iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.62245

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Learning Efficient to Converting a Video Content into Summarizing Word Document Using NLP

M. Vasuki[1], Dr. T. Amalaraj Victoire[2], Anujaa. T[3]
[1]*Associate Professor, Department of Master of Computer Application, Sri Manakula Vinayagar Engineering College Puducherry-605 107, India*
[2]*Associate Professor, Department of Master of Computer Application, Sri Manakula Vinayagar Engineering College Puducherry-605 107, India*
[3]*PG Student, Department of Master of Computer Application, Sri Manakula Vinayagar Engineering College Puducherry-605 107, India*

Abstract: ***This paper explores an efficient methodology for converting video content into summarized Word documents using Optical Character Recognition (OCR) tools. The process integrates multiple advanced technologies, including frame extraction, audio transcription, OCR for text recognition in video frames, and natural language processing (NLP) techniques for text summarization. The proposed approach begins with extracting key frames from the video and applying OCR to capture on-screen text. Simultaneously, the video's audio track is transcribed into text using speech-to-text technology. The extracted texts are then combined and processed through NLP algorithms to generate a concise summary. Finally, the summarized text is formatted into a structured and readable Word document. This methodology aims to create accurate, readable, and easily shareable text documents from video content, offering significant utility in fields such as education, corporate training, and media. The paper discusses the tools and technologies used, the challenges faced, and the potential future directions for enhancing the efficiency and accuracy of this conversion process.***

Keyword: ***Visual recognition, Attribute learning, Descriptive features, Concise representation, Image analysis, Feature extraction, Machine learning, Object detection, Classification, Computational efficiency.***

I. INTRODUCTION

Converting video content into textual transcripts offers a multitude of advantages, with enhanced accessibility standing out as a paramount benefit. By transcribing video content into text, individuals with hearing impairments are empowered to access and engage with video-based information, ensuring inclusivity and equal access to multimedia content. Moreover, textual representations enable a broader audience, including those who prefer consuming content in textual formats or are unable to watch videos in certain environments, to seamlessly interact with video content. This accessibility enhancement not only promotes inclusivity but also expands the reach and impact of video content across diverse audiences. Through the conversion of video to text, our project aims to bridge the accessibility gap, fostering a more inclusive digital environment where everyone can access and benefit from multimedia resources. Optical Character Recognition (OCR) is a technology designed to convert different types of documents, such as scanned paper documents, PDFs, or images captured by a digital camera, into editable and searchable data. OCR tools enable the automated recognition and digitization of printed or handwritten text, which can then be manipulated, searched, and stored electronically.

OCR stands for Optical Character Recognition The technology you're referring to is Optical Character Recognition (OCR). OCR is a process that converts various types of documents, including scanned paper documents, PDF files, or images captured by a digital camera, into editable and searchable data. OCR software analyzes the structure of a document image to recognize and extract text characters, allowing the content to be electronically edited, searched, and stored. It's commonly used in various applications, including document digitization, data entry automation, and accessibility for visually impaired individuals**.**

Natural Language Processing (NLP) is indeed a branch of artificial intelligence (AI) focused on the interaction between computers and human (natural) languages. It involves the development of algorithms and systems that can understand, interpret, and generate human language in a way that is both meaningful and useful. NLP combines computational linguistics with machine learning, deep learning, and other AI technologies to enable computers to process and analyze large amounts of natural language data. It enables computers to understand, interpret, and generate human language in a way that is both meaningful and useful. NLP encompasses a wide range of tasks, techniques, and applications aimed at processing and analyzing natural language data.

II. LITERATURE SURVEY

1) *Title: "Video Summarization Using Textual and Visual Information Fusion"*
Authors: Wang, M., Zhang, J., & Guo, B.
Publication Year: 2020
Summary: This paper proposes a method for video summarization by integrating textual and visual information. Textual information is extracted using NLP techniques, while visual information is captured through deep learning models. The fusion of these modalities results in more comprehensive and informative video summaries.

2) *Title: "A Survey on Video Summarization Techniques"*
Authors: Potluri, S., & Vuppala, A.
Publication Year: 2019
Summary: This survey provides an overview of various techniques and approaches for video summarization. It covers both extractive and abstractive summarization methods, as well as the use of features such as motion, audio, and text for summarization purposes.

