A Symbolic Approach for Task-related Gaze Classification

Quentin Laborde[1,2], Laurent Oudre[1] and Ioannis Bargiotas[3]

***Abstract*— This paper introduces an innovative symbolic framework for task-related gaze classification. The proposed method encodes multivariate eye-tracking feature signals into symbolic string sequences, facilitating the computation of an interpretable and robust distance metric between recordings. Applied to the ETRA 2019 dataset, the symbolic representation enables accurate clustering of visual tasks without dependence on black-box models. Experimental results demonstrate that our approach achieves performance comparable with state-of-the-art deep learning methods.**

I. Introduction

Eye movement research has a long history, dating back to [1] in the early 20^{th} century. Technological advancements have improved measurement, storage, and analysis, facilitating its widespread use in fields such as neuroscience, marketing, psychology, and medicine. Gaze behavior is shaped by factors like expertise, cognitive state, task demands, and personality. Beyond reflecting cognitive processes, gaze trajectories provide insights into task intentions, expertise, and contextual information.

Raw eye-gaze recordings can be analyzed across multiple dimensions. At a basic level, analysis focuses on extracting features from physiological eye movements such as fixations, saccades, and smooth pursuits. At a higher level, gaze trajectories are examined through scanpath analysis and visual areas of interest (AoIs), which add a semantic layer by considering what is being observed. However, these dimensions are typically analyzed independently, and most state-of-the-art methods focus on a single dimension [2].

To address this limitation, we propose a novel approach that uses symbolic representation methods to transform eye-tracking signals into string sequences, capturing multiple dimensions of visual behavior. In our method, symbolic representation involves replacing segments of complex multivariate signals — specifically, extracted features from eye-gaze recordings — with symbols, *e.g.* letters, that represent the semantic value of these sequences. Symbolization methods generally consist of two key steps: segmenting complex, multivariate signals into homogeneous chunks, followed by quantization, where each segment is mapped to a discrete value from a set of *atoms*.

This transformation allows for the construction of a meaningful distance metric between eye-tracking recordings, facilitating the clustering of these sequences. Our method provides a flexible, fully interpretable analysis pipeline, eliminating the need for black-box models. We validate our approach using the ETRA 2019 challenge dataset [3], positioning it within the context of existing state-of-the-art methods.

The following sections provide an overview of the current state-of-the-art in scanpath analysis, with a focus on string-based approaches (Section II), followed by a detailed description of our proposed method (Section III). Finally, we present the evaluation results of our approach (Section IV)

II. Related Works

Comparison of gaze trajectories is fundamental in eye-tracking research, with established methodologies [4]. A common representation, the scanpath, consists of fixations — eye pauses — and saccades — rapid transitions. For instance, Eyeanalysis, a simple method, measures scanpath similarity by mapping fixations based on spatial proximity [5], while MultiMatch evaluates multiple dimensions, including spatial location and duration. However, both focus on spatial attributes, neglecting semantic information.

To integrate spatial and semantic dimensions, string-based methods encode scanpaths as symbol sequences, enabling effective similarity measurement and cognitive state classification. This typically involves defining Areas of Interest (AoIs) within the stimulus [6]. String-editing metrics like Levenshtein and Needleman-Wunsch [7] compare these sequences but primarily capture spatial properties, overlooking transition dynamics between AoIs—key for understanding task execution and expertise levels [8].

To overcome this limitation, subsequence-based methods have emerged. SubsMatch compares scanpaths by analyzing subsequence frequency, using a sliding window to segment scanpaths into fixed-length subsequences stored in frequency dictionaries [9]. A recent advancement, Geisler et al.'s MinHash-based approach [10], also employs a sliding window but estimates similarity via the MinHash algorithm to approximate the Jaccard index. This method has been rigorously tested on the ETRA 2019 benchmark dataset.

In recent years, deep learning has advanced scanpath analysis, particularly for task-based clustering of eye-tracking signals. Bautista et al. [11] proposed an autoencoder-based model that processes raw position and velocity signals to capture fine-grained and macro-scale gaze patterns while remaining stimulus-agnostic. Similarly, Thentu et al. [12] applied convolutional neural networks (CNNs) to RGB images of fixation scanpaths from cognitive tasks. Evaluated on the ETRA 2019 dataset, these models enable direct comparison with state-of-the-art methods, highlighting the growing role of neural networks in eye-tracking research.

