Knowledge-Based Patterns of Remembering: Eye Movement Scanpaths Reflect Domain Experience

Geoffrey Underwood*, Katherine Humphrey, and Tom Foulsham

School of Psychology, University of Nottingham, UK
geoff.underwood@nottingham.ac.uk, lpxkah@nottingham.ac.uk,
lpxtf@nottingham.ac.uk

Abstract. How does knowledge of a domain influence the way in which we inspect artefacts from within that domain? Eye fixation scanpaths were recorded as trained individuals looked at images from within their own domain or from another domain. Sequences of fixations indicated differences in the inspection patterns of the two groups, with knowledge reflected in lower reliance of low-level visual features. Scanpaths were observed during first and second viewings of pictures and found to be reliably similar, and this relationship held in a second experiment when the second viewing was performed one week later. Eye fixation scanpaths indicate the viewer's knowledge of the domain of study.

Keywords: Eye movements; scanpaths; knowledge-based processing; saliency.

1 Introduction

When we are introduced to a domain of knowledge and educated about its aspects, our perceptions of the domain are changed, and our enquiries of new artefacts may reflect these altered perceptions. One tool that can be used to investigate the perceptions of knowledgeable individuals who are inspecting pictures of domain-specific artefacts is the recording of the viewer's eye movements. It is known that the eye fixations of trained drivers, for example, are different from those of novices, even when watching roadway scenes rather than engaging in the active task of driving itself [1, 2]. In normal vision, scanning behaviour consists of saccades: fast, ballistic shifts of gaze that bring regions of interest onto the area of the eye with the highest resolution (the fovea). The measurement of these movements is now common in many diverse areas of cognitive psychology including research into reading [3] the perception of pictures and scenes [4], problem solving [5] and complex behaviours such as driving [1]. A variety of measures are taken in such experiments, and these measures reflect assumptions about the functioning of the visual-cognitive system: firstly, that the accuracy with which saccades are targeted is an indication of early attentive processing based on peripheral vision (for example, in the analysis of saccade landing positions in reading, or the measurement of saccade lengths as an index

* We are grateful to the U.K. Engineering and Physical Sciences Research Council for support (EPSRC award EP/E006329/1), to Laurent Itti for the use of his saliency software, and to two anonymous reviewers for their comments.

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 125–144, 2008.

of performance in visual search); secondly that the position and duration of a fixation are reliable diagnostics of what is being processed and the difficulty of this processing. Thus fixation durations and gaze durations (the sum duration of consecutive fixations within a region) are commonly explored to indicate cognitive processing of different stimuli. What these measurements have in common is that they generally consider each eye movement event independently. However, the pattern of inspection can only be revealed by considering a sequence of successive fixations. These patterns are sometimes referred to as *scanpaths* or *scan patterns* [4].

What determines where the eyes will move to next? Early researchers considering this question recorded a large variation in the scanpaths made when observers viewed complex stimuli such as pictures, but there were patterns [6, 7]. A glimpse at some scanpaths (see Figure 2) shows that the places where people fixate, and the movements between them, are not random, and neither are they regularly distributed across space, as we might suppose if the visual system were trying to sample the whole scene uniformly. If looking at a picture containing people for example, fixations tend to be focused on the faces. In fact, fixations across many stimuli tend to cluster on regions rated informative [8] and some researchers have analysed the low-level statistical properties of these image regions [9], in an attempt to identify the bottom-up determinants of attention.

In addition to being tied to the visual features present in a stimulus, the pattern of eye movements made by an observer is known to vary according to the task being undertaken. In his often-cited early work on eye movements, Yarbus [7] highlighted the fact that scanpaths exhibited when viewing the same stimulus would be quite different if the viewer was given a different task. Two commonly studied experimental tasks are looking at a scene in order to remember it for later and searching a scene for a specific target. The between-subjects variation in scanpaths has led some to label scanpaths as distinctively idiosyncratic, presumably reflecting personal knowledge, experience or viewing strategy. To study these top-down aspects of overt attention it is useful to be able to compare scanpaths across viewings, stimuli and individuals. Before looking at this technique in more detail, we will consider a specific theory for which scanpath comparison is particularly important. It is necessary here to distinguish between scanpath theory and the measurement of observed scanpaths.

Scanpath theory is an ambitious set of ideas that were originally proposed in two papers by Noton and Stark [10, 11]. The theory describes scanpaths as controlled by internal, cognitive models representing the viewer's expectations of the scene. These models might represent the saccades involved in viewing a picture or scene as a kind of structure or syntax that binds together the features processed at fixation. When viewing the same scene again, as in the test phase of a recognition experiment, a scanpath might be re-invoked or checked against the external stimulus. The main evidence for scanpath theory came from experiments showing that scanpaths recurred when stimuli were reviewed in a recognition task. In Noton and Stark [10] this conclusion was reached based on subjective observation of the patterns shown by each subject and there was no quantification of the similarity between the scanpaths. Other researchers have examined the presence of repetitive scanpaths when imagining a previously seen image [12, 13].

