Problem Statement

Today, avid readers can access their media in a variety of ways. E-readers are portable, and audio books allow listeners to perform other tasks while following a story; however, some traditionalists prefer the physical experiences associated with paper books. Within the course of a single novel, all three formats can be advantageous. For example, a user may need to go grocery shopping while in the midst of reading a paper book. If they were to switch to the audiobook format, they could continue enjoying the novel while driving a car. On day trips, a quick skim of a digital book in a cell phone might be preferable.

Manually synchronizing these formats can be cumbersome, especially since E-readers have different page schemes depending on font size, audiobook audio files aren’t precisely divided into separate paragraphs, and the same novel by different publishers have different page schemes.

Now that large databases are easily accessible through the cloud, a means of automatically syncing all three media formats is possible. The main challenge would be matching a chapter of a physical book to a timestamp in the corresponding audiobook or paragraph of an Ebook. We propose to solve this problem with image processing.

Project Scope/Goals

We will focus on extracting features from images of a physical book’s pages and matching them to specific timestamps of an audio book. Although we plan to implement a proof of concept with every page of a single book, the challenge is in scaling this to a database of thousands of books. Scanned pages of a physical book will train the classifier while pictures taken by an Android-based smartphone will be used to search for a page given a database features.

Some previous work has been accomplished in document image retrieval. In particular, Hull provided a feature extraction technique for text documents involving the wordlengths of horizontally and vertically adjacent words [1]. Also, Locally Likely Arrangement Hashing (LLAH) has been used as a mapping features to a database of documents [2].
We intend to do more literature research on document image retrieval (especially for large databases), combine and improve them by keeping memory need small at the same time. Our goals are to discover the optimal tradeoff between processing speed, database size, and matching accuracy, and to create a demonstration of this concept on an Android-based phone. The phone would take a picture of any page of a book and automatically play back the relevant time stamp of its corresponding audiobook.

**Work to be Completed**

Figure 1: General Algorithm

The general algorithm to be implemented can be found in Figure 1.

In terms of image processing, our algorithm must extract the specially defined descriptor of taken image (word lengths or locations, neighbors or possible geometric features could be used). The same feature extraction algorithm should be applied to the all images in the database. Then it must map and evaluate these features via pairwise matching. Timestamp of the matched page must be accurately computed so that the correct timestamp is selected with an acceptable error probability.

In real-world applications, the number of books (and thus, the database), might be too large and unstable to store in a single phone. We may be required to find a means through which Android can communicate with an external database that stores the knowledge of other pages descriptors. Retrieval system for mobile device should be implemented by C or Java (according to their performance).

Lastly, as part of the demonstration, we must develop the Android GUI and the Audiobook playback scheme.

**References**


[3, 4, 1, 2]