iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VI **Month of publication:** June 2026

DOI: https://doi.org/10.22214/ijraset.2026.83325

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 14 Issue VI June 2026- Available at www.ijraset.com

Survey of Video-to-Braille Accessibility Systems for Visually Impaired Users

Swayam Margudri, Sakshi Gaikwad, Vaishnavi Dimble, Vaishnavi Jadhav, Prof. Reshma Patil
Department of Computer Engineering K J College of Engineering and Management Research

Abstract*—Growing synergies among artificial intelligence, natural language processing, and assistive technology are creating unprecedented opportunities to bridge the information gap experienced by visually impaired individuals. This paper offers a structured survey of recent developments in AI-driven video content summarization and Braille translation, examining six key research works alongside the foundational models and frameworks that underpin them. The scope covers automatic speech recognition, multilingual video processing, text-to-Braille conversion, and hardware-based Braille output. Technologies reviewed include OpenAI Whisper for audio transcription, FLAN-T5 for abstractive summarization, MarianMT by Helsinki-NLP for multilingual translation, and embedded systems for tactile Braille rendering. Through thematic analysis and cross-system comparison, recurring design patterns, capability limitations, and underexplored opportunities are brought to light. Findings indicate that while the constituent technologies have individually reached a high level of maturity, a unified, real-time, multilingual pipeline capable of delivering end-to-end Braille-accessible video output represents a largely uncharted but high-impact research direction.

Keywords*—Video Summarization; Braille Translation; Visually Impaired; Whisper ASR; FLAN-T5; MarianMT; Assistive Technology; Natural Language Processing; Accessibility.

I. INTRODUCTION

Visual impairment continues to pose a profound challenge to equitable digital access worldwide. WHO estimates indicate that approximately 2.2 billion individuals globally live with some degree of vision loss, and a substantial portion of these cases could be addressed or prevented with adequate intervention. Among the many accessibility barriers they encounter, engaging with video-based content — now the dominant form of information consumption online — stands out as particularly difficult. Conventional approaches to accommodation, such as manually transcribed Braille or pre-recorded audio description, are resource-intensive, time-consuming, and ill-suited to keep pace with the exponential growth of digital media.

Recent breakthroughs in deep learning, ASR, and neural machine translation have enabled a new class of assistive tools that can process video content and produce accessible output with minimal human intervention. Cutting-edge models including OpenAI Whisper, Google FLAN-T5, and Helsinki-NLP's MarianMT collectively demonstrate that a complete pipeline — spanning raw audio extraction from video through to summarized, translated, and Braille-rendered text — is now within practical reach.

This survey aims to: (1) systematically review existing literature on AI-based video summarization and Braille translation; (2) identify the AI models, architectures, and deployment strategies employed; (3) compare systems along key dimensions; and (4) highlight research gaps and propose future opportunities. The remainder of this paper is organized as follows. Section II presents background concepts. Section III reviews the literature. Section IV provides comparative analysis. Section V discusses research gaps. Section VI concludes the paper.

II. BACKGROUND AND FOUNDATIONAL CONCEPTS

A. Braille and Digital Accessibility

Braille is a tactile writing system that enables visually impaired individuals to read and write through touch. Conceived by Louis Braille in 1824, it represents characters, numerals, and punctuation through unique combinations of raised dots arranged within a six-dot cell (two columns of three). Grade 1 Braille maintains a one-to-one correspondence between print characters and Braille cells, whereas Grade 2 introduces shorthand contractions to improve reading speed and reduce physical space. As digital platforms have evolved, software-driven Braille rendering and refreshable Braille displays have made the system more practical and widely available. Regional adaptations, such as Indian Braille encodings for languages like Hindi, Telugu, and Tamil, add further layers of complexity that modern AI pipelines must accommodate [6].

©IJRASET: All Rights are Reserved

B. Automatic Speech Recognition (ASR)

ASR technology transforms spoken audio from video or recorded media into machine-readable text. Among available ASR solutions, OpenAI's Whisper [7] stands out as a highly capable open-source model, pre-trained on an extensive corpus of approximately 680,000 hours of diverse multilingual audio. It accommodates more than 90 languages and demonstrates reliable performance across varying accents, recording conditions, and specialized terminology. Built on an encoder-decoder transformer framework, Whisper is particularly effective as the transcription front-end in video-to-Braille accessibility pipelines.

