Fixation Maps: Quantifying Eye-movement Traces

David S. Wooding
Applied Vision Research Unit
Kingsway House West
University of Derby, Derby DE22 3HL
Tel: +44 1332 593131

d.wooding@derby.ac.uk

ABSTRACT

The analysis of eye-movement *traces* (i.e. the patterns of fixations in a search) is a powerful but often neglected area of eye-movement research. This is largely because it requires a more complex analysis than parameters such as mean fixation duration and as a result, previous attempts have focused on qualitative appraisal of the form of an eye-movement trace. In this paper, we introduce the concept of the "fixation map". We discuss its application to the quantification of similarity of traces, and the degree of "coverage" by fixations of a visual stimulus. The use of fixation maps in the understanding and communication of large numbers of eye-movement traces is also examined.

Keywords

Eye-movements, traces, analysis, fixation map, similarity.

1. INTRODUCTION

Over the winter of 2000/2001, the Applied Vision Research Unit at the University of Derby conducted the world's largest eye-movement experiment. An automated eye-tracker was left running in a room of the National Gallery, London, as part of the millennium exhibition: "Telling Time". In the three months of the exhibition, 5,638 subjects had their eye-movements successfully recorded while they viewed digitised images of paintings from the National Gallery collection. The quantity of data was unprecedented and posed many challenges both in terms of academic analysis and in communication of results to the public. Aside from the traditional measures such as mean fixation duration, number of fixations, etc., it became clear that novel methods of manipulation, analysis and representation of the large amount of data would be required.

Each participating subject had been presented with a series of questions that put them into different categories based on, for example, their sex, age and experience of art. *Where* individuals and groups of subjects had looked in an image would clearly be an important part of the analysis. It was felt that the concept of "similarity" between the eye-movement patterns of different individuals and groups should be explored, along with the degree to which their fixations had "covered" the image.

