JOURNAL OF DATA SCIENCE | Vol.2024:66
eISSN:2805-5160

Exploring Text Recognition Segmentation and Detection in Natural Scene Images

Wydyanto[1] and Maria Ulfa[2]

Faculty Science Technology Universitas Bina Darma Palembang, Sumatera Selatan 30264 Indonesia

Email: [1]wydyanto@binadarma.ac.id, [2]maria.ulfa@binadarma.ac.id

Abstract

Identification, segmentation, and recognition of fonts from real-world images are major challenges in computer vision, particularly due to subtle differences in font shapes, lighting, and backgrounds. This paper aims to provide a comprehensive review of the latest algorithms for text detection, segmentation, and recognition from natural scene images. A variety of techniques are assessed for their use in natural settings, including deep learning-based methods, region proposal, and feature-based detection. There is additional discussion of the difficulties of managing changes in text properties such as font type, size, orientation, and noise and occlusion disruptions. This survey also looks at preprocessing techniques like filtering and illumination normalization that are meant to increase the accuracy of text detection. In light of the findings of the literature analysis, this study concludes that the combination of adaptive segmentation techniques with deep learning-based recognition models offers promising performance in text recognition in natural scenery images. This survey provides a foundation for the development of more effective and robust methods for future applications in the fields of image processing and computer vision.

Keywords

Text detection, text segmentation, text recognition, natural scenery images, computer vision.

Introduction

Text detection and text segmentation are important techniques in image processing to identify and separate text from the background of an image. Text localization in images involves locating and circumscribing text occurrences with tight rectangular boxes, while text segmentation involves segmenting text lines into binary images with black characters on white background. Horizontal text is primarily considered in these processes, as it accounts for the majority of text occurrences. (Lienhart & Wernicke, 2002). Text detection identifies text areas as outer regions of an image, and during the identification step of the system, text detection recognizes text areas as outer components of the image. Text detection and text recognition are two processes performed to obtain text information. (Jung, Kim, & Jain, 2004).

Submission: 25 October 2024; **Acceptance:** 21 December 2024

Optical Character Recognition (OCR) is a technology that allows the system to recognize characters in an image and convert them into text that can be understood. (Saric, 2017) (Mittal et al., 2022). The performance of OCR systems depends on the quality of input documents, with better results seen in more constrained inputs. (Karpinski et al., 2019)(R. Jain & Gianchandani, 2019). However, OCR machines still struggle with unconstrained handwriting compared to humans, but advancements are continuously improving the technology. "License plate recognition systems are very useful in various applications such as vehicle license plate recognition and handwriting recognition. They can alert officers when a vehicle of interest is passed and record time and GPS coordinates (Binmakhashen & Mahmoud, 2019)(Anagnostopoulos, Anagnostopoulos, Loumos, & Kayafas, 2006). In a vehicle license plate recognition application, the OCR system can identify the characters on the license plate and automatically output information. License plate extraction methods involve identifying features like color, texture, and character presence to locate the license plate in an image. Various techniques such as edge detection, neural networks, and statistical analysis are used for license plate segmentation and extraction based on different features (Binmakhashen & Mahmoud, 2019) .

Thus, the vehicle identification process can be carried out quickly and accurately without the need for human intervention. "In cases where the OCR system may have difficulty recognizing characters on license plates that are too dirty or damaged, techniques such as object detection, image processing, and pattern recognition can be used to improve recognition. Challenges in license plate detection and recognition arise from variations in plate types, environments, and characters, as well as factors like occlusion and inclination (Binmakhashen & Mahmoud, 2019)(Arshad, Abidin, & Obeidy, 2017). Anomaly-driven identification techniques in vehicles can help identify new breaches and enhance cybersecurity measures to prevent potential security risks in the vehicle identification process (Koscher et al., 2010).

The continuous development of text recognition technology is expected to provide better solutions for managing information contained in images efficiently. This involves techniques for compressing and indexing textual images, differentiation between text and images in documents, and the implementation of various compression models(Palani, Venkatalakshmi, & Venkatraman, 2014). One of the solutions being developed is the use of deep learning technology in OCR systems."Facial recognition technology can accurately predict political orientation from naturalistic facial images with a 72% success rate, outperforming human accuracy and personality questionnaires. The accuracy remains high even when controlling for age, gender, and ethnicity. This technology has critical implications for privacy and civil liberties (Kosinski, 2021) (Waelen, 2023). With this technological breakthrough, it is hoped that the vehicle identification process can be carried out more efficiently and accurately, thereby minimizing potential security risks. Extracting content from images is very challenging due to image quality and background noise. There are various types of images that have text as part of them with backgrounds, such as document images, landscape images, and digital birth images. Where scenes are often filmed by the camera.

The applications of license plate detection include vehicle detection and license plate localization for Intelligent Transportation Systems (ITS) and security applications. License plate detection is achieved through various methods such as YOLO V3, DCNN, Faster-RCNN, deep learning, and character segmentation. The goal is to accurately locate license plates in complex scenarios with different illuminations and backgrounds (Mahmood et al., 2022). Given the above-mentioned issues and the characteristics of landscape text, it is more challenging to find and identify text in natural scenes compared to text in paper and digital

photographs. This document is formatted as follows: Section 1 provides the background of this paper. Section 2 describes the different types of photos that contain text and the features to consider when trying to extract text from them. Section 3 discusses several methods for finding, classifying, and identifying words in landscape photos. Section 4 lists Methods for Detecting Landscape Text. Section 5 Comparison of text detection methods based on extreme regions

Results and Discussion

Text images may be classified into three types.
A) Document image: Text images that fall into the category of image documents usually contain well-structured and organized text, such as official documents, letters, or reports. This type of document generally has a consistent format and is easy to process automatically.

B) Landscape picture: Images of text in this category usually contain text within a more complex visual context, such as posters, billboards, or street signs. The process of text detection and segmentation in this landscape image can become more challenging due to the various visual elements surrounding the text.

