Vision, Modeling, and Visualization (2018)
F. Beck, C. Dachsbacher, and F. Sadlo (Eds.)

Automatic generation of saliency-based areas of interest for the visualization and analysis of eye-tracking data

Wolfgang Fuhl[1,a], Thomas Kuebler[1,b], Thiago Santini[1,c], Enkelejda Kasneci[1,d]

[1]University of Tübingen, Department of Computer Science, Perception Engineering
[a]wolfgang.fuhl@uni-tuebingen.de, [b]thomas.kuebler@uni-tuebingen.de, [c]thiago.santini@uni-tuebingen.de, [d]enkelejda.kasneci@uni-tuebingen.de

Abstract

Areas of interest (AOIs) are a powerful basis for the analysis and visualization of eye-tracking data. They allow to relate eye-tracking metrics to semantic stimulus regions and to perform further statistics. In this work, we propose a novel method for the automated generation of AOIs based on saliency maps. In contrast to existing methods from the state-of-the-art, which generate AOIs based on eye-tracking data, our method generates AOIs based solely on the stimulus saliency, mimicking thus our natural vision. This way, our method is not only independent of the eye-tracking data, but allows to work AOI-based even for complex stimuli, such as abstract art, where proper manual definition of AOIs is not trivial. For evaluation, we cross-validate support vector machine classifiers with the task of separating visual scanpaths of art experts from those of novices. The motivation for this evaluation is to use AOIs as projection functions and to evaluate their robustness on different feature spaces. A good AOI separation should result in different feature sets that enable a fast evaluation with a widely automated work-flow. The proposed method together with the data shown in this paper is available as part of the software EyeTrace [?] http://www.ti.uni-tuebingen.de/Eyetrace.1751.0.html.

CCS Concepts

• ***Human-centered computing*** → ***Heat maps; Scientific visualization;*** *Information visualization;* • ***Computing methodologies*** → *Cross-validation;* • ***Applied computing*** → *Fine arts;*

1. Introduction

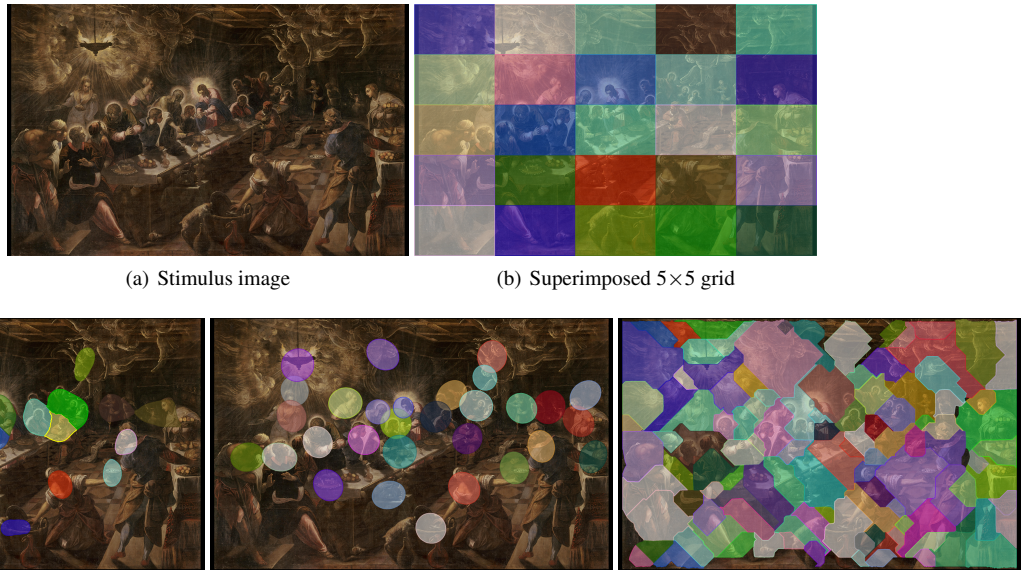
The raw gaze signal itself does not immediately reveal intentions, strategies, nor the cognitive state of a person. A recorded fixation location only suggests perception at a specific location, not perception of a specific entity. Therefore, to visualize and analyse such data, we need to add an additional semantic layer on top of the data in order to *make sense* of it. We can do so by aggregating gaze towards meaningful subregions of the stimulus, known as areas of interest. Usually one of the early steps in eye-tracking data processing is the identification of fixations, as gaze samples recorded during a saccade have different implications on perception as fixation locations [TKK*13, SFKK16].

Traditionally, such AOIs (Areas Of Interest) were manually annotated by the data analyst. However, this approach has several flaws associated with the subjective judgment of the data analyst. Furthermore, even small inaccuracies in the calibration of the eye tracker might result in fixations being wrongly assigned to a close-by AOI. Generally, the level of detail of AOI annotation is often unclear (e.g., when thinking about a face one might annotate the face as a whole or specific subregions such as the eyes and the mouth, depending on the expectations of the data analyst). Some methods for scanpath comparison therefore suggest a complete renunciation of AOIs [DNJ*12].

A partial remedy for the manual, subjective annotation of AOIs is the automated creation of AOIs based on either the recorded gaze data, the stimulus material or neither. Figure 1a shows Jacopo Robusti, called Jacopo Tintoretto, "The last supper", created in 1592-1594, and is used as stimulus for the participants. In the simplest form, an evenly sized grid is superimposed on the stimulus (Figure 1b). Data-driven approaches segment the gaze heatmap [Nys08, FKS*15] of multiple participants (Figure 1c) or cluster fixations [PS00, SD04]. These approaches lead to a certain robustness towards measurement noise and calibration inaccuracies as these are already contained within the data during AOI creation (Figure 1d). If for some reason these methods are not applicable, one can utilize image information [PS98] (Figure 1e) to generate somewhat meaningful AOIs.

