SALIENCY MAP RETRIEVAL FOR ARTISTIC PAINTINGS INSPIRED FROM HUMAN UNDERSTANDING

Răzvan George Condorovici, Ruxandra Vrânceanu, Constantin Vertan

The Image Processing and Analysis Laboratory, "Politehnica" University Bucharest
Email: {rcondorovici, rvranceanu, cvertan}@alpha.imag.pub.ro

ABSTRACT

This paper presents a simple and efficient method for detecting salient regions in digital representations of paintings. The main challenge is to model the way human eye and mind see and understand visual art. Based on a combination of features such as shape, colour, local contrast and position, the most relevant areas of a digital representation of a painting are detected. The model follows thoroughly the human interpretation of the artistic painting. The presented approach shows robustness regardless of the art movement the analyzed painting belongs to.

Keywords: *Image Segmentation, Image Classification, Feature Extraction, Computer Aided Analysis.*

1. INTRODUCTION

The rapidly grown usage of computers in daily life allows intelligent systems to enter an area not long ago considered reserved just for human beings. The art world began to use computers for tasks such as classification, archiving or analysis of paintings. One of the biggest challenges in this field is to make computers see art the same way humans do.

Previous attempts describe some of the steps required by an automatic system to analyze and understand such paintings. Thus, Gunsel et.al. [1] worked on painting classification based on the art movement. Aesthetic quality evaluation based on the art movement has been proposed by Li et al [2], emotion recognition in paintings has been dealt in [3], while non-photorealistic image rendering reproducing a certain art style is the topic of [4].

One of the most provoking aspects to overcome in order to accomplish the previous mentioned tasks or any other similar one is to semantically interpret the image. Unlike a machine, which uniformly analyzes the whole image, the human brain detects some regions of interest that are to be more thoroughly analyzed onward [5]. During the quick scan that leads to detection of the areas of interest, the human brain performs two processes: pattern recognition, meaning the labelling of an object representation based on a previous experience and the detection of atypical areas, meaning the extraction of areas with significantly different features.

This paper describes a computational approach to model the last of the two previous described processes.

These processes are well known to artists, a wide kind of techniques being used to guide the viewers' attention to some specific areas of the painting. This is the main reason for which it is strongly desirable to detect and exploit these intrinsic features of a painting during the analysis process.

1.1 Motivations

The usage of saliency information is of real help when it comes to image segmentation, where the main two issues are oversegmentation and undersegmentation. These issues can be efficiently addressed if during the segmentation process the area's saliency is taking into consideration.

Considering the fact that artists do use various techniques to increase the saliency of some areas in paintings, the feature extraction process can be improved by using saliency information.

The speed of a non-photorealistic image rendering process can be increased if saliency information is used. This way, a greater computation effort can be spent in relevant areas, while a simpler and faster processing can be apply in less perceivable regions.

1.2 Related Work

Our attempt is not unique in the literature. DeCarlo and Santella [6] present a saliency based segmentation method. The authors propose an eye tracking based solution that offers remarkable results but has the drawback of needing human interaction.

Itti, Koch and Niebur [7] use a dynamic neural network and a set of multiscale image features to create a saliency map for a natural image.

The same authors use a set of multiscale features [8] to detect and combine spatial discontinuities in intensity, colour and orientation into a saliency map.

Hence different alternatives exist for extracting the saliency map from a painting. A review and a comparison of some of the known methods may be followed in [9].

1.3 Outline of the Proposed Algorithm

In this paper an automatic method for detecting regions of interest in a digital representation of a painting has been developed. The proposed algorithm models the processes occurring in the human brain when a scene is evaluated for finding regions of interest. The input image is assumed to be represented in the RGB colour space.

The first step of the algorithm is the input image's segmentation. The second step in detecting the regions of interest consists in extracting a set of features similar to those used by the human eye for each region of the image. These steps are described in section 2. In section 3 the methodology to combine the extracted features into a final saliency map is presented. The performance of this method is discussed in section 4. The paper ends with conclusions.

