following algorithm, they track these lines and removed them, in order to produce clean document image for filled-in data extraction and OCRing. But this method cannot take care of the extraneous lines having height lesser than the expected character height. They also perform a template matching technique to eliminate the background frame information (includes lines, tables, boxes, etc.), in order to store the form information effectively. But to do this, they not only perform image registration but also perform a pixel level comparison, which make it time-consuming. Similar to this method, Yu and Jain in [93] capture the frame structure from blank forms and suppress that in filled-in forms. To represent the frame structure they use block adjacency graph (BAG). In this work, they specially focus on the horizontal line segments and analyze them to detect and separate touching characters.

Wang and Srihari in [94] focus on separating the characters as small components from large components like complex boxes and large line segments by comparing the height of the CCs with a threshold. For that purpose, they compute a histogram of components heights. A large peak in the histogram generally corresponds to the height of character components. This is because the number of character components in form documents is greater than the number of other components like lines, boxes, etc. The component height at which the cumulative probability density exceeds 95 percent is chosen as threshold. The larger components are further examined to extract the touching components. As in this method the size is considered to separate the characters from large components, the other components like choice boxes, dotted lines and noises may also be classified as small component along with characters. In addition to that, this method is developed for typed forms and thus may not work well for forms with handwritten filled-in data.

Hori and Doermann in [95] use a box-driven technique to analyze the structure of a table-form. For that purpose, they initially compute the contour of the form objects and then estimate the bounding boxes of these contours. That means all the form objects are represented by their bounding boxes. They label these boxes as characters, cells, line segments, margin, noise, etc., based on some rules, generated by considering the height and width of these boxes. After that, lines and cells are further examined to detect and separate touching characters. Like the method described in [94], this is also developed for forms with no handwritten filled-in data.

Similar to forms, in filled-in checks, text/non-text separation is mainly performed during the extraction of various filled-in data (generally handwritten) [45]. For example, in [96], to extract and recognize the courtesy amount, Dzuba et al. remove the unwanted objects like long line segments using technique based on Hough transform. They eliminate the symbols like dollar, cent, etc., from courtesy amount based on shape, size and relative positions of the components. Heutte et al. in [97] also use Hough transform to detect the guidelines and boxes for courtesy amount extraction. Knerr et al. in [98] use the thickness, direction and curvature of line to remove the lines from courtesy amount in French checks.

In addition to the courtesy amount, another most important filled-in data in checks are legal amount, which is basically written in words on long line segments. While extracting and recognizing this information, removal of these line segments is very essential. As often these lines touch the filled-in data, all methods introduced for removing these lines take care of this issue. Govindaraju et al. in [99] separate these lines using a stroke length-based technique. Before estimating the stoke length, a thinning operation is performed and then the smoothed strokes are identified. Dimauro et al. in [100] use a morphology-based technique with dynamic selection of structuring element to separate the lines from the legal amount data. Akiyama et al. in [101] and Liu et al. in [102] use Sobel operator to generate the gradient images at initial stage. Then, they apply a local thresholding technique to get the candidate edge points of the line segments. After that they detect and eliminate the lines using least square fitting and Hough transform-based methods.

3.4 Methods for Class-3 documents

Class-3 type documents typically have complex backgrounds, which poses an additional challenge for segmentation and classification. Typical region or component classification-based methods would not be useful here. For that reason researchers have given more importance to the initial intensity level processing. After an initial stage, further analysis is required on the region, component and pixel levels as well. Figure 8 presents a general framework for text/non-text separation in Class-3 type documents.

Initial intensity level processing typically includes commonly used grayscale conversion and image binarization processes, as well as color reduction and image layer separation (or color-based segmentation). For example, Vu et al. [51] present a method which performs three-stage preprocessing to improve the contrast of the grayscale image. Layer separation then uses a multi-level thresholding method to separate the text region from the non-text regions. Finally, recursive and another heuristic filters are used to localize the text region.

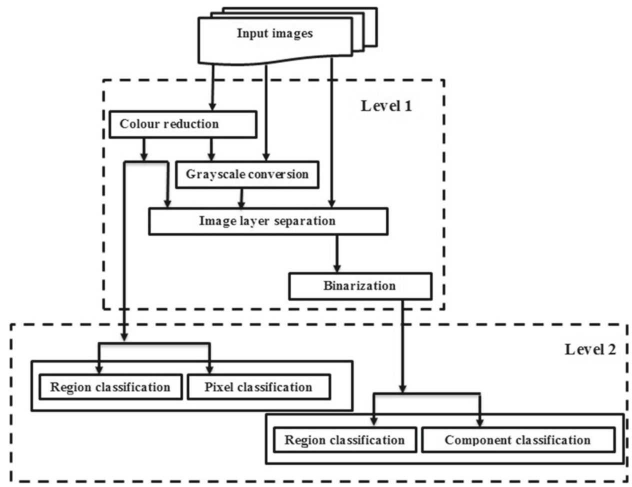


Fig. 8 A general framework for text/non-text separation in Class-3 type documents

In [48,62] Chen et al. use a multi-plane segmentation technique to perform the layer separation which decomposes the original image into different object planes for separating homogeneous objects like candidate text regions, graphics, pictures and background. Each plane is binarized for further processing. In [48], a knowledge-based system is developed to perform component classification on binarized image. Whereas method in [62] merges CCs using recursive X-Y cut into different clusters (or regions) and classification is done at cluster level.

Strouthopoulos et al. [63] perform an unsupervised adaptive color reduction using a neural network classifier and tree-search-based procedure. Sobottka et al. [52] do the same using a histogram-based clustering technique. In [63], the resultant color images are divided into different color planes. On each plane, RLSA-based segmentation technique is applied. Spatial features are extracted from the texture patterns of segmented regions to classify them as text or non-text. Text boxes are then identified from each color plane and merged, and further examined on the basis of color to separate text from background. In [52], after color reduction, top-down and bottom-up analyses are performed independently on the resultant image. During top-down analysis, the images segmented into rectangular regions and the regions with at least two colors are classified as text. Bottom-up analysis is performed using a region growing technique. Here, merging is done on the basis of color information. Finally, the outcomes of both methods are combined.

Hase et al. [103] segment a full color image into a set of representative binary images then uses a method called multi-stage relaxation to nominate candidate character strings from the resultant images. As the extracted string may contain non-text components, a likelihood of a character string for each element is defined using the Mahalanobis distance which

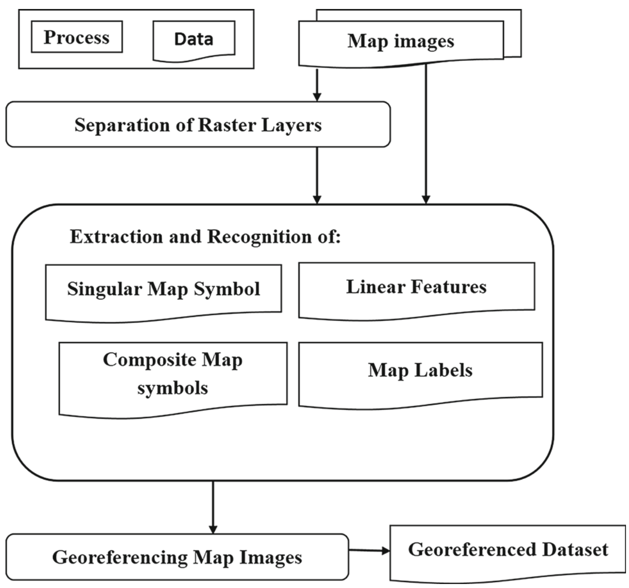


Fig. 9 A typical map processing system [15]

is calculated on the basis of density features of lines and spaces. All the elements extracted from the representative color images are superimposed and if any overlap occurs then two sets of rules are devised to handle that issue, where one set of rules is formed based on the relation of inclusion, the other one uses the relation of overlap.