3) *Title: "Learning Visual Knowledge Memory Networks for Visual Textual Entailment Recognition"*
Authors: Su, P., et al.
Publication Year: 2019
Summary: The paper presents a novel approach for visual textual entailment recognition, which involves determining whether the information conveyed in an image is entailed by a given textual statement. The proposed method utilizes memory networks to effectively capture and reason about visual and textual information.

4) *Title: "Text Detection and Recognition in Imagery: A Survey"*
Authors: Zhang, Z., & Shen, W.
Publication Year:2017
Summary: This survey provides an overview of text detection and recognition techniques in imagery, covering both traditional and deep learning-based approaches. It discusses the challenges associated with text detection and recognition in various types of images, including scenes and documents.

5) *Title: "A Survey on Deep Learning Techniques for Video Summarization"*
Authors: Sharma, V., & Kumar, S.
Publication Year: 2021
Summary: This survey explores the application of deep learning techniques for video summarization. It discusses various deep learning architectures and methodologies used for video summarization tasks, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs).

III. OCR TOOLS

1) *Pytesseract:* Pytesseract is indeed a Python wrapper for Google's Tesseract-OCR Engine. This combination allows developers to easily integrate Optical Character Recognition (OCR) capabilities into their Python applications. It allows developers to extract text from various sources such as PDFs, documents, and images. Pytesseract is easy to use and integrates well with Python applications.
2) *TensorFlow:* TensorFlow is an open-source machine learning framework developed by Google that provides a comprehensive ecosystem for building, training, and deploying machine learning models. It includes various tools and libraries that make it suitable for a wide range of applications, including Optical Character Recognition (OCR). TensorFlow enables the development of custom OCR pipelines through its powerful capabilities in text detection and recognition.
3) *Keras-OCR:* Keras-OCR is a library built on top of TensorFlow and Keras, specifically designed for OCR tasks. It provides pre-trained models and easy-to-use APIs for text detection and recognition. Keras-OCR simplifies the process of building end-to-end OCR systems using deep learning.
4) *Google Vision API:* Google Vision API offers a cloud-based OCR service that allows developers to integrate OCR capabilities into their applications easily. It supports various features such as image labeling, face detection, and optical character recognition. Google Vision API provides accurate and reliable OCR results with minimal setup.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

5) *AWS Textract:* AWS Textract is a fully managed OCR service provided by Amazon Web Services (AWS). It enables developers to extract text and data from scanned documents, PDFs, and images using machine learning algorithms. AWS Textract automatically detects text and key information from documents, making it easy to incorporate OCR into applications.

IV. ALGORITHM

Natural Language Processing (NLP) encompasses a wide range of algorithms and techniques designed to enable computers to understand, interpret, and generate human language. Here are some key algorithms commonly used in NLP:

1) *Tokenization:* Tokenization involves breaking down text into smaller units, such as words, subwords, or characters. It serves as the foundation for many NLP tasks by providing a structured representation of textual data.
2) *Statistical Language Models:* Statistical language models estimate the probability of word sequences occurring in a given language. They are used in tasks such as language generation, speech recognition, and machine translation.
3) *Part-of-Speech Tagging:* Part-of-speech (POS) tagging is indeed a crucial task in natural language processing (NLP) that involves assigning grammatical labels, such as noun, verb, adjective, etc., to each word in a sentence. Hidden Markov Models (HMMs) and Conditional Random Fields (CRFs) are two commonly used approaches for performing POS tagging
4) *Named Entity Recognition (NER):* NER identifies and categorizes named entities (e.g., person names, locations, organizations) within text. Sequence labeling algorithms like CRFs and deep learning models such as Bidirectional LSTMs are often used for NER.
5) *Text Classification:* Text classification assigns predefined categories or labels to text documents based on their content. Algorithms include Naive Bayes, Support Vector Machines (SVM), and deep learning architectures like Convolutional Neural Networks (CNNs) and Transformers.
6) *Sequence-to-Sequence Models:* Sequence-to-sequence models map input sequences to output sequences, making them suitable for tasks like machine translation, text summarization, and question answering. Recurrent Neural Networks (RNNs) and Transformer-based architectures are commonly used for sequence-to-sequence tasks.
7) *Word Embeddings:* Word embeddings are indeed representations of words as dense vectors in a continuous vector space. These vectors capture semantic relationships between words based on their contexts in large text corpora. Several techniques have been developed to learn word embeddings, with Word2Vec, GloVe (Global Vectors for Word Representation), and FastText being some of the most prominent ones.
8) *Dependency Parsing:* Dependency parsing analyzes the grammatical structure of sentences by determining the relationships between words. Dependency parsing algorithms include Transition-Based Parsing and Graph-Based Parsing.
9) *Semantic Parsing:* Semantic parsing converts natural language utterances into formal representations, such as logical forms or executable queries. It is used in tasks like question answering and dialog systems.
10) *Topic Modeling:* Topic Modeling algorithms such as Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF) are essential tools in natural language processing (NLP) for uncovering latent topics within a collection of documents.

