Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 219–223

Towards Determining Syntactic Complexity of Visual Stimuli Used in Art Therapy

Bolesław Jaskuła, Jarosław Szkoła
University of Information Technology and Management
Sucharskiego Str. 2, 35-225 Rzeszów, Poland
Email: {bjaskula, jszkola}@wsiz.rzeszow.pl

Krzysztof Pancerz
University of Management and Administration
Akademicka Str. 4, 22-400 Zamość, Poland
Email: kpancerz@wszia.edu.pl
University of Information Technology and Management
Sucharskiego Str. 2, 35-225 Rzeszów, Poland

***Abstract*—In the paper, we deal with the problem of automatic determining syntactic complexity of visual stimuli. This problem is important in case of using paintings in eye-tracking based diagnosis and therapy of some kinds of neuropsychological and emotional disorders. Our approach to solving the considered problem is based on the clustering procedure using Self Organizing Feature Maps. The clustering results are compared with the heat maps obtained in the eye-tracking process.**

I. Introduction

ART therapy is based on the idea that the creative process of art making is healing and life enhancing and is a form of nonverbal communication of thoughts and feelings. Like other forms of psychotherapy and counseling, it is used to encourage personal growth, increase self-understanding, and assist in emotional reparation, and has been employed in a wide variety of settings with children, adults, families, and groups. Art therapy supports the belief that all individuals have the capacity to express themselves creatively and that the product is less important than the therapeutic process involved [1]. However, not only art making but also art viewing can have a therapeutic influence. Participating in the arts and viewing the arts have been found to have tremendous therapeutic impact [2]. Ulrich conducted a significant study of psychiatric patients' response to art from an extensive and diverse collection of wall-mounted paintings and prints [3].

Constant research is needed for increasing effectiveness of art diagnosis and therapy. Research in the area of neuroaesthetics is an example of this type of activities. As its name implies, neuroaesthetics is an attempt to combine neurological research with aesthetics by investigating the experience of beauty and appreciation of art on the level of brain functions and mental states. The first approach relies on observation of subjects viewing art samples and inspection of the mechanism of vision, with the aim of inducing general rules about aesthetics. This is the most popular approach to neuroaesthetics proposed by Zeki [12].The second approach aims at establishing the link between certain brain areas and artistic activity. In contrast to approaches focusing on the artistic abilities and creativity, the third approach investigates aesthetic enjoyment through brain-imaging experiments on subjects looking at pictures. A fundamental methodological crux for all these approaches is whether the aesthetic judgments are perceived as bottom-up processes driven by neural primitives or as top-down processes with high-level correlates [4].

Conclusions presented above enable us to hope that the visual art can be an effective tool in the diagnosis and therapy process, for example, in the treatment of some kinds of neuropsychological and emotional disorders. In order to do that, we need research systematizing the methodology of application of the art as a therapeutic tool. The first step of our research has been presented in [11]

II. Syntactic Complexity of Visual Stimuli

As it was mentioned earlier, painting perception is connected with the activity of a number of regions of the brain. A structure of visual stimuli, i.e., its complexity, influences which regions of the brain are activated by visual stimuli (painting), i.e., which cognitive functions (basic or higher) are initiated by the patient. Therefore, an important goal of conducted research is categorization of visual stimuli according to their complexity, i.e., their usefulness for diagnosis or therapy in the treatment of different kinds of neuropsychological and emotional disorders.

Paintings can represent many things and be analyzed on various syntactic and semantic levels [5] (cf. Figure 1). Jaimes and Chang [6] developed an overview structure of these content levels for the indexing of images. Jaimes and Chang's Index Pyramid consists of four syntactic and six semantic levels, where the width of each level gives an indication of the amount of knowledge required to describe the image content on that level. Even though the lower levels consist of upper levels, single levels can be seen as individual parts.

Syntactic levels define the visual elements such as colors and lines and are in close relation to visual perception. Semantic levels are concentrated on visual concepts and they define the meanings of the visual elements and of their arrangements. As can be seen from the Pyramid, semantics can be observed on the general, specific or abstract level. Jörgensen [7] divides painting attributes into three different groups; perceptual, reactive and interpretive attributes. Perceptual attributes are closely related to visual stimuli, e.g. the color "red". Reactive attributes are peoples' personal reactions to paintings, e.g. uncertainty, confusion and liking the image.

