Susan Kuretski
skure2@illinois.edu
CS410 - Fall 2020
Technology Review

A comparison of Google Vision and Tesseract Optical Character Recognition

Introduction

The purpose of this review is to compare Google Vision and Tesseract optical character recognition, OCR. Tesseract began its development in 1984 at Hewlett-Packard as a PhD research project. After more than a decade in development, the project was sent to the University of Nevada Las Vegas for the 1995 Annual Test of OCR Accuracy. Tesseract showed great improvements in accuracy compared to other projects. After this, Tesseract collected some dust, but in 2006 it was released as open source and is now supported by Google. Since then, Tesseract has had a few major releases to improve its accuracy and ability to support different languages.

Google Vision API was released to the public in 2016. While the very detailed underpinnings of the API and OCR are proprietary, comparisons can be made from what is known.

How Does OCR Work

Many fine details of OCRs differ, but on a high level, they share some commonality. Most OCRs will perform some sort of image enhancements, adjustments, or preprocessing prior to text recognition in order to increase their chances of success. This can be done by de-skewing, contrast enhancement, line detection or removal, or scaling to name a few.

After preprocessing, text recognition begins, which is typically done in one of two ways. The first approach is matrix matching where images are compared to a library of characters pixel-by-pixel. This technique was popular among the first OCRs. The second approach is feature extraction where characters are decomposed into objects like lines and circles where orientation and direction matter. These decomposed parts of characters are then compared with

a vector representation of known characters. To find the best candidate matches, nearest neighbor classifiers are then used.[1]

Tesseract implements text recognition in two passes. The first pass is the feature extraction as described above. The second pass takes characters recognized with high confidence from the first pass to aid in recognition of the remaining unknown characters. This is known as adaptive recognition.

Google Vision has five stages to its process. First, there is the initial text detection step where a "CNN-based subsystem detects and localizes lines of text by generating a pixel-level heatmap of text likelihood that is used to generate a set of bounding boxes.[2]"[2] The bounding boxes then correspond to a line of text. The second stage is direction identification where the lines of characters are assigned north, south, east, or west. The third stage identifies the main writing system. It is assumed that each line has a dominant writing system. The fourth stage is the actual text recognition where a log-linear framework is used which combines "an inception style optical model and N-gram character based language model as the major inputs.[2]" Finally, the fifth stage performs layout analysis, which infers structure about the document such as reading order, titles, footers, and paragraphs.

Cost

Because Tesseract is open source, its usage is free which can be a huge benefit for research, education, and startups where funds may be limited.

At time of writing, the first 1000 units per month are free with Google Vision API[3]. After that, it ranges between $1.50-$3.50 per 1000 units, up to 5,000,000/month. After 5,000,000 units, prices range between $0.60 - $1.50 per 1000. For optical character recognition, the price is $1.50 and $0.60 before and after 5,000,000 respectively. The price point for 1,001+ units could quickly add up for those on a tight budget.

Accuracy

The accuracy of results depends on the input into OCR. Most systems fare well with clear, large, contrasting, English printed text. However, handwriting, heavily edited text, and non-traditional text layouts can prove to be difficult for OCRs. Language is also an important consideration as

---

[1] Wikipedia. "Optical Character Recognition." 4 11 2020, https://en.wikipedia.org/wiki/Optical_character_recognition.
[2] Walker, Jake, et al. "A Web-based OCR Service for Documents." *13th IAPR International Workshop on Document Analysis Systems*, 2018, https://das2018.cvl.tuwien.ac.at/media/filer_public/85/fd/85fd4698-040f-45f4-8fcc-56d66533b82d/das2018_short_papers.pdf.
[3] Google. "Google Cloud Vision API Documentation." *Pricing*, 22 06 2020, https://cloud.google.com/vision/pricing.

not all OCRs support the same languages. In 2018, Walker and et. al. published their performance comparison of Google Vision OCR vs Tesseract 4.0. Below is a table of results.

Language	**Books**			**Web**		
	#Lines	N-CER (%)		#Lines	N-CER(%)	
		Tesseract	**Google**		**Tesseract**	**Google**
Arabic	946	14.0	4.8	4208	54.8	19.4
English	1000	1.0	0.6	4868	44.0	15.6
Hindi	1067	5.4	2.5	3726	49.3	20.6
Japanese	773	28.0	4.9	3256	57.5	17.1
Russian	864	1.7	1.2	3883	36.2	16.7

N-CER is the normalized character error rate. Books constituted books taken from Google Books and Web constituted images taken from the Web containing text.

Across the board, Google Vision performed better in accuracy with lower rates of N-CER.

Continued Support and Updates

Both engines have continued to provide various updates and enhancements. Google Vision has had seven updates in the year 2020. In October 2018, Tesseract had a large release with version 4.0.0 which introduced an engine using a neural network system based on LSTMs (long short term memory).

Google Vision currently supports 60 languages with many others in experimental phase or are mapped to other languages. In its latest release, Tesseract has support for 129 languages.

Conclusion

While there is no clear answer to the question of which OCR to use, both Tesseract and Google Vision offer advantages. Tesseract may be a better choice for educational, research, exploratory, or startup purposes due to cost. Moreover, Tesseract offers flexibility to explore enhancements or ideas since the code base is public. One drawback compared to Google

Vision is that it is written in C++ and support for other programming languages is provided via community written wrappers.

Google Vision offers a polished, more accurate product due to its commercial support and dedicated team. It can be used with C#, Go, Java, Node.js, PHP, Python, and Ruby, giving it a commercial edge. However, due to its black box-ness and costs, it may deter potential users.

Since 1984, OCRs have made many gains in terms of sophistication, accuracy, speed, and availability. As deep learning, computer vision, and natural language processing continues to evolve, there is a promising future for optical character recognition.

Sources

Google. “Google Cloud Vision API Documentation.” *Pricing*, 22 06 2020, https://cloud.google.com/vision/pricing.

Smith, Ray. “An Overview of the Tesseract OCR Engine.” *Proc. Ninth Int. Conference on Document Analysis and Recognition (ICDAR), IEEE Computer Society*, 2007, pp. 629-633, https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/33418.pdf.

Walker, Jake, et al. “A Web-based OCR Service for Documents.” *13th IAPR International Workshop on Document Analysis Systems*, 2018, https://das2018.cvl.tuwien.ac.at/media/filer_public/85/fd/85fd4698-040f-45f4-8fcc-56d66533b82d/das2018_short_papers.pdf.

Wikipedia. “Optical Character Recognition.” 4 11 2020, https://en.wikipedia.org/wiki/Optical_character_recognition.