C) Unstructured image: Scene text editing involves replacing text in an image while maintaining its realistic appearance, facing challenges like text style transfer and background texture retention. The proposed style retention network (SRNet) breaks down the job into sub networks for text conversion, background inpainting, and fusion. (L. Wu et al., 2019). The process of text detection and segmentation in this unstructured image can become more complicated because the text is often mixed with various other visual elements. The document is nothing more than the document's picture format. Scanners and phone cameras may create document images that include text and graphics.

The correction of contrast and geometric distortions in camera-captured document images goes beyond what is needed for scanned documents. OCR systems are available for almost every language and script, with efforts to convert all existing books to electronic files. Historical documents receive less funding but research aims to improve accuracy, speed, and scope of conversion. The digitization of medical records highlights the need for integrated document recognition approaches. (Nagy, 2016). The images are converted from paper-based documents into image formats for electronic reading by scanning them ke dalam sistem sebagai foto digital dalam format TIFF. These photos are known as scanned multiple documents. (Lerum, Karlsen, & Faxvaag, 2003)

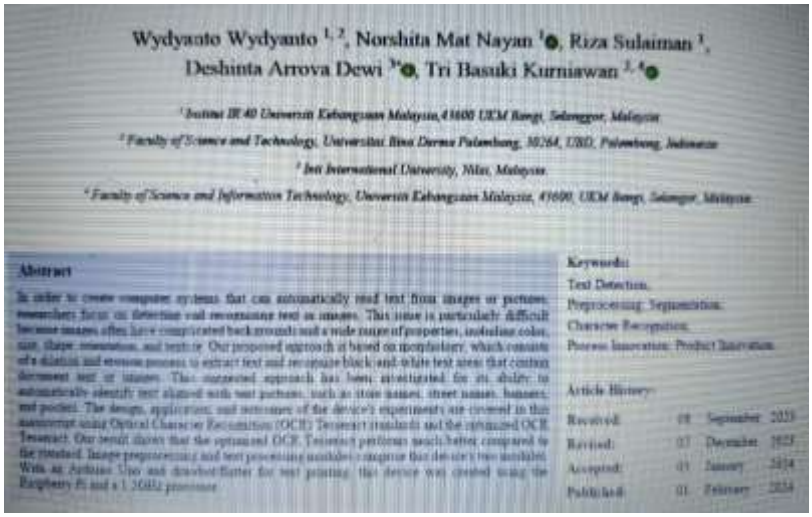
Wydyanto Wydyanto [1, 2], Norshita Mat Nayan [1], Riza Sulaiman [1], Deshinta Arrova Dewi [3], Tri Basuki Kurniawan [3, 4]

Fig. 1 Document Text Image (Wydyanto, Nayan, Sulaiman, Dewi, & Kurniawan, 2024)

To achieve this goal, we will use a dataset that includes a wide variety of landscape photographs obtained from various sources. Each image will be tested using the methodology we have developed, and the results will be analyzed in depth. Thus, we hope to demonstrate that our proposed approach is capable of addressing the challenges in text detection, segmentation, and recognition in scenic images better than existing methods. "Object detection methods are crucial for applications like navigation, object recognition, and area mapping. These methods, especially 3D object detection, provide detailed information about an object's size and location, which is essential for tasks like path planning and collision avoidance (Danial, Rosli, M.Zarar, & Jean-Marc, 2011)(Kuipers & Levitt, 1988) (Aqel, Marhaban, Saripan, & Ismail, 2016). These photos are very difficult to recognize and detect due to their complex backgrounds, which include text of varying sizes, styles, and alignments. Furthermore, perspective and lighting distortions affect the scene text. Current OCR algorithms cannot detect misaligned lines of text or complex background noise.

.

Fig. 2 Scene Text Image

Digitally born image has high clarity and sharpness, and is not affected by background noise or perspective distortion. This makes the process of text detection and recognition easier and more accurate. "Recognizing text within digitally born images can be challenging due to various handwriting styles and character-touching problems, especially in scene text with diverse styles and complex backgrounds (Chen, Jin, Zhu, Luo, & Wang, 2021). Text recognition is more difficult in different languages because text features vary (Chen et al., 2021). Text recognition in the wild involves fundamental problems like text localization, verification, detection, segmentation, and recognition, each with its unique challenges (Chen et al., 2021). Multilingual datasets, such as those including Chinese text, present additional difficulties in text recognition due to the unique characteristics of Chinese characters. (Chen et al., 2021).

Therefore, the development of text recognition techniques that can handle various types of images becomes very important in improving the performance of OCR systems for born-digital images. For example, when recognizing text in images of graffiti that often have unconventional writing styles, the OCR system must be able to overcome perspective distortion and background noise so that the text can be accurately recognized. "The development of adaptive and sophisticated text recognition techniques is crucial for enhancing OCR systems. OCR, an AI method, is used for identifying machine and handwritten text and digits from various images containing text. Researchers and companies are actively working on developing OCR solutions, with applications in automation, workflow enhancement, healthcare, finance, and more (Yin, Pei, Zhang, & Hao, 2015) (P. H. Jain et al., 2023). Advanced AI techniques have led to significant developments in OCR technologies, but challenges remain in creating OCR with human-like abilities, especially for localized or personalized handwritten text (P. H. Jain et al., 2023).

Adaptive OCR solutions can improve accuracy with individual users and clients in various sectors, including healthcare, banking, insurance, and postal services. The quality of OCR-generated text can impact downstream natural language processing tasks, emphasizing the importance of high-quality OCR above 90 percent (Miri, Abràmoff, Kwon, Sonka, & Garvin, 2017). Future research directions include exploring transformer-based applications for improving OCR solutions for handwritten text and integrating OCR with NLP solutions for sentiment analysis and socioeconomic modeling. (Parth et al., 2023). Digital photos have more disadvantages than document and landscape images, includes more complicated foreground/background, poor resolution, compression loss, and very soft edges. As a result, it is difficult to separate text from the background during text extraction..