[1]Université Paris Saclay, Université Paris Cité, ENS Paris Saclay, CNRS, SSA, INSERM, Centre Borelli, F-91190, Gif-sur-Yvette, France

[2]SNCF, Technologies Department, Innovation & Research, F-93210, Saint Denis, France

[3]Université Paris-Saclay, Inria, CIAMS, F-91190, Gif-sur-Yvette, France

III. METHODS

The overall objective of our method is to compute, using symbolic representation techniques, a distance between pairs of eye-gaze recordings. The resulting distance matrix can then be used as input for standard clustering algorithms to group the recordings according to the visual task performed. This approach aims to provide a flexible classification method for visual cognitive tasks, achieving performance comparable to deep learning-based methods while maintaining a modular processing pipeline, free from black-box models.

The proposed approach involves several steps, performed independently for each modality, outlined in the following sections. An illustration of the entire procedure, from raw eye-gaze recordings to the final consolidated distance matrix, is provided in Figure 1.

A. Step 1: Feature Extraction

The first step involves extracting features that reflect gaze behavior from a set of N eye-gaze recordings. We focus on well-established features that characterize oculomotor behavior and gaze dynamics across four analysis modalities: *fixation*, *saccade*, *scanpath*, and *area of interest* features — denoted as fix, sac, sp and aoi throughout this article. We briefly describe these modalities below.

1) Fixation Features.: A fixation is a period during which gaze remains directed at a specific location, projecting visual content onto the fovea centralis. Fixations and saccades are first identified using a Hidden Markov Model (HMM) approach [13]. Five features are extracted from fixation sequences: fixation *frequency*, *drift displacement*, *drift distance*, *drift velocity* and *bivariate contour ellipse area*.

2) Saccade Features.: Saccades are rapid, ballistic eye movements that redirect gaze toward objects of interest. Identified using the same HMM approach, saccade features describe both the shape and kinematics of these movements. The extracted features include saccade *frequency*, *amplitude*, *efficiency*, *peak velocity*, *peak acceleration*, *skewness exponent* and *peak velocity/amplitude ratio*.

3) Scanpath Features.: A scanpath is a sequence of consecutive fixations characterized by their spatial positions — horizontal and vertical coordinates — and durations. Features in this modality geometrically characterize scanpath trajectories, including scanpath *length*, *convex hull*, *bivariate contour ellipse area*, *fractal dimension*, *recurrence rate*, *recurrence determinism* and *recurrence laminarity*.

4) Areas of Interest Features.: AoIs represent regions of semantic or contextual importance within the visual field, replacing precise fixation positions with occurrences of gaze within predefined areas. AoIs are identified using a mean-shift clustering approach [14]. Once fixations are assigned to clusters — *i.e.* an area of interest — a set of features can be extracted. Specifically, we compute AoI *number*, *bivariate contour ellipse area*, *Lempel Ziv complexity*, *transition stationary entropy*, *transition joint entropy* and *transition mutual information*.

For each modality, features and descriptors are computed over fixed-length—and potentially overlapping—time windows. This feature extraction process thus generates four time series, denoted as $\mathbf{X}^{n,m}$, for each eye-gaze recording $n = 1, \ldots, N$, and each modality $m \in fix, sac, sp, aoi$. These time series have dimensions of $T^m \times D^m$, where T^m represents the number of time windows considered, and D_m indicates the number of features examined for each modality.

B. Step 2: Data Uniformization

In the absence of *a priori* knowledge about the expected distributions of the descriptors under study, experimental cumulative distribution functions (ECDFs) are computed for each feature within each modality across recordings. These ECDFs are then used to standardize the feature values to the range $[0, 1]$.

This step has two advantages: (i) for the segmentation step, which relies on changes in the mean, this standardization reduces the impact of extreme values in regions of low density within the feature distributions, and (ii) the symbolization process used in the symbolization step is similarly facilitated by reducing disparities in feature scales.

At the end of this step, we obtain normalized feature series. For simplicity, we retain the same notations. Specifically, we generate four time series, denoted as $\mathbf{X}^{n,m}$, for each eye-gaze recording $n = 1, \ldots, N$, and each modality $m \in \{fix, sac, sp, aoi\}$, with dimensions of $T^m \times D^m$.

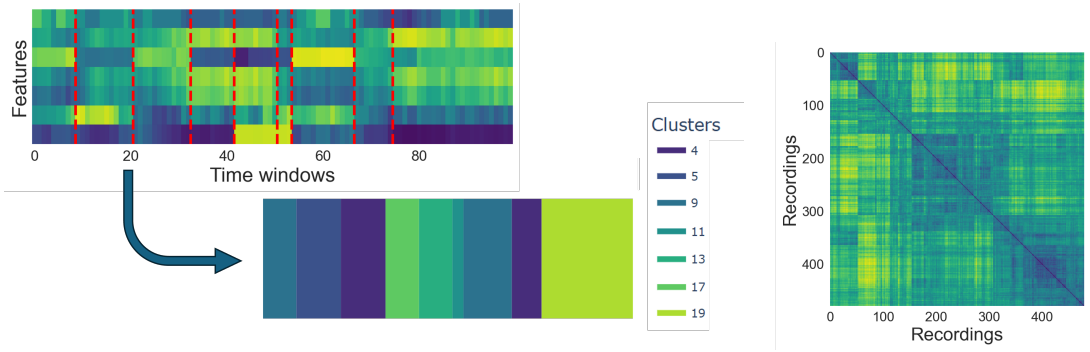
C. Step 3: Adaptive Segmentation

Adaptive signal segmentation involves applying a change-point detection algorithm to a given signal. In our case, the goal is to identify breakpoints that partition the input time series, denoted as $\mathbf{X}^{n,m}$, where $m \in fix, sac, sp, aoi$, into relatively homogeneous segments. Specifically, This process identifies the w^* unknown time instants $t_1^* < ... < t_{w^*+1}^*$ at which certain statistical properties — specifically, the mean — of the multidimensional input signal change abruptly.

We assume the number of change points, w^*, is unknown and must be estimated. Given an input time series $\mathbf{X} = \mathbf{x}_1, \ldots, \mathbf{x}_T$, the change-point estimates $\hat{t}_1, \ldots, \hat{t}_{\hat{w}+1}$ — $\hat{w}$ represents the detected number of changes — are obtained by minimizing the following optimization problem [15]

$$(\hat{w}, \hat{t}_1, \ldots, \hat{t}_{\hat{w}+1}) := \underset{(w, t_1, \ldots, t_{w+1})}{\operatorname{argmin}} \sum_{k=0}^{w+1} \sum_{t=t_k}^{t_{k+1}-1} ||\mathbf{x}_t - \overline{\mathbf{x}}_{t_k:t_{k+1}}||^2 + \lambda(w+1)$$

where $\overline{\mathbf{x}}_{t_k:t_{k+1}}$ is the empirical mean of $\{\mathbf{x}_{t_k}, \ldots, \mathbf{x}_{t_{k+1}-1}\}$ and $\lambda > 0$ is a penalization parameter. The penalized formulation above balances reconstruction accuracy—measured by the sum of squared errors—with model complexity, determined by the number of change points. The optimization problem is solved using the *Pruned Exact Linear Time* (PELT) algorithm [16]. The penalty parameter λ controls the trade-off between sensitivity and sparsity. For calibration, we adopt the standard penalty scaling $\lambda = ln(n)$ [15].



(a) Symbolization (b) Distance measure

Fig. 1: **Method Summary.** We present the pipeline for computing a distance matrix between eye-gaze recordings, demonstrated here for a single modality — the *saccade modality*. First, saccadic features are extracted and normalized across all recordings using an experimental cumulative distribution function (CDF). Next, an adaptive segmentation algorithm is applied to the normalized feature timelines — shown in the upper part of Figure 1a. These segments are then clustered across all recordings to identify distinct behavioral atoms, which serve as symbolic representations of the timelines. An example of a symbolic timeline is provided in the lower part of Figure 1a. Finally, a pairwise string-edit distance is computed between each symbolic timeline pair, generating the distance matrix shown in Figure 1b.