Clavelli and Karatzas [104] use a segmentation algorithm which creates an 8-way CCs based on color similarity. CCs are extracted from resultant image and examined on the basis of their size. Very large or small sized components are eliminated as non-text and remaining components are examined on the basis of spatial and topological characteristics of the components to classify them as text or non-text.

Methods discussed above are designed to perform text/non-text separation in book and/or magazine covers, and CD covers for example. A significant number of method are also reported in literature which perform text/non-text separation for different types of maps [53,64,105]. Figure 9 shows the stages of a typical map processing system presented by Chiang et al. in [15].

Text/non-text separation on maps generally requires two-stage processing: raster layer separation and separation of text from other geographic elements (which includes linear objects like road/contour lines, other map objects and symbols). Main purpose of the raster layer separation is to segment different color layers present in a map to facilitate further processing [15]. A few methods are devised that rely on only color information [106,107], whereas some methods use both color space and local spatial context implicitly [108,109]. There are other methods that use explicitly spatial context along with color space [110]. But research trend in this direction is beyond the scope of this paper. Generally, raster layer separation methods generate a set of image layers or map layers from original map image. These layers are bicolor image containing different map objects. As linear objects and text may appear in same layer, separation of text and non-text on maps implies the separation of texts from those linear objects.

Cao and Tan [64] initially separate the sub-layer from the map images by defining a threshold on pixel intensity. Solid graphical components are then eliminated using morphology-based method described in [111], and dashed lines are removed using total line regression. Components present in the resultant image are further examined on the basis of size, and the components with large size are considered as non-text. Non-text components are further examined to separate the touching texts. Roy et al. [53] present an improved version of the method described in [84], though this is not efficient in handling the component overlapping issue. This method, initially, separates the text form the non-text component by using a set of rules on the basis of the pixel density and height, width of the bounding box of the components. At the next level, texts attached with graphics are separated by applying clustering on the basis of color information, Hough transform and skeleton segmentation.

Raveaux et al. [112,113] present a text/non-text separation method for historical maps. An anti-fading algorithm is used to restore the lost colors, and then color space selection is done using an expectation maximization algorithm. Edge detection based on the selected color space is then used to get the CCs. Finally, extracted components are classified using context information.

Pezeshk and Tutwiler [114] design a set of morphological operations to extract linear objects with any curvature from binarized image. They have also claimed that the designed morphological operations can extract a linear object even if it is attached with text. At the second level, they have extracted the objects like buildings, dashed lines, etc. This results in a text-only image. Finally, texts are grouped using pyramid decomposition into their respected strings.

Chiang and Knoblock [115] report a general training-by-example approach (named by Chiang et al. in [15]) for separating text from non-text in heterogeneous raster maps. In this work, a user marks the text area from where text pixel information is extracted to separate the text from the non-text. In [116], they describe a system called "STRABO" which follows a similar approach to extract text pixels in map images. Once the text pixels are identified, neighboring text pixels are merged to generate character and the adjacent characters are further merged to generate text strings. But the system is not very efficient when characters touch each other. Hence, they have recently introduced an improved version in [117] where initially a potential character region around the text identified by STRABO is estimated. Then from each potential region, information including the number of foreground pixels divided by the size of the area, the number of CCs, number of holes and the presence of linear objects is extracted to classify them as text or non-text. As the dimension of the potential character region is same as the bounding box of a character, described method has successfully extracted the text, attached with graphics.

Extraction of text attached with other linear objects is the major problem of text/non-text separation in maps. Thus, most of the recently reported methods for maps have given a special attention to solve this problem.

3.5 Methods for Class-4 documents

Text/non-text separation in this category of documents has remained almost unexplored compare to other classes of documents. In the literature, methods that have been introduced are split between applications of historical and regular handwritten manuscripts.

3.5.1 Methods for historical manuscripts

In general, historical manuscripts have very noisy backgrounds due to aging, and thus, the methods reported have followed an approach similar to methods applied for Class-3 documents.

Recently, Kavitaet al. [118] develop a text/non-text separation method for Indus civilization documents. Initially, input images are sharpened using the combination of Sobel and Laplacian methods [119]. Extracted CCs of the enhanced images are then skeletonized and clustered using nearest neighbor-based method, and the clusters are classified as text or non-text by the number of branches present in each cluster. The proposed method assumes that text components are less "cursive" than non-text components, so the components with fewer branches are considered as texts. Although this assumption may work for the type of handwritten documents they have considered, it is not always true.

Cohen et al. [120] describe a method to separate text from non-text components present in handwritten historical manuscripts. The proposed method performs a multi-scale analysis on the original document image to estimate candidate text lines, and the components present in these text lines are further clustered using K-means on the basis of aspect ratio to separate out noise from texts. The estimated text lines are then used to extract the textual part from corresponding binarized image. CCs from the binarized image that overlap on the estimated text lines are examined based on the information such as aspect ratio, stroke width and distance between successive text lines to classify them as text or non-text. Non-text components are further examined to separate the background, and noise from the drawings.

Table 1 Commercial/non-commercial systems perform text/non-text separation

Name	Citation	Distribution
ABBYY Fine reader	[121,122]	Commercial product
OCRopus	[123,124]	Open source OCR system
Tesseract-OCR	[125,126]	Open source OCR system

3.5.2 Methods for regular handwritten manuscripts

Regular handwritten manuscripts do not suffer as much quality degradation as historical documents. Thus, methods introduced in the literature to date typically do not perform complex intensity level operations. Sarkar et al. in [7] present a typical CC classification-based method to separate text from non-text components for this type of document. A technique called Spiral Run Length Smearing Algorithm (SRLSA) is used to classify the extracted components from a binarized version of the input image as text or non-text. A similar method is recently proposed in [61] by Bhowmik et al. where the LBP operator is used to classify the CCs. Although these methods have considered the most complex type of documents according to our categorization, these methods have not addressed the most common issue of handwritten documents, i.e., touching components.

3.6 Other general systems

To the best of our knowledge, there is no software/tool mentioned in the literature which is solely dedicated to the perform text/non-text separation task. But there are some commercial/ non-commercial systems that carry out text/non-text separation as a part of their system, before recognizing the text. They do not expose their technique so it is difficult to include them in the taxonomy above. For reference, they are listed in Table 1 and will be included in the evaluation section.

4 Performance evaluation

The performances of the methods discussed in Sect. 3 are often evaluated on user defined datasets, although some recent competitions have driven the community to require the use of standardized datasets. A detailed description of several standard databases and commonly used evaluation protocols is described in the following subsections. Comparison of performance of selected methods is also presented on these datasets.

Table 2 Standard datasets used to evaluate the text/non-text separation methods

Database	Citation	Class	Quantity (no. of pages)	Availability	Language(s)
UCI Machine Learning Repository	[55,127]	Class-1	101	Publicly available	Russian
UW-I English Document Image Database	[80,128]	Class-1	1147	Restricted available	English
UW-II English/Japanese Document Image Database	[80,129]	Class-1	1101	Restricted available	English, Japanese
UW-III English/Technical Document Image Database	[44,130]	Class-1	1,777	Restricted available	English
ICDAR 2009 Dataset	[44,131,132]	Class-1	1240	Available by request	English
The Media Team document database II	[49,133]	Class-1 Class-3	512	Publicly available	English
UNLV Database	[82,135]	Class-1	–	Not available	–
Diotek Company Database	[4]	Class-1	–	Restricted available	Korean
6-Inch Historical Ordnance Survey maps of the United Kingdom (UK)	[117,136]	Class-3	–	Restricted available	English
Islamic Heritage project (IHP)collection	[120,134]	Class-1 Class-3 Class-4	156,000	Publicly available	Arabic
					Persian
					Ottoman Turkish
					Urdu
					Chagatai
					Malay
					Gujarati
					Several Western languages

4.1 Standard databases

In this subsection, standard databases of document images which are used to evaluate the performance of different methods are listed for comparison. A brief description of those databases is presented in Table 2.

