Received 12 January 2023, accepted 18 February 2023, date of publication 27 February 2023, date of current version 3 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3249961

RESEARCH ARTICLE

Segmentation-Less Extraction of Text and Non-Text Regions From JPEG 2000 Compressed Document Images Through Partial and Intelligent Decompression

TEJASVEE BISEN[1], MOHAMMED JAVED[1], P. NAGABHUSHAN[2], AND OSAMU WATANABE[3], (Member, IEEE)

[1]Department of Information Technology, Indian Institute of Information Technology, Allahabad, Allahabad, Uttar Pradesh 211015, India
[2]Department of Computer Science and Engineering, Vignan University, Guntur, Andhra Pradesh 522213, India
[3]Department of Electronics and Computer Systems (ECS), Takushoku University, Tokyo 112-8585, Japan

Corresponding author: Tejasvee Bisen (rsi2018001@iiita.ac.in)

ABSTRACT JPEG 2000 is a popular image compression technique that uses Discrete Wavelet Transform (DWT) for compression and subsequently provides many rich features for efficient storage and decompression. Though compressed images are preferred for archival and communication purposes, their processing becomes difficult due to the overhead of decompression and re-compression operations which are needed as many times the data needs to operate. Therefore in this research paper, the novel idea of direct operation over the JPEG 2000 compressed documents is proposed for extracting text and non-text regions without using any segmentation algorithm. The technique avoids full decompression of the compressed document in contrast to the conventional methods, where they fully decompress and then process. Moreover, JPEG 2000 features are explored in this research work to partially and intelligently decompress only the selected regions of interest at different resolutions and bitdepths to accomplish segmentation-less extraction of text and non-text regions. Finally Maximally Stable Extremal Regions (MSER) algorithm is used to extract the layout of segmented text and non-text regions for further analysis. Experiments have been carried out on the standard PRImA Layout Analysis Dataset leading to promising results and saving computational resources.

INDEX TERMS Bitdepths, DWT, JPEG 2000, MSER, partial and intelligent decompression, resolutions, text and non-text segmentation.

I. INTRODUCTION

Digital images in raw form generally occupy large storage space, and hence image compression techniques are being invented and popularly used to provide efficient storage and transmission [1], [2]. Compression of data on one hand provides efficiency for archival and communication purposes, but on the other hand its processing gets expensive. This is because, the compressed data needs to decompressed and then processed, which requires more computational resources. Further, the decompressed data after processing, again needs to be compressed for efficiency reasons.

The associate editor coordinating the review of this manuscript and approving it for publication was Badri Narayan Subudhi.

Thus, developing novel and intelligent techniques to operate directly over the compressed data without fully decompressing the data is a challenging research issue that has gained popularity in recent years [3]. In this research paper, an attempt is made to accomplish segmentation-less extraction of text and non-text regions from document images using partial and intelligent decompression of JPEG 2000 compressed document images.

JPEG 2000 [4] image encoding standard was released by JPEG committee in the year 2000, and since then, it has been very popular due its support for high level of scalability and accessibility in colour images. Document images with colour information is required in many applications such as document forensics, signature forgeries detection, as well as in many biomedical applications [5]. Documents with coloured background and text embedded in them, such as magazines and newspapers, are very common nowadays. Segmenting text and non-text from these documents is a challenging task in the field of Document Image Analysis and Recognition (DIAR). Segmentation improves the accuracy rate within the OCR process and boosts performances in the DIAR [6]. Some of the methods for segmenting text and non-text are region-based, connected component-based with features such as edge and texture, and some hybrid approaches [7]. Extracting text from documents such as bank cheques and forms are attempted by raster image analysis and vectorized image analysis methods, as reported by [8]. A text extraction method from document images of lower resolution was proposed by enhancing the resolution using the interpolation based resolution enhancement method and then extracting text by OCR [9]. Text extraction from complex colour document images using a canny edge detector for finding the edges of text regions and also using connected components and texture feature of image to separate text from non-text regions has also been reported in [10]. In [11] methods are given for extracting text from scanned, camera captured or scientific document image using gabor filter and connected components analysis. Researchers have also used Haar Discrete Wavelet Transform (DWT) for segmenting text from document images by detecting the edges and then using the line feature, vector graph based on the edge map and the stroke, and finally, the text is segmented by line feature [12]. In [13] classification of text and non-text components is performed using connected components and pixel based approach. Statistical approaches have also been used for text and non-text classification on handwritten documents [14] and works on the extraction of text and graphics from different scripts of newspapers [15].

In the literature, many works have been attempted using machine learning models besides conventional approaches for segmentation of document images [16], [17]. In [18], text extraction is done from camera captured image using features such as intensity variation and color variance, employing K-means color clustering. Text and non-text segmentation using deep learning based Convolutional Neural Network(CNN) [19], fully convolutional network (FCN) [20] from historical handwritten documents are also reported. It has been reported that segmentation of document images using discriminative learning instead of connected components have performed better results over connected components [21]. With text and non-text segmentation, it is required to understand the document structure or layout to maintain the reading order of text in the document as we find documents with a single column, multiple columns, and complex backgrounds, etc. Layout analysis is performed in two ways: structural layout analysis [22] is performed to group the document components, and functional layout analysis [23] is performed for labeling the structural blocks using domain-dependent information. For example, the first page of the research paper contains the functional blocks, the title, author, abstract, keywords, and paragraph of the text body [24]. Researchers have worked to segment the document's layout by working on the foreground and background pixels [25]. Page segmentation plays an important role in applications such as document indexing, accessibility, and document classification [26]. Along with conventional methods, machine learning methods works are also used for layout segmentation. According to a survey in [27], deep neural networks (DNNs) are used for segmenting pages of historical documents. However, the text and non-text detection methods and page layout analysis methods discussed so far only work with the fully decompressed documents, as illustrated in Figure-1, and they cannot be directly applied on compressed document images. Hence, exploring novel techniques for operating with compressed document data either directly or through intelligent partial decompression is an interesting research problem.

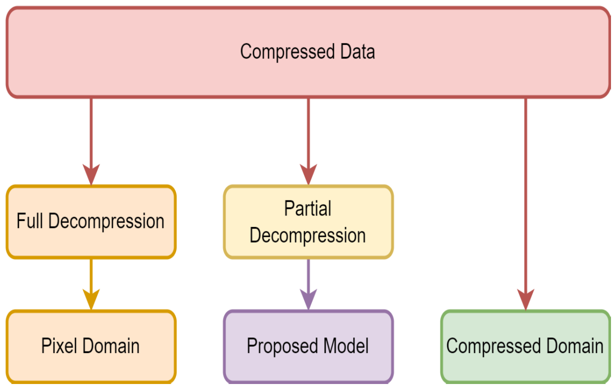

FIGURE 1. Model for processing compressed data, in the pixel domain with full decompression, on partially decompressed data, and directly in compressed domain.

In the literature, there are also attempts to segment text and non-text directly in compressed document images with formats like run length encoding (RLE), JPEG, JBIG. Text segmentation, line, word and character, document block segmentation, font size detection have been performed directly in compressed domain using RLE as compression technique [28]. Researchers have also processed image directly in JPEG compressed domain for scaling, rotating, segmenting images [29]. Also research works have been reported to segment document images directly in JBIG (Joint Bi-level Image Expert Group) compressed domain [30]. In [31] text line segmentation on handwritten document images is performed directly in Run-Lenth compressed domain. Some of works have been reported for image processing operations directly in JPEG 2000 compressed domain [32] using DWT properties. They have used techniques to modify the image and perform scalar addition, multiplication adding two images, directly in compressed domain. Recent studies have shown better results for processing images directly in compressed domain [33].

In this research paper, the main objective is to accomplish segmentation-less text and non-text extraction form JPEG 2000 compressed document images through partial

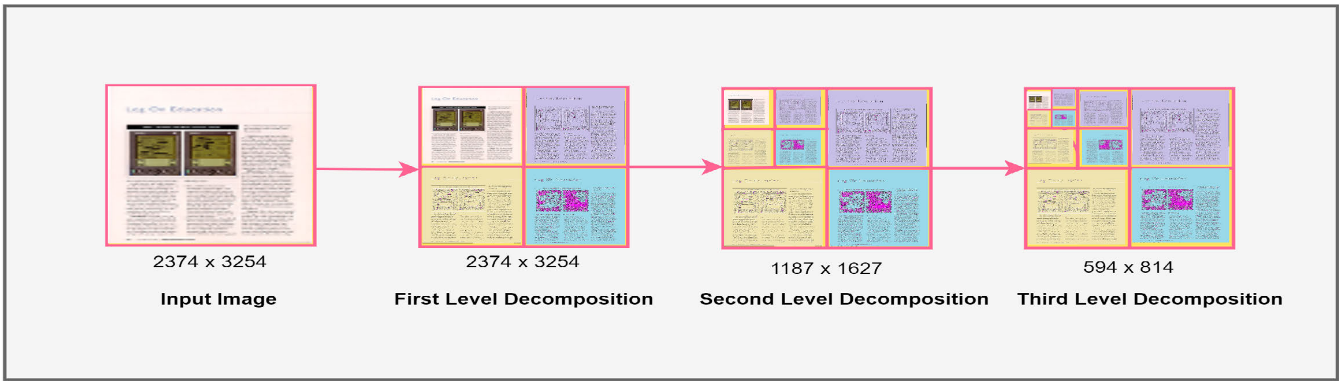

FIGURE 2. Three-level (resolution) DWT decomposition of sample image from PRImA layout analysis dataset.

and intelligent decompression. To the best of our knowledge, this is the first attempt in the literature to propose intelligent decompression on JPEG 2000 to carry out text and non-text segmentation in the compressed domain. The proposed approach is experimented with PRImA layout analysis dataset [34] and promising results have been reported.