At the end of this process, each time series $\mathbf{X}^{n,m}$ from every eye gaze recording is segmented according to w breakpoints, partitioning the series into homogeneous segments.

D. Step 4: Symbolization

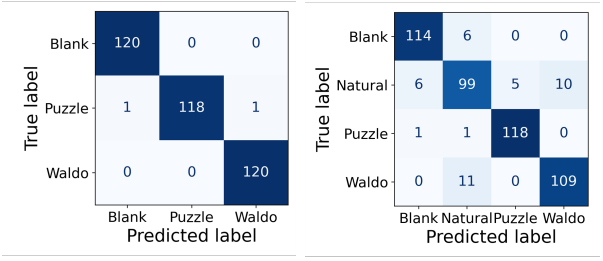
Once the segment boundaries are identified for each modality across all input recordings, the mean per segment — in dimension D^m for $m \in \{fix, sac, sp, aoi\}$ — is computed for each segment. This process produces a set of segments, each represented by a feature vector of dimension D^m. From these segments, we aim to extract a set of base *atoms* that represent *canonical behaviors*, to which all segments will be subsequently assigned. While various clustering methods could be employed for this purpose, we opted for a standard K-means algorithm, where the desired number of clusters K corresponds to the number of symbols for each modality.

Formally, for each modality $m \in \{fix, sac, sp, aoi\}$ given a set of N_s segments, denoted as $\mathbf{x}_1, \ldots, \mathbf{x}_{N_s}$, with only the segment means retained with dimension D^m, the objective is to partition the N_s segments into K clusters, $\mathbf{C} = \{C_1, \ldots, C_K\}$. This partitioning is achieved by minimizing the within-cluster distance, as

$$\underset{\mathbf{C}}{\operatorname{argmin}} \sum_{i=1}^{K} \sum_{\mathbf{x}_j \in S_i} ||\mathbf{x}_j - \mu_i||^2$$

with μ_i the mean — also called centroid — of points in C_i.

At the end of this step, each segment is assigned a symbol corresponding to the label of its associated cluster. Consequently, for each eye-gaze recording $n = 1, \ldots, N$ and each modality $m \in \{fix, sac, sp, aoi\}$, we obtain a sequence $\mathbf{S}^{n,m}$ representing the series of segments linked to specific atoms. This effectively transforms the raw gaze data into a symbolic, or string-based, representation. Furthermore, the pairwise distance between individual symbols can be computed as the Euclidean distance between their corresponding cluster centroids.



(a) ETRA 3 (b) ETRA 4

Fig. 2: **Confusion matrices.** Confusion matrices obtained using a SVM classifier for ETRA-3 task and ETRA-4 tasks.

E. Step 5: Distance Measure

A simple distance measure for symbolic sequences is the general edit distance, which calculates the minimum cost required to transform one sequence into another using three fundamental edit operations, namely *deletion*, *insertion* and *substitution*. Each operation is assigned a specific edit cost, and the total transformation cost is computed using the Wagner-Fischer algorithm [17], a dynamic programming approach for efficiently determining the edit distance.

Given two symbolic sequences $\mathbf{S}^1 = s_1^1, \ldots, s_{N_1}^1$ and $\mathbf{S}^2 = s_1^2, \ldots, s_{N_2}^2$, the Wagner-Fischer algorithm first constructs a cost matrix $\mathbf{D}$ of size $(N_1+1)\times(N_2+1)$. Each entry $\mathbf{D}(i,j)$ represents the minimum cost required to transform the first i elements of $\mathbf{S}^1$ into the first j elements of $\mathbf{S}^2$. The algorithm recursively fills the matrix according to the

TABLE I: **Accuracy.** Reported accuracies for the ETRA-4 task (Blank, Natural, Puzzle, Waldo) and the ETRA-3 task (Blank, Puzzle, Waldo). The best results are underlined. The presented method is denoted **Symbolization**.

Method	**ETRA** 4	**ETRA** 3
Symbolization	<u>0.917</u>	<u>0.994</u>
Encodji [18]	0.890	—
MinHash [10]	0.880	—
GazeMAE [11]	0.875	—
AdaBoost [12]	0.840	0.880
ConvNet [12]	0.880	0.970

following recurrence relation:

$$\left[\mathbf{D}^{(0)}\right]_{i,j} = \min \begin{cases} \left[\mathbf{D}^{(0)}\right]_{i-1,j} + \tau_{del} \\ \left[\mathbf{D}^{(0)}\right]_{i,j-1} + \tau_{ins} \\ \left[\mathbf{D}^{(0)}\right]_{i-1,j-1} + \tau_{sub}(s^1_{i-1}, s^2_{j-1}) \end{cases}$$

where, τ_{del} and τ_{ins} represent the costs for deletion and insertion, respectively. Furthermore $\tau_{sub}(s^1_{i-1}, s^2_{j-1})$ represents the Euclidean distance between the centroids of the corresponding symbols, as defined in Step 4. Finally, the last entry of the distance matrix, $\left[\mathbf{D}\right]_{N_1,N_2}$ represents the total cost to align the two sequences. This value is then normalized by $\max(\mathrm{N}_1, \mathrm{N}_2)$ to produce a dissimilarity score.

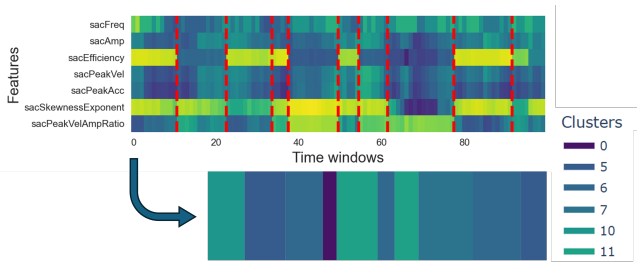
Since the segments generated in Step 3 vary in length, their length information can be incorporated into the symbolic sequences by replicating each symbol proportionally. This preserves temporal structure, giving longer segments greater weight in sequence alignment and dissimilarity.

At this stage, we obtain four distance matrices, one for each modality. The mean square across these matrices is computed element-wise to produce a single, consolidated distance matrix between the recordings. This resulting distance matrix represents the final output of our symbolization method.

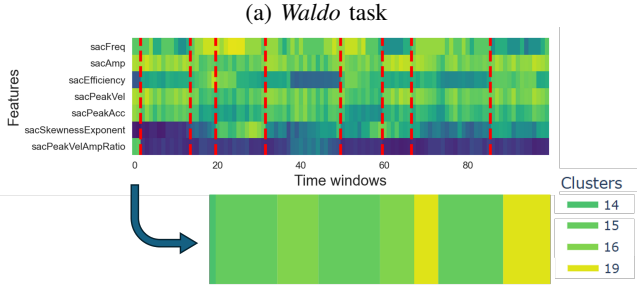
IV. Results

We evaluate our method using the Eye Tracking Research & Applications (ETRA) dataset. This dataset has been previously employed for analyzing saccades and microsaccades [3] and was featured in the ETRA 2019 data mining challenge. It consists of recordings from eight participants who viewed four distinct image types: (i) blank images, (ii) natural scenes, (iii) picture puzzles, and (iv) *"Where's Waldo?"* images. For blank and natural scenes, participants freely explored the images. Picture puzzles presented pairs of nearly identical images, where participants were tasked with identifying the differences. *"Where's Waldo?"* images depicted complex scenes, challenging participants to find the character Waldo. Each viewing session lasted approximately 45 seconds. Eye movements were tracked using an SR Research EyeLink II eye-tracker with a sampling frequency of 500 Hz, resulting in $N = 480$ eye movement recordings — 120 for each image type.

Features introduced in Step 1 were extracted, for each recording, using time windows of length 5 seconds for *fix* and *sac* features, 20 seconds for *sp* features and 30 seconds for *aoi* features. For each recording and each modality, $L^{fix} = 100$, $L^{sac} = 100$, $L^{sp} = 50$ and $L^{aoi} = 50$ time windows were examined. After adaptative segmentation step, identified segments of each recording were symbolized, using $K = 20$ symbols for each modality. Finally, a distance matrix of size $N = 480$ was computed.