4.2 Evaluation protocols

Text/non-text separation techniques available in the literature for different categories of documents are generally evaluated using one of the following ways,

- Human Vision-based visual evaluation [10,78,137]
- Quantitative evaluation [4,44,55]

In Human Vision-based visual evaluation, output images are manually examined to assess the effectiveness of the methods. Many of the earlier methods [10,78,137] have followed this approach to evaluate their systems due to the absence of an effective quantitative evaluation technique and ground truth data.

On the other hand, quantitative evaluation is the formal comparison between the output of a method and corresponding ground truth. Like classification itself, these evaluations are typically done at the pixel [13], component [44] or block [56] levels. Researcher has suggested several metrics [138] for this purpose, but most commonly used metrics are,

$$\text{Precision} = \frac{T_{\text{positive}}}{T_{\text{positive}} + F_{\text{positive}}} \tag{1}$$

Table 3 Variable definitions for quantitative metrics

Term	Class	Text: Pixel level	Component level	Block level
T_{positive}	True positive	Number of text pixels correctly predicted as text pixels	Number of text components correctly predicted as text components	Number of text blocks correctly predicted as text blocks
F_{positive}	False positive	Number of non-text pixels incorrectly predicted as text pixels	Number of non-text components incorrectly predicted as text components	Number of text blocks correctly predicted as text blocks
F_{negative}	False negative	Number of text pixels incorrectly predicted as non-text pixels	Number of text components incorrectly predicted as non-text components	Number of text blocks incorrectly predicted as non-text blocks
T_{negative}	True negative	Number of non-text pixels correctly predicted as non-text pixels	Number of non-text components correctly predicted as non-text components	Number of non-text blocks correctly predicted as non-text blocks

$$\text{Recall} = \frac{T_{\text{positive}}}{T_{\text{positive}} + F_{\text{negative}}} \tag{2}$$

$$F\text{-measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

$$\text{Accuracy} = \frac{T_{\text{positive}} + T_{\text{negative}}}{T_{\text{positive}} + T_{\text{negative}} + F_{\text{positive}} + F_{\text{negative}}} \tag{4}$$

The definition of variables in equations 1, 2 and 4 for each level of comparison is given in Table 3.

4.3 Performance comparison

In this subsection, a comparative study of the methods for separation of text and non-text regions for different document classes considered is presented in tabular form. Tables 4, 5, 6 and 7 present the performance comparison of the methods for Classes-1, 2, 3 and 4 respectively.

5 Summary and discussion

In this section, we overview how the different methods handle the major issues in separating the text and non-text regions, and summarize limitations. We also highlight possible future directions for research.

5.1 Primary issues and challenges

To develop an efficient text/non-text separation method, one has to overcome common challenges such as degraded document quality, regular and irregular noise, and text touching non-text. In addition, recent advances in document generation have introduced additional obstacles such as more complex layouts and backgrounds.

5.1.1 Complex backgrounds

Complex backgrounds are found primarily in Class-3 type documents but may occur in the documents of Class-2 due to aging or other degradations. As it is common to Class-3, methods developed for the documents of this category mostly address this issue. For example, the method described in [51] uses a multi-level thresholding of intensity values to separate the complex background from the foreground. But it has been observed that this method has not performed as expected while dealing document images with low contrast and resolution. Authors in [48,62] perform multi-plane segmentation, using automatic localized histogram multi-level thresholding and multi-level region matching and assembling. Authors claim that the method has performed well, not only for the documents with complex background but also for the doc-

Table 4 Performance comparison of recent text/non-text separation methods for Class-1 documents

Method	Database	Performance measure (*in%*)						
		Text				Non-text		
		Accuracy	Precision	Recall	F-measure	Precision	Recall	F-measure
Vil'kin et al. [56]	UCI Machine Learning Repository [56] (subset of 35 images)	84.23	NA	NA	NA	NA	NA	NA
Le et al. [44]	UW-III [130], ICDAR2009 [132] (all total a subset of 305 images)	NA	97.32	99.17	NA	81.65	73.09	NA
Tran et al. [4]	ICDAR2009 [132] Database of Diotek (all total a subset of 114 images)	NA	94.99	93.31	NA	88.99	87.22	NA
Vil'kin et al. [56]	Own (174 images)	93.20	NA	NA	NA	NA	NA	NA
Zagoris et al. [49]	Media Team document database [133] (512 images)	96.62	NA	NA	NA	NA	NA	NA
Bukhari et al. [82]	UW-III [130], ICDAR2009 [132], UNLV [135] and Own database (all total a subset of 215 images)	NA	NA	98.19	NA	NA	91.90	NA
Emmanouilidis et al. [50]	Media Team document database [133] and UW-III [130] (all total a subset of 35 images)	95.25	98.66	94.72	NA	87.43	96.61	NA
Chowdhury et al. [80]	UW-I [133] and UW-III [130] (all total a subset of 200 images)	97.50	97.90	97.00	NA	97.02	98.00	NA
Lin et al. [57]	Media Team document database [133] (subset of 50 images)	NA	96.43	NA	NA	90.51	NA	NA
Tesseract [44,126]	UCI Machine Learning Repository [127] (subset of 35 images)	NA	NA	NA	92.50	NA	NA	74.23
FineReader [44,121]	UW-III [130], ICDAR2009 [132] (all total a subset of 305 images)	NA	NA	NA	93.09	NA	NA	71.75
OCRopus [44,123]	ICDAR2009 [132] Database of Diotek (all total a subset of 114 images)	NA	NA	NA	84.18	NA	NA	51.08

NA not available

uments with degraded background or sharp variation in the illumination. In [104], a color-based segmentation has been presented using a global threshold, which is chosen experimentally. Generally, global threshold-based techniques are not very useful for the documents with textured/complex colored background. In [63], and [52], the authors have performed a color quantization-based approach. In [63], authors use a neural network classifier and tree-search procedure-based method, whereas in [52], a histogram-based clustering technique is used. The method described in [103] converts the input images into a uniform color space and then follow an adaptive color selection method based on image histogram to perform the color level segmentation.

5.1.2 Inverse or reverse video

Inverse or reverse video generally refers to the situation where a lighter foreground appears with a dark background.