ETRA'02 New Orleans Louisiana USA

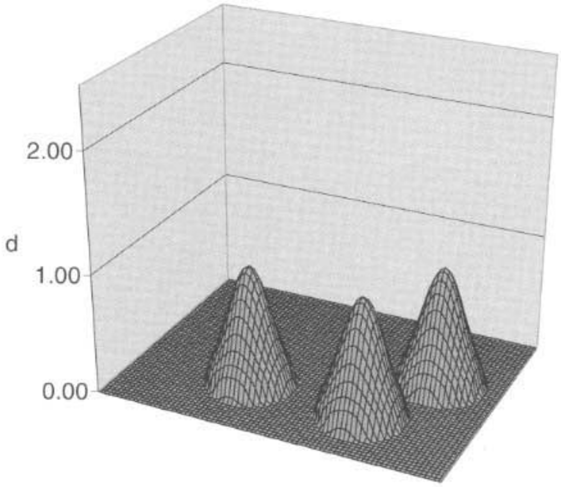


Figure 1: Three individual fixations on a fixation map

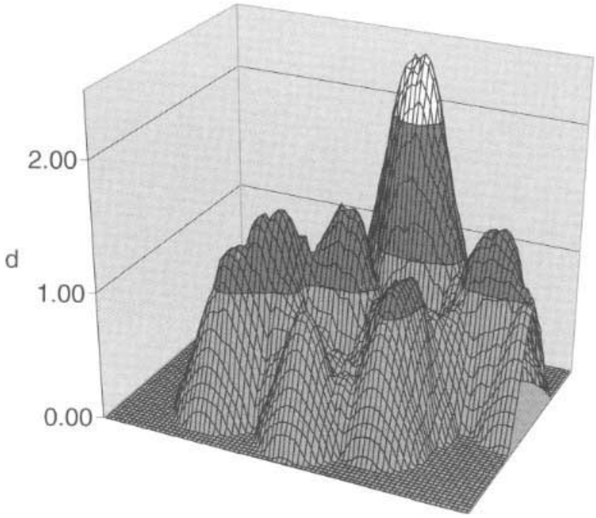
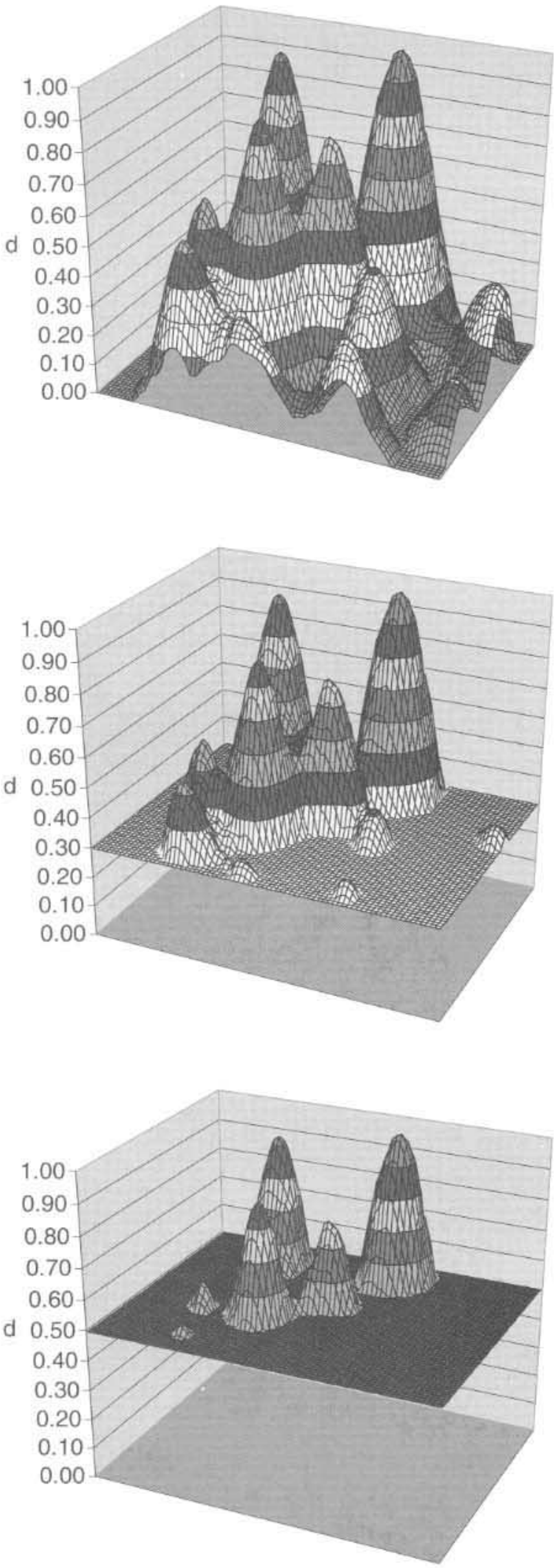


Figure 2: The appearance of the fixation map after 17 fixations, some having overlapped.



Figures 3-5: The original fixation map (top) is "flooded" to different values of d (0.3 at middle, 0.5 at bottom).

Lastly, it was clear that a means of visualisation and communication of these unwieldy data sets was necessary, to satisfy both public and academic interest in the work.

These issues are to a large extent addressed by the use of "fixation map" analysis, which represents a versatile analytical tool while retaining objectivity and minimising the number of assumptions made in the process. The methodology of the technique and its applications are described in the following sections.

2. FIXATION MAP ANALYSIS

At its simplest, the fixation map is a two-dimensional record of the locations of all the fixations being analysed, whether in an individual trace, or in a selection of traces. However, the maps developed for the analysis in this paper are three dimensional, with the third dimension being an indication of the discrimination/detection/perception (termed the property "d" in this paper) achievable from that fixation. The precise definition of "d" is deliberately left vague in this paper. The definition is not essential to the concept of the fixation map; as a parameter, "d" can be defined to suit the analysis required, but can be thought of as an extension of the concept of the soft-shelled "visual lobe" [Overington 1976]. As this property falls off with distance from the centre of the fixation, the three-dimensional form of an individual fixation is usually approximated to a 3 dimensional gaussian, or, less elegantly, a lump (figure 1).

As a result, the fixation map might be more descriptively termed a *landscape* or *terrain*, since the value at any point indicates the *height* or *amount* of property d at that point.

