**Title:**

Thickshake: Historical Image Classification System

**Author:**

Mark Shelton

**Supervisors:**

Joshua Hollick, Dr Andrew Woods, Debra Jones

**Abstract:**

The State Library of Western Australia (SLWA) Pictorial Collection consists of more than one million photographs, postcards, slides, albums, glass negatives, illustrations and art works covering all aspects of Western Australian history. Each year, thousands of items are added to this collection (32,367 in 2016-17). The public expects that items in the collection can be easily searched and browsed. To make this possible, employees and volunteers research and describe each item’s contents and context. This task is expensive and time-consuming. The total cost of adding new items to the collection is over $10 million per year ($11,501,736 in 2016-17). Recent advances in technology have opened the possibility of automating these description tasks. We have developed Thickshake, a library catalogue interface that automatically generates collection metadata. Thickshake provides three primary contributions: a metadata parsing framework (incl. geolocation and date extraction), an image processing framework (incl. face detection and text extraction), and a machine learning framework (incl. subject recognition). The image processing and machine learning frameworks have been designed for high-performance computing environments, utilising GPU acceleration and parallel datastores to process thousands of collection items quickly. For more information, visit: <https://github.com/markshelton/thickshake>

**Image:**