(a) *Waldo* task



(b) *Puzzle* task

Fig. 3: **Visual inspection.** For the same participant, this figure presents segmented time series for the saccade modality. The examples correspond to the *Waldo* task — Figure 3a — and the *Puzzle* task — Figure 3b — illustrating the differences in visual behavior between the two tasks. These differences are further emphasized in the symbolic representations derived from the segmented timelines.

Step 5 generates a pairwise distance matrix between data sequences. We then applied metric multidimensional scaling (MDS) [19] to embed these distances into a lower-dimensional space. Given the distance matrix and a predefined number of dimensions $N_{MDS} = N/4$, MDS maps each data point into an N_{MDS}-dimensional space, aiming to preserve pairwise distances. The resulting N_{MDS}-dimensional coordinates serve as input to a support vector machine (SVM) classifier. The SVM, using a Radial Basis Function (RBF) kernel, identifies the optimal hyperplane that maximally separates the data points. A grid search was performed to tune the regularization parameter $C = [0.1, 1, 10]$. Classification performance was evaluated using 5-fold cross-validation. Two classification tasks were conducted:

- **ETRA 4.** Classification of 4 visual activities, namely *Blank*, *Natural*, *Puzzle*, and *Waldo*.
- **ETRA 3.** Classification of 3 visual activities, namely *Blank*, *Puzzle*, and *Waldo*.

The resulting confusion matrices are presented in Figure 2. Mean accuracies, along with a comparison to results obtained

TABLE II: **Ablation Study.** Accuracy for the ETRA-3 and ETRA-4 tasks obtained through the sequential removal of modalities.

Modalities	ETRA 4	ETRA 3
$\{fix, sac, sp, aoi\}$	0.917	0.994
$\{sac, sp, aoi\}$	0.876	0.989
$\{fix, sp, aoi\}$	0.850	0.992
$\{fix, sac, aoi\}$	0.875	0.986
$\{fix, sac, sp\}$	0.862	0.994
$\{fix, sac\}$	0.824	0.981
$\{fix, sp\}$	0.808	0.978
$\{fix, aoi\}$	0.818	0.958
$\{sac, sp\}$	0.852	0.986
$\{sac, aoi\}$	0.853	0.975
$\{sp, aoi\}$	0.811	0.956
$\{fix\}$	0.686	0.889
$\{sac\}$	0.756	0.958
$\{sp\}$	0.735	0.933
$\{aoi\}$	0.723	0.903

from other methods in the literature, are provided in Table I.

Additionally, a visual inspection of the process is possible. Figure 3 shows segmented time series from the same participant during two different tasks, namely *Puzzle* and *Waldo*, for the *saccade* modality. In the *Waldo* task — Figure 3a — the participant exhibits a low saccade rate with small amplitude, accompanied by low peak velocity and acceleration—indicative of localized visual information acquisition, consistent with the task's requirements. In contrast, during the *Puzzle* task — Figure 3b — the participant quickly shifts from one side of the visual stimulus to the other while searching for differences between the two images. This behavior is characterized by frequent, fast, and high-amplitude saccades.

These differences in saccadic behavior are preserved after the symbolization step where the resulting symbolized sequences contain distinctly different patterns of symbols, or *atoms*, for each task.

Finally, we evaluate the effectiveness of our approach in correctly classifying tasks based on the incorporated modalities. The results, presented in Table II, indicate that accuracy decreases progressively with the sequential removal of modalities. This suggests that all modalities contribute complementarily to achieving high precision, with no apparent redundancy.

V. Discussion

We demonstrated that our approach effectively clusters visual tasks using well-established eye-tracking descriptors, achieving higher accuracy than recent deep learning methods. The method is fully transparent, avoiding black-box models while remaining modular, allowing emphasis on specific eye-tracking dimensions based on data characteristics.

This method should be seen as a general 5-step approach, where each step can be refined. For instance, the features extracted in Step 1 can be adjusted based on the specific context. Similarly, in Step 4, we opted for a simple clustering algorithm for symbolization, but more advanced approaches could be considered, such as Gaussian Mixture Copula Models or Kernel K-means clustering algorithms.