2.1 Creating a fixation map

The creation of a fixation map begins with a blank map of the stimulus presented. For digital images it is reasonable to use a blank map of the same dimensions in pixels as the original stimulus, with one value of d per pixel in the original stimulus. This arrangement will be assumed in the following description, with the individual locations in the fixation map being termed *map pixels*. The scale of the map is, though, essentially arbitrary

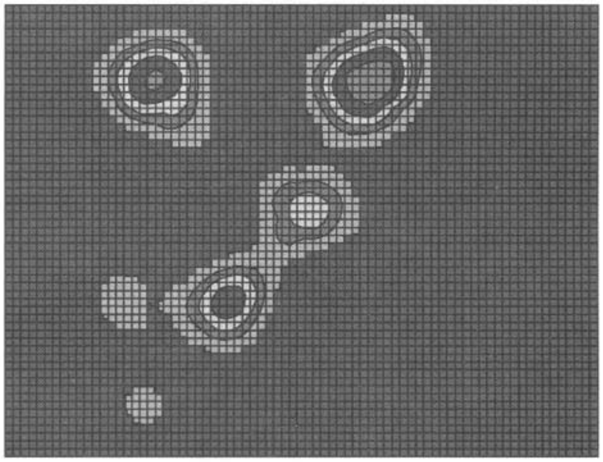


Figure 6: The contour map corresponding to figure 5.

and under the control of the experimenter.

For each fixation location, an identical 3d gaussian of unit height is dropped onto the map (figure 1). If this overlaps with an existing fixation, the height at any map pixel is added to the existing height at that point. With increasing numbers of fixations, a landscape is built up (figure 2), indicating the variation in d across the image.

With large numbers of fixations, the fixation may be approximated to a cylinder rather than the 3-d gaussian. As more cylinders overlap and superimpose, then the fixation map effectively "softens".

It should be noted that in analysing fixation data in this way the only assumptions that have been made have been:

1. To represent the property "d" associated with the fixation as a 3-dimensional gaussian
2. To determine a half-height width for the gaussian (i.e. over what area it is said to act)

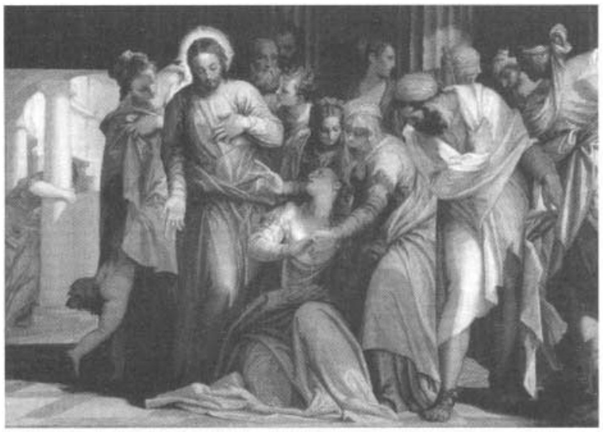
These assumptions are well-defined and repeatable, and can be quoted when reporting any results. A typical width might be the size of the fovea projected onto the stimulus. The *true* width depends on the area over which a fixation can be said to exist, and this continues to be the subject of some debate. All the fixation maps in this paper have been prepared on real data from the National Gallery exhibit, and for simplicity, a gaussian with 2 degree half height width has been used throughout. Details of the images are given below in section 4.

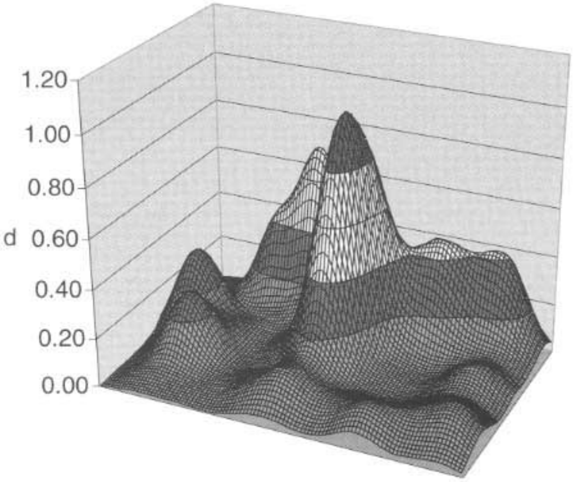
A slightly more elaborate, but useful, analysis can be achieved if the height of the 3d gaussian is in proportion to the duration of the fixation. The fixation map then becomes a "dwell map", representing not only the areas fixated, but the amount of time spent there. However, this paper will only describe the simpler fixation map, disregarding fixation duration.

