Heliyon 7 (2021) e07616

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Assessment of eye-tracking scanpath outliers using fractal geometry

Robert Ahadizad Newport *, Carlo Russo, Abdulla Al Suman, Antonio Di Ieva

Computational NeuroSurgery (CNS) Lab, Department of Clinical Medicine - Faculty of Medicine, Health and Human Sciences, Macquarie University, Sydney, NSW, Australia

ARTICLE INFO

Keywords:
Higuchi fractal dimension
Visual scanpath
Hilbert curve
Outlier
Computational neuroscience

ABSTRACT

Outlier scanpaths identification is a crucial preliminary step in designing visual software, digital media analysis, radiology training and clustering participants in eye-tracking experiments. However, the task is challenging due to the visual irregularity of the scanpath shapes and the difficulty in dimensionality reduction due to geometric complexity. Conventional approaches have used heat maps to exclude scanpaths that lack a similarity pattern. However, the typically-used packages, such as ScanMatch and MultiMatch often generate discordant results when outlier identification is done empirically. This paper introduces a novel outlier evaluation approach by integrating the fractal dimension (FD), capturing the geometrical complexity of patterns, as an additional parameter with the heat map. This additional parameter is used to evaluate the degree of influence of a scanpath within a dataset. More specifically, the 2D Cartesian coordinates of a scanpath are fitted to a space filling 1D fractal curve to characterise its temporal FD. The FDs of the scanpaths are then compared to match their geometric complexity to one another. The findings indicate that the FD can be a beneficial additional parameter when evaluating the candidacy of poorly matching scanpaths as outliers and performs better at identifying unusual scanpaths than using other methods, including scanpath matching, Jaccard, or bounding box methods alone.

1. Introduction

Following the seminal paper by Noton and Stark [1] that demonstrated how eye movements, known as scanpath sequences, may be replicated by the same viewer, a large number of papers have tackled the challenge of clustering scanpaths (Kumar et al. [2], Goldberg and Helfman [3], Fahimi and Bruce [4]). Scanpaths are geometrically complex, as seen in Fig. 1, making categorisation challenging when no obvious visual defects exist. Comparing the way different people look at something addresses the problem of perception deficiency, whether through lack of experience, or unintended obfuscation of the stimulus. Indeed, this can have relevant applications in teaching, diagnosis, and improvements in visual communications and marketing. In the last two years, neuroscience research has invested greater interest in clustering expert and novice scanpath data in a myriad of different ways (Castner et al. [5], Brunyé et al. [6], Król and Król [7]).

Naturally, when comparing a scanpath's response to a defined stimulus, outliers should be defined, although a formal definition of outlier scanpaths does not exist, even as researchers computationally strive to cluster and compare entire scanpath data sets (Burch et al. [8], Jolliffe [9]). If an outlier definition is restricted to data that falls outside the standard deviation of the mean, as it traditionally is defined to be, by excluding information, the results can be biased and the whole research is at risk of missing potentially relevant information, limiting the opportunity to yield important discoveries. On the other hand, too much inclusively could decrease statistical significance. So a delicate balance exists where a careful approach to identify irrelevant data points can greatly improve the identification of clusters.

At first glance, the myriad of points that make up a scanpath appear to be similar to nodes in a network or graph, which may include relationships between people in a social network, cat brain cortical connectivity, or airport networks. Indeed, research by Li et al. [10] has

* Corresponding author.
E-mail address: robert.newport@hdr.mq.edu.au (R.A. Newport).

https://doi.org/10.1016/j.heliyon.2021.e07616
Received 8 January 2021; Received in revised form 13 April 2021; Accepted 14 July 2021

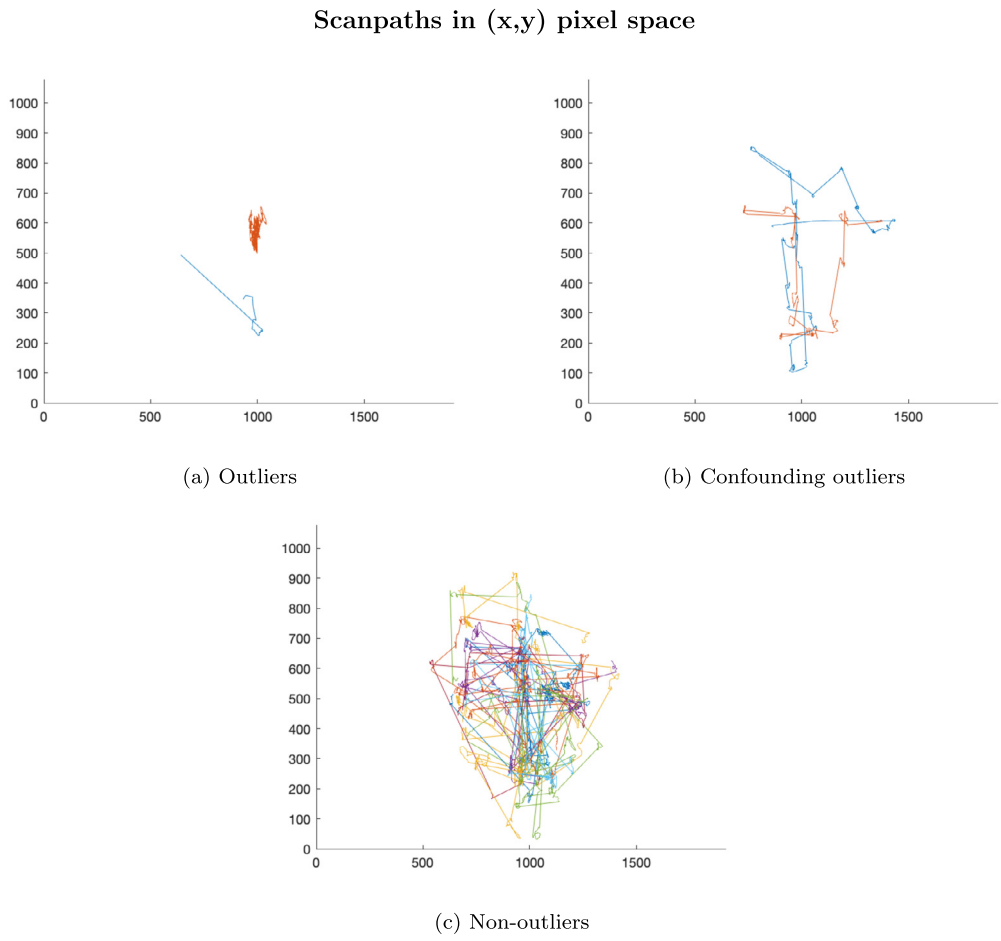


Fig. 1. (a) Two different stacked visually recognisable outliers P44 in blue and P11 in red, (b) confounding stacked outliers P20 in blue and P46 red, (c) and all other stacked visual non-outliers.

shown that these large network representations can be analysed systematically using clustering coefficients. However, scanpaths are a sequence of points along a line, whereas a graph or network exploits "relationships between constituents" of lines coming out of vertices [10]. This would mean a scanpath would always have a minimal clustering coefficient because it lacks a triangular pattern which is used to assess the connectivity in a vertex's neighbourhood.

