Visual Saliency Analysis in Paintings

Bogdan Stoica*, Laura Florea†, Alexandra Bădeanu*, Andrei Racoviţeanu* and Iulian Felea* and Corneliu Florea‡
Image Processing and Analysis Laboratory, University "Politehnica" of Bucharest, Romania
‡ corneliu.florea@upb.ro ‡ laura.florea@upb.ro

***Abstract*— Visual saliency is the property of a scene to draw the attention of different people to the same objects within it. Its investigation and development enabled many practical applications. In this paper we aim to use it for the study of visual art. As a first contribution we introduce an extensive database annotated in terms of visual saliency by considering data acquired from many users. Secondly, based on visual saliency regions, we identify some core differences between various artistic movements represented within. Thirdly, we evaluate comparatively several state-of-the-art saliency prediction algorithms, so to identify the most performing in the context of digitized paintings.**

I. Introduction

Recent developments in image recording, storage and analysis favor an eased access to works of art for general public in parallel to collecting a large corpus over the web. While in art domain it is often said that art is for humans only to understand or as Franklin Lloyd Wright said "Art is a discovery and development of elementary principles of nature into beautiful forms suitable for human use", in computer science there is a continuous effort to create autonomous systems that understand art concepts. In this paper we address the problem from a less used perspective and that is of visual saliency.

The motivation of the current paper lies in the findings of Zeki [1] and respectively Ramachandran and Herstein [2]. They, by independent psychological experiments, found that the key for understanding art is the identification of the perceptual process, rather than the analysis of the aesthetic properties. Our approach is by correlating the visual saliency regions with artistic movement (also named style) of a specific artwork.

A. Related Work

Visual saliency. Given a scene, the majority of the human's attention will be drawn by the same areas which are called saliency regions [3], [4]. More precisely human visual fixation behavior is driven by sensorial bottom-up mechanisms; visual saliency refers to the physical, bottom-up distinctness of image details [5]. In other words, the saliency is the distinct subjective perceptual quality of some items in a scene to stand out from their neighbors and to attract our attention [3].

Following the seminal paper of Itti and Koch [3], many visual saliency methods have been introduced. Various state of the art solutions are thoroughly discussed in the review by Toet [4]. Typically, ground truth saliency data is retrieved by recoding a viewer's gaze by a professional eye tracking device. The spatial geometry of the recording area leads to the retrieval of the location of the saliency maps in images [6]. Next, given the saliency ground truth data, a plethora of methods were proposed to automatically estimate these regions. A summary could start with the now-classical model of Itti and Koch [3] and finish with the very recent models that investigate the uses of perceptual color description [7], or the reverse hierarchical descriptions as they seem to mimic better the gaze behavior [8]. Concluding, the automatic saliency region maps methods are mature and with robust results.

A notable solution is the Graph-Based Saliency [9] which is based on identification of features which extract "unusual" behavior from a neighborhood to model the center-surround property of the human visual system and aggregates in a graph that models the bottom-up visual saliency. The salient regions are inherently sparse and this may be captured as the inverse of the sign of the DCT transforms over image blocks [10]. Alternatively, one may retain the amplitude of the Fourier transform as the origin map of salient regions [11]. A more recent solution integrates and optimizes the quaternion DCT and FFT for improved accuracy [12].

Saliency application in art. Among applications, it has been showed that the saliency map distribution leads to a reliable aesthetic measure of natural images [13]; this idea was further developed into a complexity measure of the artistic painting by Cacciola et al. [14]. Specifically, for paintings it was showed [15] that there exists a deep connection between the saliency map (i.e. gaze fixations), and the emotion generated by that artistic work.

Regarding the process of gaze fixation on artistic paintings, Locher et al. [16] showed that art viewers (for both representational or abstract types of art) in the first two seconds do a full sweep of the image, analyzing its "gist" (i.e. understanding everything at once, regardless of the visual complexity of the scene [17]). Only after this first explorative stage, viewers tend to focus on the more subtle aspects. Moreover, it has been shown that during this second period, bottom-up saliency plays an important role in the allocation of eye movements [18].