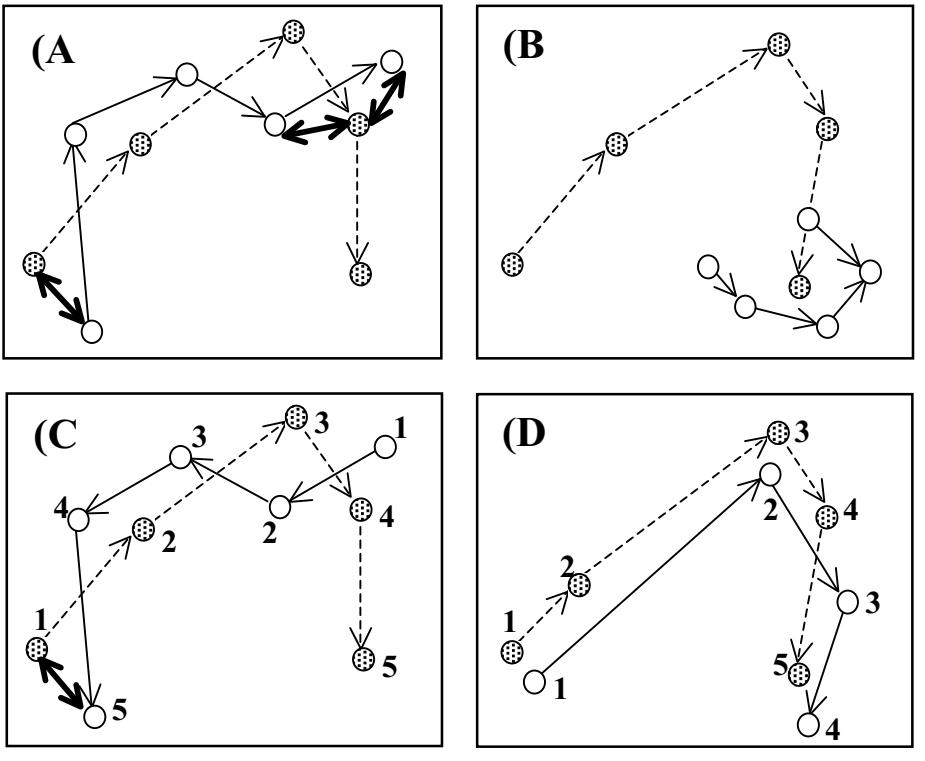


Fig. 1. Calculating the linear distance between two hypothetical scanpaths. Circles show fixation locations. (A) Each fixation is compared to its closest neighbour in the other scanpath. This distance is illustrated for three fixations (bold arrows). (B) This metric is confounded by differences in the spatial variability of the two scanpaths. All of the fixations in one scanpath (open circles) will be compared to just one in the other set, leading to a low mean distance despite very different patterns. (C) The metric ignores sequence information. Ordinal position is emphasised with numbers. Note that the first fixation is compared to the fifth fixation in the other set. (D) If each fixation is compared with that in the same serial position, small differences skew later comparisons. In the case illustrated, despite broadly similar scanpaths, the distance between second and subsequent fixations is large.

Little support has been found for scanpath theory and it has difficulty explaining some phenomena. For example, it is not necessary to move ones eyes to encode or recognise a picture, and the apparently large amount of variability within the patterns shown by a single person viewing the same stimulus also make a strong version of scanpath theory untenable [9]. As a result, some researchers prefer to use the term *scan patterns* rather than *scanpaths* [4], in order to dissociate the fixation recordings from the theory. We do not rely upon any of the assumptions of scanpath theory here. Instead, we will restrict ourselves to a discussion of how the relationships between scanpaths might be quantified, for purposes of comparing the fixation patterns recorded during the viewing of the same picture on two separate occasions.

1.1 Methods for Comparing Scanpaths

In this section, several different basic methods for comparing scanpaths will be reviewed. Figure 1 shows some hypothetical scanpaths and these will be used as examples here. Alongside the practicality of using a method on large datasets, the main criteria for evaluating these methods will be whether it appropriately captures the degree of similarity between different scanpaths and the ease of testing this statistically.

1.1.1 Distance-Based Methods

Scanpaths are inherently spatial. As a result it would seem most appropriate to measure the distance between two scanpaths superimposed on the same visual area. A metric developed by Mannan and colleagues [9, 14, 15] computes the similarity between scanpaths by measuring the distance between each fixation in one set and its nearest neighbour in the other. Scanpaths that are more similar, in the sense that they dwell on locations close to each other, will show a smaller average linear distance. Figure 1 depicts this measurement for several different comparisons. The average linear-distance is defined as D, where

$$D^2 = \frac{n_1 \sum_{j=1}^{n_2} d_{2j}^2 + n_2 \sum_{i=1}^{n_1} d_{1i}^2}{2n_1 n_2 (a^2 + b^2)} \tag{1}$$

and where n_1 and n_2 are the number of fixations in each scanpath and a and b are the dimensions of the image. d_{1i} is the distance between the *ith* fixation in the first set and its nearest fixation in the second set, and d_{2j} is the same distance for the *jth* fixation in the second set.

Computation of this measure is straightforward from the fixation coordinates that normally make up raw fixation data. In addition, it is robust to scanpaths with different numbers of fixations and is scaled relative to the size of stimulus being viewed (due to the term $a_2 + b_2$). In order to produce an estimate of the absolute degree of similarity, Mannan et al. [9] compute the similarity index, I_s, by comparing the average linear distance between two scanpaths with that between randomly generated scanpaths of the same size (D_r):

$$I_s = \left(1 - \frac{D}{D_r}\right) 100 \tag{2}$$

This gives a value between 0 (chance similarity) and 100 (identity), with negative values indicating scanpaths that are more different than expected by chance. The distance between randomly generated scanpaths (D_r) produces the normally distributed similarity that would be expected from chance or uniform scanning. This distribution was examined by Mannan et al. [9]. For a constant display size, the average random distance gets smaller as more fixations are added to the scanpath (as n_1 and n_2, which do not have to be equal, increase).

The second problem occurs when the spatial distribution in one set of locations is very different from that in the other (see Figure 1b). This leads to multiple fixations being compared to a single location in the other scanpath, potentially producing high

similarity from two scanpaths that appear very different. Similarly, one outlier will skew two otherwise very similar scanpaths. Tatler, Baddeley, & Gilchrist [16] identified these problems, pointing out that the linear distance method is fundamentally confounded by differing amounts of variability in the two scanpaths. Henderson, Brockmole, Castelhano, & Mack [17] proposed a "unique assignment" variant of the linear distance metric whereby each fixation is paired with just one other. All possible pairings are computed, and that chosen which minimises the average distance. A disadvantage of this approach, and of the serial position version, is that they require equal numbers of fixations in each set.

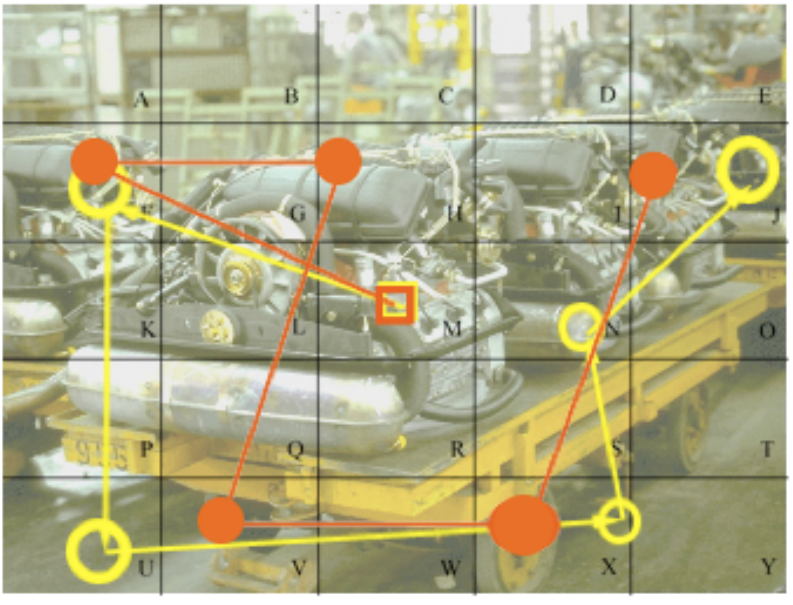
Figure 1 illustrates two specific problems with the linear distance method. Firstly, the measurement does not take into account the temporal sequence of the scanpath. Fixation locations are compared to whichever fixation is closest, regardless of when it occurred. As it ignores the order information, this metric would give extremely high similarity to the example in Figure 1b, despite the fact that in one scanpath the observer starts at the bottom left and works upwards whilst in the other they do the opposite. One way to avoid this problem might be to compute a "serial position" version, where the distance is computed between each fixation and that fixation which occurred in the same serial position in the other scanpath. However, this would be skewed by any small deviations, as illustrated in Figure 1d.