These algorithms help in identifying the underlying themes or topics present in textual data, which is valuable for tasks such as document categorization, information retrieval, and content recommendation.

V. ARCHITECTURAL DIAGRAM

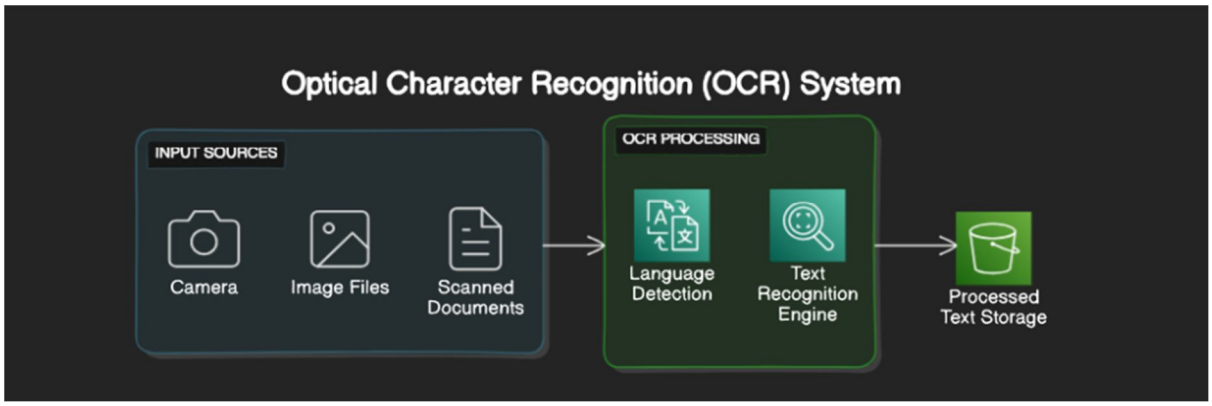


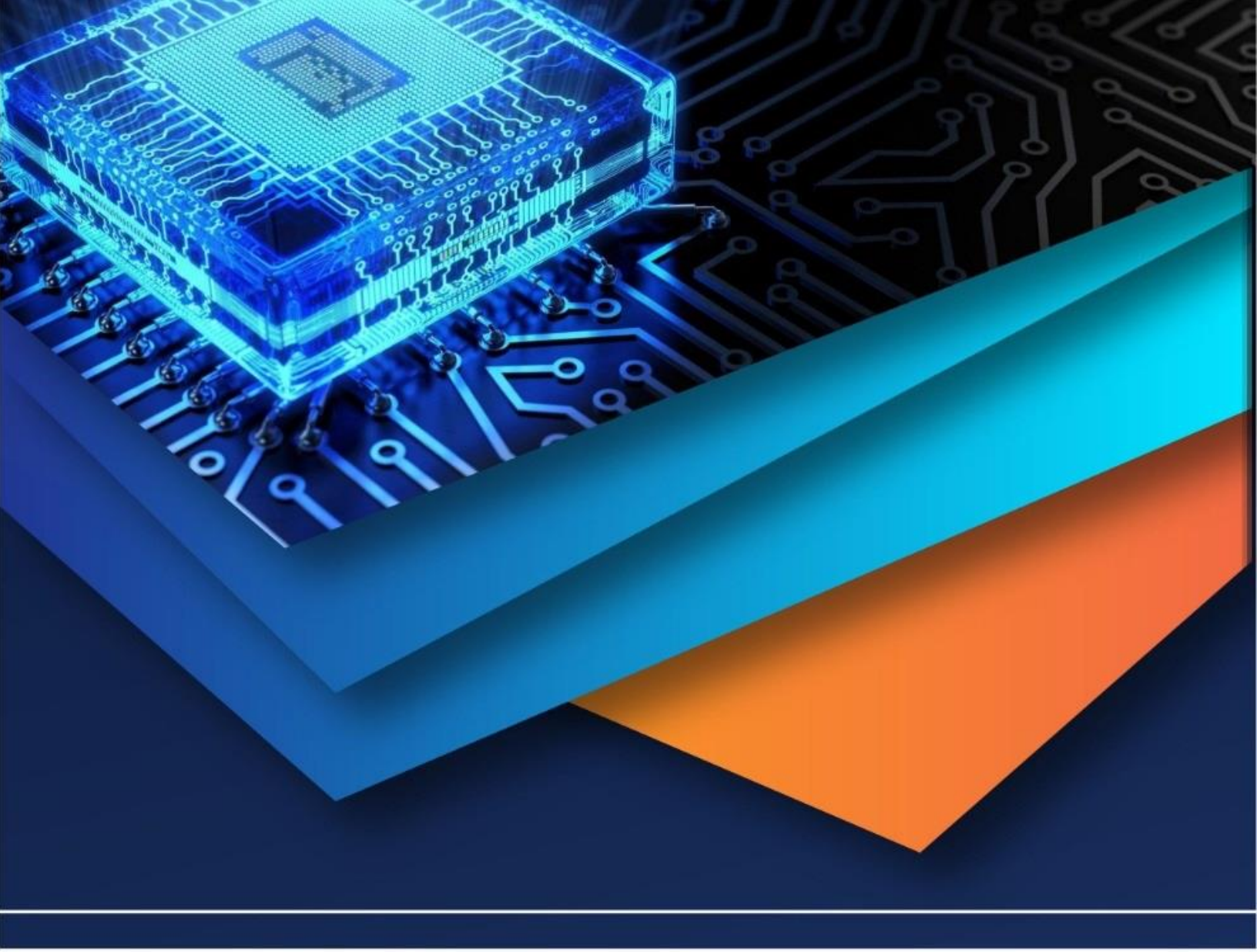
Fig:1.1

VI. CONCLUSION

In conclusion, by combining automated transcription services, NLP techniques, summarization algorithms, manual review processes, and integration with word processing tools, it's possible to efficiently convert video content into a summarizing Word document. This approach enables users to distill the key insights and information from videos into a concise and actionable format, facilitating better understanding, knowledge retention, and dissemination of information.

REFERENCES

[1] Deng, J., Ding, N., Jia, Y., Frome, A., Murphy, K., Bengio, S., Li, Y., Neven, H., & Adam, H. (2013). Learning Discriminative Attributes for Fine-grained Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
[2] Farhadi, A., Endres, I., Hoiem, D., & Forsyth, D. (2009). Describing Objects by their Attributes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
[3] Sadeghi, M. A., & Farhadi, A. (2011). Attribute-based Transfer Learning for Object Categorization with Zero/One-shot Learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
