

Aesthetic appraisal of art - from eye movements to computers

C. Wallraven¹, D. Cunningham², J. Rigau³, M. Feixas³, M. Sbert³

¹Max Planck Institute for Biological Cybernetics, Germany

²WSI-GRIS, University of Tübingen, Germany

³Graphics and Imaging Laboratory, University of Girona, Spain

Abstract

By looking at a work of art, an observer enters into a dialogue. In this work, we attempt to analyze this dialogue with both behavioral and computational tools. In two experiments, observers were asked to look at a large number of paintings from different art periods and to rate their visual complexity, or their aesthetic appeal. During these two tasks, their eye movements were recorded. The complexity and aesthetic ratings show clear preferences for certain artistic styles and were based on both low-level and high-level criteria. Eye movements reveal the time course of the aesthetic dialogue as observers try to interpret and understand the painting. Computational analyses of both the ratings (using measures derived from information theory) and the eye tracking data (using two models of saliency) showed that our computational tools are already able to explain some properties of this dialogue.

1. Introduction

The investigation of aesthetic judgments of art seems difficult at first: if scholars disagree about what constitutes art, similarly the philosophy of aesthetics has enjoyed a lively debate over the past centuries about suitable definitions of aesthetics. Here we attempt to add to this discussion by following a constructivist, empirical approach that tries to extract information about aesthetics in two ways: directly, by asking viewers about their aesthetic judgments, and indirectly, by observing how their eyes move during the aesthetic experience of watching a piece of art.

What can eye movements tell us about the processes and thoughts of the observer? In what has become the most reproduced image in the context of eye movements studies, Alfred Yarbus in 1967 analyzed the scan paths of an observer viewing I.E. Repin's painting entitled 'They Did Not Expect Him' [Yar67]. What he found is that observers were focusing their eyes on specific parts of the painting (such as persons, important objects, etc.) rather than randomly scanning the image. In addition, he found that the scanpath, i.e., the trajectories of eye movements, critically depended on the task or context with which the image was viewed: if observers were asked to look for people, their gaze focused on people only, whereas if they were asked to simply look at the painting, many more fixations occurred also in different areas. Ever since this time, researchers have used eye tracking

to investigate what kind of visual information is focused on depending on the task.

In this paper, we want to focus on what eye movements predict about how art is judged. As the previous example (and many more, see below) has shown, perception of art has always been a favorite topic in eye movement studies. Usually, however, the number of paintings that were analyzed in these studies was rather small and—whereas considerable progress has been made over the past decades—computational analyses of information that observers' eye movements could be correlated to were scarce. Our goal, therefore, is to use a larger, well-controlled set of paintings which spans multiple art periods and to correlate gaze behavior (as well as the ratings) of observers with specific measures derived from information theory. Additionally, we were interested in seeing how gaze behavior would change depending on the task which was given to the observers—does gaze behavior differ for a relatively simple task such as judging the visual complexity of a painting, compared to a higher-level task such as judging its aesthetic impression?

2. Related work

Empirical studies of aesthetics: In what is perhaps the most general framework for describing the aesthetic experience of art [LBOA04] discuss five stages: perceptual analyses, implicit memory integration, explicit classification, cognitive mastering and evaluation. The output of the frame-

work consists of both aesthetic pleasure (an emotive component) and aesthetic judgment (a cognitive component). This model inspired studies on the role of expertise in aesthetic appreciation of art [AL06], as well as on how stylistic features are extracted following content-based information [ALHC08]. In [WCF08], further steps towards an empirical investigation of the 'explicit classification' step of the model by [LBOA04] were undertaken. Participants (who were not trained in the arts or art history) had to cluster 275 images from different art periods into consistent clusters based on 'style of painting'.

Eye tracking and art: The first investigation into visual processing using eye tracking was conducted in the early thirties [Bus35]. Participants viewed photographs of various types of artworks, and the resulting scan paths were analyzed. Interestingly, the study found no differences between trained and untrained artists (although see [NLK93], who found differences between those two groups using a more detailed analysis). Furthermore, no two scan-paths were exactly identical thus demonstrating individual differences in gaze behavior, although global patterns could be derived that defined two overall search strategies of global and local scan paths. Participants also tended to focus on high-contrast regions in the foreground (including faces and people).

In perhaps the most well-known study, Alfred Yarbus in 1967 analyzed the scan paths of an observer viewing I.E. Repin's painting entitled 'They Did Not Expect Him' [Yar67]. One of the core findings was that eye movement patterns changed dramatically depending on viewing instructions: for each type of instruction, Yarbus found that the most informative regions of the image were scanned and processed. Similarly, in experiments by Molner [Mol81], participants who were told that they were going to be asked about the aesthetic qualities of a painting made longer fixations than participants who were going to be asked about the contents. Recent experiments on the role of task context on eye-movements have found that fixations fall primarily on task-relevant objects: studies of fixations made during making a cup of tea [LMR99], for example, revealed that during pouring of the tea, participants viewed the spout of the teapot.