2. FEATURES EXTRACTION

2.1 Image Segmentation

As stated in section 1, all the saliency relevant features are extracted for regions of the image which are resulted following a segmentation process. Considering that the segmentation accuracy is not extremely relevant for this purpose, any segmentation method will be adequate. In our work a Fuzzy C-Means thresholding algorithm was used [10]. A number of 10 clusters proved to be enough for any kind of input paintings. The initial clusters are randomly initialized and 100 iterations are executed. Regions belonging to each of the ten clusters are labelled, resulting in N areas.

From the N regions we discard the ones having less than $N_{min} = 10$ pixels. The pixels belonging to these regions are labelled the same as the largest neighbouring region. This way, the smaller regions that can be caused by a segmentation error and that certainly do not consist in a region of interest are discarded.

2.2 Compacity Degree

Psychological studies revealed that the human brain is trained to quickly separate the background and the foreground of an image and to pay more attention to the latter [5]. At the same time, while detecting the regions of interest in a picture, the human eye spends more time in compact areas. This, corroborated with the fact that after the segmentation process the background areas are found in wider spread regions, lead to the need of a region compacity measure.

The compacity measure, C is computed for each *i* region as the ratio between the squared perimeter's cardinality, $|\mathrm{P}|$, and the region's cardinality, $|\mathrm{R}|$, multiplied by 4π:

$$C_i = \mathrm{T} - \frac{|P_i|^2}{4\pi|R_i|} \tag{1}$$

The perimeter is computed as the external morphological gradient.

In (1) T is a threshold adequately chosen so that a negative compacity value reflects a non compact region while a positive compacity value corresponds to a compact region. The compacity based saliency map is presented in Fig. 1c.

The initial domain of definition is theoretically infinite. Empirically it was determined that an optimal value for T in (1) is 10, this value allowing to declare as non-compact all regions that have a negative compacity degree. Furthermore, most of the compacity values have been found to belong to the [-15, 15] interval. The values can therefore be safely normalized to [-1, 1]:

$$C_i^N = \max\left(-1, min\left(1, \frac{C_i}{15}\right)\right) \tag{2}$$

2.3 Local Contrast

Another important feature used by the human brain to detect regions of interest is the local contrast. A point of spatial discontinuity in colour representation is a strong attraction point for the human eye.

To model this behaviour, a local contrast measure, LC, was developed. We compute for each region *i* a measure that reflects the color discontinuity with respect to the colors of the each *j* neighbouring regions:

$$\mathrm{LC_i} = \frac{\sum_j\left(|\mathrm{R_i}|\sum_{\mathrm{p}\in\{\mathrm{R,G,B}\}}\left(\mathrm{k_{i,p}} - \mathrm{k_{j,p}}\right)\right)}{\sum_j|\mathrm{R_j}|} \tag{3}$$

, where $\mathrm{k_{m,n}}$ is the value of the p-th plane for the m-th region. The local contrast is computed as the sum of contrasts on each of the R, G and B planes. As stated in (3), a neighbouring region's contribution to the local contrast is directly proportional with the region's size. A representation of the saliency map based on local contrast can be seen in Fig. 1d.

The local contrast values are also normalized to [0, 1] through a linear transformation:

$$LC_i^N = \frac{LC_i - LC_{\min}}{LC_{max} - LC_{min}} \tag{4}$$

, where $\mathrm{LC_{min}}$ and $\mathrm{LC_{max}}$ are the minimum and maximum values of the local contrast map.

2.4 Edge influence

One of the first clues the human eye searches for when analyzing a scene are the edges. As stated in [5], the human brain is almost fully capable of understanding a scene based only on edges. The human attention for edges is perfectly normal considering that edges represent objects' boundary, translated in changes in the surrounding world, changes to which one must react.