The rest of the paper is organized as follows. In Section II we have explained the problem background which includes the brief introduction about JPEG 2000 and Maximally Stable Extremal Regions (MSER) [35] for extracting layout of segmented text and non-text regions, Section III describes our proposed approach for document image segmentation-less text and non-text separation and layout extraction method, Section IV deals with the experimental results and analysis and Section V gives concluding remark of the paper.

II. PROBLEM BACKGROUND

In this section, discussion about JPEG 2000 and MSER algorithm is provided as background for the proposed model in the current research work.

A. JPEG 2000

In JPEG 2000 [36], contents can be encoded using DWT without loss, and can be accessed and decoded in many qualities and resolutions. The key features of JPEG 2000 used in this research work while decompression is bitdepth and resolution, which helps to save decompression time by proposing partial decompression of the image till the required level. Bitdepth is color information of the image, as the number of bitdepth increases the colour information increases and also the image file size. JPEG 2000 uses DWT for compression of images which supports the decompression of images at different resolutions as shown in Figure-2 - the three-level DWT decomposition of an image. The information of image decreases as the DWT decomposition level increases. To reconstruct the image, $r (r > 0)$ sub-bands HL-$(\mathrm{R_L} - r + 1)$, LH- $(\mathrm{R_L} - r + 1)$ and HH- $(\mathrm{R_L} - r + 1)$ needs to be combined with the image at resolution $(r - 1)$. $\mathrm{R_L}$ indicates the number of resolution levels. A sample color image from PRImA layout analysis dataset in Figure-3 showing text and

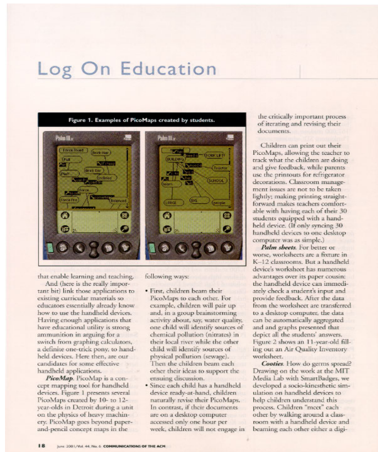

FIGURE 3. A sample colour image from PRImA layout analysis dataset showing text and non-text regions [34].

non-text regions with the dimension 2374 × 3254, 300 dpi and file size is 22 MB.

The different compression and decompression stages involved in JPEG 2000 are shown in Figure-4. In the proposed approach (shown in Figure-5) quantization step is skipped as lossless compression is used, for which reversible colour transform (RCT) is chosen, as JPEG 2000 has also the option of irreversible colour transform (ICT) for lossy compression. RCT helps for obtaining the partially decompressed image at certain bitdepths. Through Figure-6 to Figure-9, the sample document in Figure-3 is decompressed at different resolutions (6, 5, 4 and 1) and at different bitdepths for each resolution (1 to 8), just to show the significance of partial decompression in extracting text and non-text information directly from the compressed JPEG2000 stream. From the figures, it can be observed that non-text components are clearly separated at resolution-4 and bitdepth-2 from text regions of the image. However, the text regions can be extracted at Resolution-4 and bitdepth-8. With this intelligence, the text and non-text regions can be directly extracted from the JPEG2000 compressed stream which will reduce full decompression saving

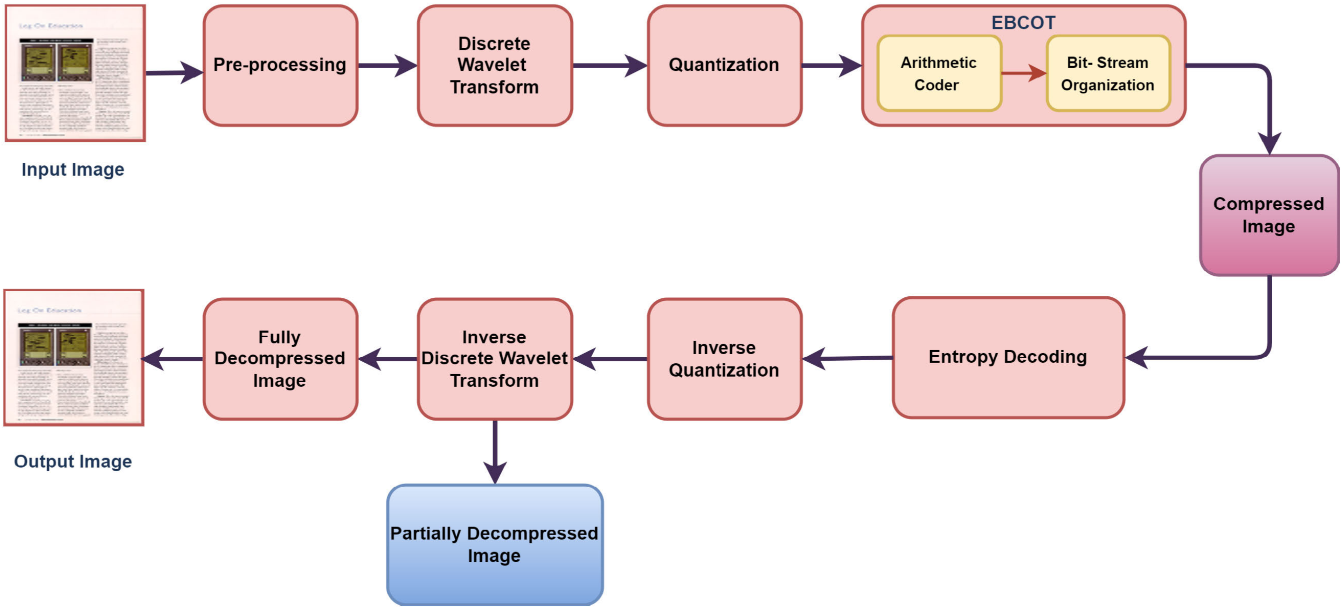

FIGURE 4. **Working Model of JPEG 2000 for compression and decompression of images.**

computation time and space. Text and non-text components in an image are separated by decomposition of components at different resolutions using wavelet transform. JPEG 2000 part I standard uses (5,3) and the (9,7) filters for dyadic decomposition [37]. An input image $I_{Image}(t)$ is dyadically decomposed to L levels by wavelet transform shown through equation 1 given below.

$$I_{Image}(t) = \sum_{n\in z} \mathrm{I}_0^L[n]\phi_n^L + \sum_{m=1}^{L}\sum_{n\in z} \mathrm{I}_1^m[m]\psi_n^L \tag{1}$$

where $\mathrm{I}_0^d[n]$ and $\mathrm{I}_1^d[n]$ are high pass and low pass band, at resolution level d respectively. The ϕ_n^L and ψ_n^L are scaling and wavelet function respectively.

Table-1 shows the partial decompression time for the sample image shown in Figure-3 at different resolutions and different bitdepths. It is clearly observed that partial decompression at lower resolution saves lot of computation time in comparison with full decompression of the image at the highest resolution. Also, it is noted that the computation time across different bitdepths remain all most same for a particular resolution. Resolution-6 which is the highest resolution takes more time to decompress than other lower resolutions. Other lower resolutions provide image contents with some loss of information; therefore depending on the type of application and analysis required a particular resolution and bitdepth can be chosen.

B. MSER

MSER stands for Maximally Stable Extremal Regions [35] is used for layout extraction of coloured document image. It's a blob detection method in images.MSER detects co-variant regions from images. It is a stable connected component of the grayscale images. It works by connecting the approximately same regions using a range of thresholds. MSER has been used to extract the text using the property stroke width.

III. PROPOSED METHOD

The block diagram of the proposed model for accomplishing segmentation-less extraction of text and non-text region and subsequent layout extraction with partial and intelligent decompression is shown in Figure-5. The proposed approach uses some parameters for obtaining the partially decompressed image for extracting text and non-text components without using any segmentation algorithms. The parameters used during compression and decompression are mentioned in Table-2. The proposed approach uses one of the significant parameter bitdepth of OpenJPEG, to perform segmentation-less extraction of text and non-text components. It works for coloured document images at different bitdepths, which is the colour information of the image. The input to the model is a coloured document image compressed with JPEG 2000 (6 DWT level decomposition) using parameters during compression. To extract text and non-text regions from the compressed image, partial decompression at different resolutions and bitdepths is explored. In the Figure-5 shown, the compressed image is partially decompressed at resolution-4 and bitdepth-2 for extracting non-text components, whereas for extracting text components it is partially decompressed at resolution-4 and bitdepth-8. After obtaining the partially decompressed images at different bitdepths, the subtraction operation is performed between image I_{PD} partially decompressed image obtained at bitdepth-8 and non-text component image I_{NT} at bitdepth-2 and both at resolution-4 to obtain the

TABLE 1. Partial Decompression time of a coloured Sample magazine image of PRImA Layout Analysis Dataset at different resolutions and Bitdepths (Time in milliseconds).