Summarizing, gaze pattern and aspects related to user behavior when viewing art were intensively studied. Yet, the use of saliency for the description of the painting content was overlooked.

Contribution. Given the direct relation between saliency map distribution and complexity [14], we anticipated that connections between the gaze trajectory and artistic movement can be determined and can form a sustainable element of analysis for painting.

In this direction, we have used a professional gaze tracker, to record multiple users, and so to build the ground truth saliency maps (as recorded in the second fixation stage, after the gist period) for a digitized painting collection.

Given the saliency reference, we evaluated several state of the art saliency methods to determine the best eye fixations predictor for the case of paintings. We perform an initial test that shows a connection between the saliency maps and the paintings art movement.

The remainder of the paper is organized as follows: Section II describes the protocol and the structure of the database; section IV describes various experiments performed directly on the saliency data. In section III we describe the performance achieved with state of the art solution for saliency prediction so that the paper will end with discussion and conclusions.

II. Database

In this section we present our efforts in gathering visual saliency data for paintings. The images were selected from the Pandora18k database [19] so to span all 18 movement illustrated. In order to express the relationship between the two databases, we name the current visual saliency dataset Pandora-saliency.

The protocol of the experiment is inspired from the work of Judd et al. [20]. The selection was made by an art expert so to use representative paintings for each movement. Also we ensure that the digitized versions of the paintings are qualitative in the sense that they have sufficient resolution and they do not suffer from JPEG compression artifacts. From 24 to 40 people have looked at each painting (as detailed in table I).

For the actual recording we tried Tobii Pro X2-30 and a Gazepoint GP3; initial tests indicated superior performance of the first one. The system is based a dark pupil eye tracking system that uses infrared illumination and computer-based image processing to register movements of the eyes in realtime. We used a 9 point calibration system. In our experiment, users were presented with each image for 9 seconds and before presenting each image the observer fixates on a cross in the center of the gray screen for half a second.

The database contains 30 images from romanian artists that although can be included in some movement, for the moment are left outside. It is worth saying that people that looked at the images are non art experts, being at most peoples with affinity with the topic. Very small scale experiment performed with art experts lead to consistently different result. The explanation is probably related to the fact that art experts have previously looked at the paintings and now are searching for detailed aspects.

Figure 1 shows some examples of the results obtained using the eye tracker. The fixation data will be available on the project page[1].

Comparison with other saliency databases. A more detailed comparison of the proposed database with other similar attempts may be followed in table II. It should be noted that only Paintings 91 [21] reports saliency data in conjunction with art. It should also be noted the highest number of users that looked at an image is achieved by our database. Initially we emphasized for the larger number as viewed images are masterpieces and thus assumed to be complex scenes that may produce confusing results if only few people looked at it. Furthermore, as discussed in section IV, the large number of user allows us to study some hypothesis about correlation between human gaze behavior and artistic movement.

[1]The database and fixation data will be available at `http://imag.pub.ro/pandora/pandora_download.html`

TABLE I

Structure of the Pandora-saliency dataset with respect to subjects viewing the paintings and artistic movement distribution. Each set contains 45 images.

Set	Viewers
1	40
2	37
3	34
4	33
5	24

Movement	Paints
Byzantine	4
Early Ren.	6
North Ren.	5
High Ren.	7
Baroque	9
Roccoco	11
Romant.	8
Realism	10
Impress	8
Post-Imp.	30
Fauvism	30
Express.	8
Symbol.	8
Cubism	30
Surreal.	30
Naive-art	30
Pop-art	6
Romanian	30

TABLE II

Comparison with other visual saliency databases.

Name	No. Images	Subjects	Image nature
FIFA [22]	180	8	Natural color images
Toronto [18]	120	11	In/out color images
MIT [20]	1003	15	Natural color images
NUSEF [23]	758	25	Natural color images
Paintings91 [21]	182	10	Paintings
PandoraSaliency	270	40	Paintings

III. Predicting Visual Saliency

To evaluate the performance of the saliency algorithms we use the area under the Pearson Correlation Coefficient between the ground truth, obtained as the average of the maps for all users for one image, and subject prediction. We recall that the coefficient, CORR, is computed as:

$$CORR = \rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (1)$$

where x_i are the samples from the ground truth image (obtained by averaging the images from all subjects on a given painting), while y_i is the subject saliency map, obtained either from a specific user, or from a saliency prediction method.

