COSC451 2009 Assignment 1 (10%)

There are two options for this assignment. The first involves programming; the second is an essay.

Option 1: Implementation of a simple saliency map

In this option, your task is to implement a simple saliency map, which takes an image and returns a modified image file, indicating the 'most interesting regions' within the input image. You should submit two things: the code which implements the saliency map, and a report.

Code

Your are free to implement the saliency map as you want. However, as a starting point, it is useful to know about how to implement simple 'feature detectors' which pick out 'blobs' or 'edges' in an input image. A simple implementation is to use **convolution kernels** to filter the input image. A convolution kernel is a square matrix of numbers with an odd number of rows and columns. Here's a simple one:

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

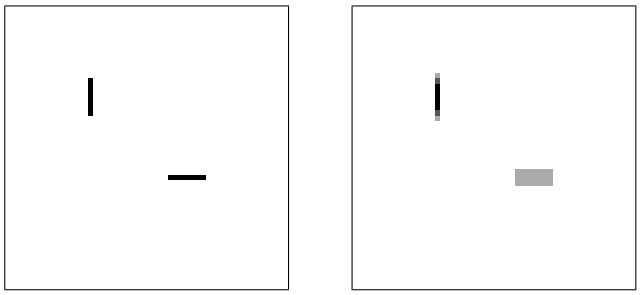
Applying the kernel to a pixel in the image returns a value for that pixel which is based on a sum of the pixel and its surrounding pixels, weighted according to the numbers in the matrix. Say we apply the kernel to a pixel P_5 whose neighbours are defined as follows:

P_1	P_2	P_3
P_4	P_5	P_6
P_7	P_8	P_9

The value returned will be a weighted sum of the values of pixels around P_5:

$$0\times P_1 + 1\times P_2 + 0\times P_3 + 0\times P_4 + 1\times P_5 + 0\times P_6 + 0\times P_7 + 1\times P_8 + 0\times P_9 \div W$$

(where W is the sum of the numbers in the kernel). When an image is **convolved** with a kernel, the kernel is applied to each pixel, generating a new image as output. Different convolution kernels result in different features of the input image being highlighted. For instance, the above kernel emphasises vertical lines (with a certain spatial frequency). If we convolve the image on the left with this kernel, we get the image on the right:



(Notice that the horizontal line is relatively greyed out compared to the vertical one. A thresholding operation could get rid of it altogether.)

To implement convolution, you can use whatever software you want. One useful resource on Linux is **imagemagick**, which provides a set of command-line image processing tools. For instance, you can run the convolution operation just described using the following command:

```
convert input-file.png -convolve "0,1,0,0,1,0,0,1,0" output-file.png
```

For more details, see `http://www.imagemagick.org/`.

Another option is MATLAB.[1] There are some existing MATLAB programs for computing saliency maps, in particular Dirk Walther's 'saliency toolbox' which you can consult, though I want you to write your own code. The basic convolution operations are available in MATLAB too; there's a link to an introduction to image processing in MATLAB on the 451 webpage. (MATLAB also allows you to implement more complex oriented filters, in particular **Gabor filters**, which you might want to try.)

The way to implement a saliency map is to run lots of different oriented filters detecting different features at different spatial frequencies, and then compute a weighted sum of the results. I have put a link to a paper by Itti *et al.* with some technical details of their own model on the 451 webpage. You don't need to reimplement their whole algorithm, but it might give you some ideas about what to do. The aim is to build a filter that picks out the kinds of features of natural images that a human would naturally pick out.

Generating a saliency map which reproduces human intuitions is a somewhat open-ended problem. As a baseline, you should aim for your saliency map code to find the three pictures of Obama in the noisy background shown in Figure 1. The image can be downloaded from the 'resources' page of the course website.

Report

You should also submit a report about your code. The report should contain four sections:

- How the code works;
- How to run the code (so I can test it);

[1] You can run MATLAB from the department Linux machines. If you want it on Mac or Windows, it must be installed individually; just ask.

Figure 1: A test image

- Some examples of its results (including its performance on the Obama test image and on some other 'natural images' which you choose);
- A brief evaluation of how well it does, and what might be done to improve it and make it more human-like.

Submission and marking

You should submit the code and the assignment by email to me (`alik@cs.otago.ac.nz`) by 5pm on **Friday of Week 7** (i.e. on **Friday April 23rd**). 10% of available marks will be deducted for each day late.

I'll give equal weight to the code and the report. (But the 'report mark' is likely to reflect the quality of the code, so in that sense the implementation is the main focus.)

Option 2: Essay

The second option for the assignment is an essay, with the following title:

> Summarise and assess the evidence for the proposal that there are separate neural pathways for determining the location of a visual stimulus and determining its category. Your answer should include a discussion of processing in early visual cortices (V1–V4), the inferotemporal cortex (IT), parietal cortex and frontal eye fields.

To answer this question, you should use the following reading list.[2]

- Logothetis *et al.* (1995): a study of the object categorisation pathway in inferotemporal cortex.
- Colby and Goldberg (1999): a survey of representations of location in parietal cortex.
- Thompson and Bichot (2005): an overview of evidence that the frontal eye fields (FEF) represent salient locations.
- Moore and Armstrong (2003): evidence for modulation of V4 cells by FEF.
- DiCarlo and Maunsell (2003): some evidence that IT cells do retain some information about location.

You're not expected to read everything on the reading list. The best thing to do is to download all the papers, then skim them all and see which ones look most interesting. You're free to make reference to other papers too, e.g. ones which you find on Web of Science (looking at papers which cite the above papers could be useful), or papers cited in the course textbook.

[2]You can get these papers online from Web of Science. Go to the Otago Library webpage, then follow links to 'Article Databases' → 'W' → 'Web of Science via Web of Knowledge'. This is a fantastic resource—if you don't already know about it, you should try it!

Submission and marking

You should submit the essay by email to me (`alik@cs.otago.ac.nz`) by 5pm on **Friday of Week 7** (i.e. on **Friday April 23rd**). 10% of available marks will be deducted for each day late.

You will be marked on the clarity and organisation of the essay, and on evidence that you have gone into some detail in the assigned readings.

References

Colby, C. and Goldberg, M. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, **22**, 399–349.

DiCarlo, J. and Maunsell, J. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *Journal of Neurophysiology*, **89**, 3264–3278.

Logothetis, N., Pauls, J., and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, **5**(5), 552–563.

Moore, T. and Armstrong, K. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature*, **421**(6921), 370–373.

Thompson, K. and Bichot, N. (2005). A visual salience map in the primate frontal eye field. *Progress in Brain Research*, **147**, 251–262.