Recently, Locher et al. have reported [LKMTN07] evidence for a two-stage model, in which viewers of both representational and abstract art in the first two seconds tend to do a global sweep of the image. This sweep covers roughly a quarter of the display after which viewers concentrate on finer details. The overall coverage increases during this subsequent viewing time to 38% by 7 s and to 46% by 30s - the latter being a non-significant increase (for a further review of research done by Locher et al., see [Loc06]). These findings indicate a 'coarse-to-fine' strategy in which viewers first analyze the 'gist' of the image and then later focus on finer details - further evidence for this is provided in the study by data indicating that participants provide very similar descriptions of the image given only a short viewing time as opposed to having unlimited time for visual exploration.

Finally, in [WKK*07], eye tracking was used to determine how viewers would interpret so-called 'indeterminate art' - art which at first glance seems to have a clear interpretation but at a closer look does not even consist of coherent objects. Participants were instructed either to report whether such an image contained a person or to which category the image belonged. Analysis of the eye movements revealed clear differences between the two tasks: a preference for a more global search strategy in the categorization task compared to the person detection task (this is in accordance with current theories of scene processing). In addition, high-contrast regions were preferred over low-contrast regions with only immediately relevant parts of the scene being fixated following the coarse-to-fine search pattern identified by [LKMTN07].

Computational studies of eye tracking: All of the above studies have mainly focused on descriptive analyses of fixation patterns - computational analyses of the visual information contained in fixations of natural images [KRH*00] have found that in comparison to random regions, fixations contain higher spatial variance, i.e., a higher information content. Those regions often correspond to second-order features such as curves, corners, and occlusions rather than first-order changes in intensity such as straight lines and edges. In one of the most well-known models of attentional processing [IK01], several computational features (based on color, intensities, and orientations of image gradients) are integrated into a so-called 'saliency map' which predicts salient regions in an image—that is, regions which are likely to draw attention to them based on their low-level properties. The model has been shown to account for a significant proportion of fixations participants made while free-viewing different images [PLN02]. Correlation decreased, however, over time as top-down processes in image exploration took over.

3. Experimental design and methods

In the following, we will describe the eye tracking experiment that was conducted to investigate the appraisal of art. In order to investigate both low-level and high-level dimensions of the aesthetic experience, we asked participants to view a set of paintings and to solve two different visual tasks: rate the visual complexity of the painting (low-level) and rate the aesthetic appeal of the painting (high-level). We then conducted a number of analyses comparing the behavioral results to computational measures of complexity and information content, as well as comparing the eye tracking results to both the behavioral and the computational measures.

3.1. Stimuli

The data set is taken from [WCF08] in which categorization of artworks was investigated. It includes 11 art periods spanning larger, cultural movements in art history: Gothic, Renaissance, Baroque, Rokoko, Classicism, Romanticism, Realism, Impressionism, Expressionism, Surrealism, and Postmodernism. For each of these art periods,



Figure 1: Example images for each art period from Gothics to Postmodernism.

5 major artists were identified, and for each artist, 5 representative paintings were chosen. All images were taken from www.prometheus-bildarchiv.de - an online library which provides access to a variety of large, online collections of art. Figure 1 shows example paintings from the 11 different art periods.

3.2. Experimental design

A total of 275 paintings were shown to a total of 20 participants split into two groups of 10 participants each. Participants, all of them with normal or corrected-to-normal vision, viewed the paintings from a distance of 56cm while resting their head on a chin-rest. Eye tracking data was collected with an 'Iview'-Eye-tracker using a master PC for stimulus display with the software package Psychtoolbox [Bra97] and a synchronized control PC for recording of the data. All images were shown centred on a 21" monitor at a resolution of 1024x768 pixels - as the paintings were of different formats, images were resized to a maximum width of 800 pixels, and a maximum height of 600 pixels, respectively - on average, images subtended $\approx 25^\circ \times 25^\circ$ of visual angle.

Before each experiment, a calibration of the eye-tracker had to be done in order to align the camera of the 'Iview'-System to each participant. For this, we used a standard 9-point calibration sheet, which consisted of nine points with fixed pixel positions that were distributed on the screen and which participants were asked to fixate in a fixed order.

Each of the 275 (randomized) trials consisted of a fixation cross that was shown for 2 seconds followed by an image that was shown for 6 seconds. Participants were asked not to direct their gaze outside of the image area during this period. Afterwards, a question for each of the two tasks was shown on the screen and participants had to answer by selecting the appropriate item on a list. For the complexity task, the question was 'How complex was the image visually', whereas for the aesthetics task, the question was 'How aesthetically appealing was the image?'—participants rated the two values on a scale of 1 (not complex/aesthetic) to 7 (highly complex/aesthetic). We did not impose a time limit

for their answer—response time was nevertheless recorded. A keypress started the next trial. In order to prevent fatigue and to ensure that participants were able to maintain their position on the chin rest, the 275 trials were interrupted by four breaks. After each break, the calibration was repeated.

After the experiment, participants were asked to fill out a questionnaire which gathered data about how difficult the experiment was perceived, what strategies were used in answering the two tasks, and how well participants knew the different art periods that occurred in our data set. For the analysis of the experiment, in the following we will discuss behavioral data (consisting of two different ratings) and eye tracking data (individual fixations for each image).

3.3. Computational measures

The first set of measures used for correlation with the ratings are based on the diversity of the palette, the compressibility of the image, the structural complexity of the image, and the information associated to the colors of the image. The *palette entropy* H expresses the average uncertainty or information content of a pixel and $N \times H$ represents the information content of an image (in bits). The Shannon entropy can be seen as a measure of the *palette complexity* [RFS08]. The *Kolmogorov complexity* $K(x)$ of an image x is the length of the shortest program to compute x on an appropriate universal computer [LV97] and can be interpreted as a measure of the descriptonal complexity of an image. Due to the non-computability of $K(x)$, JPEG compression is used for estimation—in our experiments, $K(x)$ has been normalized by the initial image size. The *excess entropy* E [FC03] of pixel intensities is a measure of their correlation or structural complexity and can be interpreted as the degree of predictability of a pixel given its neighbors. In our experiments, the computation of E is done on 256 intensity values and blocks of 3 consecutive pixels.

In order to compare the eye tracking data with predictions made by saliency models, we implemented two different versions of models. The first model is based on the well-known Itti and Koch model of visual saliency [IK01] and computes

three different saliency maps from an image [WK06]: images are first resampled into a Gaussian pyramid at multiple resolutions; each image is separated into an intensity channel, and two color-opponencies Red-Green and Blue-Yellow; the intensity image is then filtered with a Gaussian kernel at four different orientations to obtain orientation maps; centre-surround operations then combine two resolution maps into so-called feature maps. These feature maps are then summed over all resolutions and integrated into one saliency map. For all of these computations, we retained the default parameters set by the implementation of [WK06].

The second model is given by Color Mutual Information (CMI) [RFS08] and measures the degree of dependence between the color c and the regions of the painting and describes information or saliency associated with the color c . For a given value of mutual information (in our case, 50%), a BSP-partitioning algorithm subdivides an image into quasi-homogeneous regions, yielding a saliency map based on information content.

On a given saliency map, a winner-take-all neural network is used to determine the first attended location (that is, the first fixation made by the computer). In the next iteration, a 70x70 pixel region around the object gets suppressed and the next-highest location in the saliency map is obtained until 25 fixations have been made by the computer.

4. Analysis of Behavioral Data

Complexity Task: Figure 2a shows the ratings for the complexity task. Overall, images were rated as being moderately complex with a value of 3.96 ± 0.06 SEM (standard error of the mean). Ratings ranged from 1.3 to 6.4 for the least and most complex painting, respectively which are shown in Figure 2b,c. As one might have expected, the minimalist art by Newman was judged as relatively simple, whereas the extremely intense and dense depiction of the 'Ballsouper' by Menzel was seen as very complex. Inter-rater agreement as measured by correlating all possible pairs of raters and calculating the average Pearson- r : $r = 0.45$ (all associated p -values were highly significant) which implies a significant agreement among participants. Given that we did not explicitly anchor the complexity scale, this value is still very high compared to other typical rating results: average inter-rater agreement for judging personality traits ranges, for example, is $r = 0.27$ [FC88].

An analysis of variance (ANOVA) revealed a highly significant effect of art period ($F(10,90)=6.12$, $p < 0.001$). The post-hoc Scheffe criterion revealed any difference of 0.44 between two means to be significant—the Postmodernist period, for example, was consistently ranked to be of lowest complexity, whereas the Gothic, Renaissance, Romanticism, and Surrealist periods had the highest ratings. Interestingly, the *variance* of the ratings was highest for the Gothic, Expressionist, and Surrealist periods. Overall, however, the strongest effect on the ratings was the drop in complexity

for the Postmodernist period—showing that the abstract and minimalist works of that time clearly have a common 'look'.

Aesthetics Task: Figure 3a plots the ratings for the aesthetics task. Again, images were overall rated as being moderately aesthetically pleasing with a value of 3.71 ± 0.09 SEM. Ratings varied from 1.7 to 6.0 for the least and most aesthetically pleasing painting, respectively (shown in Figure 3b,c). Inter-rater agreement for this task was highly significant, but decidedly lower at $r = 0.28$ showing that the inherent difficulty of the task.