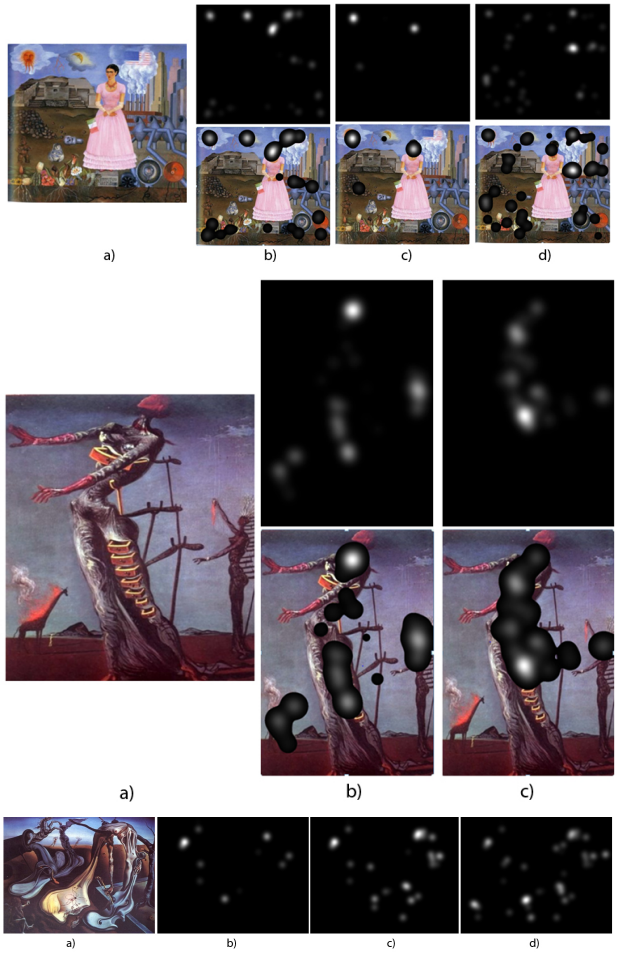

Fig. 1. a) Paintings from PANDORA database, b) Saliency map after 4.5 seconds, c) Saliency map after 9 seconds , d) Final saliency map.

Compared to other used measures such as ROC [20] or Normalized Saliency Scanpath - NSS, the CORR measure is more robust and assumes a more thorough match [24].

We have tested the following methods: Graph-Based Saliency (GBS–[9]), signature based (SIGN–[10]), the FFT based ([11]) also adjusted for DCT and the quaternion (QUAT–[12]). The obtained results may be followed in table III. As a baseline, we report the average subject, best subject and worst subject.

If one analyzes the results reported in table III, he will note the best performing saliency prediction methods reach higher values than the worst case for a human subject. Under this circumstance, we claim that results obtained by annotating a much larger corpus of digitized paintings by automatic methods is a feasible direction that may lead to interesting results.

IV. Saliency Data Experiments

In this paper we will process the saliency map holistically, without breaking into spatial regions. Thus we have not investigated various differences in trajectory which may had lead to the same final result.

Compared to other visual saliency datasets, we have recorded the gaze trajectory in two steps: one after 4.5 seconds and one after 9 seconds from the beginning. The second includes the trajectory of the first. The reason for this detailing is be able to perform a more thorough analysis of regions that captures subject attention. Regarding the motivation, let us recall the intuitive observation of Cacciola et al. [14] that the more complex a scene is, the more diversity will exist among various subjects while the recoding of the saliency map.

