2015 13th International Conference on Document Analysis and Recognition (ICDAR)

# Text-Graphics Separation to Detect Logo and Stamp from Color Document Images: A Spectral Approach

Amit Vijay Nandedkar
School of Information Technology
IIT Kharagpur, India
Email: amitnandedkar@sit.iitkgp.ernet.in

Jayanta Mukhopadhyay
Dept. of Computer Science & Engg.
IIT Kharagpur, India
Email: jay@cse.iitkgp.ernet.in

Shamik Sural
School of Information Technology
IIT Kharagpur, India
Email: shamik@sit.iitkgp.ernet.in

***Abstract*—Text and graphics separation is an important task in the field of document image processing. This work aims at detecting graphics such as logos and stamps in a scanned document image. A novel spectral filtering based text-graphics separation algorithm (SFTGS) is presented here. The property of text that it is the major source of high spatial frequency components in a document image, is exploited in this algorithm. Accordingly high frequency filtering is used to separate the text symbols. This is followed by a segmentation process for delineating residual text and the graphics. The main advantage of SFTGS is that it works in a single pass, and can discriminate graphics and text without supervised training. Subsequently, the graphics segments are further categorized into two different classes, namely logos and stamps. In this case, we assume that these are the two classes of graphical objects present in the documents. The technique is evaluated using publicly available document dataset consisting of graphics as stamps and logos. The result is compared with existing approaches reported in the literature, and it is found that the proposed method performs superior to them. An overall performance of 89.1% recall and 96.9% precision is obtained for SFTGS.**

## I. Introduction

Presently digital document images are so widely used that innovative techniques are required to organize these document image databases according to their contents. A scanned document may consist of graphics such as computer generated artwork, rendered images of a graphics system, digital images captured by a camera, logos, stamps, etc. Suitable techniques to extract graphics and text are required to interpret the content of a document. Such techniques are also helpful for an optical character recognition (OCR) system for enhancing its proficiency. Typically separation of logos and stamps in document images (Figs. 1(a) and (b)) has drawn attention of several researchers.

Logos and stamps are commonly used in most of the official documents to represent their legitimacy. Hence, digitization and processing of the official documents requires extraction of logos and stamps from the scanned images. In the past, several efforts to detect logos and stamps are reported in [1]–[8]. A few techniques were proposed using block segmentation and classification to partition a document into text, horizontal or vertical lines, and pictures [8]. Wahl et al. [9] suggested a run-length smoothing algorithm for classifying regions into text and images. Fisher et al. [10] developed a rule-based algorithm on the basis of statistical properties of connected components like height, aspect ratio, density, perimeter, and width, etc., for document image segmentation. Wang and Srihari [11] used the run-length matrix for newspaper classification. Fletcher and Kasturi [12] adopted Hough transform to group connected components into a logical character string to facilitate its discrimination from graphics. Shi and Chen [13] proposed adaptive document block segmentation and classification. Jain et al. [14] suggested a neural network based system for page segmentation using texture features.


Fig. 1. (a) Stamp and (b) Logo examples of the document image dataset [1].

These techniques [9]–[14] are applicable to gray scale documents, and suffer from misclassification when the font size of characters or the scanning resolution is varied. Cote et al. [15] suggested document pixel classification based on texture sparseness criteria using support vector machine (SVM). Hoang et al. [16] proposed text extraction from graphical documents using morphological component analysis (MCA). Different techniques are proposed in the past for segmentation of stamps from document images either based on color or geometric features. Micenkova et al. [1] proposed a stamp detection technique based on clustering in $YC_bC_r$ color space and geometrical features. The technique is independent of shape and form of stamps, but it is limited to only stamps that are not black in color. Ahmed et al. [2] presented a stamp detection method using key point descriptors and geometric features that needs training to detect colored as well as black stamps. However, it reported low average precision and recall rates, and its inability to detect severely overlapped stamps.

