Everything I learned about cataloguing I learned from watching James Bond

Paul Hagon
Senior Web Designer
National Library of Australia
phagon@nla.gov.au

Abstract:

Is it enough to rely on human interpretation for cataloguing our collections, or like James Bond, will we require high tech gadgets to get the job done? Image searches now analyse the pixel-by-pixel colour values of an image, allowing searching by colour. Consumer-grade photo management software incorporates facial recognition, allowing us to identify individuals. When researching, will we be querying the content to get the metadata or will we still be relying on querying the metadata to deliver the content? Will these technologies filter down into the way we catalogue items within our collections, or will this be a layer that enhances our traditional cataloguing methodologies or provides an additional service for our users?

Introduction

The search capabilities of our libraries' collections are currently only as good as the metadata that is used to catalogue them. Any errors that are made when cataloguing an item have a flow on effect that makes the object more difficult for a user to find at a later date. Errors could be through a simple misspelling, or incorrect categorisation. The ranking algorithms that our catalogues use to display search results rely on accurate information to be in the system.

For many of our collections, the metadata that needs to be recorded is obvious and is part of the physical object. Books by their very nature have a basic, defined metadata structure that is printed on the cover and first pages of the item - Title, Author/s, Publisher, Year of publication, pages, etc.

Other collection items are not quite as clear cut as this. For example, a photograph has minimal amounts of obvious metadata attached to the object: the type of print it is, size or maybe a date stamp. Providing even the basic metadata elements to attach to this type of item requires the expert knowledge of a 3rd party, such as the curator, cataloguer or someone associated with the image, either the creator or subject of the image who can provide some context to it.

Over the past few years we have seen the introduction of more widespread metadata generation, particularly user generated metadata from tagging and annotation. We have also seen the introduction of automated tagging systems such as Open Calais that extracts meaningful terms from data to create additional pathways for discovery.

A search of the National Library of Australia's catalogue for The Sydney Morning Herald returns results letting us know that the newspaper is available. That is the limit of what we can do. If we visit the National Library of Australia's Australian Newspaper service, we can search through the contents of the Sydney Morning Herald. The actual content provides much richer results, while the issue, page number and the other associated metadata fields have been relegated to secondary information sources. Suddenly we are starting to directly query the content of the item, rather than the metadata surrounding the item.

The process enabling this is known as full text search. It is a process typically used for printed works, where physical item is scanned and converted to a digital object. This digital object is run through a process such as optical character recognition. Optical character recognition analyses a digital image, compares the shapes of groups of like-coloured pixels with a database of what a typical letter of the alphabet should look like and determines what character of the alphabet the pixels are likely to be representing. Depending upon the quality of the print, the accuracy of the recognition can be variable. The recognition process can also be augmented by comparing the converted text against a known dictionary to further reduce errors.

Full text search for a printed collection item is an easy concept to understand. What about the concept of a full text search for collection items that are from other sources, such as audio or image collections? Is it possible to search the contents of these media types in a similar manner?

Audio

An audio waveform can be analysed for unique patterns that identify one sound from another.

One use of this technology is in the field of ornithology. Neil Boucher is using computerised audio waveform analysis to try to locate the rare Coxen's Fig Parrot in South East Queensland. He is using a computerised recording system to extract bird noises from recordings made in the field. By comparing the waveforms of the bird calls in these recordings to the audio waveforms for known bird calls, he is able to determine which species are living in the area.

Another example of a 'full text' audio search is a service like Shazam. In 2002, Shazam launched a service where users could identify music that was playing, by simply using their mobile phone to pass a snippet of the song's audio waveform to the Shazam service, where the waveform is compared to a database of known audio waveforms, and the metadata associated with that particular waveform, the artist and song title, is returned to the user.

Images

Digitised images are made up of a finite grid of pixels. By analysing the colours and patterns made by individual pixels and groups of pixels, patterns in the image can be recognised and information can be extracted.

One of the simplest methods of image analysis is determining the colours that make up an image. The colour of each pixel in an image is made up of a combination of red, green and blue values (for a typical 3-channel RGB image). An 8-bit image will have 256 values (2^8=256) for each of the red, green and blue channels. Pixel values of common colours are:

Colour	Red value	Green value	Blue value
Black	0	0	0
White	255	255	255
Grey	128	128	128
Red	255	0	0
Green	0	255	0
Blue	0	0	255

By analysing an image at a pixel-by-pixel level and extracting the colour values of the individual pixels, it is possible to determine the colour range of the image and software can mathematically determine the dominant colour of an image. Once the colour is determined, a search can be carried out across a series of images, looking for a particular colour. A search for blue would likely return images containing a lot of sky or water based upon the values for a range of blue pixels (blue pixels typically being in the range 70,70,180 to 0,0,255).

