August 28, 2018

AUGMENTED SEARCH FOR PHYSICAL DOCUMENTS

HP INC

Augmented Search for Physical Documents

In recent years, we are used to searching words in electronic documents such as PDFs, Powerpoints, Word Documents, and so on. Whenever we have a large document and we need to find a specific word or sentence - with a few keyboard commands - we can easily find it, and all occurrences are highlighted to make locating it easier. On the other hand, when we think about physical documents (e.g. printed documents, magazines), there is no such mechanism that allows us to search words or sentences on it. Our method helps creating a search query on physical documents by allowing the user to point a camera to the document, search for words or sentences, and the solution will highlight them with augmented reality technology.

1   Introduction

When a user reads a text in a paper he cannot look for specific words or sentences unless he reads the entire paper or if the paper has some words or topics highlighted in a visual manner that calls the user's attention. This invention aims to save user time when he doesn't want to read an entire paper just to get a piece of information that's underlying in the text. It enables:

- Searching words or sentences on documents in a physical media;
- Highlighting areas of the physical paper via an augmented reality mechanism;
- Richer interaction in cases where the user is flipping pages.

For instance, let's say that the user has a scientific paper on his desk and that he wants to find a specific topic definition written somewhere in that paper. He could open an application that makes use of our invention, type the search query in a textbox (or maybe say it by voice), point the camera to the document, and wait for the processing to happen. The parts of the text that match such query in the document would then be highlighted in the application. If the user changes the current page or points to another direction, the process will then be repeated for that new page.

If the user is flipping pages of a book for example, we could use our solution to enable the user to search the book and highlight the content using a platform that contains a

scanning and projection system. The user could then flip pages and the solution would continue executing the given search in the next pages and highlighting the content accordingly.
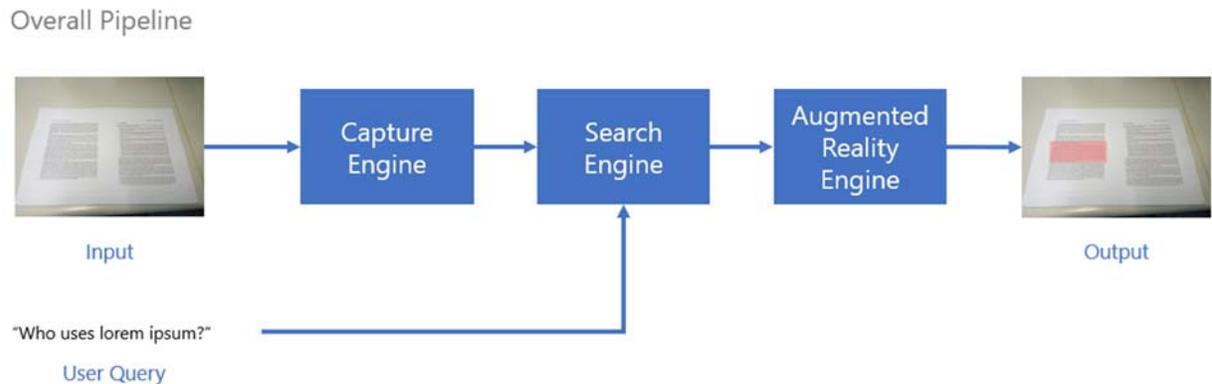
2   Description

Figure 1. Overview of the invention pipeline.

Our invention can run on a device such as a smartphone that has a camera and a display, and it can be described with three major components (**Figure 1**), namely:

- A Capture Engine, which can be divided in two parts
  - Actual "document capture", which involves taking a picture of the document, detecting the contours, dewarping it, and cropping it.
  - OCR, which processes the preprocessed captured image and extracts the text from it.
- A Search Engine, which can be divided in two parts
  - Word and sentence search, which is done via word/sentence matching between the query and the text - this search may be powered by Machine Learning.
- An Augmented Reality Engine, which, based on the search engine results, finds the position of the answer (matched text or sentence) on the image and highlights it in the real-world document representation on the device's screen, through augmented reality

○ The Augmented Reality Engine can also be seen as a projection system. Therefore we see it as both a projection system (document is highlighted physically) or an augmented reality one (document is highlighted digitally).

Capture Engine

The Capture Engine is depicted in **Figure 2**. After capturing the image of the document (e.g. via a smartphone or any other device that contains a camera), the image is preprocessed. Such preprocessing is to prepare to the OCR process. Hence, things such as dewarping and alignment are applied to make the document more readable. Next, OCR is applied to the given image, where the text and its position are extracted. Such position is going to be afterwards used by the Augmented Reality Engine to find the most suitable position to highlight the terms that were found.
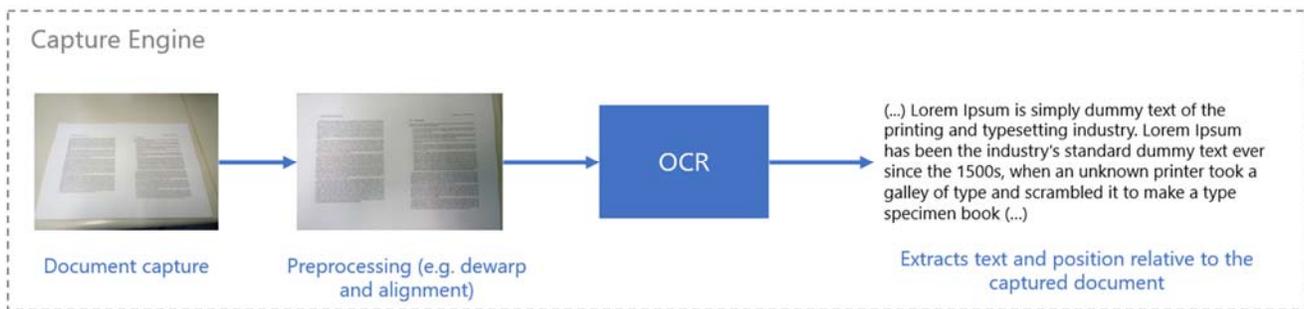

Figure 2. Overview of Capture Engine.

