John bums sheep - an initial investigation

Fintan S. Nagle

July 4, 2013

Fire is a complex and dynamic phenomenon. Characterised by rapidly shifting patterns of both first- and second-order motion, it evokes rich visual percepts as well as aesthetic responses. We report an investigation of the visual features useful for discrimination between similar fires, the neurocognitive architecture responsible for comparing them, and the specificity of the neural representations involved.

From an evolutionary point of view, mastery of fire was key to human development. Being able to control fire allowed early humans to cook food, defend themselves from predators and survive in cold, challenging environments. Fire was the first of a long line of technologies which release stored energy from fuel and turn it to human purposes; the earliest archaeological evidence of fire use dates back 1.8 million years (REF), with frequent use found from (FILL) million years onwards (REF). Even before this, hominids regularly encountered flame in the form of bushfires, although these were perceived as a threat, not a controllable, exploitable entity.

The evolving human visual system has therefore been exposed to a large amount of flamelike stimuli in the last 1.8 million years. These stimuli have often appeared in dangerous or life-threatening contexts, either posing a threat or aiding survival. In sufficiently extreme situations, such as extreme cold or heavy predation, those early humans who could successfully control fire had an increased chance of survival.

It is therefore natural to enquire whether the human visual system has become adapted in any way to the perception of flamelike stimuli. Does the visual system employ any specific representations or specialised models when attending to fire, does it use the same general-purpose systems employed when observing a novel moving stimulus?

This question recalls the ongoing debate concerning the specialisation of face perception. We find increased activation of the fusiform face area and inferior temporal sulcus while viewing faces; this can be explained either by innate specialisation or learned proficiency. In the same way, observation of fire may recruit neurons and systems which respond preferentially to, and perform better on, flame stimuli. On the other hand, observing fire may stimulate the same neural populations as observing other moving stimuli.

This report aims to answer two questions. Firstly, what types of visual information are important when observing flames; which details are represented, and which thrown away? This choice defines the upstream visual subsystems (for example, that of motion detection, edge detection or shape detection) which are activated by fire stimuli. Secondly, are any of these systems specific to fire? We do not address the question of whether any specificity is learned or innate.

We begin with a survey of the ecological context in which fire is observed and processed. We then describe several yes/no and 2AFC tasks designed to activate the visual subsystems which process fire and measure task performance under removal of different types of visual information from the stimulus.

1 Fire in the natural environment

Here we swiftly review the basic characteristics of fire and the contexts in which early humans interacted with it. It is essential to consider the motivations and tasks which lead the visual system to process and represent fire.

We can define the set of dynamic visual stimuli termed "fire" as the light fields emitted from combusting objects. In terms of luminance, they display rapid variation from gentle glows to dazzling flares, which covers virtually the whole photopic range of the visual system; one can imagine that scotopic vision is not employed except for very dim, distant fires. In terms of size, they may subtend any angle from the whole visual field (for a large, close fire) to a point (for a small, distant fire).

In terms of spatial frequency, fire is characterised by

FFT stuff here

It is important to consider the context in which flames were observed by early humans. We can imagine two main situations: a) avoiding large natural wildfires, and b) controlling small artificial hearth fires for cooking and defence. Good performance at both tasks is essential for survival, but we are not as interested in visual development supporting task performance in situation a), as it is not specific to humans; all large land animals in an environment at risk of wildfires must possess this skill. We consider only the anthrospecific task of controlling small artificial fires.

Fire is employed in many situations, from defensive to culinary to ritualistic. In order to avoid being overly selective, we do not consider a particular task, but the general context in which they can all be placed: the extraction of useful information (or affordance) from visual stimuli. In order to properly control a fire, the observer must estimate properties such as the temperature of various parts, whether the fire is growing or shrinking, whether it is likely to spark or flare up, and in which direction it will spread. Knowing these details accurately is especially important in the early stages of firelighting, when the flame is small and at risk of extinguishing.

We therefore restrict this work to the study of small, artificial fires which do not pose an imminent threat to survival.