An example of using this method of searching is the Multicolr Search Lab from Idée Labs. This application uses the Flickr API to search over 10 million Creative Commons licensed photographs that people have uploaded to their Flickr accounts. The search is based upon colours selected by the user. A user can select up to 10 colours to match against from a palette of 120 colours.

Having the ability to search by colour was made even more mainstream in April 2009 when Google added a colour filter to their Google Images search application, enabling a user to limit the colour of images returned for a search term to one of 12 colours. A search for ‘roses’ would return all roses of all colours, but selecting a red colour would limit the search to display only red roses.

This mathematical analysis can be extended beyond just the colours in an image, to the shapes that make up an image. By analysing the locations where pixels with similar colour values occur, the shapes contained within an image can be determined. A common scenario for this is the process used in optical character recognition that has previously been mentioned. Optical character recognition is a relatively simply process, as there are clearly defined foreground (text) and background objects. A photograph is a much more complex environment, as it contains many more patterns, and determining which are foreground and background components of the image can be more difficult. There are some common subjects in photographs, which can start to have consistent patterns, such as faces.

Being able to analyse a face in a photograph is more complex than analysing a comparatively closed data set, such as the two-dimensional letters of the alphabet. Facial features are three dimensional, although they are rendered as two-dimensional in a photograph. The same face at the same angle can appear to be vastly different, depending upon the direction of a light source, as different shadows alter the light and dark areas of a face. Faces themselves are not uniform, they change between individuals, race, even down to the expression on the face (whether they are smiling or frowning, for instance). Overall, however, faces tend to have consistent patterns: dark shadows where the eyes are, highlights where the cheek bones are, a vertical highlight and shadows at the nose and horizontal highlights and shadows around the mouth. By comparing the brightness values of pixels (and groups of pixels) and their locations relative to other areas of brightness or darkness, a computer can attempt to detect if an image contains a face.

In August 2006, Google acquired image technology company Neven Vision, which specialised in facial and object recognition. In June 2007, Google incorporated this technology into their Google image search and added a content type filter to their search interface. This allows a search to be restricted to various types of images: faces, clip art, and drawings. The search restriction is not relying on any entered metadata for determining the type of image; it is being determined by carrying out an analysis of the pixels that make up the image.

This is not just the realm of large corporations that are being exposed to these technologies. Open source projects, such as OpenCV, are being worked on by individuals interested in computer vision applications.

At a consumer level, users are starting to use facial recognition services on their own digital photo collections, through web based applications like Picasa (owned by Google) and standalone applications like iPhoto. These applications allow a user to go through their photo library and assign a name to a face in a photo. The application then uses facial recognition software to analyse the remainder of the user's photo collection and offers suggestions on photos that may contain the same person: an engaging method of cataloguing.

The experiments

To test if the facial recognition algorithms used in these programs could be adapted to cataloguing historic photographic collections, I carried out two experiments. For the first experiment, I used the facial recognition features of iPhoto 8.1 to analyse a series of photographs from the National Library of Australia's photographic and manuscripts collections. For the second experiment, I used the facial detection algorithms in Open CV to detect if the photos contained a face.

A small series of 184 photographs of 13 Prime Ministers of Australia were selected. The selection was made to cover as much variety of photo options as possible. This included close up portraits, group shots and photographs taken from various angles. I also attempted to include photographs covering different ages of the Prime Ministers, from their early career, through their time as Prime Minister and into their retirement.

The photographs used were of a medium resolution, 1500 pixels on their longest edge. Five photographs were not available at this higher resolution, so lower-resolution images of 600 pixels on their longest edge were used instead. The quality of the images was variable, depending upon the source image. Some images had very poor contrast while others were not sharp. Sixteen of the 184 photographs were in colour; the remainder were black and white (or sepia-toned).

For the first experiment using iPhoto, the photographs were imported into the application and a name was assigned to one photograph for each Prime Minister. The image that was initially chosen for each Prime Minister was the best quality portrait from the selected images. iPhoto was then used to analyse the remainder of the photo library and offered suggestions for the faces it recognised in the remainder of the photographs.