Fig. 3 gambar terkahir digital

Heterogeneous text image:
"The sources discuss the technological transformation of images in digital culture and the implications for photography, visual culture, and interactive pornography. They also mention the use of relevance feedback in image retrieval and the challenges in image segmentation and shape matching (Jung et al., 2004). However, there is no specific mention of scene image text, text images, document image text, and digital images in the text provided. (Lister, 2013) (Datta, Joshi, Li, & Wang, 2008).

For example, in heterogeneous text, text detection, localization, and extraction are critical steps in processing images with printed text. Text detection identifies the presence of text, localization pinpoints the location of text and offers bounding boxes, and extraction extracts text from the background. This extracted text image is then enhanced and converted to plain text using OCR technology (Jung et al., 2004). (Jung et al., 2004)(Ye & Doermann, 2015) The process of identifying and extracting text from this type of image can become more complex due to variations in the type and placement of text within those images. Therefore, the development of text recognition methods that can handle various types of images becomes very important in the field of image processing and computer vision.

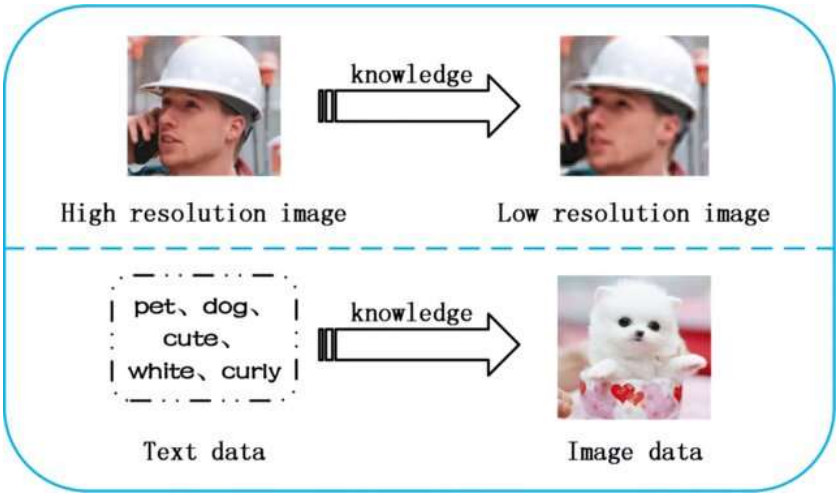

Fig. 4 Image of Heterogeneous Text

These photos are available in a variety of formats, including digital birth photos, document photos, text photos, and landscape photos. (Jung et al., 2004)

Factors to Consider When Detecting Text from Images
1. font type, thickness (stroke width), and size (height, width);
2. Coordinates (X, Y) or position on the image;
3. The background as well as the color and texture of the foreground;
4. Camera position that can cause geometric distortion;
5. Orientation;
6. Coordination;
7. Symbols, integers, and non-text content;
8. Explanation;
9. Language;
10. Resolution;
11. Contrast;
12. Blur and noise.

The content provided in the sources discusses the extraction of text information from images and videos, which involves various processes such as detection, localization, tracking, and recognition of text. The taxonomy of relationships between images and text focuses on how images and text interact, identifying 49 relationships grouped into three categories based on the closeness of the conceptual relationship between image and text. Both sources highlight the importance of understanding the relationship between text and images for effective communication and document design (Jung et al., 2004) (Marsh & White, 2003):
a) Scene text is text that appears in an image but does not reflect the main picture content
b) Artificial text, or text created separately from an image

unlike scene text, is an important source of information as well as an object of indexing and retrieval. It is a very difficult challenge to identify, categorize, recognize, and extract text from photos with accuracy and reliability of image content..

Previous research has attempted to address this challenge using various approaches, such as deep learning, image segmentation, and image processing. However, there are still shortcomings in text recognition on images with complex backgrounds. This project intends to develop a novel method that is more effective in recognizing text in photographs with complicated and diverse backgrounds. The proposed method will utilize deep learning techniques and image segmentation to improve text recognition accuracy, especially in difficult conditions. Many ways for text detection from landscape pictures have been developed in recent years; in this section, we will briefly discuss various techniques, as well as their benefits and drawbacks for text detection, text segmentation, and character recognition.

Methodology

Over the past few years, many techniques for text detection from landscape images have been developed; in this section, we will briefly discuss these techniques, along with their advantages and disadvantages, for text detection, text segmentation, and character recognition.

1) Window-based technique

The Sliding Window-based Method is one of the basic techniques often used in computer vision to detect objects, including text, in images. This approach involves moving a fixed-size window across the entire image to examine every potential location and scale where the target object might be, also known as the area-based method. The method described in the query uses a sliding window approach for text detection in images and then utilizes machine learning techniques for text identification. This method combines MSERs and sliding-window based methods to handle complex text information efficiently (Huang, Qiao, & Tang, 2014) (Ye, Huang, Gao, & Zhao, 2005). This method is slow because the photos have to be processed at different scales. This approach limits the text search to a rectangular portion of the image. This reduces the number of portions checked for text.

2) Component-based connected method (CC), (CC),

This method uses a modular approach in software development. In this method, a large system is divided into smaller interconnected components. Each component in a system-of-systems has specific tasks and functions that contribute to achieving the overall goals of the system. Components can interact with each other to ensure the successful operation of the entire system (Maier, 1999). By using a connected component-based method, developers can accelerate the development process, increase flexibility, and facilitate overall system maintenance. Text extraction from natural scene images involves three stages: detection and localization, text

enhancement and segmentation, and optical character recognition (OCR). The challenges include variations in font size, color, text alignment, illumination, and reflections. Algorithms for text extraction are categorized based on methodology, performance, and execution time. No single method can address all variations in scene texts (Zhang, Zhao, Song, & Guo, 2013). The shortcomings are evident in some high-light nature photos and very poor text contrast. There are three types of techniques for extracting related elements from images.

3) Edge-based Method

The method mentioned is based on factors such as the edge of the character, which is a reliable feature of the text regardless of color/intensity, layout, orientation. Unsupervised text detection techniques include edge-based methods which exploit the high contrast between text and its background (Mirza, Zeshan, Atif, & Siddiqi, 2020).