2.2 Normalisation

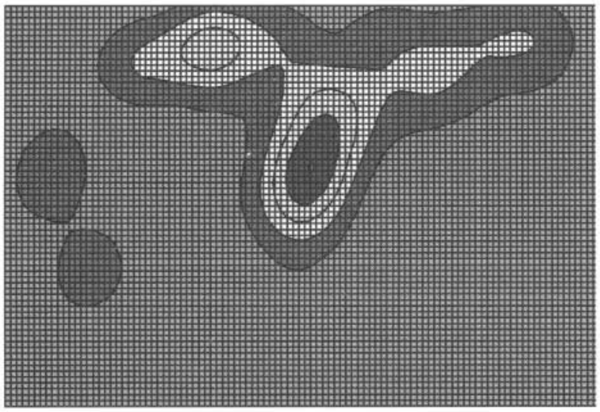
The map can represent the fixations of one trace by one individual or, just as easily, those of many traces by one or by many individuals. As described, the map is a representation of the cumulative fixations used in its construction.

If the map is to be compared with another map to determine differences in location of fixations, it is desireable to normalise





Figures 7, 8, 9 and 10 (reading left to right): the original image N0931 (see section 4); map of fixations of 131 traces; corresponding contour plot; and the image redrawn with areas receiving higher numbers of fixations appearing brighter.



the map so that the maximum value of d (the highest peak) is given a value of 1.

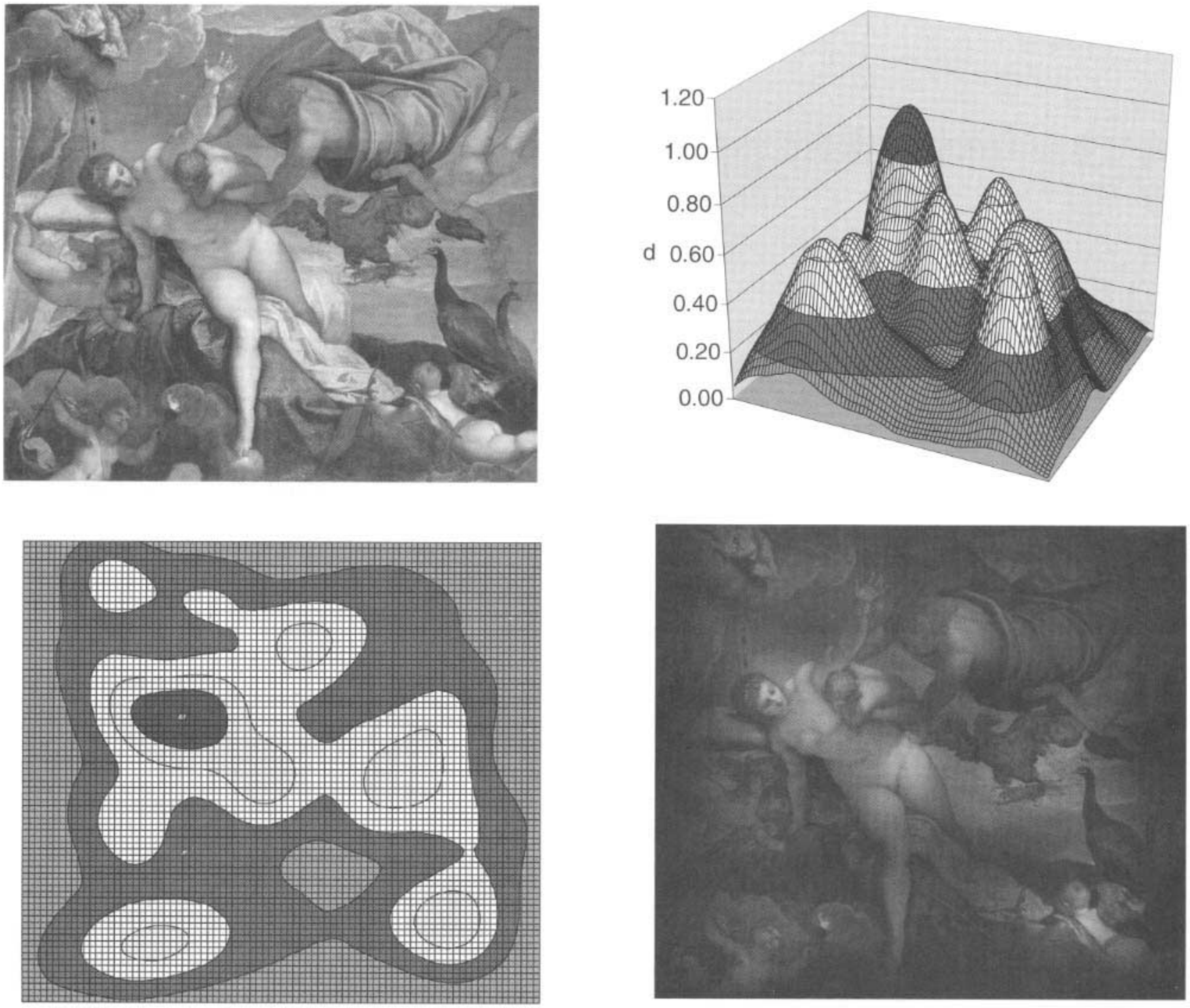
If the comparison is to determine whether one map contains more clusters of fixations and another a more even distribution of fixations, then the absolute totals of d may well be important and it would be desirable to leave the map un-normalised.