For the second experiment, using Open CV, the same photographs were imported into the application and each image was given a rating depending upon how successful the facial detection was.

Results

The performance of the facial recognition software within iPhoto could, at best, be described as fair, as results were extremely variable. In some instances, manually identifying a Prime Minister in their one key photograph resulted in 35% of the remaining photographs of that individual being correctly identified. In other cases, no other photographs of that individual were being correctly identified. After this first manual pass, only 24% of the photographs had been correctly identified (the 13 ‘key photographs’ and 32 that were recognised by the software).

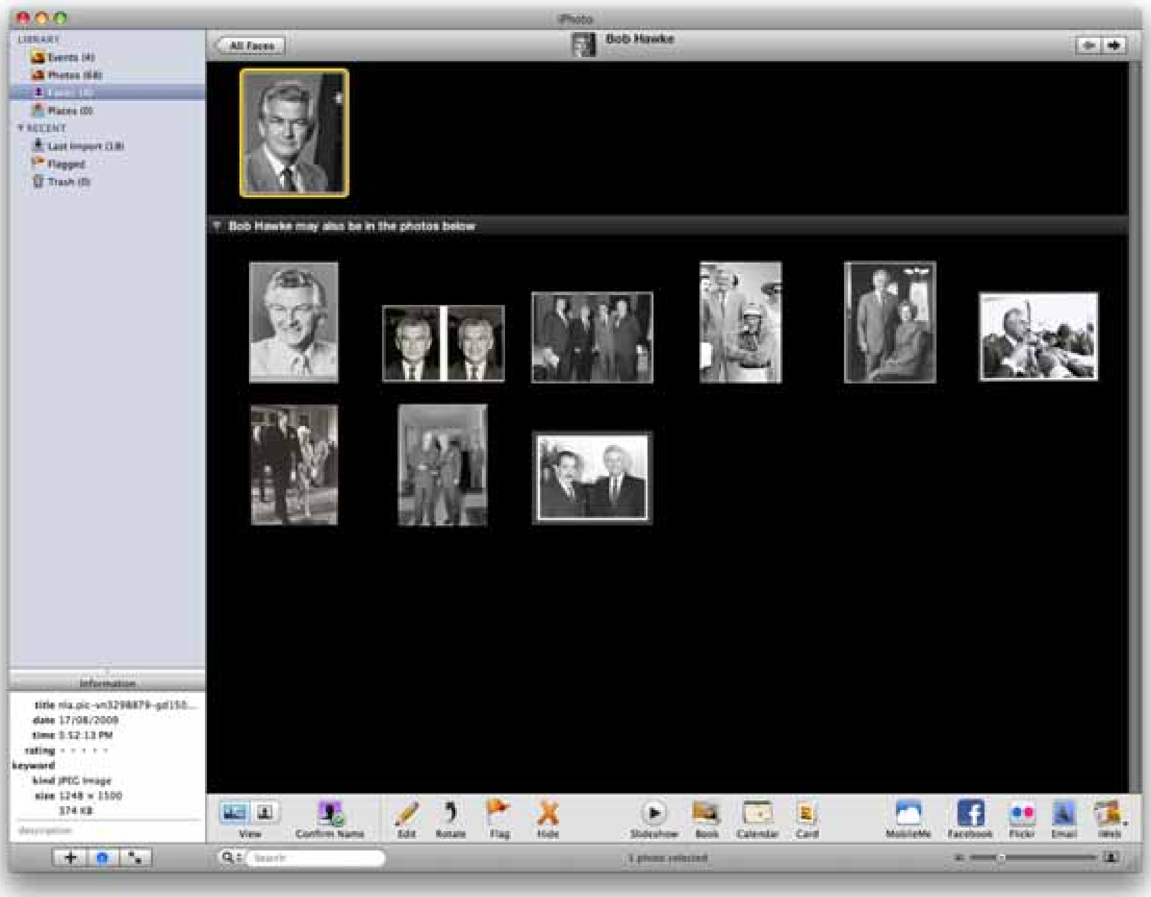

Figure 1: iPhoto facial recognition

A second correction pass was applied, where the choices from the first pass were confirmed and the software then had more information to base its recognition upon. This again proved highly unreliable. The best result for an individual was another 31% of photographs of one Prime Minister being identified correctly. Six of the 13 Prime ministers had no further photographs identified. After two passes, only 32% of the Prime Ministers in the photographs had been correctly identified.

