Motivation

- Pro Quo Books is an online, used book seller.
- They buy bulk books and list them for sale on various web sites.
- Upon receipt, books are fed through a conveyor belt system.
- Books with barcodes are scanned and processed automatically.
- Books without barcodes are manually identified, a slow and expensive process.
- Our system automates the identification of the non-barcoded books.

Lifecycle Model

- Evolutionary Delivery process
  - A mix between staged delivery and pure prototyping
  - Two-week intervals of prototyping followed by two-week intervals improving the original algorithms

System Architecture

- Book ID
- Identification Library
- Database
- Candidate Results
- Query Dataset
- Original Image
- Processed Image Data
- New Content
- Preprocessor Library
- Preprocess New Book
- Update Complete
- Database Updates

Future Work

- Use Optical Character Recognition (OCR) to identify books based on titles
- Provide a way to automatically generate and deploy new configurations and libraries
- Improve identification accuracy

Test Application – Example Results

- Use Optical Character Recognition (OCR) to identify books based on titles
- Provide a way to automatically generate and deploy new configurations and libraries
- Improve identification accuracy

Technologies

- MySQL
- Microsoft .NET
- AForge.NET

Step 1: Take Picture

- Books moving on the conveyor belt are weighed and have their picture taken. This information is sent to the identification system.

Step 2: Preprocessing

- The conveyor belt is cropped out of the image and image statistics, used in the identification step, are calculated.

Step 3: Set Pruning

- The weight and image statistics are used to quickly filter out books that are not matches. This prunes the dataset of 3 million books to a more manageable size.

Step 4: Identification

- The preprocessed imagery is compared with the images remaining in the dataset. Each book is given a confidence rating expressing how closely it matches the original book.

Step 5: Result

- The dataset book with the highest confidence rating is selected as the most likely match. Now identified, the original book can now be sorted and ultimately shipped to a customer.