Partial Decompression Time									
Resolution	Bitdepth-1	Bitdepth-2	Bitdepth-3	Bitdepth-4	Bitdepth-5	Bitdepth-6	Bitdepth-7	Bitdepth-8	Average Time
6	1495	1499	1500	1503	1505	1506	1507	1518	1504.125
5	408	410	412	417	417	417	418	418	414.75
4	95	97	98	101	102	102	103	103	100.125
3	35	37	37	38	37	38	38	39	37.375
2	14	14	15	15	16	15	15	15	14.875
1	8	7	7	8	8	8	8	9	7.875

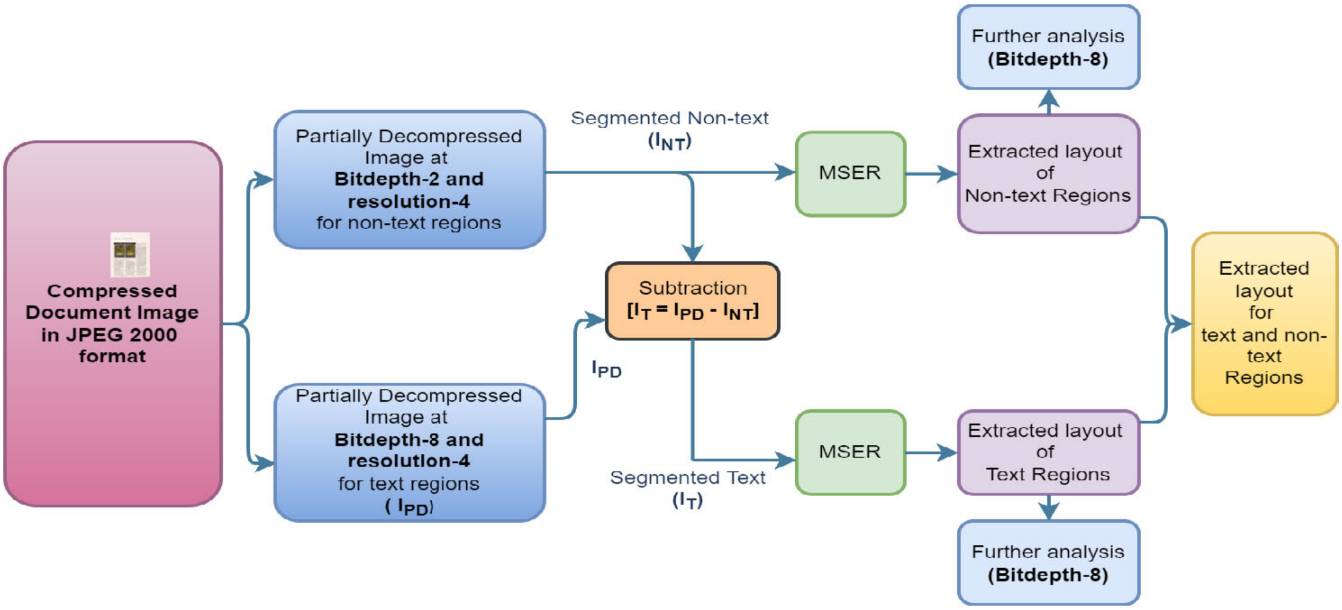

FIGURE 5. Block Diagram for segmentation-less text and non-text separation and layout extraction and further analysis (symbol spotting, word spotting) of document images using partial decompression and MSER.

TABLE 2. Parameters used while compression and decompression of Image.

Compression Parameters	Decompression Parameters
-n (number of resolutions-6)	-r (resolution 0-5)
-r (Compression ratio-10)	-p (Bitdepth 1-8)

text component image I_T as shown in equation 2.

$$I_T = I_{PD} - I_{NT} \tag{2}$$

However, for further processing of segmented text and non-text components, bitdepth-8 is preferred as the quality of image contents at bitdepth-2 is not good. For layout segmentation, MSER is applied to the partially decompressed text and non-text components of the image extracted at different resolutions and bitdepths. MSER provides a layout to both text and non-text parts of the image by a bounding box. The proposed model, depending on the requirements of the DIAR application, provides the flexibility of carrying out layout extraction separately for text regions and non-text regions, as well as the combination of both text and non-text regions as illustrated in Figure-5. The segmented images can be used for further analysis such as information spotting tasks (symbol spotting, word spotting) after extracting text and non-text regions. So our approach provides a segmentation-less extraction of text, non-text components, and layout from JPEG 2000 compressed document images using partial and intelligent decompression.

IV. EXPERIMENTAL RESULTS

The proposed approach has been implemented using the ''OpenJPEG'', [38] open-source software for JPEG 2000 implementation for compression and decompression, and Spyder 3.3.4 for text and non-text separation and layout extraction, with 8GB RAM and i5 CPU Processor.The proposed method has been tested on the JPEG 2000 compressed version of the publicly available dataset PRImA Layout Analysis Dataset [34] comprising of competition datatset images RDCL2017 [39] and RDCL2019 [40]. The dataset has 478 consists of coloured document images comprising of magazines and technical/scientific articles respectively. These datasets are developed by PRImA Research

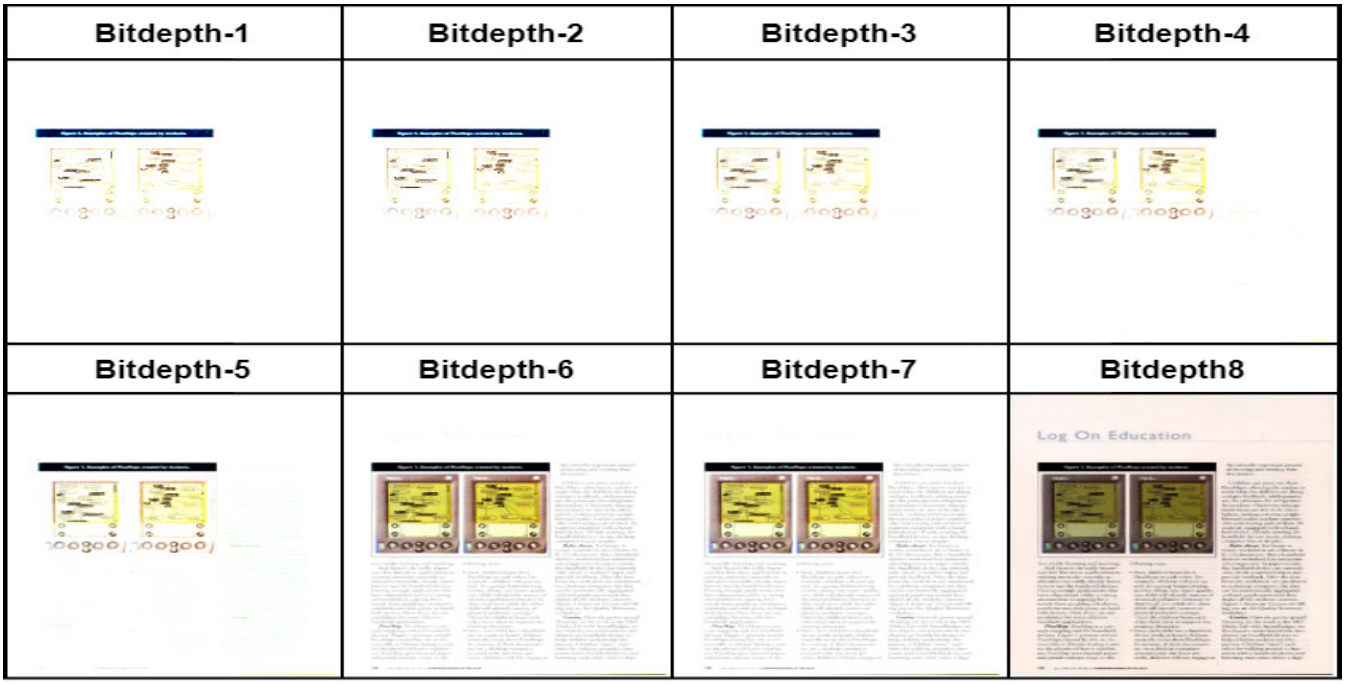

FIGURE 6. A sample document image of PRImA Layout Analysis Dataset partially decompressed at resolution-6 and various different bitdepths.

