International Journal of Engineering

Journal Homepage: www.ije.ir

Enhancing Book and Document Digitization from Videos: A Feature Fusion-Based Approach

G. Buddhawar*[a], K. Jariwala[a], C. Chattopadhyay[b]

[a] *Computer Science and Engineering Department, Sardar Vallabhbhai National Institute of Technology, Surat, Gujarat, India*
[b] *School of Computing and Data Sciences, FLAME University, Pune, Maharashtra, India*

PAPER INFO

Paper history:
Received 10 August 2023
Received in revised form 02 December 2023
Accepted 10 December 2023

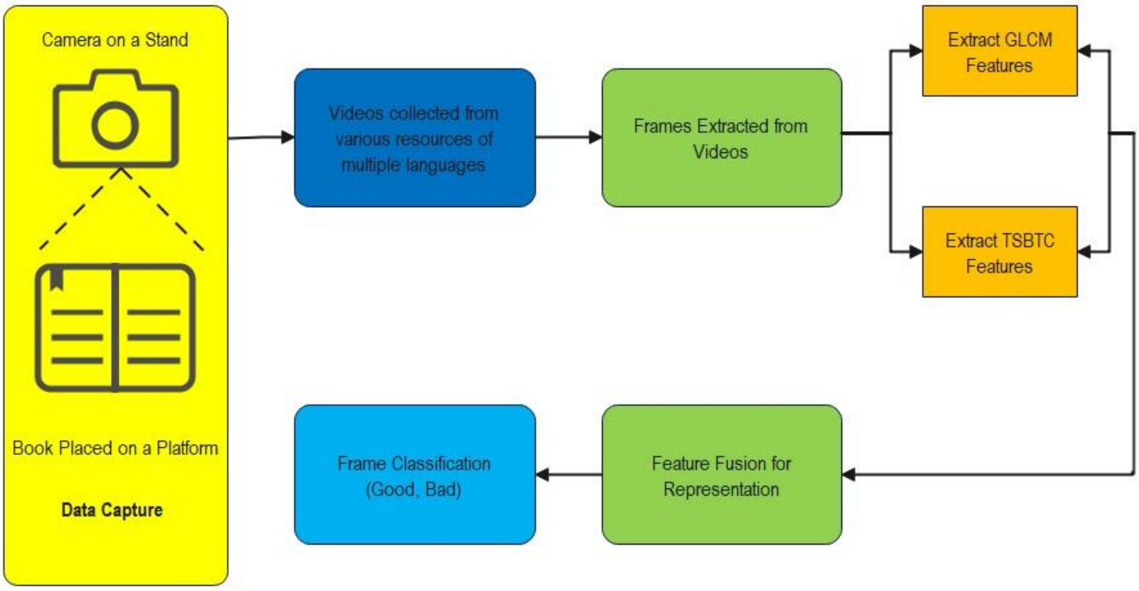
Keywords:
Video Document
Multilingual
Digitized Book
Information Retrieval
Digitization

ABSTRACT

In an age where preserving knowledge and information from books and documents is crucial, traditional manual scanning methods are tedious and error-prone. It involves a lot of human intervention and, as a result, sometimes results in erroneous digitization, which makes the downstream tasks, such as optical character recognition, difficult. Therefore, innovative techniques are required to be proposed that not only reduce human effort in terms of digitization but also give highly accurate results over the recently proposed state-of-the-art techniques. We proposed a novel computer vision-based algorithm that combines Gray-Level Co-occurrence Matrix (GLCM) features with Thepade's 10-ary texture features (TSBTC) for video frame classification. This hybrid approach significantly enhances frame selection accuracy, ensures high-quality digitization, and accommodates multiple languages and document types. We also proposed a dataset of 54,000 diverse images to demonstrate our algorithm's effectiveness in real-world scenarios and compare it to existing methods, making a valuable contribution to document digitization. The proposed dataset can be utilized for several document image analysis tasks.

***doi**: 10.5829/ije.2024.37.03c.11*

Graphical Abstract


Camera on a Stand
Book Placed on a Platform
Data Capture
Videos collected from various resources of multiple languages
Frames Extracted from Videos
Extract GLCM Features
Extract TSBTC Features
Feature Fusion for Representation
Frame Classification (Good, Bad)

*Corresponding author email: d18co005@coed.svnit.ac.in (G. Buddhawar)

Please cite this article as: Buddhawar G, Jariwala K, Chattopadhyay C. Enhancing Book and Document Digitization from Videos: A Feature Fusion-Based Approach. International Journal of Engineering, Transactions C: Aspects. 2024;37(03):538-45.