[4] Parikh, D., & Grauman, K. (2011). Learning Semantic Attributes for Object Recognition and Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10), 1992-2008.
[5] Li, L.-J., Su, H., Xing, E. P., & Fei-Fei, L. (2013). Learning Attributes and Parts for Object Categorization and Description. *International Journal of Computer Vision*, 103(2), 136-153.
[6] Torralba, A., & Efros, A. A. (2011). Learning Visual Attributes. In *Proceedings of the IEEE International Conference on Computer Vision*.
[7] Socher, R., Ganjoo, M., Manning, C. D., & Ng, A. Y. (2013). Zero-Shot Learning through Cross-Modal Transfer. In *Advances in Neural Information Processing Systems*.
[8] Yang, Y., Hospedales, T. M., Xiang, T., & Gong, S. (2015). Attribute-based Generalization Beyond the Seen Categories. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
[9] Tang, R., Shao, Z., Tang, S., & Wang, L. (2019). Attribute-aware Semantic Segmentation of Indoor Images. *IEEE Transactions on Image Processing*, 28(3), 1481-1495.
[10] Zhu, Z., Shi, K., Sheng, L., & Yu, M. (2019). Attribute-guided Convolutional Neural Networks for Face Age Estimation. *Pattern Recognition Letters*, 125, 290-297.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY