Graph-based clustering for identifying region of interest in eye tracker data analysis

Kanghang He, Cheng Yang, Vladimir Stankovic, and Lina Stankovic
Department of Electronic and Electrical Engineering
University of Strathclyde, Glasgow, G1 1XW, UK
Email: {kanghang.he, cheng.yang, vladimir.stankovic, lina.stankovic}@strath.ac.uk

Abstract—Localization of a viewer's region of interest (ROI) on eye gaze signal trajectories acquired by eye trackers is a widely used approach in scene analysis, image compression, and quality of experience assessment. In this paper, we propose a novel clustering approach for ROI estimation from potentially noisy raw eye gaze data, based on signal processing on graphs. The clustering approach adapts graph signal processing (GSP)-based classification by first cleverly selecting a starting data sample, and then classifying the remaining samples. Furthermore, Graph Fourier Transform is used to adjust GSP parameters on-the-fly to maximise accuracy. Experimental results show competitive clustering accuracy of our proposed scheme compared to Density-based spatial clustering of applications with noise (DB-SCAN), Distance-Threshold Identification (I-DT), and Mean-Shift on publicly available Shape Dataset and the potential of estimating ROI accurately on true eye tracker data[1].

I. Introduction

When an image or scene is viewed, the eye gaze tends to pause on small regions within the image, called fixation areas. On average, fixations last for around 200 ms during the reading of linguistic text, and 350 ms during the viewing of a scene [1]. Existing approaches for detecting Region of Interest (ROI) in the viewed image first represent the centre of a fixation area as a fixation point [2], and then use clustering to group these fixation points from all fixation areas into spatial regions, identified as ROI. Various clustering approaches have been used to detect ROI, such as k-means and distance threshold [3], [4], Density-based spatial clustering of applications with noise (DB-SCAN) [5], Distance-Threshold Identification (I-DT) [6] and Mean-shift [7]. The gaze data, acquired by commercial eye trackers, is normally affected by high level of measurement noise and contains missing data due to eye blinks and occasional head movements. This motivates the use of Graph Signal Processing (GSP), an emerging field used to represent irregular data structures on graphs [8], [9], for robust gaze data clustering.

GSP is proposed for dataset classification in [10], where each dataset sample is associated with a graph vertex. The underlying graph is then designed to capture dependency between the data samples, by connecting samples/vertices that are highly correlated with high-weight edges. GSP-based classification is competitive to other machine learning based classification approaches, such as Support Vector Machines, when the data samples are noisy and/or training dataset is of poor quality or size, since GSP generates a graph based on intuition instead of relying on training (see [11]). GSP-based classification is used for image and text classification [10], energy disaggregation [12], motion classification [13], and many other signal/image processing tasks (see [9], [12], [14] and references therein). In [14] the GSP-based classification is extended to clustering for energy disaggregation by searching for vertices that have high similarities and grouping them into a cluster.

To identify ROI from noisy eye tracker data, in this paper, we first pre-process the eye tracker measurements, comprising the time-stamped eye gaze locations in the viewed image, by filtering the data to ensure convergence to locations of higher density, similarly to [7], and then cleverly choosing a data sample as a starting point for clustering. Then we propose a GSP-based iterative clustering method, for spatial clustering of pre-processed eye tracker data to detect ROI. Clustering is performed on the graph, where each graph vertex is associated to one spatial gaze measurement, that defines horizontal and vertical position of the gaze, and weights of the edges reflect spatial correlation between the measurements.

Note that, we focus on detection of ROI in still images, where an ROI is a group of gaze measurement spatially concentrated regardless of the time information [7]. Having this in mind, and due to GSP's resilience to noise, we bypass the traditional step of first finding time-dependent fixation points prior to ROI detection, making the proposed method robust to timing jitter and synchronization problems. Moreover, the method is inherently robust to measurement noise and eye blinks, and no denoising or data cleaning steps, common for eye tracker data processing, are needed.

Section II describes the proposed method, including pre-processing, selection of the starting point for iterative clustering, the proposed GSP-based clustering algorithm, and autonomously optimizing clustering parameters. Section III presents experimental results of the proposed clustering algorithm on a public dataset and captured eye tracker data, benchmarked using three state-of-the-art clustering methods for ROI detection. The last section concludes and highlights future work.