1. INTRODUCTION

Books and documents are valuable sources of information and knowledge that need to be preserved and accessed for various purposes. Digitizing books and documents are the processes of converting them from physical to digital form, which enables easier storage, retrieval, analysis, and dissemination of their content. However, digitizing books and documents is not a trivial task, as it involves several challenges, such as dealing with different languages, formats, layouts, fonts, and the quality of the original documents. Moreover, digitizing books and documents can also bring many benefits, such as enhancing their readability, accessibility, searchability, and usability. One of the common methods for digitizing books and documents is manual scanning, which is a tedious, costly, and error-prone process. A more efficient and convenient way is to record a video of the book and later extract the pages, but this requires a robust and accurate computer vision-based algorithm to perform the digitization. Therefore, developing efficient and effective methods for digitizing books and documents is an important and relevant problem.

Several methods have been proposed for digitizing books and documents from video recordings (1-3). These methods can be broadly classified into two categories: frame-based and video-based. Frame-based methods treat each frame of the video as an independent image and apply image processing techniques to extract the document content. Video-based methods exploit the temporal information of the video and use motion analysis techniques to track the document regions and capture the optimal frames. Frame-based methods are simpler and faster, but they are more sensitive to noise, blur, and distortion in the frames. Video-based methods are more robust and accurate, but they are more complex and computationally intensive. A detailed literature survey of these methods will be provided in section 2.

In this paper, we proposed a novel computer vision-based algorithm to automatically digitize books and documents from video recordings. Our algorithm is a feature-fusion based approach that combines two methods and achieves high accuracy in identifying the optimal frames to capture from the video, avoiding flipping instances, and ensuring high-quality digitization. Our algorithm can handle multiple languages and various types of documents, such as books, magazines, newspapers, and reports. We also introduce a new dataset of 5400 document images in different languages, which we used to train and evaluate our algorithm. We demonstrate the effectiveness of our algorithm on various real-world scenarios and compare it with existing methods.

Digitized documents can be used for various applications of information extraction and retrieval (4-8). Information extraction is the process of automatically extracting structured information from unstructured documents (8). Information retrieval is the process of finding relevant and useful information from large collections of documents (4). Some examples of information extraction and retrieval techniques are indexing, searching, summarizing, categorizing, and analyzing the document content (4-6). These techniques can help users find the information they need, as well as discover new insights and patterns from the document data. However, the quality and accuracy of information extraction and retrieval depend on the quality and accuracy of digitization (5, 6). Therefore, our algorithm can facilitate these tasks and improve their performance by providing high-quality digitization of books and documents from video recordings.

Based on the purpose mentioned in the previous paragraph, the current version of the paper makes the following key research contributions:

- A Dataset of 54000 Frames with their Ground Truth.
- A novel hybrid algorithm for video frame classification that combines GLCM features with Thepade's 10-ary texture features.
- Performance comparison with existing methods.

The rest of the paper is organized as follows: Section 2 provides a literature survey of the existing methods. Section 3 describes the details of the new dataset, while the proposed algorithm and its components in detail are given in section 4. Section 5 presents the experimental results, while section 6 concludes the paper and suggests some directions for future work.

2. LITERATURE SURVEY

Digitizing books and documents from video recordings are a challenging but valuable task, involving the extraction and reconstruction of textual and graphical content to enhance information accessibility. This task can be useful for various applications, such as digital libraries, e-learning, cultural heritage preservation, and personal archiving (7-9). However, it also faces several difficulties (10). To address these challenges, various methods have been proposed in the literature (11-13) that can be broadly classified into two categories: frame-based and video-based methods. Frame-based methods operate on individual or selected frames of the video, while video-based methods exploit the temporal information and redundancy of the video sequence.

Frame-based methods assume that the video frames contain sufficient information to reconstruct the document content and typically consist of three steps: frame selection, frame rectification, and frame stitching (14). Brown et al. (15) proposed a minimal solution for panoramic stitching based on homography estimation. Chhajed and Gargb (16) developed a smartphone-based system for book digitization using frame selection and rectification. Bouguet (17) implemented a camera calibration toolbox for frame rectification.

