

A Vector-based, Multidimensional Scanpath Similarity Measure

Halszka Jarodzka*
Knowledge Media Research Center
Tuebingen, Germany

Kenneth Holmqvist†
Marcus Nyström‡
Lund University Humanities Lab
Lund, Sweden

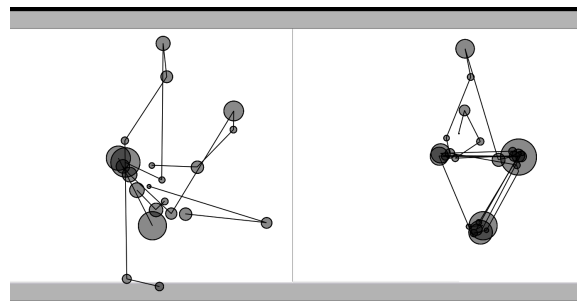


Figure 1: Are these scanpaths similar? Scanpaths from two participants looking at the same stimulus.

Abstract

A great need exists in many fields of eye-tracking research for a robust and general method for scanpath comparisons. Current measures either quantize scanpaths in space (string editing measures like the Levenshtein distance) or in time (measures based on attention maps). This paper proposes a new pairwise scanpath similarity measure. Unlike previous measures that either use AOI sequences or forgo temporal order, the new measure defines scanpaths as a series of geometric vectors and compares temporally aligned scanpaths across several dimensions: shape, fixation position, length, direction, and fixation duration. This approach offers more multifaceted insights to how similar two scanpaths are. Eight fictitious scanpath pairs are tested to elucidate the strengths of the new measure, both in itself and compared to two of the currently most popular measures - the Levenshtein distance and attention map correlation.

CR Categories: I.5.3 [Pattern Recognition]: Clustering—Similarity Measures; I.5.4 [Pattern Recognition]: Applications—Signal Processing J.4 [Social and Behavioral Sciences]: Psychology

Keywords: scanpath, sequence analysis, Levenshtein distance, string edit, vector

1 Introduction

*e-mail: h.jarodzka@iwm-kmrc.de

†e-mail: kenneth@lucs.lu.se

‡marcus.nystrom@humlab.lu.se

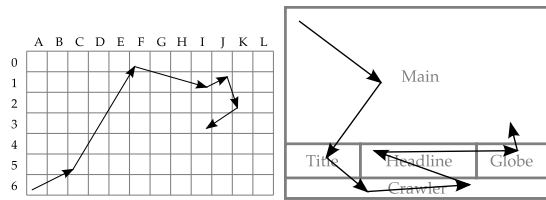
The data emerging from an eye-tracker recording are very information dense. Before data can be sensibly interpreted by eye-tracking researchers, this rich information must be simplified. For instance, area of interest (AOI) measures replace precise fixation positions with imprecise AOI hits, in effect a quantization of space. Attention maps and other point-based measures entirely collapse data over time, losing all sequence information.

Although previous measures that severely simplify the scanpaths provide valuable results for some research questions, they are not sufficient for others. Thus, many researchers stress the urgent need to investigate scanpaths [Underwood et al. 2008; Ehmke and Wilson 2007; Josephson and Holmes 2006; Tzanidou et al. 2005; Josephson and Holmes 2002]. In practice, the need to compare and analyze scanpaths often results in considerable manual data analysis and description of surface features of single scanpath visualizations [Ehmke and Wilson 2007; Josephson and Holmes 2006]. Statistical testing of scanpath similarity between groups has been considered difficult to the level of unachievable [Tzanidou et al. 2005], and the solutions recently invented are bordering to computationally intractable [Feusner and Lukoff 2008].

Previous attempts to replace manual analysis with computer algorithms dominantly fall into one of two categories of which the *string-edit* (a.k.a. *the Levenshtein*) measure [Levenshtein 1966] is the most common. It quantizes spatial information by replacing fixations with a character for the AOI they hit. Thus, the representation of a scanpath is a character string. Similarity is reduced to counting edit operations (insertions, deletions, or substitutions) in pairs of character strings. This measure produces only one value, which can be normalized on the longer, the shorter, or on both strings. The *attention map measures* are a broad class of related measures which all ignore sequence information. Calculation of similarity can be done in a number of ways and results in only one value.

In contrast, we propose a measure that represents scanpaths with geometric vectors. This allows the scanpath to retain a sequence of fixations and saccades and measure similarity using geometry. This approach was not used previously. Moreover, a vector representation of scanpaths gives the opportunity to simplify scanpaths in more controlled and flexible ways than previous, cruder methods.

What the string-edit method and the new measure presented in this paper have in common is that they *align the entities* (characters



(a) Using gridded AOIs to approximate fixation position with AOI hits. The first fixation to approximate fixation position is approximated with the AOI with AOI hits. Both first and last fixation are approximated with AOI “Main”.

Figure 2: Gridded vs. semantic AOIs for representing scanpaths.

or fixations/saccades, respectively) in one scanpath with the corresponding entities in the other scanpath. The representation format and the alignment process are closely interlinked, and together determine the nature of the measure.

It can be argued that the long-term goal for scanpath similarity measures includes a sensitivity for differences in at least *shape*¹, *position*, *length*, *direction* and *duration* between two scanpaths. Two obvious prerequisites of such measures are, first of all, that the scanpath retains information along these dimensions, and second, that the information is used in the calculation of the measures.