The model for this human behaviour is obtained by analyzing the gradient absolute value in the immediate proximity of the region. To define the immediate proximity of the area, a map, M, consisting in the sum of external and internal morphological gradients, is computed. The morphological gradients are computed using a square structuring element, K, having the size of $0.025 \cdot \min(\mathrm{H},\mathrm{W})$, where H and W are the image's height and width:

$$\mathrm{M_i} = (\mathrm{R_i} \oplus \mathrm{K} - \mathrm{R_i}) + (\mathrm{R_i} - \mathrm{R_i} \ominus \mathrm{K}) \tag{5}$$

The edges' influence degree, EI, is computed as the decimal logarithm of the sum of the absolute value of the horizontal and vertical image gradients, Gx and Gy, corresponding to the previously defined map, divided by the map's cardinality:

$$\mathrm{EI_i} = \log\left(\frac{\sum_{\mathrm{p}\in\mathrm{M_i}}\left(\mathrm{abs}(\mathrm{Gx_p}) + \mathrm{abs}(\mathrm{Gy_p})\right)}{|\mathrm{M_i}|}\right) \tag{6}$$

Like the local contrast map, the edge influence map is also normalized to [0, 1]. An example of saliency map is presented in Fig. 1e.

2.5 Uniqueness Degree

Another important point of interest for the human eye in a given context represents the points of uniqueness. Translated