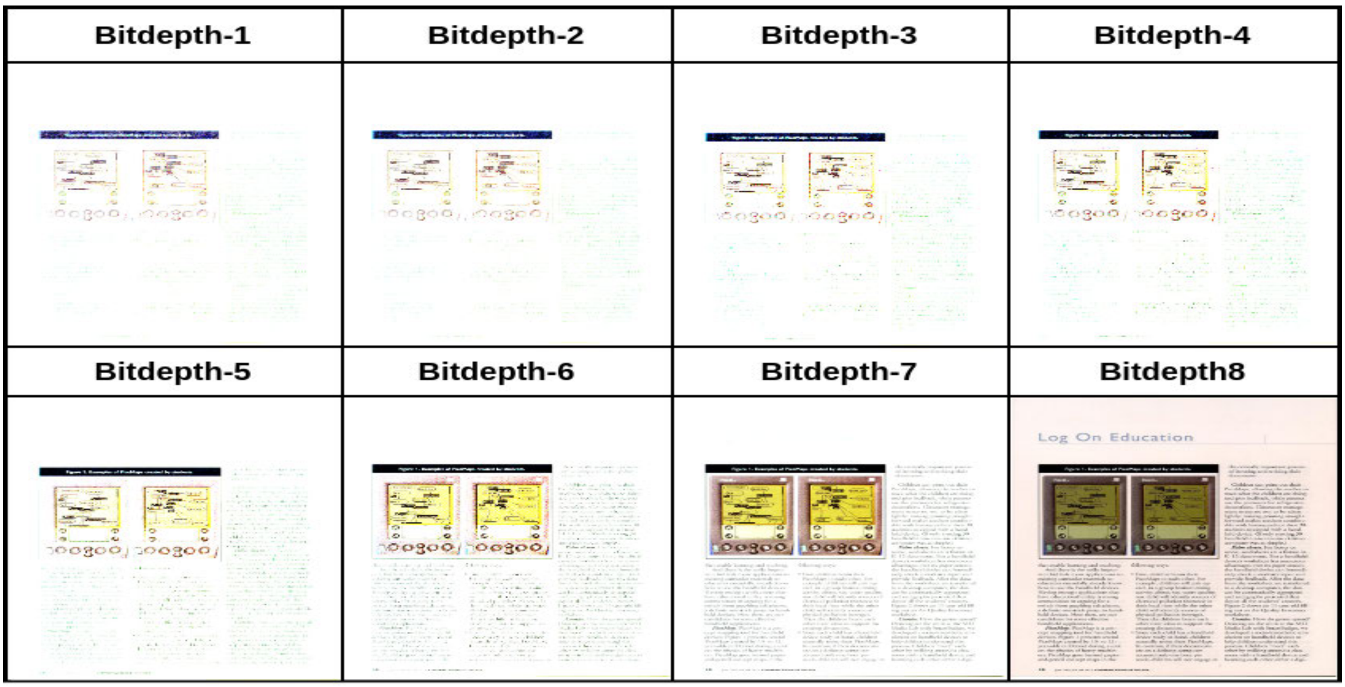

FIGURE 7. A sample document image of PRImA Layout Analysis Dataset partially decompressed at resolution-5 and various different bitdepths.

Lab, University of Salford, UK. The documents of these datasets are diverse and complex. They have different types of regions like text, inverted text,non-text (images, tables, equations, graphs, charts etc). with different shape and size, which make it very challenging for the segmentation of the text and non-text components. The output of the proposed approach on sample image of PRImA Layout Analysis Dataset is depicted in Figure18. Different evaluation metrics

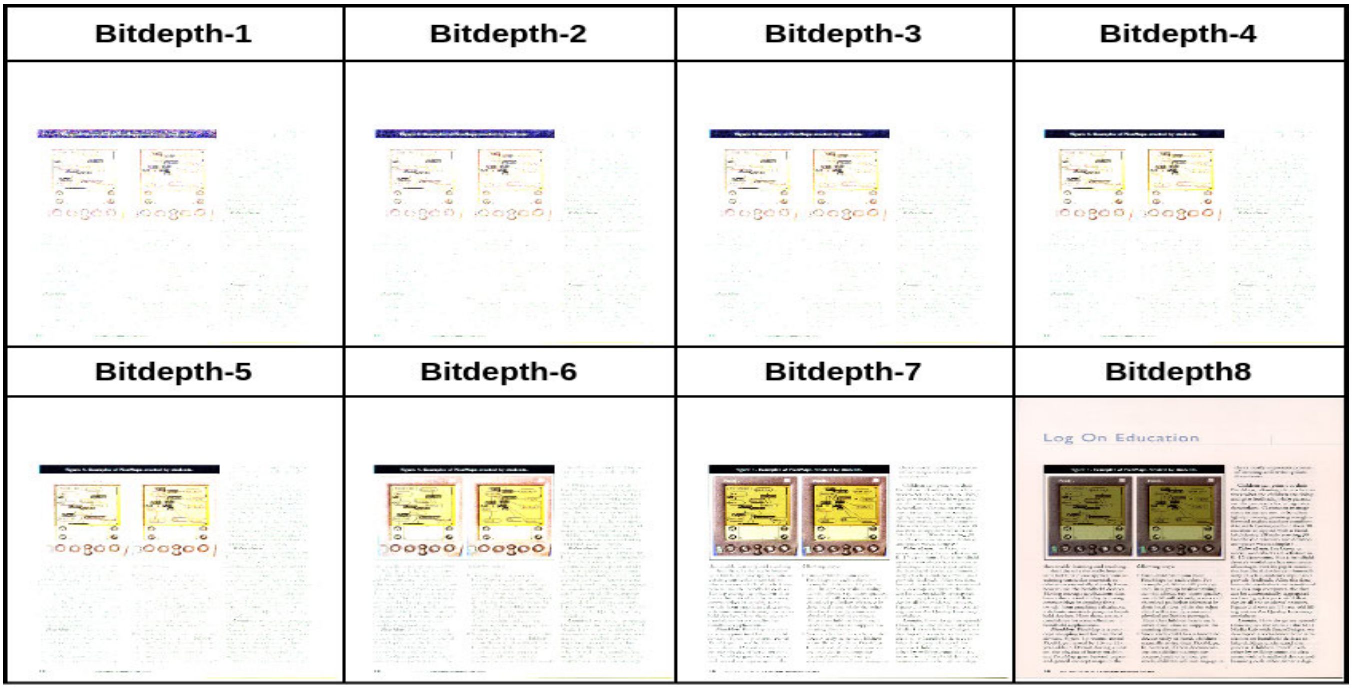

FIGURE 8. A sample document image of PRImA Layout Analysis Dataset partially decompressed at resolution-4 and various different bitdepths.

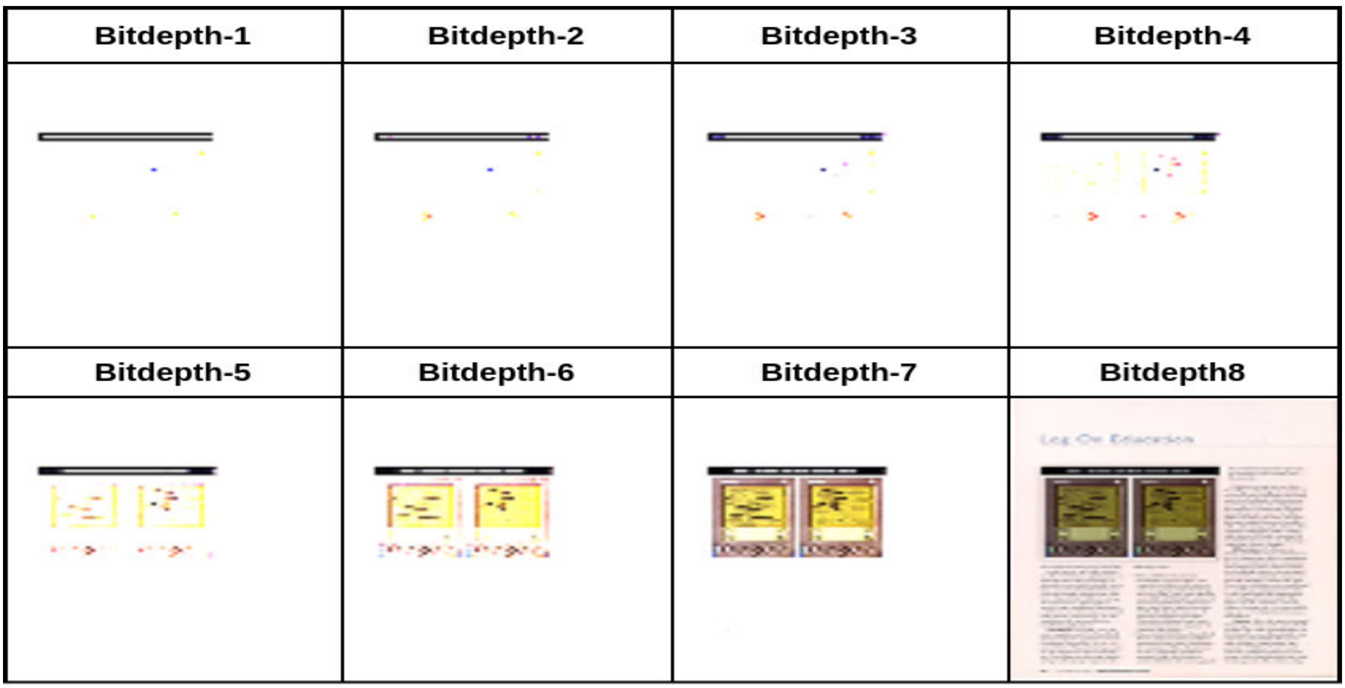

FIGURE 9. A sample document image of PRImA Layout Analysis Dataset partially decompressed at resolution-1 and various different bitdepths.

used for performance evaluation of the proposed approach on coloured document image are precision, recall, F1 Score and Segmentation accuracy.

The Precision defines the proportion of correctly predicted positive values to the total predicted positive values, the Recall defines the proportion of correctly predicted positive

TABLE 3. Performance evaluation of proposed approach for text segmentation, non-text segmentation, and text and non-text segmentation accuracy at different bitdepths (at resolution-4) for the coloured magazine images of PRImA Layout Analysis Dataset.