2 String-based methods

Several scanpath measures, the most known of which is the Levenshtein string-edit measure, represent scanpaths with a string of AOI hits. When using string-based methods, it is important to distinguish between *gridded AOIs* and *semantic AOIs*. Figure 2 shows both varieties. The gridded AOIs are constructed by putting a grid of equally sized areas across the stimulus, ignoring whatever semantics there is in the stimulus. When a scanpath runs over the gridded AOIs, each fixation is replaced by the name of the AOI it hits. For example, the scanpath in figure 2(a) will be represented by the string A6 – C5 – F0 – I1 – J1 – K2 – I3.

Each such AOI name is an approximation of a fixation point by a whole area, or in other words: We have quantized space. By definition, quantization introduces noise; a small difference in gaze pattern would be enough to alter the string, but interestingly, other small differences can result in the *same* string. That is, some differences matter, but other differences are ignored.

A division of stimulus space into semantic AOIs adapts to the natural meaningful parts of the stimulus. In figure 2, the semantic AOIs are taken from Josephson and Holmes [2006], who used them to analyse viewing behaviour on a television screen. Semantic AOIs have different sizes, and therefore the approximation of fixation position can be very coarse. In figure 2, the scanpath example will be represented with the string M – M – T – C – C – H – G – M (where M = ‘Main’ etc.). The three M’s in the string represent fixations with very different positions in the stimulus, which means we retain a very coarse position approximation. This can be motivated if the ‘Main’ AOI is a semantically homogeneous area from the viewpoint of the theory, hypothesis and analysis in the study using the measure. Otherwise, very different scanpaths will be represented as equal when using semantic AOIs.

¹Shape in this paper refers to a vector \mathbf{u} , and the difference in shape refers to the difference between two geometric vectors, $\mathbf{u} - \mathbf{v}$

A third, less common way to define AOIs is through clustering of eye-movement data [Santella and DeCarlo 2004]. Then the AOIs comprise regions that contain proportionally many fixations from both scanpaths. The disadvantage is that those AOIs may be rather used for post-hoc data explorations than for hypothesis testing.

In the string M – M – T – C – C – H – G – M, we have repetitive AOIs for M and C (‘Crawler’), since our sample scanpath has two successive fixations in each of ‘Main’ and ‘Crawler’. Brandt and Stark [1997] use a variety of the AOI-string representation of scanpaths that ignores repetitions, or in other words: They represent dwells in AOIs rather than fixations. The dwell-based AOI-string for the same scanpath would be M – T – C – H – G – M. Removing repetitions, sometimes referred to as compressing the string, is in fact a sort of clustering method: Several fixations (C – C) are replaced by one dwell (C).

The major advantages of the AOI string representation of scanpaths is that it retains an approximate *sequence* of fixations, and is a rough representation of the *shape*, *length* and *direction* of the scanpath. Fixation *durations* are difficult to represent. The major drawback, is the reliance on AOI segmentation, in essence a spatial quantization which necessarily introduces artefacts in the measures relying on it.

3 Attention maps and other point-based methods

Without movement forward, can there be a scanpath? The *temporal sequence*, the order in which the scan progresses, suggests itself as a necessary property of scanpaths. Without it, two scanpaths with reversed order but otherwise identical would be indistinguishable. Nevertheless, the attention map measures and others that ignore order – such as the Mannan distance measure [Mannan et al. 1996] – are often referred to as “scanpath metrics” in literature [Cerf et al. 2009; Boccignone et al. 2005]. Similarly, in their overview over what they refer to as scanpath metrics, Underwood et al., [2008] discuss both the Mannan distance measure and fixation maps, before settling for the Levenshtein string-edit analysis.

An attention map is calculated on the basis of a set of points (raw data sample or fixations) with no order. Typically, the superposition of Gaussian functions centred on these points forms a probability landscape and, when represented with colour, a “heat map”. There are many measures for calculating similarities between two such landscapes, heat maps, or sets of points, including simple subtraction [Wooding 2002], the earth mover distance [Dempere-Marco et al. 2006], the correlation coefficient [Ouerhani et al. 2003; Rajashekar et al. 2008], the Kullback-Leibler Divergence [Rajashekar et al. 2004; Tatler et al. 2005], and average landing altitude [Itti 2005]. All these methods reduce similarity to a single scalar.

Attention maps represent the spatial *positions* and the scanpath *shape* (except sequence). The *length* and *direction* of the scanpath are ignored, as their calculation would depend on sequence. Fixation durations may or may not be included in the representations, depending on how the attention maps are precisely calculated.

4 The proposed method

4.1 Representing Scanpaths as Geometrical Vectors

We propose to represent the scanpath with vector mathematics. A vector $\mathbf{u} = [x, y]$ is a mathematical entity with direction and length, and thus suitable for representing an idealized saccade². The vector representation does keep the length and the direction of each

²A saccade that takes the shortest distance between two points.

saccade fully intact. An advantage of this representation is that fixations can be represented inbetween those saccadic vectors. Thus, the *position of fixations* are the precise start and ending positions of the saccades. Fixations can of course have durations. Together, fixations and saccades can be used to represent a scanpath, and is very analogue to the scanpath itself.