II. Proposed Method

A. System Overview

Let N be the total number of samples from eye tracker gaze data, and $(\mathbf{x}, \mathbf{y}) = \{(x_1, y_1), ..., (x_N, y_N)\}$ be the spatial locations of the samples in the viewed image. The objective is to group samples into clusters $m \in \{1, \ldots, M\}$, where M is the total number of clusters which is unknown. Let $\mathbf{c} = \{c_1, \ldots, c_N\}$ where $c_i = m$ if sample (x_i, y_i) belongs to cluster m.

Fig. 1 shows the block diagram of the proposed GSP-based clustering method. First, similarly to [7], we perform preprocessing by shifting the input eye gaze data to make all samples move simultaneously towards locations of higher density (Section II-B). Then, one sample is cleverly chosen as the starting point to clustering (Section II-C). Next, a binary GSP-based classification is performed, similarly to [10] and [14], to classify all data samples into one of two classes: belonging to the same class as the starting point or not (Section II-C). All samples classified to the class of the starting point form Cluster $m = 1$, and are removed from the dataset, a new starting point is chosen, m is incremented, and the process is repeated until no samples remain unclustered.

The cluster accuracy a (Section II-E), is then used to evaluate the *quality* of the formed clusters. If the accuracy improved compared to the previous iteration, then the underlying graph used for GSP processing is adjusted, the clustering labels are reset and the process is repeated until accuracy cannot be further improved.
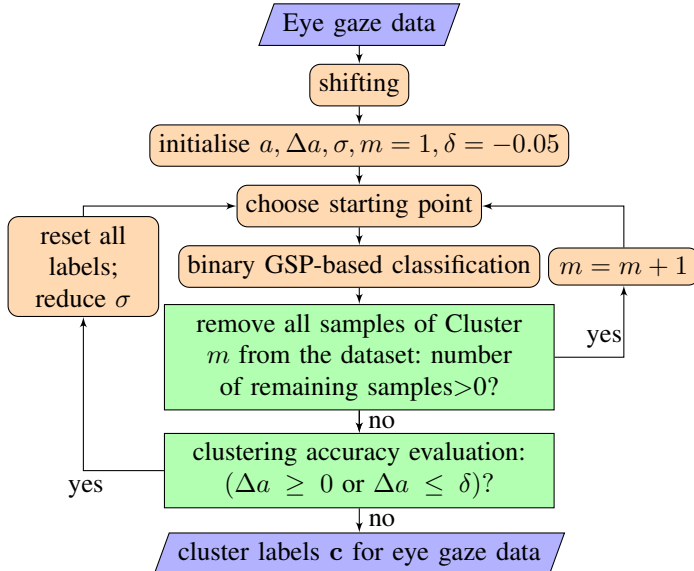

Fig. 1. The proposed GSP-based clustering algorithm.

The pseudocode of the proposed method is shown in Algorithm 1. The following subsections provide a detailed description of each step.

B. Shifting Samples

Before clustering, a preprocessing operation, *shifting*, is performed to make clustering robust to outliers, e.g., sac-

cade points. This shifting step aims to move all samples to higher density locations, making samples in each cluster more compact. Let $(\mathbf{x}^*, \mathbf{y}^*) = \{(x_1^*, y_1^*), ..., (x_N^*, y_N^*)\}$ denote the shifted data. For each sample (x_i, y_i), we define (x_i^*, y_i^*) as:

$$x_i^* = \frac{\sum_{n=1}^b x_n}{b}, \tag{1}$$

$$y_i^* = \frac{\sum_{n=1}^b y_n}{b}, \tag{2}$$

where b is the number of most correlated neighbours to the vertex v_i, i.e., we take the b highest-weight connections to vertex v_i. Fig. 2 shows an example of the raw data samples and pre-processed, shifted samples. It is clear that the shifted points become more concentrated in each cluster.
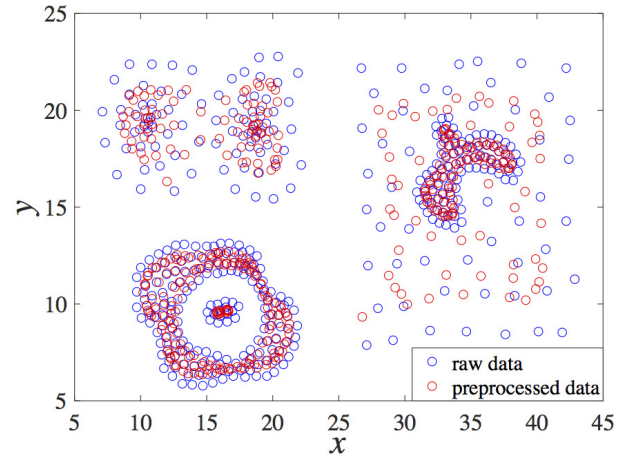

Fig. 2. Comparison between shifted, preprocessed data and raw data.

C. Setting the Starting Point

Many ROI detection methods are based on randomly choosing the starting point (see [7] and references therein); hence bad initial positions (e.g., saccade points) will rapidly reduce the clustering accuracy. We thus propose a method to refine the initial, random selection of the starting point. Let (x_s, y_s) be a randomly chosen starting point. The shifted starting point x_s^* and y_s^* is calculated and updated using (1), (2) and until the difference between the points before and after shifting cannot be reduced anymore. The refined starting point $(\hat{x}_s, \hat{y}_s)$ is the nearest sample to the final (x_s^*, y_s^*). Note that this way we ensure high point density around the selected starting point, hence the starting points is unlikely to be an outlier.

D. GSP-based Clustering

We use binary GSP-based classification to find samples that belong to the same class as $(\hat{x}_s, \hat{y}_s)$. We do this by first constructing a graph $\mathcal{G} = (\mathcal{V}, \mathbf{A})$, where for $i = 2, \ldots, N + 1$ each vertex $v_i \in \mathcal{V}$ is associated with one data sample (x_i^*, y_i^*), v_1 is associated with $(\hat{x}_s, \hat{y}_s)$, and $\mathbf{A}$ is the weighted adjacency matrix of $\mathcal{G}$ [9]. As commonly done in the GSP literature [9], each entry $A_{i,j}$ of $\mathbf{A}$, *i.e.*, the edge weight between nodes v_i

and v_j, is defined using Euclidean distance with a Gaussian kernel:

$$A_{i,j} = \exp\left\{ -\frac{(x_i^* - x_j^*)^2 + (y_i^* - y_j^*)^2}{\sigma^2} \right\}, \qquad (3)$$

where σ is a scaling factor. Next, a *graph Laplacian* is defined as follows:

$$\mathbf{L} = \mathbf{D} - \mathbf{A}, \qquad (4)$$

where $\mathbf{D}$ is a degree matrix and $D_{k,k} = \sum_{j=1}^{N} A_{j,k}$.

Starting from $m = 1$, we assume $\hat{x}_s$ and $\hat{y}_s$ belong to Cluster m and classify the remaining N samples as belonging to Cluster m or not. In particular, we initialize an $(N + 1)$-length vector ϱ as graph signal: $\varrho = [\varsigma \ \mathbf{s}^m]^\top$, where $\varsigma = 1$ is associated with starting point $(\hat{x}_s, \hat{y}_s)$, $\mathbf{s}^m$ is an N-length row vector initialised as all zeros.

Similar to [12], we adopt a Laplacian regularizer $\varrho^\top \mathbf{L}\varrho$ to measure the variation in signal $\mathbf{s}^m$ with respect to the underlying graph, with the objective to find an $\mathbf{s}^{m*}$ that minimizes the variation in the graph signal. The optimization problem

$$\mathbf{s}^{m*} = \arg\min_{\mathbf{s}^m} ||\varrho^\top \mathbf{L}\varrho||_2^2 \qquad (5)$$

has the following closed-form solution [15], [16], [17]:

$$\mathbf{s}^{m*} = -\mathbf{L}^\#_{2:N+1,2:N+1} \varsigma \mathbf{L}^\top_{1,2:N+1}, \qquad (6)$$

where $\mathbf{L}^\#_{2:N+1,2:N+1}$ is the pseudo-inverse of $\mathbf{L}_{2:N+1,2:N+1}$, and $s^{m*} \in [0, 1]$. If s_i^{m*} is close to 1 (based on a heuristically set distance threshold), we designate that (x_i^*, y_i^*) belongs to the same cluster m as $(\hat{x}_s, \hat{y}_s)$, which we label as $c_i = m$.

Next, we remove all clustered points, increment m and repeat the procedure starting with selecting a new starting point (Sec. II-C) until all samples are labelled with a cluster number. Finally, the clusters that contain small fraction of samples, where the fraction parameter is denoted as α, are assumed to be outliers and all grouped to Cluster 0. The cluster with coordinate $(0, 0)$ is also labelled as 0 since it contains all lost data caused by eye blinks and noise.