Table 5 Performance comparison of text/non-text separation methods for Class-2 documents

Method	Database	Performance measure ($in\%$)						
		Text				Non-text		
		Accuracy	Precision	Recall	F-measure	Precision	Recall	F-measure
Do et al. [88]	Own (5 images of engineering drawings)	NA	NA	92.25	NA	NA	NA	NA
Tombre et al. [84]	Own (5 images of engineering drawings)	NA	NA	73.00	NA	NA	NA	NA
Dori and Wenyin [16]	Own (engineering drawings)	94.00	NA	NA	NA	NA	NA	NA
Lu et al. [42]	Own (4 images of engineering drawings)	NA	NA	81.56	NA	NA	NA	NA
Dori and Velkovitch [91]	Own (14 images of engineering drawings)	89.00	NA	NA	NA	NA	NA	NA

Table 6 Performance comparison of recent text/non-text separation methods for Class-3 documents

Method	Database	Performance measure ($in\%$)						
		Text				Non-text		
		Accuracy	Precision	Recall	F-measure	Precision	Recall	F-measure
Nazari et al. [117]	6-inch Historical Ordnance Survey maps of the United Kingdom [136] (10 map images)	NA	77.00	86.00	NA	NA	NA	NA
Vu et al. [51]	Own(100 images of book and magazine covers)	94.30	NA	NA	NA	NA	NA	NA
Chiang and Knoblock [54]	Own(15 map images)	NA	92.70	87.90	NA	NA	NA	NA
Chen et al. [48]	Own(50 images of book and magazine covers)	NA	98.90	99.30	NA	NA	NA	NA
Clavelli and Karatzas [104]	National Archive of Catalonia in Spain (50 poster images)	NA	76.90	73.20	NA	NA	NA	NA
Chen and Wu. [62]	Own(65 images of book and magazine covers)	NA	NA	99.30	NA	NA	NA	NA
Roy et al. [53]	Own(10 map images)	NA	NA	98.80	NA	NA	83.80	NA
Cao and Tan [64]	Own(24 map images)	NA	87.10	99.40	NA	NA	NA	NA
Sobottka et al. [52]	Own(16 images of book and magazine covers)	70.00	98.33	51.27	NA	NA	NA	NA

Table 7 Performance comparison of recent text/non-text separation methods for Class-4 documents

Method	Database	Performance measure ($in\%$)						
		Text				Non-text		
		Accuracy	Precision	Recall	F-measure	Precision	Recall	F-measure
Bhowmik et al. [61]	Own(100 images of unconstrained handwritten pages)	90.80	NA	NA	NA	NA	NA	NA
Kavitha et al. [118]	Own(500 images of Indus documents)	89.00	NA	NA	NA	NA	NA	NA
Cohen et al. [120]	Islamic Heritage project (IHP) collection [134] (subset of 252 images)	NA	99.80	89.60	NA	89.60	94.90	NA
Sarkar et al. [7]	Own(9000 components)	83.30	NA	NA	NA	NA	NA	NA

Some approaches which deal with reverse video in scene images and video frames are reported in [139,140], while the methods described in [51] and [62] deal with reverse/inverted text in offline document images. Although the methods in [52,63] can generally handle this issue , but most of the earlier methods such as [77,141] developed for Class-1 documents do not consider this case. In recent work, researchers have developed methods [142,143] for the documents belonging to Class-1, where they have considered this situation as a special case.

5.1.3 Overlapping or touching components

This is the most important issue in text/non-text separation and appears almost all classes of documents except Class-1, where the chance of overlap is very low. But the nature of overlapping text varies by application, so researchers tend to develop custom techniques.

While developing methods for Class-3 type documents, researchers rely primarily on color level information, along with some topological information, including orientation. For example, in [117], authors have considered two potential character regions at the sides of detected text using orientation. Each potential character region is then classified as text or non-text. These steps have been performed repeatedly until two consecutive regions are classified as non-text or a region is found with too few foreground pixels. The proposed method performs poorly when text overlaps with too many non-text regions. This method may also suffer if orientation of the detected text is not estimated correctly. The method described in [53] uses a color segmentation-based approach to separate the text having a different color than the touching non-text. A Hough transform and a skeleton segmentation are then performed to separate the text having same color as the touching non-text. But it has been observed that the method fails at color segmentation when one color gets mixed with another and it also faces difficulty during skeleton segmentation if a long line touches a dense blob.

Authors in [64] perform a thinning operation on the touching components, and the intersection points are identified to segment the components into fragments. The slope of each fragment is estimated using line fitting. The fragments connected to same junction with a similar slope are merged into a line. After this decomposition, fragments are filtered using size to separate texts from the non-texts. Removal of non-texts can damage texts at the touching area so the average width of the line is computed to restore the damaged area. It is assumed that the strokes of text are shorter than those of graphics. But with this assumption, system may face problems when text attaches to a dashed line. In [84], authors have specified a search region which depends on detected text not attached with a graphics and uses its orientation. Text attached to graphics resides in those search areas has then been skeletonized and segmented. Each segmented area is reconstructed using an inverse distance transform. The major limitation of this method is that it cannot separate the text strings completely connected with a graphics.

In addition to the above technique, there are a significant number of methods available in the literature that address the touching characters in filled-in forms and check images. Maderlechner et al. in [144] use a skeleton analysis-based method to separate the touching character(s) from the form structure. Casey et al. [92] perform the separation task using a line tracking technique. Doermann and Rosenfeld [145] solve this issue using cross section computation. Hori and Doermann in [95] check the presence of distortion at contours to locate the touching character. In addition to these, many other methods [93,146] are also available to take care of this issue in form and check images. But all these methods are only useful when the character components are attached with a long and straight line like component.

5.1.4 Handling of documents of degraded quality

In addition to the above issues, quality degradation is another common concern to handle in document image processing research. Thus, a large collection of methods [147–153] are available for the purpose of handling such documents.

A lthough most of the reviewed methods for Class-2 and Class-3 documents have addressed these issues while handling documents with complex background and overlapping or touching components, challenges remain.

5.2 Future research

From this survey it is clear that researchers have shown a lot of interest in text/non-text separation in the recent past. As a result, new methods are being proposed or earlier solutions are being modified to make them more applicable to this problem. For example, in last few years, researchers have started to apply deep learning-based solution to various document analysis problems. These methods have proven more accurate than traditional methods in a number of areas including scene text extraction [135–137]. Similar solutions can be applied to the problem of text/non-text separation in general offline documents with complex layouts as well.

While this survey reflects the evidence of successful progress in the development of an effective text/non-text separation method for printed documents, it also reveals that almost no effort has been made for purely handwritten manuscripts. Even though few methods for this category of documents are discussed in our survey, these do not address the most crucial and frequent issues, i.e., touching or overlapping components.

Most of the methods available in the literature that deal with touching components for Class-3 type documents are not very useful for handwritten documents. The main reason for this is that most of the proposed methods either use the color information, the orientation of the text or the height and width of the detected text. But, in general, regular handwritten manuscripts are written using single color of ink, so color differences may not be useful. In unconstrained handwritten manuscripts, orientation, height and width are less reliable. Although a significant number of methods are available which address this issue for Class-2 documents, they are only developed for handling a specific type of touching scenario, where one or more characters from filled-in data touch a long and straight line. While dealing with unconstrained handwritten documents, an arbitrary shaped text component may overlap with another arbitrary shaped text or non-text component. In contrast to these methods, some research work also available in the literature which consider touching character issue in completely unconstrained handwritten text-only documents. One such method is introduced by Kang and Doermann [154], where they use template-based method to separate the touching components in handwritten text line. This is basically a recognition-based segmentation technique. In addition to the recognition-based methods [155,156], many other methods are also present in the literature which segment the touching components based on contour, skeleton, projection profile analysis [157,158]. Although these methods are built for handwritten text-only documents these can be extended to solve the issue of touching component for the documents considered under Class-4.

Another problem often seen handwritten manuscripts is annotations that are not part of the normal flow of the document. The lack of an effective method for separation of text/non-text present in the handwritten documents has created a huge vacuum in the domain of document image processing. Thus, developing an appropriate text/non-text classification system is desired to design a complete document processing system for the document images particularly in handwritten.

