Region of interest generation algorithms for eye tracking data

Wolfgang Fuhl
University Tuebingen
Tuebingen, Baden-Wuerttemberg
wolfgang.fuhl@uni-tuebingen.de

Thomas Kuebler*
University Tuebingen
Tuebingen, Baden-Wuerttemberg
thomas.kuebler@uni-tuebingen.de

Hanna Brinkmann
University Vienna
Vienna, Vienna
hanna.brinkmann@univie.ac.at

Raphael Rosenberg
University Vienna
Vienna, Vienna
raphael.rosenberg@univie.ac.at

Wolfgang Rosenstiel
University Tuebingen
Tuebingen, Baden-Wuerttemberg
Wolfgang.Rosenstiel@uni-tuebingen.de

Enkelejda Kasneci
University Tuebingen
Tuebingen, Baden-Wuerttemberg
Enkelejda.Kasneci@uni-tuebingen.de

ABSTRACT

Using human fixation behavior, we can interfere regions that require to be processed at high resolution and where stronger compression can be favored. Analyzing the visual scan path solely based on a predefined set of regions of interest (ROIs) limits the exploration room of the analysis. Insights can only be gained for those regions that the data analyst considered worthy of labeling. Furthermore, visual exploration is naturally time-dependent: A short initial overview phase may be followed by an in-depth analysis of regions that attracted the most attention. Therefore, the shape and size of regions of interest may change over time. Automatic ROI generation can help in automatically reshaping the ROIs to the data of a time slice. We developed three novel methods for automatic ROI generation and show their applicability to different eye tracking data sets. The methods are publicly available as part of the EyeTrace software at http://www.ti.uni-tuebingen.de/Eyetrace.1751.0.html

CCS CONCEPTS

• **Human-centered computing** → **Visualization design and evaluation methods**; *Visualization toolkits*; • **Applied computing** → Fine arts;

KEYWORDS

ROI generation, AOI generation, Eye tracking analysis, Visualization, Expertise classification, Abstract art

ACM Reference Format:
Wolfgang Fuhl, Thomas Kuebler, Hanna Brinkmann, Raphael Rosenberg, Wolfgang Rosenstiel, and Enkelejda Kasneci. 2018. Region of interest generation algorithms for eye tracking data. In *ETVIS'18: 3rd Workshop on Eye Tracking and Visualization , June 14–17, 2018, Warsaw, Poland.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3205929.3205937

*Work of the author is supported by the Institutional Strategy of the University of TÄijbingen (Deutsche Forschungsgemeinschaft, ZUK 63)

ETVIS'18, June 14–17, 2018, Warsaw, Poland

ACM ISBN 978-1-4503-5787-6/18/06…$15.00
https://doi.org/10.1145/3205929.3205937

1 INTRODUCTION

Presently, eye tracking can be found in different research areas, such as human-computer interaction, medicine, neuroscience, psychology, and many more [Duchowski 2002; Toker et al. 2013]. While one focus of eye tracking research addresses human behavior, another growing field is its application. These applications allow for control and comfort while the subject while the subjects interacts unconstrained with the world. For both research and application, large amounts of data are handled [Blascheck et al. 2014]. These quantities continue to grow due to better and cheaper devices that allow for large scale studies: With more subjects and longer recording durations. But in order to gain insights from these enormous data collections, we need to reduce the complexity.

The first structuring of eye tracking data was done in the years 1879 to 1920 [Duchowski 2002], where the raw gaze points were grouped by their eye movement behaviors into fixations and saccades. This structuring process led to the finding that certain eye movement patterns indicate, for example, the task performed by a subject [Tatler et al. 2010]. These patterns consist of a sequence of fixations, where the eye is held still and directed towards the perceived areas, and saccades, or fast eye movements, during which the visual input is suppressed.

A common method to further simplify the data is performed by defining regions of interest (ROI) or areas of interest (AOI), i.e., areas with a specific semantic meaning. The terms will be used synonymously throughout this paper. The semantic meaning of ROIs are usually determined as having special interest to the researcher. For instance, studies on web page organization and design [Goldberg et al. 2002; Pan et al. 2004] as well as graph reading [Strobel et al. 2018], product design [Mawad et al. 2015] and dwell time on facial areas for children with autism [Auyeung et al. 2015] have relied on ROI definitions to interpret eye movement behavior. For data analysis and statistics, information for each ROI can be performed separately: It can be, though is not limited to, average dwell time, or the number of fixations on the ROI. Also, connections between ROIs, such as transition probabilities, can be investigated.

Manual ROI annotation is necessarily a subjective step. Given the large amount of data and the desire to analyze changes in ROIs and ROI shapes associated with specific time segments, it quickly becomes laborious. Therefore, automatic ROI generation based on the data of different viewers and variable time segments is a useful and supportive automation process. It not only supports the

researcher, but also allows for determining ROIs in a data-driven, objective way: For instance, ROI difference comparison between subjects, groups of subjects, or different time slices. Analysis techniques based on ROIs can be extended by the automation such as the development of circular ROI transitions [Blascheck et al. 2013] which can be inspected over time and visualized as a video.