Table 1: iPhoto 8.1 Facial recognition results

Prime Minister	**Total number of photos**	**Number correctly identified 1st pass**	**Number correctly identified 2nd Pass**
Robert Menzies	17	1	0
Sir John Gorton	8	1	0
Gough Whitlam	24	5	6
Bob Hawke	18	6	2
Edmund Barton	17	4	1
Paul Keating	14	1	1
John Howard	22	3	1
Malcolm Fraser	9	1	0
John Curtin	17	3	2
Harold Holt	15	2	0
Ben Chifley	6	2	1
James Scullin	10	3	0
William McMahon	7	0	0
Total	**184**	**32**	**14**

The identification was highly dependent upon the type of photograph and the position of the individual's face within the photograph. Close up portraits where the individual was looking directly at the camera performed the best. As soon as the face of the individual was tilted to the side, up or down, the software struggled to identify the person correctly.

Although iPhoto failed to identify an individual correctly most of the time, it excelled in identifying that the images contained faces. More than 90% of faces that were looking face-on to an image were correctly identified as being faces. Faces that were not recognised tended to be when the lighting on the individual's face was poor and they were in shadow (in one case caused from wearing a hat). Faces that were in profile tended not to be recognised by the software, and were ignored.

The resolution of the images did not appear to be a factor in how well the software performed, as the software correctly identified that faces existed within the five smaller (600-pixel wide) images and correctly identified the individual in three of the cases.

Due to the small numbers of colour images and overall inaccuracy of the software, this test was unable to determine if detection rates were better in black and white or colour images.

After iPhoto's success in detecting if a face was in an image, a second experiment was carried out using a different software package to duplicate the first test. A program was written incorporating the open source package Open CV. This application made no attempt at facial recognition, it only determined if a face was present in an image.

The same set of images used in the first test were loaded into the application and classified into three categories: successful detection, successful detection with errors and failure. When Open CV successfully detected a face, it drew a square around the face in the image (see Figure 2).

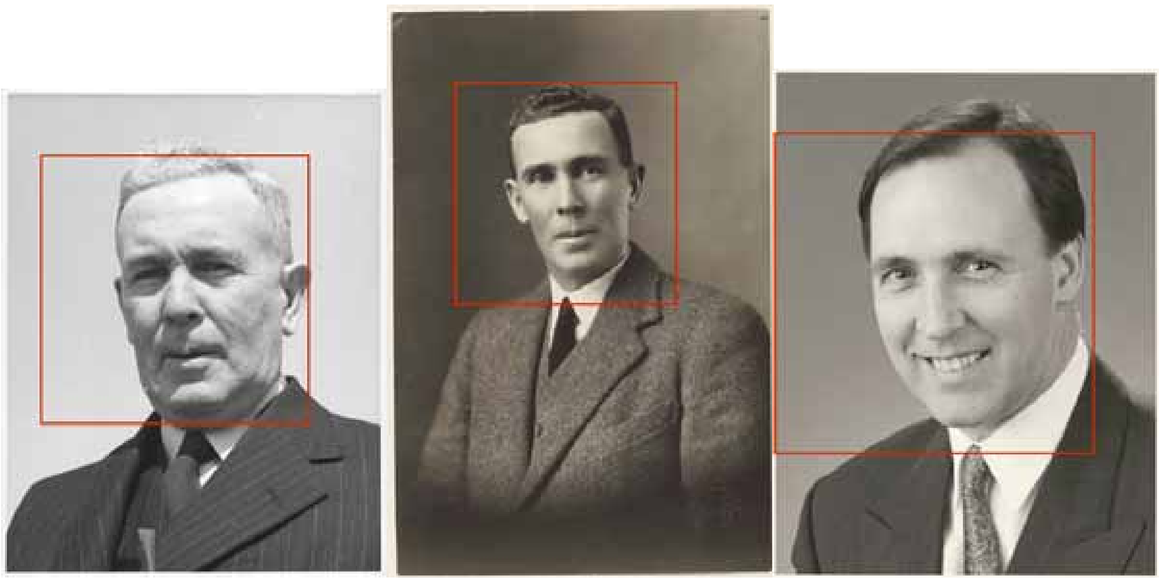

Figure 2: Examples of successful facial detection

The category of detection with errors was included as sometimes the software correctly detected that an image contained a face, but would incorrectly show the location of the face and positioned the square elsewhere in the image (see Figure 3)

Figure 3: Example of detection with errors.