Search Engine

Search Engine is the component responsible for searching user queries in the extracted text. **Figure 3** shows an example of word search in a text extracted in the previous step (by the Capture Engine). It takes as input the raw text and a user query (in this case, a single word or a sentence) and outputs matches of the query - where the text was found. The search engine component can be based on, for instance, Machine Learning algorithms, which are able to handle more complex scenarios. For example, we could preprocess the sentence and use Word2Vec[2] as the representation for individual words or Paragraph2Vec[3] for entire sentences. By using Word2Vec, for instance, we can look for not only exact matches but also for synonyms and similar words. Such cases where we

match a synonym or a related word/sentence but not the word/sentence itself, could be highlighted with different colors when showing it to the user through the device.
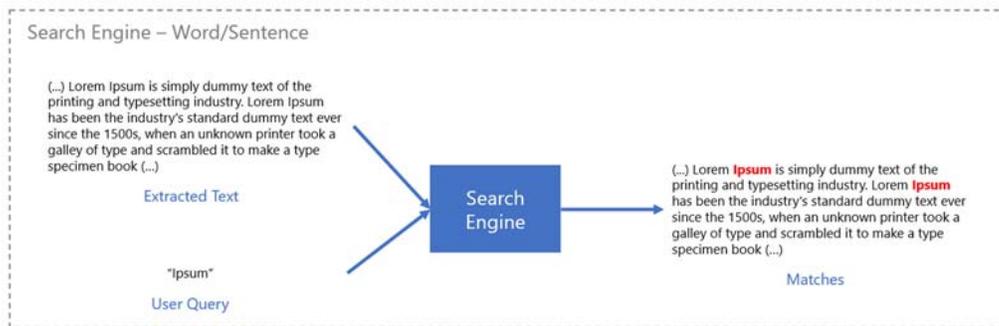

Figure 3. Search Engine example with word search.

Augmented Reality Engine

Finally, the last component is responsible for obtaining the answer/result of the query and highlighting it on the text. It takes as input the result of the Search Engine (answer and relative position to the document) and the current view of the document (same one that was used to execute the OCR), and highlights the text in the application. Such flow is exemplified in **Figure 4.** This can be done "live", through streaming the document view from the device's camera to the device's display so that the highlighted area would follow the document representation as the user moves the device with the camera.
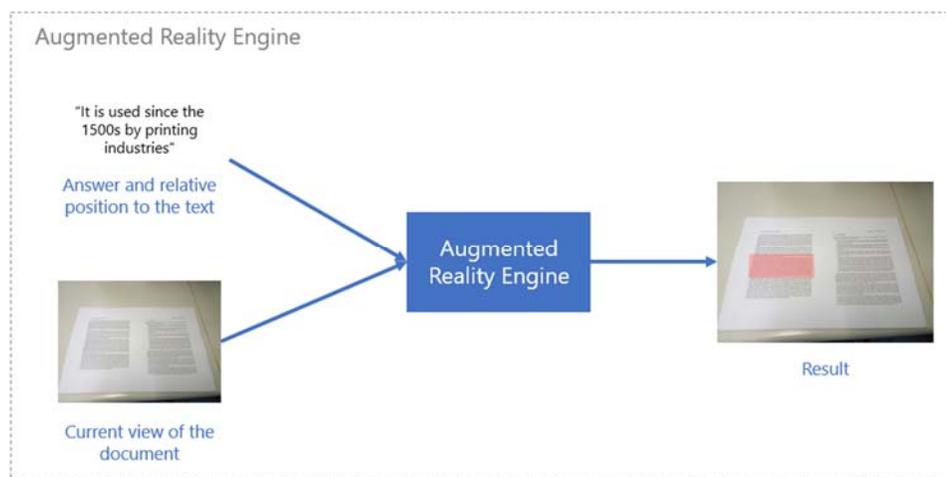

Figure 4. Augmented Reality Engine example.

The Augmented Reality Engine can be a projection system as well, where the highlight will be a projection of a different color over the physical document. We envision that it would be very useful for a scenario with scanning/projection system, where the user could place the document under the scanning device, ask a question via voice or type into a textbox, and our solution would parse such query and highlight the document via the projection system. It would, for example, enable the user to flip pages and see the results over different parts of the document.

3   Discussion and Conclusions

This invention solves a problem that affects many people namely doing search in physical documents. It is very interesting for a variety of applications, specially for mobile applications that process documents. It can also be embedded in different platforms that contain a scanning/projection interface and devices similar to Google Glass, where the user query is performed via voice and the interface highlights the matches in the user's field of view. We also believe that the same flow could be extended and applied to images, using text to perform image search.

Disclosed by Thomas Paula, Wagston Staehler, and Alexandre Santos, HP Inc.