JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

IntelliPresent: AI-Powered Smart Presentation System

Prof.Deepali Deshpande
Information Technology
Vishwakarma Institute of Technology
Pune, India
deepali.deshpande@vit.edu

Soham Pingat
Information Technology
Vishwakarma Institute of Technology
Pune, India
soham.pingat22@vit.edu

Sujal Powar
Information Technology
Vishwakarma Institute of Technology
Pune, India
sujal.powar22@vit.edu

Sameer Raut
Information Technology
Vishwakarma Institute of Technology
Pune, India
sameer.raut22@vit.edu

Vaibhav Pokale
Information Technology
Vishwakarma Institute of Technology
Pune, India
vaibhav.pokale22@vit.edu

***Abstract*— Traditional presentations require human intervention, which can interrupt the flow of information and reduce audience interaction. To address these problems, we introduce IntelliPresent—a rich, AI-powered presentation system to enhance presentations in terms of interaction, engagement, and effectiveness. IntelliPresent uses voice tracking through Google Speech-to-Text, OpenCV-hand gesture detection, and image text extraction through PyTesseract. IntelliPresent also uses AI-content summarization for effective information transmission. In addition to these features, the system dynamically creates slides, automatically visualizes data, provides context to the subject, and incorporates an integrated model for presentation generation with editable templates. Built as a web app using Node.js, Express, and HTML/CSS, IntelliPresent transforms traditional presentations into a hands-free, interactive experience. By overcoming the limitations of traditional tools, IntelliPresent is a paradigm shift in presentation automation and user experience.**

Keywords : AI-powered presentations, Voice tracking, Hand gesture recognition, Text extraction, AI-content summarization, Presentation automation.

I. INTRODUCTION

Success in presentation is a critical factor in information presentation in the education, business, and public speaking sectors of today's era. Human user interaction is a common shortcoming of conventional presentation approaches, which prevents the smooth flow of information and discourages interactive participation from the audience. IntelliPresent aims to overcome such challenges by proposing an AI-based system that improves presentation processes, fosters audience participation, and makes content management easy.

This article documents the essential characteristics, technological application, and impact on presentation practices in use today, demonstrating IntelliPresent's offering of a sequence of features based on Artificial Intelligence to provide increased interactivity and effectiveness. The system uses a solution that not only improves conventional methods of presenting information but extends to capture document analysis, real-time summaries, and dynamic content generation.

II. PROBLEM STATEMENT

Most traditional presentations use physical interaction for navigating and editing content, breaking the flow of the presenter and disengaging the audience. There is a demand for an intuitive, interactive, and smooth presentation experience that has become stronger, especially with heightened calls for dynamic content customization and operation without the use of hands. Capitalizing on recent advancements in AI, IntelliPresent answers the above problems and sets out to revolutionize presentation delivery and reception.

Traditional presentation techniques are not interactive nor flexible, and they involve manual navigation and fixed content. These shortcomings decrease interaction and efficacy, particularly in dynamic contexts such as business and education. A smarter system that incorporates voice tracking, gesture detection, real-time extraction of content, and AI-driven slide adaptation in order to establish a hands-free, interactive, and flexible presentation process is required.

III. LITERATURE SURVEY

[1] Barawal and Arora surveyed several approaches to text extraction from images, such as region-based, edge-based, and texture-based methods. Their study presents the effectiveness of morphological and edge-based approaches for text location. They applied their results by means of PyTesseract and OpenCV with strong performance on various types of images. The methods were beneficial for document processing, image indexing, and video summarization applications.

[2] S. S. Patil and A. M. Pawar concentrated on designing a text summarization system with Term Frequency-Inverse Document Frequency (TF-IDF). The summarizer performed well in tasks like question answering and text classification. They suggested future improvements such as topic modeling

and machine learning to enhance accuracy in domain-specific applications.

[3] Yalniz and Manmatha introduced a Markov Random Field (MRF) model to address the limitations of traditional OCR systems. Their method integrates visual features, significantly enhancing text search accuracy across various scripts and noisy documents. The model supports arbitrary text queries and offers real-time search capabilities.

[4] In a subsequent paper, Manmatha and Yalniz applied their MRF-based text extraction technique to street-level images. The technique handles noisy scenes and multi-scripts such as English, Telugu, and Ottoman well and is thus extremely helpful in challenging retrieval situations.