Method based on Color

Color clustering is performed by grouping pixels with the same or similar colors and forming candidate regions. The candidate area is then divided into subregions with similar colors, and the average color for each subregion is calculated and assigned to each pixel of the region (Member & Sun, 2000).

Method based on Edge and Color Combination

The method of combining Method 1 and Method 2 to detect edges and text colors has achieved better results by leveraging the convolutional structure of CNNs to process the complete image at once rather than applying CNN classifiers to every proposed cropped character (Thesis et al., 2014).

Hybrid method

This method uses an area detector to identify text candidates, and the corresponding components are recovered as character candidates through local binarization. Non-characters are then removed using the Conditional Random Fields model, which allows characters to be grouped together as text. The CRF model separates text and non-text components in natural landscape photos for segmentation and analysis (X. Wang, Song, Zhang, & Xin, 2015) (Y. Wang, Shi, Xiao, Wang, & Qi, 2018). Disadvantage difficult to categorize text.

Texture-based Method

[1]: This method relates to the text area as a special texture. "Text extraction from natural scene images involves various challenges such as changes in camera angles causing perspective distortions that affect extraction performance. Different methods like edge-based, texture-based, and connected component-based approaches are used for text detection and localization. The performance measure used is the detection rate, which is the ratio of detected text to all text contained in the image. The proposed statistical unified approach for text extraction from hybrid textual images uses carefully selected features to discriminate between text and non-text regions, showing promising results compared to other methods (Zhang et al., 2013) (Liu, Zhang, & Lu, 2008). In this method, a text region detector is designed based on texture. This can be used to estimate the probability of text position and scale and then analyzed into text regions or not.

D) Angle-Based Method

The use of corner points for text detection was prompted by the observation that most characters in text have multiple corner points. Corner points are regarded as key features for

text detection because of their stability and discriminative qualities. (Zhao et al., 2011). This method is used to describe the text region formed by corner points using several discriminative features. Angle-based methods are faster compared to texture-based methods, but their performance is less satisfactory (Guo et al., 2019) (Joshi & Patil, 2019)

E) Method based on Semi-automatic Ground Truth Generation

Texts with different orientations and languages are included in the semi-automatic ground truth generation system for text detection and recognition. If the automated approach produces inaccurate findings, the technology allows the user to manually correct the underlying truth (Kim, 2015). Eleven word-level attributes are used in this method to assess its performance: region, content, script type, orientation information, text type (caption/scene), text condition (distortion or no distortion), start frame, end frame, line index, word index, bounding box coordinate values, and area.

F) The method used to request images using scene text

Text verification and localization are performed in two steps. Preprocessing and development of candidate text regions are the two stages of text localization. Image blocks are located using morphological processes such as erosion, dilation, opening, and closing. Shah, Mehta, Roy, Khachane, and Mody (2018) (Gupta, Marina, Sethi, Asharif, & Khosravy, 2017). Candidate letters are then generated by applying stroke width transformation with some adjustments. To identify lines of text, these letters are concatenated, and the lines are then divided into words.

During the text verification stage, a Support Vector Machines (SVM) classifier is trained using the Histogram of Oriented Gradient (HOG) feature to determine whether a sub-image contains a single line of text in characters like Roman or not. The T-HOG descriptor, an improvement of the R-HOG, is optimized for this task and has shown to outperform other text detection systems. It efficiently characterizes images of single-line texts and can be used in various text detection applications (Minetto, Thome, Cord, Leite, & Stolfi, 2013)(Zaitoun & Aqel, 2015).

G) Method based on the Extremal Character Region

Character candidates are extracted by calculating the probability of each extreme region becoming a character, and ERs with the highest local pr. Character segmentation involves dividing a word image into individual character images using methods like contour analysis, vertical projection profiles, and connected components. Segmentation can be aided by pitch and character size estimation, with heuristic rules often used for locating segmentation points. Combining segmentation and recognition based on local shape information has shown promising results (Ho, 1992).

Word recognition methodologies typically involve character segmentation, recognition, and contextual postprocessing steps, with contextual constraints playing a crucial role in improving accuracy (Banafa, 2023). The Latest Technique for Text Segmentation in Adegan Fungsi energi (atau biaya) baru diperkenalkan pada piksel gambar dokumen, yang didefinisikan sebagai variabel acak dalam Markov Random Field (MRF) dalam penelitian terkini. Each variable receives a label on the front or back, and the energy function determines the quality of the combination. As a result, the energy function is reduced, allowing for a more advanced graphic design strategy to determine the appropriate balance. The sources discuss visual-text inpainting and the challenges of reconstructing damaged text images. They mention that existing methods struggle to accurately recover text within scene text images due to the complexity of text content. Different approaches like image inpainting and diffusion inpainting

are explored, but text completion from damaged images remains a challenge. The proposed CLII model aims to predict missing characters in damaged text images by leveraging visual attributes. Proposed a new method for embedded text segmentation. This method is based on two assumptions of the embedded text:

The color of text pixels follows a Gaussian distribution; ii) the local part of the embedded text has the same color distribution as the global part. With these two assumptions, he developed a two-step text segmentation approach: in the first step, the rough segmentation step, he used a 1-D Gaussian function to create a model of the text pixel color distribution. Parameter model and parameter yang ditentukan dari heuristic procedure yang dibuat dan diperoleh dengan mengekstrak bagian teks yang meyakinkan using operator punch. T. Wakaharad obtained a number of sub-gambars after using K-means in the HIS space. Then, using network characteristics and SVM classification thresholds, he assesses the likelihood of character images from sub-images that have been segmented in a systematic manner.

Existing Methods for Scene Character Recognition

The proposed character recognition system uses a segmentation algorithm with a multi-size sliding window to generate a template image. It does not require explicit font identification, instead utilizing a font-independent approach for character segmentation and recognition (Qaroush et al., 2022)(Saabni, 2014). The proposed character recognition method using local features with several desired properties involves the fusion of multiple feature sets, specifically Gabor wavelets and Local Binary Patterns (LBP), to achieve better performance than using either feature set alone. This method aims to improve robustness to spatial deformations and noise in uniform regions, as well as enhance performance under difficult illumination conditions (Cicolani, 2018).