ACKNOWLEDGMENT

Part of this work has been founded by the Industrial Data Analytics And Machine Learning chairs of ENS Paris-Saclay

References

[1] R. Dodge and T. S. Cline, "The angle velocity of eye movements." *Psychological Review*, vol. 8, no. 2, p. 145, 1901.

[2] W. F. Bischof, N. C. Anderson, and A. Kingstone, "Temporal methods for eye movement analysis," *Eye movement research: An introduction to its scientific foundations and applications*, pp. 407–448, 2019.

[3] J. Otero-Millan, X. G. Troncoso, S. L. Macknik, I. Serrano-Pedraza, and S. Martinez-Conde, "Saccades and microsaccades during visual fixation, exploration, and search: foundations for a common saccadic generator," *Journal of vision*, vol. 8, no. 14, pp. 21–21, 2008.

[4] N. C. Anderson, F. Anderson, A. Kingstone, and W. F. Bischof, "A comparison of scanpath comparison methods," *Behavior research methods*, vol. 47, no. 4, pp. 1377–1392, 2015.

[5] S. Mathôt, F. Cristino, I. D. Gilchrist, and J. Theeuwes, "A simple way to estimate similarity between pairs of eye movement sequences," *Journal of Eye Movement Research*, vol. 5, no. 1, pp. 1–15, 2012.

[6] W. Fuhl, T. C. Kübler, T. Santini, and E. Kasneci, "Automatic generation of saliency-based areas of interest for the visualization and analysis of eye-tracking data." in *VMV*, 2018, pp. 47–54.

[7] F. Cristino, S. Mathôt, J. Theeuwes, and I. D. Gilchrist, "Scanmatch: A novel method for comparing fixation sequences," *Behavior research methods*, vol. 42, pp. 692–700, 2010.

[8] N. Castner, E. Kasneci, T. Kübler, K. Scheiter, J. Richter, T. Eder, F. Hüttig, and C. Keutel, "Scanpath comparison in medical image reading skills of dental students: distinguishing stages of expertise development," in *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, 2018, pp. 1–9.

[9] T. C. Kübler, E. Kasneci, and W. Rosenstiel, "Subsmatch: Scanpath similarity in dynamic scenes based on subsequence frequencies," in *Proceedings of the Symposium on Eye Tracking Research and Applications*, 2014, pp. 319–322.

[10] D. Geisler, N. Castner, G. Kasneci, and E. Kasneci, "A minhash approach for fast scanpath classification," in *ACM Symposium on Eye Tracking Research and Applications*, 2020, pp. 1–9.

[11] L. G. C. Bautista and P. C. Naval, "Gazemae: general representations of eye movements using a micro-macro autoencoder," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 7004–7011.

[12] S. Thentu and N. Attar, "Investigating classification methods using fixation patterns to predict visual tasks," *IFAC-PapersOnLine*, vol. 55, no. 29, pp. 19–24, 2022.

[13] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *Proceedings of the 2000 symposium on Eye tracking research & applications*, 2000, pp. 71–78.

[14] A. Santella and D. DeCarlo, "Robust clustering of eye movement recordings for quantification of visual interest," in *Proceedings of the 2004 symposium on Eye tracking research & applications*, 2004, pp. 27–34.

[15] C. Truong, L. Oudre, and N. Vayatis, "Selective review of offline change point detection methods," *Signal Processing*, vol. 167, p. 107299, 2020.

[16] R. Killick, P. Fearnhead, and I. A. Eckley, "Optimal detection of changepoints with a linear computational cost," *Journal of the American Statistical Association*, vol. 107, no. 500, pp. 1590–1598, 2012.

[17] R. A. Wagner and M. J. Fischer, "The string-to-string correction problem," *Journal of the ACM (JACM)*, vol. 21, no. 1, pp. 168–173, 1974.

[18] W. Fuhl, E. Bozkir, B. Hosp, N. Castner, D. Geisler, T. C. Santini, and E. Kasneci, "Encodji: encoding gaze data into emoji space for an amusing scanpath classification approach," in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, 2019, pp. 1–4.

[19] I. Borg and P. J. Groenen, *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media, 2005.