Aggregating gaze data to AOIs allows for a more robust and meaningful subsequent analysis by inspecting statistical information (such as the number of fixations and average fixation duration) distributed towards each AOI [CMTG10, DNJ*12]. It is also possible to analyze the transition matrix between AOIs. This matrix contains the probability of gaze traversing from a specific AOI

DOI: 10.2312/vmv.20181252



(a) Stimulus image (b) Superimposed 5×5 grid

(c) Heatmap segmentation (d) Fixation clusters (e) Saliency segmentation

Figure 1: *a) the stimulus image used. b) AOIs created with a superimposed 5x5 grid. c) AOIs created using a heatmap of fixations. d) AOIs created using mean shift clustering. e) AOIs resulting from the proposed saliency segmentation.*

to another specific AOI. One often uses a string encoding for this purpose, where a letter is assigned to each fixation based on the AOI it has been assigned to. The result is a *scanpath word*. Due to the inequality in length of eye-tracking recordings and different sampling rates of eye trackers their comparison is a challenging task [DDJ*10, CMTG10, **?**].

The definition and consistency of AOIs represents a challenge as well, especially when one strives for scenes and pictures with no obvious shape or texture present [HKvdBH16]. An example is abstract art where the AOIs cannot be defined based on objects or surfaces as it is done for figurative or concrete artwork. We will address this problem by defining AOIs based on image features of the stimulus. These are motivated by a computational recreation of properties of the human visual pathway (saliency maps). We demonstrate how regions can be extracted from saliency maps by existing algorithms for heatmap segmentation. Thereby we generate consistent regions for a given stimulus and can define a consistent level of detail as well via quantifiable parameters of the used algorithm [HKvdBH16].

For evaluation of the quality of generated AOIs, we cross-validate support vector machine classifiers with the task of separating scanpaths of art experts from those of novices. This is done for different key metrics calculated on the generated AOIs. The comparison is based on the overall classifier performance for the different AOI generation algorithms. The motivation for this evaluation is to use AOIs as projection functions and to evaluate their robustness on different feature spaces. A good AOI separation should result in different feature sets that enable a fast evaluation with a widely automated work-flow.

The proposed work-flow can be subdivided in two major steps. First, we compute a saliency map of the stimulus image. In the second step, we treat the saliency map like a gaze heatmap and apply a state-of-the-art algorithm for AOI extraction (an example is shown in Figure 1e).

The remainder of this paper is organized as follows. Section 2 gives an overview on the state-of-the-art. Our method is presented in Section 3. Section 4 and 5 present and discuss the collected eye-tracking data and the evaluation results. Section 6 concludes this paper.

2. Related work

In this section, we describe existing approaches for creating AOI in an automated way. Therefore, we group methods into four categories namely shape, data and stimulus based as well as the their scientific advancement saliency maps. Shape-based methods are preliminary defined regions that are not related to the image or the data. In the category of stimulus-based methods, only the stimulus image is used to define the AOIs. The last category contains data-driven algorithms. Hybrid approaches of the aforementioned categories do exist as well. As the field of saliency maps is very vivid and hundreds of different approaches do exist, we focus on the description of the approach utilized in this work and refer the reader to [BI13] for a broader overview.

2.1. Shape-based AOI generation

Shape-based AOIs are commonly applied for their simplicity. For stimuli that distribute their content equally over the whole area, their application can be justified. However, the borders of such

AOIs will not correspond to meaningful objects and the interpretation of results can be difficult. Popular shapes are squares, rectangles or ellipses. By the definition of shape sizes the data analyst implies some assumptions on the data. These can be quantified by reporting the shape size, as contrary to manual annotation at different detail levels in different stimulus regions.

2.2. Stimulus-based AOI generation

Privitera et al. [PS98] suggested to segment a stimulus image into coherent regions. Ten different algorithms for image feature extraction and clustering are analyzed and compared against human fixations. The aim was to improve image compression quality by compressing aware of regions relevant to a human. Other approaches for stimulus-based AOIs originate in image segmentation [SM00, AMFM11]. The main interest here is the exact regions corresponding to objects or entities like persons or cars. The main goal is to automate the manual labeling process. As there are stimulus materials that do not contain objects or separated areas, this approach is not always applicable. Therefore, a more general extraction of features motivated by the low-level processes of human vision is of interest. Most prominent approaches for such low-level processes are saliency maps [IKN98, HZ07].

2.3. Data-based AOI generation

Methods from this realm utilize the semantic information contained within the eye-tracking data itself in order to generate AOIs. [Woo02, Nys08] proposed thresholding approaches for fixation heatmaps. In [FKS*15], a gradient-based segmentation is applied in order to avoid local heatmap maxima that would result in less viewed AOIs being omitted. A technically different approach uses mean-shift clustering on fixation locations to achieve similar results [PS00, SD04]. The advantage of these methods is that small calibration errors can be compensated as they are contained within the data during AOI creation. Human knowledge about possible stimulus segmentations can be extracted from the eye-tracking data. However, enough data for a fully converged heatmap needs to be available (usually at least 30 recordings). Furthermore, this approach cannot cope with spatially overlapping AOIs. In that case, they might be fused to one large, continuous AOI. As such a distinction might be relevant for specific stimulus materials.