These unique geometric characteristics are exploited in shape comparisons to define outlier scanpaths by Burch et al. [8] Jaccard and bounding box methods. Other researchers empirically find outliers using scanpath comparison heatmaps by excluding ones that match poorly with others in their group as seen in Davies et al. [11]. However, simply being a poor match does not make a strong case for exclusion, especially if the scanpath may be a small cluster representing an underpopulated label. Leaving out unusual yet influential scanpaths that have a large effect is defined by Jolliffe [9] as a data set perturbation. These odd scanpaths may have valid saccades and fixation points within the stimulus boundary, but lack any of the features that a majority of other scanpaths share, as would be the case if the participant was distracted or inattentive. In some cases, a judgement call must be made: does this participant's scanpath represent an artefact of the experimental process, whether through inattention, lack of expertise, or a software glitch? Or should this scanpath represent an edge case, potentially influencing boundaries during clustering? Indeed, a hallmark of a robust analysis is to remove data artefacts while keeping data that are unusual but still significant.

This paper proposes that these ambiguities can be mitigated by a more detailed geometric complexity metric achieved via the fractal dimension. First, by fitting the 2D scanpath into a 1D fractal space filling curve, then by measuring the sequence of fixations using the Higuchi fractal dimension (HFD). The outliers from HFD analysis are compared to the non-matching outliers from heatmaps analysis with scanpath comparison packages to determine if there is a connection. This research shows that the outliers using HFD correlate well with both outliers found empirically through heatmaps using scanpath matching, and outliers found through bounding box and Jaccard methods.

1.1. Bounding box and Jaccard method

Davies et al. [11] defines outliers as scanpaths that are markedly similar or different to one another in a 2017 paper analysing the visual behaviour of clinicians in electrocardiology. In that study, several methods were used to compare scanpaths, with the final determination of outlier status done empirically using shades of blue to red in plots to illustrate how well the participants compare to one another. The appearance of fixations outside the stimulus boundary as empirical indicators of outlier status leads to the question of whether a bounding box containing only the minimal limits of a scanpath could be used as a simplification for comparison and measurement. This question prompts the need for an additional parameter when accessing outlier status. Burch et al. [8] proposed a method of defining outliers using a bounding box and Jaccard algorithm as a means of finding scanpaths that do not follow the common behavioural characteristics within a cluster. Burch et al. [8] described the bounding box as a normalised value covering a scanpath's minimal area. The Jaccard method uses the intersecting size of common elements divided by the union size of unique elements in two scanpaths.

1.2. ScanMatch and MultiMatch

The ScanMatch and MultiMatch are prevalent methods used within the last decade for scanpath comparison methods. An examination of how these methods compare and contrast should be attempted in order to understand the role that these methods play in making matches and excluding outliers.

The researchers Cristino et al. [12] for ScanMatch used a method developed by Needleman and Wunsch [13] for matching sequences. ScanMatch comparison is done through quantised spatial and temporal binning of (x, y) coordinates into string representations of time, order, and fixation location. A substitution matrix is used to maximize the similarity score by providing a connection between any useful dimension, including semantic space or gridded distance, with the coded letters. A penalty gap is used to minimize the score as compared scanpaths differ over wider string pairs in the substitution matrix. The issue of fixation durations lost in single letter substitutions was addressed by temporal binning, where a string substituted scanpath DEF with increasing fixation durations could be represented as $DEEFFF$. However, the issue of spatial quantisation remains due to division of the observation boundary into squares representing a string. When a fixation falls into this area, the proximity from the centre to the boundary are entirely replaced by the string, quantising edges to the centre.

MultiMatch by Dewhurst et al. [14] attempted to mitigate this through its vector based approach. It quantises a scanpath through iterative agglomerated sequential fixations within a threshold of distance. MultiMatch uses geometric vectors to compare scanpaths aligned sequentially over many dimensions. These include: *vector similarity*, where aligned saccade differences are measured, normalised, and averaged to determine geometric shape similarity; *length similarity*, where only the amplitude of aligned saccades are measured, normalised, and averaged without regard for duration, direction, or location; *direction similarity*, where the angle differences between aligned saccades are normalised by π and averaged without regard for location or amplitude; *positional similarity*, where Euclidean distances are measured, normalised, and averaged between aligned fixations, taking into account direction and amplitude; and *durational similarity*, where aligned fixation durations are normalised, averaged and measured without regard for amplitude and position. The vector nature of the scanpath representation allows for more refined simplification compared to gridded methods.

Both ScanMatch and MultiMatch transform their Cartesian coordinates into a lower dimensional string format conducive to comparison. Fig. 3 illustrates how a grid, similar to those used in map books, is used to quantise fixation points into boxes represented by letters on both the x and y axis. These letters are concatenated to represent a singular one dimensional coordinate for the quantised fixation's location. A major limitation in MultiMatch is that it only accounts duration in one (the duration comparison score) out of five of its comparison algorithms. ScanMatch attempts to capture fixation duration coarsely by approximating durations in its string representation with letter repetition. However, this further increases the degree of quantisation in the model through both positional and temporal generalisation. The many different comparison parameters used in both ScanMatch and MultiMatch makes consistent outlier identification difficult because scanpaths displaying outlier characteristics using one approach may masks those outlier attributes in another.

