

Text and Speech Recognition Technologies



OCR Optical Character Recognition

The procedure of converting handwritten or typewritten text in digital images into editable and searchable data (Chaudhuri et al., 2016; Sarika et al., 2021).

Applications (Chaudhuri et al., 2016; Sarika et al., 2021)

- Assistance for the visually impaired
- Automated license plate recognition
- Automated cartography
- Language translation
- Banking applications
- Document digitization

Challenges

Difficulty achieving high correct recognition due to

- complex scenes,
- varying lighting conditions,
- skewness and rotation,
- blurring and degradation,
- diverse fonts and styles,
- multilingual settings,
- and damaged documents (Awel & Abidi, 2019).

Educational Implications

- Text digitalization
- Accessibility and inclusivity
- Document administration
- Automated grading
- Language learning

History (Chaudhuri et al., 2016; OCRology, 2021)

Mid-1940s

The inception of the first character recognizer



Mid-1950s

OCR machines became commercially available

1960-1965

First generation of OCR systems

Mid-1960s & early 1970s

Second generation of OCR systems



Mid-1970s

Third generation of OCR systems

Till 1986

A few thousand OCR machines were sold

During the 1990s

Substantial progress in OCR systems occurred, empowered by the continuous growth of information technologies



Nowadays

The advancement of OCR systems continues with more powerful technologies

ASR Automatic Speech Recognition

A technology that converts spoken language into written text. The goal of ASR is "for a computer to be capable to perceive speech, recognize, and take action upon spoken words" (Vadwala et al., 2017, p.31).

History (Wang et al., 2019)

In 1952

The first truly complete speech recognizer

During the 1950s

Speech recognition systems were limited to recognizing single words or vowel

During the 1960s

Specialized hardware was constructed by Japanese laboratories for speech recognition tasks



Around the 1970s

Rapid evolution of speech recognition for speaker-specific tasks with isolated words and small vocabulary

Mid-1980s

HMM technology made a breakthrough in Large Vocabulary Continuous Speech Recognition (LVCSR)



During the 1990s & early 2000s

The HMM-GMM framework dominated the field of speech recognition

More recently

Deep learning techniques brought remarkable improvements.

In 2011

New Framework (CD)-DNN-HMM achieved significant performance gains

Applications

- Voice command systems (Vadwala et al., 2017)
- Dictation systems
- Accessibility services (Fendji et al., 2022)
- Telecommunications (Ibrahim & Varol, 2020)
- Transcription services (Ibrahim & Varol, 2020)
- Language learning

Challenges

To attain optimal precision, proficient speech recognition systems need to address challenges associated with

- vocabulary,
- channel variability,
- diverse utterance approaches,
- diverse utterance styles,
- and speaker models (Vadwala et al., 2017).

Educational Implications

- Remote learning
- Accessibility and inclusivity
- Multilingual education
- Automated grading
- Language learning

References

- Awel, M. A., & Abidi, A. I. (2019). Review on optical character recognition. International Research Journal of Engineering and Technology (IRJET), 6(6), 3666-3669.
- Chaudhuri, A., Mandaviya, K., Badelia, P., & Ghosh, S. K. (2016). Optical character recognition systems. In Studies in fuzziness and soft computing (pp. 9-41). https://doi.org/10.1007/978-3-319-50252-6_2
- Fendji, J. L. K. E., Tala, D. C., Yenke, B. O., & Atemkeng, M. (2022). Automatic speech recognition using limited vocabulary: A survey. Applied Artificial Intelligence, 36(1), 2095039.
- Ibrahim, H., & Varol, A. (2020, June). A study on automatic speech recognition systems. In 2020 8th International Symposium on Digital Forensics and Security (ISDFS) (pp. 1-5). IEEE.
- OCRology. (2021, December 10). A quick history of OCR - OCRology - Medium. Medium. <https://medium.com/ocr-ology/a-quick-history-of-optical-character-recognition-ocr-c916d58e2170>
- Sarika, N., Sirisala, N., & Velpuru, M. S. (2021, January). CNN based optical character recognition and applications. In 2021 6th International conference on inventive computation technologies (ICICT) (pp. 666-672). IEEE.
- Vadwala, A. Y., Suthar, K. A., Karmakar, Y. A., Pandya, N., & Patel, B. (2017). Survey paper on different speech recognition algorithm: challenges and techniques. Int J Comput Appl, 175(1), 31-36.
- Wang, D., Wang, X., & Lv, S. (2019). An overview of End-to-End Automatic Speech recognition. Symmetry, 11(8), 1018. <https://doi.org/10.3390/sym11081018>