References

1. Lucas, S.M., et al.: ICDAR 2003 robust reading competitions: entries, results, and future directions. Int. J. Doc. Anal. Recognit. **7**(2–3), 105–122 (2005)
2. Shahab, A., Shafait, F., Dengel, A.: ICDAR 2011 robust reading competition challenge 2: reading text in scene images. In: 2011 International conference on document analysis and recognition (ICDAR), pp. 1491–1496 (2011)
3. Yin, X.C., Zuo, Z.Y., Tian, S., Liu, C.L.: Text detection, tracking and recognition in video: a comprehensive survey. IEEE Trans. Image Process. **25**(6), 2752–2773 (2016)
4. Tran, T.-A., Na, I.-S., Kim, S.-H.: Separation of text and non-text in document layout analysis using a recursive filter. KSII Trans. Internet Inf. Syst. **9**(10), 4072–4091 (2015)

5. Yu, C., Song, Y., Zhang, Y.: Scene text localization using edge analysis and feature pool. Neurocomputing **175**, 652–661 (2016)
6. Asif, M.D.A., Tariq, U.U., Baig, M.N., Ahmad, W.: A novel hybrid method for text detection and extraction from news videos. Middle East J. Sci. Res. **19**(5), 716–722 (2014)
7. Sarkar, R., Moulik, S., Das, N., Basu, S., Nasipuri, M., Kundu, M.: Suppression of non-text components in handwritten document images. In: 2011 International Conference on Image Information Processing (ICIIP)
8. Wahl, F.M., Wong, K.Y., Casey, R.G.: Block segmentation and text extraction in mixed text/image documents. Comput. Graph. Image Process. **20**(4), 375–390 (1982)
9. Bunke, H.: Automatic interpretation of lines and text in circuit diagrams. In: Kittler, J., Fu, K.S., Pau, L.F. (eds.) Pattern Recognition Theory and Applications, pp. 297–310. Springer, Berlin (1982)
10. Fletcher, L.A., Kasturi, R.: A robust algorithm for text string separation from mixed text/graphics images. IEEE Trans. Pattern Anal. Mach. Intell. **10**(6), 910–918 (1988)
11. Kamel, M., Zhao, A.: Extraction of binary character/graphics images from grayscale document images. CVGIP Graph. Models Image Process. **55**(3), 203–217 (1993)
12. Wendling, L., Tabbone, S.: A new way to detect arrows in line drawings. IEEE Trans. Pattern Anal. Mach. Intell. **26**, 935–941 (2004)
13. Bukhari, S.S., Azawi, A., Ali, M.I., Shafait, F., Breuel, T.M.: Document image segmentation using discriminative learning over connected components. In: Proceedings of the 9th IAPR International Workshop on Document Analysis Systems, pp. 183–190 (2010)
14. Okun, O., Drmann, D., Pietikainen, M.: Page segmentation and zone classification: the state of the art, DTIC document (1999)
15. Chiang, Y.-Y., Leyk, S., Knoblock, C.A.: A survey of digital map processing techniques. ACM Comput. Surv. **47**(1), 1 (2014)
16. Dori, D., Wenyin, L.: Automated CAD conversion with the machine drawing understanding system: concepts, algorithms, and performance. IEEE Trans. Syst. Man Cybern. Part A **29**(4), 411–416 (1999)
17. Ye, Q., Doermann, D.: Text detection and recognition in imagery: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **37**(7), 1480–1500 (2015)
18. Van Phan, T., Nakagawa, M.: Text/non-text classification in online handwritten documents with recurrent neural networks. In: 2014 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), pp. 23–28 (2014)
19. Karatzas, D., Antonacopoulos, A.: Colour text segmentation in web images based on human perception. Image Vis. Comput. **25**, 564–577 (2007)
20. Van Phan, T., Nakagawa, M.: Combination of global and local contexts for text/non-text classification in heterogeneous online handwritten documents. Pattern Recognit. **51**, 112–124 (2016)
21. Delaye, A., Liu, C.-L.: Contextual text/non-text stroke classification in online handwritten notes with conditional random fields. Pattern Recognit. **47**(3), 959–968 (2014)
22. Delaye, A., Liu, C.-L.: Multi-class segmentation of free-form online documents with tree conditional random fields. Int. J. Doc. Anal. Recognit. **17**(4), 313–329 (2014)
23. Degtyarenko, I., Radyvonenko, O., Bokhan, K., Khomenko, V.: Text/shape classifier for mobile applications with handwriting input. Int. J. Doc. Anal. Recognit. **19**(4), 369–379 (2016)
24. Bresler, M., Průša, D., Hlaváč, V.: Online recognition of sketched arrow-connected diagrams. Int. J. Doc. Anal. Recognit. **19**(3), 253–267 (2016)
25. Lucas, S.M.: ICDAR 2005 text locating competition results. In: Eighth International Conference on Document Analysis and Recognition, 2005. Proceedings, pp. 80–84 (2005)
26. Yao et al., C.: Incidental Scene Text Understanding: Recent Progresses on ICDAR 2015 Robust Reading Competition Challenge 4. arXiv Prepr. arXiv1511.09207 (2015)
27. Zhong, Y., Zhang, H., Jain, A.K.: Automatic caption localization in compressed video. IEEE Trans. Pattern Anal. Mach. Intell. **22**(4), 385–392 (2000)
28. Ye, Q., Huang, Q., Gao, W., Zhao, D.: Fast and robust text detection in images and video frames. Image Vis. Comput. **23**(6), 565–576 (2005)
29. Haritaoglu, I.: Scene text extraction and translation for handheld devices. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, vol. 2, pp. II-408 (2001)
30. Chen, X., Yang, J., Zhang, J., Waibel, A.: Automatic detection and recognition of signs from natural scenes. IEEE Trans. Image Process. **13**(1), 87–99 (2004)
31. Sermanet, P., Chintala, S., LeCun, Y.: Convolutional neural networks applied to house numbers digit classification. In: 2012 21st International Conference on Pattern Recognition (ICPR), pp. 3288–3291 (2012)
32. He, Z., Liu, J., Ma, H., Li, P.: A new automatic extraction method of container identity codes. IEEE Trans. Intell. Transp. Syst. **6**(1), 72–78 (2005)
33. Lienhart, R., Wernicke, A.: Localizing and segmenting text in images and videos. IEEE Trans. Circuits Syst. Video Technol. **12**(4), 256–268 (2002)
34. Lu, S., Chen, T., Tian, S., Lim, J.-H., Tan, C.-L.: Scene text extraction based on edges and support vector regression. Int. J. Doc. Anal. Recognit. **18**(2), 125–135 (2015)
35. Weinman, J.J., Learned-Miller, E., Hanson, A.R.: Scene text recognition using similarity and a lexicon with sparse belief propagation. IEEE Trans. Pattern Anal. Mach. Intell. **31**(10), 1733–1746 (2009)
36. Clark, P., Mirmehdi, M.: Recognising text in real scenes. Int. J. Doc. Anal. Recognit. **4**(4), 243–257 (2002)
37. Perantonis, S.J., Gatos, B., Maragos, V., Karkaletsis, V., Petasis, G.: Text area identification in web images. In: Vouros, G.A., Panayiotopoulos, T. (eds.) Methods and Applications of Artificial Intelligence, pp. 82–92. Springer, Berlin (2004)
38. Lopresti, D., Zhou, J.: Locating and recognizing text in WWW images. Inf. Retr. Boston **2**(2–3), 177–206 (2000)
39. Brown, M.K., Glinski, S.C., Schmult, B.C.: Web page analysis for voice browsing. In: Proceedings of the 1st International Workshop on Web Document Analysis (WDA'2001), pp. 59–61 (2001)
40. Penn, G., Hu, J., Luo, H., McDonald, R.: Flexible web document analysis for delivery to narrow-bandwidth devices. In: Document Analysis and Recognition (ICDAR), p. 1074 (2001)
41. Antonacopoulos, A., Hu, J.: Web Document Analysis: Challenges and Opportunities, vol. 55. World Scientific, Singapore (2003)
42. Lu, Z.: Detection of text regions from digital engineering drawings. IEEE Trans. Pattern Anal. Mach. Intell. **20**, 431–439 (1998)
43. Haralick, R.M., Queeney, D.: Understanding engineering drawings. Comput. Graph. Image Process. **20**(3), 244–258 (1982)
44. Le, V.P., Nayef, N., Visani, M., Ogier, J.-M., De Tran, C.: Text and non-text segmentation based on connected component features. In: 13th International Conference on Document Analysis and Recognition (ICDAR), pp. 1096–1100 (2015)
45. Jayadevan, R., Kolhe, S.R., Patil, P.M., Pal, U.: Automatic processing of handwritten bank cheque images: a survey. Int. J. Doc. Anal. Recognit. **15**(4), 267–296 (2012)
46. Hönes, F., Lichter, J.: Layout extraction of mixed mode documents. Mach. Vis. Appl. **7**(4), 237–246 (1994)
47. Handwritten Document Images. https://goo.gl/images/WyKY7e. Accessed: 08 Sept 2017
48. Chen, Y.-L., Hong, Z.-W., Chuang, C.-H.: A knowledge-based system for extracting text-lines from mixed and overlapping