1.1.2 String Edit-Distance

In order to capture the temporal order of scanpaths, several researchers have utilised a method designed specifically for sequence analysis: the Levenshtein distance [18], or simply string edit-distance [12, 19]. This algorithm is an extension of the Hamming distance that gives the difference between two strings of symbols in terms of how many positions are identical. The edit-distance is defined as the number of editing operations (deletions, insertions and replacements) required to turn one string into another, and this distance will decrease as strings become more similar. A method (based on discrete dynamic programming) is available which computes the minimum number of operations required, and this distance has been used for comparing a range of different strings of items, from DNA sequences to birdsong [20]. Figure 2 illustrates how this method can be applied to eye movement sequences. The visual stimulus is divided into regions, each of which is allocated a letter. Each 2-dimensional scanpath can now be transformed into a character string, and the edit distance between the two can be computed. It is often desirable to compare similarity across scanpaths of different lengths, so the distance can be normalised by the number of fixations, and an index of similarity, which here we call s calculated from its reciprocal distance between two scanpaths is calculated as the minimum number of steps required to transform one string into another:

$$s = 1 - \frac{d}{n} \tag{3}$$

where d is the edit-distance between two scanpaths and n is the number of fixations within the longest scanpath. This metric is equivalent to *Ss*, the first of three string-based similarity measures identified by Privitera [21]. *Ss* is the sequential similarity, whilst *Sp* (locus similarity) represents the number of characters shared by both



String 1: *FUXNJ* (open circles) String 2: *FHVXJ* (filled circles)

Editing cost in transforming String 1 into String 2:

String item 2:	*U* to *H*	= one replacement
String item 3:	*V* inserted	= one insertion
String item 4:	*N* deleted	= one deletion

Total editing "cost" = 3 changes in a string of 5 items
Normalised difference = 3/5 = 0.6
String similarity = 1 – 0.6 = 0.4

Fig. 2. Each circle indicates an eye fixation, and here two sequences of fixations are represented. Fixation durations are suggested by variations in the sizes of the circles. A string-editing procedure is used to evaluate scanpath similarity by calculating the "editing cost". The distance between two scanpaths is calculated as the minimum number of steps required to transform one string into another. This edit cost can be normalised and converted into a standardised similarity score, where a score of 1 represents two identical strings.

strings, giving an index of the positions both scanpaths dwell on, regardless of order. The final metric mentioned is *St* (transition similarity), which consists of Markov matrices of region transitions.

There are several important decisions to be made if using the edit distance method. Firstly, how is the region schema produced? In some cases there are clear areas of interest that can be predefined by the researcher. These might correspond to areas of a display or particular objects in a scene. However, in other cases it might be desirable to look at scanpaths over the whole image and to use regions of a constant size. In this situation the image can be divided into a grid, although this raises the question of how large these regions should be. A third possibility is to use the fixation data themselves to produce the regions, using statistical clustering techniques, for example Privitera & Stark [22]. Thus the region schema could be found which divided the

image into the desired number of regions. It might be useful to perform the analysis with several different sets of different regions, perhaps of varying sizes. Similarity that is present at several scales and robust to changes in the region organisation is likely to be the most reliable. Choi et al. [19] argue that the estimates of similarity that they give are robust, whether using 10 or 15 regions. Figure 3 illustrates the consequences of varying the grid size, and can be used to suggest an optimal size in which increasing the number of grids has no further discernible effect on the Levenshtein distance value.

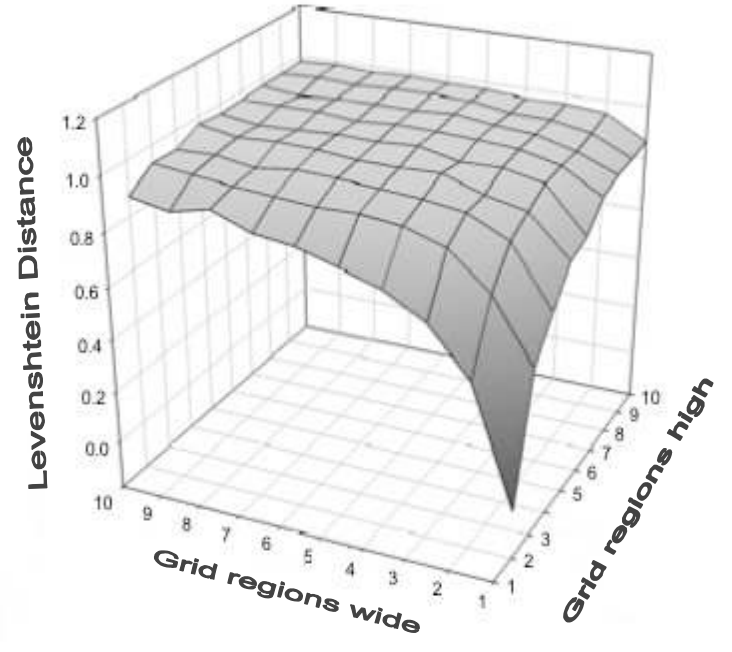


Fig. 3. The normalised Levenshtein string-editing distance between the fixations in two randomly generated scanpaths varies with the size and number of regions. Data are based on scanpaths over an area divided into a grid of regions with dimensions varying from 1 x 1 (only one region where all fixations are evaluated as equal) to 10 x 10 (100 regions). Note that the distance expected by chance increases as a finer grid is used.

A second decision that is commonly made is to condense consecutive fixations on the same region, which would result in repeated symbols in a string, into a single character. Groner, Walder, & Groner [23] make a distinction between local and global scanning, with the former consisting of small readjustment saccades, of less theoretical interest. Thus the coarser scale movements between regions may be more useful. Of course, in combination with decisions regarding the size and shape of regions this will have a profound effect on estimates of scanpath similarity.

How can we calculate the significance of *s*? As with distance-based methods this problem amounts to comparing experimentally derived similarity with a random or chance estimate. The chance similarity can be easily calculated as the probability that

any two characters will be the same. Thus for a 5 by 5 grid there is a 1/25 chance that any two fixations will be in the same place. The similarity of randomly generated scanpaths could be used as the denominator in an estimate of absolute similarity analogous to equation (2). Randomly generated similarity varies with the size and number of regions (as demonstrated in Figure 3), so the same grid would need to be used as that used with experimentally derived data. The random model might also need to be adjusted to take into account other biases present in the experimental sample. For example scanpaths might be constrained to start or finish in a particular place, and this would affect similarity estimates.