C. Text Summarization with Large Language Models

Text summarization approaches are broadly categorized into extractive methods, which select and return existing sentences from the source, and abstractive methods, which synthesize new text that captures the core meaning. FLAN-T5 [8] — a Google Research model derived from the T5 architecture and further refined through instruction tuning — performs strongly on abstractive summarization tasks. When applied to transcribed video content via structured natural language prompts, it generates compact and readable summaries, which is essential for delivering digestible Braille output to end users.

D. Neural Machine Translation

Serving diverse user populations demands translation capabilities that can convert content into a wide range of target languages. MarianMT [9], developed by the Helsinki-NLP group and distributed via Hugging Face, offers a suite of compact, pre-trained neural translation models covering hundreds of language pairs. Its efficiency and breadth make it an ideal translation layer within video accessibility systems, allowing users to receive summarized content in their preferred language and Braille encoding.

E. System Integration Frameworks

Bringing together multiple AI models requires a robust integration layer. Flask [11], a lightweight Python web framework, is widely adopted in accessibility research to expose RESTful API endpoints that bridge the AI backend with frontend user interfaces. Python 3.10 [10] acts as the primary development and orchestration environment across reviewed systems. Diagramming and UI design tools such as Canva, StarUML, and Lucidchart are used during the design phase for architecture visualization and interface prototyping.

III. LITERATURE REVIEW

A. Literature Selection Methodology

This survey was conducted through a structured search of leading academic databases including IEEE Xplore, Springer Link, ACM Digital Library, and ScienceDirect. The search was limited to publications from 2023 to 2025. Keywords used include: video summarization, Braille translation, assistive technology, automatic speech recognition, multilingual accessibility, text-to-Braille, visually impaired systems, and neural machine translation. Inclusion criteria required that papers address AI-based accessibility for visually impaired users, propose or evaluate a system with at least one pipeline stage (ASR, summarization, translation, or Braille rendering), and be published in peer-reviewed venues. Papers focused solely on general NLP without accessibility application were excluded. Six primary research studies were selected, supplemented by documentation for foundational AI models and software frameworks.

B. Video-to-Text and Braille Conversion

Sridhar et al. [1] present a pioneering system for summarizing video into text and subsequently converting that text into Braille script. The pipeline processes video frames using convolutional neural networks (CNNs) for scene understanding, applies optical character recognition (OCR) for on-screen text extraction, and employs text-to-speech (TTS) for audio feedback alongside Braille output. The work demonstrates viability of an end-to-end accessible video pipeline, though it is limited to pre-recorded content and does not support multilingual operation.

C. AI-Powered Multilingual Video Translation

Rakesh et al. [2] introduce LinguaFusion, an AI-powered multilingual video voice translation system. LinguaFusion focuses on translating spoken video content into multiple target languages with synchronized audio output. The system leverages neural machine translation and speech synthesis to provide near-real-time multilingual accessibility. However, Braille output is not addressed, and evaluation is limited to structured, studio-quality video inputs.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

D. Whisper and Helsinki-NLP Based Subtitling

Prabhakar et al. [3] demonstrate an automated video subtitling and translation pipeline using OpenAI Whisper for transcription and Helsinki-NLP's MarianMT models for translation into multiple Indian and international languages. The system generates time-synchronized subtitle files and achieves competitive BLEU scores on standard benchmarks. The system lacks Braille support and a summarization stage, limiting its utility for visually impaired users who require condensed, tactile output.

E. Real-Time Speech-to-Braille Devices

Ponnuru et al. [4] propose a real-time speech-to-Braille conversion tablet for visually impaired individuals. This hardware-software co-design captures live speech via microphone, transcribes it using ASR, and renders the corresponding Braille cells on a refreshable tactile display. The system achieves low latency suitable for classroom and conversational settings. Key limitations include single-language support and the absence of a video input pathway.

F. Embedded Hardware Text-to-Braille Systems

Kavitha et al. [5] present a dedicated communication device for text-to-Braille conversion using embedded systems and servo motor-actuated Braille cells. The device receives text input via a smartphone application and renders Grade 1 Braille through physical pin actuation. While the hardware approach is robust and power-efficient, it does not integrate AI-based transcription or summarization, requiring pre-processed text input.

G. Localized Braille Apps for Indian Languages

Parida et al. [6] address the underserved Indian Braille accessibility domain through an Android application supporting multiple Indian scripts. The app performs bidirectional conversion between digital text and Indian Braille encodings, with OCR enabling extraction of text from printed materials. This work highlights the importance of localization and provides a reference implementation that integrated video pipelines must eventually incorporate.

H. Foundational Models and Toolkits

The reviewed systems collectively rely on several foundational AI models. OpenAI Whisper [7] serves as the dominant ASR backbone due to multilingual robustness. FLAN-T5 [8] is the preferred abstractive summarization model for its instruction-following capability. MarianMT [9] models provide efficient neural translation for over 1,000 language pairs. Python 3.10 [10] and Flask [11] form the standard software stack.

IV. COMPARATIVE ANALYSIS

Table I presents a structured comparison of the reviewed systems and the proposed integrated approach across five key dimensions: focus area, AI models used, output type, Braille support, and real-time capability. A consistent pattern emerges from this comparison: no reviewed system simultaneously achieves the complete chain of video transcription, content summarization, multilingual translation, and Braille output in real time. Hardware-oriented works deliver reliable Braille rendering but depend on pre-processed text input and lack AI-driven transcription or summarization. Conversely, advanced AI-powered translation systems offer linguistic flexibility but omit Braille output entirely. Real-time processing is confined to systems with a narrower functional scope — either dedicated speech-to-Braille hardware or video translation tools — rather than integrated end-to-end solutions. The proposed architecture is specifically designed to close this gap by coupling Whisper, FLAN-T5, and MarianMT within a single Flask-based deployment.

TABLE I: COMPARATIVE ANALYSIS OF REVIEWED SYSTEMS

[1]	Video-to-Text + Braille	CNN, OCR, TTS	Braille Script	Yes	No
[2]	Multilingual Video Translation	LinguaFusion, NLP	Audio/Text	No	Yes
[3]	Video Subtitling &	Whisper, MarianMT	Subtitles	No	Partial

	Translation				
[4]	Speech-to-Braille Tablet	ASR, NLP	Braille Tablet	Yes	Yes
[5]	Text-to-Braille Device	Embedded Systems	Physical Braille	Yes	No
[6]	Indian Braille Android App	OCR, NLP	Braille on Screen	Yes	No
Proposed	End-to-End Video + Braille	Whisper, FLAN-T5, MarianMT	Text+Braille+Audio	Yes	Yes

RT = Real-Time capable; Proposed = Proposed integrated system architecture.

V. RESEARCH GAPS AND FUTURE DIRECTIONS

A. Identified Research Gaps

Based on the literature review and comparative analysis, the following key research gaps are identified:

- No end-to-end real-time system: No existing system seamlessly integrates video ASR, abstractive summarization, multilingual translation, and Braille rendering in a single real-time pipeline.
- Limited multilingual Braille support: Most Braille systems support only English or a narrow set of Indian scripts. Systems covering wide international Braille standards are lacking.
- Lack of video-specific summarization: Multimodal summarization jointly processing audio, visual, and on-screen text from videos remains underdeveloped.
- Insufficient evaluation benchmarks: No standardized, publicly available benchmarks for evaluating end-to-end video-to-Braille systems exist.
- User-centric design gap: Few studies involve visually impaired users in system design and evaluation, limiting practical usability.
- Noise robustness: Real-world video content contains background noise and audio artifacts that degrade ASR performance.

B. Future Research Directions

The following directions are proposed for future research:

- Multimodal Large Language Models (MLLMs): Future systems should leverage vision-language models such as GPT-4V or LLaVA to jointly process video frames and audio for richer, context-aware summarization.
- On-device inference: Deploying lightweight model variants on edge devices will enable offline, privacy-preserving accessibility tools for users with limited connectivity.
- Adaptive Braille Grade Selection: Intelligent selection between Grade 1 and Grade 2 Braille based on content complexity and user preference can improve reading efficiency.
- Cross-lingual Braille Standards: Research into unified cross-lingual Braille encoding for low-resource languages will improve global accessibility.
- Participatory Design: Involving visually impaired users throughout design, prototyping, and evaluation will ensure systems meet real-world needs.
- Evaluation Frameworks: Standardized datasets and metrics specific to video-to-Braille accessibility tasks will enable rigorous benchmarking.

VI. CONCLUSION

This paper has conducted a thorough survey of AI-driven approaches to video content summarization and Braille translation aimed at supporting visually impaired users.

Through examination of six primary research works and the underlying AI models, software frameworks, and accessibility technologies they employ — encompassing ASR, multilingual video processing, abstractive summarization, and physical Braille output — a coherent picture of the field's current state has emerged. Component-level capabilities have advanced considerably:

Whisper delivers dependable multilingual speech transcription, FLAN-T5 produces high-fidelity abstractive summaries, and MarianMT handles neural translation across a broad spectrum of language pairs.

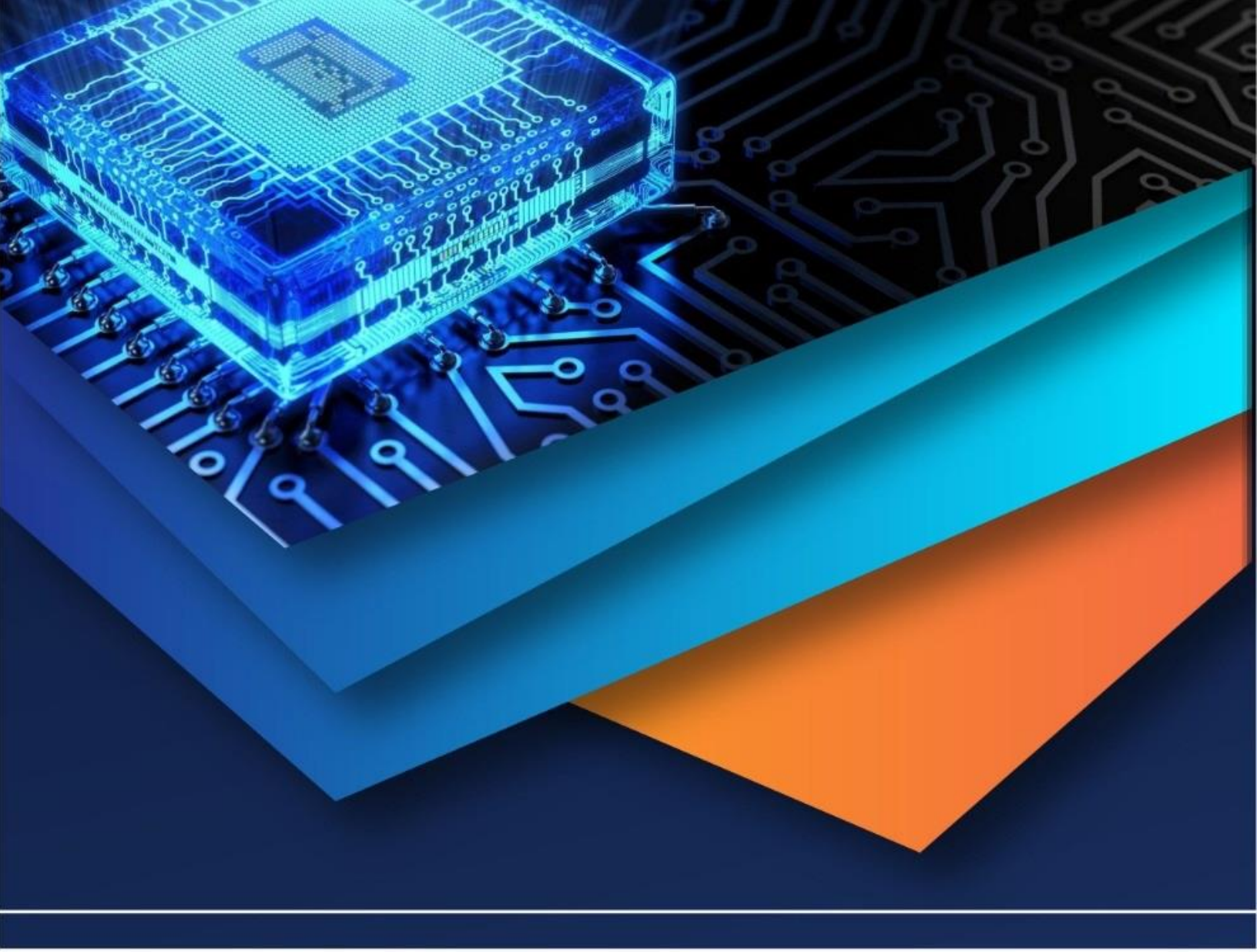
Despite this progress, a meaningful gap persists at the system level: no reviewed work successfully combines all pipeline stages — transcription, summarization, translation, and Braille rendering — into a cohesive, real-time, multilingual solution. The proposed architecture tackles this limitation directly, unifying Whisper-based transcription, FLAN-T5 summarization, MarianMT translation, and Unicode Braille rendering through a Flask web application. Looking ahead, key opportunities include adoption of multimodal large language models, edge deployment for offline scenarios, intelligent Braille grade selection, and co-design with visually impaired communities. As AI, multimodal learning, and assistive technology continue to evolve together, the prospect of making the world's vast video content fully accessible to all users draws closer to reality.