	Text Extraction			Non-text Extraction			
Bitdepths	F1 Score	Precision	Recall	F1 Score	Precision	Recall	Accuracy(%)
1	94.63	94.42	94.83	93.23	90.69	95.92	95.37
2	94.56	94.25	94.88	93.25	90.72	95.92	95.40
3	93.84	92.81	94.89	93.27	90.78	95.92	95.41
4	93.28	91.87	94.92	93.20	90.82	95.71	95.32
5	92.74	90.57	95.02	91.21	90.812	95.74	95.38
6	92.59	90.23	95.08	91.46	90.84	92.10	93.59
7	92.59	90.22	95.10	91.33	90.83	91.83	93.46
8	92.93	90.22	95.82	91.00	90.82	91.19	93.51

TABLE 4. Performance comparison of proposed approach for text/non-text separation of document images with existing algorithms in pixel domain.

		Text			Non-text		
Methods	Accuracy(%)	F1 Score	Precision	Recall	F1 Score	Precision	Recall
Bhowmik et al., [13]	92.22	NA	NA	NA	NA	NA	NA
Khan et al., [41]	87.59	NA	NA	NA	NA	NA	NA
Sriman et al., [42]	NA	NA	98	70	NA	98	70
Ghosh et al. [43]	91.96	NA	NA	NA	NA	NA	NA
Bukhari et al. [44]	NA	99.433	NA	NA	99.668	NA	NA
Bai et al. [45]	NA	94.6	93.7	95.4	94.6	93.7	95.4
Sah et al. [46]	96.44	94.7	92.5	97	97.3	98.5	96.2
Tran et al. [47]	NA	94.99	93.31	NA	88.99	87.22	NA
Diem et al. [48]	NA	94.35	NA	NA	94.58	NA	NA
Bukhari et al. [49]	NA	NA	NA	98.19	NA	NA	91.90
Nagabhushan et al. [10]	NA	NA	94.036	97.77	NA	NA	NA
Umer et al. [50]	94.15	NA	NA	NA	NA	NA	NA
Tran et al. [51]	84.90	NA	NA	NA	NA	NA	NA
Li et al. [52]	95.29	NA	NA	NA	NA	NA	NA
Ours (Partially decompressed dataset)	95.32	93.28	91.87	94.92	93.20	90.82	95.71

TABLE 5. Processing time for the proposed approach (partial decompression + text and non-text separation time) for different resolutions and bitdepths for the sample document shown in Figure-3 image of PRImA Layout Analysis Dataset (Time in milliseconds).

Resolution	Bitdepth-1	Bitdepth-2	Bitdepth-3	Bitdepth-4	Bitdepth-5	Bitdepth-6	Bitdepth-7
6 (2374x3254)	1574.304	1578.901	1665.48	1668.51	1690.50	1789.12	1799.63
5 (1187x1627)	561.60	492.05	484.82	476.91	475.86	473.78	470.63
4 (594x814)	140.89	126.83	124.83	120.95	118.91	116.92	112.62

values to the all values in actual positive class. The F1Score defined as the average of Precision and Recall, it also takes false positive and false negative. The segmentation Accuracy is average percentage of text classified positively as text accuracy and non-text classified positively as non-text accuracy. The mathematical equations for the above metrics are represented as Equations 3 - 6, where $True_{Positive}$ is the value predicted as positive correctly, means the values of predicted and actual class are both positive. $True_{Negative}$ is the value predicted as negative correctly, means the value of predicted and actual class are both negative. $False_{Positive}$, when the value predicted is negative and the actual class is positive. $False_{Negative}$, when the value predicted is positive and actual class is negative.

$$Precision = True_{Positive}/Total_{positive} \tag{3}$$

TABLE 6. Proposed approach processing time (partial decompression, text and non-text separation time) comparison of text/non-text separation of document images of PRImA Layout Analysis Dataset with existing algorithms in pixel domain (Time in milliseconds).

Author	Approach	Processing Time (in ms)
Bhowmik et al. [13]	BINYAS (a complex document layout analysis system)	18240
Bai et al. [45]	MSP-Net	130
Li et al. [9]	Interpolation-based resolution enhancement (RE) algorithm	1040
Song et al. [18]	Colour and intensity variance method and Support Vector Machine (SVM)	1296.8
Raju et al. [11]	Anisotropic filtering, Block Energy Analysis (BEA)	30000
Chaudhuri et al., [53]	Thresholding, connected component labelling	1350
Sobottka etal. [54]	Clustering algorithm, Topdown and Bottom-up Analysis	Top-down analysis- 1050 and bottom-up analysis - 20260
Suen et al. [55]	Sobel edge detector, Constrained Run Length Algorithm (CRLA)	72000
Umer et al. [50]	Convolutional neural network	569.1
Ours	Partial decompression of images at bitdepth-4 and bitdepth-8	126.83 (proposed approach generalized to work on bitdepth-2 and lowest resolution -4)

$$\text{Recall} = \text{True}_{\text{Positive}}/\text{Total}_{\text{actualpositive}} \tag{4}$$

$$\text{F1Score} = 2 * (\text{Recall} * \text{Precision}) /(\text{Recall} + \text{Precision}) \tag{5}$$

$$\text{Accuracy} = \frac{\text{Recall}l_{\text{Text}} + \text{Recall}_{\text{Non-text}}}{2} \tag{6}$$

Table-3 shows the performance of our approach in terms of different evaluation metrics such as precision, recall, F1 score, and segmentation accuracy for the PRImA layout analysis dataset. The performance of the proposed approach has also been compared with the existing algorithms of text and non-text separation in the pixel domain or uncompressed domain as reported in Table-4. From the results shown in Table-4, it is observed that the proposed approach performs better in comparison to existing approaches. The proposed approach performs better for segmenting text and non-text components on the JPEG 2000 compressed document images through partial and intelligent decompression without using any segmentation algorithm. In Table-5 comparative study is shown in terms of processing time(partial decompression, text, and non-text separation time).

It is observed that the highest resolution which is at resolution-6, takes more time in comparison with lower resolutions 5 and 4. The proposed approach for text and non-text extraction from compressed coloured document images works at resolution 4 and bitdepth-2. This is because, at the next lower resolutions (3, 2, and 1), the text and non-text components are not visible due to the decrease in the quality of resolution of the image. In Figure-10 and Figure-11, we present the performance of the proposed approach in terms of time and space for separating text and non-text components. The graphs in Figure-10 and Figure-11 represent that it takes less time and space as resolution and bitdepth decrease, which makes the proposed approach efficient in terms of time and space.

In Section-II, it was discussed that during compression, JPEG 2000 decomposes the input image at different resolutions using dyadic decomposition. A similar pattern is also observed during partial decompression as illustrated through the graph in Figure-12. The graph shows that, as the resolution of an image is increased, the size of the text and non-text segmented components (in terms of area and perimeter) also increases. Therefore, the proposed model in this paper can segment text and non-text components through partial and intelligent decompression without employing a segmentation algorithm at resolution-4. At resolution-4, it is also ensured that the extracted text and non-text components are of good quality or resolutions so that they can be subjected to further DIAR processing such as word spotting, symbol spotting.

TABLE 7. Layout Extraction Time for a sample document image shown in Figure-3 from PRImA Layout Analysis Dataset at different resolutions and bitdepths (Time in milliseconds).

Resolution	Bitdepth-1	Bitdepth-2	Bitdepth-3	Bitdepth-4	Bitdepth-5	Bitdepth-6	Bitdepth-7
6	1676.375	1660.196	1656.244	1670.286	1706.352	1773.721	1828.765
5	447.216	446.448	443.578	460.302	479.307	482.976	482.170
4	104.188	109.225	112.789	115.32	110.374	115.944	121.877
3	38.927	40.556	39.778	41.619	40.962	41.781	44.208
2	15.68	14.918	16.279	17.189	17.474	16.271	16.994
1	8.909	7.996	8.294	9.020	9.004	9.991	9.997

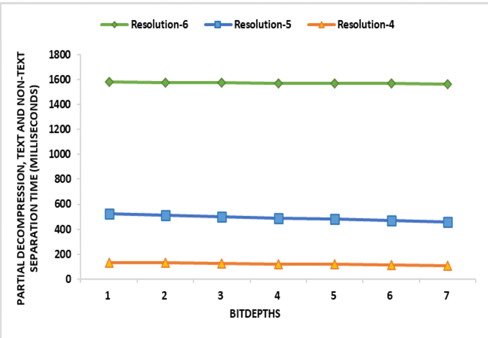

FIGURE 10. Visual representation of the processing time (Partial Decompression of text and non-text extraction time) at different bitdepths and resolution for the proposed approach.

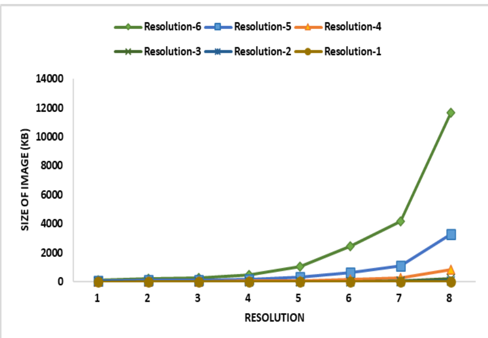

FIGURE 11. Visual representation of space analysis of proposed approach at different bitdepths and resolutions for the proposed approach.

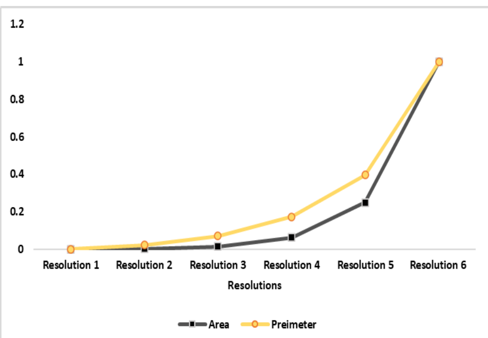

FIGURE 12. Visual representation for the increasing area and perimeter of text and non-text components on different resolutions of the proposed approach.