The representation of scanpaths with vector mathematics preserves:

1. the shape (vector representation) of the scanpath.
2. the length of the scanpath saccades (ignoring saccadic curvature).
3. the direction of the scanpath saccades.
4. the position of fixations.
5. the duration of fixations.

In the following sections, we will implement and evaluate the proposed scanpath comparison method using vector representations.

4.2 Scanpath simplification

Besides providing an intuitive description of a scanpath, a vector based representation makes it easy to simplify a scanpath, so that it retains information that facilitates the alignment of, and comparison between two scanpaths.

For example, a scanpath representation that aims at assessing the overall shape of a scanpath should not include every single saccadic vector, especially if a whole series of small saccades looks like a single point or a single line. String editing techniques for scanpath comparison, for example, are often preceded by simplification procedures where repetitions in letter strings are removed [Brandt and Stark 1997]. This section describes how to cluster a vector based scanpath representation from a local (fine) to a global (coarse) level, but with the constraint of removing as little vital information about the scanpaths as possible. Obviously, this is a trade-off; if too much is removed, the scanpath loses its meaning, but if too little is removed, the scanpaths may be difficult to align with each other, and small saccades can become unproportionally influential in the similarity calculations. To meet this trade-off, we present a two-step method: First, within a scanpath, several small saccadic vectors usually occur in one area, which differ in their directions. In this case, the purpose is not to focus on those differences, but rather to focus on the fact that the person looked at this area. Thus, any group of consecutive saccade vectors $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}, m > 1$ with amplitudes smaller than a threshold, T_{amp} are replaced with a new vector $\mathbf{u}' = \mathbf{u}_1 + \mathbf{u}_2 + \dots + \mathbf{u}_m$ (amplitude-based clustering). Consequently, very short vectors that have different directions are clustered, because they are likely to represent the inspection of one stimulus feature and therefore do not belong to the overall shape of the scanpath. Second, within a scanpath several consecutive saccades often have similar directions. Analogous with the amplitude based clustering, any two consecutive saccades that are directed at most T_ϕ radians apart will be added into one larger saccade (direction-based clustering). As a consequence, consecutive vectors having a very similar direction are added, because they do not by themselves contribute to the overall global shape.

Fixation durations associated with the merged vectors are still represented in the simplified scanpath; each fixation at the beginning of a merged vector is added to the beginning of a global vector replacing it. This way, the scanpaths retain most of the information about the dimensions used to evaluate their similarity. Figure 3 illustrates what amplitude and direction based clustering do to a hypothetical scanpath.

This type of clustering extends prior methods for clustering of local and global scanpaths [Groner et al. 1984], because saccadic vectors are not only clustered according to their distance, but also according to their direction.

4.3 Temporal Alignment of Simplified Scanpaths

In this step the scanpaths are temporally aligned to each other based on their shape. In order to do that, we used the clustered (simplified) scanpaths. Thus, we compare two scanpaths that are characterized as a series of fixations and vectors.

Comparison between two scanpaths, $S_1 = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$ and $S_2 = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$, is performed in the following steps (illustrated in Figure 4):

1. According to a similarity metric, calculate how similar each element i , e.g., saccade amplitude, in one scanpath is compared to each element j in the other scanpath. Let $\omega_{(i-1)m+j}$ denote the similarity between elements i and j . Collect the results in a matrix $M(i, j)$.
2. Create an adjacency matrix $A(k, l), \{k, l\} = 1, 2, \dots, mn$ for $M(i, j)$, i.e., define a set of rules of how matrix elements in $M(i, j)$ are connected. Associate each connection with a weight $\omega_{(i-1)m+j}$.
3. Create a graph representation from the adjacency matrix, where matrix elements are called nodes.
4. Find the *shortest path* from the node 1 to (the last) node mn in the Graph using Dijkstra algorithm [Dijkstra 1959].
5. Align scanpaths along the shortest path.

In this paper, we always align scanpaths based on their shapes. However, the alignment may be performed on other dimensions, if they are of greater interest in a particular study.

4.4 Scanpath comparison

For each pair of fixation and saccade vectors in the simplified and aligned scanpaths, the following measures are computed (average):

1. difference in shape between saccade vectors $\mathbf{u}_i - \mathbf{v}_j$,
2. difference in amplitude (length) between saccade vectors $\|\mathbf{u}_i - \mathbf{v}_j\|$,
3. distance between fixations,
4. difference in direction (angle) between saccade vectors,
5. difference in duration between fixations.

Consequently, we end up with five measures indicating how similar two scanpaths are along different dimensions. To obtain similarity values between $[0, 1]$, the first three measures are normalized by the screen diagonal. Direction is normalized by π , whereas the difference for each fixation duration is normalized against the maximum duration of the two durations being compared.