E. Self Parameter Tuning

Initial testing shows large dependency of the accuracy of the results on the scaling factor σ defined in (3) that weights the relationship between the data samples. Large σ leads to large $A_{i,j}$ indicating high correlation between the samples i and j, which would result in many sample points being clustered together. Low σ has the opposite effect: small clusters would be formed comprising only highly correlated samples.

Since the best σ, that is, the one that maximizes accuracy, is signal dependent, we propose a method for finding the optimal σ based on the signal samples. First, σ is set to be a very high value, which leads to rough clustering (i.e., a small number of large clusters) for the giving dataset. After all samples are labelled following the procedure from the previous subsection, we define a graph signal for Cluster m, $\mathbf{g}^m$ as:

$$g_i^m = \mathbb{1}(c_i == m). \qquad (7)$$

where $\mathbb{1}$ is an indicator function that returns 1 if the condition is true and 0 otherwise. We then calculate the graph Laplacian $\mathbf{L_G}$, which is symmetric, thus the signal value decomposition (SVD) of $\mathbf{L_G}$ is given by [18]:

$$\mathbf{L_G} = \mathbf{U\Lambda U}^\top, \qquad (8)$$

where $\mathbf{\Lambda}$ is a diagonal matrix with $\{\lambda_0, \lambda_1, \ldots, \lambda_{N-1}\}$ as eigenvalues of $\mathbf{L_G}$ on the diagonal, and $\mathbf{U}$ is a set of eigenvectors. The Graph Fourier Transform (GFT) of $\mathbf{g}^m$ is then given by:

$$\hat{\mathbf{g}^m} = \mathbf{U}^\top \mathbf{g}^m. \qquad (9)$$

The eigenvalues of $\mathbf{L_G}$ act as the graph frequencies and corresponding eigenvectors act as the graph harmonics [19], [20], [21]. Small λ's carry information about low frequency components of the signal, while high frequencies (details) are carried by large λ's. Motivated by the fact that high energy in the high frequencies indicate bad cluster quality, we estimate the frequency content of $\hat{g}_i^m$ as follows. Let $f = (\lambda_0 + \lambda_{N-1})/2$ and $j^* = \arg\min_j |f - \lambda_j|$, then λ_i for $i \leq j^*$ would carry information about energy content in the lower half of the frequency spectrum.

Let r_m be the ratio of the total number of low/high frequency components in $\hat{\mathbf{g}^m}$ that are above/below a heuristically set threshold γ a *i.e.*,

$$r_m = \frac{\sum_{i=j^*+1}^{N} \left(\mathbb{1}(|\hat{g}_i^m| > \gamma)\right)}{\sum_{i=1}^{j^*} \left(\mathbb{1}(|\hat{g}_i^m| > \gamma)\right)}. \qquad (10)$$

$r_m > \tau$ indicates a good cluster, where τ is a chosen parameter. If not, all samples in this cluster are considered as incorrectly clustered. In addition, the samples with cluster label equal to 0 are also counted as incorrect samples. The estimated clustering accuracy is calculated as:

$$a = 1 - \frac{\kappa}{N}, \qquad (11)$$

where κ is the total number of incorrectly clustered samples that is given by:

$$\begin{aligned}
\kappa = &\sum_{i=1}^{N} \left(\mathbb{1}(x_i^* = 0 \ \& \ y_i^* = 0)\right) \\
&+ \sum_{i=1}^{N}\sum_{m=1}^{M} \left(\mathbb{1}\left((c_i = m) \ \& \ \left(\frac{\sum_{i=1}^{N}\left(\mathbb{1}(c_i = m)\right)}{N} \leq \alpha\right)\right)\right) \\
&+ \sum_{i=1}^{N}\sum_{m=1}^{M} \left(\mathbb{1}\left((c_i = m) \ \& \ (r_m \leq \tau)\right)\right),
\end{aligned}$$

$$(12)$$

where the first line captures all samples in Cluster 0, the second line includes clusters that have very low number of samples (below α), and the third line comes from all clusters that give $r_m < \tau$. The scaling factor σ is reduced by small decrements until there is no improvement in a anymore.

Algorithm 1: Proposed GSP-based spatial clustering.

Input: $(\mathbf{x}, \mathbf{y})$;
Output: $\mathbf{c}$;

1 **set** x_i^* via (1), y_i^* via (2), $a = 0$, $\Delta a = 0$, $\sigma = 20$, $\beta = 0.99$, $\tau = 5$, $b = 10$, $\delta = -0.05$, $\alpha = 0.03$;
2 **while** $\Delta a \geq 0$ *or* $\Delta a \leq \delta$ **do**
3   **set** $\sigma = \sigma - 1$, $m = 1$, $\mathbf{c} = 0$;
4   **while** *number of remaining samples > 0* **do**
5   **randomly set** (x_s, y_s);
6   **compute** $(\hat{x}_s, \hat{y}_s)$ as in Sec. II-C.;
7   **compute** $\mathbf{A}$, $\mathbf{L}$ with $(\hat{x}_s, \hat{y}_s)$, $(\mathbf{x}^*, \mathbf{y}^*)$, (3), (4);
8   **compute** $\mathbf{s}^{m*}$ via (6);
9   **set** $\mathbf{c}($find $\mathbf{s}_i^{m*} > \beta) = m$, $m = m + 1$;
10   **remove from the dataset** samples i with $\mathbf{s}_i^{m*} > \beta$;
11   **compute** Δa via (10), (11), (12);
12 **return** $\mathbf{c}$;

III. EXPERIMENTAL RESULTS

In this section, we first validate the proposed spatial clustering algorithm on a public clustering dataset with known ground-truth labels, and show how the proposed clustering algorithm compares with DB-SCAN, I-DT and Mean-shift algorithms. Then we present the results with true eye tracker data to compare the accuracies of the four aforementioned methods in detecting ROI. Table I shows all parameters used for the proposed method in all experiments, which were heuristically obtained and are used for all datasets.

TABLE I
PARAMETER SETTINGS FOR THE PROPOSED METHOD USED IN ALL THE EXPERIMENTS.

symbol	parameter	setting
b	neighbouring samples	10
τ	cluster quality	5
σ	scaling factor for $\mathbf{A}$	initially set 20
α	sample fraction	0.03
β	labelling threshold	0.99
δ	accuracy difference threshold	-0.05
γ	frequency response	$\frac{1}{2}\max\{|g_i^{\hat{m}}|\}$

A. Results with Shape Dataset

The four algorithms are first tested on the public *Shape dataset* [22], which is often used to assess accuracy of spatial clustering methods. The images in this dataset are scatter diagrams with labels, indicating the cluster index for each point, where the points close to each other are assigned to the same cluster.

Table II shows the clustering results of the four methods on the Shape dataset. The accuracy is measured as a ratio of the number of correctly clustered samples to the total number of samples. In DB-SCAN, the minimum number of points required to form a cluster and the number of neighbourhood samples of a point are denoted as ϵ and ζ, respectively. The distance threshold in I-DT is denoted as η. The distance threshold and the number of neighbourhood samples of a point

TABLE II
CLUSTERING ACCURACY OF THE PROPOSED METHOD, DB-SCAN, I-DT AND MEAN-SHIFT WITHOUT PREPROCESSING.

parameter	Proposed	DB-SCAN	I-DT	Mean-shift
	Self-adaptable σ	$\epsilon = 2$ $\zeta = 14$	$\eta = 5$	$\sigma_s = 7$ $\rho = 10$
R15	0.89	0.53	0.13	0.20
D31	0.76	0.06	0.30	0.35
Aggregation	0.79	0.88	0.87	0.89
Toy	0.30	0.91	0.46	0.48
Compound	0.84	0.90	0.69	0.78
Pathbased	0.80	0.85	0.67	0.66
Flame	0.95	0.35	0.95	0.91
Mean	**0.76**	**0.64**	**0.58**	**0.61**

in Mean-shift are denoted as σ_s and ρ, respectively. Parameters in DB-SCAN, I-DT and Mean-shift are tuned and fixed using a greedy search scheme to get best performance for the whole dataset. The proposed approach can self tune itself to find the best parameters for each image pattern.