[5] Li et al. introduced a text detection and digital video tracking system. The method combines a scale-space feature extractor, a neural network, and an SSD-based tracking module. The system can process both graphic and scene text in dynamic motion, and it is applicable for content-based video retrieval.

[6] Suarez and Murphy provided a comprehensive overview of depth-based hand gesture recognition systems, categorizing hand localization and gesture classification techniques. Their work demonstrates the utilization of low-cost sensors such as Kinect to increase accessibility and references challenges such as environmental variability.

[7] Oudah et al. were working on AI-based approaches to recognize hand gestures for human-computer interaction, medical, and home automation applications. The study explains that implementation of machine learning can significantly enhance non-verbal communication systems and the solution to occlusions and environmental problems.

[8] (Extended Work)Oudah et al. authored an extended review on computer vision-based gesture recognition developments. The article outlined major challenges, including illumination, occlusion, and computational inefficiency, and proposed methods for overcoming these limitations for other applications.

[9] Graves et al. proposed an end-to-end speech recognition system using Recurrent Neural Networks (RNNs). By reducing the traditional multi-stage pipeline to its simplest form, their model experienced dramatic accuracy gains, particularly in noisy and conversational environments, through the application of sequence-to-sequence learning.

[10] Chan et al. proposed the Listen, Attend, and Spell (LAS) model, which combined attention mechanisms and encoder-decoder models. The model outperformed conventional systems in dealing with long sequences and complex linguistic structures and was thus particularly well-suited to large-vocabulary speech recognition.

[11] Fu et al. introduced DOC2PPT, a system to facilitate the automation of presentation slide generation from scientific papers. By extracting and organizing key content in slideable structures, this system greatly facilitated the preparation of presentations.

[12] Bandyopadhyay et al. enhanced slide generation with Large Language Models (LLMs) using a multi-stage method. With the integration of topic modeling, summarization, and design optimization, their method generated better quality and coherence in automatically generated slides.

[13] Devlin et al. presented BERT, a bidirectional transformer model that has set new state-of-the-art performance in natural language processing tasks like question answering and language inference. BERT's capacity to learn bidirectional context was a significant contributor to NLP applications.

[14] Khan et al. constructed a real-time low-resolution image facial expression recognition system. With pattern recognition, their system was resilient under difficult conditions, with the potential for surveillance and human-computer interaction.

This literature review introduces important developments in text extraction, gesture recognition, speech recognition, and automated slide generation, which are the pillars of IntelliPresent. Our accurate text extraction with PyTesseract and OpenCV is guided by the findings of [1,3,4]. For AI-driven content summarization for efficient information sharing, [2] directs us. For hand-free gesture navigation, we are inspired by [6,7,8], employing computer vision-based hand gesture recognition for hand-free navigation. [9,10] direct our voice tracking and Google Speech-to-Text based speech recognition for convenient interaction. [11,12] direct our automated slide generation for ease of content organization.

By combining these methods, IntelliPresent converts regular presentations into a complete automated, AI-driven, and interactive presentation, avoiding the drawbacks of the traditional tools and boosting engagement.

IV. METHODOLGY

4.1 System Architecture

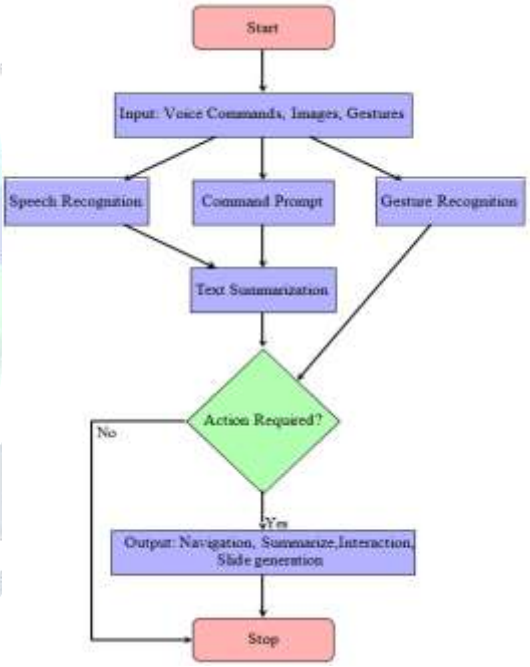

Fig.1 Flow Diagram

IntelliPresent is an Internet-based, modular system, with the following significant components:

Fig.1 Environment For making Presentation

- **Voice Tracking**: Applies Google Speech-to-Text API to voice-activate slides and live phrase highlighting in presentations.