Using insights from natural language processing and presenting a Markov chain framework for parsing images. A composition machine is a generative, probabilistic image model that embodies a hierarchy of part/whole relationships, distinct from Markovian models. It perturbs a Markov backbone to achieve greater selectivity and accommodate contextual relationships (Ya & Geman, 2006) (Ya & Geman, 2006) (Yin et al., 2015). This approach allows for the reuse of components among many entities and non-Markovian distributions. The proposed method combines visual n-grams and MKL strategies to improve the BoVW representation for image classification tasks. Visual n-grams capture complex visual patterns and spatial context in images, enhancing classification performance. MKL fusion strategies are used to integrate visual words and visual n-grams for better classification results (López-Monroy, Montes-y-Gómez, Escalante, Cruz-Roa, & González, 2016)

Comparison of text detection methods based on extreme regions

Maximally Stable Extremal Regions (MSER) is a method used to detect text from natural scene images. (Yan & Gao, 2014) (Ait Bella, El Rhabi, Hakim, & Laghrib, 2022). The MSER detector has the highest repeatability score for different sizes of detected regions compared to Harris-Affine, Hessian-Affine, and IBR. The distinctiveness of the regions is important for matching and clustering in practical applications, with the SIFT descriptor providing the best matching results(Qijie Zhao1, 2014). Tthe MSER-based method is categorized as a connected component-based method, and it has reported promising performance on the widely used ICDAR 2011 dataset (Wei et al., 2017).

Table 1 below compares various text detection methods based on MSER, along with their advantages and disadvantages.

Detection Method	Excess	Lack
MSER (Maximally Stable Extremal Regions)	- Very good at detecting naturally segmented text from the background. - Capable of detecting text under varying lighting conditions. - Fast and efficient in simple image processing.	- Difficulty detecting text with high noise. - Less effective on text with low contrast backgrounds. - Not optimal for text with significant size or shape variations.
MSER + Stroke Width Transform (SWT)	- More accurate in detecting individual characters in text. - Better at detecting text of different sizes and fonts.	- Slower than pure MSER. - Sensitive to very large variations in text thickness.
MSER + Edge Detection	- Good at detecting text on complex or detailed backgrounds. - Produces clearer text contours.	- Performance decreases on images with a lot of noise or with non-uniform lighting..
MSER + Connected Component Analysis (CCA)	- Better at separating adjacent or touching characters. - Effective on text written in blocks or paragraphs.	- Difficulty detecting text in images that vary greatly in size or orientation. - Can incorrectly detect non-text objects with characteristics similar to text.
MSER + Machine Learning (SVM, CNN, etc.)	- Very accurate in detecting and classifying text, especially when used with advanced features of CNN or SVM. - Capable of learning from large datasets and handling text of various shapes, sizes, and languages.	- Requires heavier computation and long training time. - Requires large datasets to train the model effectively.

Brief Explanation:

a. MSER specifically works well for text that is highly contrasted from the background and stable in various lighting conditions**.** However, this approach can fail when dealing with more complex images or those with high noise.
b. Combining MSER with other methods, such as SWT or Edge Detection, can improve accuracy, especially in more complex images or when text is present in various sizes and orientations.
c. Machine Learning-based methods (e.g., CNN) offer the most accurate results, but with high computational costs and large dataset requirements.

APPLICATION

1. Text detection, segmentation, and extraction from complex images can be applied to various fields where information needs to be analyzed and understood. Various methods such as using Gabor filters for texture segmentation, detecting text lines using horizontal spatial variance, and segmenting text into binary images with black characters on white background have been proposed in research literature (Waelen, 2023) (Lienhart & Wernicke, 2002), (Zhong, Karu, & Jain, 1995) (Jung et al., 2004).

2. Content-based image filtering can help detect image spam, pornography, hate speech, and fraud easily (Gómez-Aguilar & Atangana, 2018) (Jung et al., 2004)

3. Text extraction technology can be applied to detect scene text from images taken with laptops, mobile phones, and other equipment. The process involves text localization to identify text regions and text recognition to convert image-based information into text

codes. However, automatic localization of text regions from camera-captured images with complex backgrounds remains a challenge(Merino-Gracia, Lenc, & Mirmehdi, 2012)

4. By identifying text on road signs from movies, text extraction can be used to track traffic in real time. Several academics have built models to recognize deformed text from photographs, while previous research has focused on leveraging edge-based features to detect scene text from still images. A proposed framework incrementally detects text on road signs from video by locating road signs before detecting text within candidate areas. The framework exploits spatiotemporal information in video and employs a two-step strategy for text detection. The framework considers the appearance of a road sign in a video as a pyramid of sign image patches along the timeline, utilizing discriminative points detection(W. Wu, Chen, & Yang, 2005).

5. Optical character recognition (OCR) is a technology that converts bitmap images back into text. Different fonts can pose challenges for simple template matching algorithms. More complex systems segment text, sort it into lines and angles, and break down characters. Memory in computers includes Random Access Memory (RAM) for quick access and Long Term Memory (LTM) for storage on disks like magnetic or optical disks (Shah et al., 2018).

6. In order for postal automation to guarantee fast delivery from sender to recipient, automatic geocoding and localization of postal addresses on envelopes are essential. The speed and method of delivery of goods vary between types of postal services, such as courier, express, and universal postal services. The main component of postal traffic optimization is the postal network, which includes technical resources, postal facilities, and human resources. The United States Post Office utilizes sorting machines for mail processing, with single and double coding methods being debated for address coding (*Igor Baltić.pdf,* n.d.)

