

## the real face of white australia

**Author :** Tim Sherratt

**Tagged as :** [facial detection](#), [invisibleaustralians](#), [python](#)

**Date :** September 21, 2011

In many of the presentations I've given in recent times I've managed to include a question raised by Tim Hitchcock in his chapter in *The Virtual Representation of the Past*. Tim asks:

What changes when we examine the world through the collected fragments of knowledge that we can recover about a single person, reorganised as a biographical narrative, rather than as part of an archival system?

The idea of turning archival systems on their head to expose the people rather than the bureaucracy is what motivates Kate Bagnall and I in our attempts to make the [Invisible Australians](#) project into a reality.

*Invisible Australians* aims to liberate the lives of those who suffered under the restrictions of the White Australia Policy from the rich archival holdings of the National Archives of Australia and elsewhere.

We always knew that the portrait photographs, included on a range of government documents, would provide a compelling perspective on these lives, but we weren't quite sure how we were going to extract them. Up until last weekend, I'd assumed that we'd develop a crowdsourcing tool that contributors would use to mark-up the photos.

Now I'm not so sure.

In the space of a couple of days I've extracted over 7,000 photographs and built an application to browse them — here is [the real face of White Australia](#)...



How did I do it? Paul Hagon, at the National Library of Australia, [gave a presentation](#) last year in which he explored the possibilities of facial detection in developing access to photographic collections. The idea lodged in my brain somewhere and a few days ago I started to poke around looking to see how practical it might be for *Invisible Australians*.

It didn't take long to find [a python script](#) that used the [OpenCV library](#) to detect faces in photographs. I tried the script on a few of the NAA documents and was impressed — there were a

## discontents

working for the triumph of content over form, ideas over control, people over systems  
<http://discontents.com.au>

---

few false positives, but the faces were being found!

So then the excitement kicked in. I modified the script so that instead of just finding the coordinates of faces it would enlarge the selected area by 50px on each side and then crop the image. This did a great job of extracting the portraits. I tweaked a few of the settings as well to try and reduce the number of false positives. Eventually, I developed a two-pass system that repeated the detection process after the image had been cropped and it's contrast adjusted. This seemed to weed out a few more errors. You can [find the code](#) on GitHub.

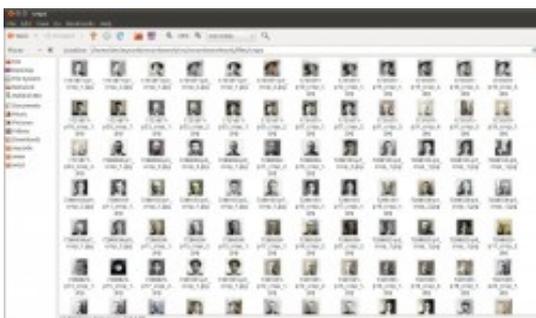
Once the script was working I had to assemble the documents. I already had a basic harvester that would retrieve both the file metadata and digitised images for any series in the NAA database. Acting on Kate's advice, I pointed it at series [ST84/1](#) and downloaded 12,502 page images.

All I then had to do was loop the facial detection script over the images. Simple! The only problem was that my 3-year-old laptop wasn't quite up to the task. As it's CPU temperature rose and rose, I was forced to employ a special high-tech cooling system.



Keeping my laptop alive...

But after running for several hours, my faithful old laptop finally worked it's way through all the documents. The result was a directory full of 11,170 cropped images.



The results

There were still quite a lot of false positives and so I simply worked my way through the files, manually deleting the errors. I ended up with 7,247 photos of people. That's a strike rate of nearly 65% which seems pretty good. The classifier, which does the actual facial detection, was probably trained on conventional photographs rather than on the mixed-format documents I was feeding it.

Then it was just a matter of building a web app to display the portraits. I used Django for the

backend work of managing the metadata and delivering the content, while the interface was built using a combination of [Isotope](#), [Infinite Scroll](#) and [FancyBox](#).

It's important to note that the portraits provide a way of exploring the records themselves. If you click on a face you see a copy of the document from which the photo was extracted. A link is provided to examine the full context of the image in RecordSearch. This is not just an exhibition, it's a finding aid.

What next? There are many more of these documents to be harvested and processed (and many more still yet to be digitised). I will be adding more series as I can (though I might have to wait until I can afford a new computer!). I'd also like to explore the possibilities of facial or object detection a bit more. Could I train my own classifier? Could I detect handprints, or even classify the type of form?

In the meantime, I think our experimental browser helps us to understand why the *Invisible Australians* project is so important — you look at their faces and you simply want to know more. Who are they? What were their lives like?

UPDATE: For more on the photos and the issues they raise, see [Kate Bagnall's posts](#) over at the [Tiger's Mouth](#).

### Share this:

- [Click to email this to a friend \(Opens in new window\)](#)
- [Click to print \(Opens in new window\)](#)
- [Click to share on Twitter \(Opens in new window\)](#)
- [Share on Facebook \(Opens in new window\)](#)
- [Click to share on Google+ \(Opens in new window\)](#)
- 

This work is licensed under a [Creative Commons Attribution 4.0 International License](#).