TABLE 8. Parameter Compression Ratio influence the size of compressed image without affecting the partially decompressed image segmentation accuracy.

Compression Ratio	Size of Image	Accuracy
1	10.7	95.32
2	10.7	95.32
3	7.7	95.32
4	5.8	95.32
5	4.6	95.32
6	3.9	95.32
7	3.3	95.32
8	2.9	95.32
9	2.6	95.32
10	2.3	95.32
11	2.1	95.32
12	1.9	95.32
13	1.8	95.32
14	1.7	95.32
15	1.5	95.32

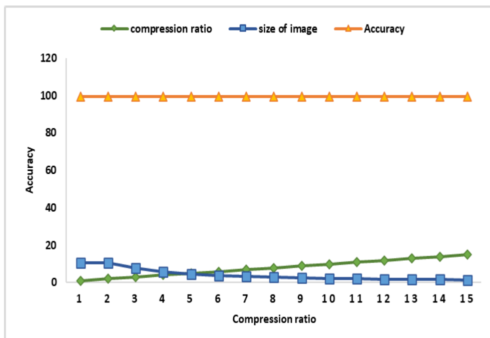

FIGURE 13. Visual representation of the increasing the parameter compression ratio of JPEG2000 while accuracy being almost same at different resolutions.

Here in the Table-6 shows the comparative analysis of the proposed approach in terms of processing time with the existing algorithm's processing time in pixel domain or uncompressed domain, which shows that the proposed approach outperforms compared other existing approaches in the uncompressed domain. To the best of our knowledge, there is no similar comparative study reported in JPEG 2000 compressed domain, and hence comparative study in this paper are reported taking research works reported in uncompressed documents. The proposed approach also

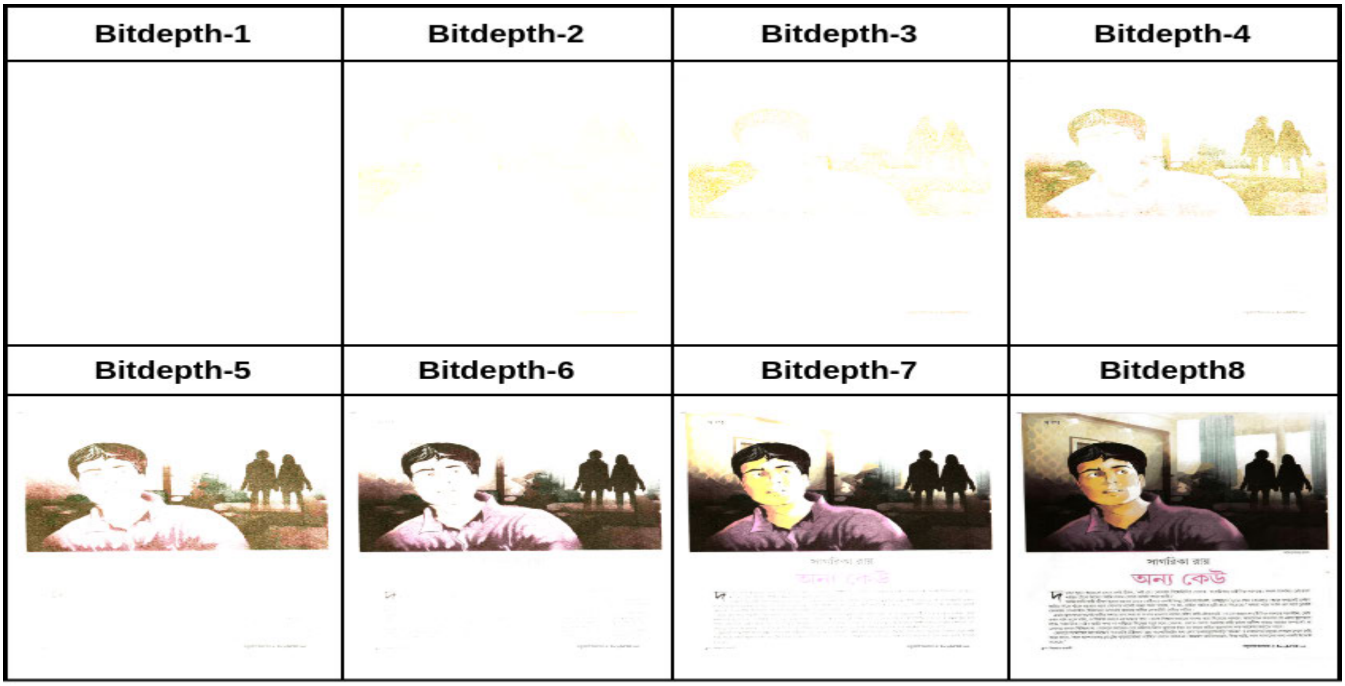

FIGURE 14. A sample Bengali coloured magazine image from "Kishore Bharti" magazine at different bitdepths at resolution-6.

FIGURE 15. A sample greyscale document image from ICDAR- 2009 dataset at different bitdepths at resolution-4.

completes the layout extraction of the document images, Table-7 shows the layout extraction time of document images at different bitdepths and resolutions. To show the robustness of the proposed model, the technique has also experimented on the Bengali magazine "Kishore Bharti" image as shown in Figure-14. It works similar to English scripts or documents as of "PRImA Layout Analysis Dataset" images. So the proposed approach is capable of working on different scripts. As shown in Figure-15, the proposed approach is also applied to a sample image of the ICDAR 2019 dataset, and it is noted that the text and non-text from the images are not extracted. This is because, it is a grayscale image where bitdepth is 1 bit during JPEG 2000 compression, and therefore proposed model is not helpful to segment the text and non-text components in such documents. The proposed approach has also mislabeled the text as non-text (text within non-text objects) and vice-versa to some extent, due to the same colour information in both text and non-text regions as shown in Figure-16. The proposed approach also extracts some small non-text components at resolution-4 and bitdepth-2 which is illustrated in Figure-17. Despite the issue of mislabeling the text as non-text and vice-versa, it can be said that the proposed approach performs better as compared to conventional approaches. By exploring the bitdepth and resolution features of the image it brings novelty to the proposed approach, by performing segmentation-less extraction of text, non-text components, and layout extraction.

The graphs in Figure-10 and Figure-11, show the processing time (partial decompression + text and non-text separation time), and it is observed that as resolution decreases,

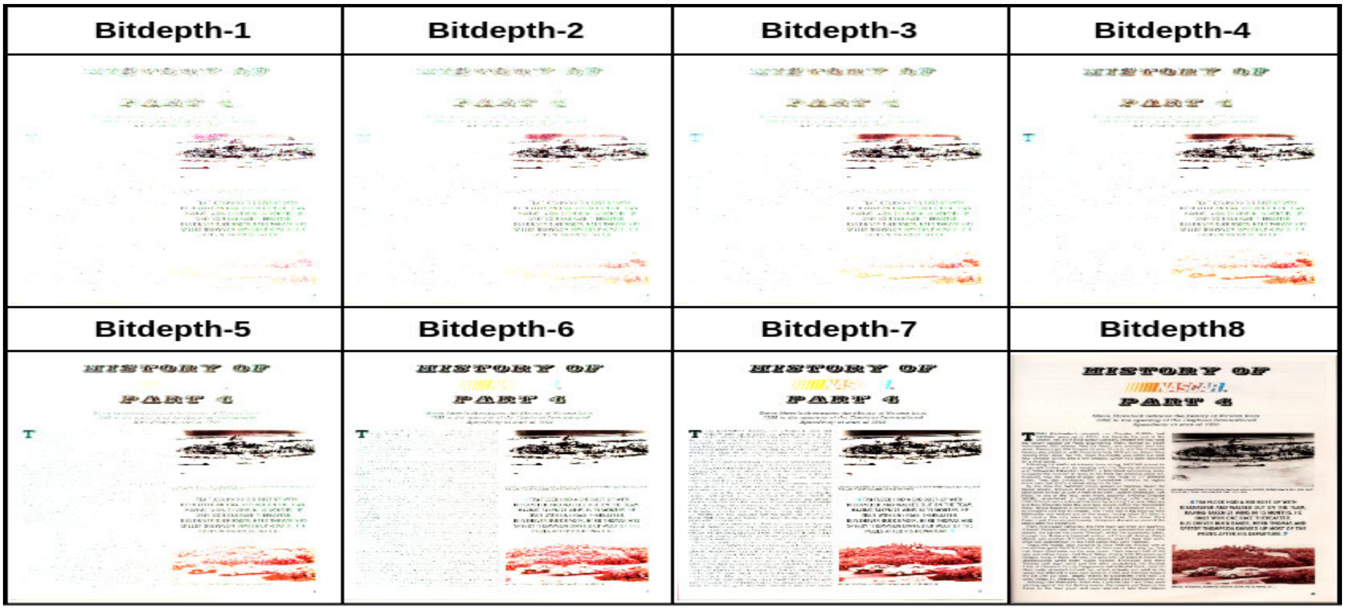

FIGURE 16. A sample document image of PRImA Layout Analysis Dataset partially decompressed at resolution-4 and various different bitdepths which mislabels text as non-text components.