7. Text extraction in video sequences involves detecting, localizing, tracking, and extracting text from images and video frames. Various techniques exist to address this problem, including text detection, localization, tracking, and binarization. The extraction of text from video can be challenging due to variations in text size, style, orientation, and alignment, as well as complex backgrounds (Kumar & Ramakrishnan, 2018) (Crandall, Antani, & Kasturi, 2003)(Jung et al., 2004).

8. Wearable applications: "Wearable devices like glasses, phones, and cameras are being used to assist blind and visually impaired individuals through various technologies such as AI-based sign language interpreters and audio assistance for teachers. However, specific information about text detection and conversion into speech for blind people was not found in the provided sources (Tahoun, Awad, & Bonny, 2019).

9. Online shopping applications using mobile phones allow customers to type in the name of the item and obtain the necessary information. Younger shoppers under age 35 prefer using technology like two-way text chat and videoconferencing when shopping online. Men are more interested in using various types of technology in the shopping process, while women focus more on price, promotions, and coupons. Women are also more interested in features that increase shopping convenience, such as electronic shopping lists and speaking to customer service representatives (Ye et al., 2005) (Burke, 2002).

Conclusion

Real-world approaches for text detection, segmentation, and recognition are reviewed in this work along with their characteristics. The main concepts, advantages, disadvantages, and applications of text detection algorithms are also explained in this work. Detecting and recognizing text from landscape photographs is more difficult than other types of images. This is due to the complex background, irregular shapes, and varying sizes of text in natural landscapes. Despite these challenges, recent advances in deep learning algorithms have shown promising results in improving the accuracy and efficiency of text detection in such images. In addition, the integration of semantic information and contextual cues has further improved the performance of text recognition systems in natural scenes. In conclusion, the development of robust text detection and recognition techniques for natural landscapes has great potential for a variety of applications in areas such as automated driving, augmented reality, and environmental monitoring. This is influenced by a number of factors, including blurriness, font design, lighting effects, and orientation. Although there are many algorithms, no single method works for every application. Therefore, there is a lot of room for experimentation with text extraction, recognition, segmentation, and detection from natural landscape images. In addition, text from other languages that are different from English in some respects can be detected.

References

Ait Bella, F. Z., El Rhabi, M., Hakim, A., & Laghrib, A. (2022). An innovative document image binarization approach driven by the non-local p-Laplacian. *Eurasip Journal on Advances in Signal Processing*, *2022*(1). https://doi.org/10.1186/s13634-022-00883-2

Anagnostopoulos, C. N. E., Anagnostopoulos, I. E., Loumos, V., & Kayafas, E. (2006). A license plate-recognition algorithm for intelligent transportation system applications. *IEEE Transactions on Intelligent Transportation Systems*, *7*(3), 377–391. https://doi.org/10.1109/TITS.2006.880641

Aqel, M. O. A., Marhaban, M. H., Saripan, M. I., & Ismail, N. B. (2016). Review of visual odometry: types, approaches, challenges, and applications. *SpringerPlus*, *5*(1). https://doi.org/10.1186/s40064-016-3573-7

Arshad, H., Abidin, R. Z., & Obeidy, W. K. (2017). Identification of vehicle plate number using optical character recognition: A mobile application. *Pertanika Journal of Science and Technology*, *25*(S6), 173–180.

Banafa, A. (2023). Edge Computing Paradigm. *Quantum Computing and Other Transformative Technologies*, 81–85. https://doi.org/10.1201/9781003339175-21

Binmakhashen, G. M., & Mahmoud, S. A. (2019). Document layout analysis: A comprehensive survey. *ACM Computing Surveys*, *52*(6). https://doi.org/10.1145/3355610

Burke, R. R. (2002). Technology and the customer interface: What consumers want in the physical and virtual store. *Journal of the Academy of Marketing Science*, *30*(4), 411–432. https://doi.org/10.1177/009207002236914

Chen, X., Jin, L., Zhu, Y., Luo, C., & Wang, T. (2021). Text Recognition in the Wild: A Survey. *ACM Computing Surveys*, *54*(2). https://doi.org/10.1145/3440756

Cicolani, J. (2018). Beginning Robotics with Raspberry Pi and Arduino. In *Beginning Robotics with Raspberry Pi and Arduino*. https://doi.org/10.1007/978-1-4842-3462-4

Crandall, D., Antani, S., & Kasturi, R. (2003). Extraction of special effects caption text events from digital video. *International Journal on Document Analysis and Recognition*, *5*(2–3), 138–157. https://doi.org/10.1007/s10032-002-0091-7

Danial, M. N., Rosli, O., M.Zarar, M. J., & Jean-Marc, O. (2011). Image Segmentation and Text Extraction: Application To The Extraction Of Textual Information In Science Images. *International Seminar on Application of Science Mathematics 2011*, 1–8.

Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, *40*(2), 1–60. https://doi.org/10.1145/1348246.1348248

Gómez-Aguilar, J., & Atangana, A. (2018). Fractional Derivatives with the Power-Law and the Mittag–Leffler Kernel Applied to the Nonlinear Baggs–Freedman Model. *Fractal and Fractional*, *2*(1), 10. https://doi.org/10.3390/fractalfract2010010

Guo, J., Liu, Y., Wu, L., Liu, S., Yang, T., Zhu, W., & Zhang, Z. (2019). A geometry- and texture-based automatic discontinuity trace extraction method for rock mass point cloud. *International Journal of Rock Mechanics and Mining Sciences*, *124*(February), 104132. https://doi.org/10.1016/j.ijrmms.2019.104132

Ho, T. K. (1992). *A Theory of Multiple Classi er Systems And Its Application to Visual Word Recognition*.

Huang, W., Qiao, Y., & Tang, X. (2014). Robust scene text detection with convolution neural network induced MSER trees. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8692 LNCS*(PART 4), 497–511. https://doi.org/10.1007/978-3-319-10593-2_33

Igor Baltić.pdf. (n.d.).