5 Results and Discussion

The proposed algorithm is compared to the Levenshtein distance [Levenshtein 1966] and the correlation coefficient r between attention maps, which are two of the most common measures for scanpath comparison. The comparison is made on eight pairs of fictitious scanpaths constructed to illustrate where the proposed measure could be particularly useful. Prior to computing the Levenshtein distance, fixations are quantized on a 5×5 grid, and the re-

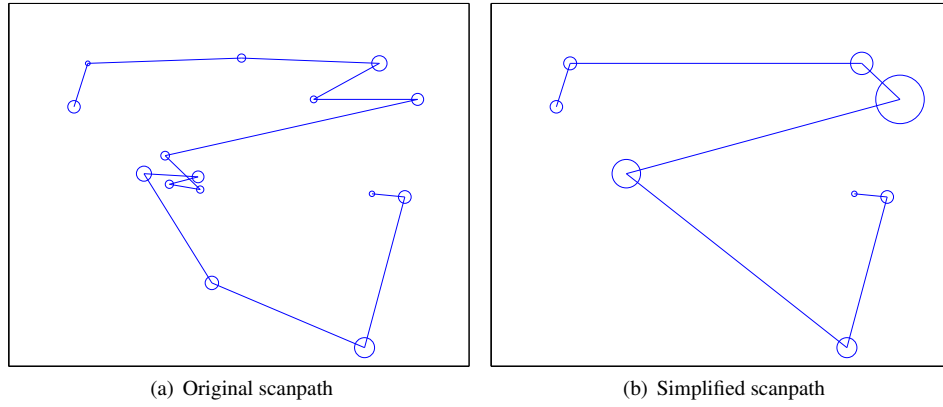


Figure 3: Example of scanpath simplification. Notice how saccade vectors with similar direction and/or short saccade vectors at similar positions are merged. Fixation durations associated with small, merged vectors are added to the beginning of a new, larger vector.

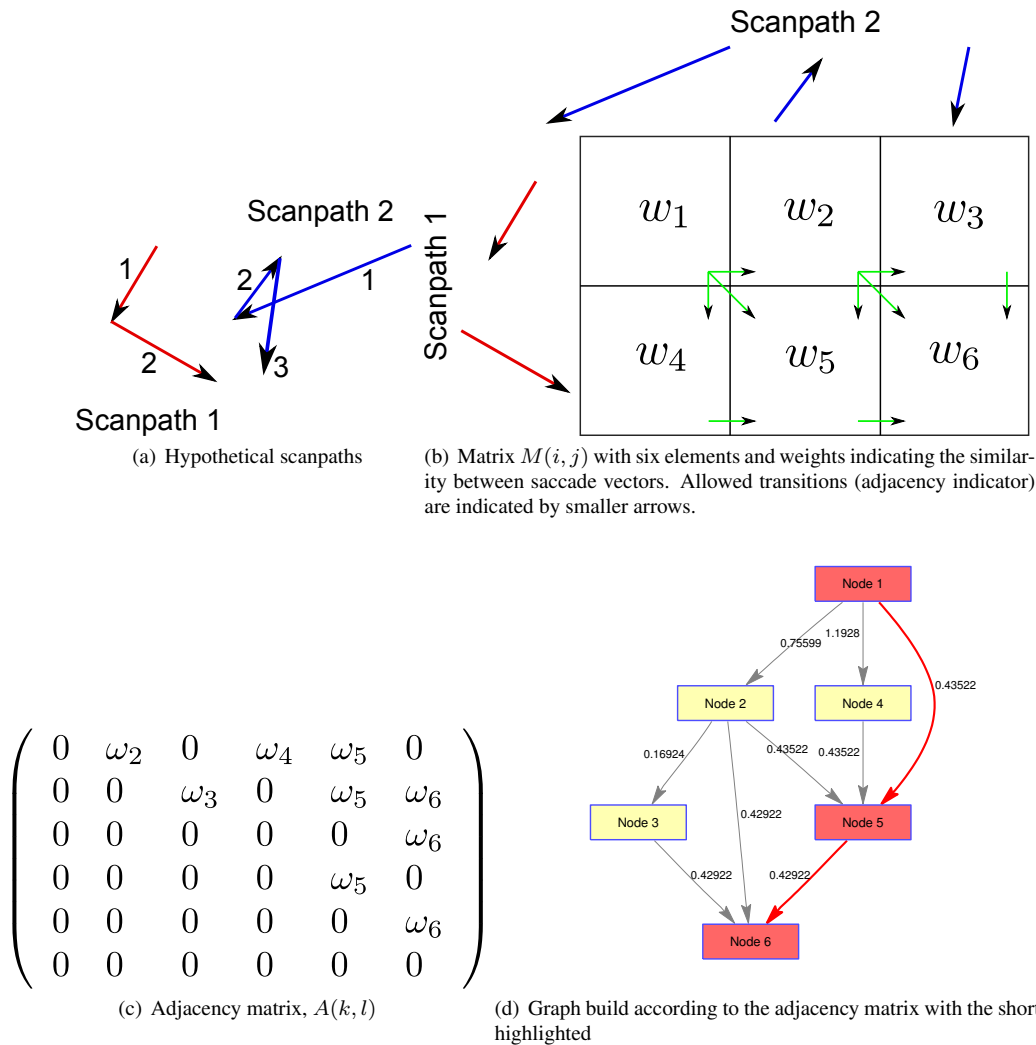


Figure 4: General principle of the proposed method for scanpath comparison. (a) The scanpaths. (b) A matrix where adjacency is indicated by small arrows. Each matrix element is a node i , which contains a value ω_i indicating the similarity between two saccade vectors. To align two scanpaths, Dijkstra algorithm is used to find the minimum cost path from node $i = 1$ to node $i = 6$. As a result, all saccade vectors in one scanpath are matched with saccade vectors in the other scanpath. Aligned saccade vectors (and fixation) can then be compared across different features. The similarity metric in this example was saccade vector difference (i.e., ‘shape’).