This work proposes a novel *spectral filtering based text-graphics separation algorithm (SFTGS)* that works on an unsupervised learning paradigm. It aims to separate the color document contents into two groups namely, textual region and graphical region consisting of logos and stamps, by using their distinctive spectral characteristics. Further, it processes

978-1-4799-1805-8/15/$31.00 ©2015 IEEE

the identified graphical region to detect stamp and logo. The main contribution of this work is to separate text and graphics regions from document image without performing complex operations, such as text-size dependent block segmentation [9]–[13], projection profile analysis [17], the supervised training [14, 15] for classification, etc. It uses the simple fact that spectral characteristics of textual regions differ from most of the graphical regions. It is a non-filter bank based approach [18] followed by a mean-shift segmentation [19] processing to separate textual and graphical regions in the document image. The experimental evaluation of this system is performed using a publicly available scanned document image database [1] with well-defined ground truths [1]. These document images comprise of complex graphical regions such as stamps of various colors, shapes and multi-colored, multi-shaped logos as depicted in Figs. 1(a) and (b), respectively. The proposed SFTGS algorithm can detect and discriminate graphical objects such as logos and stamps irrespective of their shapes and colors.

## II. SPECTRAL FILTERING BASED TEXT-GRAPHICS SEPARATION SYSTEM

The proposed algorithm is depicted in Fig. 2. The SFTGS algorithm is broadly divided into three steps: 1) Suppression of text contents by frequency selective image formation, 2) Identification and separation of graphical and textual regions, and 3) Logo and stamp separation from the detected graphical regions. It is quite evident that the text symbols contribute to high spatial frequencies in a document image. If the high-frequency components get filtered, it suppresses the visual importance of textual regions more compared to most of the graphical regions. The SFTGS algorithm utilizes filtering operation based on the concept that the human vision system performs frequency selective operations [18]. The frequency selective filtered document image contains pixels contributing to relatively low spatial frequencies. This filtered image is further processed to detect graphics using a sequence of operations, such as median filtering, blurring, quantization, and finally mean shift segmentation [19]. The details are discussed in following subsections.

### A. Frequency selective image formation

The primary goal in this step is to suppress text symbols in the document. The notion is to select pixels in the area with relatively lower spatial frequency characteristics into the layer $F_L$. The spectral selection process removes the pixels in regions contributing high spatial frequencies. The gray scale image $X$ of the input color document image $I$ (of size $N \times M$) is used in the layer formation process because it consists of all structural and textural details of an image. The gray component is filtered first by a 2D Gaussian filter. The high pass filtered image $Y$ is:

$$Y = |X - G(X)| \tag{1}$$

where $Y$ is the high pass filtered image, $G(X)$ is Gaussian filtered image of $X$.

[1]Authors would like to thank Mr. Soumyadeep Dey, Research Scholar, IIT Kharagpur, for providing the groundtruths of logos, stamps, text components of the StaVer [1] dataset using the groundtruth-tool developed by him.

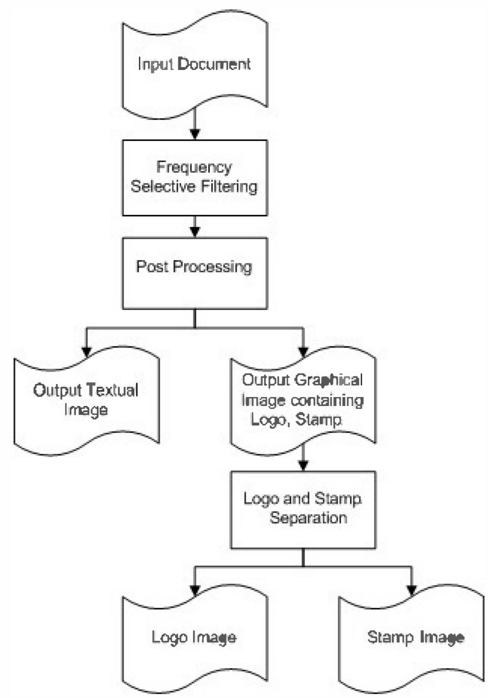



Fig. 2. The SFTGS to detect logo and stamp.