Jain, P. H., Kumar, V., Samuel, J., Singh, S., Mannepalli, A., & Anderson, R. (2023). Artificially Intelligent Readers: An Adaptive Framework for Original Handwritten Numerical Digits Recognition with OCR Methods. *Information (Switzerland)*, *14*(6). https://doi.org/10.3390/info14060305

Jain, R., & Gianchandani, P. D. (2019). A hybrid approach for detection and recognition of traffic text sign using MSER and OCR. *Proceedings of the International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2018*, 775–778. https://doi.org/10.1109/I-SMAC.2018.8653761

Joshi, K. N., & Patil, B. T. (2019). Effect of Illumination Systems on Statistical Texture Parameters Based Clustering and Discrimination of Machined Surfaces Using Machine Vision. *Mapan - Journal of Metrology Society of India*, *34*(2), 197–205. https://doi.org/10.1007/s12647-018-0279-z

Jung, K., Kim, K. I., & Jain, A. K. (2004). Text information extraction in images and video: A survey. *Pattern Recognition*, *37*(5), 977–997. https://doi.org/10.1016/j.patcog.2003.10.012

Karpinski, R., Lohani, D., Belaid, A., Karpinski, R., Lohani, D., Belaid, A., … Loria, U. D. L. (2019). *Metrics for Complete Evaluation of OCR Performance To cite this version : HAL Id : hal-01981731 Metrics for Complete Evaluation of OCR Performance*.

Khosravy, M., Gupta, N., Marina, N., Sethi, I. K., & Asharif, M. R. (2017). Morphological filters: An inspiration from natural geometrical erosion and dilation. In *Modeling and Optimization in Science and Technologies* (Vol. 10). https://doi.org/10.1007/978-3-319-50920-4_14

Kim, K. B. (2015). Image binarization using intensity range of grayscale images. *International Journal of Multimedia and Ubiquitous Engineering*, *10*(7), 139–144. https://doi.org/10.14257/ijmue.2015.10.7.15

Koscher, K., Czeskis, A., Roesner, F., Patel, S., Kohno, T., Checkoway, S., … Savage, S. (2010). Experimental security analysis of a modern automobile. *Proceedings - IEEE Symposium on Security and Privacy*, 447–462. https://doi.org/10.1109/SP.2010.34

Kosinski, M. (2021). Facial recognition technology can expose political orientation from naturalistic facial images. *Scientific Reports*, *11*(1), 1–7. https://doi.org/10.1038/s41598-020-79310-1

Kuipers, B. J., & Levitt, T. S. (1988). Navigation and Mapping in Large Scale Space. *AI Magazine*, *9*(2), 25. Retrieved from http://www.aaai.org/ojs/index.php/aimagazine/article/view/674

Kumar, D., & Ramakrishnan, A. G. (2018). 32.Methods for text segmentation from scene images (Vol. 13). https://doi.org/10.5565/rev/elcvia.591

Lerum, H., Karlsen, T. H., & Faxvaag, A. (2003). Effects of Scanning and Eliminating Paper-based Medical Records on Hospital Physicians' Clinical Work Practice. *Journal of the American Medical Informatics Association*, *10*(6), 588–595. https://doi.org/10.1197/jamia.M1337

Lienhart, R., & Wernicke, A. (2002). Localizing and segmenting text in images and videos. *IEEE Transactions on Circuits and Systems for Video Technology*, *12*(4), 256–268. https://doi.org/10.1109/76.999203

Lister, M. (2013). The photographic image in digital culture, Second edition. In *The Photographic Image in Digital Culture, Second Edition* (Vol. 9780203797563). https://doi.org/10.4324/9780203797563

Liu, Y., Zhang, D., & Lu, G. (2008). Region-based image retrieval with high-level semantics using decision tree learning. *Pattern Recognition*, *41*(8), 2554–2570. https://doi.org/10.1016/j.patcog.2007.12.003

López-Monroy, A. P., Montes-y-Gómez, M., Escalante, H. J., Cruz-Roa, A., & González, F. A. (2016). Improving the BoVW via discriminative visual n-grams and MKL strategies. *Neurocomputing*, *175*(PartA), 768–781. https://doi.org/10.1016/j.neucom.2015.10.053

Mahmood, Z., Khan, K., Khan, U., Adil, S. H., Ali, S. S. A., & Shahzad, M. (2022). Towards Automatic License Plate Detection. *Sensors*, *22*(3), 1–19. https://doi.org/10.3390/s22031245

Maier, M. W. (1999). Architecting Principles for. *The Aerospace Corporation; John Wiley & Sons, Inc. Syst Eng*, *1*, 267–284.

Marsh, E. E., & White, M. D. (2003). A taxonomy of relationships between images and text. *Journal of Documentation*, *59*(6), 647–672. https://doi.org/10.1108/00220410310506303

Member, S., & Sun, Y. (2000). *A Hierarchical Approach to Color Image*. *9*(12), 2071–2082.

Merino-Gracia, C., Lenc, K., & Mirmehdi, M. (2012). A head-mounted device for recognizing text in natural scenes. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7139 LNCS*, 29–41. https://doi.org/10.1007/978-3-642-29364-1_3

Minetto, R., Thome, N., Cord, M., Leite, N. J., & Stolfi, J. (2013). T-HOG: An effective gradient-based descriptor for single line text regions. *Pattern Recognition*, *46*(3), 1078–1090. https://doi.org/10.1016/j.patcog.2012.10.009

Miri, M. S., Abràmoff, M. D., Kwon, Y. H., Sonka, M., & Garvin, M. K. (2017). A machine-learning graph-based approach for 3D segmentation of Bruch's membrane opening from glaucomatous SD-OCT volumes. *Medical Image Analysis*, *39*, 206–217. https://doi.org/10.1016/j.media.2017.04.007

Mirza, A., Zeshan, O., Atif, M., & Siddiqi, I. (2020). Detection and recognition of cursive text from video frames. *Eurasip Journal on Image and Video Processing*, *2020*(1). https://doi.org/10.1186/s13640-020-00523-5

Mittal, H., Pandey, A. C., Saraswat, M., Kumar, S., Pal, R., & Modwel, G. (2022). A comprehensive survey of image segmentation: clustering methods, performance parameters, and benchmark datasets. *Multimedia Tools and Applications*, *81*(24), 35001–35026. https://doi.org/10.1007/s11042-021-10594-9

Nagy, G. (2016). Disruptive developments in document recognition. *Pattern Recognition Letters*, *79*, 106–112. https://doi.org/10.1016/j.patrec.2015.11.024

Palani, D., Venkatalakshmi, K., & Venkatraman, E. (2014). *I Mplementation & C Omparison O F D Ifferent S Egmentation a Lgorithms F or*. *3*(3), 1217–1223.