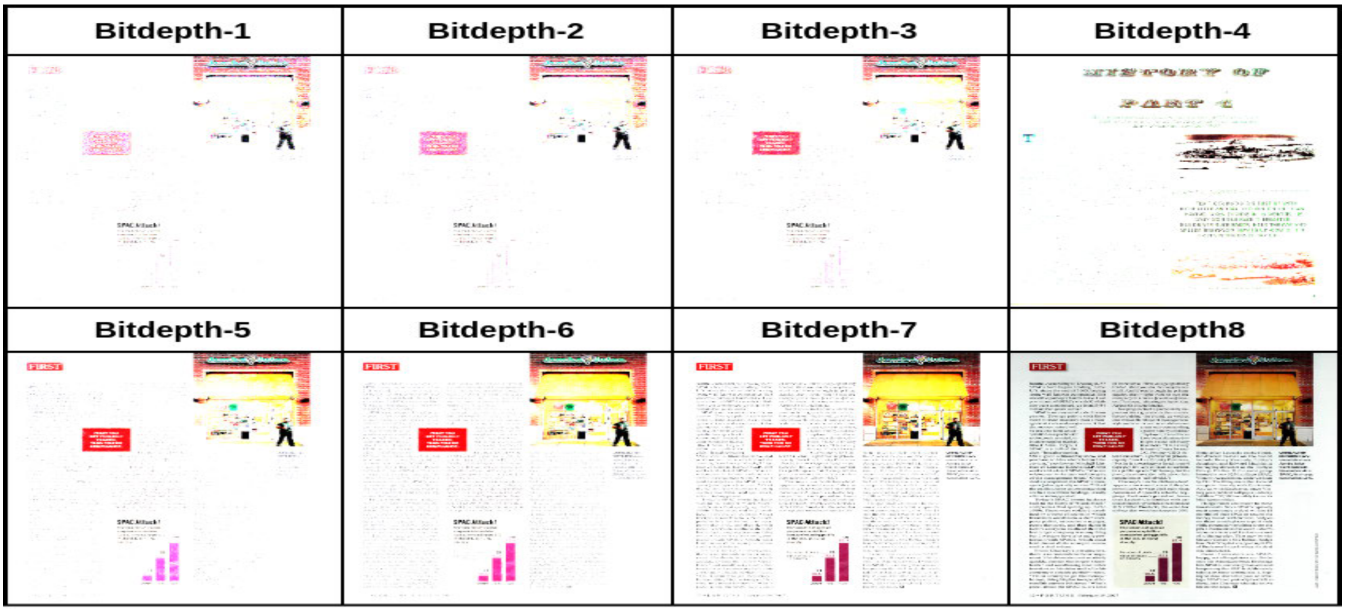

FIGURE 17. A sample document image of PRImA Layout Analysis Dataset partially decompressed at resolution-4 and various different bitdepths which detect small non-text components at resolution-4 and bitdepth-2.

the processing time also decreases. So we can perform the text and non-text separation of coloured document images at a lower resolution (resolution-4) to save decompression time, space and quality. Layout extraction for the entire document can also be performed at lower resolutions by partially decompressing the image. The result for text and non-text separation, layout extraction of coloured document image at different bitdepths for resolution-4 are shown in Figure-18. It shows that the proposed approach successfully separates the non-text components at bitdepth-2 and text regions at bitdepth-8 of the coloured document images at resolution-4. The proposed approach has also been experimented with using the parameter compression ratio during compression which reduces the size of the compressed image, which is shown in the Table-8 and is visualized in Figure-13. The proposed approach works independently from the parameter compression ratio and gives promising results. The results of the proposed approach show that it performs reasonably well for extracting text, non-text and layout in terms of time and space while maintaining the accuracy of the approach.

FIGURE 18. Segmented image of text and non-text components and layout extraction of coloured document image at bitdepth-2 and resolution-4, where first row is bit depth image, second row is non-text Image, third row is text image, fourth row is final layout extracted image.

V. CONCLUSION

Text and non-text segmentation, layout extraction is a significant step in many DIAR processing. The performance of this step has a further impact on the system. In the present work, we have presented a novel research work on segmentation-less extraction of text and non-text components through partial and intelligent decompression for JPEG 2000 compressed documents images. It utilizes the unexplored features bitdepth and resolution of JPEG2000 compressed images. Document images are partially decompressed using both features, at bitdepth-2 and resolution-4 to accomplish the task of segmentation. Layout extraction is also performed with the help of MSER to provide a bounding box to the regions of the partially decompressed image. The performance of the proposed approach is evaluated on PRImA Layout Analysis Dataset comprising competition dataset images of RDCL2017 and RDCL2019 and compared with existing methods. The results show that partial decompression of image (at lower resolution and bitdepth) and segmentation-less separation of text and non-text saves time and space, which helps in reducing the time taken by DIA operations. The proposed approach performs better in terms of time, space and segmentation accuracy.

REFERENCES

[1] K. Sayood, *Introduction to Data Compression*. Oxford, U.K.: Newnes, 2012.

[2] N. Khan, I. Yaqoob, I. A. T. Hashem, Z. Inayat, W. K. M. Ali, M. Alam, M. Shiraz, and A. Gani, "Big data: Survey, technologies, opportunities, and challenges," *Sci. World J.*, vol. 2014, Jul. 2014, Art. no. 712826.

[3] M. Javed, P. Nagabhushan, and B. B. Chaudhuri, "A review on document image analysis techniques directly in the compressed domain," *Artif. Intell. Rev.*, vol. 50, no. 4, pp. 539–568, 2017.

[4] M. Rabbani, "JPEG2000: Image compression fundamentals, standards and practice," *J. Electron. Imag.*, vol. 11, no. 2, p. 286, 2002.

[5] K. Jung and R. Seiler, "Segmentation and compression of documents with JPEG2000," *IEEE Trans. Consum. Electron.*, vol. 49, no. 4, pp. 802–807, Nov. 2003.

[6] V. P. Le, N. Nayef, M. Visani, J.-M. Ogier, and C. D. Tran, "Text and non-text segmentation based on connected component features," in *Proc. 13th Int. Conf. Document Anal. Recognit. (ICDAR)*, Aug. 2015, pp. 1096–1100.

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

[8] S. Bhowmik, R. Sarkar, M. Nasipuri, and D. Doermann, "Text and non-text separation in offline document images: A survey," *Int. J. Document Anal. Recognit.*, vol. 21, nos. 1–2, pp. 1–20, Jun. 2018.

[9] Z. Li and J. Luo, "Resolution enhancement from document images for text extraction," in *Proc. 5th FTRA Int. Conf. Multimedia Ubiquitous Eng.*, Jun. 2011, pp. 251–256.

[10] P. Nagabhushan and S. Nirmala, "Text extraction in complex color document images for enhanced readability," *Intell. Inf. Manage.*, vol. 2, no. 2, pp. 120–133, 2010.

[11] S. Raju S, P. B. Pati, and A. G. Ramakrishnan, "Gabor filter based block energy analysis for text extraction from digital document images," in *Proc. 1st Int. Workshop Document Image Anal. Libraries*, 2004, p. 233.

[12] S. Audithan and R. Chandrasekaran, "Document text extraction from document images using Haar discrete wavelet transform," *Eur. J. Sci. Res.*, vol. 36, no. 4, pp. 502–512, 2009.

[13] S. Bhowmik, S. Kundu, and R. Sarkar, "BINYAS: A complex document layout analysis system," *Multimedia Tools Appl.*, vol. 80, no. 6, pp. 8471–8504, Mar. 2021.

[14] B. Pravalpruk and S. Watcharabutsarakham, "Statistical approach for text and non-text classifier in off-line handwritten document," in *Proc. 17th Int. Conf. Electr. Eng./Electron., Comput., Telecommun. Inf. Technol. (ECTI-CON)*, Jun. 2020, pp. 644–647.

[15] R. P. Kaur, M. Jindal, and M. Kumar, "Text and graphics segmentation of newspapers printed in Gurmukhi script: A hybrid approach," *Vis. Comput.*, vol. 37, pp. 1637–1659, Jul. 2020.

[16] R. Batista das Neves Junior, L. F. Verçosa, D. Macêdo, B. L. D. Bezerra, and C. Zanchettin, "A fast fully octave convolutional neural network for document image segmentation," 2020, *arXiv:2004.01317*.

[17] A. Sheshkus, D. Nikolaev, and V. L. Arlazarov, "Houghencoder: Neural network architecture for document image semantic segmentation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 1946–1950.

[18] Y. J. Song, K. C. Kim, Y. W. Choi, H. R. Byun, S. H. Kim, S. Y. Chi, D. K. Jang, and Y. K. Chung, "Text region extraction and text segmentation on camera-captured document style images," in *Proc. 8th Int. Conf. Document Anal. Recognit. (ICDAR)*, 2005, pp. 172–176.

[19] K. Chen, M. Seuret, J. Hennebert, and R. Ingold, "Convolutional neural networks for page segmentation of historical document images," in *Proc. 14th Int. Conf. Document Anal. Recognit. (ICDAR)*, vol. 1, Nov. 2017, pp. 965–970.

[20] Y. Xu, W. He, F. Yin, and C.-L. Liu, "Page segmentation for historical handwritten documents using fully convolutional networks," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, Nov. 2017, pp. 541–546.