An analysis of variance (ANOVA) revealed a highly significant effect of art period ($F(10,90)=11.14$, $p < 0.001$). The post-hoc Scheffe criterion showed that any difference of 0.52 between the two means would become significant. Art periods with a high degree of aesthetic appeal were, for example, Romanticism, Realism, and Impressionism. Many of the most aesthetically appealing paintings contained a rich, colorful palette (such as 3c) and depicted natural outdoor scenes. We observed a remarkable historical trend in that aesthetic appeal starts to rise steadily from Gothic to Romanticism, forming a plateau up to the Impressionist period, after which aesthetic scores drop drastically, with the lowest appeal for the Postmodern period. This pattern first of all, shows a general trend in our participant group towards classical (that is, realistic) art with the modern art periods being less aesthetically pleasing. This finding also corresponds well to the results obtained in [WCF08], where participants' sorting behavior of artworks clearly separated pre-Impressionist from post-Impressionist/modern art—something that might be explained using these aesthetic preferences.

Task Comparison: As is already evident from Figures 2 and 3, the dimension of complexity seems ill-suited to fully explain the dimension of aesthetic appeal. A closer look at single items does reveal some cases in which images that were rated as more complex also were rated as more aesthetically pleasing, but the opposite also happens frequently. This finding is confirmed by calculating the correlation between the ratings of the two experiments: overall, we obtained a non-significant correlation of $r = 0.1$ ($p = 0.23$). In terms of response times, a post-hoc t-test showed that the aesthetics task did not take participants significantly longer than the complexity task ($t(18)=0.64$, $p=0.65$)—this demonstrates that judging the aesthetic appeal of an image is as easy as it is to judge its visual complexity.

Questionnaires: Overall, both tasks were rated as being easy with an average difficulty of 2.3 (on a scale of 1 to 5). None of the participants reported to have a problem with identifying aesthetically pleasing images and everyone found this and the complexity task to be intuitive. Average knowledge of/about art was low for both participant groups (2.4 in the complexity group, and 1.8 in the aesthetics group) showing that we were querying truly 'naive' participants.

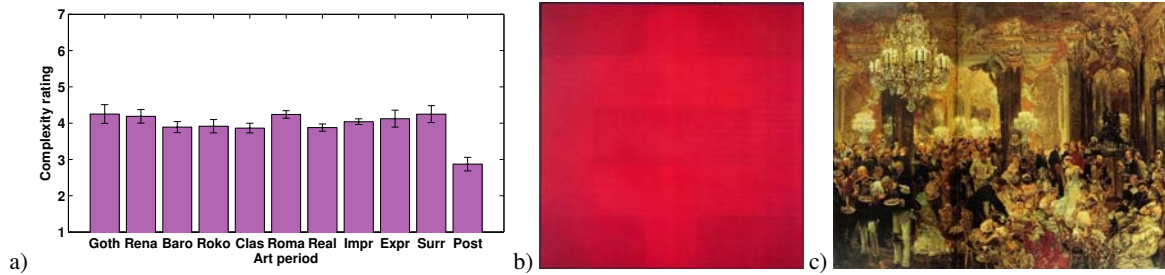


Figure 2: a) Complexity ratings broken down by art period, b) least and c) most complex artwork.

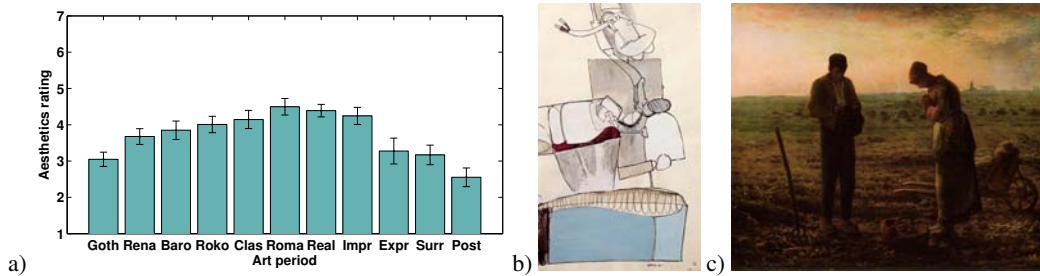


Figure 3: a) Aesthetics ratings broken down by art period, b) least and c) most aesthetically appealing artwork.

In terms of strategies that participants used to identify visually complex images, the features that correlated with a high complexity were (number in brackets indicate how many participants mentioned this feature): large number of objects (7), high level of detail (7), more contrast and/or colors (5), visibility of the brushstrokes (4), representational paintings being more complex than abstract ones (2). The analysis of visual complexity thus relied on both mid-level features (such as identifying objects) as well as on lower-level features (such as color palette, or brushstrokes)—see also [WCF08].