2.3 Areas of interest

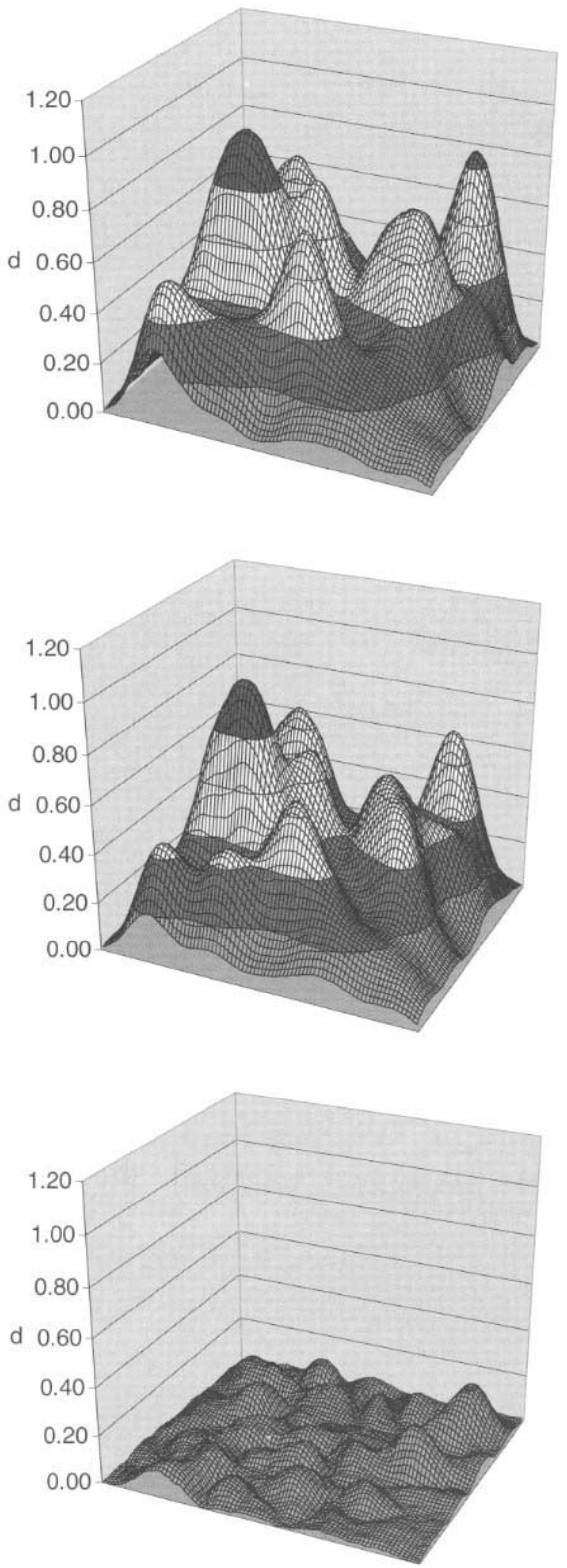
Fixation map analysis affords an opportunity objectively to define the principal "areas of interest" of observers when viewing an image. This can be on the basis of all areas in which the value of d is greater than a critical value, d_{crit}. To visualise this, if we return to the "landscape" metaphor, the landscape is flooded to the level of d_{crit} leaving only the highest peaks (the areas of interest) as islands. Alternatively, the "top five" areas (for example) can be determined by gradually increasing d_{crit} from zero until the required number of areas remain on the map. Figures 3, 4 and 5 illustrate the initial map (top figure) and the same map "flooded" with two values of d_{crit}. Figure 6 is a two dimensional contour map of the same data as the terrain in figure 5 to illustrate the areas of interest.