The results of the second experiment using Open CV showed promise with 85% of images correctly detected as having faces within the images:

Table 2: Open CV Facial detection results

Result	Number of images detected	Percentage
Successful detection	121	66%
Successful detection, but with errors	34	19%
Failure	28	15%

Results using Open CV were better than those obtained using iPhoto. Open CV was able to correctly detect faces from a wider variety of angles, from front on through to a side profile. It was even able to detect faces where there were obstructions (see Figure 4), but it still struggled to detect faces where a face was partially hidden by a hat or a beard. Like iPhoto, Open CV had problems detecting faces where the images were of poor quality and contained little contrast (see Figure 5).

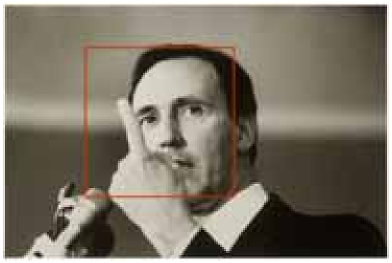

Figure 4: Correct detection of partially obscured face

Figure 5: Examples of images where faces were unable to be detected

Open CV has an advantage over iPhoto for facial detection in that it can be used in custom built applications to automate the detection process. The detection process using iPhoto is a manual process that is not suitable for large numbers of images.

Conclusion

It was an interesting experiment to carry out. iPhoto performed poorly at the task of identifying Prime Ministers from photographs based purely on facial recognition. After two passes with manual confirmation, only 59 of the 184 photographs (32%) had been correctly identified. Given that the tests were attempting to detect only 13 individuals, it proved disappointing to have such a small success rate on such a limited sample size.

A more promising result was the ability of iPhoto and especially Open CV to recognise that an image contained a face. Further investigations using Open CV are warranted. Further tests will need to be carried out over a wider range of images, such as landscapes and objects, to judge whether this can reliably be used to facial detection purposes. If further studies prove to be reliable, it may be feasible to automatically categorise items from the National Library of Australia's photographic and manuscript collection as having faces and create a search limit within our catalogue.

As at September 2009, the facial recognition technology contained within consumer grade applications is not reliably accurate for any large-scale cataloguing purposes for historical photographic collections. However, the software required for face detection is accurate enough to warrant further testing across a wider range of images.

References

Apple Inc, Cupertino <http://www.apple.com/ilife/iphoto/#organize>

Boucher, NJ 2002, 'The SoundID Autonomous Sound Recorder', *SoundID*, viewed 21 August 2009, <http://soundid.net/SoundID/Papers/The%20SoundID%20Autonomous%20Sound%20Recorder.pdf>.

Chan, S 2008, Powerhouse Museum, Sydney, OPAC2.0 – OpenCalais meets our museum collection / auto-tagging and semantic parsing of collection data, *Fresh + new(er)*, viewed 2nd April 2008 <http://www.powerhousemuseum.com/dmsblog/index.php/2008/03/31/opac20-opencalais-meets-our-museum-collection-auto-tagging-and-semantic-parsing-of-collection-data/>

Graham, A 2006, Google, Mountain View, A better way to organize photos?, *Official Google Blog,* viewed 10 April 2009 <http://googleblog.blogspot.com/2006/08/better-way-to-organize-photos.html>

Google, Mountain View, <http://images.google.com/advanced_image_search>

Google, Mountain View, <http://picasa.google.com/features-nametags.html>

Idée inc, Toronto, <http://labs.ideeinc.com/multicolr/>

National Library of Australia, Canberra, <http://catalogue.nla.gov.au/Record/2300554?lookfor=the%20sydney%20morning%20herald>

National Library of Australia, Canberra, <http://newspapers.nla.gov.au>

Open CV, viewed 6 July 2009 <http://opencv.willowgarage.com/wiki/>

Shazam Entertainment Ltd, London, <http://www.shazam.com/music/web/home.html>

Tanguay, D 2009, Google, Mountain View, Search the rainbow, *Official Google Blog*, viewed 10 April 2009 <http://googleblog.blogspot.com/2009/04/search-rainbow.html>