text/graphics compound document images. Expert Syst. Appl. **39**(1), 494–507 (2012)
49. Zagoris, K., Chatzichristofis, S.A., Papamarkos, N.: Text localization using standard deviation analysis of structure elements and support vector machines. EURASIP J. Adv. Signal Process. **2011**(1), 1–12 (2011)
50. Emmanouilidis, C., Batsalas, C., Papamarkos, N.: Development and evaluation of text localization techniques based on structural texture features and neural classifiers. In: 10th International Conference on Document Analysis and Recognition, ICDAR'09, pp. 1270–1274 (2009)
51. Vu, H.N., Tran, T.A., Na, I.S., Kim, S.H.: Automatic extraction of text regions from document images by multilevel thresholding and k-means clustering. In: 2015 IEEE/ACIS 14th International Conference on Computer and Information Science (ICIS), pp. 329–334 (2015)
52. Sobottka, K., Kronenberg, H., Perroud, T., Bunke, H.: Text extraction from colored book and journal covers. Int. J. Doc. Anal. Recognit. **2**(4), 163–176 (2000)
53. Roy, P.P., Llados, J., Pal, U.: Text/graphics separation in color maps. In: Proceedings—International Conference on Computing: Theory and Applications, ICCTA 2007 (2007)
54. Chiang, Y.-Y., Knoblock, C.A.: Recognizing text in raster maps. Geoinformatica **19**(1), 1–27 (2015)
55. Oyedotun, O.K., Khashman, A.: Document segmentation using textural features summarization and feedforward neural network. Appl. Intell. **45**, 1–15 (2016)
56. Vil'kin, A.M., Safonov, I.V., Egorova, M.A.: Algorithm for segmentation of documents based on texture features. Pattern Recognit. Image Anal. **23**(1), 153–159 (2013)
57. Lin, M.W., Tapamo, J.-R., Ndovie, B.: A texture-based method for document segmentation and classification. S. Afr. Comput. J. **36**(1), 49–56 (2006)
58. Shih, F.Y., Chen, S.S.: Adaptive document block segmentation and classification. IEEE Trans. Syst. Man Cybern. Part B **26**(5), 797–802 (1996)
59. Park, H.C., Ok, S.Y., Cho, H.: Word extraction in text/graphic mixed image using 3-dimensional graph model. ICCPOL **99**, 171–176 (1999)
60. Antonacopoulos, A., Ritchings, R.T.: Representation and classification of complex-shaped printed regions using white tiles. In: Proceedings of the 3rd International Conference on Document Analysis and Recognition, August 14–16, pp. 1132–1135 (1995)
61. Bhowmik, S., Sarkar, R., Nasipuri, M.: Text and non-text separation in handwritten document images using local binary pattern operator. In: Proceedings of the First International Conference on Intelligent Computing and Communication, pp. 507–515 (2017)
62. Chen, Y.L., Wu, B.F.: A multi-plane approach for text segmentation of complex document images. Pattern Recognit. **42**, 1419–1444 (2009)
63. Strouthopoulos, C., Papamarkos, N., Atsalakis, A.E.: Text extraction in complex color documents. Pattern Recognit. **35**(8), 1743–1758 (2002)
64. Cao, R., Tan, C.L.: Text/graphics separation in maps. In: Blostein, D., Kwon, Y.B. (eds.) Graphics Recognition Algorithms and Applications, pp. 167–177. Springer, Berlin (2001)
65. Chiang, Y.Y., Knoblock, C.A.: An approach for recognizing text labels in raster maps. In: Proceedings—International Conference on Pattern Recognition (2010)
66. Velázquez, A., Levachkine, S.: Text/graphics separation and recognition in raster-scanned color cartographic maps. In: Lladós, J., Kwon, Y.-B. (eds.) Graphics Recognition. Recent Advances and Perspectives, pp. 63–74. Springer, Berlin (2003)
67. Baird, H.S., Jones, S.E., Fortune, S.J.: Image segmentation by shape-directed covers. In: Pattern Recognition, 1990. 10th International Conference on Proceedings, vol. 1, pp. 820–825 (1990)
68. Ha, J., Haralick, R.M., Phillips, I.T.: Recursive XY cut using bounding boxes of connected components. Proceedings of the Third International Conference on Document Analysis and Recognition **2**, 952–955 (1995)
69. Sun, H.-M.: Page segmentation for Manhattan and Non-Manhattan layout documents via selective CRLA. In: Eighth International Conference on Document Analysis and Recognition (ICDAR'05), pp. 116–120 (2005)
70. Agrawal, M., Doermann, D.: Voronoi++: a dynamic page segmentation approach based on voronoi and docstrum features. In: 2009 10th International Conference on Document Analysis and Recognition, pp. 1011–1015 (2009)
71. Ferilli, S., Basile, T., Esposito, F.: A histogram-based technique for automatic threshold assessment in a run length smoothing-based algorithm. In: Proceedings of the 9th IAPR International Workshop on Document Analysis Systems, pp. 349–356 (2010)
72. Pan, Y., Zhao, Q., Kamata, S.: Document layout analysis and reading order determination for a reading robot. In: TENCON 2010–2010 IEEE Region 10 Conference, pp. 1607–1612 (2010)
73. Jain, A.K., Yu, B.: Document representation and its application to page decomposition. IEEE Trans. Pattern Anal. Mach. Intell. **20**(3), 294–308 (1998)
74. Smith, R.W.: Hybrid page layout analysis via tab-stop detection. In: 2009 10th International Conference on Document Analysis and Recognition, pp. 241–245 (2009)
75. Chen, K., Yin, F., Liu, C.L.: Hybrid page segmentation with efficient whitespace rectangles extraction and grouping. In: 2013 12th International Conference on Document Analysis and Recognition, pp. 958–962 (2013)
76. Fan, K.-C., Wang, L.-S.: Classification of document blocks using density feature and connectivity histogram. Pattern Recognit. Lett. **16**(9), 955–962 (1995)
77. Antonacopoulos, A., Ritchings, R.T.: Representation and classification of complex-shaped printed regions using white tiles. Proceedings of the Third International Conference on Document Analysis and Recognition **2**, 1132–1135 (1995)
78. Pavlidis, T., Zhou, J.: Page segmentation and classification. CVGIP Graph. Models Image Process. **54**(6), 484–496 (1992)
79. Drivas, D., Amin, A.: Page segmentation and classification utilising a bottom-up approach. Proceedings of the Third International Conference on Document Analysis and Recognition **2**, 610–614 (1995)
80. Chowdhury, S.P., Mandal, S., Das, A.K., Chanda, B.: Segmentation of text and graphics from document images. In: Ninth International Conference on Document Analysis and Recognition, ICDAR 2007, vol. 2, pp. 619–623 (2007)
81. Bloomberg, D.S.: Multiresolution morphological approach to document image analysis. In: Proceedings of the International Conference on Document Analysis and Recognition, Saint-Malo, France (1991)
82. Bukhari, S.S., Shafait, F., Breuel, T.M.: Improved document image segmentation algorithm using multiresolution morphology. In: IS&T/SPIE Electronic Imaging, pp. 78740D–78740D (2011)
83. Ablameyko, S.V., Uchida, S.: Recognition of engineering drawing entities: review of approaches. Int. J. Image Graph. **7**(4), 709–733 (2007)
84. Tombre, K., Tabbone, S., Pélissier, L., Lamiroy, B., Dosch, P.: Text/graphics separation revisited. In: Lopresti, D., Hu, J., Kashi, R. (eds.) Document Analysis Systems V, pp. 200–211. Springer, Berlin (2002)
85. Lin, S.-C., Ting, C.-K.: A new approach for detection of dimensions set in mechanical drawings. Pattern Recognit. Lett. **18**(4), 367–373 (1997)
86. Ah-Soon, C., Tombre, K.: Variations on the analysis of architectural drawings. Proceedings of the Fourth International Conference on Document Analysis and Recognition **1**, 347–351 (1997)