[21] S. S. Bukhari, M. I. A. Al Azawi, F. Shafait, and T. M. Breuel, "Document image segmentation using discriminative learning over connected components," in *Proc. 9th IAPR Int. Workshop Document Anal. Syst.*, Jun. 2010, pp. 183–190.

[22] A. M. Namboodiri and A. K. Jain, "Document structure and layout analysis," in *Digital Document Processing*. Cham, Switzerland: Springer, 2007, pp. 29–48.

[23] R. Kasturi, L. O'Gorman, and V. Govindaraju, "Document image analysis: A primer," *Sadhana*, vol. 27, no. 1, pp. 3–22, Feb. 2002.

[24] L. O'Gorman and R. Kasturi, *Document Image Analysis*, vol. 39. Los Alamitos, CA, USA: IEEE Computer Society Press, 1995.

[25] H. Alheritiere, F. Cloppet, C. Kurtz, J.-M. Ogier, and N. Vincent, "A document straight line based segmentation for complex layout extraction," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, Nov. 2017, pp. 1126–1131.

[26] D. He, S. Cohen, B. Price, D. Kifer, and C. L. Giles, "Multi-scale multi-task FCN for semantic page segmentation and table detection," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, Nov. 2017, pp. 254–261.

[27] B. Liebl and M. Burghardt, "An evaluation of DNN architectures for page segmentation of historical newspapers," 2020, *arXiv:2004.07317*.

[28] M. Javed, P. Nagabhushan, and B. B. Chaudhuri, "Direct processing of document images in compressed domain," 2014, *arXiv:1410.2959*.

[29] R. L. de Queiroz, "Processing JPEG-compressed images and documents," *IEEE Trans. Image Process.*, vol. 7, no. 12, pp. 1661–1672, Dec. 1998.

[30] E. E. Regentova, S. Latifi, D. Chen, K. Taghva, and D. Yao, "Document analysis by processing JBIG-encoded images," *Int. J. Document Anal. Recognit.*, vol. 7, no. 4, pp. 260–272, Sep. 2005.

[31] R. Amarnath and P. Nagabhushan, "Text line segmentation in compressed representation of handwritten document using tunneling algorithm," 2019, *arXiv:1901.11477*.

[32] F. Chebil, M. K. Bel Hadj Miled, A. Islam, and K. Willner, "Compressed domain editing of JPEG2000 images," *IEEE Trans. Consum. Electron.*, vol. 51, no. 2, pp. 710–717, May 2005.

[33] J. Mukhopadhyay, *Image and Video Processing in the Compressed Domain*. Boca Raton, FL, USA: CRC Press, 2011.

[34] A. Antonacopoulos, D. Bridson, C. Papadopoulos, and S. Pletschacher, "A realistic dataset for performance evaluation of document layout analysis," in *Proc. 10th Int. Conf. Document Anal. Recognit.*, Dec. 2009, pp. 296–300.

[35] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vis. Comput.*, vol. 22, no. 10, pp. 761–767, 2004.

[36] P. Schelkens, A. Skodras, and T. Ebrahimi, *The JPEG 2000 Suite*, vol. 15. Hoboken, NJ, USA: Wiley, 2009.

[37] M. D. Adams and R. Ward, "Wavelet transforms in the JPEG-2000 standard," in *Proc. IEEE Pacific Rim Conf. Commun., Comput. Signal Process.*, vol. 1, Aug. 2001, pp. 160–163.

[38] A. Descampe. (Feb. 23, 2015). *OpenJPEG: An Open-Source JPEG 2000 Reference Implementation*. [Online]. Available: https://www.openjpeg.org/

[39] C. Clausner, A. Antonacopoulos, and S. Pletschacher, "ICDAR2017 competition on recognition of documents with complex layouts–RDCL2017," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, Nov. 2017, pp. 1404–1410.

[40] C. Clausner, A. Antonacopoulos, and S. Pletschacher, "ICDAR2019 competition on recognition of documents with complex layouts–RDCL2019," in *Proc. Int. Conf. Document Anal. Recognit. (ICDAR)*, Sep. 2019, pp. 1521–1526.

[41] T. Khan and A. F. Mollah, "Text non-text classification based on area occupancy of equidistant pixels," *Proc. Comput. Sci.*, vol. 167, pp. 1889–1900, Jan. 2020.

[42] B. Sriman and L. Schomaker, "Multi-script text versus non-text classification of regions in scene images," *J. Vis. Commun. Image Represent.*, vol. 62, pp. 23–42, Jul. 2019.

[43] S. Ghosh, D. Lahiri, S. Bhowmik, E. Kavallieratou, and R. Sarkar, "Text/non-text separation from handwritten document images using LBP based features: An empirical study," *J. Imag.*, vol. 4, no. 4, p. 57, Apr. 2018.

[44] S. S. Bukhari, A. Gupta, A. K. Tiwari, and A. Dengel, "High performance layout analysis of medieval European document images," in *Proc. 7th Int. Conf. Pattern Recognit. Appl. Methods*, 2018, pp. 324–331.

[45] X. Bai, B. Shi, C. Zhang, X. Cai, and L. Qi, "Text/non-text image classification in the wild with convolutional neural networks," *Pattern Recognit.*, vol. 66, pp. 437–446, Jun. 2017.

[46] A. K. Sah, S. Bhowmik, S. Malakar, R. Sarkar, E. Kavallieratou, and N. Vasilopoulos, "Text and non-text recognition using modified HOG descriptor," in *Proc. IEEE Calcutta Conf. (CALCON)*, Dec. 2017, pp. 64–68.

[47] T.-A. Tran, I.-S. Na, and S.-H. Kim, "Separation of text and non-text in document layout analysis using a recursive filter," *KSII Trans. Internet Inf. Syst.*, vol. 9, no. 10, pp. 4072–4091, 2015.

[48] M. Diem, F. Kleber, and R. Sablatnig, "Text classification and document layout analysis of paper fragments," in *Proc. Int. Conf. Document Anal. Recognit.*, Sep. 2011, pp. 854–858.

[49] S. S. Bukhari, F. Shafait, and T. M. Breuel, "Improved document image segmentation algorithm using multiresolution morphology," *Proc. SPIE*, vol. 7874, Jan. 2011, Art. no. 78740D.

[50] S. Umer, R. Mondal, H. M. Pandey, and R. K. Rout, "Deep features based convolutional neural network model for text and non-text region segmentation from document images," *Appl. Soft Comput.*, vol. 113, Dec. 2021, Art. no. 107917.

[51] H. T. Tran, N. Q. Nguyen, T. A. Tran, X. T. Mai, and Q. T. Nguyen, "A deep learning-based system for document layout analysis," in *Proc. 6th Int. Conf. Mach. Learn. Soft Comput.*, Jan. 2022, pp. 20–25.

[52] M. Li, M. Bai, and Y. Lv, "Text segmentation by integrating hybrid strategy and non-text filtering," *Multimedia Tools Appl.*, vol. 81, pp. 44505–44522, Jun. 2022.

[53] A. R. Chaudhuri, A. K. Mandal, and B. B. Chaudhuri, "Page layout analyser for multilingual Indian documents," in *Proc. Lang. Eng. Conf.*, Dec. 2002, pp. 24–32.

[54] K. Sobottka, H. Bunke, and H. Kronenberg, "Identification of text on colored book and journal covers," in *Proc. 5th Int. Conf. Document Anal. Recognit. (ICDAR)*, Dec. 1999, pp. 57–62.

[55] H. M. Suen and J. F. Wang, "Text string extraction from images of colour-printed documents," *IEE Proc. Vis., Image Signal Process.*, vol. 143, no. 4, pp. 210–216, Aug. 1996.

TEJASVEE BISEN is currently pursuing the Ph.D. degree with the Department of Information Technology, Indian Institute of Information Technology, Allahabad, Uttar Pradesh, India. Her research interests include image processing, compressed domain processing and analysis, and deep learning.

MOHAMMED JAVED received the Ph.D. degree from the University of Mysore, with a focus on detecting text in compressed document images. He was a visiting Faculty Researcher with the Institute of AI and Data Science, University at Buffalo, Buffalo, NY, USA, with Prof. David Doermann. He is currently an Assistant Professor with the Department of Information Technology, Indian Institute of Information Technology, Allahabad, Uttar Pradesh, India. His current research interests include image processing, pattern recognition, data compression, compressed domain processing and analysis, and big data analytics.

P. NAGABHUSHAN received the Ph.D. degree from the University of Mysore, with a focus on detecting text in compressed document images. He is currently the Vice-Chancellor with Vignan University, Guntur. His research interests include image processing and pattern recognition in compressed images, compressed data analytics, dimensionality reduction integrated classification, and character recognition by dimensionality reduction.

OSAMU WATANABE (Member, IEEE) received the B.S., M.S., and Ph.D. degrees from Tokyo Metropolitan University, in 1999, 2001, and 2004, respectively. In 2004, he joined the Faculty of Takushoku University as a Research Associate of electronics and computer systems, where he has been a Professor with the Department of Electronics and Computer Systems, Since 2021. His research interests include image coding and image processing. He is a member of IEICE and ITE. He received the IPSJ/ITSCJ Standardization Contribution Award, in 2021. He is also an Expert of ISO/IEC/JTC1/SG29/WG1 Committee "Joint Photographic Experts Group (JPEG)." Since October 2020, he has been the JPEG Image Coding and Quality Group Co-Chair. He is an Associate Editor of *ITE Journal*.