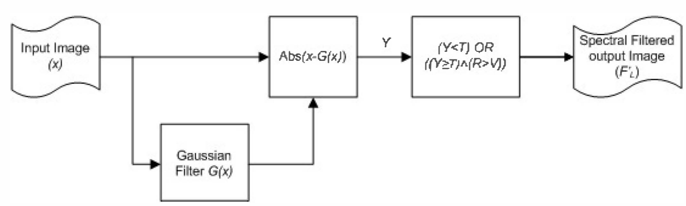


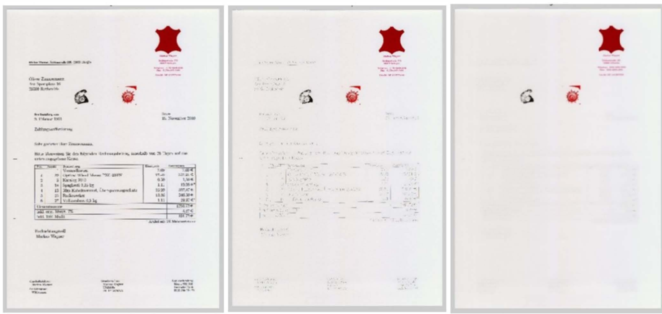

Fig. 3. Frequency selective image layer formation method for document images.

A simple frequency selective filtering is:

$$F_L(i,j) = \begin{cases} I(i,j), \ If \ Y(i,j) < T \\ B_k \ \ , \ Otherwise \end{cases} \tag{2}$$

where $1 \leq i \leq N$, $1 \leq j \leq M$. $T$ is the frequency selective threshold value used in document image layer formation and $I(i,j)$ indicates color information at coordinates $(i,j)$. $B_k$ is background color. It is an average color present at all background pixel locations of $I$ identified by binary thresholding of $X$.

It may be noted that only frequency selective operation is not suitable to detect chromatic graphical objects such as stamps and logos as the filtering operation may lose some high frequency information from graphical objects as well. So, we choose to retain pixels with high spatial frequency and high chromaticity components into the frequency selective filtered image and call it as $F_L'$ as shown in Fig. 3. This process helps to retain maximum number of pixels from stamp and logo regions and suppresses the textual symbols in the document. The intermediate color-layered output image retaining the candidate graphical objects with relatively low spatial frequencies along with the preservation of high chromatic pixel information in



(a) Input Document (b) Filtering using $F'_L$ (c) Median and Gaussian Filtered

Fig. 4. Example of document image and its filtered version by frequency selective method, and Median, Gaussian filtered image.

the region of graphical objects is:

$$F'_L(i,j) = \begin{cases} I(i,j),\ If\ (Y(i,j) < T)\ or\ ((Y(i,j) \geq T) \wedge (R(i,j) > V)) \\ B_k\ \ ,\ Otherwise \end{cases} \quad (3)$$

where $1 \leq i \leq N$, $1 \leq j \leq M$.
$T$ is the frequency selective threshold value used in document image layer formation.
$I(i,j)$ is color information at coordinates $(i,j)$.
$R$ is the chromaticity matrix of the input document image $I$ in $YC_bC_r$ color space, $R(i,j) = \sqrt{C_b(i,j)^2 + C_r(i,j)^2)}$ [1], and $B_k$ is background color.

It is empirically observed that Gaussian filter with $\sigma = 6$, frequency selective threshold $T = 35$, and the chromaticity threshold $V = 190$, perform well for text and graphics separation. The removal of these pixels with corresponding high magnitude values in $Y$ results in the suppression of visual perception of textual symbols in the document image. The loss of visual information is more stringent for text symbols than the graphical regions. This process helps to suppress text symbols and locate prominent graphical areas present in the document image.

Fig. 4(b) shows an example of filtered document image by frequency selective method using $F'_L$ of the document image shown in Fig. 4(a). The visual importance of the textual symbols is drastically reduced compared to graphical regions present in the document image. The proposed frequency selective filtering using $F'_L$ is promising as it can retain pixels of graphical regions with high spatial frequency and high chromaticity properties.