Although the editing-distance method successfully captures the temporal sequence of scanpath data, it reduces all spatial information to a binary choice where fixations are either in the same region or they are not. This seems intuitively unsatisfactory and leads to some comparisons being equivalent despite large differences (see Figure 1b). This problem makes the regions used, often a fairly arbitrary decision, critical, as a fixation which lies just over the region border will be counted the same as one which is on the other side of the image. In an extreme case, a fixation might be computed as more similar to a fixation that lies in the same region than one that is spatially closer but outside the region's bounds. In one sense, using more regions provides a more accurate representation with a higher resolution of the movements made. However, more regions also make the analysis less tolerant of small deviations that might otherwise seem negligible.

1.2 Other Methods

We have outlined two main methods for comparing scanpaths, but there are several other ways of analysing such patterns. Some researchers have used Markov matrices [1, 24, 25], which show the transition probabilities from one predefined region to another. However, while this may be useful for short scanpath segments, the matrices explode exponentially when longer chains are explored, making them impractical. They also require the same decisions regarding which regions to use as the edit-distance.

Fixation maps such as those shown in Figure 4 are a useful way of displaying eye movements, particularly those from large populations [26]. In these maps, fixations are represented by a two-dimensional Gaussian centred on the fixation location. The width and height of the Gaussian can be varied, and multiple fixations summed, forming an attentional landscape. Comparing two fixation maps is then possible, and as fixations are essentially distributed this may be an efficient way of computing the spatial similarity between two scan patterns which avoids some of the confounds associated with linear distance. Two maps could be correlated or a difference map could be produced (perhaps after normalising the height of the peaks). Spatially identical scanpaths would give a completely flat difference map. Standard fixation maps hold no information regarding sequence, although it might be possible to introduce a temporal element, either by combining maps derived from different time periods or by varying the height of fixation peaks over time. The fixation map approach also provides a way of identifying the regions of interest from the data (for use in the edit-distance method, for example). A threshold or critical value can be chosen, and all

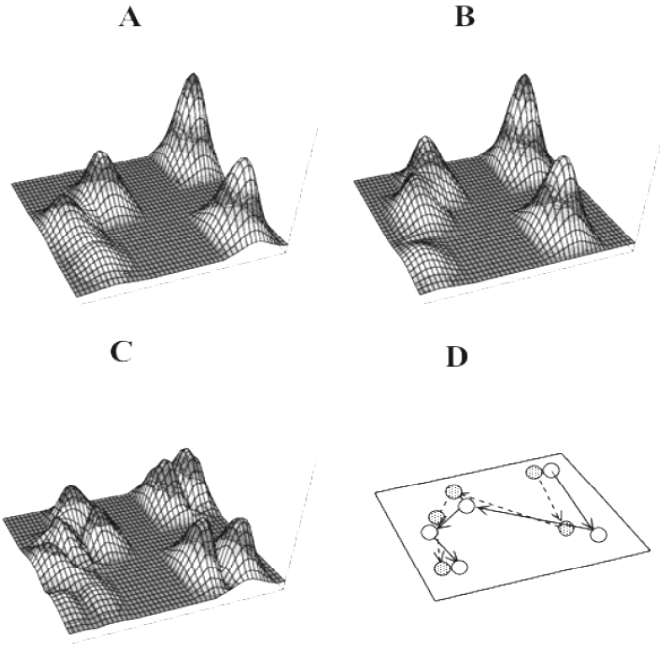


Fig. 4. Fixation maps can be used to represent and compare scanpaths. Each fixation is represented by a 3D Gaussian. The z-axis or height could represent fixation duration or another index. Alternatively, in order to encode sequence, height can represent when the fixation occurred. Early fixations produce a high peak, whilst later ones give a progressively lower distribution. Multiple fixations on the same area are summed together, producing an attentional landscape. A) and B) show two fairly similar scanpaths, with z normalised to range from 0 to 1. The absolute difference between the two can be represented in the same way as a "difference" map. Identical scanpaths would show a totally flat difference map, whilst peaks indicate areas where fixation allocation differed between the two. A 2D schematic of the two scanpaths is also shown (D).

areas receiving more attention, those whose peaks lie above the threshold, can be selected for use in further analyses. Alternatively the threshold could be gradually adjusted until the required number of discrete regions is selected. Tatler et al. [16] also avoid contaminating their measure of similarity by looking at the full distribution of fixations across the image. In their method fixations are binned into 2˚ by 2˚ squares and a spatial probability distribution derived. The difference between two scanpaths is then given by a measure from information theory, the Kullback-Leiber divergence, which computes the difference between the corresponding probability distributions. The Kullback-Leiber divergence gives the number of additional bits of information needed to describe one distribution given another. Thus a low value indicates similar scan patterns. A disadvantage of this technique is that it requires large amounts of data, so it is best used when the fixations from many trials and observers are being examined.

In summary, distance-based methods are useful for averaging commonalities in *where* people look, though they are confounded by differences in spatial variability and do not reflect the temporal order of serial scanpaths. The edit-distance approach captures sequence at the expense of spatial resolution but requires somewhat arbitrary decisions such as which region scheme to use. Levenshtein's [18] string editing method is implemented easily and has been used successfully elsewhere, and so will be used to compare fixation sequences in the following two experiments.

2 Experiment 1: Domain Knowledge and Fixation Scanpaths

Fixation scanpaths are more similar over multiple viewings of a picture than would be expected by chance. Independently it has also been found that low-level visual saliency has a large influence on the locations of the first few fixations. However, bottom-up processes such as these may be overridden by top-down cognitive knowledge in the form of domain proficiency, suggesting that fixation locations are determined by multiple factors. In the present experiment domain specialists were asked to look at a set of photographs in preparation for a memory test. They were then given a second set of pictures and were asked to identify each one as being from the previous set, or new (not seen during the encoding phase). The experiment investigates the stability of fixation scanpaths between the first and second viewing of a picture, and the influence of the viewer's own knowledge of the domain from which the pictures were selected.

Regular patterns of fixations may result from fixation on the most conspicuous regions. Each time the picture is inspected, perhaps the viewer looks first at the most conspicuous region, then at the next most conspicuous region, and so on. The conspicuity or saliency distributions do not change between viewings of course, and so the sequence of fixations would not change either. Itti and Koch's [27] algorithm enables the measurement of the visual saliency of an image on the basis of its physical properties, by the identification of peaks in the distribution of intensity and changes in colour and orientation. The algorithm builds an overall "saliency map" of the image to determine the ordinal allocation of attention to the regions of the display. The effect of salience on eye fixation locations has been supported by Parkhurst, Law, and Niebur [28] who showed participants a range of images, including photographs of home interiors, buildings, and natural environments. Saliency strongly predicted fixation probability during the first two or three fixations, and the model performed above chance throughout each trial. Parkhurst et al. concluded that a purely bottom-up account of visual attention was sufficient to account for fixation behaviour. Further support comes from Underwood, Foulsham, van Loon, Humphreys, and Bloyce, [29] who found that when viewers inspected the scene in preparation for a memory task, objects higher in saliency were more effective in attracting early fixations.