2.4 Coverage

A measure of the amount of the original stimulus covered by the fixations included in the analysis is easily obtained from the fixation map. The analysis requires that a critical threshold value of d, d_{crit} is defined, as in the previous section. Clearly, the value of d_{crit} is in some way implied when setting a width for the 3d gaussian that represents a fixation. As d falls off from the centre of the fixation, it will reach a critical level at which it is not high



Figures 11, 12, 13 and 14 (reading left to right): the original image N1313 (see section 4); map of fixations of 130 traces; corresponding contour plot; and the image redrawn with areas receiving higher numbers of fixations appearing brighter.



Figures 15-17. Fixation maps for two different groups of 65 subjects viewing the same image, and (bottom) the similarity map obtained by subtracting one map from the other.

enough for detection (for example).

To determine coverage, for each map pixel, if d is less than d_{crit}, then d is set to zero, and if it is greater than d_{crit}, d is set to 1. The sum of values of d across the map divided by the area of the map therefore gives the proportion of the map effectively covered by fixations. This is equivalent to the fractional area occupied by the "islands" in figure 6 (17%).

2.5 Graphical descriptions

The principal goal in undertaking fixation map analysis might be said to be to answer the question: "Where in the image did people tend to look?" This is particularly the case when communicating results to a non-academic audience. The simplest and most direct solution is to change the image in some way so as to reflect the variation in height of the fixation map.

The chosen method was to change the luminance of each pixel in the original image to reflect the value of d from the fixation map at the point. In other words, those parts of the image which received more fixations appeared brighter than those areas which received fewer fixations.

Various other methods were considered. The degree of blurring of parts of the image can be altered depending on the value of d at that point, with those areas receiving the highest density of fixations appearing more clear. It was felt that this method was open to misinterpretation, particularly with fixation maps created from multiple traces, as the graphical representation might be construed as a representation of visual function.

Results of the luminance method are shown for two images: figures 7, 8, 9 and 10 for image N0931; and figures 11, 12, 13 and 14 for image N1313. In each block of four images the top left image is the original image; the top right is the fixation map for around 130 individual traces; the bottom left image is the contour plot corresponding to the fixation map; and the bottom right image is the original image modified so that areas receiving larger numbers of fixations appear brighter. Figure 10 illustrates how the distribution of fixations with image N0931 mainly occurred in a very small informative part of the image, leaving the remainder of the image in darkness. Figure 14 illustrates that image N1313 provided a greater number of areas of interest to the observer, and these were well distributed around the image.

In part this paper contributes to the growing body of work on information visualisation. Thomas et al. [2001] describe the advantages of such descriptions of data: researchers use their own visual systems to detect emerging patterns and trends in graphical representations of data sets, determining where best to employ further statistical analysis. Unlike the mapping of complex text data by Thomas et al., the spatial relationship between different fixations is not a construction of the analysis used, rather it is their real spatial separation relative to the stimulus used. This makes the analysis of fixation data by fixation maps both powerful and more straightforward. Visualisation is only one *application* of fixation map analysis, and is the least quantitative of the applications described here. It does, however, provide a useful and intuitive means of describing and communicating the underlying patterns of otherwise overwhelming data sets.

2.6 Similarity

Another application of fixation map analysis is the quantification of the similarity between eye-movement traces. Nearly all previous studies examining similarity in eye-movement traces have relied on a subjective decision by the authors as to the degree of similarity, or more usually, the general "family" of similar traces to which a particular pattern belongs.