Features that participants used to judge the aesthetic appeal of a painting were: holistic "I like this painting" (6), appealing content (5), appealing choice of colors (5), abstract less appealing than representational (4), skillful execution of painting (2). These results fit very well with the overall data discussed above in that abstraction was seen as less aesthetically pleasing than representational images where interpretation of the image content was easier. In addition, many participants mentioned that a vivid choice of colors was more appealing to them—meaning, for example, that Gothic paintings with their limited palette were seen as less aesthetically pleasing. Similarly to the complexity task, participants mentioned a whole range of features from low-level (colors) to mid-level (skillful execution of the painting, ready interpretation). In contrast to the previous task, a 'holistic' overall impression was mentioned many times which demonstrates that even though participants found it easy to do the task, the verbalization of the important dimensions was sometimes difficult.

Correlation with computational measures: In the following, correlation analyses were conducted between the com-

putational features H, K, E (and some combinations thereof) and the behavioral data averaged over all participants for each of the two tasks. Perhaps not surprisingly given the complexity of the task, of the three measures, H was the only one to correlate weakly with $r_{H,aes} = 0.14$ with the aesthetics ratings. All measures, however, correlated significantly with the complexity ratings: $r_{H,comp} = 0.48$, $r_{K,comp} = -0.44$, $r_{E,comp} = 0.40$. These correlations were not fully dependent, however, such that a simple average of all three measures correlated better at $r_{HKE,comp} = 0.54$.

More importantly, however, and similarly to the behavioral data, the correlation values were highly variable across art periods and measures: color complexity correlated significantly with ratings on 7 art periods with high values for the Gothic and Classicism periods, compressibility correlated with 4 art periods with a high value for the Postmodern period, and excess entropy correlated extremely well with only the Gothic and Postmodern periods. None of the measures correlated with Expressionism or Surrealism. This demonstrates that *different* criteria were used by participants to evaluate complexity thus confirming the questionnaires results in which a variety of image features were mentioned.

5. Analysis of eye tracking data

The raw data of the eyetracker still contains saccades and eye blinks which need to be removed for the final analysis. For this, we used a toolbox for Matlab called 'Ilab' [Git02] which processes the raw eye tracking data to extract the fixations. This requires the specification of two parameters: the minimum duration t_{fix} ms for which the eye gaze stays within a given region of radius r_{fix} pixels. Here, we chose $t_{fix} = 100$ ms and $r_{fix} = 5$ pixels, where the latter corresponds to a fixation region of $\approx 1^\circ$ of visual angle.

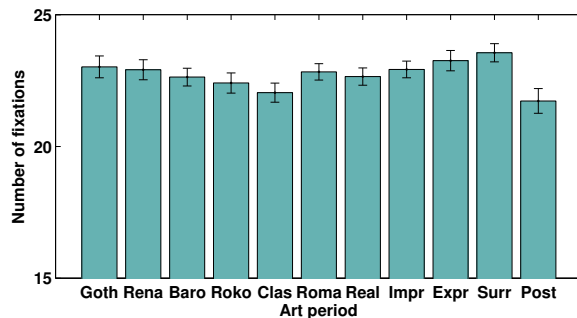


Figure 4: Number of fixations broken down by art period.

The fixation analysis is visualized in this paper using the following procedure: First, for each painting and each task, we place at all fixation locations for a single participant a Gaussian 'blob' of diameter $\approx 2^\circ$ visual angle which represents a rough approximation of the foveated area. These individual maps are then averaged over all 10 participants and visualized with a 'jet' color palette ranging from blue (no fixations) to red (many fixations). Finally, in order to aid interpretation of the fixated areas, each of these maps is alpha-blended with the original painting.

Overall results: The first analysis concerns the number of fixations and the lengths of fixations for both tasks. On average, participants did 22.8 ± 0.2 fixations with an average length of 204 ± 3 ms in the complexity task and 22.6 ± 0.2 fixations with an average length of 209 ± 3 ms in the aesthetics task. Accordingly, a post-hoc ANOVA comparing the two experiments found no effect of task on either number or length of fixations ($F(1,18)=0.01/0.2$; $p=0.78/0.64$ respectively). We did find, however, a strong effect of art period on both measures ($F(10,180)=10.6/13.3$; $p<0.001$, resp.)—using the post-hoc Scheffe criterion we found that Postmodernist and Classicist images received less and at the same time longer lasting fixations than, for example, the Surrealist paintings (see Figure 4). Fixation lengths and number of fixations were highly anti-correlated with $r = -0.86$.