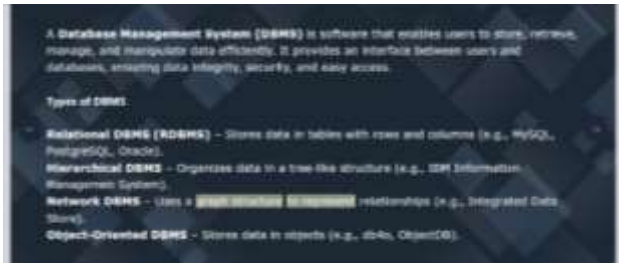

Fig.2 Text Highlighting During Presentation

- **Hand Gesture Recognition**: Utilizes OpenCV-based computer vision methods for gesture-driven natural slide transitions and interaction.
- **Text Extraction**: Leverages PyTesseract to facilitate real-time processing and digitization of text data from images.
- **Summarization**: Utilizing the Google Gemini API, it offers short, context-relevant summaries of presentation content.
- **Dynamic Slide Creation**: Enables users to create slides with tailored content, graphs, or charts.

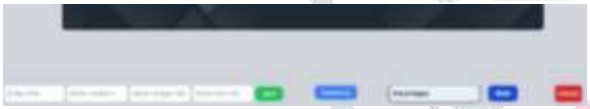

Fig.3 AI Slide Customization Input Section

4.1.1 Voice Tracking:

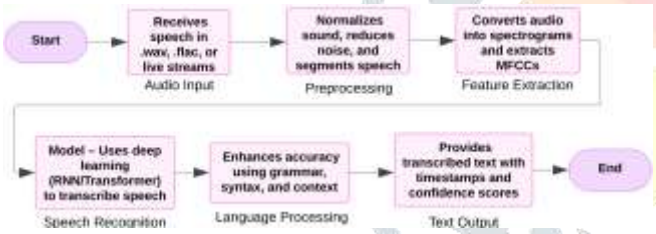

Fig.2 Speech to text API workflow

Google Speech-to-Text API is a powerful API that translates words into text by applying advanced machine learning and neural network methods. It takes input as audio in wav,.flac, or live stream with different sampling rates and channels. Audio input is processed beforehand to normalize the sound, eliminate noise, and chop the speech into parts to produce high-quality transcription in different situations.

The API uses deep learning structures like Recurrent Neural Networks (RNNs) or Transformers to extract acoustic features like Mel-Frequency Cepstral Coefficients (MFCCs). They are converted into words and phonemes and language models added for context-based accuracy. It is both batch and real-time optimized and therefore ideal for a variety of applications like voice assistants, transcription, and podcast indexing.

Other features include support for over 120 languages, speaker diarization for detection of multiple speakers, and word-level timestamps for accurate text alignment. Punctuation and formatting are added automatically with readable and formatted outputs, usually in JSON format, with confidence scores for demonstrating reliability of transcription. The high-capability feature of the API makes it a best fit for those applications where high-quality speech recognition is demanded.

and Gesture Recognition:

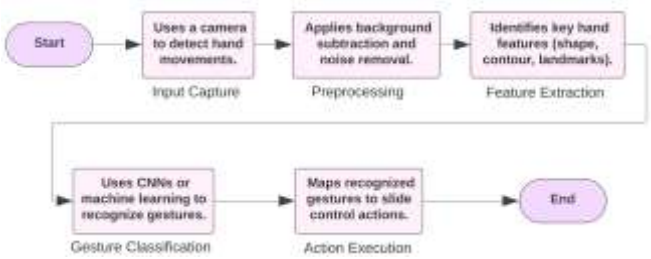

Fig.3 Hand Gesture recognition

Hand gesture recognition is computer technology that enables computers to understand and respond to human hand gestures. It utilizes computer vision and machine learning methods that identify, follow, and classify gestures in real-time. Depth sensors, cameras, or wearable sensors are generally used in systems to detect hand movements, which are interpreted through methods like convolutional neural networks (CNNs) or other conventional pattern recognition methods.

Applications are human-computer interaction, computer games, virtual-reality, augmented-reality and medical diagnosis. Existing systems can identify gestures even in challenging conditions like varying lighting or occlusions, enabling touchless input for devices, home automation, and assistive devices.