A two-level semi-supervised clustering method (3)incorporates labelled and unlabelled data simultaneously on various documents. A dictionary-independent method (6) for text language identification was proposed and tested on 31000 texts from 31 different languages. Fadaei (7) proposed a new dominant color descriptor to improve accuracy. The authors use a GAN-based resolution network (8) for face recognition. Parnak et al. (10) proposed the forgery detection mechanism for the extraction of features from the image using Benford's law as a benchmark and on CASIA datasets. Rashno and Fadaeib (14) also proposed an image restoration model based on the convex set feature of the images. Chhajed and Gargb (16) proposed the work related to binary images in based on a histogram-based decision tree that performs better with applications like steganography and watermarking. Charoqdouz and Hassanpour (18) proposed an approach for face images on documents.

Video-based methods are based on the idea that the video sequence contains redundant and complementary information that can be exploited to improve document reconstruction. These methods typically consist of four steps: video segmentation, video rectification, video mosaicing, and video enhancement (11). The advantages of video-based methods are their ability to handle camera motion and illumination variation, to reduce the number of frames required to cover the whole document, and to produce high-resolution and high-quality document images. However, they also have some disadvantages, such as being complex, time-consuming, and prone to errors in video segmentation and mosaicing. Charoqdouz and Hassanpour (18) proposed a robust book page extraction algorithm using boundary growing, and Chen et al. (19) developed a robust text detection method using edge-enhanced maximally stable extremal regions. Ulges et al. (20) presented a document capture system using stereo vision. Edge-based methods were also proposed by Kantarcıoğlu et al. (21), Firouzi et al. (22), Dixit and Shirdhonkar (23). We summarize their main features, advantages, and disadvantages in Table 1.

TABLE 1. Comparison of frame-based and video-based approaches

Method	Features	Advantages	Disadvantages
Frame-based	- Frame level - Consist of frame selection, frame rectification, and frame stitching	- Simple, efficient, - Robust to occlusion and page curling	- Require many frames - Sensitive to camera motion and illumination - Produce artifacts or inconsistencies
Video-based	- Video segmentation, - Video rectification, - Video mosaicing, - Video enhancement	- Handle motion and illumination - Require fewer frames - Produce high-resolution and quality images	- Complex and slow - Prone to errors in segmentation and mosaicing

Understanding the textual component of a natural scene (24) is useful for various applications, such as navigation and translations. It faces challenges such as complex background, low resolution, varying font, size, color, orientation, and style of text, occlusion, and distortion (25). The methods for this task can be classified into two categories: top-down and bottom-up methods. Top-down methods (26, 27) use global cues to locate and segment the text regions, while bottom-up methods use local cues to group and merge the text components or characters. Table 2 summarises some of the key research in this domain. Based on the literature survey, we found that the drawbacks of the methods discussed require an innovative way to digitise books and documents from video recordings. We suggest an innovative approach and overcome their drawbacks in the following part. Our approach and its experimental findings are described.

3. DATASET PREPARATION

The first step in this research is to prepare a dataset for book flip recognition. The proposed dataset has 54000 frames (.jpeg images) with their ground truth (good and bad frames marked). We have scanned these textbooks and documents in Hindi, English, and Marathi. Book Scanner with a mobile stand was used to perform the above operation. Frames are extracted from the recorded video using the ffmeg tool. Each image is 1024x768 pixels. All the images are saved as colour images with

TABLE 2. A summary of state-of-the-art algorithms, datasets, and performance metric available in the literature in the said domain

Sr.no	Algorithm	Dataset	Metric
1	Deep Learning-Based Optical Character Recognition (9)	MNIST	Character Recognition Accuracy
2	Document Layout Analysis (13)	COCO text	Word recognition accuracy
3	Document Analysis Systems (28)	IAM handwriting dataset	Text detection metric
4	Attention-Based Models for OCR (29)	Synthetic dataset	layout analysis metrics
5	Sparse Coding and Dictionary Learning (30)	Books magazines and newspaper	End to End evaluation metric

RGB channels. Figure 1 depicts some of the extracted frames, where the odd numbers ((i), (iii), and (v) represent the good frames, while the even number frames, i.e., (ii), (iv), and (vi), are the bad frames. The dataset is used to evaluate the proposed approach for improving the performance of machine learning classifiers and will be made publicly available for research purposes.

4. PROPOSED APPROACH

In this section, we are going to describe the proposed approach in detail. Figure 2 depicts a block diagram of the proposed approach. The proposed approach is discussed in the following sub-sections.

4. 1. TSBTC Algorithm in Book Digitization

Thepade's Sorted Block Truncation Coding (TSBTC) algorithm (12) performs complex procedures to identify the key frames from a set of video frames captured in a controlled or uncontrolled scenario. Initialising frame counts for "bad" and "good." A 500x9 grid-based multidimensional array is initialised. This array stores frame colour component average values. A video stream picture, 'A,' is processed during frame processing. Red (R), Green (G), and Blue (B) values are carefully retrieved from this picture. Colour components are key to the frame's aesthetic character. These extracted values are listed in ascending order. Then, mean values for each component are calculated and placed in the array.

A key feature of the TSBTC algorithm is its ability to distinguish between good and bad frames. A threshold-based mean colour value evaluation distinguishes this. If the estimated mean values for R, G, and B components fall below a threshold, the frame is considered "good" and the count is increased. However, if the mean values are above the threshold, the frame is considered "bad", and the count of "bad" frames is increased. In the TSBTC algorithm, the RGB got divided. Simultaneously, to find out the feature vector again, the frame values were divided into 10 parts. Per part, the centroid is calculated. Equations 1 to 4 in Table 3 show the criteria for the R value and it is the same for the remaining colors. The feature vector size is calculated using TSBTC 10-ary.

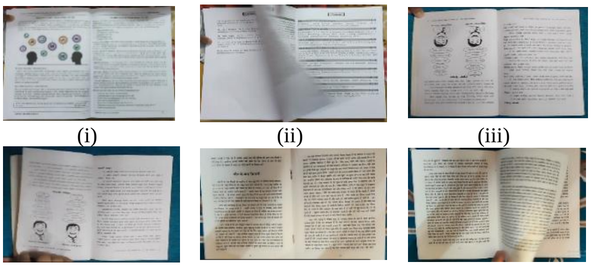

(i) (ii) (iii)
(iv) (v) (vi)

Figure 1. Sample images of the proposed dataset

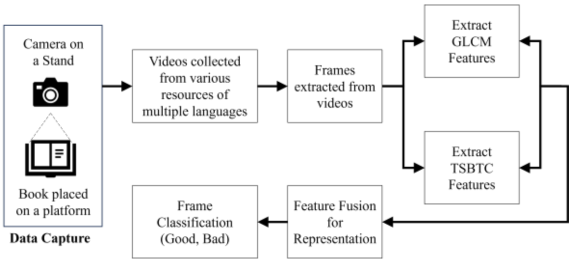

Camera on a Stand
Book placed on a platform
Data Capture
Videos collected from various resources of multiple languages
Frames extracted from videos
Extract GLCM Features
Extract TSBTC Features
Feature Fusion for Representation
Frame Classification (Good, Bad)

Figure 2. Block diagram of the proposed approach

4. 2. GLCM Feature Extraction in Book Digitization

The use of Gray-Level Co-occurrence Matrix (GLCM) feature extraction is pivotal in the proposed book digitization process. It effectively captures visual characteristics, especially during sequential page flipping recorded in curated videos. GLCM computation utilizes advanced statistical techniques to reveal spatial correlations among grayscale pixel intensity pairs in image 'e1'. It visually represents intricate patterns and textural details in digital book pages by capturing frequency distributions of pixel pairings at various distances and angles. The process includes preprocessing steps like converting 'q1' to grayscale 'w1' and resizing 'e1' to a 128x128-pixel resolution, improving efficiency and uniformity. Equations 5 to 8 in Table 4 demonstrate how to compute the GLCM features. The systematic transformation of dynamic video data into statistical co-occurrence matrices enables spatial correlation analysis among pixel intensities.

4. 3. Synergistic Fusion of TSBTC and GLCM Approaches in Book Digitization

The breakthrough in the domain of book digitization is achieved through the combination of TSBTC and GLCM feature extraction. Algorithm 1 presents the broader steps of the proposed hybrid approach.

The initialization of 'good' and 'bad' frame counters initiates this procedure. A multidimensional array, 'arr,' records mean color values from the Red (R), Green (G),

TABLE 3. Equations used to calculate red component values in the TSBTC technique.

Feature	Formula	
lR	$\left[\frac{4}{m \times n}\right] \times \sum_{i=1}^{\frac{m \times n}{4}} sortedSDR(i)$	(1)
muR	$\left[\frac{4}{m \times n}\right] \times \sum_{i=\frac{(m \times n)}{4}+1}^{\frac{m \times n}{4}} sortedSDR(i)$	(2)
mlR	$\left[\frac{4}{m \times n}\right] \times \sum_{i=(m \times n)/2+1}^{\frac{m \times n}{4}} sortedSDR(i)$	(3)
uR	$\left[\frac{4}{m \times n}\right] \times \sum_{i=(m \times n \times 3)/4+1}^{\frac{m \times n}{4}} sortedSDR(i)$	(4)