The performance of DB-SCAN on some images, such as *Compound*, is good. There are many individual points that are isolated in *Compound* dataset. They are all clustered as a single cluster in ground truth which DB-SCAN consider them as noise. However, since DB-SCAN is a density-based spatial clustering, it cannot adapt well to different point density characteristics of the images, leading to poor performance in some cases. I-DT is a distance-based method, where results are highly influenced by the size of clusters in the image. In Mean-shift all points are repetitively moved until converged to positions with high density [7]. Then, a distance threshold is applied to cluster the shifted points, while the size of the clusters is depended on σ_s. Thus, the overall performance is poor due to variations in cluster sizes across the images.

For some images, the ground-truth clusters are connected with consecutive points. In GSP-based clustering, these clusters will be treated as piecewise smooth signals since the weight is defined based on the distance between samples, and thus these clusters are incorrectly merged into the same cluster. Shift pre-processing can move these connecting points closer to their closest high density centres, disconnecting in this way the clusters, and leading to more effective clustering.

Table III shows the results of our proposed GSP-based clustering method compared with DB-SCAN, I-DT and Mean-Shift after shift preprocessing is applied on the images prior to running all 4 clustering algorithms. We again use a greedy search scheme to get the optimal parameters for all competing schemes. Overall performance for all methods except Mean-shift has improved compared to clustering the raw data without pre-processing. This proves that the shift preprocessing can significantly improve the clustering accuracy. For Mean-shift, the preprocessing does not improve the performance, since the effect of the proposed shift preprocessing is similar to the operation that is already done in the Mean-shift. Indeed, Mean-shift uses the weighted mean of nearby points based on the kernel function to make the samples compact.

B. Results on the Eye Tracker Dataset

The algorithms are also tested on true eye tracker data recorded by Eye Tribe [23] at sampling rate of 30Hz, to

	Proposed	DB-SCAN	I-DT	Mean-shift
parameter	Self-adaptable σ	$\epsilon = 1$ $\zeta = 5$	$\eta = 5$	$\sigma_s = 7$ $\rho = 10$
R15	0.99	0.53	0.52	0.20
D31	0.93	0.23	0.48	0.35
Aggregation	0.95	0.97	0.64	0.90
Toy	0.93	0.90	0.33	0.47
Compound	0.73	0.86	0.92	0.79
Pathbased	0.77	0.73	0.63	0.68
Flame	0.98	0.93	0.97	0.91
Mean	**0.90**	**0.74**	**0.64**	**0.61**

assess the accuracy of the methods in detecting ROI in the viewed images. Experiments are performed in a laboratory with moderate artificial light conditions, which remain unchanged for the duration of all trials. Ten subjects participated in the experimens, aged between 25 and 45 years old, both male and female, all with normal vision. The subjects were sitting in front of a DELL P2414 screen with a resolution of 1920x1080 pixels, at about 70 cm distance from the eye tracker, which is located under the screen. Calibration was performed using OGAMA [24] calibration process, whereby subjects are asked to follow a coloured dot moving in the corners and centre of the screen. This calibration process was included before each trial. Note that OGAMA is an open source software for recording and analyzing eye gaze and mouse data for experiments with screen based slide show stimuli. OGAMA does not generate ROI information.

Two different experiments are set. In Experiment 1, four objects are displayed on a blank white-coloured background and shown to viewers for 5 seconds. Two slides are shown:

1) A white/Blank background image with four words sparsely distributed.
2) A white/Blank background image with four small coloured icons spread out across the slide.

The viewers are asked to focus their attention on the four objects, one at the time. Hence, the clustering algorithms should result in four distinct clusters each pointing to one object. Examples of the ROI identification with the proposed approach overlapped with the displayed image is shown in Fig. 3. The ellipses are drawn to emphasise all samples of a cluster to make the visual clustering results clearer.
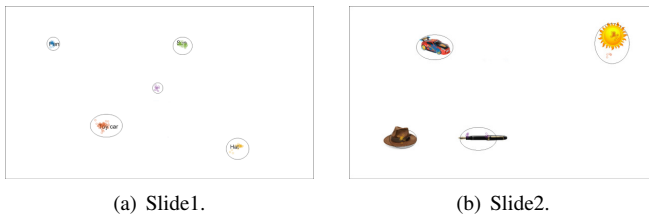

(a) Slide1.   (b) Slide2.

Fig. 3. ROI estimation validation in Experiment 1 for Subject 2 with (a) words distributed, (b) icons distributed (Enlarge slightly in colour).

The comparison results between the four methods are shown in Table IV, where CD is the number of correctly detected ROIs and ID stands for the number of incorrectly detected

ROIs. Both CD and ID are averaged over all subjects. $CD = 20$ if all ROIs are detected correctly. An ROI is correctly detected if at least half of the samples in the resulting cluster overlap with the target object, and there are no other clusters that overlap with the target.

One can see from the table, that the proposed method leads to the highest CD and lowest ID indicating the highest accuracy. Generally, all methods perform well, since the objects are clear, the background is white, and the objects are far away from one another. In order to test the ROI detection accuracy in a more challenging scene, we set Experiment 2.

	Proposed		DB-SCAN		I-DT		Mean-shift	
	CD	ID	CD	ID	CD	ID	CD	ID
Slide1	18	3	14	6	14	7	15	8
Slide2	19	2	15	5	16	7	13	6

In Experiment 2, 10 slides are shown to the subjects, all full of icons (around 70) [25]:

1) Slide1. Blank background; 2 sec on each 4 target icons.
2) Slide2. Blank background; 5 sec on each 4 target icons.
3) Slide3. Blank background and the target icons are very small (compared to other icons in the image).
4) Slide4. Blank background and the target icons are large.
5) Slide5. Blank background; and the whole slide is noisy.
6) Slide6. Nature image as background; 2 sec on each 4 target icons.
7) Slide7. Nature image as background; 5 sec on each 4 target icons.
8) Slide8. Nature image as background and the target icons are very small.
9) Slide9. Nature image as background and the target icons are very large.
10) Slide10. Nature image as background and the whole slide is noisy.

The subjects are informed about the positions of the target icons in the slides before the experiment. During the experiment, the subjects are told to focus their attention on those icons. The ROI will be the target icons that the subjects are asked to focus on. The saccades while finding the target icons are noise. The numerical comparison results are shown in Table V. Figs. 4 and 5 show two examples obtained with the proposed method.

If all ROIs are correctly detected and no redundant ROIs are found, $CD = 30$ and $ID = 0$. Table V indicates that the proposed GSP-based clustering method gives highly accurate ROI detection results in all situations. The incorrectly detected ROI are very few which shows the competitiveness of the proposed method. DB-SCAN and I-DT cannot adapt to the changes in the slides, producing often poor results. Only considering the density or distance is the main reason why DB-SCAN and I-DT cannot provide as good results as the

	Proposed		DB-SCAN		I-DT		Mean-shift	
	CD	ID	CD	ID	CD	ID	CD	ID
Slide1	28	1	24	12	21	5	25	5
Slide2	29	2	25	10	22	10	22	8
Slide3	26	1	20	12	18	19	20	7
Slide4	27	4	21	9	17	21	15	25
Slide5	29	2	20	5	19	6	21	3
Slide6	27	3	18	11	22	10	19	6
Slide7	30	1	20	13	21	9	19	7
Slide8	28	2	15	11	19	11	17	9
Slide9	28	3	14	10	21	7	11	23
Slide10	26	2	22	7	21	9	21	9
Mean	**27.8**	**2.1**	**19.9**	**10**	**20.1**	**10.7**	**19**	**10.2**

proposed method. As for Mean-shift, the results are relatively better than DB-SCAN and I-DT for most slides except Slides 4 and 9. The target icons in these two slides are much larger than other icons. Therefore the size of ROIs are also relative large. Mean-shift incorrectly breaks the ROI into more than one cluster.
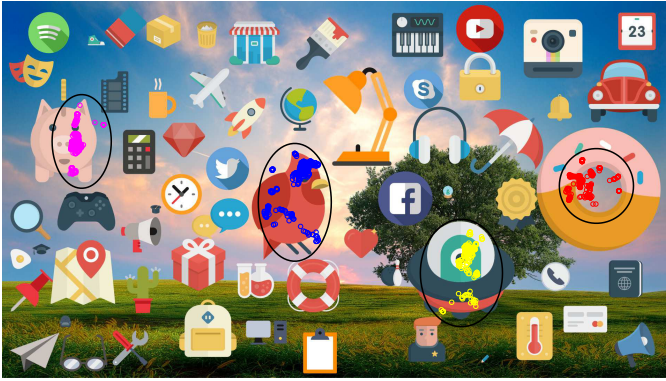

Fig. 4. ROI estimation validation for Subject3 and Slide9.

Fig. 5. ROI estimation validation for Subject3 and Slide10.