For each image we have computed the correlation between each subject and the average over all subjects. Given the previous emphasized conclusions [14], one will expect that movement that aim to more complex scene and composition to exhibit higher variability. To study this aspect we have plotted the variance of the correlation of the saliency maps obtained from individual subjects with respect to the max-min difference. While for some movements we have insufficient data to draw any conclusion, yet, in figure 2, one may see a clustering between movements. For instance for expressionism or cubism movements, due to the abstraction, it is hard to get in one gist the meaning of the painting. On the other hand, some movements which although have a richness of details, project a straight forward interpretation of the theme/scene. This initial test shows that indeed the saliency correlation can give information about the type of art movement of the painting and is worth to be further investigated.

V. Discussion and Conclusions

In this paper we have described our efforts to investigate the problem of saliency in digital paintings. We have acquired eye fixation data for a subset of Pandora18k dataset, by recoding data from a number of subjects significantly larger than in other cases. We have evaluated several state of-the-art saliency prediction solutions for estimating regions of fixation in art images. Our results show that some of these standard saliency prediction algorithms provide performance for saliency estimation comparable to average human subject while looking at art images.

The high number of subjects used allowed us to study the possible existence of correlations between how people look at artistic paintings and the movement from which it is assigned. Indeed, the results showed that those movements which are characterized by a higher level of abstraction lead to more diversity in areas fixated by user.

As further extension we aim to expand the study by incorporating the romanian paintings, annotate with predicted saliency a larger corpus of the database and to determine if further aspects may be revealed.

Acknowledgment

This work is supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS – UEFISCDI, number PN-II-RU-TE-2014-4-0733. The authors wish to express their gratitude to Iuliana Dumitru for her help in selecting the most interesting and representative works art to be viewed.

TABLE III

COMPARISON OF DIFFERENT SALIENCY PREDICTION APPROACHES ON THE PAINTINGS. NOTE THAT MOST OF THE SALIENCY PREDICTION ALGORITHMS PROVIDE IMPROVED PERFORMANCE COMPARED TO BASELINE. FOR SUBJECTS WE CONSIDER ONLY THOSE WHO LOOKED TO ALL IMAGES.

Method	**Prediction**					**Subjects**		
	GBS–[9]	SIGN–[10]	FFT–[11]	DCT–[11]	QUAT–[12]	Average	Worst	Best
CORR, ρ	0.454	0.373	0.112	0.142	0.385	0.525	0.416	0.608

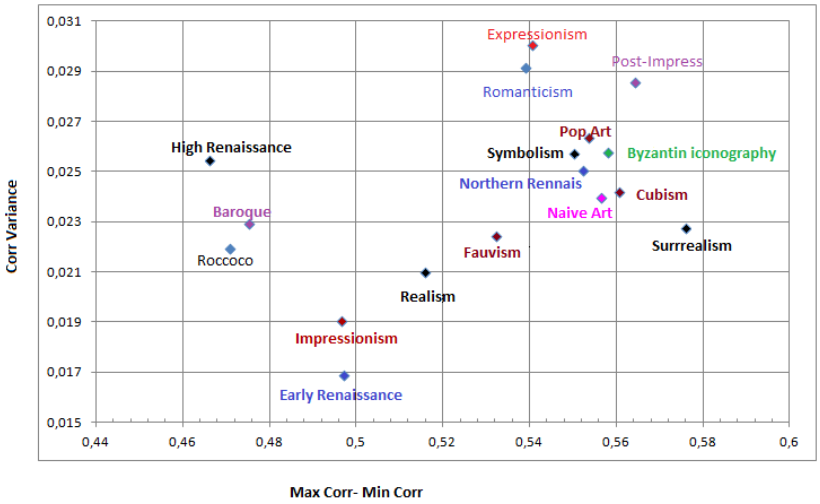

Fig. 2. Max-min difference with respect to in-movement variance of the correlation coefficent for illustrated movements. Note the two main clusters.

REFERENCES

[1] S. Zeki, *Inner Vision: an exploration of art and the brain*. Oxford University Press, 1999.

[2] V. S. Ramachandran and W. Hirstein, "The science of art: A neurological theory of aesthetic experience," *Journal of Consciousness Studies*, vol. 6, no. 6-7, pp. 15–51, 1999.