A bottom-up explanation for similarities in scanpaths at encoding and recognition could therefore be that fixation locations are at least partly determined by saliency, as this remains constant over viewings. Repeated patterns of fixations may be a product of viewers repeatedly looking at the most conspicuous regions of the picture, and so similar scanpaths may not result from a memory of the first viewing but from the visual characteristics of the picture itself.

Top-down influences are known to reduce the effects of visual saliency on fixation behaviour, and so it is possible that the bottom-up effect of saliency could be moderated by an increase in the viewer's top-down knowledge of the scene. If an effect of domain knowledge on saliency influences eye movements, it would be interesting to see if it was consistent over repeated viewings. This has not been specifically investigated in non-search tasks, although there have been studies that have found a cognitive override of saliency in search tasks [29-31].

The eye movements of experts differ from non-experts; for example, experienced football players have been found to have a higher search rate, involving more fixations of shorter duration than novice players [32]. However, there is little evidence of how the eye movements of domain-specialists and non-specialists vary with the saliency map of an image. Furthermore, if eye movements are related to memory, then the overriding effect of domain knowledge should be constant over time, producing similar scanpaths on multiple viewings of the same stimulus. Specialists are consistently more accurate with the recognition of domain specific targets [33] and they consistently produce scanpaths reliably different from non-specialists [34]. We hypothesise that the fixation scanpaths of non-specialists will be influenced by low-level visual saliency, but that domain specialists will produce different eye-movements to non-specialists on the same picture, and would provide support for the overriding effects of domain knowledge. If domain specialists look at images in ways that reflect their knowledge, then the possibility arises of recording a student's eye movements as an implicit assessment of their knowledge.

2.1 Method

Three groups of students were recruited: 15 Engineers, 15 American Studies students and 15 non-specialists (control group).

Eye position was recorded using an SMI iVIEW X Hi-Speed eye tracker. Ninety high-resolution digital photographs were used as stimuli, sourced from a commercially available CD-ROM collection. Thirty of the pictures were engineering-specific, 30 were Civil War specific, and 30 were of natural scenes such as gardens, parks and landscapes. Half of the pictures in each category were shown in both the first viewing (encoding) and in the second viewing (recognition) phases, while the other half were shown only as part of the recognition test test.

Itti and Koch's [27] model was used to generate saliency maps for the first five most salient regions for each picture (see Figure 5) – the regions of greatest aggregate intensity, colour and orientation variation. The only further criterion for stimuli was that all 5 salient regions were non-contiguous; those pictures where the same or overlapping regions were re-selected within the first 5 shifts were replaced.

Following a practice at the task, the first stage of the experiment began, with 45 pictures shown to all participants (15 engineering pictures, 15 Civil War pictures and 15 natural scenes), presented in a randomised order. Each picture was preceded by a fixation cross, which ensured that fixation at picture onset was in the centre of the screen, and each picture was presented for 3 sec. During this time participants freely inspected any aspects of the picture they chose. After all 45 stimuli had been presented, participants were informed that they were going to see a second set of pictures and had to decide whether each picture was new or old, using the computer keyboard

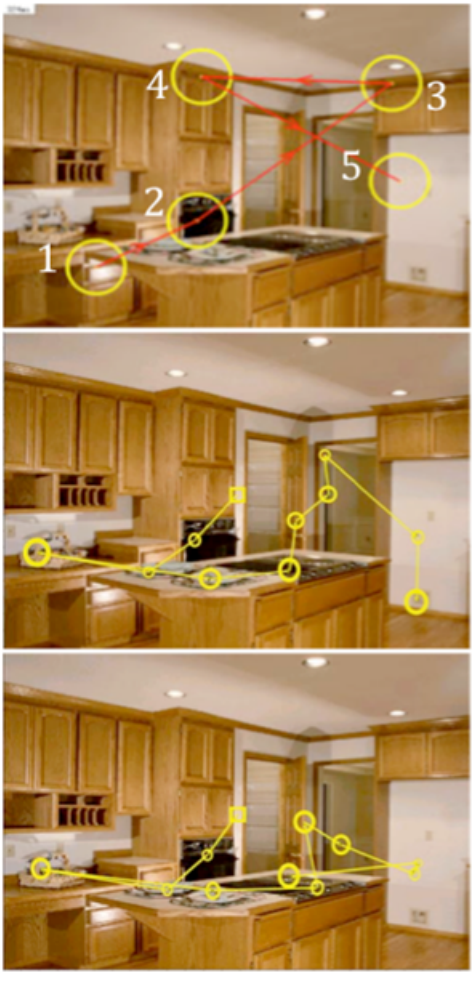


Fig. 5. Saliency map predictions of the first five fixations on a picture from the neutral category (top panel), based on the five most salient regions. The saliency map algorithm identifies the most conspicuous regions of an image on the basis of changes in colour, orientation and brightness. The most conspicuous region is first identified (the brightly illuminated draw in the lower left quadrant of the top picture, numbered "1" above), and then the next most conspicuous region, and so on. A process of inhibition of return prevents the same region being selected repeatedly. The rank orderings of conspicuous regions form the basis for a prediction of the order of the first few fixations on a picture. The top panel, showing the five most salient regions of the picture, and their rank orderings serve as the model-predicted order of fixations, for comparison with actual scanpaths. The centre panel and the bottom panel are exceptionally similar scanpaths from the first and second viewings by one participant.

to indicate whether each picture had been shown previously. During this phase, 90 stimuli were presented in a random order; 45 of these were old and 45 new. Each picture was again shown for 3 sec.

2.2 Results and Discussion

The analyses focused on string analyses, to compare scanpaths during encoding and the second viewing, and comparing encoding scanpaths to the sequences predicted by the saliency model. String editing was used to analyse the similarity between scanpaths produced on encoding and second viewing. A 5 by 5 grid was overlaid onto the stimuli (see Figure 2). The resulting 25 regions were labeled with the characters A to Y from left to right. Fixations were then labeled automatically by the program, according to their spatial coordinates, resulting in a character string representing all the fixations made in this trial.

The scanpaths generated from encoding of a picture were compared with the scanpaths during recognition, and with those predicted by the saliency model [27]. An example of a predicted scanpath is shown in Figure 5 (top panel). Each scanpath was given a score depending on the similarity of the eye fixations on first viewing each picture with the scanpaths predicted by the saliency model. The two lower panels of Figure 5 show examples of observed scanpaths, during encoding (centre panel) and recognition (bottom panel). The average similarity scores (maximum value of 1) for each group of participants were compared for each type of stimulus using an ANOVA.