1.3. HFD and Hilbert method

Burn and Mandelbrot [15] proposed the problem of measuring the coastline of an island where, as the detail of the measurement moves closer to the nooks and crannies of an increasingly detailed shoreline, so too does the perimeter of the measured sand, rocks, and other details. Alternatively, in an effort to measure the coastline, a proposal to define its geometric complexity was introduced using a similar concept of infinite length found in fractals. As the scale of a coastline is increased, so too does its length, which in nature become impractical to measure as scales approach smaller sizes e.g. in a grain of sand along the perimeter of the coastline. A calculation for measuring the coastline was proposed using a grid of boxes, where intersections between the boxes and coastline are counted at one scale, and compared to the same intersections when both boxes and coastline are scaled up proportionately. This became known as the box counting dimension, or D_B Minkowski–Bouligand Dimension. Box counting can be done to determine a fractal dimension measurement of the Cartesian geometry encompassing a complete scanpath. However, this measurement would ignore the temporal aspect of a scanpath if box counting is applied just on the scanpath's 2D shape. Therefore, an FD measurement incorporating a temporal dimension should be implemented. Fractal analysis is well suited to physiological timeseries, e.g., electrocardiograms and electroencephalograms, where the correlation function, standard deviation, mean value, and other statistical attributes will change over time. This is particularly useful in brain diseases, e.g. the metabolic encephalopathy, where renal and hepatic disease can significantly affect higher brain function (Faigle et al. [16]), as well as in brain tumours and other central nervous system pathologies (Di Ieva et al. [17]).

The Box-Counting FD calculation method described above to measure a coastline cannot be used to measure the interdependent temporal sequences found in dynamic, non-stationary sequences like the heart rate in the cardiovascular system. To address this shortcoming, the HFD is a non-stationary fractal analysis method, and can be used in dynamic physiological systems (Di Ieva et al. [17]) by approximating the box counting value of a time series. Jacob et al. [18] proposed that the HFD's main strength is its direct calculation in the time domain. Indeed, as outlined by Klonowski [19], HFD is also more robust with noisy signals compared to linear methods such as correlation dimension, which needs thousands of points compared to HFD which needs as few as 100. Moreover, HFD was used as an additional feature set to increase the accuracy of a Support Vector Machine (SVM) classifier for the automatic diagnosis of encephalopathy based on EEG (Jacob et al. [18]).

Peano [20] wrote a seminal paper (in a time when fractal geometry was not characterised yet) on a way to continuously map a curve on a unit interval, that could pass through every point on a unit square, using a fractal shape. The range of these curves could fill entire n-dimensional hypercubes filling a Euclidean space without endpoints. This created an avenue for David Hilbert's discovery of a fractal curve known as "Hilbert curve" in 1891. Furthermore, normalisation of multidimensional data in a $1D$ space would lead to increased precision as coordinate space resolution increases. Thus, as it increases in size, its ratio of detail to scale remain constant.

2. Methods

This paper proposes an additional parallel step in tandem with traditional similarity heatmaps made with ScanMatch and MultiMatch, as illustrated in Fig. 2, where the geometric complexity of the scanpath is measured using the HFD. The goal of the HFD is as a measure of the scanpath's geometric complexity over a common window scale rather than it being forcefully fitted into a mathematical fractal framework. To facilitate the HFD measurement, a novel process for reducing scanpath dimensionality is used where 2D Cartesan points are measured as a 1D Hilbert curve distances, as seen in the Fig. 3. The HFD is calculated using each scanpath's Hilbert curve distances measured against their durations. A scatter plot of all HFDs is then constructed plotting a boundary marking two standard deviations from the mean. Scanpaths outside this boundary are deemed outliers. Parallel to this process, a similarity heatmap is constructed using ScanMatch and MultiMatch, as seen in the bottom middle box in Fig. 2. Finally, in the last step, outliers in both the HFD scatter plot and heatmap of matches are compared, with scanpaths that share outlier status in both approaches being considered as final outliers.

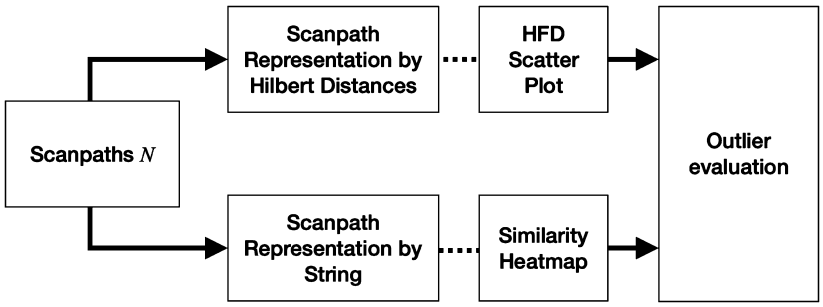


Fig. 2. Proposed outlier scanpaths detection framework. The Higuchi fractal dimension (HFD) scatter plot is generated using a novel Hilbert curve distance technique from the scanpaths. The similarity heatmap is generated from the scanpaths using traditional scanpath comparison packages like ScanMatch and MultiMatch. The final outlier scanpaths are then selected based on matching outlier status between the outliers in both approaches.

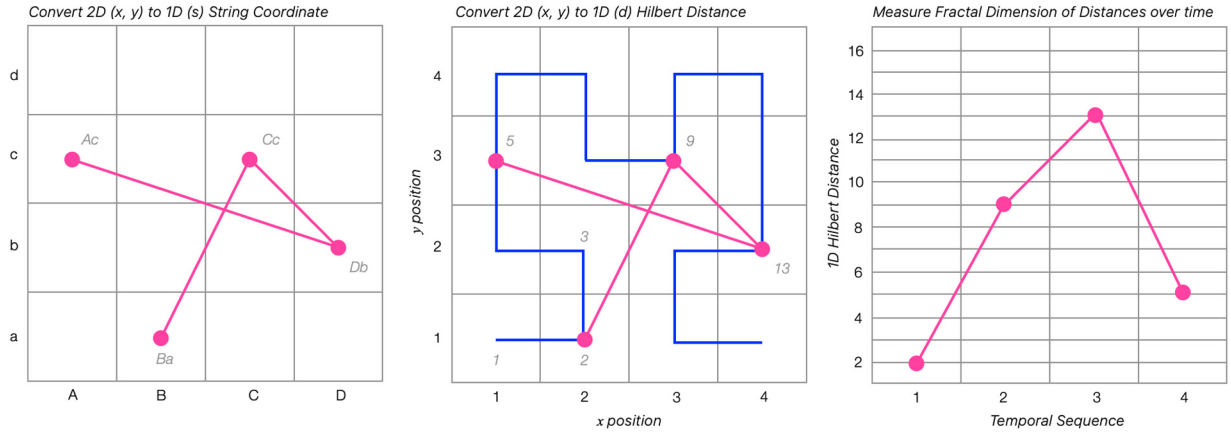


Fig. 3. A pink scanpath demonstrates how a theoretical collection of four fixations could be represented over Cartesian, String, and Hilbert curve distances. All three figures represent four theoretical fixation points over a 4 by 4 unit space. The left figure illustrates how ScanMatch and MultiMatch reduce the 2D Cartesian coordinates to Ba, Cc, Db, and Ac. The middle figure illustrates a novel method for reducing the same Cartesian coordinates to Hilbert distances 2, 9, 13, and 5. A blue Hilbert curve overlay demonstrates the Hilbert curve distance's path. The right figure shows the Hilbert curve distances to the left plotted against their temporal sequence.

2.1. Stimulus and participants

Experiments were conducted by a trained researcher and approved through The Faculty Ethics Subcommittees at Macquarie University in accordance with the Australian National Statement on Ethical Conduct in Human Research. The 53 participants, labelled P01 through to P53, included medical professionals with varying MRI interpretation experience. Exposure to stimulus is preceded and followed by exposure to noise. They are asked to examine an MRI of a normal, non-pathological scan, illustrated by Fig. 4. Participant gaze is captured by the eye-tracker EyeLink® 1000 Plus (SR Research, Ottawa, Ontario, Canada) and saved into a matrix consisting of the trial number, participant ID, eye fixations, saccades, and blinks, and a timestamp for each captured event. To reduce unwanted data, post processing reduces these values to three columns representing right eye (x, y) coordinates and a timestamp for when a fixation is made.

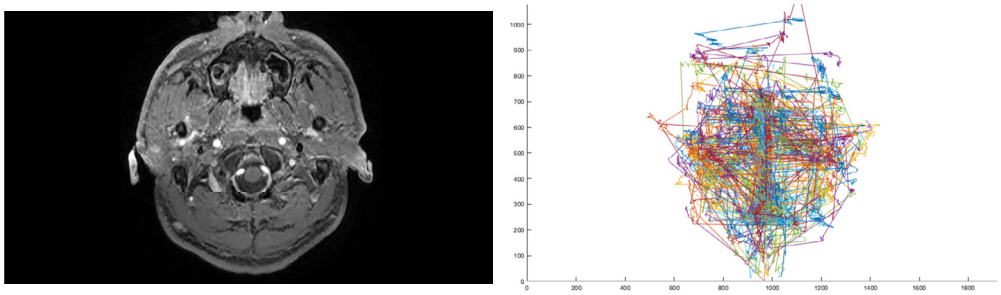
2.2. Scanpath representation by Hilbert distances

Fig. 3 illustrates how a simplified 1D Hilbert curve distance path would appear juxtaposed to a 2D Cartesian map. To simplify the demonstration, a theoretical stimulus with a measurement of 4 units by 4 units for a total maximum number of 16 locations is viewed by a participant creating a scanpath as shown with the pink line and points with four fixations at (2,4), (3,2), (4,3), and (1,2). Converting the 2D locations to 1D Hilbert curve distances result in four fixations at 2, 9, 13, and 5. Even though dimensionality is reduced from 2D to 1D, locality is preserved because the Hilbert path travels over all locations in the map, making 1D conversion back to 2D possible without loss. Furthermore, plotting 1D Hilbert curve distances against their duration at each fixation point can provide 2D data useful when calculating the HFD.

2.3. HFD and scatter plot

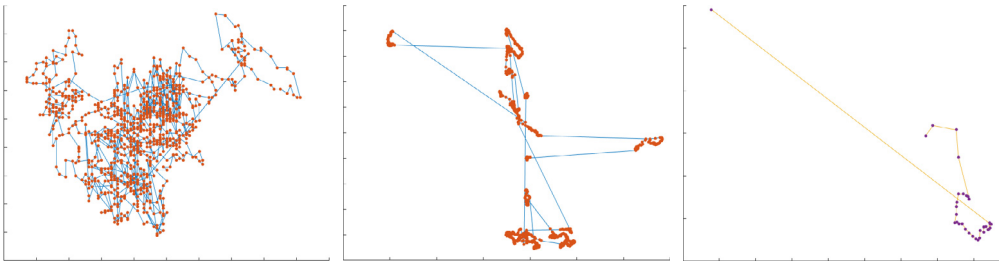
After converting scanpath locations to Hilbert curve distances, the FD is calculated by taking the distances versus timestamps as input. Higuchi [21] developed the HFD to measure FD in a time series with a range of 1 to 2. To understand this empirically, a normal curve with no perturbations has a lower limit HFD of 1 or 0% occupation of the 2D plane; at the upper limit, pure noise fills 100% of the signal volume with a HFD of 2 or 100% 2D space occupancy. To find the percentage of 2D space dominated by the curve, normalisation of the limits $(HFD-1)$ and multiplication by 100 can be performed. Fig. 5 illustrates three very different HFD scanpaths where the left is P11 with HFD 1.9596, the middle is P01 with HFD 1.5672, and the right is P44 with HFD 1.1515. Both P11 and P44 are on the extreme bounds of HFD complexity which typically exists within a range of 1 to 2. Signals measured directly in the time domain can be between these two limits, with their HFD used as a metric for correlation evaluations. We have used the HFD due to its suitability for non-stationary analysis. Though, other popular methods for calculating FD exist, including Yard Stick or Compass Method, Box Counting Method, Variation Method, and Structure Function Method [22]. However, these methods can introduce computational artefacts based on the conversion resolution when incorporating temporal attributes into its geometrical 2D shape, and hence not suitable in this context.

A scatter plot of HFD values from all participants can be used as a benchmark to compare geometric complexity between scanpaths to see



(a) Example of an MRI scan as stimulus (b) All scanpaths

Fig. 4. Normal MRI scan (a) and all participant scanpaths (b) viewing the scan.



(a) High complexity (b) Typical complexity (c) Low complexity

Fig. 5. Scanpaths illustrating the relationship between geometrical complexity and HFD values, (a) high complexity HFD = 1.9596, (b) average complexity HFD = 1.5672, and (c) low complexity HFD = 1.1515.

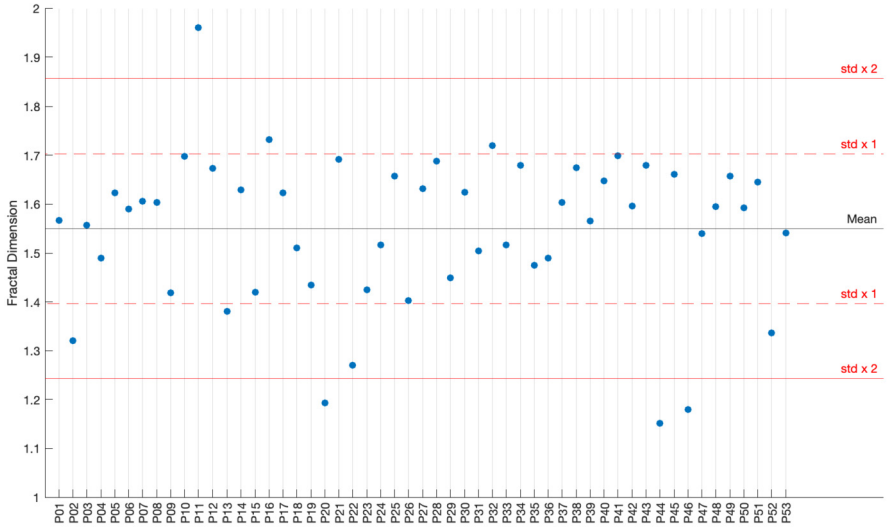


Fig. 6. Scatter plot with two standard deviations from the mean of all participants' Higuchi fractal dimensions (HFDs).

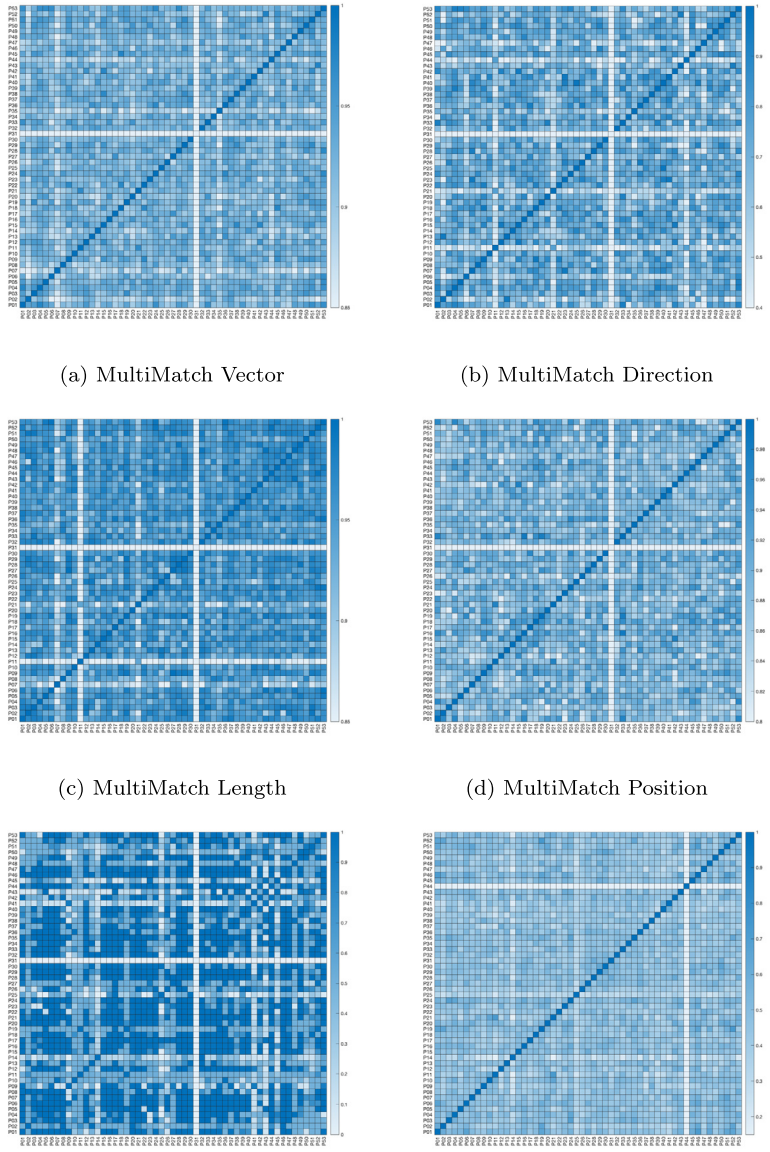
which fall outside one and two standard deviations away from the mean as illustrated in Fig. 6. Two standard deviations are a close approximation to a 95% confidence interval, with demarcations for each standard interval used as a means of establishing distance from the mean, used to empirically gauge geometric complexity when comparing participants' HFD scores to other scores.

2.4. Similarity heatmaps

A heatmap is constructed using a comparison of all participant scans versus all other scans to see which ones have few or no matches by exploiting ScanMatch and MultiMatch methods. The ScanMatch and MultiMatch can provide one and five types of heatmaps, respectively. This provides an empirical estimation of which scanpaths may be outliers due to their inability to cluster, as seen in the white lines representing no matches in Fig. 7. However, using this method alone may produce false positive outliers, where a scanpath that matches poorly with others may actually represent a valid cluster by itself, rather than a bad scanpath produced by artefacts in the experimentation process, especially when participant scanpath sample sizes are low. For this reason, the addition of the HFD parameter is used to strengthen or weaken the candidacy of poor scanpath matches as outliers.

2.5. Outlier evaluation

Table 1 outlines quantitative outlier candidates evaluated using ScanMatch, MultiMatch, bounding box, Jaccard, and HFD methods. Scanpath comparison outliers are divided into two categories, as seen in Fig. 7. The "Almost no matches" category is designated for all scanpaths that contain 1 or fewer matches other than itself. The "Few matches"



(a) MultiMatch Vector (b) MultiMatch Direction

(c) MultiMatch Length (d) MultiMatch Position

(e) MultiMatch Duration (f) ScanMatch

Fig. 7. Different heatmaps using ScanMatch and MultiMatch methods: (a) MultiMatch Vector displays poor matches for P31, and weak matches for P07, P35, and P44. (b) MultiMatch Direction displays poor matches for P31, and weak matches for P11, P21, and P44. (c) MultiMatch Length displays poor matches for P11 and P31, and weak matches for P07, P21. (d) MultiMatch Position displays poor matches for P31. (e) MultiMatch Duration displays poor matches for P31, and weak matches for P41, P43, and P45. (f) ScanMatch displays very poor matches for P44, weak matches for P25, and P14.

Table 1. Quantitative outliers identified by different methods.

	Almost no matches	Few matches
MultiMatch Vector	P31	P07, P35, P44
MultiMatch Direction	P31	P11, P21, P44
MultiMatch Length	P11, P31	P07, P21
MultiMatch Position	P31	-
MultiMatch Duration	P31	P41, P43, P45
ScanMatch	P44	P14, P25
	Geometrically different	
Bounding box	P11, P21, P44	
Jaccard	P11, P44	
Fractal Dimension	P11, P20, P44, P46	
	Correctly Identified	Confounding
Identified Outliers	P11, P44	P31

category is for scanpaths that appear visually paler to a noticeable degree compared to other samples in the heatmap. Similarly, bounding box and Jaccard method outliers are visually identified as being noticeably paler than other samples, as seen in Fig. 9. The HFD outliers are distinguished in a scatter plot of values outside the two standard deviations from the HFD mean, as seen in Fig. 6. The final outliers are selected by our proposed method by analysing both reoccurring and confounding data intersecting the results in both the scanpath comparison methods and geometric methods as seen in Table 1. Both P11 and P44 are seen as poorly matching results in MultiMatch vector, direction, and length. ScanMatch was able to detect P44 but not P11. Bounding box detected both P11 and P44 but returned confounding results with P21. The Jaccard method correctly identified P11 and P44 with no confounding results. The HFD method also correctly identified P11 and P44 but also produced two results not identified elsewhere with P20 and P46. These confounding HFD results prompt the researcher to vi-

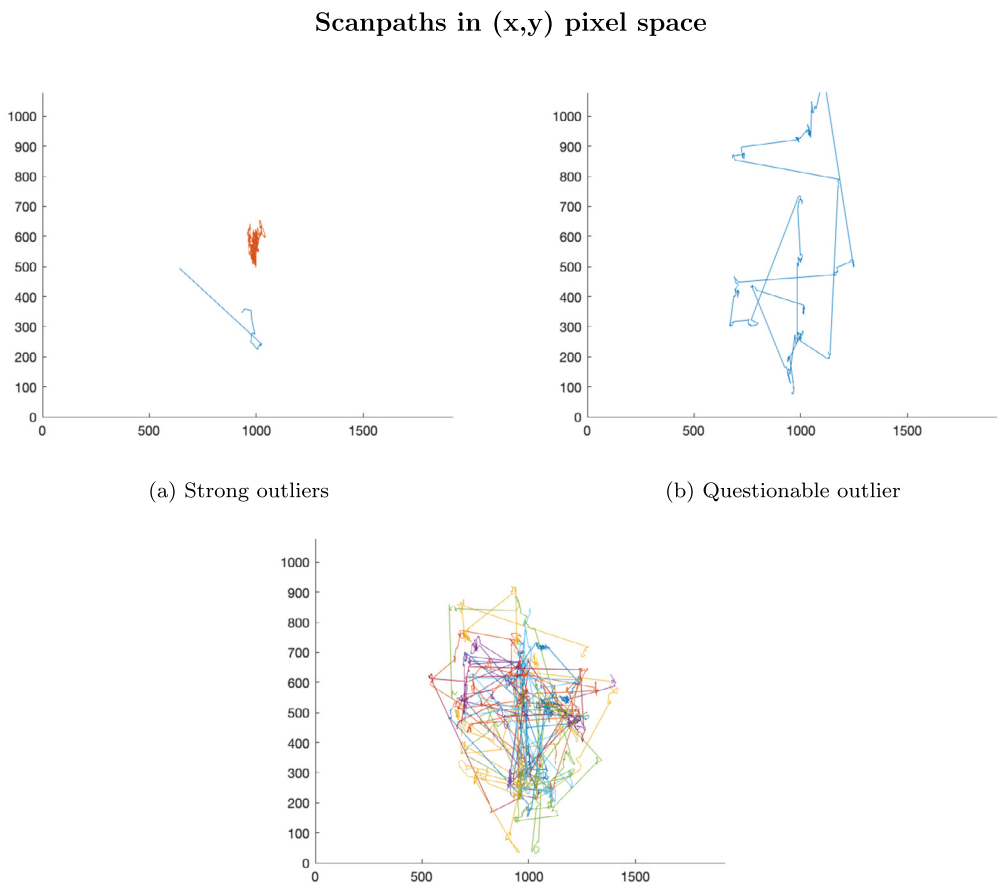


Fig. 8. (a) Visually recognisable outliers P44 in blue and P11 in red, (b) visually questionable outlier P31, (c) and non-outliers P01-P10 and P12-P16.

sually examine the confounding scanpaths for P20 and P46 seen as red and blue in Fig. 1 (b), which indeed appear unusual in geometry to others. Further study can be made to determine if these HFD confounding results indeed point to deeper insight into whether they represent artefacts in the experimental process or are indeed false outlier predictions.

3. Results

The candidate outliers identified by ScanMatch to cluster poorly with other participants include P44, P25, and P14, as shown by pale matches in Fig. 7. MultiMatch returns five matching scores, with P31 being a strong, consistent outlier result, and P07, P21, and P44 returned as weak matches illustrated by Vector, Direction and Length scores. MultiMatch Position and Duration only provided strong clues for P31 and weak scores for P41, P43, and P45 which were not revealed as outliers through any other method.

Interestingly, P31 performed strongly as an outlier in MultiMatch but was not found as an outlier using any other method, including ScanMatch, bounding box, Jaccard, or HFD methods. Upon visual inspection of the scanpath for P31 as seen in Fig. 8 (b), it appears visually unremarkable compared to P44 shown in Fig. 8 (a), which was identified by ScanMatch as being a strong outlier.

MultiMatch Length was able to identify P11, which was recognised by bounding box, Jaccard, and HFD methods as an outlier, and also appeared unusual during visual inspection demonstrated in Fig. 8(a). Burch et al. [8] would identify P44, P21, and P11 using the bounding box, and P44 and P11 using the Jaccard method as outliers as illustrated in Fig. 9.

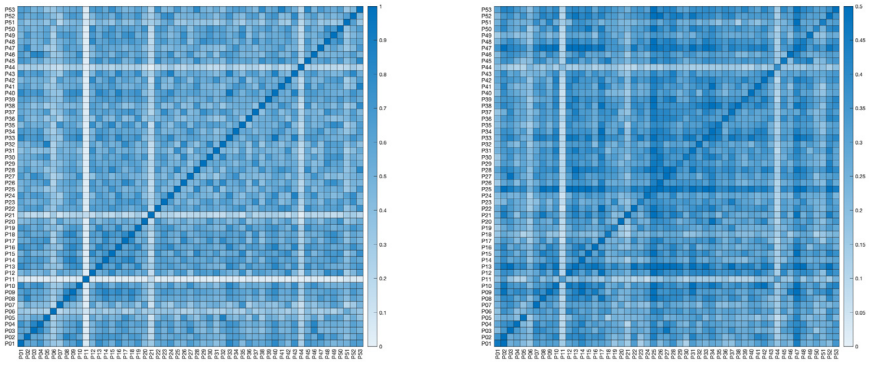
3.1. Performance analysis

Using the FD as a parameter in addition to ScanMatch results illustrates that the geometric complexity of P44 is outside 2 standard deviations from the mean, as illustrated by Fig. 6, while P14 and P25 are within one. This indicates that P44 is a strong candidate as outlier due to it being both a poor match with other scanpaths in the heatmap and a poor match in geometric complexity as demonstrated in the scatter plot.

Additionally, when comparing FD with MultiMatch results, we find that the length score agreed with HFD method in identifying P11 as an outlier, seen in Fig. 6 well above 2 standard deviations from the mean, but also aligned with other MultiMatch scores for weak matches with P31, which was found by FD method to be normal.

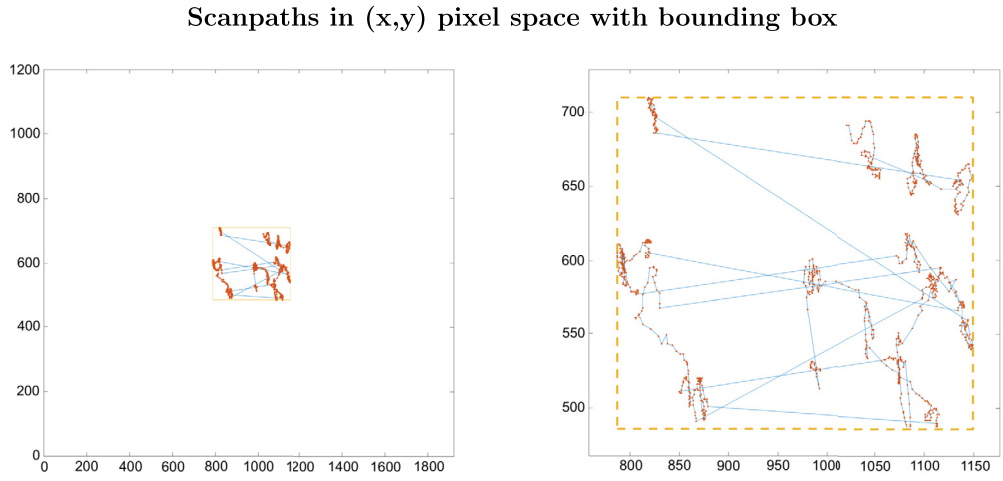
Indeed, P31 was a strong candidate as outlier in all MultiMatch results but appeared unremarkable visually, as seen in Fig. 8 (b), and geometrically, as measured by Burch et al. [8] methods in Fig. 9, and in the HFD scatter plot in Fig. 6. This poor matching outlier may be an example of the scanpath represented as a solitary cluster rather than as a true scanpath anomaly. Slightly weaker candidates for poor matches, MultiMatch length, direction and the bounding box method identified P21 as an outlier.

However, both the Jaccard and HFD method found P21 to be geometrically similar to others, making this a possible false outlier identified by both MultiMatch and the bounding box method. The Burch et al. [8] bounding box method identified P11, P21, and P44 as outliers while the Jaccard method identified P11 and P44. As described above, the bounding box method identified P21 as an outlier which did



(a) Bounding Box (b) Jaccard

Fig. 9. Bounding box (a) method displays outliers being P44, P21, and P11. Jaccard method (b) displays outliers being P44 and P11.



(a) Confounding Outlier (b) Confounding Outlier close-up

Fig. 10. (a) Confounding outlier P21 to scale with stimulus size, (b) close-up of confounding outlier P21 with orange dotted bounding box, detailed fixation points in red, and scanpath in blue.

not appear visually unusual or geometrically unique using HFD or Jaccard. Fig. 10 illustrates that the back and forth scanning and fixation gaze pattern closely resembles the typical complexity pattern seen in the dataset demonstrated by Fig. 5 (b). However, when viewing all the scanpaths in the dataset, as seen in Fig. 4 (b) the size of the bounding box and its position are out of place compared to others. This explains why the bounding box method and MultiMatch Direction and Length heatmaps are sensitive to its comparatively unusually small boundary. However, because the geometric complexity of the gaze data appear to be similar to other samples, this may point to intentional visual fixations of a specific detail in the stimulus made by the participant rather than an artefact of distraction or inattention. The Jaccard method did correctly identify the most obvious outliers as being P44 and P11; however, the more granular method of using Jaccard binning versus boxing appeared to attenuate the contrast in outliers as illustrated by the flat appearance of the Jaccard heatmap in Fig. 9.

The results indicate that both P44 and P11 are strong candidates for exclusion from any clustering analysis due to both their differences in geometric complexity and their poor matches with other scanpaths using heatmaps. Poor candidates including P21, P31, P07, and P14 are more difficult to deem as outliers due to their geometric similarity to well matching scanpaths.

4. Discussion

Burch et al. [8] introduced geometric complexity as a means of distinguishing outliers in scanpath data, and this paper's method of reducing dimensionality using Hilbert curve distances followed by the distillation of its geometry using the HFD is shown to provide an additional parameter that can be used to spot anomalous scanpaths.

Shortcomings to this approach include the statistically mundane way in which FD values are found using two standard deviations from the mean. A more robust approach may include calculating the Gaussian of the distribution of values to find trailing values at the tails. Regarding locality preservation using Hilbert curves, a weakness exists in the 'neck' of the curve, as seen in Hilbert curve distance 2 and 15, or coordinate $(2,4)$ and $(3,4)$ in Fig. 3. These coordinates represent the weakest locality preservation points in a Hilbert 2D to 1D conversion. Another weakness could exist due to no accounting for scale. The geometric complexity of fine perturbations found in the fixations of a participant viewing a small dot could be interpreted similarly to sweeping eye jerks and fits over a larger canvas captured by a very sensitive eye tracking device. This can be mitigated through quantising out the fine perturbations found in more sensitive devices.

Furthermore, some scanpaths exist for which geometric complexity acts as good method for detecting outliers but ScanMatch and Multi-

Match methods do not work. These may be artefacts of quantisation in the ScanMatch or MultiMatch comparison algorithm but may also point to unique cases where scanpath geometry does not play a factor in two scanpaths being similar. Further study in these edge cases may provide insight into how important a role geometric complexity plays in both the comparison of scanpaths and the identification of its outliers. The quantitative nature of this research opens the door to a deeper qualitative study investigating the scanpath attributes that cause these participants to match poorly, and what, if any, connections there are between participant expertise, geometric complexity outliers, and poor matchability.

Indeed, observations by Jones and Sibson [23] regarding the proportions of variances within data representing an "index of interestingness" imply that robust outlier methods for pruning perturbations while maintaining "interestingness" found in unusual, yet valid, samples will increase accuracy of unsupervised clustering results that rely on exploratory multivariate analysis. Similarly, this research attempts to capture "interesting" deviations through the distribution of fixations made via their geometric complexity, as seen in the sequence of Hilbert distances. However, unlike the projection pursuit where components are removed as data elements are reduced, this method both preserves locality in each point during dimensionality reduction by using Hilbert distances, and also evaluates entire sequences without reduction by using their Fractal dimension. Machine learning methods including Support Vector Machines and k-Nearest Neighbor would immediately benefit from outlier reduction due to their sensitivity ([24, 25]) to being mislead by extraneous artefacts.