TABLE 4. Equations used to calculate the feature extraction using GLCM features

Feature	Formula	
Contrast	$\sum_{i=1}^{N}\sum_{j=1}^{N}(i-j)^2\, P(i,j)$	(5)
Entropy	$-\sum_{i=1}^{N}\sum_{j=1}^{N}\, P(i,j)\lg P(i,j)$	(6)
Correlation	$\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\,(i-\bar{x})(j-\bar{y})\,P(i,j)}{\sigma_x\,\sigma_y}$	(7)
Energy	$\sum_{i=1}^{N}\sum_{j=1}^{N} P(i,j)^2$	(8)

and Blue (B) components of each frame's color channels. The algorithm processes each frame iteratively, extracting colour values and calculating their mean. The frame is then classified based on predefined thresholds for each colour component. Frames falling within the defined thresholds are considered 'good,' while those exceeding the thresholds are labelled as 'poor.'

The inclusion of GLCM feature extraction enhances the algorithm's ability to identify complex textual patterns. This is accomplished by calculating the 'r1' gray-level co-occurrence matrix. During this phase, the algorithm processes the grayscale representation 'w1,' derived from the video dataset frame 'q1'. Resizing the image to a standard size of 128 by 128 pixels ensures computational consistency.

The 'r1' is generated through sophisticated statistical computations and stores the gray-level co-occurrence matrix from w1. This matrix represents the frequency distribution of pixel intensity pairings within the 'e1' image and captures textural patterns. This characterizes pixel relationships in terms of spatial proximity and orientation, yielding a nuanced understanding of the textual complexities of digitized book pages.

The amalgamation of TSBTC and GLCM enables the algorithm not only to classify frames based on color attributes but also to explore intricate textural nuances. By uniting these techniques, the algorithm not only enhances the preservation of book content but also reveals new insights into the textual and visual essence of the original material.

4. 3. 1. The Proposed Hybrid Approach

The algorithm merges TSBTC and GLCM, their individual attributes combining synergistically. Initially, TSBTC classifies frames into 'excellent' or 'poor' quality using color attributes. Concurrently, GLCM's texture analysis computes the gray-level co-occurrence matrix 'r1,' capturing nuanced textural patterns in frames. Each step contributes to this orderly process by learning from its predecessor. Our earlier rationale serves as the foundation for this handcrafted feature engineering, which embodies a holistic approach that transcends particular methodologies by combining the best of both the features. Such feature engineering allows the proposed algorithm to capture both visual aspects and the essence of textural patterns in digitized book pages, aligning seamlessly with our study's core goals.

Algorithm 1: Hybrid Approach

```
# Step 1: Extract intensity values and color components
   image = load_image(image_path)
   color_comps = extract _components(image)
# Step 2: Create feature vectors
   feature_vec = create_feature_vec(color_comps)
# Step 3: Sort feature vectors
   sorted_vec = sort_feature_vect(feature_vec)
# Step 4: Divide sorted feature vectors
   parts = divide_feature_vecs(sorted_vec,
desired_variation="ternary")
# Step 5: Compute representative values
   repr_val = compute_representative_values(parts)
# Step 6: Create comprehensive feature vector
   feature_vec = create_comp_feature_vec(repr_val)
# Step 7: Compute GLCM features
   GLCM_features = compute_GLCM (feature_vec)
# Step 8: Compute texture measures
   txtr_msr = comp_txtr_meas(GLCM_features)
# Step 9: Create GLCM features array
   GLCM_array = create_GLCM_array(txtr_msr)
# Step 10: Extract key frames
   key_frames = extract_key_frames(GLCM_array)
   return key_frames
```

5. EXPERIMENTAL RESULTS AND DISCUSSION

The findings of this study provide an approach for finding the representative frame of the Open Page Image (OPI) and removing unwanted frames from the video stream of the book being flipped. The code was executed on a Jupiter notebook, and the use of GPUs is done extensively. Hyperparameters in GLCM, like the number of gray levels and offset or distance, are used, while in TSBTC, the intensity value is used.

5. 1. Quantitative Results

In this section, we present the quantitative results of our comprehensive analysis, depicting the performance of various approaches across different classifiers. It is very common to use the F1 measure for binary classification. The mathematical formulation of the performance metric is given in the following set of equations:

$$Accuracy = \left(\frac{TP+TN}{TP+TN+FP+FN}\right) * 100 \qquad (9)$$

$$F_1 = \left(\frac{TP}{TP+0.5(FP+FN)}\right) \qquad (10)$$

where,
TP => True Positives (Good Frame predicted as Good)
TN => True Negatives (Bad Frames predicted as Bad).
FP => False Positives (Bad Frames predicted as Good)
FN => False Negatives (Good Frames predicted as Bad)

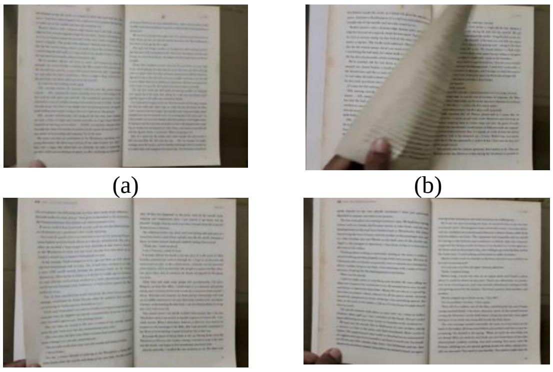
TABLE 5. Performance Comparison of the accuracy of all digitization techniques with various classifiers

Classifier	**Features**		
	TSBTC (%)	**GLCM (%)**	**Hybrid (%)**
ZeroR	**85.32**	**77.66**	**58.21**
Attribute Select Classifier	96.17	99.79	100
Bayes Net	95.31	99.68	99.88
Naïve Bayes	94.31	95.76	77.13
Naïve Bayes Multinomial	77.66	94.62	57.99
Logistic	88.23	99.06	87.21

The results are presented in Table 5, with approaches represented as columns and classifiers as rows. Notably, the best performing values, highlighted in *italics*, offer valuable insights into the efficacy of each approach. Within this context, our findings shed light on the enhanced performance of the hybrid model, further affirming its superiority. Additionally, we have experimented with Convolutional Neural Network (CNN) based on the LeNet-5 design (18) to measure the performance. There were 2,500 labelled good frames and 12,750 bad frames used for training set. The training process aimed to minimize the binary cross-entropy loss, and model accuracy was a key evaluation metric. A comparison between the performance of CNN and the proposed method is shown in Table 6.

5. 2. Qualitative Results

The qualitative results of our study are depicted in Figure 3, where a 2x2 grid format encapsulates the results at different stages of processing. Each quadrant of the grid presents a distinct image, showing a specific stage. Figure 3(a) portrays the 'good' frame, embodying the desired outcome. Figure 3(b) provides a glimpse of a 'half-flipped' frame, while Figure 3(c) delves into a 'twisted' representation that signifies an unwanted deviation. Lastly, Figure 3(d) resurfaces the ideal 'good' frame. Notably, the best-performing result will be the one where a page is completely flipped open and the textual portion of the page is completely visible, thereby making the visual comprehension of effectiveness seamless. It can be observed that Figures 3(a) and 3(d) satisfy these criteria. Although in both cases a part of the thumb is visible, however, it is not obstructing the text. To be more critical, Figure 3(a) can be considered as best among these four frames, as the portion of the thumb is the minimum in that frame. This shows that the proposed approach is able to achieve the desired result efficiently.



(a) (b)

(c) (d)

Figure 3. Qualitative results of the key frame extraction using the proposed hybrid approach

TABLE 6. Performance Comparison between the proposed model and CNN

Sr. No	**Approach**	**% Accuracy**
1	CNN	95%
2	Proposed Method	99%

5. 3. Applications and Limitations

The proposed work has valuable real-life applications in digitizing books and documents from videos, benefiting fields such as education, research, and preservation. Examples of potential downstream real-life applications of the proposed include digitizing historical documents, archival research, education and eLearning, content indexing, work for the visually impaired and protecting fragile documents. However, it faces limitations related to video quality, text recognition accuracy, language support, and legal considerations. Examples of such limitations are the quality of source videos, variability of text, distorted images, multiple language text, copyrights and permissions from various book authors, managing a large digital document, the privacy of data, and digitization process. Addressing these limitations and ensuring the quality and ethical use of digitized content are essential in the future.

6. CONCLUSION AND FUTURE SCOPE

This paper presents a computer vision-based system that identifies key frames from video of a book being flipped under the field of view of a camera. The system fuses two algorithms (TSBTC and GLCM) to select the best frames and avoid page flipping, achieving high accuracy. The system supports many languages and document types and is evaluated on a new dataset of 54,000 images. The system outperforms existing methods and opens up new possibilities for digitizing textual resources. Future work can concentrate on refining the algorithm's precision, optimising its computational efficiency, and integrating advanced techniques such as deep learning to further enhance its accuracy. By improving the algorithm, we foresee a future in which digitization becomes more accessible, efficient, and accurate.