The quantification of the similarity between one eye-movement trace and another initially appears a simple task. It is easy, however, to create a number of artificial traces which do not *appear* similar, while one trace comprises fixations with identical *locations* to those in the other trace. This raises two important questions:

1. How important is the path in assessing similarity of traces?
2. What importance should be given to differences in the path? For example, if one trace involves two saccades in travel between two locations, how similar is this to a trace in which the journey is made with only one saccade?

The answers to these questions will, of course, depend on the analysis being undertaken. However, for many purposes it is the *location* of fixations not the *order* which is important. This is fortunate, as the addition of order would markedly increase the complexity of the analysis, assuming it is possible to come by a solution that is sufficiently objective. A subjective interpretation, unavoidably taking into account the path and drawing on an unknown internalised set of weightings in the mind of the individual undertaking the analysis, will lead to a misleading and probably unrepeatable answer.

Attempts at quantifying similarity are rare. Mannan et al. [1995] concentrate solely on fixation location without regard to the order in which the fixations were made. They use a least squares index of similarity, I_S, which examines the distance between a fixation in one trace and its nearest neighbour in the other trace. Their index compares a single trace with another. In order to give similarity measures for populations of traces (e.g. in the calculation of intra- or inter- group similarity) it is necessary to average the similarities of separate trace pairs.

At the present time a reliable method of calculating the similarity of two fixation maps is still being developed. Clearly fixation map analysis promises to be a powerful tool in this area, as it allows the comparison of fixation maps of individuals and of groups. At the simplest level, two normalised maps (figures 15 and 16) can be subtracted from each other. A third *similarity* map is thereby created (figure 17), in which the d value for each map pixel is the magnitude of the *difference* of the d values for the corresponding map pixels in the two original maps. The value of d in the similarity map will range from zero (for identical) to 1 (for completely dissimilar). In the example given in figures 15 to 17, 130 subjects who had viewed the same image were randomly assigned to two groups of 65. Figure 15 is the fixation map for the individual traces of the subjects in the first group, with figure 16 being the fixation map for the subjects in the second. Figure 17 (the magnitude of the difference between the two maps in figures 15 and 16) shows that there is little difference in regions fixated by the two groups, and the mean value of d per map pixel in this similarity map is appropriately low at 0.04.

3. CONCLUSIONS

Fixation map analysis represents an objective method of quantifying aspects of eye-movement traces. It can be used to define various parameters of the eye-movement trace, including the degree of coverage and areas of interest. Fixation map analysis can provide a method of comparing the similarity of traces, or of populations of traces. Finally, it has proved itself to be a useful technique in the visualisation and communication of large eye-movement data sets.

4. IMAGES USED

The images of paintings used in this study all appear with the permission of the National Gallery, and remain © National Gallery, London, with annotations © IBS, University of Derby.

- N0931: *Christ addressing a kneeling woman.* Paolo Veronese (about 1546). National Gallery, London. Digitised and scaled to 1024 x 734 pixels and 16-bit colour, presented at 22.5 x 15.5 degrees for 20s to each of 131 individuals.
- N1313: *The Origin of the Milky Way.* Jacopo Tintoretto (probably 1575 to 1580). National Gallery, London. Digitised and scaled to 858 x 768 pixels and 16-bit colour, presented at 18.9 x 16.2 degrees for 20s to each of 130 individuals.

5. ACKNOWLEDGMENTS

- The National Gallery "Telling Time" Exhibit was conceived and constructed by David Wooding, Mark Mugglestone, Kevin Purdy and Alastair Gale of the Applied Vision Research Unit, University of Derby, UK.
- The exhibit was conducted in cooperation with the National Gallery, London (exhibition curator: Alexander Sturgis), and with the support of Derby University Enterprises Limited (DUEL).
- The analysis of data is part-funded by the British Academy (grant reference 31757).

6. REFERENCES

OVERINGTON, I. 1976. *Vision and Acquisition.* Pentech Press, London.

MANNAN, S., RUDDOCK, K.H. AND WOODING, D.S. 1995. Automatic Control of Saccadic Eye Movements made in Visual Inspection of briefly presented 2-D images. *Spatial Vision, 9(3)*, 363-386.

THOMAS, J., COWLEY, P., KUCHAR, O., NOWELL, L., THOMSON, J., WONG, P.C. 2001. Discovering Knowledge Through Visual Analysis. *International Conference on Knowledge Management*, Graz, Austria. July 12-13, 2001.