2 How is fire represented?

3 Are these representations specialised?

Its presence in the environment of the evolving

We are thinking on a level between the neural and the cognitive, a level insufficiently detailed to model individual neurons but more specific than descriptions of qualia or percepts. Our main primitive is the concept of the *neural*

representation, a description of how a particular idea, metric or concept is coded by neural activity or architecture. It is illustrative to provide some examples.

A place cell, which fires when an animal occupies a particular area of its environment, is a representation of that particular location. Simiarly, a head direction cell represents orientation.

Visual information is encoded in many different representations as it passes from retina to cortex.

A The UCL Vision Research Lab's expression space pipeline

This section describes how a single still face-image (and, by extension, a video formed of a sequence of images) can be represented as a series of coefficients which (together with a specially defined "face space") can reproduce a given face-image with a high degree of accuracy and a great deal of compression; a face-image can be specified by the coefficients alone (provided the face space is already given), information which is orders of magnitude smaller than the bitmapped version of the image.

Heavy use is made of the technique of *principal component analysis* or *PCA*[**?**][1], a technique which highlights the axes of largest variance in a set of multivariate point data. Consider an d-dimensional space containing n points. Each point is defined by d coordinates along the normal axes of the space. PCA imposes b *new axes* on the space (the choice of b is down to the operator, subject to $b < d$) and expresses each point by b coordinates along each of these new axes.

Each new axis is chosen so that it spans the maximum possible variance among the points, given the important constraint that the new axes (like the old ones) must be orthonormal. For example, imagine a data set which looks like a vaguely cylindrical point cloud in the original d-space. The axis of greatest variance will be along the longditudinal axis of the point cloud; this, therefore, will be the first axis of b-space. See figure **??** for an illustration. Mathematically, the transformation can be done by finding the eigenvectors and eigenvalues of the covariance matrix of the initial variables.

After PCA each point is described by a b-vector, along with the origins and directions of the b new axes (which are the same for every point and must therefore only be given once for the data set). If $b = d$, PCA merely rotates the data set and every point is still described fully; information is not lost. If $b < d$, some information about the precise location in d-space of each point is lost. The larger b, the smaller the information loss and the smaller the compression. PCA is therefore user-parametrisable in terms of b, allowing an adjustable tradeoff between compression and accuracy.

The first research on the application of PCA to faces was done by Sirovich *et al* in 1987[**?**], who evaluated the feasibility of the technique on a set of example faces. Taking each face (a 128×128 pixel greyscale picture) as a point in D-dimensional space, with $D = 128^2 = 2^{14}$, they performed PCA to extract a series of B basis vectors spanning the subspace in which the example faces were situated. They noted that each face can be expressed by adding together a series of "eigenfaces" (points in B-space, each one generated by extending a basis vector by a certain coefficient). To make their method work, Sirovich *et al* had to perform a substantial degree of cropping and normalisation before applying PCA; their test set was also composed exclusively of young Caucasian males.

A major problem with feeding a linearised image directly into the PCA procedure, then linearly combining eigenfaces into an output face, is that linear combinations produce inherent blur; combining two widely differing eigenfaces

[1] Also known as the discrete Karhunen-Loève transform, the Hotelling transform or proper orthogonal decomposition.

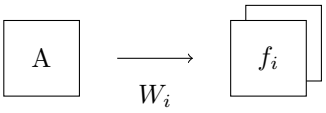
will not produce a realistic-looking face but a blurred combination of the two. This is a reflection of the fact that faces are not pure mathematical artefacts but physiological objects whose shape is constrained by their underlying musculoskeletal structure.

Work over the intervening years has extended this simplistic application of PCA into a more mature method capable of dealing with a wider variety of faces and expressions in full colour. The following procedure is typical of the processing pipeline used by the UCL Vision Research Lab.

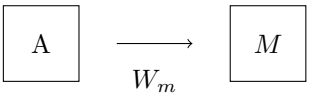
To take into account shape information as well as texture information, all frames are first conceptually shape-standardised using a warping operation[2] The input to the PCA process consists of *a)* RGB or greyscale pixel data specifying a base image, and *b* a vector flow field allowing this base image to be warped into a final output image. This separates texture data from shape data (even though this division is to some degree arbitrary, as discussed later).

The input to the procedure is a sequence of frames f_i and a reference frame A (selected so as to contain colour components allowing realistic warping to a large amount of expressions, such as an open mouth showing teeth and a dark area between the teeth. Warping can remove colours from an image by reducing their area to zero, but not add them).