Repeating the analysis only for the *first* fixation, we found again no difference between the complexity and the aesthetics task. Additionally, the ANOVA showed that the first fixations on Postmodernist paintings were significantly longer than for all other periods. We thus found no effect of task (or top-down) information on either number or length of fixations. Instead, we found a highly significant (more bottom-up) effect of image content: the first fixation on abstract paintings took longer. In our view, this can have one of two reasons: either participants try to interpret the image content during this time [WKK*07], or abstract images in general do not afford so many fixations due to their limited complexity.

The two left panels of Figure 5 show the averaged fixation maps for the complexity and the aesthetics task. First of all, it becomes evident that participants tend to fixate the center quite often—this is a well-known finding (see, e.g.,

[PLN02]) and reflects both the typical bias of image content towards the center as well as general search patterns in human eye movements (the pattern still looks the same if we do not include the first fixation after the fixation cross, which might have biased participants). Furthermore, participants tend to scan images preferably in the vertical dimension—a closer look reveals that this is mostly due to the number of portraits that are included in the database where participants first try to localize the face and the facial features with their initial fixations. Finally, we examined the correlation between the two fixation maps: taking the overall data, we find almost perfectly correlation of $r = 0.98$ showing that the fixation patterns are highly similar across the two tasks.

When looking at the overall *spatial distribution*, however, we do find differences between the tasks: by summing all normalized fixations maps for all images, we can estimate the total area that is covered by the fixations. Even though the spatial distribution looks very similar across both tasks (compare the two left panels in Figure 5), the area that is covered by $\approx 91\%$ of all fixations is 17.2% in the complexity task, and 19.4% in the aesthetics task. Even though significant, this difference is small and testifies to the fact that both tasks seem to require global search strategies.

In order to investigate the build-up of the search pattern, we repeated the fixation analysis for a time window of ≈ 2 s. By counting the area covered by the fixations, we found that fixations covered only 10.1% —a significantly smaller proportion of the image. This increase in coverage over time thus conforms well to the two-stage model by [Loc06, LKMTN07].

Art period / image analysis: As shown in Figure 5, the fixation maps exhibit a strong center bias with relatively little coverage of larger areas. By averaging fixation data over the 11 art periods, we found that different periods show different fixation patterns: the most narrow distribution was obtained for the Postmodern period followed by the Rokoko and Classicism periods, whereas the broadest distributions were obtained for the Gothic and Expressionist periods. These patterns clearly reflect the different motifs used in the different periods: Rokoko and Classicism have the highest proportion of portraits in the database (11 resp. 10 out of 25 paintings), whereas especially the Expressionist period contains many densely painted scenes covering the whole canvas.

Different fixation strategies depending on image content are shown in Figure 6. First of all, if images contained faces (whether human or animal, portrait or several figures), these drew attention immediately and repeatedly (see Figure 6a) showing the high degree of salience for our observers. A very interesting tendency in the eye movements concerns the 'desire to interpret': the painting by Magritte in Figure 6b contains a broken window in which part of the scene visible through the 'window' has broken off in the shards. This Surrealist twist very reliably drew attention of our participants as they were interpreting the image. Similarly, in Figure 6c,

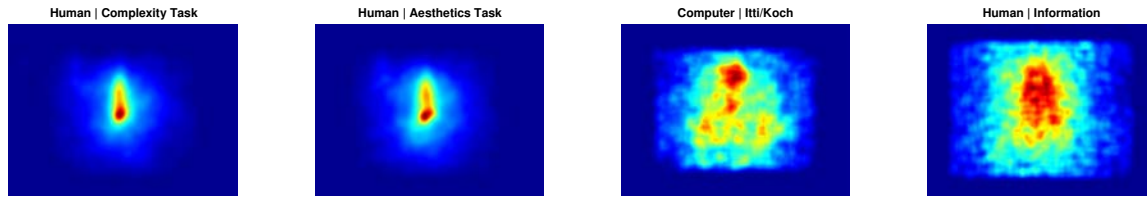


Figure 5: Average fixation maps for the complexity and aesthetics task and the two computational models.

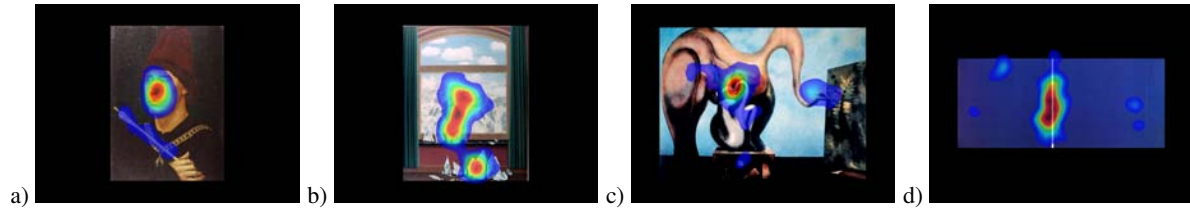


Figure 6: Examples of fixation maps for four different images.

which depicts a fantasy construct 'painting' a scene, fixations immediately focus on the 'hand' as the critical feature for making the interpretation of the figure clear to an observer. Finally, Figure 6d shows the fixations for a fully abstract image: fixations immediately move towards and then follow the white vertical line as the single most non-uniform feature. This tendency to follow suggested axes of objects can also be seen in Figure 6a with the quill pen that is held by the person.