There was a reliable between-groups difference in viewing Civil War stimuli [$F(2,42) = 4.068$, $MSe = 0.012$, $p<0.05$], with engineers (similarity score of 0.098) and control students (score of 0.107) matching the saliency model closer than American Studies undergraduates (score of 0.054). Paired contrasts indicated that both of these comparisons were reliable at $p<0.05$. There was also a between-groups difference in viewing Engineering stimuli, [$F(2,42) = 16.249$, $MSe = 0.026$, $p<0.05$], with American Studies (score of 0.122) and control participants (0.105) showing scanpath similarity scores closer to the predictions of the saliency model than those of the engineering students (score of 0.0426). The engineers had lower similarity scores than both of the other groups (both comparisons were reliable at $p<0.001$). There was no between-groups difference in viewing neutral stimuli [$F(2,42) = 2.739$, $MSe = 0.015$, $p<0.05$]. When students looked at pictures of artefacts from within their domain of knowledge they were less likely to look at the visually most salient features, and this held for American Studies participants inspecting Civil War photographs and for engineering students inspecting pictures of motors, turbines and industrial production facilities.

The scanpaths generated from encoding of a picture were compared with those on second viewing during the recognition test using an ANOVA analysis. Overall, there was a string similarity of 0.238 for non-specialists, 0.245 for Engineers and 0.268 for American Studies undergraduates. Randomly generated strings would give a value of approximately 0.0417, making the string similarities for all three groups of participants reliably greater than chance ($p<0.05$). There was no difference between the participant groups on the similarity of scanpaths at encoding and recognition [$F(2,42) = 0.522$, $MSe = 0.004$, $P = 0.597$].

Does knowledge of a domain interact with the influence of saliency on scanpaths when viewers look at images from within their domain? All the fixations made on a particular stimulus were compared to the five most salient areas of that picture. In

previous research, saliency effects have been found when memory tasks were performed. In this experiment, it was found that the specialist groups made fewer fixations on the salient areas when the pictures were from their own domain – engineering students made fewer fixations on the visually salient areas of Engineering pictures, and American Studies undergraduates made fewer fixations in salient areas of Civil War pictures. There was no significant difference between groups when looking at neutral pictures, and non-specialists showed no significant difference across stimuli types. This lends some support for the saliency map theory, in that saliency is again shown to have a large influence on eye-movements. However, it does partly argue against the position that a *purely* bottom-up account of visual attention is sufficient to account for fixation behaviour. This is not the case, as salient features had less of an effect when the picture fell into the student's specialist domain. Engineers make reliably fewer fixations in highly salient regions when viewing engineering pictures and American Studies undergraduates made fewer fixations on the salient regions of Civil War pictures. This apparent cognitive-override of saliency may seem intensified because the interesting parts of the stimuli perhaps were of particularly low saliency, and thus it is almost like they were actively seeking out low-salient regions, which would not have been of interest to non-specialist.

The cognitive-override effect that has been found is consistent with previous investigations of saliency influences [29-32] in a search task, but when an encoding task was used, as here, the saliency map did predict fixation locations. This was only found for non-domain specialists here.

Overall, scanpaths produced on encoding of a picture compared to those produced on second viewing were more similar than would be expected by chance. This was consistent across all participants, despite group or stimulus type. Scanpath theory [10] suggests that visual patterns are represented in memory as a network of features and the attention shifts between them. This network is then replayed and compared to the external stimulus when recognising the image later. By this account, the scanpaths at encoding were similar to those at recognition because they were stored and recalled top down, to determine the scanning sequence. Although the similarity seen here is statistically reliable, the scanpaths are far from identical, and there is still a large amount of variance unaccounted for. Previous demonstrations of scanpath similarity have largely used simple patterns, with fewer and larger regions of interest. It is likely that the much more complex photographs used here resulted in reduced scanpath repetition, possibly due to a greater influence of knowledge-based inspection strategies.

In conclusion, saliency does have a strong influence on eye movements, shown by the similarity of actual scanpaths to those predicted by the saliency model [27]. However, domain-specific knowledge can act as an overriding factor, decreasing the influence of saliency on driving eye movements. This effect has been shown to be stable over time.

3 Experiment 2: Delaying the Interval between Viewings

In Experiment 1 the scan patterns recorded during a recognition test were more similar to the patterns seen during the first inspection of the picture than would be expected by chance. The saliency model also had success in predicting fixation locations, but only for

viewers who were relatively unfamiliar with the domain from which the pictures were selected. Comparing performance during a recognition test against performance during encoding may have obscured the analysis, however, because encoding and recognition are different tasks with different cognitive processes. Viewers would have been inspecting the pictures for very different purposes during encoding and recognition. A different estimate of scanpath similarity might be obtained if a comparison was made of scan patterns on two successive viewings of a picture where the purpose of inspection is held constant. In Experiment 2 we again show students pictures from their own domain of knowledge and from another, with eye fixations recorded during encoding and recognition, but also tested recognition a week after the initial viewing, so that scanpaths could be compared during two recognition tests. The control participants revealed no interesting patterns of fixations in Experiment 1, and so only domain specialists were compared here.

3.1 Method

The participants were 15 American Studies and 15 Engineering students from the same source as those tested in Experiment 1, the same pictures were used, and eye position was again recorded using an SMI iVIEW X Hi-Speed eye tracker. Saliency maps were generated using Itti and Koch's [27] model with standard parameters. The experiment used a two-by-three mixed design, with two specialist groups of participants and three specific types of stimuli. All participants viewed the same stimuli under the same task conditions. Test pictures were inspected under one of three viewing conditions: encoding, immediate recognition, and delayed recognition.

Participants were not told to look for anything in particular in any of the pictures but were asked to look at them in preparation for a memory test. Following a practice phase with five pictures not otherwise used in the experiment, the first stage of the experiment began. There were 45 stimuli (15 engineering pictures, 15 Civil War pictures and 15 natural scenes) presented in a randomised order. Each picture was preceded by a fixation cross, which ensured that fixation at picture onset was in the centre of the screen. Each picture was presented for 3 seconds, during which time participants moved their eyes freely around the screen. This presentation format is the same as was used in Experiment 1.

After all 45 pictures had been presented, participants were informed that they were going to see a second set of pictures and had to decide whether each picture was new (never seen before) or old (from the previous set of pictures) by making a keyboard response. During this phase, 90 pictures were presented in a random order; 45 of these were old and 45 new (though the participants were not informed of this fact). In order to facilitate an ideal comparison between encoding and test phases, each picture was again shown for 3 seconds. One week after the original recognition test, participants returned to the laboratory and were shown the 90 test pictures again, with task again being to say whether they had seen each picture during the original encoding phase.