sulting string of letters is 'compressed' [Brandt and Stark 1997]. The attention maps (AM) are built from 2D Gaussian functions

$$G(x, y) = \exp\left(\frac{(x - x_0)^2}{2\sigma} + \frac{(y - y_0)^2}{2\sigma}\right) \quad (1)$$

where (x_0, y_0) represents one fixation position in the scanpath. $\sigma = 0.10W$ was set to span 10% of the screen width (W). To generate an attention map, Gaussian functions from all fixations in a scanpath are superimposed. Thus, the attention maps do not take fixation duration into account. A minimum fixation duration of 40 ms was used in the constructed scanpaths. Furthermore, T_{amp} and T_ϕ were set to $0.2W$ and $\pi/4$ deg, respectively. The exact values do not critically effect the results in this paper, but should be selected with care when applied to real eye-tracking data.

Table 1 shows the results. For all three measures the similarity values are between 0 and 1. However, a high value in the correlation coefficient between attention maps corresponds to a high similarity between scanpaths, whereas for the remaining two measures a low value corresponds to a high similarity between scanpaths.

Example 1 is composed of two identical scanpaths, an observation that all three measures robustly confirm. However, scanpaths may have a similar location, but occur in a *reversed order* (example 2). Since only one letter is shared between letter strings in this case, a string-editing measure finds such scanpaths to be very different (here: 0.80). In contrast, attention-map-based measures assume such scanpaths to be identical (1.00). The new measure provides a more detailed picture of the (dis)similarity of the two scanpaths. We see the highest similarity for saccade amplitude (0.16), whereas the similarities for shape (0.31), position (0.34), direction (0.26), and duration (0.43) – all compared in reversed order – are lower.

One application that causes severe problems for current spatial similarity measures is the comparison of visual/mental imagery scanpaths (example 3). Such scanpaths may be very similar in their shape by preserving the overall pattern. At the same time, they may introduce some variability and distortion and hence be very different on a coordinate level [Brandt and Stark 1997; Laeng and Teodorescu 2002; Zangemeister and Liman 2007]. One problem that may occur in comparing such scanpaths is *offset in the spatial dimension*. In that, the scanpath that occurred during inspecting the stimulus may be centered, whereas the scanpath that occurred during mental imagery may be shifted to one side of the screen. This is due to the fact that the eyes have no spatial reference point while looking at a blank screen. Although both scanpaths have identical shapes, current spatial similarity measures estimate them to be quite different. Results show that both the attention-map-based measure (-0.23) and the Levenshtein distance (1.00) estimate the two scanpaths as very dissimilar. Again, the new measure provides a more multifaceted description: the shape, the length, and the directions of both scanpaths are very similar (better than 0.01). What does differ in both scanpaths are the positions of their fixations. Now, the researcher must decide whether this difference is meaningful for his research question (e.g., comparing two persons' scanpaths on a picture with many closely located details) or not (e.g., in a mental imagery study). Moreover, both scanpaths differ in terms of fixation duration. Again, the research question should decide whether this observation is relevant (cf. example 8).

Another problem that might occur in calculating scanpath similarity is *temporal offset* (cf. example 4). Temporal offset stands for the fact that a part of one scanpath is very similar to one within another scanpath, however later in time. For instance, by inspecting static pictures one person might orient her/himself longer than another, but starting with the same inspection strategy after a longer orientation phase. Similar problems occur in comparing mental/visual

imagery scanpaths to real object inspection scanpaths. Before an observer starts to describe the already seen stimulus, s/he might need some time to organize her/his thoughts before s/he starts the description and thus the actual scanpath. The organizing phase is hence off-topic to the actual scanpath. Such an off-topic beginning does not occur in the inspection of the stimulus, because it can be inspected immediately. Thus, the mental/visual imagery scanpath is shifted in time compared to the real stimulus inspection scanpath. A comparable pattern occurs for differences in viewing time or fixation number, which easily can occur in scanpaths. Both result in a *different scanpath length*. Therefore, the occurrence of different scanpath lengths is almost inevitable, and a very important property of a scanpath similarity measure is to deal with scanpaths of different lengths [Josephson and Holmes 2006; Zangemeister and Liman 2007]. Results show that the attention-map-based measure assumes the two scanpaths to be relatively similar (0.62). In contrast, the Levenshtein distance confound similarity completely with string length. That is, a long scanpath is per default dissimilar to a short one, even though the shorter is a substring of the longer. Moreover, the string-edit measure has no possibility to deal with a temporal offset of even one character but, as example four illustrates, judges such strings as being quite different (0.40). The new measure provides a broader perspective on the type of (dis)similarity: the saccade lengths of both scanpaths are highly similar (0.01) whereas their shapes (0.11), the positions of their fixations (0.16), and the directions of their saccades (0.12) differ more. For the high difference in fixation durations (0.32) see example 8. As already described above, the researcher has to decide which difference is meaningful for his specific research question.