8. REFERENCES

1. Obiora KU, Okeke IE, Onwurah B. Digitization of library resources in university libraries: A practical approach, challenges and prospects. 2015. 10.1109/ETTLIS.2015.7048210
2. Azim N, Mat Yatin S, Jensonray R, Ayub Mansor S. Digitization of records and archives: Issues and Concerns. International Journal of Academic Research in Business and social sciences. 2018;8(9):170-8. 10.6007/IJARBSS/v8-i9/4582
3. Sadjadi S, Mashayekhi H, Hassanpour H. A two-level semi-supervised clustering technique for news articles. International Journal of Engineering, Transactions C: Aspects. 2021;34(12):2648-57. https://doi.org/10.5829/ije.2021.34.12C.10
4. HS C, Shenoy MK. Advanced text documents information retrieval system for search services. Cogent Engineering. 2020;7(1):1856467. 10.1080/23311916.2020.1856467
5. Lillis D, Scanlon M, editors. On the benefits of information retrieval and information extraction techniques applied to digital forensics. Advanced Multimedia and Ubiquitous Engineering: FutureTech & MUE; 2016: Springer. 10.1007/978-981-10-1536-6_83
6. Hassanpour H, AlyanNezhadi M, Mohammadi M. A Signal Processing Method for Text Language Identification. International Journal of Engineering, Transactions C: Aspects. 2021;34(6):1413-8. 10.5829/ije.2021.34.06c.04
7. Fadaei S. New dominant color descriptor features based on weighting of more informative pixels using suitable masks for content-based image retrieval. International Journal of Engineering, Transactions B: Applications. 2022;35(8):1457-67. 10.5829/ije.2022.35.08b.01
8. Shahbakhsh MB, Hassanpour H. Empowering face recognition methods using a gan-based single image super-resolution network. International Journal of Engineering, Transactions A: Basics. 2022;35(10):1858-66. 10.5829/ije.2022.35.10a.05
9. Buddhawar G, Jariwala KN, Chattopadhyay C, editors. Some Aspects of Text Recognition from Video Document in Education 4.0. 2021 Emerging Trends in Industry 40 (ETI 40); 2021: IEEE. 10.1109/ETI4.051663.2021.9619427
10. Parnak A, Baleghi Damavandi Y, Kazemitabar S. A Novel Image Splicing Detection Algorithm Based on Generalized and Traditional Benford’s Law. International Journal of Engineering, A: Basics; 2022;35(4):626-34. 10.5829/ije.2022.35.04a.02
11. Kumar V. Region completion in a texture using multiresolution transforms. International Journal of Engineering, Transactions B: Applications; 2014;27(5):747-56. 10.5829/idosi.ije.2014.27.05b.10
12. Kekre H, Thepade SD, Lohar AT, editors. Image retrieval using block truncation coding extended to color clumps. 2013 International Conference on Advances in Technology and Engineering (ICATE); 2013: IEEE. 10.1109/ICAdTE.2013.6524769
13. Binmakhashen GM, Mahmoud SA. Document layout analysis: a comprehensive survey. ACM Computing Surveys (CSUR). 2019;52(6):1-36. 10.1145/3355610
14. Rashno A, Fadaei S. Image restoration by projection onto convex sets with particle swarm parameter optimization. International Journal of Engineering, Transactions B: Applications; 2023;36(2):398-407. 10.5829/ije.2023.36.02b.18
15. Brown M, Hartley RI, Nistér D, editors. Minimal solutions for panoramic stitching. 2007 IEEE conference on computer vision and pattern recognition; 2007: IEEE. 10.1109/CVPR.2007.383082
16. Chhajed G, Garg B. Novel Scheme for Data Hiding in Binary Images using Cover Pattern Histogram. International Journal of Engineering, Transactions B: Applications; 2023;36(11):2124-36. 10.5829/ije.2023.36.11b.16
17. Bouguet J-Y. Camera calibration toolbox for matlab. http://www vision caltech edu/bouguetj/calib_doc/. 2004. 10.22002/D1.20164
18. Charoqdouz E, Hassanpour H. Feature Extraction from Several Angular Faces Using a Deep Learning Based Fusion Technique for Face Recognition. International Journal of Engineering, Transactions B: Applications; 2023;36(8):1548-55. 10.5829/ije.2023.36.08b.14
19. Chen H, Tsai SS, Schroth G, Chen DM, Grzeszczuk R, Girod B, editors. Robust text detection in natural images with edge-enhanced maximally stable extremal regions. 2011 18th IEEE international conference on image processing; 2011: IEEE. 10.1109/ICIP.2011.6116200
20. Ulges A, Lampert CH, Breuel T, editors. Document capture using stereo vision. Proceedings of the 2004 ACM symposium on Document engineering; 2004. 10.1145/1030397.1030434
21. Kantarcıoğlu M, Xi B, Clifton C. Classifier evaluation and attribute selection against active adversaries. Data Mining and Knowledge Discovery. 2011;22:291-335. 10.1007/s10618-010-0197-3
22. Firouzi M, Fadaei S, Rashno A. A new framework for canny edge detector in hexagonal lattice. International Journal of Engineering, Transactions B: Applications; 2022;35(8):1588-98. 10.5829/IJE.2022.35.08B.15
23. Dixit U, Shirdhonkar M. An Improved Fingerprint-based Document Image Retrieval using Multi-resolution Histogram of Oriented Gradient Features. International Journal of Engineering, A: Basics; 2022;35(4):750-9. 10.5829/IJE.2022.35.04A.15
24. Mishra A, Alahari K, Jawahar C, editors. Top-down and bottom-up cues for scene text recognition. 2012 IEEE conference on computer vision and pattern recognition; 2012: IEEE. 10.1109/CVPR.2012.6247990
25. Roy S, Roy PP, Shivakumara P, Louloudis G, Tan CL, Pal U, editors. HMM-based multi oriented text recognition in natural scene image. 2013 2nd IAPR Asian Conference on Pattern Recognition; 2013: IEEE. 10.1109/ACPR.2013.60
26. Shivakumara P, Bhowmick S, Su B, Tan CL, Pal U, editors. A new gradient based character segmentation method for video text recognition. 2011 International conference on document analysis and recognition; 2011: IEEE. 10.1109/ICDAR.2011.34
27. Singh M, Kaur A, editors. An efficient hybrid scheme for key frame extraction and text localization in video. 2015 International conference on advances in computing, communications and informatics (ICACCI); 2015: IEEE. 10.1109/ICACCI.2015.7275784
28. Hamdan M, Cheriet M. ResneSt-Transformer: Joint attention segmentation-free for end-to-end handwriting paragraph recognition model. Array. 2023:100300. 10.1016/j.array.2023.100300
29. Xiao Z, Nie Z, Song C, Chronopoulos AT. An extended attention mechanism for scene text recognition. Expert Systems with Applications. 2022;203:117377. 10.1016/j.eswa.2022.117377
30. Gao F, Deng X, Xu M, Xu J, Dragotti PL. Multi-modal convolutional dictionary learning. IEEE Transactions on Image Processing. 2022;31:1325-39. 10.1109/TIP.2022.3141251