Identifying fixations and saccades, clustering fixation locations, and determining ROIs are important algorithmic steps that have a major influence on key eye tracking metrics, such as the average fixation duration or the proportion of gaze directed towards a ROI. Therefore, we believe that a researcher should always be aware of the implications of one of these specific analysis steps. In order to bring awareness researchers on the effects algorithm choice and parameterization, we provide implementations as well as an overview of several algorithms for generating ROIs and compare them to manually labeled ROIs. This work focuses on generating ROIs in an automated, data-driven way. Three new methods, two that generate ROIs from a heatmap representation and a new overlap clustering of fixations, are introduced and compared to mean-shift clustering of fixations ([Privitera and Stark 2000; Santella and DeCarlo 2004]).

2 RELATED WORK

Initially, researchers may be tempted to solve the ROI generation problem by image segmentation. This method stems from the field of computer vision and identifies coherent regions in an image. However, such an approach to process and segment the stimulus image [Privitera and Stark 1998] is still challenging because semantic knowledge of the objects in the stimulus could be required.

Fortunately, information about stimulus semantics can be extracted directly from the gaze data. Data-driven algorithms can utilize the semantic segmentation of the stimulus provided by the viewer's behavior reflected in the eye tracking data. An example of such a method is the mean-shift clustering of fixation locations [Privitera and Stark 2000; Santella and DeCarlo 2004]. Locations that are looked at more frequently are likely to be semantically meaningful and good candidates for ROIs.

Possibly the closest method to our approach is by [Wooding 2002]. Where the authors constructed a three-dimensional fixation map (two-dimensional location plus the gaze density calculated by summing up normal distributions centered at each fixation) and "flooded [it], leaving only the highest peaks as islands" [Wooding 2002]. This approach can be derived from a shadow map visualization of the fixation map. Depending on the level of "flooding", meaning the fixation density threshold where the shadow map is rendered transparent, different ROIs can be identified.

Nyström [Nyström 2008] proposed to perform hierarchical segmentation by thresholding the fixation map at half of its maximal density. The segmented ROIs can then be subdivided by repeating the procedure with a new threshold, relative to the maximum of the segmented region. The process can be repeated an arbitrary amount of times, which results in a more accurate ROI segmentation. What is most remarkable in this algorithm is that it is able to even capture large area ROIs. Where they generally show a higher spread of gaze locations and are therefore less likely to sum up to high fixation counts in the fixation map. This hierarchical approach is likely to contain these ROIs somewhere within the hierarchy. In contrast, the aforementioned flooding approach may result in these areas being disregarded.

3 GENERATING ROIS FROM HEATMAPS

In the following sections, three different methods for automatically generating ROIs in a data-driven way are proposed: Local maxima thresholding, heatmap gradients, and overlap clustering.

3.1 Threshold based ROI algorithm

Figure 1(a) shows the workflow of the algorithm. It starts with pre-thresholding the fixation heatmap. Here, areas with a density smaller than the pre-threshold are irrelevant for further computations. For all images in Figure 1, we used a prethreshold of 1% of the maximum density of the heatmap to remove invalid gaze points.

In the next step, we search for local maxima in the density of the heatmap. This step is motivated by the observation that two high density ROIs that are spatially close to each other get easily fused to one bigger ROI when a simple density threshold is applied. Then, the spread of the Gaussian applied at every fixation location which makes the heatmap smooth and fusing of ROIs easier. Considering local maxima as candidate ROIs enables us to treat close-by maxima as separate ROIs and to fuse them afterwards, if applicable. Depending on the data recorded (and the measurement error of the eye-tracker), the heatmap contains many local maxima. To determine relevant maxima, a sliding window of a user-defined size W is applied. Only the largest of all local maxima within the window is considered a valid candidate. Based on the size of the window, candidate ROIs can be fused and smaller ROIs discarded. The size of the window is highly stimulus dependent, and determines the desired detail level of the analysis.

$$LocMax(x_i, y_i, W) = \begin{cases} 1, & I(x_i, y_i) > I(x_i + x_k, y_i + y_k) \\ & \forall x_k, y_k \in W \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Equation 1 calculates the candidate selection with x_i, y_i being the pixel position in the heatmap, W the set of allowed deviations in the search window and $I(x_i, y_i)$ is the density at location x_i, y_i in the heatmap. The equation assigns the value 1 to valid candidate local maxima, 0 otherwise.

In the following step, a threshold, based on a percentage of the heatmap density at the local maximum, is applied to each region. This step is similar to the second iteration in Nyström's approach [Nyström 2008], where a 50% is applied. All pixels both above the threshold and connected to the local maximum are assigned to a new ROI. An example of the thresholding as well as how the percentage parameter influences the size of the generated ROI is shown in Figure 2. Offshoots, as shown in Figure 1(f)(the red area on the right side, close to the center), or very small local maxima can combine ROIs. Afterwards, polygons borders are calculated for all ROIs.

The pseudocode in Algorithm 1 shows the threshold procedure that is based on a region growing approach. The algorithm searches for pixels neighboring the current ROI and adds them to the ROI if their density value is above the threshold. This step is repeated to convergence.