Section 2 discussed the problem of setting AOIs, in particular as grids, in detail. Example 5 provides an extreme case, where two scanpaths look almost identical, but reside in very different AOIs, because the fixations are located close to AOI boarders. While the attention-map based measure estimates the scanpaths as being very similar (0.90), the Levenshtein distance estimates them to be totally different (1.00). Thus, the Levenshtein measure is completely biased by the definition of AOIs. The new vector-based measure instead estimates the scanpaths as very similar on the following dimensions: shape (0.05), position (0.04), length (0.04), direction (0.01). The difference that the new measure detects is in the fixation durations (0.57, for a discussion on differences in durations see example 8). Another bias that may occur is an *exaggeration or a miniaturization* of one scanpath compared to another (example 6). In that, the imagery scanpath may be expanded compared to the original stimulus inspection scanpath, or vice versa. Moreover, a mental/visual imagery scanpath may be more locally restricted compared to global viewing of real pictures [Zangemeister and Liman 2007]. Both common measures estimate the scanpaths as very dissimilar: an attention-map correlation coefficient of $r = -0.36$ and a Levenshtein distance of 1.00. While the new measure detects a dissimilarity, it is far smaller for position (0.28), length (0.32), duration (0.32), and shape (0.32). Most notably, the measure detects a perfect similarity in direction of the saccades (0.01).

Example 7 illustrates two scanpaths where long (global) saccades are used to get to a new region, whereafter this region is inspected in more detail using smaller (local) saccades. Intuitively, such typical scanpaths are quite similar. However, the two scanpaths are estimated as very dissimilar by the attention-map-based measure (-0.05) and the Levenshtein measure (0.76). The new measure better captures this intuitive similarity in all dimensions.

Scanpaths may differ in terms of the duration of certain fixations. While solving a mathematical problem, for instance, consider a person that makes two short fixations on a formula, followed by two longer fixations on its corresponding graph. Then another person makes four fixations in the same order, but instead directs the longer

Metric	① Equal	② Reversed	③ Spatial offset	④ Temporal offset
$r(AM_1, AM_2)$	1.00	1.00	-0.23	0.62
Levenshtein distance	0.00	0.80	1.00	0.40
Proposed:				
Shape	0.00	0.31	0.01	0.11
Position	0.00	0.34	0.15	0.16
Length	0.00	0.16	0.00	0.01
Direction	0.00	0.26	0.00	0.12
Duration	0.00	0.43	0.32	0.32
Metric	⑤ AOI border problem	⑥ Scaled	⑦ Local/Global	⑧ Duration
$r(AM_1, AM_2)$	0.90	-0.36	-0.05	1.00
Levenshtein distance	1.00	1.00	0.76	0
Proposed:				
Shape	0.05	0.32	0.29	0.00
Position	0.04	0.28	0.39	0.00
Length	0.04	0.32	0.23	0.00
Direction	0.01	0.01	0.33	0.00
Duration	0.57	0.32	0.33	0.88

Table 1: Similarity between fictitious scanpaths using the 1) correlation between attention maps ($r \in [-1, 1]$) 2) Levenshtein distance and 3) the method proposed in this paper. For the two latter methods, ‘0’ indicates a high similarity. The proposed method outputs differences with respect to certain aspects of the scanpaths: Shape (difference between saccade vectors), (fixation) position, saccade length, saccade direction, and fixation duration. All differences are computed on all pairs of aligned saccades and fixations, and only the average is reported. In all examples, alignment of scanpaths is done according shape. All scanpaths except the reversed one in example 2 start in the upper left corner.

fixations toward the formula. The first person is rather studying the graph, whereas the second person is rather studying the formula. Such a situation is presented in example 8. The attention-map-based measure considers both scanpaths to be exactly similar (1.00)³. A string-editing measure may be even more biased; not only does it consider the two scanpaths completely similar (0), but it would still do so if the difference in fixation time increased even further. The new measure sees the similarity of both scanpaths in shape, position, length, and direction (all 0). However, it detects the high difference in fixation durations of both scanpaths (0.88), which may be very important in certain scenarios.

6 Conclusions and Future Work

This paper presents an alternative to common measures that assess the similarity of two scanpaths, like the Levenshtein distance or correlating attention-maps. The new measure uses a representation of scanpaths as mathematical vectors. This type of representation preserves important features of the real scanpath: the overall shape, the direction and the amplitude of its saccades, the position and the duration of its fixations. In a second step the measure simplifies the scanpath representation by amplitude- and direction-based clustering. The simplified scanpath representation is compared to another one with regard to their shape. This is done by calculating vector differences within a recurrence matrix. The found path of least resistance within this matrix is the basis for further calculations. On this path, the position and the duration of the fixations of both scanpaths are compared. Moreover, the length and the direction of the saccades of both scanpaths are compared. As a result the measure delivers five normalized values, one for each scanpath dimension.

This new measure was compared to the currently most common measures: the Levenshtein distance and correlation of attention maps. In general, the presented examples demonstrate that the new measure provides more detailed information on the type of (dis)similarity of two scanpaths according to the five dimensions. In particular, the new measure provides more valid results for specific cases compared to the other measures. Since the new measure does not require to define AOIs, it does not get biased by fixations falling into AOI border regions as it is the case for the Levenshtein distance. Moreover, the new measure is capable of describing the (dis)similarity of spatially shifted or scaled scanpaths. In contrast, the other two measures simply estimate such cases as dissimilar. Furthermore, the new measure deals also with temporal issues. Not only does it take into account fixation durations, but it also deals successfully with shifts in time and different scanpath lengths. Note that the difference in scanpath length is a severe problem, in particular for string-editing. Some measures are often not applicable at all, like the Hamming distance, or “unique assignment” variants of linear distance metrics [Henderson et al. 2007].