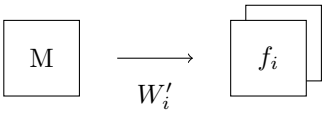
1. Using the multichannel gradient model (MCGM) algorithm[**?**], a neurally plausible motion-detection and warp-finding algorithm, generate warp fields W_i from each A to each f_i. Each warp field is simply a vector field showing how to deform A to obtain f_i.



2. Average these fields, generating a mean warp W_m.
3. Warp the reference image with the mean warp to generate the warp mean M.



4. Generate new warp fields W_i' which warp the warp mean to each frame f_i (not to the reference image). These warp fields are generated by taking the reverse of the mean warp, then doing the initial warps W_i. These new fields W_i' each symbolise the warp required to reshape the warp mean into the shape of the face in frame f_i.



[2] *Warping* simply spatially deforms an image; *morphing* combines spatial deformation with blurring between a start and finish image.

5. Apply the reverse warps of fields W_i' to each frame. This results in a set of frames f_i' which all have the *shape* of the warp mean, but different texture information.
6. For each frame, store f_i' and W_i'. Each original frame can be reconstructed by applying W_i' to f_i'.
7. Serialise this information (by concatenating the lists r_i, g_i, b_i, x_i, y_i containing red, green, blue and x, y-warp information for each pixel) for each frame. Each frame is now symbolised by a high-dimensional vector containing colour and warp information; these vectors exist in an f-dimensional *frame space*. Frame space, of course, contains all possible images, including those which do not represent faces.
8. Perform PCA on the resulting point cloud. The space spanned by the e principal components chosen is an e-dimensional *expression space*; each point therein represents an expression.
9. Reconstruct expressions by taking a point in e-space (either representing a real expression, or an artificial one), projecting it into frame space, warping its image data with its warp data, and displaying it.

This process is summarised in figure **??**.

B Source code

The following pages show Matlab source code for the two main functions of the report: building a second-order PCA space and reclaiming a video therefrom. Much functionality is included in external functions.

C Video files

The included CD contains video files showing various comparative reconstructions generated during the experiment.

- One clip reconstructed with different values of s.
- Side-by-side comparisons of original and reconstructed clips, s=7.
- Some random dynamic expressions, s=7.
- Comparison of the clips produced in section **??** showing the effect of increasing or decreasing each principal component.

D Jokes used to generate the smiles during filming

- What did the ham say to the doctor?
  I'm cured!
- Why did Bambi start smoking?
  Deer pressure!
- Why did Little Jack Horner sit in the corner?
  Because his bum was square.
- Why was the electron depressed?
  It felt a bit negative.
- A horse walks into a bar. "Why the long face?" the barman says. "My wife has left me and I'm an alcoholic," says the horse.
- A hole has been found in a naturist camp wall. The police are looking into it.
- What do you call a man with seagulls nesting in his head?
  Cliff!
- What do you call a man with rabbits burrowing into his head?
  Warren.
- Some trees were stolen from my local forest. The police are stumped.
- Later the same day, all the toilet seats were stolen from my local police station. The police have nothing to go on.
- On the news today... police were called to my local nursery school where a two-year-old was resisting a rest.
- When Prince William joined the army he disliked the use of the phrase "Fire at Will!"
- We have a saying where I come from. Time flies like an arrow. Fruit flies like a banana.
- What's a prisoner's favourite punctuation mark?
  The full stop. It marks the end of his sentence.
- Did you hear about the soldier who survived attacks of mustard gas and pepper spray? He was a seasoned veteran.
- When is a door not a door?
  When it's a jar.
- When is a car not a car?
  When it turns into a side street.
- Did you hear about the short fortune-teller who escaped from prison? He was a small medium at large.
- The Energizer bunny was arrested for being violent. He was charged with battery.
- Have you heard about the new practice of coffee-stealing? They call it "mugging."
- A fire ripped through the campsite. The heat was intense.
- My friend went on holiday to Egypt, but couldn't accept that he was really there. I think he was in denial.
- Why do tennis players rarely marry?
  Because love means nothing to them.