87. Lai, C.P., Kasturi, R.: Detection of dimension sets in engineering drawings. IEEE Trans. Pattern Anal. Mach. Intell. **16**(8), 848–855 (1994)
88. Do, T.-H., Tabbone, S., Ramos-Terrades, O.: Text/graphic separation using a sparse representation with multi-learned dictionaries. In: 2012 21st International Conference on Pattern Recognition (ICPR), pp. 689–692 (2012)
89. De, P., Mandal, S., Bhowmick, P., Das, A.: ASKME: adaptive sampling with knowledge-driven vectorization of mechanical engineering drawings. Int. J. Doc. Anal. Recognit. **19**(1), 11–29 (2016)
90. Favreau, J.-D., Lafarge, F., Bousseau, A.: Fidelity vs. simplicity: a global approach to line drawing vectorization. ACM Trans. Graph. **35**, 120 (2016)
91. Dori, D., Velkovitch, Y.: Segmentation and recognition of dimensioning text from engineering drawings. Comput. Vis. Image Underst. **69**(2), 196–201 (1998)
92. Casey, R., Ferguson, D., Mohiuddin, K., Walach, E.: Intelligent forms processing system. Mach. Vis. Appl. **5**(3), 143–155 (1992)
93. Yu, B., Jain, A.K.: A generic system for form dropout. IEEE Trans. Pattern Anal. Mach. Intell. **18**(11), 1127–1134 (1996)
94. Wang, D., Srihari, S.N.: Analysis of form images. Int. J. Pattern Recognit. Artif. Intell. **8**(5), 1031–1052 (1994)
95. Hori, O., Doermann, D.S.: Robust table-form structure analysis based on box-driven reasoning. Proceedings of the Third International Conference on Document Analysis and Recognition **1**, 218–221 (1995)
96. Dzuba, G., Filatov, A., Gershuny, D., Kil, I., Nikitin, V.: Check amount recognition based on the cross validation of courtesy and legal amount fields. Int. J. Pattern Recognit. Artif. Intell. **11**(4), 639–655 (1997)
97. Heutte, L., et al.: Multi-bank check recognition system: consideration on the numeral amount recognition module. Int. J. Pattern Recognit. Artif. Intell. **11**(4), 595–618 (1997)
98. Knerr, S., Anisimov, V., Baret, O., Gorski, N., Price, D., Simon, J.-C.: The A2iA intercheque system: courtesy amount and legal amount recognition for French checks. Int. J. Pattern Recognit. Artif. Intell. **11**(4), 505–548 (1997)
99. Govindaraju, V., Srihari, S.N.: Separating handwritten text from interfering strokes. In: From Pixels to Features. III Front. Handwritten Recognition, pp. 17–28 (1992)
100. Dimauro, G., Impedovo, S., Pirlo, G., Salzo, A.: Removing underlines from handwritten text: an experimental investigation. In: Impedovo, S., Downton, A.C. (eds.) Progress in Handwriting Recognition, pp. 497–501. World Scientific (1997)
101. Akiyama, T., Hagita, N.: Automated entry system for printed documents. Pattern Recognit. **23**(11), 1141–1154 (1990)
102. Liu, K., Suen, C.Y., Cheriet, M., Said, J.N., Nadal, C., Tang, Y.Y.: Automatic extraction of baselines and data from check images. Int. J. Pattern Recognit. Artif. Intell. **11**(4), 675–697 (1997)
103. Hase, H., Shinokawa, T., Yoneda, M., Suen, C.Y.: Character string extraction from color documents. Pattern Recognit. **34**(7), 1349–1365 (2001)
104. Clavelli, A., Karatzas, D.: Text segmentation in colour posters from the Spanish civil war era. In: 10th International Conference on Document Analysis and Recognition, ICDAR'09, pp. 181–185 (2009)
105. Li, L., Nagy, G., Samal, A., Seth, S., Xu, Y.: Integrated text and line-art extraction from a topographic map. Int. J. Doc. Anal. Recognit. **2**, 177–185 (2000)
106. Dhar, D.B., Chanda, B.: Extraction and recognition of geographical features from paper maps. Int. J. Doc. Anal. Recognit. **8**(4), 232–245 (2006)
107. Cordeiro, A., Pina, P.: Colour map object separation. In: Proceedings of the ISPRS Mid-Term Symposium 2006, Remote Sensing: From Pixels to Processes, pp. 243–247 (2006)
108. San, L.M., Yatim, S.M., Sheriff, N.A.M., Isrozaidi, N.: Extracting contour lines from scanned topographic maps. In: International Conference on Computer Graphics, Imaging and Visualization, CGIV 2004. Proceedings, pp. 187–192 (2004)
109. Chen, Y., Wang, R., Qian, J.: Extracting contour lines from common-conditioned topographic maps. IEEE Trans. Geosci. Remote Sens. **44**(4), 1048–1057 (2006)
110. Leyk, S.: Segmentation of colour layers in historical maps based on hierarchical colour sampling. In: International Workshop on Graphics Recognition, pp. 231–241 (2009)
111. Kasturi, R., Bow, S.T., El-Masri, W., Shah, J., Gattiker, J.R., Mokate, U.B.: A system for interpretation of line drawings. IEEE Trans. Pattern Anal. Mach. Intell. **12**(10), 978–992 (1990)
112. Raveaux, R., Eugen, B., Locteau, H., Adam, S., Héroux, P., Trupin, E.: A graph classification approach using a multi-objective genetic algorithm application to symbol recognition. In: International Workshop on Graph-Based Representations in Pattern Recognition, pp. 361–370 (2007)
113. Raveaux, R., Burie, J.-C., Ogier, J.-M.: Object extraction from colour cadastral maps. In: The Eighth IAPR International Workshop on Document Analysis Systems, DAS'08, pp. 506–514 (2008)
114. Pezeshk, A., Tutwiler, R.L.: Automatic feature extraction and text recognition from scanned topographic maps. IEEE Trans. Geosci. Remote Sens. **49**, 5047–5063 (2011)
115. Chiang, Y.-Y., Knoblock, C.A.: Recognition of multi-oriented, multi-sized, and curved text. In: International Conference on Document Analysis and Recognition, pp. 1399–1403 (2011)
116. Chiang, Y.-Y., Knoblock, C.A.: A general approach for extracting road vector data from raster maps. Int. J. Doc. Anal. Recognit. **16**(1), 55–81 (2013)
117. Nazari, N.H., Tan, T., Chiang, Y.-Y.: Integrating text recognition for overlapping text detection in maps. Electron. Imaging **2016**(17), 1–8 (2016)
118. Kavitha, A.S., Shivakumara, P., Kumar, G.H., Lu, T.: Text segmentation in degraded historical document images. Egypt. Inform. J. **17**, 189–197 (2016)
119. Shivakumara, P., Sreedhar, R.P., Phan, T.Q., Lu, S., Tan, C.L.: Multioriented video scene text detection through Bayesian classification and boundary growing. IEEE Trans. Circuits Syst. Video Technol. **22**(8), 1227–1235 (2012)
120. Cohen, R., Asi, A., Kedem, K., El-Sana, J., Dinstein, I.: Robust text and drawing segmentation algorithm for historical documents. In: Proceedings of the 2nd International Workshop on Historical Document Imaging and Processing, pp. 110–117 (2013)
121. ABBYY Fine reader. https://finereaderonline.com/en-us. Accessed Oct 2017
122. Mendelson, E.: ABBYY finereader professional 9.0. PC Mag. (2008)
123. OCRopus. https://en.wikipedia.org/wiki/OCRopus. Accessed Oct 2017
124. Breuel, T.M.: The OCRopus open source OCR system. Electron. Imaging **2008**, 68150F–68150F (2008)
125. Smith, R.W.: History of the Tesseract OCR engine: what worked and what didn't. In: IS&T/SPIE Electronic Imaging, p. 865802 (2013)
126. Tesseract-OCR. https://github.com/tesseract-ocr/tesseract/wiki. Accessed Oct 2017
127. Lichman, M.: UCI Machine Learning Repository, University of California, School of Information and Computer Science, Irvine, CA, (2013). http://archive.ics.uci.edu/ml. Accessed Oct 2017
128. UW-I English Document Image Database. http://isis-data.science.uva.nl/events/dlia//datasets/uwash1.html. Accessed Oct 2017