3.2 Results and Discussion

The scan patterns generated from first viewing of a picture were first compared to respective scan patterns predicted by the saliency model [27] again using the string

edit-distance method. Observed scan patterns were found to be more similar to those predicted by the model when stimuli were *not* domain-specific. When stimuli *were* domain specific, the similarity score dropped to the estimated chance level. There was a reliable between-groups difference in viewing Civil War stimuli [$F1,28=52.50$, $MSe=0.099$, $p<0.001$], with a string similarity score of 0.027 for American Studies and 0.142 for engineering participants. There was a reliable between-groups difference in viewing Engineering stimuli [$F1,28=48.75$, $MSe=0.067$, $p<0.001$], with a string similarity score of 0.033 for American Studies and 0.128 for engineering participants. There was no difference between groups when viewing neutral stimuli [$F1,28=1.40$]. Students were less likely to perform according to the predictions of the model when viewing pictures from their domain of interest.

The scan patterns observed during the encoding phase and during the immediate recognition test were also quantified by the edit-distance method, and the resultant similarity scores compared to the estimated chance level, using a one-sample t-test. All comparisons showed the string similarities between encoding and test to be reliably greater than chance. Strings of fixation locations were similar for American Studies undergraduates inspecting Civil War pictures (observed similarity score between encoding and recognition of 0.155, $t14=9.54$, $p<0.001$), engineering pictures (similarity score of 0.147, $t14=10.89$, $p<0.001$), and neutral scenes (similarity score of 0.181, $t14 = 9.89$, $p<0.001$). The two scan patterns were also similar for Engineers looking at Civil War pictures (score of 0.196, $t14=8.69$, $p<0.001$), at engineering pictures (score of 0.247, $t14=13.23$, $p<0.001$), and at neutral pictures (score of 0.203, $t14=8.11$, $p<0.001$).

A comparison was also made between scanpaths in the immediate and delayed picture recognition tests. A mixed-model ANOVA found reliable a main effect of type of picture [$F2,28=31.84$, $MSe=0.077$, $p<0.001$], and no effect of participant group: [$F<1$], but there was an interaction between these two factors [$F2,56=68.25$, $MSe=0.164$, $p<0.001$]. Pairwise comparisons found that for both American Studies undergraduates and Engineers, scan patterns were reliably more similar between encoding and delayed test when they inspected pictures that were within their domain of interest. That is, American Studies participants had higher string similarity scores for their two recognition viewings of Civil War pictures (similarity score of 0.220) than they did for engineering pictures (0.112) or for neutral scenes (0.099), with both comparisons reliable at $p<0.001$. Engineering students showed the opposite pattern, with greater string similarities for engineering pictures (score of 0.278) than for Civil War pictures (0.093) or for neutral pictures (0.091), and both comparisons were again very reliable ($p<0.001$). All six average similarity scores were greater than the value estimated for chance (all contrasts were reliable at $p<0.001$).

4 Discussion and Conclusions

When we look at a picture a second time, do we look at the same features, and in the same order? In each experiment a set of images of real-world scenes was shown to participants who were familiar with the domain from which they were selected, or not. If scanpath similarity depends upon the viewer's knowledge of the domain of the picture then scanpaths could, in principle, be used to assess a viewer's knowledge. The pictures were shown for a few seconds, and then a recognition test performed,

with the students deciding whether they had previously seen each picture. Eye fixations were recorded throughout, and fixation scanpaths were quantified for comparison between the two viewings of the picture. In the second experiment there was a notable addition to the procedure: the recognition test was repeated after a week so that scanpaths could be compared across two viewings that had the same purpose. We also asked about the value of visual saliency in attracting fixations and fixation sequences: do viewers look at visually conspicuous regions in a scene?

The delayed recognition test was introduced because encoding and recognition engage different cognitive processes and so a comparison of scan patterns may reflect these differences rather than any differences between inspection processes. In both experiments, viewers' scanpaths at encoding were more similar to those made when inspecting the same picture at test than would be expected if fixation locations were made randomly. Importantly, scanpaths were similar in successive recognition tests.

The scanpath comparison in the first experiment was between fixation sequences made during encoding and during recognition, and because these are different tasks engaging different cognitive processes, we introduced an additional test in Experiment 2, a second recognition task. This resulted in a new finding, in which a comparison between two recognition tasks, performed a week apart, eliminated the differences between cognitive processes that are inherent in these memory tasks. When comparing the two recognition tasks, in which the same viewers looked at the same pictures on separate occasions, there was a similarity between the scan patterns of the first few fixations. The stability of the scan patterns over time and when the same task is used is of note here.

During both encoding and recognition in the experiments, the regions identified as being visually conspicuous by the saliency map model were fixated more often than other regions, but only for those participants who were not familiar with the content of the picture. Support for the saliency map model is qualified by the extent to which engineering and American Studies undergraduates looked at pictures taken from their own domains of interest and at other pictures. The same pictures were presented to both groups of participants, to eliminate the possibility of any results being attributable to differences between pictures. The saliency map model correctly predicted high proportions of fixations on salient regions for viewers looking at pictures from other domains of interest, but when they looked at pictures from within their domain, there was very little correspondence between the locations of their fixations and the locations of conspicuous items. When an engineer looked at a picture of an engineering plant or a turbine the tendency was to inspect the feature of domain interest rather than the bright, colourful components – the content dominated inspection and the picture's visual characteristics were secondary. Similarly, when American Civil War specialists looked at uniforms, weapons and other artefacts they also responded to the meaning of the items depicted.

The two experiments confirm the predictions of the saliency map hypothesis in the locations of early fixations on pictures of real-world scenes, at least for viewers unfamiliar with the content. Conspicuous regions are fixated more than other regions during the first few seconds of inspection when viewers were encoding the pictures on first viewing. It has been previously reported that fixations are guided to these regions during encoding but not in search tasks [29-31], and the present results extend these conclusions to take into account the prior knowledge of the viewer.

This result is important because it is evidence of individual scan patterns that are picture-based rather than being the product of general scanning strategies or of saliency-driven scanning. If scanpaths were the product of a habitual and stereotypical saccade-generator routine, or indeed of the low-level visual characteristics of an image, then we would expect invariant similarity scores. Instead, when comparing fixation sequences for pictures from different domains of interest, the similarity between scan patterns varied. There was sensitivity to the content of the picture here, with individuals varying their fixations behaviour according to what they were looking at. The knowledge of the viewer influences the way that they inspect a picture, and this raises the possibility of the assessment of their domain knowledge through the observation of their fixation scanpaths. A knowledgeable student will inspect a picture according to its content, whereas an unfamiliar picture will be inspected according to its low-level visual characteristics. Perhaps a student's level of knowledge could be assessed implicitly by observing their eye movements.