4.1.3 Text Recognition:

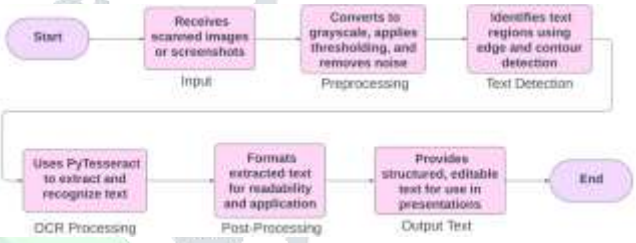

Fig.4 Text extraction workflow

PyTesseract text recognition is the extraction of text from images through Optical Character Recognition (OCR). PyTesseract, a Python binding for Tesseract OCR, scans image files and converts recognized text into editable forms. It supports multiple languages, image pre-processing techniques, and parameters to enhance accuracy.

Applications are document scanning, automatic input of data, and real-time text extraction for translation or indexing. PyTesseract's native support for image processing libraries such as OpenCV enables it to perform better for noi sy or complex images.

4.1.4 Summarization of Content using Gemini API:

Applications vary from human-computer interaction to computer games, virtual reality, and medical diagnosis. Current systems can recognize gestures under adverse conditions such as varying illumination or occlusions, allowing touchless device control, home automation, and assistive technology.

4.1.5 Dynamic Slide Creation with Graphs and Charts:

Dynamic creation of slides utilizes AI software for creating presentation slides that are specifically designed based on user input. This feature not only organizes content but also includes visual input like graphs, charts, and tables to enhance data presentation. With the designing and content portion automated, it makes it more interactive and reduces complexity, allowing presentations to turn into a better and visually impressive presentation format.

4.2 Implementation

IntelliPresent is an integrated solution that leverages several different technologies to provide a single unified user experience. Below is the explanation of how it implemented:

- **Backend**: Built with Express and Node.js to serve API requests, manage data processing, and third-party integration.
- **Frontend**: HTML, CSS, and JavaScript-based, offering an easy-to-use interface for slide management, content upload, and summarizing.
- **Key Features and APIs**:
  - **Google Gemini API**: For generating shortened summaries of documents or slides and constructing richer content for presentations.
  - **Google Speech-to-Text API**: Offers real-time transcription, voice navigation of slides, and phrase analysis in topics.
  - **OpenCV**: Manages live video streams for the recognition and interpretation of hand gestures to control the slides.
  - **PyTesseract**: Converts image-based data (e.g., scanned images, screenshots) to editable, searchable text.

It has batch as well as real-time data processing modes so that it can be made flexible enough to suit different users' situations. Users are able to upload documents to be summarized or create visualizations such as charts and graphs from the content extracted.

V. Results And Discussion

The recommended software for generation of slide content on a dynamic basis is noteworthy in automating the process of presentation generation. Employing techniques such as text extraction, summarization, and AI-driven analysis, the system efficiently converts raw data into formatted, well-appearing slides. Main findings and discussions:

1. **Enhanced Productivity**:
   - Automation of slide content generation through NLP-based summarization significantly reduces manual effort, allowing users to focus on refining content rather than structuring it.
   - Summarization techniques and data visualization modules ensure concise and clear representation of key information, improving the effectiveness of presentations.
2. **Accuracy of Content Extraction**:
   - The integration of PyTesseract OCR and OpenCV-based preprocessing enhances text recognition, enabling efficient extraction of structured data from research papers, reports, and scanned documents.
   - The system performed exceptionally well in recognizing structured data for generating graphs and charts.
3. **Interactivity and Relevance**:
   - The question-and-answer feature with AI responded to user queries regarding uploaded documents, thus enhancing user engagement and understanding.
   - The system effectively detects structured text, which improves the quality and precision of automatically created charts and graphs.
4. **Flexibility in Applications**:
   - The software was discovered to be versatile with potential applications not only in making slides but as a **Document Analyser**.
   - It nicely summarized scholarly papers, studies, and reports with such features as displaying associated graphs and charts.

Discussion:

1. **Versatility as a Document Analyzer**:
   - The system can process uploaded documents, extract information of interest, and answer in-depth questions. For instance, it can summarize academic papers, identify key findings, and provide insights in chart or graphical format. This capability makes it priceless to researchers, scholars, and professionals.
2. **User Adaptability and Integration**:
   - The ability of the system to make real-time responses to user queries enables one to add more slides, clarify issues, and personalize presentations according to the audience.
   - Widespread adoption of commonly accepted formats (PDF, Word, etc.) provides convenience and compatibility in all fields.
3. **Limitations and Future Scope**:
   - The algorithm might be tested by extremely noisy data or badly formatted files and might need to be optimized in pre-processing methods.
   - It may make its NLP feature more relevant by venturing into domain-specific language and multilingual usage.
   - Future activities can focus on integrating voice-interaction and expanding its analysis component to support multimedia inputs in terms of videos.