[3] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 20, no. 11, pp. 1254–1259, 1998.

[4] A. Toet, "Computational versus psychophysical bottom-up image saliency: A comparative evaluation study," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 33, no. 11, pp. 2131–2146, 2011.

[5] J. H. Fecteau and D. P. Munoz, "Salience, relevance, and firing: a priority map for target selection," *Trends in cognitive Sciences*, vol. 10, no. 8, pp. 382–390, 2006.

[6] T. Judd, F. Durand, and A. Torralba, "Fixations on low-resolution images," *Journal of Vision*, vol. 11, no. 4, pp. 14–14, 2011.

[7] J. Kim, D. Han, Y.-W. Tai, and J. Kim, "Salient region detection via high-dimensional color transform," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 883–890.

[8] T. Shi, M. Liang, and X. Hu, "A reverse hierarchy model for predicting eye fixations," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 2822–2829.

[9] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proceedings of the 19th International Conference on Neural Information Processing Systems (NIPS)*, 2006, pp. 545–552.

[10] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 34, no. 1, pp. 194–201, 2012.

[11] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 1597–1604.

[12] B. Schauerte and R. Stiefelhagen, "Quaternion-based spectral saliency detection for eye fixation prediction," in *European Conference on Computer Vision (ECCV)*, 2012, pp. 116–129.

[13] L.-K. Wong and K.-L. Low, "Saliency-enhanced image aesthetics class prediction," in *16th IEEE International Conference on Image Processing (ICIP)*, 2009, pp. 997–1000.

[14] M. Cacciola, G. Occhiuto, and M. F, "Artistic complexity and saliency: Two faces of the same coin?" *International Journal of Information Acquisition*, vol. 9, no. 2, pp. 1 350 010 – 10, 2013.

[15] V. Yanulevskaya, J. Uijlings, E. Bruni, A. Sartori, E. Zamboni, F. Bacci, D. Melcher, and N. Sebe, "In the eye of the beholder: employing statistical analysis and eye tracking for analyzing abstract paintings," in *Proceedings of the 20th ACM International Conference on Multimedia*. ACM, 2012, pp. 349–358.

[16] P. Locher, E. A. Krupinski, C. Mello-Thoms, and C. F. Nodine, "Visual interest in pictorial art during an aesthetic experience," *Spatial Vision*, vol. 21, no. 1, pp. 55–77, 2007.

[17] A. Oliva, *Neurobiology of Attention*. Elsevier, 2005, ch. Gist of the scene, pp. 251–256.

[18] N. Bruce and J. Tsotsos, "Saliency, attention, and visual search: An information theoretic approach," *Journal of Vision*, vol. 9, no. 3, pp. 5–5, 2009.

[19] C. Florea, R. Condorovici, C. Vertan, R. Butnaru, L. Florea, and R. Vrânceanu, "Pandora: Description of a painting database for art movement recognition with baselines and perspectives," in *24th European Signal Processing Conference (EUSIPCO)*. IEEE, 2016, pp. 918–922.

[20] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *IEEE International Conference on Computer Vision (ICCV)*, 2009, pp. 2106–2113.

[21] F. S. Khan, S. Beigpour, J. van de Weijer, and M. Felsberg, "Painting-91: a large scale database for computational painting categorization," *Machine Vision and Applications*, vol. 25(6), pp. 1385–1397, 2014.

[22] M. Cerf, E. Frady, and K. C., "Faces and text attract gaze independent of the task: Experimental data and computer model," *Journal of Vision*, vol. 12, no. 10, p. 115, 2009.

[23] R. Subramanian, H. Katti, N. Sebe, M. Kankanhalli, and T. S. Chua, "An eye fixation database for saliency detection in images," in *European Conference on Computer Vision (ECCV)*, vol. 6314, 2010, pp. 30–43.

[24] L. Elazary and L. Itti, "A bayesian model for efficient visual search and recognition," *Vision Research*, vol. 50, no. 14, pp. 1338–1352, 2010.