## B. Post Processing: Identification and separation of graphical and textual regions

Next, identification and separation of graphical and text regions are performed on the filtered image. The filtered image is first median filtered, and then, blurred using Gaussian filter to reduce the visual perception of textual residues as well as to fuse regions present in graphical areas. Finally, it is color quantized and segmented using mean-shift segmentation. Median filtering helps to suppress noise due to text-residuals and fills the missing pixels in the graphical regions using dominant color of the neighborhood. Gaussian filter helps to suppress text residuals further. The output of Gaussian blurred image is color quantized to minimize the effect of color variations, and to facilitate the color segmentation operation. Color quantization to 16 colors is empirically found to provide good performance. Let us denote this post-processed image as $P$. Fig. 4(c) shows example of median and Gaussian filtered image of the frequency selective filtered document shown in Fig. 4(b). It may be noted that pixels belonging to most of the text symbols are removed to a great extent.

In the next step, the mean shift segmentation [19] is carried out on $P$ to identify the candidate graphical regions of the document image. Let the segmented image be $S$, and $r$ be the total number of segments present in $S$. Smaller segments are ignored at this stage, and a threshold value on the sizes of segments is used for this purpose. The largest segment is considered to be background $B$. The rest of the segmented regions form the set of candidates for graphical regions. Let this set be denoted by $F$:

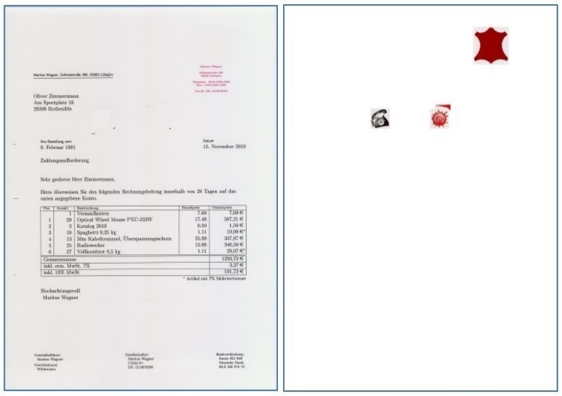
$$F = \{f_1, f_2, ......, f_{r-1}\} \quad (4)$$

The segmented document image is $S = F \cup \{B\}$. The foreground graphical regions are used to create the binary mask $G_{\text{MASK}}$. In the next step, for each foreground region present in $F$, a separate processing on $G_{\text{MASK}}$ is performed to remove textual residuals. $G_{\text{MASK}}$ is refined using bounding boxes formed around foreground regions and by applying geometrical constraints on bounding box area, height, and width. Any bounding box with area less than $N \times M \times T_1$ is removed from $G_{\text{MASK}}$, where $N$ is width, $M$ is height of the document image, and $T_1$ is a constant. Apart from this any bounding box near the logo centers (discussed in section II.C), with width $W_i$ and height $H_i$ such that $W_i \leq T_2 \times N$ and $H_i \leq T_3 \times M$ is removed from $G_{\text{MASK}}$. It is observed that such bounding boxes contain chromatic textual residuals near logo. The values of $T_1 = 2.31 \times 10^{-5}$, $T_2 = 0.203$, and $T_3 = 0.013$ are empirically chosen.

The refined $G_{\text{MASK}}$ is used to form bounding boxes for candidate graphical regions, and it is used to extract graphical regions from document image $I$. Let the document image containing only graphics segments, obtained by the above process, be denoted as $D_{\text{GRAPHICAL}}$. Similarly, the textual $D_{\text{TEXTUAL}}$ document image is created by considering complemented binary mask $\bar{G}_{\text{MASK}}$. Figs. 5(a) and (b) depict text-graphics separated images of the document shown in Fig. 4(a). The separation of logo and stamp from the graphics is discussed in the following section.