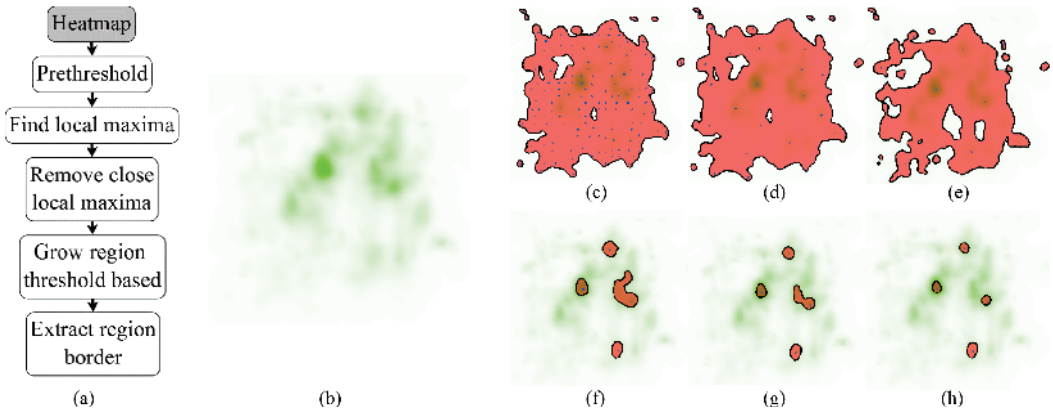


Figure 1: (a) shows the workflow of the threshold based ROI algorithm. (b) shows the fixation heatmap, which is the foundation of the following calculations. (c,d,e) visualize local maxima in the heatmap density as blue dots when using different window sizes (50px, 150px, 250px). The area highlighted in red is the pre-threshold of 1% of maximum density. The bottom row contains the results for different values of the threshold, where the red areas are extracted ROIs at (f) 50%, (g) 60% and (h) 70% of the density at the local maximum from (e).

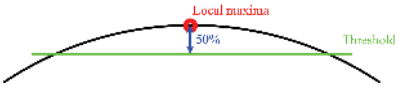


Figure 2: Calculation of the cutoff threshold for one local maximum. The black line represents the density distribution within the heatmap, the red dot is the local maximum, and the green line is the calculated threshold based on 50% of the density at the local maximum.

Algorithm 1 Thresholding algorithm. *ROI* holds the ROI region and is initialized as the point of the local maximum. More pixels are added during the iterations. *TH* is the threshold density, calculated from a percentage of the density at the local maximum. a and b are 2d pixel locations within the heatmap.

Require: ROI, TH, I
  function $Grow_{region}(ROI, TH, I)$
    while $\exists a \in I | distance(a,b) < 2 \; \forall b \in ROI$ **do**
      if $TH < (I(a))$ **then** $add(a, ROI)$
      end if
    end while
    return ROI
  end function

3.2 Gradient based ROI algorithm

The workflow of the gradient based ROI algorithm is shown in Figure 3(a). The first step is a pre-thresholding, just as in the threshold based algorithm. The impact of the pre-thresholding parameter is shown in Figure 3. Generally, ROIs get smaller, and low density ROIs are discarded for an increasing pre-threshold. In a second step, the density gradient is calculated (Figure 4(b)).

When the gradient (the first derivative of the density) crosses zero, there is no slope in the heatmap (Figure 4(b) red and grey circles). A derivative of zero implies a local maximum or minimum or a saddle point in the original function (see Figure 4). If we progress from each local maximum towards the next point where the gradient crosses zero, we have found the borders of our ROI. At the bottom of Figure 4, the assignment to ROIs is shown.

In our implementation, we used directional gradients pointing to the highest value in their 8-connected neighborhood and the position itself. Afterwards, we start from those values that do not point to any neighbor, but have the highest density within their neighborhood. All neighbor pixels with a gradient pointing towards this location are added to the new ROI. This step is repeated to convergence till convergence.

This procedure allows ROIs to be completely contained within larger ROIs. Those contained ROIs can easily be identified by their enclosing polygon outline being within a larger polygon (considering the enclosing border polygon $ROI_i \subset ROI_j | i \neq j$). If such a ROI is found, it is joined to the enclosing ROI (considering the pixel position values $ROI_i \cup ROI_j | i \neq j$).

$$Dir(x_i, y_i, W) = \begin{cases} (x_k, y_k), & Max(I(x_i + x_k, y_i + y_k)) \\ & \forall x_k, y_k \in W \end{cases} \quad (2)$$

Equation 2 describes the gradient calculation. I is the intensity value, W contains all position shifts [-1, 0, 1] in pixels for each direction (8 neighbors and pixel itself) and x_i, y_i is the starting location. The formula returns the vector to a neighboring maximum or $(0, 0)$ if the position itself is the maximum in its neighborhood. The algorithm for growing the region via the gradient is shown in Algorithm 2.

3.3 Overlap clustering

In [Kübler et al. 2015], the authors describe how to fit ellipses to samples recorded during a fixation and also how the ellipsoid shape can be used for data quality assessment, as the actual spatial extent of the fixation is represented by the ellipse axes.

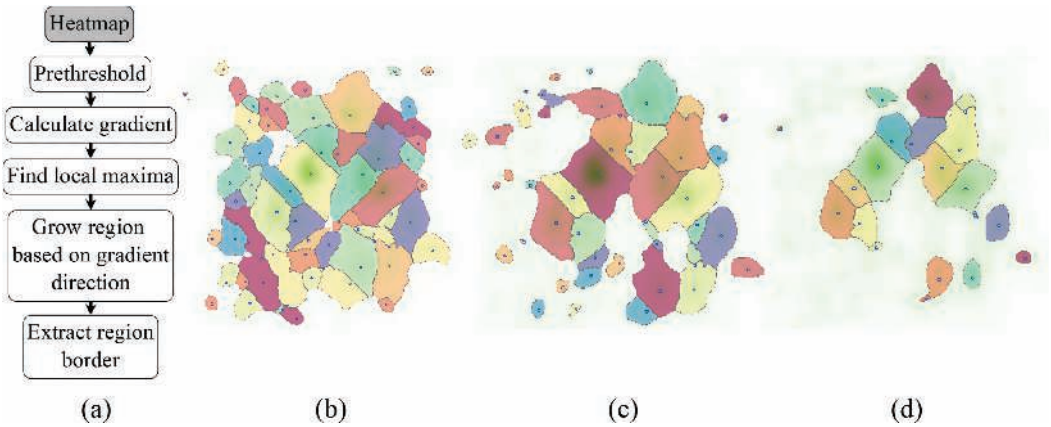


Figure 3: (a) shows the workflow of the gradient based ROI algorithm. In (b) the pre-threshold is applied at 1% of the maximum density level. (c) and (d) at 5% and 10%, respectively. Individual ROIs are highlighted in different colors.