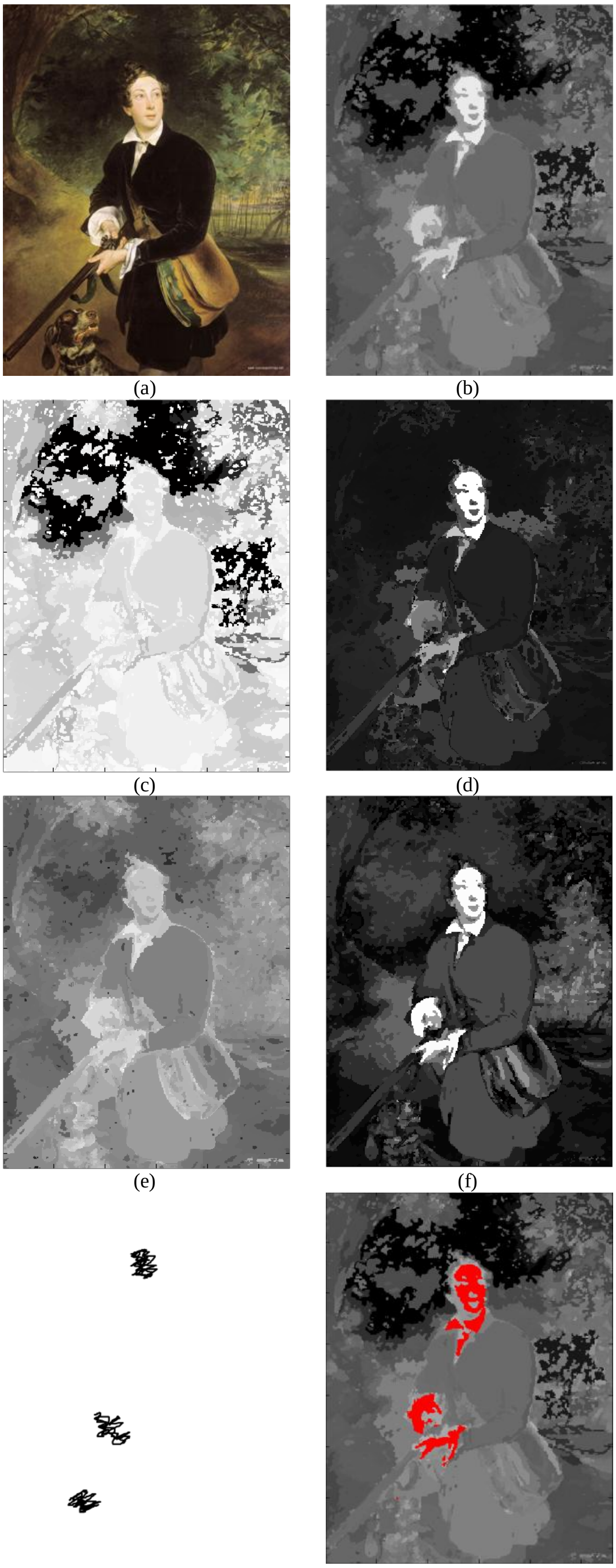

Figure 1: (a) Original Image: “Portrait Of The Poet and Playwright Aleksey Konstantinovich Tolstoy”, Alexander Briullov. (b) Final Saliency Map (c) Compacity Saliency Map. (d) Local Contrast Saliency Map. (e) Edge Influence Saliency Map. (f) Uniqueness Saliency Map. (g) User Defined Saliency Map. (h) Thresholded Saliency Map

into paintings world, the human eye is greatly attracted by a light region in an overall dark painting or by a dark region in an overall light painting.

To reflect this behaviour, a uniqueness measure, U, is derived as the sum of absolute differences between the current region values and the image’s mean values for each of the R, G, and B planes:

$$U_i = \sum_{p \in \{R,G,B\}} abs\left(k_{i,p} - \overline{k_p}\right) \tag{7}$$

, where $\overline{k_p}$ is the mean value for the p-th plane of the image. A representation of the saliency map based on uniqueness degree can be seen in Fig. 1f.

For the values of uniqueness degree features, the theoretical domain of definition is [0, 765] for the extreme case when the mean values of the images are 0 or 255 and the current region values are respectively 255 or 0. Considering that the mean values are apriori known, the domain of definition, D, can be further restrained to:

$$D = \left[0; 3 \cdot \max\left(255 - \overline{k}_{min}, \overline{k}_{max}\right)\right] \tag{8}$$

This last feature is also normalized to [0, 1]:

$$U_i^N = \frac{U_i}{\left(3 \cdot \max\left(255 - \overline{k}_{min}, \overline{k}_{max}\right)\right.} \tag{9}$$

3. FEATURES INTERPRETATION

The next step after extracting the saliency maps is to combine them into a single final map of the most representative areas of the painting.

Unfortunately it is of extreme difficulty to determine the weights the human brain applies for each of the four features when evaluating a scene.

Having all features in [0, 1] or [-1, 1] intervals, the next step is to combine them into a single saliency map. A naive and rigid approach would be to compute the final map as the sum of the feature maps. Although this solution offers acceptable results, a more flexible approach is to weight each feature independently in the addition process:

$$S_i = \alpha \cdot C_i^N + \beta \cdot LC_i^N + \gamma \cdot EI_i^N + \delta \cdot U_i^N \tag{10}$$

Using a training set, after an optimization process was determined that the best results are obtained with $\alpha = 0.15$, $\beta = 0.35$, $\gamma = 0.2$ and $\delta = 0.3$. For obtaining a list of the most relevant regions, the previously generated saliency map is thresholded. A normalized cumulative histogram of saliency values is created. Each region having a saliency value higher than 70% of the maximum value is declared as salient.

An example of final saliency map can be seen in Fig. 1b.

4. RESULTS

In order to test the described method, 6 sets of 10 paintings each were evaluated. Each set contained paintings from a certain art movement. The evaluated art movements were: Impressionism, Modernism, Postimpressionism, Realism, Romanticism and Renaissance.

Considering the fact that our solution models the human brain behaviour, the definition of some objective quality measures is practically impossible. To overcome this lack of

Artistic movement	Detection Rate	
	Proposed Solution	Harris Corner Detector Based Solution
Impressionism	64%	46%
Modernism	56%	42%
Postimpressionism	50%	48%
Realism	75%	55%
Romanticism	66%	49%
Renaissance	70%	53%
Overall	63,5%	48,8%

Table I – Detection rates

evaluation criteria a subjective testing methodology was developed [6].