Computational Analysis The two right panels of Figure 5 show the predicted fixation pattern using the Itti/Koch model (IK) and using the saliency maps based on color mutual information (CMI). The correlation of these overall measures with the human data is $r_{IK-Comp}=0.78$, $r_{IK-Aest}=0.81$ and $r_{CMI-Comp}=0.81$, $r_{CMI-Aest}=0.83$, respectively. The IK-map shows a much more point-focused overall distribution, whereas the CMI-map has a more (realistic) centered distribution. Nevertheless, compared to the human-human correlation of $r_{Aest-Comp}=0.98$, the correlations are much lower and clearly do not capture the fixation pattern well enough.

In addition, we found large variations in correlations across art periods as Figure 7 shows for the two models: both models have a clear peak for the earlier art periods of Rokoko and Classicism, with the CMI model also performing well in the Baroque period. This finding can be explained again by the large number of portraits in these periods. All portraits usually contain a bright face set against a darker background - this low-level saliency is well captured by both models with a slight, but significant advantage for the CMI model which takes the spatial distribution of color information into account.

Finally, Figure 8 shows the fixation predictions of the IK and CMI models for two images of Figure 6. For the first image, the IK model predicts the eye region as well as the two end points of the quill pen as salient fixation locations (Fig-

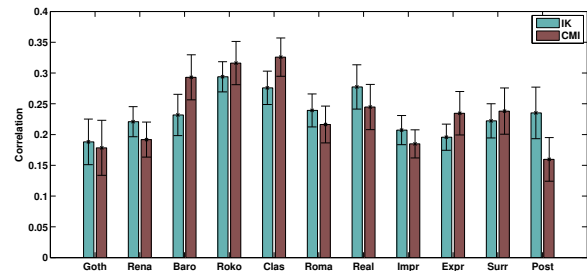


Figure 7: Correlations of the IK and CMI models with human data across art periods.

ure 8a), whereas the CMI model selects the whole face, as well as the whole quill pen as salient (Figure 8b). The CMI model accordingly results in a much better fit to the human data: a) $r_{IK-Aest}=0.45$, b) $r_{CMI-Aest}=0.62$. Conversely, the IK model predicts for the abstract painting a highly localized fixation pattern along the white line which corresponds to a very good fit with the human data, whereas the broader distribution of fixation patterns predicted by the CMI model selects too many candidate regions: c) $r_{IK-Aest}=0.83$, d) $r_{CMI-Aest}=0.39$.

6. Conclusions

In this study, we evaluated participants' aesthetic experience using ratings of complexity and aesthetic appeal as well as eye tracking analyses. We found clear preferences for art periods and single paintings both in terms of complexity as well as in terms of aesthetic appeal. Surprisingly, participants overall clearly rated classical and realistic paintings as much more aesthetic than more modern art stemming from the 20th century. In their evaluation, they used the whole range of image features from low-level (distribution of color) to higher-level (global impression) for both tasks with a tendency towards more global features for the aesthetics task.

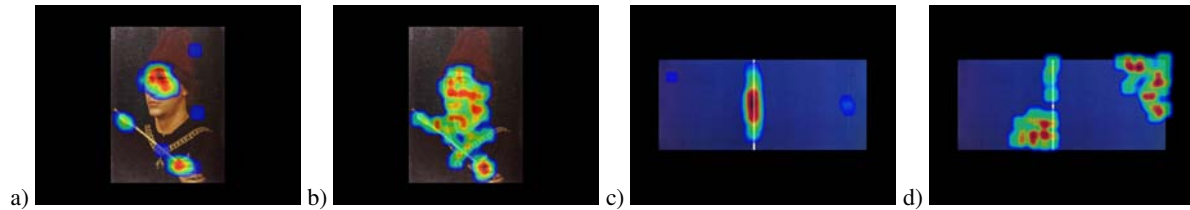


Figure 8: Examples of fixation maps for two of the images in Figure 6 with predictions of the a,c) IK model, b,d) CMI model.

As one might have expected, computational features are still not able to fully explain aesthetics ratings: a correlation with several ratings derived from information theoretic considerations produced significant results for the complexity task but not for the aesthetics task. The dependency of these correlations on the art period provides confirmation of the questionnaire analysis in that participants used *different* image features for their ratings ranging from the color palette (*H*) to the degree of organization (*E*).