## C. Logo and Stamp Separation

The logo provides a perceptual identity of the document source to the viewer. It is considered that the graphical output produced by the SFTGS algorithm consists of logo and stamp regions, and logos are placed only at some predefined range of positions in the document image. In the logo and stamp separation process, we have considered the possible logo-positions knowledge in the document image dataset. This information is very useful in logo detection as mentioned in [5]. In the proposed SFTGS algorithm, an unsupervised $k$-means clustering [20] is used to cluster the logo-centroid $c(x,y)$ and logo spread $w$ from the ground truths. Here, $w$ is the maximum spread of logo from centroid either in horizontal or vertical direction. The output of $k$-means clustering generates $k$ number of cluster centers as $C(x_k, y_k, w_k)$ in 3-dimensional normalized space. The SFTGS algorithm



(a) (b)

Fig. 5. Text-graphics separated document images of Fig. 4(a).

clusters logo positions from ground truth into $k$=5 clusters. This number of clusters is selected empirically by observing the performance of this technique for the StaVer dataset [1]. However, prior knowledge of the number of different logos, and their possible placements in documents may provide an estimate of $k$. Fig. 6 shows computed logo cluster centers in (width, height, spread) space in the given set of documents.

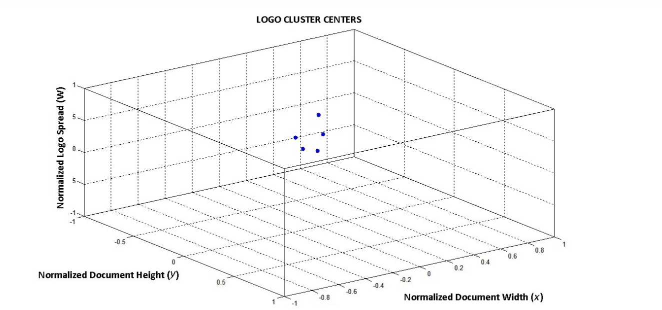
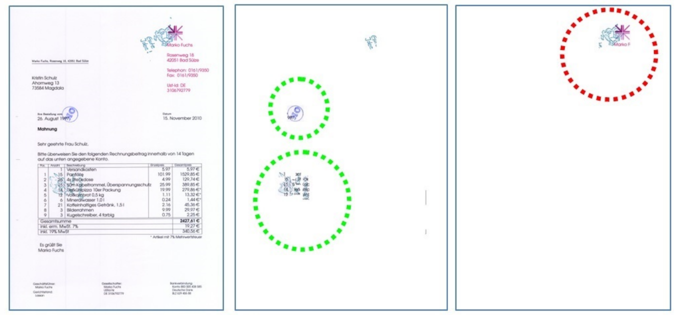


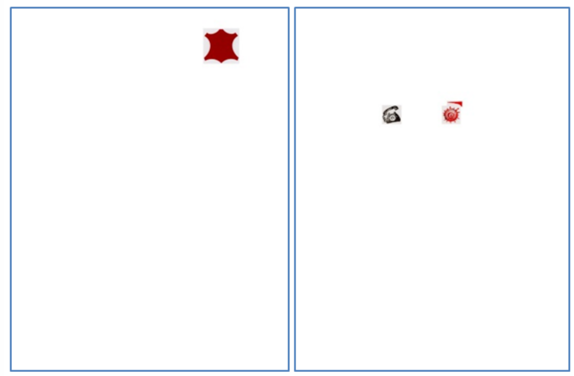

Fig. 6. Logo Cluster centers.

The SFTGS algorithm can accommodate a new dataset with variation in logo positions by re-clustering logo groundtruth data. In the surrounding of each candidate logo center $(x_k, y_k)$ , the number of foreground pixels $h_k$ is computed within a window of size $(2w_k \times 2w_k)$ in $D_{\text{GRAPHICAL}}$ image. A candidate logo with $h_k$ greater than a threshold value is accepted as a logo. Otherwise it is rejected and declared as a stamp. Note that, the SFTGS algorithm is not able to separate the stamps that are fully overlapped with logo region. Fig. 8 shows an example of logo and stamp separated images obtained from the processed image shown in Fig. 5(b). The failure to detect severely overlapped stamp with text and logo region is demonstrated in Fig. 7(a). It may be noted that detection of overlapped stamp and logo regions is a challenging problem. Some of the previous efforts [2] also acknowledge this limitation. It is observed that the SFTGS algorithm efficiently detects stamps that are severely overlapped with textual regions as shown in Fig. 7(b) using green marked circles with reference to document image shown in Fig. 7(a). However, the SFTGS algorithm is not able to detect the stamp which is completely overlapped with logo. The SFTGS algorithm considers such stamps as a part of logo-



(a) Input Document (b) Detected Stamp Image (c) Detected Logo Image

Fig. 7. Example of the detected stamps and logo in severely overlapping condition by SFTGS. Here green marking indicates the detected severely overlapped stamps and red marking indicates detected logo overlapped by stamp.