IV. CONCLUSION

This paper proposes a spatial clustering method for ROI detection using the emerging concept of GSP. A shift preprocessing approach is utilised to further improve the clustering accuracy. Graph Fourier Transform is applied to evaluate the cluster quality, and thereby adjust the GSP parameter. The proposed method can provide highly accurate clustering results on public shape clustering dataset. It also shows excellent ROI detection performance for true eye tracker data in a range of challenging scenes. Future work will consist of further improving ROI detection accuracy by adding another term in adjacency matrix definition and applying time in the graph weight to detect fixation points.

REFERENCES

[1] X. Chen and Z. Chen, "Visual attention identification using random walks based eye tracking protocols," in *2015 IEEE Global Conf. Signal and Inform. Proc. (GlobalSIP)*, Dec 2015, pp. 6–9.

[2] D. D. Salvucciand and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *The 2000 ACM Symp. Eye Tracking Research & Applications*, 2000, pp. 71–78.

[3] C. M. Privitera and L. W. Stark, "Algorithms for defining visual regions-of-interest: comparison with eye fixations," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 22, no. 9, pp. 970–982, Sep 2000.

[4] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.

[5] M. Ester, H. Kriegel, J. Sander, and X.Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *KDD*, vol. 96, no. 34, 1996, pp. 226–231.

[6] O. Špakov and D. Miniotas, "Application of clustering algorithms in eye gaze visualizations," *Inform. Techn. and Control*, vol. 36, no. 2, 2015.

[7] A. Santella and D. DeCarlo, "Robust clustering of eye movement recordings for quantification of visual interest," in *The 2004 ACM Symp. Eye Tracking Research & Applications*, 2004, pp. 27–34.

[8] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Proc.*, vol. 61, pp. 1644–1656, Apr. 2013.

[9] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Sig. Proc. Magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[10] A. Sandryhaila and J. Moura, "Classification via regularization on graphs." in *GlobalSIP*, 2013, pp. 495–498.

[11] C. Zhang, D. Florêncio, and P. A. Chou, "Graph signal processing–a probabilistic framework," *Microsoft Res., Redmond, WA, USA, Tech. Rep. MSR-TR-2015-31*, 2015.

[12] K. He, L. Stankovic, J. Liao, and V. Stankovic, "Non-intrusive load disaggregation using graph signal processing," *IEEE Transactions on Smart Grid*, to be published, 2016.

[13] C. Yang, A. Kerr, V. Stankovic, L. Stankovic, P. Rowe, and S. Cheng, "Human upper limb motion analysis for post-stroke impairment assessment using video analytics," *IEEE Access*, vol. 4, pp. 650–659, 2016.

[14] B. Zhao, L. Stankovic, and V. Stankovic, "On a training-less solution for non-intrusive appliance load monitoring using graph signal processing," *IEEE Access*, vol. 4, pp. 1784–1799, 2016.

[15] C. Yang, Y. Mao, G. Cheung, V. Stankovic, and K. Chan, "Graph-based depth video denoising and event detection for sleep monitoring," in *IEEE Int. Workshop Multimedia Signal Proc. (MMSP)*, Sept 2014, pp. 1–6.

[16] C. Yang, G. Cheung, V. Stankovic, K. Chan, and N. Ono, "Sleep apnea detection via depth video and audio feature learning," *IEEE Trans. Multi.*, vol. 19, pp. 822–835, Apr. 2017.

[17] S.Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[18] G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," *Numerische mathematik*, vol. 14, pp. 403–420, 1970.

[19] R. Singh, A. Chakraborty, and B. S. Manoj, "Graph fourier transform based on directed laplacian," in *2016 Int. Conf. Signal Processing and Commun. (SPCOM)*, June 2016, pp. 1–5.

[20] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Processing*, vol. 62, no. 12, pp. 3042–3054, June 2014.

[21] A. Sandryhaila and J. Moura, "Discrete signal processing on graphs: Graph fourier transform," in *2013 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, May 2013, pp. 6167–6170.

[22] *Clustering datasets*. [Online]. Available: https://cs.joensuu.fi/sipu/datasets/

[23] *The eye tribe*. [Online]. Available: https://theeyetribe.com/

[24] *OGAMA open gaze and mouse analyzer*. [Online]. Available: http://www.ogama.net/

[25] *Flaticons*. [Online]. Available: https://www.flaticon.com/