Analysis of the eye movements has revealed only small differences between the two tasks: both the complexity and the aesthetics task favored a more global search strategy in which participants scanned the whole image. We were also able to replicate the general search pattern identified by [Loc06, LKMTN07] of a coarse, global scan path followed by more small-scale local fixations to some degree. Two computational models of saliency were correlated with the human data producing a reasonable degree of fit with a slight advantage of the information-theory based model over the standard Itti/Koch model. As with the rating correlations, however, there is much room for improvement in terms of explaining all degrees of freedom underlying human fixation patterns. Overall, we can conclude that low-level saliency measures based on the 'simple method of pixel counting' were nevertheless surprisingly effective in capturing part of the human aesthetic experience.

References

- [AL06] AUGUSTIN D., LEDER H.: Art expertise: a study of concepts and conceptual spaces. *Psychology Science* 48 (Jan 2006), 135–156.
- [ALHC08] AUGUSTIN M., LEDER H., HUTZLER F., CARBON C.: Style follows content: On the microgenesis of art perception. *Acta Psychologica* 128 (Jan 2008), 127–138.
- [Bra97] BRAINARD D. H.: The Psychophysics Toolbox. *Spatial Vision* 10, 4 (1997), 433–436.
- [Bus35] BUSWELL G.: *How people look at pictures: A study of the psychology of perception in art*. Univ. Chicago Press, Chicago, 1935.
- [FC88] FUNDER D. C., COLVIN C. R.: Friends and strangers: acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of personality and social psychology* 55, 1 (Jul 1988), 149–58.
- [FC03] FELDMAN D., CRUTCHFIELD J.: Structural information in two-dimensional patterns: Entropy convergence and excess entropy. *Physical Review E* 67 (2003).
- [Git02] GITELMAN D. R.: ILAB: a program for postexperimental eye movement analysis. *Behavioral Research Methods Instruments Computers* 34, 4 (Nov 2002), 605–612.
- [IK01] ITTI L., KOCH C.: Computational modelling of visual attention. *Nature Reviews Neuroscience* 2, 3 (March 2001), 194–203.
- [KRH*00] KRIEGER G., RENTSCHLER I., HAUSKE G., SCHILL K., ZETZSCHE C.: Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics. *Spatial Vision* (2000).
- [LBOA04] LEDER H., BELKE B., OEBERST A., AUGUSTIN D.: A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology* 95, Pt 4 (Nov 2004), 489–508.
- [LKMTN07] LOCHER P., KRUPINSKI E. A., MELLO-THOMS C., NODINE C. F.: Visual interest in pictorial art during an aesthetic experience. *Spatial Vision* 21, 1-2 (Jan 2007), 55–77.
- [LMR99] LAND M., MENNIE N., RUSTED J.: The roles of vision and eye movements in the control of activities of daily living. *Perception* 28, 11 (1999), 1311–1328.
- [Loc06] LOCHER P. J.: The usefulness of eye movement recordings to subject an aesthetic episode with visual art to empirical scrutiny. *Psychology Science* 48 (2006).
- [LV97] LI M., VITÁNYI P. M. B.: *An Introduction to Kolmogorov Complexity and Its Applications*. Graduate Texts in Computer Science. Springer-Verlag, 1997.
- [Mol81] MOLNAR F.: *Advances in Intrinsic Motivation and Aesthetics*. Plenum Press, New York, 1981, ch. About the role of visual exploration in aesthetics, pp. 385–413.
- [NLK93] NODINE C., LOCHER P., KRUPINSKI E.: The role of formal art training on the perception and aesthetic judgment of art compositions. *Leonardo* 26 (1993), 219–227.
- [PLN02] PARKHURST D., LAW K., NIEBUR E.: Modeling the role of saliency in the allocation of overt visual attention. *Vision Research* 42, 1 (Jan 2002), 107–123.
- [RFS08] RIGAU J., FEIXAS M., SBERT M.: Informational aesthetics measures. *Computer Graphics and Applications* 28, 2 (March–April 2008), 24–34.
- [WCF08] WALLRAVEN C., CUNNINGHAM D. W., FLEMING R.: Perceptual and computational categories in art. In *International Symposium on Computational Aesthetics in Graphics, Visualization, and Imaging* (Aire-la-Ville, Switzerland, 06 2008), Brown P., (Ed.), Eurographics Association, p. 1.
- [WK06] WALTHER D., KOCH C.: Modeling attention to salient proto-objects. *Neural Networks* 19, 9 (Jan 2006), 1395–1407.
- [WKK*07] WALLRAVEN C., KAULARD K., KÜRNER C., PEPPERELL R., BÜLTHOFF H. H.: Psychophysics for perception of (in)determinate art. In *APGV 2007* (New York, NY, USA, 07 2007), ACM Press, pp. 115–122.
- [Yar67] YARBUS A.: *Eye movements and vision*. Plenum Press, New York, 1967.