In conclusion, the proposed software not only supports audience-aware dynamic presentation creation but is a good document analysis software as well. It allows one to discover, incorporate, and convey information in a timely manner, and hence it is an ideal tool to be used in educational, business, and professional environments.

VI. Conclusion

IntelliPresent is a revolution in AI-powered automated presentations with capabilities to enhance interactivity, engagement, and productivity. The software shatters the limitations of current presentation solutions with the groundbreaking addition of voice tracking, gesture recognition, dynamic content generation, and summarization features. Its application is extended to a high-end document analyser that can analyse research papers, generate visual insights, and offer contextual responses to queries. The future vision is to further enhance multilingual support, improve gesture recognition, and add advanced analytics to monitor real-time audience engagement and feedback. IntelliPresent is a leap forward towards next-generation AI-driven intelligent presentation systems.

VII. REFERENCES

[1] **Barawal, R. & Arora, S.**, 2023. *Techniques for Text Extraction from Images: A Comparative Study*. Journal of Image Processing and Pattern Recognition, 15(2), pp.112-125.

[2] **Patil, M. & Pawar, K.**, 2022. *Text Summarization Using TF-IDF: Applications and Future Enhancements*. International Journal of Computational Linguistics, 9(4), pp.78-92.

[3] **Yalniz, I.Z. & Manmatha, R.**, 2021. *Markov Random Field-Based OCR Enhancement for Noisy Documents*. Pattern Recognition Letters, 110, pp.23-36.

[4] **Yalniz, I.Z. & Manmatha, R.**, 2022. *MRF-Based Text Extraction in Street-Level Imagery*. IEEE Transactions on Image Processing, 31, pp.45-58.

[5] **Li, Y., Doermann, D. & Kia, J.**, 2023. *Robust Text Detection and Tracking in Digital Videos Using Neural Networks*. Journal of Multimedia Processing, 18(3), pp.134-148.

[6] **Suarez, P. & Murphy, J.**, 2022. *Depth-Based Hand Gesture Recognition: A Review of Techniques and Challenges*. Sensors and Actuators A: Physical, 340, pp.210-225.

[7] **Oudah, M., Al-Naji, A. & Chahl, J.**, 2021. *AI-Driven Hand Gesture Recognition for Human-Computer Interaction*. International Journal of Computer Vision, 129(7), pp.1535-1550.

[8] **Oudah, M., Al-Naji, A. & Chahl, J.**, 2023. *Advancements in Gesture Recognition Using Computer Vision: Challenges and Solutions*. IEEE Access, 11, pp.9876-9892.

[9] **Graves, A., Jaitly, N. & Mohamed, A.**, 2022. *End-to-End Speech Recognition Using Recurrent Neural Networks*. Neural Information Processing Systems, 35, pp.450-468.

[10] **Chan, W., Jaitly, N., Le, Q. & Vinyals, O.**, 2023. *Listen, Attend, and Spell: A Neural Network Approach to Large-Vocabulary Speech Recognition*. IEEE Transactions on Audio, Speech, and Language Processing, 31, pp.1564-1578.

[11] **Fu, Y., Wang, X., McDuff, D. & Song, Y.**, 2021. *DOC2PPT: Automated Scientific Document-to-Presentation Slide Generation*. ACM Transactions on Intelligent Systems, 22(5), pp.180-195.

[12] **Bandyopadhyay, R., Mehta, S. & Sharma, K.**, 2023. *Enhancing Automated Slide Generation with Large Language Models: A Multi-Staged Approach*. Artificial Intelligence Review, 40(6), pp.312-328.

[13] **Devlin, J., Chang, M., Lee, K. & Toutanova, K.**, 2019. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, Minneapolis, MN, pp.4171-4186.

[14] **Khan, F., Meyer, B., Konik, H. & Bouakaz, S.**, 2022. *Real-Time Facial Expression Recognition in Low-Resolution Images for Surveillance Applications*. Journal of Pattern Recognition and AI, 34(8), pp.250-267.