Persian Abstract

چکیده

در عصری که حفظ دانش و اطلاعات از کتاب ها و اسناد بسیار مهم است، روش های اسکن دستی سنتی خسته کننده و مستعد خطا هستند. این کار مستلزم مداخلات انسانی زیادی است و در نتیجه گاهی اوقات منجر به دیجیتالی شدن اشتباه می شود که کارهای پایین دستی مانند تشخیص کاراکترهای نوری را دشوار می کند. بنابراین، باید تکنیک‌های نوآورانه‌ای پیشنهاد شود که نه تنها تلاش انسان را از نظر دیجیتالی کردن کاهش می‌دهد، بلکه نتایج بسیار دقیقی را نسبت به تکنیک‌های پیشرفته اخیر ارائه می‌دهد. ما یک الگوریتم جدید مبتنی بر بینایی کامپیوتری را پیشنهاد کردیم که ویژگی‌های ماتریس هم‌وضعیت سطح خاکستری (GLCM) را با ویژگی‌های بافت ۱۰-اری Thepade (TSBTC) برای طبقه‌بندی فریم‌های ویدئویی ترکیب می‌کند. این رویکرد ترکیبی به طور قابل توجهی دقت انتخاب فریم را افزایش می دهد، دیجیتالی شدن با کیفیت بالا را تضمین می کند و چندین زبان و انواع سند را در خود جای می دهد. ما همچنین مجموعه داده‌ای از ۵۴۰۰۰ تصویر متنوع را برای نشان دادن اثربخشی الگوریتم خود در سناریوهای دنیای واقعی و مقایسه آن با روش‌های موجود پیشنهاد کرده‌ایم که سهم ارزشمندی در دیجیتال‌سازی اسناد دارد. مجموعه داده پیشنهادی را می توان برای چندین کار تجزیه و تحلیل تصویر سند مورد استفاده قرار داد.