2.4. Saliency maps

One of the first and most renowned biologically inspired computational model of human vision was proposed in [IKN98]. It is based on a bottom-up architecture described in [KU87] and uses features such as color and edge orientation. However, in [CMH*15] the authors showed that the result of this saliency model is generally blurry and overemphasizes small local features. This can be problematic for segmentations or detection algorithms using this saliency map as an input.

Frequency-based approaches Other approaches work in the frequency domain of an image (its Fourier transform [Bra78]). Examples are the frequency-tuned salient region detection [AHES09] or spatio-temporal saliency detection [GMZ08]. These approaches determine the saliency based on the amplitude and phase spectrum. This preserves the high-level structure of an image with the disadvantage that they tend to highlight object boundaries.

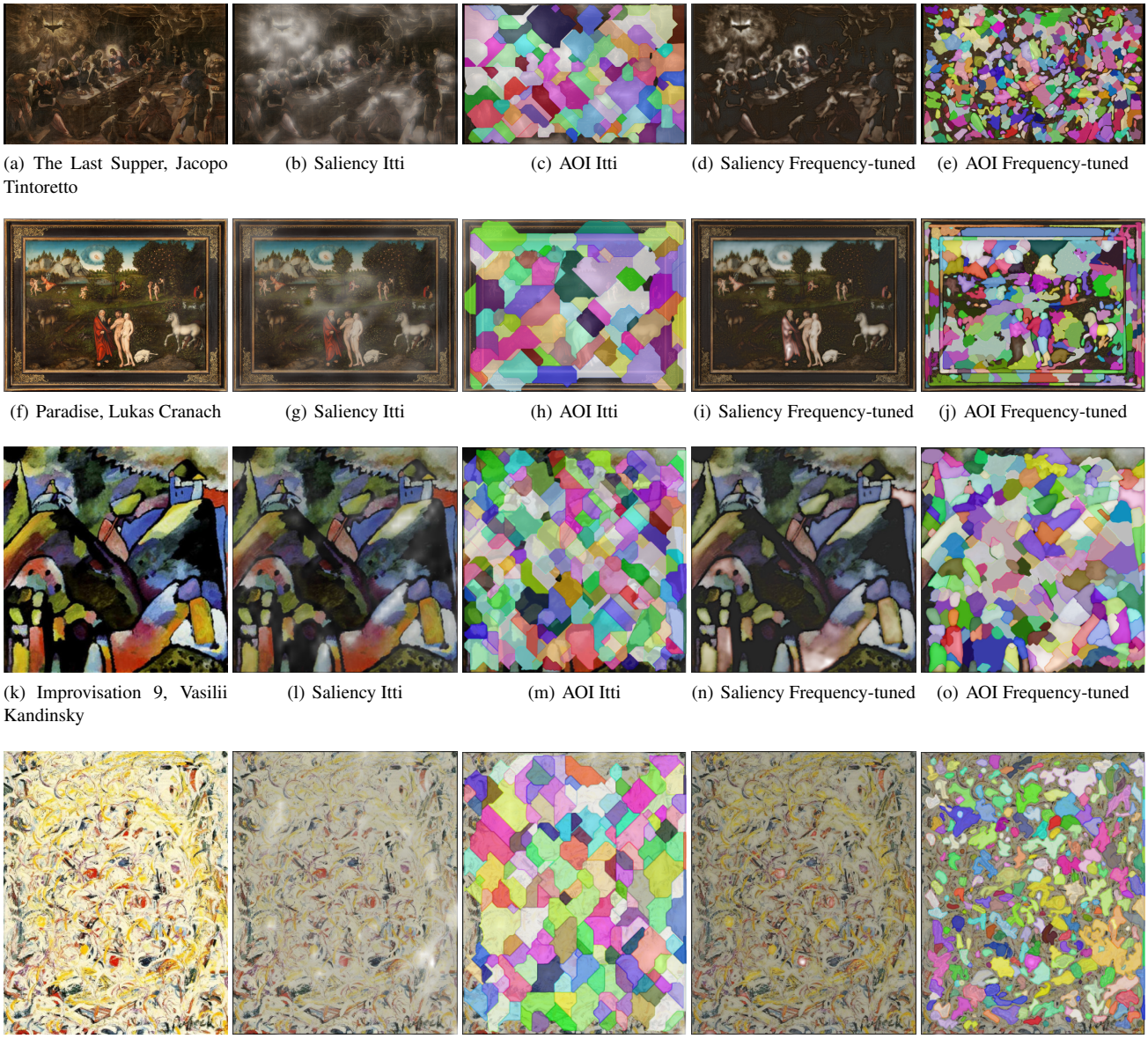
Local and global methods For color-based methods it is also possible to distinguish between local and global computation models. In local models, the surrounding region of each pixel is investigated. Based on local color contrast a saliency value for each pixel is computed [MZ03, IB05]. Local methods produce less blurry saliency maps but miss global relations between regions and textures. They are sensitive to edges or noise. Global models in contrast include the correlations and juxtapositions of different image regions [LYS*11, GZMT12, WKI*11]. The results of global methods are consistent in terms of image structures but the computational cost increases due to the involved combinatorial possibilities. Therefore, those approaches can only be applied to low resolution images, resulting in a loss of small but salient content. It has to be mentioned that the approach in [AHES09] uses both local and global relations to create a per pixel saliency map. The global part is the dissimilarity to the average image color, representing a global color contrast. In addition, the image is blurred to decrease the influence of noise for the local contrast computation done by Differential of Gaussian filters.

Conditional random fields (CRF) are a machine learning approach that can learn to extract local features and to position them in a global context. A network of nodes is distributed over the image where each node influences its neighboring nodes. This network is structured into layers superimposed on each other. Higher layers represent global relations, lower layers the local feature extraction. This was used in [Low03] and combines multi-scale and local contrast together with the regional context and the spatial color distribution. An extension of this approach was proposed in [RWKP04, FKR07] where image segmentation is used additionally to group regions more accurately.

3. Method

Modern methods for AOI creation from eye tracking data utilize the semantic knowledge of the viewer about the stimulus material. Distinct fixation targets likely correspond to distinct AOIs. However, there are also cases where this association fails: In case AOI regions overlap (such as a face partially occluding another face) or in the case of an overall low tracking accuracy compared to the resolution of the stimulus material, the blurring operation that is required for the construction of a smooth heatmap can easily lead to a fusion of multiple AOIs. This is especially the case, if one of the regions is very dominant and thereby *consumes* weaker ones. A similar effect is associated with the choice of a minimum gaze frequency cutoff value for heatmap: As we are generally only interested in regions that are frequently looked at, we have to dispose of a lot of small, seldom looked at potential AOIs.

Due to those limitations, our idea is to integrate early features of the human visual system into determining AOI boundaries. Such a method generally over-segments the image (as can be seen in Figure 1e by the many spurious AOIs), but thus it is possible to distinguish AOIs even when one region is very dominant. The over-segmentation can easily be resolved once we apply the actual eye



(a) The Last Supper, Jacopo Tintoretto (b) Saliency Itti (c) AOI Itti (d) Saliency Frequency-tuned (e) AOI Frequency-tuned

(f) Paradise, Lukas Cranach (g) Saliency Itti (h) AOI Itti (i) Saliency Frequency-tuned (j) AOI Frequency-tuned

(k) Improvisation 9, Vasilii Kandinsky (l) Saliency Itti (m) AOI Itti (n) Saliency Frequency-tuned (o) AOI Frequency-tuned

(p) Shimmering Substance, Jackson Pollock (q) Saliency Itti (r) AOI Itti (s) Saliency Frequency-tuned (t) AOI Frequency-tuned

Figure 2: *The first column shows the stimulus images, the second column contains the respective generated [IKN98] saliency maps, and the third column presents the AOIs created based on these. Columns four and five contain the frequency-tuned [AHES09] saliency maps and their corresponding AOIs.*

tracking data to the AOIs, as many of them will contain only very few fixations and can therefore be filtered. The two steps of the method are:

1: Calculate the saliency map based on the stimulus image.
2: Compute AOIs based on gradients in the saliency map.

Figure 2 shows some stimulus images that we used to generate AOIs from. The first artwork (Figure 2a) is the famous "The Last Supper" by Jacopo Tintoretto. It illustrates a complex dark scene with bright spots at the gloriole and the hanging oil lamp. Those two regions are the most salient areas for both algorithms but due to the gradient-based AOI generation other parts are extracted as well (shown in Figure 2c and 2e). The second representative for artworks is "Paradise" by Lukas Cranach shown in Figure 2f. The generated AOIs (Figure 2h and 2j) of both algorithms separate the persons in the center of the image well. The regions extracted by [IKN98] are less detailed caused by a downscaling of the original image during the computation of the saliency map.

In addition to these classical paintings, some examples for abstract art are shown in (Figure 2k and 2p)), namely "Improvisation 9" by Vasilii Kandinsky and "Shimmering Substance" by Jackson Pollock. Figure 2m, 2o), 2r and 2t) show that the AOI generation based on the saliency over-segments the scene.

In the following we will compare the proposed approach with the following methods: Fixation mean-shift clustering [PS00, SD04], generating AOIs from a gaze heatmap based on its gradient [FKS*15] and based on thresholding [Woo02, Nys08].

4. Data description

For evaluation we used eye-tracking data recorded by [Ros14] at the University of Vienna. It consists of 40 participants, 20 experts in art and 20 novices. Data was recorded with an *IViewX RED 120* eye tracker on a 30-inch monitor with 2560×1600 pixels. Participants viewed the artwork for 2 minutes from a distance of 0.9 meters.

AOIs were generated jointly on data of all participants. A gaze heatmap of all participants is shown in Figure 3a. A visual comparison of Figures 3b and 3c show a strong overall similarity with only subtle differences. For the evaluation we calculated the following eye tracking key metrics for each AOI and participant (where a gaze point corresponds to a single eye tracker sample and a fixation to one physiological eye movement event): *time to first gaze (S_1), amount of gaze points (S_2), gaze points per minute (S_3), share of gaze points (S_4), total time of gaze points (S_5), minimal consecutive time of gaze points (S_6), maximal consecutive time of gaze points (S_7), average consecutive time of gaze points (S_8), time of first fixation (S_9), amount of fixations (S_{10}), fixations per minute (S_{11}), percentage of fixations (S_{12}), total time of fixations (S_{13}), minimal consecutive time of fixations (S_{14}), maximal consecutive time of fixations (S_{15}), average consecutive time of fixations (S_{16}).* Power of the following classification step could be increased by using transition features between the AOIs, which we omitted in this evaluation. Transitions would increase the number of features exponentially since they are also applicable in different kinds (global, relative to AOI, incoming or outgoing, transitions as saccade or scan path etc.). All statistics were collected based on the *average eye*, an average gaze position of the left and right eye. In addition, the *average eye* contains only data where both eyes have been detected by the tracker, resulting in good data quality.

In Figure 3d, 3e, 3f, 3g and 3h the generated AOIs for all compared methods are shown. As can be seen in Figure 3d, 3e and 3f, clustering and generating AOIs from the heatmap produces similar results for the most prominent AOIs. Figure 3g and 3h show the results of the two saliency maps.

5. Evaluation and Results

For evaluation of the quality of generated AOIs, we cross-validate support vector machine classifiers with the task of separating scanpaths of art experts from those of novices. Based upon the results of these classifications we evaluate the AOI segmentation quality. We employed MATLAB 2015b's support vector machine (SVM) configured as *'Standardize'=false, 'KernelFunction'='linear', 'KernelScale'='auto', 'OutlierFraction'=0.0, 'Nu'=0.5*, 100 as initial random seed, and a 20-fold cross-validation (20 experts and 20 novices). The features $\{S_1,\ldots,S_{16}\}$ were evaluated in different combinations, where k represents the amount of combined features. This means that for $k=3$ all possible triple combinations of features are evaluated. For automatic AOI generation five methods where analyzed:

Table 1: *Maximal classification result per k for each AOI generation algorithm (using all AOIs). Bold numbers represent the best performers per k.*

k	CLU	$HEAT_G$	$HEAT_T$	SAL_I	SAL_{FT}
1	0.60	0.5500	**0.7500**	0.6750	0.6250
2	0.6750	0.5750	**0.7500**	0.6750	0.6250
3	0.6750	0.5500	**0.7500**	0.6750	0.6250
4	0.6750	0.5500	**0.7500**	0.6750	0.5750
5	**0.6750**	0.5500	**0.6750**	**0.6750**	0.5750
6	**0.6750**	0.5250	0.6500	**0.6750**	0.5250

CLU Fixation mean-shift clustering [PS00, SD04].
HEAT_G Heatmap Gradient [FKS*15].
HEAT_T Heatmap Threshold [Woo02, Nys08].
SAL_I The proposed method using the saliency maps generated with Itti et al.'s method [IKN98].
SAL_FT The proposed method using the saliency maps generated with the Frequency-Tuned method [AHES09].

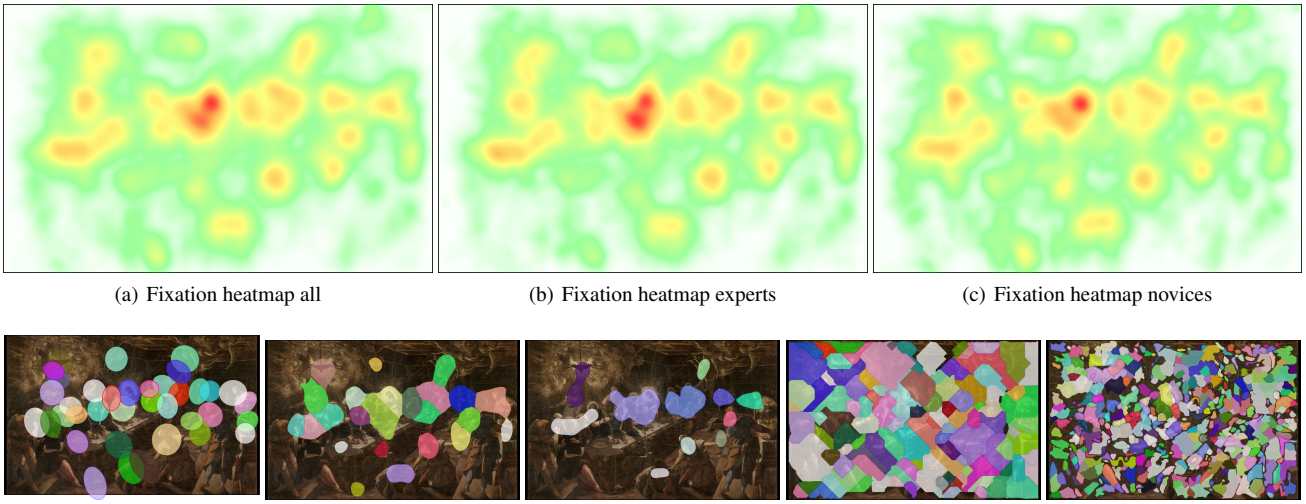
The simplest and most intuitive approach would be the comparison based on the best classification results. Those are shown in table 1. $HEAT_T$ has the best linear classification result for $k=1--4$ mainly due to feature S_12 (*share of fixations*) with the tightly fitted AOIs (shown in Figure 3f).

While the heatmap threshold method reaches superior performance in Table 1, it should be noted that the saliency-based methods can keep up with the other data-driven methods. Instead of looking at the overall best classification accuracy, we could also investigate the robustness of the classification for different feature sets. This is an important consideration, as we want AOIs that show us as many inter-group effects as possible, not only the strongest ones.

Another way to evaluate the quality of the AOIs is therefore their stability across different feature sets. This means that we analyze the capability of the AOIs to extract information out of the statistical values. Therefore, we use the entire set of combinations possibilities per k and compute the mean and standard deviation. We evaluated all combination possibilities up to a maximum of $k=6$. This was done to reduce the amount of evaluations which increase exponentially. The mean score for a method is then given by

$$S_\mu(AOI_m,CS,k)=\frac{\sum_{i=1}^{\binom{|S|}{k}} CVS(AOI_m,CS_i)}{\binom{|S|}{k}}, \quad (1)$$

Where m represents the AOI generation method, $|S|$ is the amount of statistical values, CS_i the Combination Set of statistical values, $\binom{|S|}{k}$ the binomial coefficient and $CVS()$ the Cross-Validation Score for the classification. With the mean $S_\mu(AOI_m,CS,k)$ we can



(a) Fixation heatmap all (b) Fixation heatmap experts (c) Fixation heatmap novices

(d) AOIs from clusters (e) AOIs from heatmap gradient (f) AOIs heatmap threshold (g) AOIs from Itti (h) AOIs frequency-tuned

Figure 3: *The first row shows fixation heatmaps. a) all participants, b) only the experts and c) only novices. Determining differences in these heatmaps is non-trivial. The second row shows the generated AOIs for d) the fixation clustering, e) gradient heatmap segmentation, f) threshold heatmap segmentation, g) Itti saliency map and h) frequency-tuned saliency map.*

Table 2: *The calculated score (mean) for each method using equation 1 (using all AOIs). Bold numbers represent the highest mean value per k.*

k	CLU	$HEAT_G$	$HEAT_T$	SAL_I	SAL_{FT}
1	0.4328	0.4516	**0.5391**	0.5016	0.4438
2	0.4290	0.4525	0.5073	**0.5198**	0.4148
3	0.4250	0.4508	0.4885	**0.5283**	0.4085
4	0.4252	0.4505	0.4780	**0.5375**	0.4072
5	0.4284	0.4502	0.4696	**0.5470**	0.4068
6	0.4335	0.4500	0.4608	**0.5564**	0.4060

Table 3: *The calculated standard deviation for each method using equation 2 (using all AOIs). Bold numbers represent the lowest standard deviation per k.*

k	CLU	$HEAT_G$	$HEAT_T$	SAL_I	SAL_{FT}
1	**0.0884**	0.0716	0.1208	0.1192	0.0892
2	0.0849	0.0645	0.1077	0.1002	**0.0520**
3	0.0820	0.0531	0.0925	0.0837	**0.0328**
4	0.0790	0.0454	0.0835	0.0716	**0.0227**
5	0.0769	0.0413	0.0791	0.0641	**0.0189**
6	0.0753	0.0394	0.0764	0.0603	**0.0182**

define the standard deviation as in equation 2.

$$\sqrt{\frac{\sum_{i=1}^{\binom{|S|}{k}} (CVS(AOI_m, CS_i) - S_\mu(AOI_m, CS, k))}{\binom{|S|}{k} - 1}} \tag{2}$$

The best result for one feature (S_1) using Equation 1 is reached by $HEAT_T$ which are the most restrictive AOIs (first row Table 2). For the other feature combinations (S_{2--16}) it is outperformed by SAL_I. It can also be observed in Table 2 that the only method continuously improving its score is SAL_I while the others decrease in linear classification performance. This means that SAL_I is the overall most robust AOI set and constantly over chance level. In Table 3 the standard deviations are shown. Those values hold information about the reliability of the results from Table 2 which is indicated by a hardly fluctuating value and therefore a low standard deviation. As can be seen the reliability for higher feature combination increases for all AOI generation algorithms. In addition the Whisker plot for all feature combinations ($k = 1 - -6$) are shown in Figure 4. Here the higher feature combinations dominate the result. This is because more features allow more combination possibilities and are therefore evaluated more often.

Again it can be seen that SAL_I outperforms the other approaches especially if the 25% and 75% percentiles are considered.

These results indicate that AOIs generated on the saliency map proposed by [IKN98] can be applied to art viewing experiments. As they can be computed even before the data is recorded, they could be used for online classification scenarios without the need of recording large amounts of data per piece of art.

6. Conclusion

We proposed a procedure for generating AOIs based on saliency maps. In addition, we showed that our approach is also applicable to scenes and images where it is not possible to manually annotate AOIs. This means that our approach is applicable to abstract art.

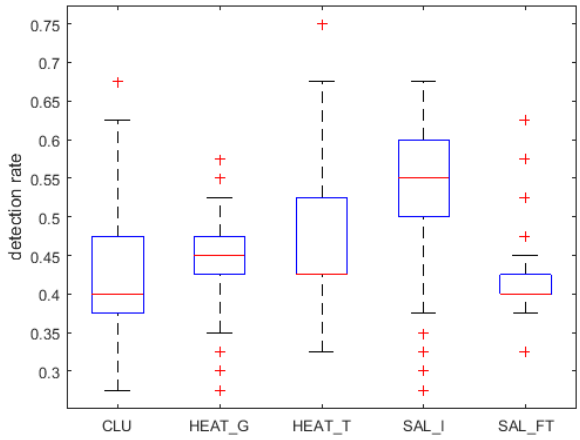


Figure 4: *Whisker plot over all feature combinations S_{1--16}. The red line is the median. The blue box represents the 25% and 75% percentiles. Black horizontal lines are the minimum and maximum of the evaluated data and the red crosses represent outliers. It has to be mentioned that higher feature set combinations are over-represented due to the higher combination possibilities.*

Figure 5: *Wassily Kandinskys "Untitled", dated 1910, is one of the first abstract artworks.*

In such artworks, it is impossible for a human to define objects or areas (Figure 5).

Our evaluation showed that saliency AOIs can reach maximal classification accuracy similar to data-driven AOIs. The data-driven approach needs large amounts of data to generate generalized AOIs. For the saliency-based AOI generation, this limitation does not apply. In addition, the proposed evaluation procedure does not only account for the best possible result but also investigates the quality and stability of key metrics. In case of this stability the saliency model from Itti outperforms the frequency based approach and the data-driven models. In practice, this stability over different input combinations is important as the evaluation of all possibilities is too costly for large amounts of data. This applies in particular if the parameters for the AOI generation algorithm have to be estimated too (no parameters for [FKS*15]). In our evaluation only statistical features are used. Modern feature extractors like SubsMatch [KRS*17b] or ScanGraph [DP16] consider also the graph of the scan path and sequences of fixations. Therefore, the obtainable classification results using those approaches are higher. In this paper we evaluated the stability of AOIs computed on saliency maps which allows a completely automatic extraction which is the basis for the aforementioned algorithms to be applicable online. Along with image features (SIFT [Low99], ORB [RRKB11], MSER [MCUP04]) to assign AOIs from different images, Saliency based AOIs are even applicable for dynamic scenes like videos.

Future research will go into the usability of our approach for online classification scenarios on abstract, figurative and concrete art. This will give further insight into the composition of museum visitors and may help with the arrangement of art and their presentation. Additionally, this insight could indicate how art or in general images can be grouped by human attention. This is only one scenario which is strongly related to art. Other scenarios like expertise level prediction or attention level are also applicable by preliminary and automatically defined AOIs. A combinatorial approach where the AOIs adapt to the new data with the saliency-based AOIs as initial estimates is also conceivable to improve online classification further.

References

[AHES09] ACHANTA R., HEMAMI S., ESTRADA F., SUSSTRUNK S.: Frequency-tuned salient region detection. In *IEEE Conference on Computer Vision and Pattern Recognition* (June 2009), pp. 1597–1604. doi:10.1109/CVPR.2009.5206596. 3, 4, 5, 8

[AMFM11] ARBELAEZ P., MAIRE M., FOWLKES C., MALIK J.: Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 33*, 5 (May 2011), 898–916. doi:10.1109/TPAMI.2010.161. 3

[BI13] BORJI A., ITTI L.: State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence 35*, 1 (Jan 2013), 185–207. doi:10.1109/TPAMI.2012.89. 2

[Bra78] BRACEWELL R.: *The Fourier Transform and its Applications*, second ed. McGraw-Hill Kogakusha, Ltd., Tokyo, 1978. 3

[CMH*15] CHENG M. M., MITRA N. J., HUANG X., TORR P. H. S., HU S. M.: Global contrast based salient region detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence 37*, 3 (March 2015), 569–582. doi:10.1109/TPAMI.2014.2345401. 3

[CMTG10] CRISTINO F., MATHÔT S., THEEUWES J., GILCHRIST I. D.: Scanmatch: A novel method for comparing fixation sequences. *Behavior Research Methods 42*, 3 (2010), 692–700. 1, 2

[DDJ*10] DUCHOWSKI A. T., DRIVER J., JOLAOSO S., TAN W., RAMEY B. N., ROBBINS A.: Scanpath comparison revisited. In *Proceedings of the 2010 Symposium on Eye-Tracking Research and Applications* (New York, NY, USA, 2010), ETRA '10, ACM, pp. 219–226. doi:10.1145/1743666.1743719. 2

[DNJ*12] DEWHURST R., NYSTRÖM M., JARODZKA H., FOULSHAM T., JOHANSSON R., HOLMQVIST K.: It depends on how you look at it: Scanpath comparison in multiple dimensions with multimatch, a vector-based approach. *Behavior Research Methods 44*, 4 (Dec 2012), 1079–1100. doi:10.3758/s13428-012-0212-2. 1

[DP16] DOLEZALOVA J., POPELKA S.: Scangraph: A novel scanpath comparison method using visualisation of graph cliques. *Journal of Eye Movement Research 9*, 4 (2016). 7

[FKR07] FRINTROP S., KLODT M., ROME E.: A real-time visual attention system using integral images. In *International Conference on Computer Vision Systems* (2007), pp. 1–10. 3

[FKS*15] FUHL W., KÜBLER T. C., SIPPEL K., ROSENSTIEL W., KASNECI E.: Arbitrarily shaped areas of interest based on gaze density gradient. In *European Conference on Eye Movements* (08 2015). 1, 3, 5, 7

[GMZ08] GUO C., MA Q., ZHANG L.: Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2008), IEEE Computer Society. 3

[GZMT12] GOFERMAN S., ZELNIK-MANOR L., TAL A.: Context-aware saliency detection. *IEEE Pattern Analysis and Machine Intelligence 34*, 10 (2012), 1915–1926. 3

[HKvdBH16] HESSELS R. S., KEMNER C., VAN DEN BOOMEN C., HOOGE I. T. C.: The area-of-interest problem in eyetracking research: A noise-robust solution for face and sparse stimuli. *Behavior Research Methods 48*, 4 (Dec 2016), 1694–1712. doi:10.3758/s13428-015-0676-y. 2

[HZ07] HOU X., ZHANG L.: Saliency detection: A spectral residual approach. In *IEEE Conference on Computer Vision and Pattern Recognition* (June 2007), pp. 1–8. doi:10.1109/CVPR.2007.383267. 3

[IB05] ITTI L., BALDI P.: Bayesian surprise attracts human attention. In *Neural Information Processing Systems* (2005), pp. 547–554. 3

[IKN98] ITTI L., KOCH C., NIEBUR E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell. 20*, 11 (Nov. 1998), 1254–1259. doi:10.1109/34.730558. 3, 4, 5, 6, 8

[KRS*17a] KÜBLER T. C., ROTHE C., SCHIEFER U., ROSENSTIEL W., KASNECI E.: Subsmatch 2.0: Scanpath comparison and classification based on subsequence frequencies. *Behavior Research Methods 49*, 3 (2017), 1048–1064. doi:10.3758/s13428-016-0765-6.

[KRS*17b] KÜBLER T. C., ROTHE C., SCHIEFER U., ROSENSTIEL W., KASNECI E.: Subsmatch 2.0: Scanpath comparison and classification based on subsequence frequencies. *Behavior Research Methods 49*, 3 (Jun 2017), 1048–1064. doi:10.3758/s13428-016-0765-6. 7

[KU87] KOCH C., ULLMAN S.: *Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry*. Springer Netherlands, Dordrecht, 1987, pp. 115–141. doi:10.1007/978-94-009-3833-5_5. 3

[Low99] LOWE D. G.: Object recognition from local scale-invariant features. In *IEEE international conference on Computer Vision* (1999), Ieee, pp. 1150–1157. 7

[Low03] LOWE D.: Distinctive image features from scale-invariant keypoints. In *International Journal of Computer Vision* (2003). 3

[LYS*11] LIU T., YUAN Z., SUN J., WANG J., ZHENG N., TANG X., SHUM H.-Y.: Learning to detect a salient object. *IEEE Pattern Analysis and Machine Intelligence 33*, 2 (2011), 353–367. 3

[MCUP04] MATAS J., CHUM O., URBAN M., PAJDLA T.: Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing 22*, 10 (2004), 761–767. 7

[MZ03] MA Y.-F., ZHANG H.: Contrast-based image attention analysis by using fuzzy growing. In *ACM Multimedia* (2003), Rowe L. A., Vin H. M., Plagemann T., Shenoy P. J., Smith J. R., (Eds.), ACM, pp. 374–381. 3

[Nys08] NYSTRÖM M.: *Off-line Foveated Compression and Scene Perception: An Eye-Tracking Approach*. PhD thesis, Lund University, 2008. 1, 3, 5

[PS98] PRIVITERA C. M., STARK L. W.: Evaluating image processing algorithms that predict regions of interest. *Pattern Recognition Letters 19*, 11 (Sept. 1998), 1037–1043. doi:10.1016/S0167-8655(98)00077-4. 1, 3

[PS00] PRIVITERA C. M., STARK L. W.: Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22*, 9 (Sept. 2000), 970–982. doi:10.1109/34.877520. 1, 3, 5

[Ros14] ROSENBERG R.: Blicke messen: Vorschläge für eine empirische bildwissenschaft. *Jahrbuch der Bayerischen Akademie der Schönen Künste 27* (2014), 71–86. doi:10.11588/artdok.00003028. 5

[RRKB11] RUBLEE E., RABAUD V., KONOLIGE K., BRADSKI G.: Orb: An efficient alternative to sift or surf. In *IEEE international conference on Computer Vision* (2011), IEEE, pp. 2564–2571. 7

[RWKP04] RUTISHAUSER U., WALTHER D., KOCH C., PERONA P.: Is bottom-up attention useful for object recognition? In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (June 2004), pp. 37–44. doi:10.1109/CVPR.2004.1315142. 3

[SD04] SANTELLA A., DECARLO D.: Robust clustering of eye movement recordings for quantification of visual interest. In *Proceedings of the 2004 Symposium on Eye Tracking Research and Applications* (New York, NY, USA, 2004), ETRA '04, ACM, pp. 27–34. doi:10.1145/968363.968368. 1, 3, 5

[SFKK16] SANTINI T., FUHL W., KÜBLER T., KASNECI E.: Bayesian identification of fixations, saccades, and smooth pursuits. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications* (New York, NY, USA, 2016), ETRA '16, ACM, pp. 163–170. doi:10.1145/2857491.2857512. 1

[SG00] SALVUCCI D. D., GOLDBERG J. H.: Identifying fixations and saccades in eye-tracking protocols. In *Symposium on Eye Tracking Research & Applications* (New York, NY, USA, 2000), ETRA '00, ACM, pp. 71–78. doi:10.1145/355017.355028.

[SM00] SHI J., MALIK J.: Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22*, 8 (Aug. 2000), 888–905. doi:10.1109/34.868688. 3

[TKK*13] TAFAJ E., KÜBLER T. C., KASNECI G., ROSENSTIEL W., BOGDAN M.: Online classification of eye tracking data for automated analysis of traffic hazard perception. In *International Conference on Artificial Neural Networks* (Berlin, Heidelberg, 2013), Mladenov V., Koprinkova-Hristova P., Palm G., Villa A. E. P., Appollini B., Kasabov N., (Eds.), Springer Berlin Heidelberg, pp. 442–450. doi:10.1007/978-3-642-40728-4_56. 1

[WKI*11] WANG M., KONRAD J., ISHWAR P., JING K., ROWLEY H. A.: Image saliency: From intrinsic to extrinsic context. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2011), IEEE Computer Society, pp. 417–424. 3

[Woo02] WOODING D. S.: Eye movements of large populations: Ii. deriving regions of interest, coverage, and similarity using fixation maps. *Behavior Research Methods 34*, 4 (2002), 518–528. doi:10.3758/BF03195481. 3, 5

7. Appendix

Table 4: *Used parameters for calculations.*

Fixation	Method: Standard Minimum duration: 50ms Maximum radius: 100px Outliers allowed: 2
Cluster	Method: Mean-Shift Minimum fixations: 80 Search radius: 50px
Heatmap	Input: Gaze points Gauss σ: 50
AOI gradient	Prethreshold: 15% Min region size: 50px
AOI threshold	Prethreshold: 20% Threshold: 50% Min region size: 50px Window size: 200px
Saliency map [IKN98]	Local maxima: 14 Gabor size: 14 Gabor σ: 100
Saliency map [AHES09]	Gauss size: 250px Gauss σ: 90