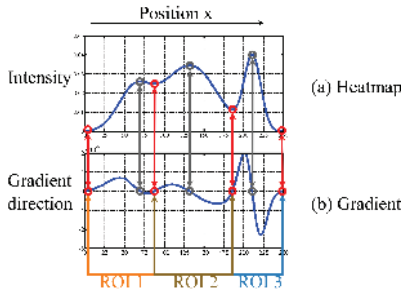


Figure 4: The heatmap density is shown in (a), with its first derivative, the gradient, in (b). Red circles mark breaking points of ROIs and the green circles indicate the starting points for growing a new ROI.

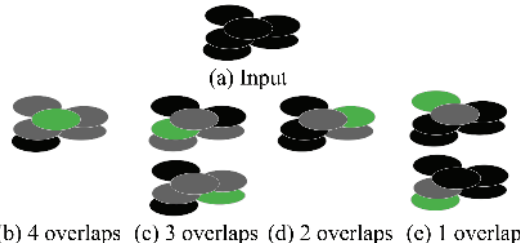


Figure 5: Overlap clustering procedure. (a) Ellipse representation of a set of fixations. In (b,c,d,e), all fixations that overlap the ellipse of the fixation currently under consideration (green) are highlighted (in gray). Each set of a fixation and its overlaps (green plus gray ellipses) is considered a cluster. Fixations that do not participate in the cluster are shown in black. The clusters are ordered from left to right in descending order of cluster size.

Algorithm 2 Gradient based region growing. *ROI* holds the ROI region and is initialized as the point of the local maximum. More pixels are added during the iterations from neighboring positions *a*.

Require: ROI, I
  function $Grow_{region}(ROI, I)$
    while $\exists a \in I \mid Dir(a) + a \in ROI$ **do** $add(a, ROI)$
    end while
    return ROI
  end function

Algorithm 3 First clustering step: F is a list containing all fixations, AC will contain all found clusters after the algorithm is run.

Require: F, AC
  function $Init_{Cluster}(F, AC)$
    for $a \in F$ **do**
    $add(a, C)$
      for $b \in F$ *and* $a \neq b$ **do**
        if $a \subseteq b$ **then** $add(b, C)$
        end if
      end for
    $add(C, AC)$
    end for
    return AC
  end function

In this chapter, we will utilize the elliptic shape for the clustering of fixations. Therefore, intersections between the ellipses are calculated and overlapping ellipses are merged to clusters. The process is visualized in Figure 5.

In the second step, overlapping clusters are merged together. This starts from the largest cluster, i.e., the one shown in Figure 5(b), and searches for other clusters that overlap with it. If two clusters overlap, they are merged together. This process is repeated until no cluster extension is possible. Afterwards the process is repeated with the largest of the remaining clusters that can still be extended.

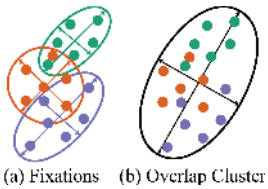


Figure 6: (a) Fixations are the ellipses outlines and the dots are gaze points. Dots and ellipses in the same color belong together. (b) The black ellipses represents the overlap cluster calculated based on all contained points. Arrows outgoing from the center of the ellipses are the axis calculated from the principal component analysis.

This merging process is done by calculating the mean minor and major axis using the principal component analysis on all gaze points belonging to the fixation ellipses included in the cluster. As shown in Figure 6, (a) illustrating the fixations and their gaze points are drawn in the same color and in (b) the resulting cluster shape is shown. The axis of all ellipses (indicated by arrows) are the vectors calculated with the principal component analysis.

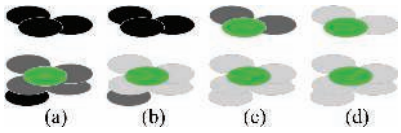


Figure 7: (a) is the starting cluster, which is the largest possible cluster from the first step. The green ellipses are the starting point of each cluster and the dark gray clusters belonging to it. In (b), the first merge is applied as visualized by the new dark gray ellipses and the light gray ellipses already contained in the cluster. (c) shows that the bottom cluster can not grow any more and is therefore finished. The next starting cluster in (c) is the top cluster as visualized by the new green ellipse and the ellipses belonging to it (dark gray). (d) is the result of the overlap clustering for the given ellipses.

Figure 7 visualizes the merging procedure. The largest cluster of the first step is chosen as a starting point and is successively enlarged by overlapping clusters. In the example, one minor cluster is merged with the larger one ((d) new dark gray ellipses). Once the cluster cannot grow anymore, we proceed with the next cluster that has not been merged yet (c). The two final clusters are shown in (d).