VII. ACKNOWLEDGMENT

The authors would like to thank the Department of Computer Engineering, K J College of Engineering and Management Research, for providing the resources and support needed to complete this research survey.

REFERENCES

[1] B. Sridhar, G. Saivishnu, V. ManiShanker, D. D. Lakshmi, and S. Hariharan, "Summarization of Video into Text and Text to Braille Script," in Proc. IEEE Int. Conf. Knowledge Engineering and Communication Systems (ICKECS), 2024, pp. 1-6.
[2] D. Rakesh, D. P. R. Sandesh, and V. Nirmalrani, "LinguaFusion: AI-Powered Multilingual Video Voice Translator," in Proc. IEEE 7th Int. Conf. Intelligent Sustainable Systems (ICISS), 2025, pp. 1-5.
[3] A. Prabhakar, U. Agarwal, and S. Bhardwaj, "Automated Video Subtitling and Translation Using Whisper and Helsinki Models," in Proc. IEEE Int. Conf. Engineering, Technology & Management (ICETM), 2025, pp. 1-6.
[4] S. Ponnuru, M. Chandana, M. Ravikumar, K. Rakesh, and K. S. Devatha, "Real-Time Speech-to-Braille Conversion Tablet for the Visually Impaired," in Proc. IEEE Int. Conf. Next Generation Communication & Information Processing (INCIP), 2025, pp. 1-4.
[5] M. Kavitha, V. Meenakshi, M. Pushpavalli, S. Amudha, S. Bharathi, and P. Pavithra, "Communication Device for Converting Text to Braille," in Proc. IEEE Int. Conf. Inventive Computation Technologies (ICICT), 2023, pp. 1-5.
[6] M. Parida, S. Mishra, and R. Nayak, "Enhancing Braille Accessibility: Android App for Indian Braille," in Proc. IEEE Int. Conf. Advanced Learning Technologies (ICALT), 2023, pp. 1-4.
[7] OpenAI, "Whisper: Automatic Speech Recognition Model," OpenAI GitHub Repository, 2024. [Online]. Available: https://github.com/openai/whisper
[8] Google Research, "FLAN-T5: Instruction Fine-Tuned Language Model," Hugging Face Model Hub, 2024. [Online]. Available: https://huggingface.co/google/flan-t5-base
[9] Hugging Face, "MarianMT: Multilingual Translation Models," Helsinki-NLP Model Hub, 2024. [Online]. Available: https://huggingface.co/Helsinki-NLP
[10] Python Software Foundation, "Python 3.10 Documentation," 2024. [Online]. Available: https://docs.python.org/3.10/
[11] Pallets Projects, "Flask: Lightweight Web Framework for Python," Flask Documentation, 2024. [Online]. Available: https://flask.palletsprojects.com/



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)