Qaroush, A., Jaber, B., Mohammad, K., Washaha, M., Maali, E., & Nayef, N. (2022). An efficient, font independent word and character segmentation algorithm for printed Arabic text. *Journal of King Saud University - Computer and Information Sciences*, *34*(1), 1330–1344. https://doi.org/10.1016/j.jksuci.2019.08.013

Qijie Zhao1. (2014). *12.Image Capturing and Segmentation Method for Characters Marked on Hot Billets*.

Saabni, R. (2014). Efficient recognition of machine printed Arabic text using partial segmentation and Hausdorff distance. *6th International Conference on Soft Computing and Pattern Recognition, SoCPaR 2014*, 284–289.

https://doi.org/10.1109/SOCPAR.2014.7008020

Saric, M. (2017). *9.Scene Text Segmentation using Low Variation Extremal Regions and Sorting Based Character Grouping*.

Shah, M., Mehta, S., Mody, P., Roy, A. Sen, & Khachane, S. P. (2018). *Handwriting Recognition of Diverse Languages*. *7*(4), 109–114.

Tahoun, N., Awad, A., & Bonny, T. (2019). Smart assistant for blind and visually impaired people. *ACM International Conference Proceeding Series*, 227–231. https://doi.org/10.1145/3369114.3369139

Thesis, K. S., Breuel, T., Prof, T. U. K., Dengel, A., Kaiserslautern, T. U., Berns, K., & Kaiserslautern, T. U. (2014). *Optical Character Recognition - A Combined ANN / HMM Approach Dissertation Sheikh Faisal Rashid*.

Waelen, R. A. (2023). The struggle for recognition in the age of facial recognition technology. *AI and Ethics*, *3*(1), 215–222. https://doi.org/10.1007/s43681-022-00146-8

Wang, X., Song, Y., Zhang, Y., & Xin, J. (2015). Natural scene text detection with multi-layer segmentation and higher order conditional random field based analysis. *Pattern Recognition Letters*, *60–61*, 41–47. https://doi.org/10.1016/j.patrec.2015.04.005

Wang, Y., Shi, C., Xiao, B., Wang, C., & Qi, C. (2018). CRF based text detection for natural scene images using convolutional neural network and context information. *Neurocomputing*, *295*, 46–58. https://doi.org/10.1016/j.neucom.2017.12.058

Wei, Y., Zhang, Z., Shen, W., Zeng, D., Fang, M., & Zhou, S. (2017). Text detection in scene images based on exhaustive segmentation. *Signal Processing: Image Communication*, *50*, 1–8. https://doi.org/10.1016/j.image.2016.10.003

Wu, L., Han, J., Zhang, C., Liu, J., Bai, X., Liu, J., & Ding, E. (2019). Editing text in the wild. *MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia*, 1500–1508. https://doi.org/10.1145/3343031.3350929

Wu, W., Chen, X., & Yang, J. (2005). Detection of text on road signs from video. *IEEE Transactions on Intelligent Transportation Systems*, *6*(4), 378–390. https://doi.org/10.1109/TITS.2005.858619

Wydyanto, W., Nayan, N. M., Sulaiman, R., Dewi, D. A., & Kurniawan, T. B. (2024). A Hybrid Approach to Detect and Identify Text in Picture. *Emerging Science Journal*, *8*(1), 218–238. https://doi.org/10.28991/ESJ-2024-08-01-016

Ya, J., & Geman, S. (2006). Context and hierarchy in a probabilistic image model. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2*, 2145–2152. https://doi.org/10.1109/CVPR.2006.86

Yan, J., & Gao, X. (2014). Detection and recognition of text superimposed in images base on layered method. *Neurocomputing*, *134*, 3–14. https://doi.org/10.1016/j.neucom.2012.12.070

Ye, Q., & Doermann, D. (2015). Text Detection and Recognition in Imagery: A Survey. *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, *37*(7), 1480–1500. https://doi.org/10.1109/TPAMI.2014.2366765

Ye, Q., Huang, Q., Gao, W., & Zhao, D. (2005). Fast and robust text detection in images and video frames. *Image and Vision Computing*, *23*(6), 565–576. https://doi.org/10.1016/j.imavis.2005.01.004

Yin, X. C., Pei, W. Y., Zhang, J., & Hao, H. W. (2015). Multi-Orientation Scene Text Detection with Adaptive Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(9), 1930–1937. https://doi.org/10.1109/TPAMI.2014.2388210

Zaitoun, N. M., & Aqel, M. J. (2015). Survey on Image Segmentation Techniques. *Procedia Computer Science*, *65*(Iccmit), 797–806. https://doi.org/10.1016/j.procs.2015.09.027

Zhang, H., Zhao, K., Song, Y. Z., & Guo, J. (2013). Text extraction from natural scene image: A survey. *Neurocomputing*, *122*, 310–323. https://doi.org/10.1016/j.neucom.2013.05.037

Zhao, X., Lin, K. H., Fu, Y., Hu, Y., Liu, Y., & Huang, T. S. (2011). Text from corners: A novel approach to detect text and caption in videos. *IEEE Transactions on Image Processing*, *20*(3), 790–799. https://doi.org/10.1109/TIP.2010.2068553

Zhong, Y., Karu, K., & Jain, A. K. (1995). Locating text in complex color images. *Proceedings of the International Conference on Document Analysis and Recognition*, *ICDAR*, *1*(10), 146–149. https://doi.org/10.1109/ICDAR.1995.598963