A second potential application of the findings is with the user-centered design of computer interfaces. To assess the usability of an interface developers may use rapid prototyping with the early involvement of end users who provide feedback that often involves providing verbal commentaries or reports [35]. Holzinger has argued that we cannot take users' verbal reports at face value, however, and that their actual non-verbal behaviour would provide a preferable measure of interface usability [36]. By comparing the fixation patterns of end users with the patterns provided by expert users, an evaluation could be developed that does not rely upon verbal reports or questionnaires, and that could provide a direct index of a system's intended usability.

References

1. Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., Crundall, D.: Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. Ergon 46, 629–646 (2003)
2. Underwood, G., Chapman, P., Bowden, K., Crundall, D.: Visual search while driving: skill and awareness during inspection of the scene. Trans. Res. F: Psychol. Behav. 5, 87–97 (2002)
3. Rayner, K.: Eye movements in reading and information processing: 20 years of research. Psychol. Bull. 124, 72–422 (1998)
4. Henderson, J.M.: Human gaze control during real-world scene perception. Trends In Cog. Sci. 7, 498–504 (2003)
5. Knoblich, G., Öllinger, M., Spivey, M.J.: Tracking the eyes to obtain insight into insight problem solving. In: Underwood, G. (ed.) Cognitive Processes in Eye Guidance, pp. 355–376. Oxford University Press, Oxford (2005)
6. Buswell, G.T.: How people look at pictures: A study of the psychology of perception in art. University of Chicago Press, Chicago (1935)
7. Yarbus, A.L.: Eye movements and vision. Plenum, New York (1967)
8. Mackworth, N.H., Morandi, A.J.: The gaze selects informative details within pictures. Percept. Psychophy. 2, 547–552 (1967)
9. Mannan, S., Ruddock, K., Wooding, D.: The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. Spat. Vis. 10, 65–188 (1996)

10. Noton, D., Stark, L.: Scanpaths in saccadic eye movements while viewing and recognizing patterns. Vis. Res. 11, 929–942 (1971)
11. Stark, L., Noton, D.: Scanpaths and pattern recognition. Sci. 173, 753 (1971)
12. Brandt, S.A., Stark, L.W.: Spontaneous eye movements during visual imagery reflect the content of the visual scene. J. Cog. Neurosci. 9, 27–38 (1997)
13. Laeng, B., Teodorescu, D.S.: Eye scanpaths during visual imagery reenact those of perception of the same visual scene. Cog. Sci. 26, 207–231 (2002)
14. Mannan, S., Ruddock, K., Wooding, D.: Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. Spat. Vis. 9, 363–386 (1995)
15. Mannan, S., Ruddock, K., Wooding, D.: Fixation patterns made during brief examination of two dimensional images. Percept. 26, 1059–1072 (1997)
16. Tatler, B.W., Baddeley, R.J., Gilchrist, I.D.: Visual correlates of fixation selection: effects of scale and time. Vis. Res. 45, 643–659 (2005)
17. Henderson, J.M., Brockmole, J.R., Castelhano, M.S., Mack, M.L.: Visual saliency does not account for eye movements during search in real-world scenes. In: van Gompel, R., Fischer, M., Murray, W., Hill, R.W. (eds.) Eye Movements: A Window On Mind And Brain, pp. 537–562. Elsevier, Oxford (2007)
18. Levenshtein, V.: Binary codes capable of correcting deletions, insertions and reversals. Soviet Physice-Doklady 10, 707–710 (1966)
19. Choi, Y.S., Mosley, A.D., Stark, L.: Sting editing analysis of human visual search. Opt. Vis. Sci. 72, 439–451 (1995)
20. Sankhoff, D., Kruskal, J.B. (eds.): Time Warps, String Edits And Macromolecules: The Theory And Practice Of Sequence Comparison. Addison-Wesley, Reading (1983)
21. Privitera, C.M.: The scanpath theory: its definition and later developments. In: Proc. SPIE 6057 (2006)
22. Privitera, C.M., Stark, L.W.: Algorithms for defining visual regions-of-interest: Comparison with eye fixations. IEEE Trans. Patt. Anal. Mach. Intell., 22970–22982 (2000)
23. Groner, R., Walder, F., Groner, M.: Looking at faces: local and global aspects of scanpaths. In: Gale, A., Johnson, F. (eds.) Theoretical And Applied Aspects Of Eye Movement Research, pp. 523–533. Elsevier, Amsterdam (1984)
24. Stark, L., Ellis, S.R.: Scanpaths revisited: cognitive models direct active looking. In: Fisher, D.F., Monty, R.A., Senders, J.W. (eds.) Eye Movements: Cognition And Visual Perception, pp. 193–227. Lawrence Erlbaum, Hillsdale (1981)
25. Underwood, G., Phelps, N., Wright, C., van Loon, E., Galpin, A.: Eye fixation scanpaths of younger and older drivers in a hazard perception task. Ophth. Physiol. Opt. 25, 346–356 (2005)
26. Wooding, D.: Eye movements of large populations: II. Deriving regions of interest, coverage and similarity using fixation maps. Behav. Res. Meth. Instr. Comp. 34, 518–528 (2002)
27. Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. Vis. Res. 40, 1489–1506 (2000)
28. Parkhurst, D., Law, K., Niebur, E.: Modelling the role of salience in the allocation of overt visual attention. Vis. Res. 42, 107–123 (2002)
29. Underwood, G., Foulsham, T., van Loon, E., Humphreys, L., Bloyce, J.: Eye movements during scene inspection: A test of the saliency map hypothesis. Euro. J. Cog. Psychol. 18, 321–342 (2006)
30. Foulsham, T., Underwood, G.: How does the purpose of inspection influence the potency of visual saliency in scene perception? Percept. 36, 1123–1138 (2007)

31. Underwood, G., Foulsham, T.: Visual saliency and semantic incongruency influence eye movements when inspecting pictures. Quart. J. Exp. Psychol. 59, 1931–1949 (2006)
32. Williams, A.M., Davids, K.: Visual search strategy, selective attention, and expertise in soccer. Res. Quart. Exer. Sport. 69, 111–128 (1998)
33. McCarley, J.S., Kramer, A.F., Wickens, C.D., Vidoni, E.D., Boot, W.R.: Visual skills in airport-security screening. Psychol. Sci. 15, 302–306 (2004)
34. Manning, D., Ethell, S., Donovan, T., Crawford, T.: How do radiologists do it? The influence of experience and training on searching for chest nodules. Radiography 12, 134–142 (2006)
35. Holzinger, A.: Rapid prototyping for a virtual medical campus interface. IEEE Software 21, 92–99 (2004)
36. Holzinger, A.: Usability engineering methods for software developers. Comm. ACM 48, 71–74 (2005)