It must be mentioned that neither the Levenshtein distance measure nor the correlation of attention maps are the newest or the best measures available. We nevertheless chose those two, because they are currently the most popular ones. Still, several measures exist that have solved some of the addressed problems. One attempt to face the problem of how to take temporal shift into account, is the ClustalG method [Wilson et al. 1999]. This algorithm only makes binary comparison between AOIs: A0 is just as far away from A1 than it is to F6, so $A0 = A1$ is just as false as $A0 = F6$. Thereby, the scanpaths are temporally aligned based on their most similar part. Although this measure faces successfully the problem of temporal shift, it still has two main disadvantages. First, it does not take different scanpath lengths into account, which is a severe drawback,

³It should be noted, however, that certain attention map implementations take fixation duration into account.

since most scanpaths are of different lengths. Second, it carries all the problems that are associated with using AOI-based string input as described above. Another measure uses quantized saccade direction and amplitude rather than stimulus space. Gbadamosi [2000] as well as Zangemeister and Liman [2007] developed a version of the string representation for scanpaths, in which each saccade is represented with a pair of (hexadecimal) numbers where the first number is for direction and the second for amplitude. Such a representation of scanpaths does not require a segmentation of space into AOIs; in fact, it completely abandons position information, and instead focusses on representing *the overall shape*. It is much more precise than the original AOI string representation, more true to our subjective perception of scanpath shape, but still based on segmentation and therefore sensitive to quantization effects.

Still, the new measure proposed in this paper has at least two drawbacks. First, this measure compares only two scanpaths. Sometimes, the overall aim is to compare whole groups of participants with each other. One solution for performing the according statistical comparison is delivered by Feusner and Lukoff [2008]. They present a method to combine many pairwise comparisons in a meaningful and computational efficient way. Second, the new measure presumes fixations and saccades to occur, however, smooth pursuit is not handled. Indeed, the representation of scanpaths with geometrical vectors allows in principle to handle smooth pursuit. In such case, smooth pursuit would have to be represented as a series of short vectors, which must not be clustered into one long vector.

Despite these drawbacks, the proposed measure is capable of dealing with important issues concerning scanpath similarity. However, open questions for future research still remain. One question is about simplifying the scanpath representation. In this paper, thresholds for amplitude- and direction-based clustering are set rather arbitrarily. Of course, a change in these would influence the final output. It is difficult to decide from a theoretical point of view how to decide the thresholds. Thus, further research that quantifies the influence of different thresholds is required.

Moreover, the comparison dimensions need to be investigated in depth. Some dimensions are not important for certain tasks and thus should play a minor role in the overall judgement. For instance, the results from the reversed order example show a relative low similarity between the two scanpaths. However, for certain studies this might not be important. Thus, in further it should be possible to match two scanpaths in the other direction for specific research questions. On the other hand, other dimensions should sometimes be handled more strictly. Temporal alignment is not always appropriate. For instance, in comparing scanpaths on dynamic material, like movies. In that, temporal misalignment is a meaningful difference between two scanpaths. Imagine a dynamic scene, where several persons are walking across a street. If one participant looks at person A and another participant looks at person B, both will produce scanpaths that are quite similar, but temporally shifted. This temporal shift is, however, meaningful saying that both participants look at different objects. In this case temporal alignment would lead to almost identical scanpaths, although they are from a semantical point of view diverse. Thus, the temporal shift is not a bias in this case. In future, the scanpath similarity measure should allow for weighting of sub-measures influencing the similarity index.

Irrespective of the content of the single dimensions, they need to be investigated further as a total. The measure presented so far delivers five separate values for five dimensions of the scanpath that can be used independently to compare different aspects of scanpath similarity. Alternatively, the five dimensions could be added up to give one precise value for scanpath similarity. However, in this case further research is needed on how those dimensions should be weighted. Most probably, different kinds of weightings are neces-

sary for different applications. Still, the five dimensions in separate allow for the application of multilevel analyses. Another important issue is the question, whether the dimensions are independent and if not, how they relate to each other. This should be investigated in further research by calculating the measure on real data. Thereupon, a factorial analysis would deliver the dependencies of the dimensions. Again, those dependencies might differ between task and/or stimuli types. Consequently, in future actual experiments must be conducted where the differences in performance between the new measure and established metrics are presented based on real data of a number of subjects on several stimuli types.