5. Conclusion

In this research, a novel quantitative approach to detect outlier scanpaths using scatter plot of FD and heatmaps by scanpath comparison packages has been introduced. The main novelty of the proposed method is the computation of FD as an additional parameter. In addition, the scanpath representation by Hilbert curve distances is a further innovation which might mitigate the quantisation problem of traditional scanpath comparison packages. The effectiveness of the proposed method has been validated through visual results.

However, some unresolved problems remain where work can be continued. Eye tracking devices continue to increase the resolution of both the (x, y) space and (t) duration of captured fixations. This increased detail may make it harder to discriminate perturbations found in a singular small cluster of fixations versus those found in the random visual search patterns of the entire stimulus. Introducing a more granular approach to time within the temporal sequence of fixations may aid in the discrimination of aberrant clustered versus stimulus-wide geometric complexity.

Future directions in this research could include an extension of the Hilbert method to develop a novel scanpath matching framework. The strength of the Hilbert method's comparatively quantitative nature may provide increased precision during quantisation. This approach may prove to be more robust than the gridded binning methods used in ScanMatch and MultiMatch.

Declarations

Author contribution statement

R.A. Newport: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

C. Russo, A. Al Suman: Analyzed and interpreted the data.

A. Di Ieva: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Funding statement

This work was funded by the Centre for Elite Performance, Expertise & Training (Macquarie University, Sydney) Seeding Grant awarded to Nalepka, Carrigan, & Di Ieva in 2019 (080619) and by an Australian Research Council (ARC) Future Fellowship granted to Prof Di Ieva in 2019 (FT190100623).

Data availability statement

Data will be made available on request.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

For the data collection, we would like to thank Dr Ann Carrigan and Dr Patrick Nalepka.

References

[1] D. Noton, L. Stark, Scanpaths in saccadic eye movements while viewing and recognizing patterns, Vis. Res. (1971).

[2] A. Kumar, N. Timmermans, M. Burch, K. Mueller, Clustered eye movement similarity matrices, in: Eye Tracking Research and Applications Symposium (ETRA), 2019.

[3] J.H. Goldberg, J.I. Helfman, Scanpath clustering and aggregation, in: Eye Tracking Research and Applications Symposium (ETRA), 2010.

[4] R. Fahimi, N.D. Bruce, On metrics for measuring scanpath similarity, Behav. Res. Methods (2020).

[5] N. Castner, T.C. Küebler, K. Scheiter, J. Richter, T. Eder, F. Hüettig, C. Keutel, E. Kasneci, Deep semantic gaze embedding and scanpath comparison for expertise classification during OPT viewing, in: Eye Tracking Research and Applications Symposium (ETRA), 2020.

[6] T.T. Brunyé, B.K. Nallamothu, J.G. Elmore, Eye-tracking for assessing medical image interpretation: a pilot feasibility study comparing novice vs expert cardiologists, Perspec. Med. Educ. (2019).

[7] M. Król, M. Król, Learning from peers' eye movements in the absence of expert guidance: a proof of concept using laboratory stock trading, eye tracking, and machine learning, Cogn. Sci. (2019).

[8] M. Burch, A. Kumar, K. Mueller, T. Kervezee, W. Nuijten, R. Oostenbach, L. Peeters, G. Smit, Finding the outliers in scanpath data, in: Eye Tracking Research and Applications Symposium (ETRA), 2019.

[9] I.T. Jolliffe, Principal Component Analysis, second edition, Encyclopedia of Statistics in Behavioral Science, 2002.

[10] Y. Li, Y. Shang, Y. Yang, Clustering coefficients of large networks, Inf. Sci. 382–383 (2017) 350–358, https://www.sciencedirect.com/science/article/pii/S0020025516320527.

[11] A. Davies, M. Vigo, S. Harper, C. Jay, The visualisation of eye-tracking scanpaths: what can they tell us about how clinicians view electrocardiograms?, in: Proceedings of the 2nd Workshop on Eye Tracking and Visualization, ETVIS 2016, 2017.

[12] F. Cristino, S. Mathôt, J. Theeuwes, I.D. Gilchrist, ScanMatch: a novel method for comparing fixation sequences, Behav. Res. Methods (2010).

[13] S.B. Needleman, C.D. Wunsch, A general method applicable to the search for similarities in the amino acid sequence of two proteins, J. Mol. Biol. (1970).

[14] R. Dewhurst, M. Nyström, H. Jarodzka, T. Foulsham, R. Johansson, K. Holmqvist, It depends on how you look at it: scanpath comparison in multiple dimensions with MultiMatch, a vector-based approach, Behav. Res. Methods (2012).

[15] R.P. Burn, B.B. Mandelbrot, The fractal geometry of nature, Math. Gaz. 68 (1984) 71.

[16] R. Faigle, R. Sutter, P.W. Kaplan, Electroencephalography of encephalopathy in patients with endocrine and metabolic disorders, 2013.

[17] A. Di Ieva, F.J. Esteban, F. Grizzi, W. Klonowski, M. Martín-Landrove, Fractals in the neurosciences, part II: clinical applications and future perspectives, Neuroscientist 21 (2015) 30–43.

[18] J.E. Jacob, G.K. Nair, A. Cherian, T. Iype, Application of fractal dimension for EEG based diagnosis of encephalopathy, Analog Integr. Circuits Signal Process. (2019).

[19] W. Klonowski, Fractal analysis of electroencephalographic time series (EEG signals), 2016.
[20] G. Peano, Sur une courbe, qui remplit toute une aire plane, Math. Ann. (1890).
[21] T. Higuchi, Approach to an irregular time series on the basis of the fractal theory, Phys. D: Nonlinear Phenom. (1988).
[22] L. Zhong, F. Zeng, G. Xu, Comparison of fractal dimension calculation methods for channel bed profiles, in: Procedia Engineering, 2012.
[23] M.C. Jones, R. Sibson, What is projection pursuit?, J. R. Stat. Soc. A, General (1987).
[24] X. Yang, Q. Song, A. Cao, Weighted support vector machine for data classification, in: Proceedings of the International Joint Conference on Neural Networks, 2005.
[25] G. Bhattacharya, K. Ghosh, A.S. Chowdhury, kNN Classification with an Outlier Informative Distance Measure, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2017.