(a) Logo region image (b) Stamp region image

Fig. 8. An example of logo and stamp separated graphical output images.

detected image as shown in Fig. 7(c) using red marking.

## III. Experimental Results

The proposed SFTGS algorithm is evaluated using publicly available stamp verification dataset StaVer [1] with ground truths identifying the text, logo and stamp segments in each image. The dataset consists of 400 document images with stamps of various colors as well as multicolor logos available in both 200 dpi and 300 dpi resolutions. The aim of the experiment is to evaluate the performance of the SFTGS algorithm in terms of recall and precision metrics in identifying text and graphics as well as logo and stamps. Its performance is also compared with the techniques reported by Micenkova et al. [1] and Ahmed et al. [2] to detect stamps in a document. The performance metrics recall and precision are defined as follows:

$$Recall = \frac{True\,positives}{True\,positives + False\,negatives} \times 100\% \quad (5)$$

$$Precision = \frac{True\,positives}{True\,positives + False\,Positives} \times 100\% \quad (6)$$

Average recall and precision of Micenkova et al. [1], oriented FAST and rotated BRIEF (ORB) method proposed by Ahmed et al. [2], and the proposed SFTGS algorithm on StaVer [1] dataset are presented in Table 1. Note that the stamp detection performance of the SFTGS algorithm is better than Micenkova et al. [1] and Ahmed et al. [2] which aim at

TABLE I. PERFORMANCE EVALUATION OF THE PROPOSED SFTGS

| | Detected Region | Micenkova et al. [1] | | Ahmed et al.[2] | | Proposed SFTGS | |
|---|---|---|---|---|---|---|---|
| | | Recall | Precision | Recall | Precision | Recall | Precision |
| 300 dpi | Text | – | – | – | – | 99.6% | 98.9% |
| | Graphics (Logo,Stamp) | – | – | – | – | 87.3% | 95.9% |
| | Logo | – | – | – | – | 87.8% | 97.6% |
| | Stamp | 83.4% | 83.8% | – | – | 84.4% | 94.1% |
| 200 dpi | Text | – | – | – | – | 99.6% | 98.6% |
| | Graphics (Logo,Stamp) | – | – | – | – | 84.6% | 96.9% |
| | Logo | – | – | – | – | 86.1% | 97.9% |
| | Stamp | 82.7% | 82.8% | 60.3% | 66.3% | 83.0% | 95.1% |
| | | | | | Overall | 89.1% | 96.9% |

detecting stamps only. A significant improvement in precision is also observed for the SFTGS algorithm compared to [1]. The results on 200 dpi resolution dataset are approximately close to the results on 300 dpi resolution dataset. Hence, the SFTGS algorithm also works for document images with lower resolutions. Performance of the SFTGS algorithm on text-graphics separation is shown in Table 1, and it is found to be satisfactory in terms of recall and precision. The SFTGS algorithm has been implemented using Matlab on a windows 7 system with 64-bit $i5$ processor @ 3.10 GHz and 4 GB RAM. The average execution time for processing a document image of size 2472×3489 (300 dpi) is 131.1 seconds and for a document image of size 1632×2302 (200 dpi) is 41.1 seconds.

## IV. CONCLUSION

The paper presents a new spectral filtering approach to detect graphics such as logos and stamps in a color document image. The proposed SFTGS algorithm provides a mechanism to detect logo and stamps separately from a scanned document. The proposed system explores the spectral characteristics of document image to separate text from graphical regions, and is found to be efficient. It is a single pass algorithm, and works on an unsupervised learning paradigm. This property makes it suitable in applications such as OCR, logo and stamp based document verification and document image retrieval. The algorithm is tested on the publicly available document image dataset and found to be suitable to separate text and graphics regions to detect logos and stamps.