References

- BOCCIGNONE, G., CAGGIANO, V., MARCELLI, A., NAPOLENTANO, P., AND DI FIORE, G. 2005. An architectural model for combining spatial-based and object-based information for attentive video analysis. In *CAMP '05: Proceedings of the Seventh International Workshop on Computer Architecture for Machine Perception*, 116–121.
- BRANDT, S. A., AND STARK, L. W. 1997. Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience* 9, 1, 27–38.
- CERF, M., HAREL, J., HUTH, A., EINHAEUSER, W., AND KOCH, C. 2009. Decoding what people see from where they look: predicting visual stimuli from scanpaths. *Lecture Notes in Artificial Intelligence* 5395, 15–26.
- DEMPERE-MARCO, L., HU, X.-P., ELLIS, S. M., HANSELL, D. M., AND YANG, G.-Z. 2006. Analysis of visual search patterns with emd metric in normalized anatomical space. *IEEE Trans. Medical Imaging* 25, 8 (Aug.), 1011–1021.
- DIJKSTRA, E. 1959. A note on two problems in connexion with graphs. *Numerische mathematik* 1, 1, 269–271.
- EHMKE, C., AND WILSON, C. 2007. Identifying web usability issues from eye-tracking data. In *People and Computers XXI – HCI... but not as we know it: Proceedings of HCI 2007*, BCS Press, London, P. of HCI-07, Ed.
- FEUSNER, M., AND LUKOFF, B. 2008. Testing for statistically significant differences between groups of scan patterns. In *ETRA '08: Proceedings of the 2008 symposium on Eye tracking research applications*, ACM, New York, NY, USA, 43–46.
- GBADAMOSI, J. 2000. *Visual Imagery und Mikro-Sakkaden bei Hemianopsie-Patienten*. PhD thesis, Fachbereich Medizin der Universität Hamburg.
- GRONER, R., WALDER, F., AND GRONER, M. 1984. Looking at faces: Local and global aspects of scanpaths. In *Theoretical and Applied Aspects of Eye Movement Research*, Elsevier Science Publishers B.V., North-Holland, A. G. Gale and F. Johnson, Eds., 523–533.
- HENDERSON, J. M., BROCKMOLE, J. R., AND MACK, M. S. C. M. L. 2007. Visual saliency does not account for eye movements during search in real-world scenes. In *Eye Movement Research: Insights into Mind and Brain*, Elsevier: Oxford, R. van Gompel R., M. Fischer, W. Murray, and R. W. Hill, Eds., 537–562.
- ITTI, L. 2005. Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition* 12, 6, 1093–1123.
- JOSEPHSON, S., AND HOLMES, M. E. 2002. Visual attention to repeated internet images: testing the scanpath theory on the world wide web. In *ETRA '02: Proceedings of the 2002 symposium on Eye tracking research & applications*, ACM, New York, NY, USA, 43–49.
- JOSEPHSON, S., AND HOLMES, M. E. 2006. Clutter or content? how on-screen enhancements affect how tv viewers scan and what they learn. In *ETRA '06: Proceedings of the 2006 symposium on Eye tracking research & applications*, ACM, New York, NY, USA, 155–162.
- LAENG, B., AND TEODORESCU, D.-S. 2002. Eye scanpaths during visual imagery reenact those of perception of the same visual scene. *Cognitive Science: A Multidisciplinary Journal* 26, 2 (March), 207–231.
- LEVENSHTAIN, V. 1966. Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics – Doklady* 10, 707–710.
- MANNAN, S., RUDDOCK, K., AND WOODING, D. 1996. The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision* 10, 3, 165–188.
- OUERHANI, N., VON WARTBURG, R., HÜGLI, H., AND MÜRI, R. M. 2003. Empirical validation of the saliency-based model of visual attention. *Electronic Letters on Computer Vision and Image Analysis* 3, 1, 13–24.
- RAJASHEKAR, U., CORMACK, L., AND BOVIK, A. 2004. Point of gaze analysis reveals visual search strategies. In *Human Vision and Electronic Imaging*, vol. 5292, Human Vision and Electronic Imaging, 296–306.
- RAJASHEKAR, U., VAN DER LINDE, I., BOVIK, A., AND CORMACK, L. 2008. Gaffe: A gaze-attentive fixation finding engine. *IEEE Trans. Image Processing* 17, 4, 564–573.
- SANTELLA, A., AND DECARLO, D. 2004. Robust clustering of eye movement recordings for quantification of visual interest. In *Proceedings of the 2004 symposium on Eye tracking research & applications*, vol. 22, Citeseer, 27–34.
- TATLER, B., BADDELEY, R., AND GILCHRIST, I. 2005. Visual correlates of fixation selection: effects of scale and time. *Vision Research* 45, 5, 643 – 659.
- TZANIDOU, E., MINOCHA, S., AND PETRE, M. 2005. Applying eye tracking for usability evaluations of e-commerce sites. In *workshop on 'Commercial Uses of Eye tracking' held at the 19th British HCI Group Annual Conference, Napier University, Edinburgh*.
- UNDERWOOD, G., HUMPHREY, K., AND FOULSHAM, T. 2008. *HCI and Usability for Education and Work*, vol. Volume 5298/2008. Springer Berlin / Heidelberg, ch. Knowledge-Based Patterns of Remembering: Eye Movement Scanpaths Reflect Domain Experience, 125–144.
- WILSON, C., HARVEY, A., AND THOMPSON, J. 1999. Clustalg: Software for analysis of activities and sequential events. In *IATUR Conference Proceedings*.
- WOODING, D. 2002. Fixation maps: quantifying eye-movement traces. In *Eye Tracking Research & Applications*, ACM Press, New York, New Orleans, LA, 31–36.
- ZANGEMEISTER, W. H., AND LIMAN, T. 2007. Foveal versus parafoveal scanpaths of visual imagery in virtual hemianopic subjects. *Computers in Biology and Medicine* 37, 7, 975 – 982. Vision and Movement in Man and Machines.