129. UW-II English/Japanese Document Image Database. http://isis-data.science.uva.nl/events/dlia//datasets/uwash2.html. Accessed Oct 2017
130. UW-III English/Technical Document Image Database. http://isis-data.science.uva.nl/events/dlia//datasets/uwash3.html. Accessed Oct 2017
131. Antonacopoulos, A., Bridson, D., Papadopoulos, C., Pletschacher, S.: A realistic dataset for performance evaluation of document layout analysis. In: 10th International Conference on Document Analysis and Recognition, ICDAR'09, pp. 296–300 (2009)
132. ICDAR 2009 Dataset. http://www.primaresearch.org/dataset/index.php. Accessed Oct 2017
133. The MediaTeam document database II. http://www.mediateam.oulu.fi/downloads/MTDB/. Accessed Oct 2017
134. Islamic Heritage Project (IHP) Collection. http://ocp.hul.harvard.edu/ihp/. Accessed Oct 2017
135. UNLV Database. http://www.isri.unlv.edu/ISRI/OCRtk. Accessed Oct 2017
136. 6-Inch Historical Ordnance Survey Maps of the United Kingdom (UK). http://maps.nls.uk/os/6inch-england-and-wales/info2.html. Accessed Oct 2017
137. Wang, D., Srihari, S.N.: Classification of newspaper image blocks using texture analysis. Comput. Vis. Graph. Image Process. **47**(3), 327–352 (1989)
138. Liu, W.Y., Dori, D.: A proposed scheme for performance evaluation of graphics/text separation algorithms. Graph. Recognit. Algorithms Syst. **1389**, 359–371 (1998)
139. Shafait, F.: Camera-Based Document Analysis and Recognition. Springer, Berlin (2007)
140. Kasar, T., Kumar, J., Ramakrishnan, A.G.: Font and background color independent text binarization. In: Second International Workshop on Camera-based Document Analysis and Recognition, pp. 3–9 (2007)
141. Drivas, D., Amin, A.: Page Segmentation and Classification Utilising Bottom-Up Approach. In: Proceedings of the 3rd International Conference on Document Analysis and Recognition, Aug 14–16, pp. 0–4 (1995)
142. Tran, T.A., Oh, K., Na, I.-S., Lee, G.-S., Yang, H.-J., Kim, S.-H.: A robust system for document layout analysis using multilevel homogeneity structure. Expert Syst. Appl. **85**, 99–113 (2017)
143. Tran, T.A., Na, I.S., Kim, S.H.: Page segmentation using minimum homogeneity algorithm and adaptive mathematical morphology. Int. J. Doc. Anal. Recognit. **19**(3), 191–209 (2016)
144. Maderlechner, G.: Symbolic subtraction of fixed formatted graphics and text from filled in forms. In: Proceedings of the IAPR Workshop on Machine Vision and Applications, Tokyo, November 1990, pp. 457–459 (1990)
145. Doermann, D.S., Rosenfeld, A.: The interpretation and reconstruction of interfering strokes. Front. Handwrit. Recognit. **3**, 41–50(1993)
146. Tang, Y.Y., Suen, C.Y., De Yan, C., Cheriet, M.: Financial document processing based on staff line and description language. IEEE Trans. Syst. Man. Cybern. **25**(5), 738–754 (1995)
147. Moghaddam, R.F., Cheriet, M.: A multi-scale framework for adaptive binarization of degraded document images. Pattern Recognit. **43**(6), 2186–2198 (2010)
148. Hedjam, R., Moghaddam, R.F., Cheriet, M.: A spatially adaptive statistical method for the binarization of historical manuscripts and degraded document images. Pattern Recognit. **44**(9), 2184–2196 (2011)
149. Howe, N.R.: Document binarization with automatic parameter tuning. Int. J. Doc. Anal. Recognit. **16**(3), 247–258 (2013)
150. Mitianoudis, N., Papamarkos, N.: Document image binarization using local features and Gaussian mixture modeling. Image Vis. Comput. **38**, 33–51 (2015)
151. Mandal, S., Das, S., Agarwal, A., Chanda, B.: Binarization of degraded handwritten documents based on morphological contrast intensification. In: Third International Conference on Image Information Processing (ICIIP), pp. 73–78 (2015)
152. Adak, C., Maitra, P., Chaudhuri, B.B., Blumenstein, M.: Binarization of old halftone text documents. In: TENCON 2015–2015 IEEE Region 10 Conference, pp. 1–5 (2015)
153. Das, B., Bhowmik , S., Saha, A., sarkar, R.: An adaptive foreground-background separation method for effective binarization of document images. In: 8th International Conference on Soft Computing and Pattern Recognition (2016)
154. Kang, L., Doermann, D.: Template based segmentation of touching components in handwritten text lines. In: International Conference on Document Analysis and Recognition (ICDAR), pp. 569–573 (2011)
155. Liu, C.-L., Sako, H., Fujisawa, H.: Effects of classifier structures and training regimes on integrated segmentation and recognition of handwritten numeral strings. IEEE Trans. Pattern Anal. Mach. Intell. **26**(11), 1395–1407 (2004)
156. Wang, Y., Liu, X., Jia, Y.: Statistical modeling and learning for recognition-based handwritten numeral string segmentation. In: 10th International Conference on Document Analysis and Recognition, ICDAR'09, pp. 421–425 (2009)
157. Papavassiliou, V., Stafylakis, T., Katsouros, V., Carayannis, G.: Handwritten document image segmentation into text lines and words. Pattern Recognit. **43**(1), 369–377 (2010)
158. Wshah, S., Shi, Z., Govindaraju, V.: Segmentation of Arabic handwriting based on both contour and skeleton segmentation. In: 10th International Conference on Document Analysis and Recognition, ICDAR'09, pp. 793–797 (2009)