Human subjects were asked to watch each painting for five seconds. After the five seconds are over, the subject is presented a blank surface, the same size as the original painting, on which he or she has to mark the areas considered to be the most relevant. Given the short period of analysis the subject is offered, it is likely that she or he did not have enough time to judge the painting from a semantic point of view.

The same digital representation of the painting is processed with the proposed solution. A supervised comparison is realized between the human marked map and the computer generated saliency map. Statistics can be seen in Table I. The areas marked as salient in both of the two maps are marked as true positives. The other areas are marked as False Positive or False Negative. In the example presented in Fig. 1g-h the true positive detection rate is 66%, the false positive rate 33% and the false negative rate is also 33%. Due to the fact that the human generated map is not accurate enough the comparison process between the two maps cannot be automated and is realized by a human being.

As can be seen in Table I, the detection rate varies from one artistic movement to another, but the overall results are satisfactory given the nature of the problem. A higher detection rate can be observed for Realist, Renaissance and Romantic paintings while for Impressionist, Postimpressionist and Modernist paintings the detection rate is decreased. This observation is sustained by art theory, being well known the fact that painters belonging to the first three art movements offered more clues for the viewer's eye.

For comparison purpose, a saliency map retrieval solution based on Harris corner detector was implemented [11]. The solution was tested on the same database. As can be seen in Table I, our solution offered better results by modelling the human brain behaviour.

5. CONCLUSIONS

This paper presents an efficient method of determining the saliency map for a digital representation of a painting. The algorithm manages to model the human behaviour when analyzing a new scene by identifying the most relevant regions of a painting based on a set of colour, shape and edges features.

Further research will be made to extend the used set of features and to find a better classification method. The final aim of the research is a method capable of matching as closely as possible the human behaviour related to region of interest detection in paintings.

ACKNOWLEDGMENT

This work has been partially supported by the Sectoral Operational Program Human Resources Development (SOP HRD) 2007-2013, financed from the European Social Fund and by the Romanian Government under the contract number POSDRU/107/1.5/S/76903.

REFERENCES

[1] B. Gunsel, S. Sariel, and O. Icoglu, "Content-based access to art paintings," Proceedings of IEE ICIP, Vol. 2, pp. 558-561, Genoa, 2005.

[2] C. Li and T. Chen, "Aesthetic visual quality assessment of paintings," IEEE Journal of Selected Topics in Signal Processing, vol. 3, issue 2, pp. 236–252, March 2009.

[3] E. Brinkman, "Recognition of the emotion in abstract expressionistic paintings using a machine learning approach", MSc Thesis, Department of Human Media Interaction, Twente University, Enschede, The Netherlands, January 2008, unpublished.

[4] J.P. Collomosse and P.M. Hall, "Cubist style rendering from photographs". IEEE Transaction on Visualization And Computer Graphics, vol. XX, No. Y, 2002.

[5] R.L. Solso, "Cognition and the visual arts," MIT Press, Bradford Books Series in Cognitive Psychology, 1994.

[6] D. De Carlo and A. Santella, "Stylization and abstraction of photographs", Proceedings of ACM SIGGRAPH 2002, Vol. 21, Issue 3, pp. 769-776, San Antonio, USA, July 2002.

[7] 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, Vol. 20, No. 11, pp. 1254-1259, November 1998.

[8] L. Itti, C. Koch, "A comparison of feature combination strategies for saliency-based visual attention systems", Journal of Electronic Imaging, Vol. 10, pp. 161-169, 1999.

[9] L. Itti, "Quantitative modelling of perceptual salience at human eye position", Journal of Visual Cognition, Vol. 14, pp. 959-984, 2006.

[10] Y.W. Limn and S.U. Lee, "On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques," Journal of Pattern Recognition, Vol. 23, Issue 9, pp. 935-952, 1990

[11] V. Gouet, N. Boujemaa, "About optimal use of color points of interest for content-based image retrieval", Research Report RR-4439, INRIA, 2002.