With Eyetrace, it is possible to adjust the overlap clustering by specifying a minimum count of overlapping fixations required to create a cluster. This minimum fixation threshold can either be applied to all data together or to the data of each subject separately. This way the algorithm can also cope with large and dense data.

Figure 8 shows each of the proposed methods applied to the same data. Naturally, the generated ROIs are not identical. Every method has it's advantages and disadvantages. For example, the gradient and mean shift approach generate many ROIs, not only based on the nature of the methods, but also on the parameter choice. This

Algorithm 4 Cluster merging step: *AC* contains the clusters found by the first step, merged clusters are stored in *GC*.

Require: AC, GC
  function $Merge_{Cluster}(AC, GC)$
    while $AC > 0$ **do**
    $C = max(AC)$
    $remove(C, AC)$
      while $\exists a \in C | a \in AC$ **do**
      $C_{sub} = get(a, AC)$
      $add(C_{sub}, C)$
      $remove(C_{sub}, AC)$
      end while
    $add(C, GC)$
    end while
    return GC
  end function

parameterization can be a wanted result, or a disadvantage based on the investigated part of human gaze behavior. In EyeTrace, ROIs can be removed, manually added, and or modified: For cases where the automatically generated ROIs are to large/small or unavailable. Figure 8(e) shows an example for large ROIs, where (c) separates the ROI of the painters head from his drawing hand. The usefulness of a shape based ROI depends always on the quality of the recorded data as well as the goal of the study or task. The data quality refers to the size of the ROI needed to capture the gaze attention of a subject and sufficient enough to see similarities or gaze behavior and excluding unrelated behavior. For example, data with low precision requires larger ROIs, whereas in contrast, large ROIs could induce an error in high precision data. In case of low accuracy data, which induces a localization problem, automatic ROI generation can help visualizing the offset.

4 EXPERIMENT

In the following sections, we will refer to cumulative clusters simply as clusters. Typically in research literature on visual behavior, clusters are calculated based on the data of a single subject. More specifically, for cumulative clusters, the data of more than one subject is combined and treated as a single subject.

In this exemplary evaluation, we aim to compare the proposed method with the state-of-the-art meanshift algorithm [Privitera and Stark 2000; Santella and DeCarlo 2004] and our improved version of hierarchical segmentation by thresholding a heatmap [Nyström 2008]. The first part (Subsection 4.1) of the evaluation concerns the impact of automatic ROI generation on statistical values. Therefore, we investigate the generated ROIs on the results of common ROI statistics and compare them to manually annotated ROIs. The used recordings are from nine subjects viewing the painting "The Art of Painting" presented on a screen for one minute. Almost one thousand fixations were recorded with the EyeTribe eye-tracker at 25 Hz sampling rate.

In the second part (Subsection 4.2), we evaluate the generated ROIs in a classification task. The task concerns differentiating between art experts and novices for a viewing experiment based on calculated ROI statistics. For this classification evaluation, we used

(a) Heatmap (b) Meanshift (c) Overlap (d) Threshold (e) Gradient

Figure 8: (a) shows the fixation heatmap where red is the highest value. In (b,c,d,e), calculated regions of interest are shown using meanshift clustering (b), overlap clustering (c), threshold based (d) and gradient based methods (e).

the data recorded in [Rosenberg 2014] at the University of Vienna. It contains 40 subjects, where 20 participants are experts and the other 20 are novices. The used eye-tracker model was "IViewX RED 120" with 120 Hz. The recordings where performed while the subject was sitting in front of a 30 inch monitor with a screen resolution of 2560x1600 pixels. Each subject viewed the artwork for 2 minutes with a head distance of 0.9 meters($\tilde{3}$ feet). In the supplementary material, all parameters used are reported.

4.1 Automatic vs. annotated statistics

The ROIs evaluated are shown in Figure 9. Where applicable, we chose the parameters to get the best ROI representations for the face of the woman, the painter's head, the mask, and the chandelier. This parameter choice was because we would expect these regions to attract most of the gaze (as can be seen in Figure 8(a)). We identified fixations in the data with the Bayesian Mixture Model proposed in [Tafaj et al. 2012], with a minimum duration threshold of 70ms.

In Table 1, the resulting fixation and gaze point statistics are shown. As key metrics for the statistics, we used the amount of gaze points (GP), the average gaze point duration (GP dur), the amount of fixations, and the average fixation duration. All values are averaged over all nine subjects. As apparent, the mean statistical values between the meanshift approach and overlap clustering are similar, except for the head ROI. Here, the shape is rather vertically stretched in the case of overlap clustering and rather horizontally positioned (overlapping the painting ROI) for the meanshift approach, see Figure 9(d). The same effect can be observed for the gradient approach on the fixation heatmap, which also separates the head ROI vertically (Figure 9(h)).

4.2 ROI classification comparison

To compare the different methods, we extracted statistics based on the generated ROIs and trained a support vector machine (SVM) classifier. The evaluation was done using a 20-fold cross validation with the SVM from Matlab 2015b. Common practice with SVM classification is to evaluate different parameters and then select the best performing result. We evaluated the kernel scale parameter in the range $1-10$ with a step width of 0.1. The evaluated kernels are *'linear'*, *'Gaussian'*, and *'polynomial'*. We evaluated each kernel function with data standardization from Matlab. The used statistical values for each ROI are: time of first entry (S_{X1}), amount (S_{X2}), per minute (S_{X3}), share (S_{X4}), total time (S_{X5}), minimal consecutive time (S_{X6}), maximal consecutive time (S_{X7}), and average consecutive time (S_{X8}). Here, X stands for gaze point (G) and fixation (F). From those 16 statistical values, ($S_{G1-8} + S_{F1-8}$) we evaluated all possible subsets of $1-4$, and the best result for classification were selected. Each evaluation was a 20-fold cross validation to ensure generality of the classifier.

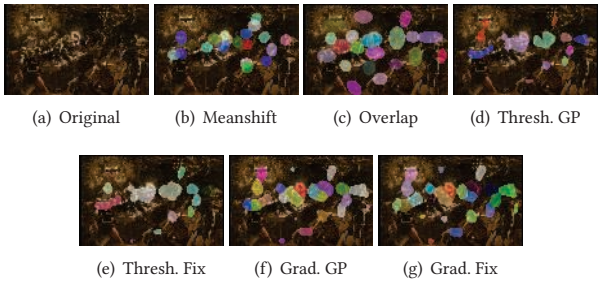
Table 2 shows the results for each method in combination with a specific kernel. As can be seen, the highest score is achieved by the meanshift clustering. This result is due to the overlapping clusters

Table 1: Averaged statistic results over all nine subjects (mean gaze position of both eyes) for generated ROIs and annotated ones. The first column specifies the annotated object and in the second column, the used method. Columns three through six contain statistics, e.g. amount of gaze points (GP), average gaze point duration (GP dur) in ms, amount of fixations (Fix) and average fixation duration (Fix dur) in ms.

ROI	Method	GP	GP dur	Fix	Fix dur
Face	Manually	83.3	61.2	5	455.1
	Meanshift	96.3	140.4	7	357.1
	Overlap	89.6	121.5	6.3	372.4
	TH GP	175.2	295.3	10	549.6
	TH Fix	168.4	267.8	9.7	554.8
	Grad	232.7	371	14.2	489.3
	Grad Fix	242.6	412.3	14.6	490.8
Head	Manually	63.1	90.3	5.3	329
	Meanshift	79	219.5	5.5	454.1
	Overlap	46.4	135	3.4	446.5
	TH GP	175	264.1	10.8	497.7
	TH Fix	156.3	203.7	10.3	454.4
	Grad	196.7	298.1	12.3	481.8
	Grad Fix	92.8	105.8	7.2	334.3
Mask	Manually	48.2	32	1.8	480.4
	Meanshift	194.7	359.1	11.2	546
	Overlap	141.7	252.7	7.8	542.5
	TH GP	101	393	5.8	614.8
	TH Fix	93.3	213.4	5.3	612.1
	Grad	103.4	435.8	6	604.1
	Grad Fix	98.1	229.3	5.5	605
Chandelier	Manually	84.4	150.2	6.3	454.2
	Meanshift	112.8	368.7	6.3	477.3
	Overlap	104.6	411.2	5.5	485.8
	TH GP	59.7	159.3	4.1	324.7
	TH Fix	62.4	157.8	4.4	328.2
	Grad	61.7	171.1	4.2	336.6
	Grad Fix	65.4	184.7	4.6	325.1

(a) Original (b) Manually (c) Meanshift (d) Overlap (e) Thresh. GP (f) Thresh. Fix (g) Grad. GP (h) Grad. Fix

Figure 9: Shows the used clusters for experiments in 4.1. (a) is the original image ,and in (b) the manually annotated ROIs with labels are shown. In (c), the clusters found by meanshift, (d) overlap clustering, (e) threshold and (g) gradient based gaze point heatmap ROIs, (f) threshold and (h) gradient based fixation heatmap ROIs.



(a) Original (b) Meanshift (c) Overlap (d) Thresh. GP

(e) Thresh. Fix (f) Grad. GP (g) Grad. Fix

Figure 10: Shows the used clusters for experiments in 4.2. (a) is the original image. In (b) the clusters found by meanshift, (c) overlap clustering, (d) threshold, and (e) gradient based gaze point heatmap ROIs, (f) threshold and (g) gradient based fixation heatmap ROIs.

Figure 11: Wassily Kandinsky, Study for Composition VII, 1913, retrospectively dated 1910 and earlier published as "First Abstract Watercolor" [public domain 1910] with cumulative clusters (red ellipses).

Table 2: The classification results for all ROI generation algorithms with the three kernels.

Method	Linear	Gaussian	Polynomial
Meanshift	72.5%	85%	77.5%
Overlap	75%	80%	75%
Threshold GP	75%	75%	75%
Threshold Fix	80%	77.5%	77.5%
Gradient GP	70%	72.5%	72.5%
Gradient Fix	70%	72.5%	67.5%

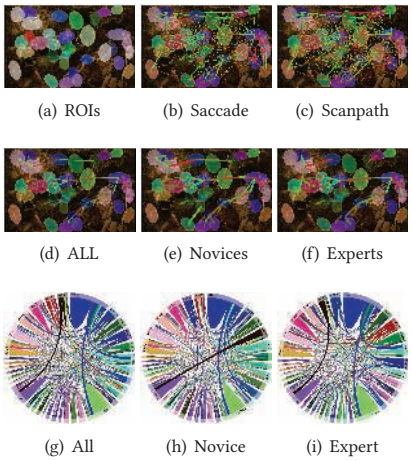
and the more centered localization in the image. The lowest results are obtained by the gradient based ROIs for this scenario. Overall, it can be seen that all methods can be used to achieve results above chance level (50%).

5 APPLICATION

When it comes to abstract paintings, a top-down definition of ROIs is difficult, because the lack of semantically meaningful objects (e.g. persons) depicted. Due to the interplay of pure lines and colors, the layers of paint, and the compositional structure in the image, saliency maps often fail. In such cases, an automated ROI generation can be very helpful. Jackson Pollock's famous work "Convergence", for example, is perceived as a very dynamic and agitated painting [Commare and Brinkmann 2016]. Yet, it is almost impossible to predict where an observer will fixate, because the colors are smeared on the canvas with the purpose to offer no specific place to focus on. A cumulative cluster analysis allows the investigation of those regions in the painting that were most fixated by the observers. Figure 11, shows cumulative clusters calculated by overlap clustering, with a minimum of 120 fixations per cluster. This visualization of attention depicts an average of 40 subjects who looked at the artwork for one minute. The regions where many fixations are accumulated have a high contrast and seem to attract the observer's attention.

5.1 Transitions

Another application related to ROIs are the transitions. Here saccades and scanpath can be analyzed. Figure 12 shows the transitions calculated between clusters. The first row shows the clusters and transitions based on saccades and scanpath as overlay on the image. The second row in Figure 12 shows the saccades as transitions for all subjects and separated by group (experts and novices). In this visualization it can be seen that both groups differ significantly. For a better visualization of this we highlighted two regions from Figure 12(e) and (f) in Figure 13. It can be seen that region 1 (R1) is different due to the strength difference of the vertical line on the left side. It has to be noted that the width of a line indicates the global proportion of a connection. In addition the thin horizontal line on the left is not present for the experts. The reverse case for the thin horizontal line on the right side. In region 2 (R2) the difference is obvious. Another impressive visualization technique is shown



(a) ROIs (b) Saccade (c) Scanpath

(d) ALL (e) Novices (f) Experts

(g) All (h) Novice (i) Expert

Figure 12: In (a) the calculated clusters are shown. (b) and (c) show the transitions. In the second row the saccade transitions are shown using a normed threshold for all, novices and experts. The last row shows chord diagrams where the ROIs are the outline and the saccades build the connections.

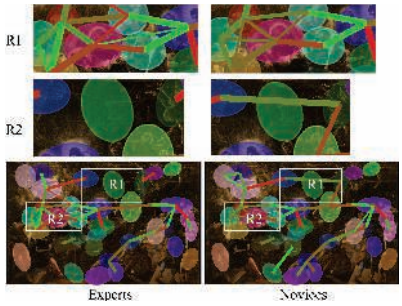


Figure 13: Shows the difference between experts and novices for images (e) and (f) from Figure 12.

by the chord diagrams in Figure 12. Without filtering the main differences and similarities between experts and novices are visible. One example is the skewed dark line only visible for novices and the dark curve on the left side of the diagram, which is only visible for experts. A clear similarity is the blue vertical curve, present for both groups.

6 DISCUSSION OF ADVANTAGES AND DISADVANTAGES

In Figure 14, some strong and weak points mentioned in table 3 for each method are shown. (a) shows the over segmentation of the gradient approach, which occurs if the prethreshold is set very low. Also, the threshold approach can segment large ROIs, as shown in Figure 14(b). This effect can be disadvantageous if the ROIs, which should be separated, intermingle, as shown in Figure 14(e) (face and mask).

(a) Grad. (b) Thre. (c) Over. (d) Mean. (e) Thre. (f) Over.

Figure 14: Exemplary weak and strong point visualizations for the different methods. (a) over segmentation, (b) large segment ROI, (c,d) overlapping example, (d) large clusters, (e) intermingle ROIs, and (f) fixed size.

Table 3: Advantages and disadvantages for each method.

Method		characteristics
Meanshift	-	four parameters
	+	size adjustable
	+	finds clusters with low gaze activity
	-	relies on fixations
	-	clusters can overlap
Overlap	+	one parameter
	-	size fixed
	+	finds clusters with low gaze activity
	-	relies on fixations
	+	minimum fixations can be applied per subject
	-	clusters can overlap
Threshold	-	three parameters
	+	size adjustable
	-	problems finding clusters with low gaze activity
	+	different input data
	+	delivers a non overlapping segmentation
Gradient	+	one parameter
	+	size adjustable
	+	finds clusters with low gaze activity
	+	different input data
	+	delivers a non overlapping segmentation
	-	tends to over segmentation
	-	ROIs are in close contact to each other

Figure 14(c,d) show the overlapping that may occur for the meanshift and overlap approach. This effect is due to the result of the principle component analysis and the region approximation as ellipses. Another weak point of the overlap clustering is shown in Figures 14(c) and (f), where it can be seen that the cluster size is fixed. This effect is due to the fixed size of the fixation ellipses.

7 CONCLUSION

We proposed three novel methods for ROI generation. Each of these methods reduce the configuration amount while maintaining ROI quality that is comparable to those achieved by state-of-the-art approaches. The methods are statistically evaluated and compared against each other. Afterwards, we showed and exemplary application to abstract art and discussed the strong and weak points of all ROI generation algorithms. Further research will inspect the gain of automatic ROI generation for gaze based rendering [Mantiuk et al. 2013], image compression [Privitera and Stark 1998] and user interface evaluations [Hutchinson et al. 1989; Oh et al. 2002].

REFERENCES

Bonnie Auyeung, Michael V Lombardo, Markus Heinrichs, Bhisma Chakrabarti, A Sule, Julia Brynja Deakin, RAI Bethlehem, L Dickens, Natasha Mooney, JAN Sipple, et al. 2015. Oxytocin increases eye contact during a real-time, naturalistic social interaction in males with and without autism. *Translational psychiatry* 5, 2 (2015), e507.

T Blascheck, K Kurzhals, M Raschke, M Burch, D Weiskopf, and T Ertl. 2014. State-of-the-art of visualization for eye tracking data. In *Proceedings of EuroVis*, Vol. 2014.

Tanja Blascheck, Michael Raschke, and Thomas Ertl. 2013. Circular heat map transition diagram. In *Proceedings of the 2013 Conference on Eye Tracking South Africa*. ACM, 58–61.

L. Commare and H. Brinkmann. 2016. Aesthetic Echoes in the Beholder's Eye? Empirical evidence for the divergence of theory and practice in the perception of abstract art.. In *M. Zimmermann (Ed.): Vision in Motion. Streams of Sensation and Configurations of Time*. Diaphanes.

Andrew T Duchowski. 2002. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers* 34, 4 (2002), 455–470.

Joseph H Goldberg, Mark J Stimson, Marion Lewenstein, Neil Scott, and Anna M Wichansky. 2002. Eye tracking in web search tasks: design implications. In *Proceedings of the 2002 symposium on Eye tracking research & applications*. ACM, 51–58.

Thomas E Hutchinson, K Preston White, Worthy N Martin, Kelly C Reichert, and Lisa A Frey. 1989. Human-computer interaction using eye-gaze input. *IEEE Transactions on systems, man, and cybernetics* 19, 6 (1989), 1527–1534.

Thomas C Kübler, Katrin Sippel, Wolfgang Fuhl, Guilherme Schievelbein, Johanna Aufreiter, Raphael Rosenberg, Wolfgang Rosenstiel, and Enkelejda Kasneci. 2015. Analysis of eye movements with Eyetrace. In *Biomedical Engineering Systems and Technologies*. Springer, 458–471.

Radoslaw Mantiuk, Bartosz Bazyluk, and Rafal K. Mantiuk. 2013. Gaze-driven Object Tracking for Real Time Rendering. *Computer Graphics Forum* 32, 2 (2013), 163–173.

Franco Mawad, Marcela Trías, Ana Giménez, Alejandro Maiche, and Gastón Ares. 2015. Influence of cognitive style on information processing and selection of yogurt labels: Insights from an eye-tracking study. *Food Research International* 74 (2015), 1–9.

Marcus Nyström. 2008. *Off-line Foveated Compression and Scene Perception: An Eye-Tracking Approach*. Ph.D. Dissertation. Lund University.

Alice Oh, Harold Fox, Max Van Kleek, Aaron Adler, Krzysztof Gajos, Louis-Philippe Morency, and Trevor Darrell. 2002. Evaluating look-to-talk: a gaze-aware interface in a collaborative environment. In *CHI'02 Extended Abstracts on Human Factors in Computing Systems*. ACM, 650–651.

Bing Pan, Helene A Hembrooke, Geri K Gay, Laura A Granka, Matthew K Feusner, and Jill K Newman. 2004. The determinants of web page viewing behavior: an eye-tracking study. In *Proceedings of the 2004 symposium on Eye tracking research & applications*. ACM, 147–154.

Claudio M Privitera and Lawrence W Stark. 1998. Evaluating image processing algorithms that predict regions of interest. *Pattern Recognition Letters* 19, 11 (1998), 1037–1043.

Claudio M Privitera and Lawrence W Stark. 2000. Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22, 9 (2000), 970–982.

public domain. 1910. Wassily Kandinski first abstract watercolor. https://en.wikipedia.org/wiki/File:First_abstract_watercolor_kandinsky_1910.jpg. First abstract watercolor, painted by Wassily Kandinsky, 1910.

Raphael Rosenberg. 2014. Blicke messen: Vorschläge für eine empirische bildwissenschaft. *Jahrbuch der Bayerischen Akademie der Schönen Künste* 27 (2014), 71–86.

Anthony Santella and Doug DeCarlo. 2004. Robust clustering of eye movement recordings for quantification of visual interest. In *Proceedings of the 2004 symposium on Eye tracking research & applications*. ACM, 27–34.

Benjamin Strobel, Marlit Annalena Lindner, Steffani Saß, and Olaf Köller. 2018. Task-irrelevant data impair processing of graph reading tasks: An eye tracking study. *Learning and Instruction* (2018).

E. Tafaj, G. Kasneci, W. Rosenstiel, and M. Bogdan. 2012. Bayesian Online Clustering of Eye Movement Data. In *Proceedings of the Symposium on ETRA*. ACM.

Benjamin W Tatler, Nicholas J Wade, Hoi Kwan, John M Findlay, and Boris M Velichkovsky. 2010. Yarbus, eye movements, and vision. *i-Perception* 1, 1 (jan 2010), 7–27. https://doi.org/10.1068/i0382

Dereck Toker, Cristina Conati, Ben Steichen, and Giuseppe Carenini. 2013. Individual user characteristics and information visualization: connecting the dots through eye tracking. In *proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 295–304.

David S Wooding. 2002. Eye movements of large populations: II. Deriving regions of interest, coverage, and similarity using fixation maps. *Behavior Research Methods, Instruments, & Computers* 34, 4 (2002), 518–528.