## ACKNOWLEDGMENT

This work is partially sponsored by the Ministry of Communications Information & Technology, Government of India; Ref.: MCIT 11(19)/2010HCC (TDIL) dt. 28-12-2010.

## REFERENCES

[1] B. Micenkova, J. van Beusekom, "Stamp detection in color document images," in *Int. Conf. on Doc. Analysis and Recognition (ICDAR)*, Sept. 2011, pp. 1125–1129.

[2] S. Ahmed, F. Shafait, M. Liwicki, and A. Dengel, "A generic method for stamp segmentation using part-based features," in *Int. Conf. on Doc. Analysis and Recognition (ICDAR)*, Oct. 2013, pp. 708–712.

[3] P. P. Roy, U. Pal, and J. Llados, "Seal object detection in document images using GHT of local component shapes," in *Proc. of the 2010 ACM Symposium on Applied Computing, SAC 10*, New York, USA, 2010, pp. 23–27.

[4] S. Dey, J. Mukherjee, S. Sural, and P. Bhowmick, "Colored Rubber Stamp Removal from Document Images," in *5th Int. Conf. on Pattern Recognition and Mach. Intell.*, 2013, pp. 545–550.

[5] G. Zhu, D. Doermann, "Automatic document logo detection," in *Int. Conf. on Doc. Analysis and Recognition (ICDAR)*, Sept. 2007, pp. 864–868.

[6] H. Wang, Y. Chen, "Logo detection in document images based on boundary extension of feature rectangles," in *Int. Conf. on Doc. Analysis and Recognition (ICDAR)*, July 2009, pp. 1335–1339.

[7] D. Doermann, K. Tombre, *Handbook of Document Image Processing and Recognition*, Springer-Verlag London, 2014.

[8] F. Y. Shih, *Image Processing and Pattern Recognition-Fundamentals and Techniques*, IEEE Press, John Wiley & Sons, 2010.

[9] F. M. Wahl, K. Y. Wong, and R. G. Casey, "Block segmentation and text extraction in mixed text/image documents," *Comput. Vis. Graph. Image Process.*, vol. 20, pp. 375–390, 1982.

[10] J. L. Fisher, S. C. Hinds, and D. P. D'Amato, "A rule-based system for document image segmentation," in *IEEE Proc. of the tenth Int. Conf. on Pattern Recognition*, Atlantic City, NJ, June 1990, pp. 567–572.

[11] D. Wang, S. N. Srihari, "Classification of newspaper image blocks using texture analysis," *Comput. Vis. Graph. Image Process.*, vol. 47, no. 3, pp. 327–352, 1989.

[12] L. A. Fletcher, R. Kasturi, "A robust algorithm for text string separation from mixed text/graphics images," *IEEE Trans. Pattern and Mach. Intell.*, vol. 10, no. 6, pp. 910–918, Nov. 1988.

[13] F. Y. Shih, S. S. Chen, "Adaptive document block segmentation and classification," *IEEE Trans. On Sys. Man and Cybern.*, vol. 26, no. 5, pp. 797–802, Oct. 1996.

[14] A. K. Jain, Yu Zhong, "Page segmentation using texture analysis," *Pattern Recognition*, vol. 29, no. 5, pp. 743–770, 1996.

[15] M. Cote, A. B. Albu, "Texture sparseness for pixel classification of business document images," *Int. Journal on Doc. Analysis and Recognition*, vol. 17, pp. 257–273, 2014.

[16] T. V. Hoang, S. Tabbone, "Text extraction from graphical document images using sparse representation," in *Int. Workshop on Document Analysis Systems*, 2010, pp. 143–150.

[17] J. Ha, R. Haralick, and I. Phillips, "Recursive X-Y cut using bounding boxes of connected components," in *Int. Conf. on Doc. Analysis and Recognition (ICDAR)*, 1995, pp. 952–955.

[18] G. Qiu, K. M. Lam, "Frequency layered color indexing for content based image retrieval," *IEEE Trans. Image Processing*, vol. 12, no. 1, pp. 102–113, Jan. 2003.

[19] D. Comaniciu, P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. on Pattern Analysis and Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[20] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. John Wiley & Sons (Asia), 2002.