978-1-4673-4471-5/$25.00

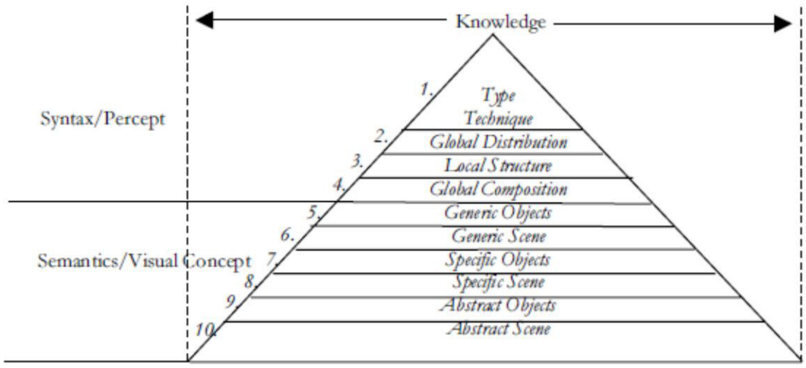


Fig. 1. The ten-level visual structure

Interpretive attributes include features like sentiment, abstract concepts and content elements like action and function.

The image complexity is a sum of syntactic and semantic complexity. At the most basic level, we are interested in the general visual characteristics of the image or the video sequence. Descriptions of the type of image or video sequence or the technique used to produce it are very general, but prove to be of great importance. Images, for example, may be placed in categories such as painting, black and white, color photograph, and drawing [6]. Global distribution aims to classify images or video sequences based on their global content and is measured in terms of low-level perceptual features such as spectral sensitivity (color), and frequency sensitivity (texture). Individual components of the content are not processed at this level. Global distribution features, therefore, may include global color (e.g., dominant color, average, histogram), global texture (e.g., coarseness, directionality, contrast), global shape (e.g. aspect ratio), global motion (e.g. speed, acceleration, and trajectory), camera motion, global deformation (e.g. growing speed), and temporal/spatial dimensions (e.g. spatial area and temporal dimension), among others. In contrast to Global Structure, which does not provide any information about the individual parts of the image or the video sequence, the Local Structure level is concerned with the extraction and characterization of the image's components. At the most basic level, those components result from low-level processing and include elements such as the Dot, Line, Tone, Color, and Texture. At this level, we are interested in the specific arrangement of the basic elements given by the local structure, but the focus is on the Global Composition. In other words, we analyze the image as a whole, but use the basic elements described above (line, circle, etc.) for the analysis.

One of the key goals realized by the observer at the syntactic level is to detect image contours. One common attribute of paintings in a broad array of artistic traditions, starting from the earliest surviving depictions on cave walls, is the use of boundary lines to depict the edges of objects. There are no actual contour lines dividing real objects from their backgrounds in most cases, which raises the question of why contour lines are so ubiquitous and effective in depiction. One theory is that line drawings are a convention that are imposed within a particular culture and passed down through learning [8].

Recapping, image complexity at the syntactic level is the degree of cognitive effort to which the observer is forced by the structure (arrangement) of the painting obtained by means of adequate painting techniques at the level of receiving physical stimuli. The degree of this type of complexity can be determined in terms of uncertainty or redundancy.

III. Procedure

Human eyes are unable to observe the whole image with the equal acuity. The area of acute vision covers a field of less than 1.5° of arc. A standard eye-tracking examination is performed within 50 cm of the screen. An image has a horizontal resolution of 1200 - 1600 pixels. Therefore, the area of acute vision covers a field with a diameter of 25 pixels. In the performed examinations, the size of a segment has been selected proportionally to the image with a horizontal resolution of 1200 pixels. Art perception is dependent on physical features of the human eye. Therefore, splitting the painting into smaller parts is important in syntactic processing. Research carried out using eye-tracking shows that data processing is realized in the form of fixations followed by saccades (i.e., transitions between fixations). We have created a specialized computer tool for the objective analysis of painting complexity - entropy. The painting with a low information content has a small value of entropy. According to information theory, the smaller probability of object occurrence, the greater information. Results obtained by means of the created tool have been compared with results obtained by means of eye-tracking in the form of the so-called heat maps. Heat maps represent the fixation locations and the duration of the fixations. The regions indicated by our tool, in most cases, are covered by experimental results obtained in the eye-tracking process. On this basis, one can determine that in the syntactic analysis of the painting, most vision fixations concern regions whose structures differ from other regions of the same painting. It is the foundation of the syntactic analysis. The main advantage of the computer tool is independence from individual features. Moreover, results are not sensitive to subjective factors.

The approach presented in this paper is based on a clustering procedure using Self Organizing Feature Maps. The concept of a Self-Organizing Feature Map (SOM) was originally developed by Kohonen [9]. SOMs are neural networks composed of a two-dimensional grid (matrix) of artificial neurons that attempt to show high-dimensional data in a low-dimensional structure. Each neuron is equipped with modifiable connections. Self-Organizing Feature Maps possess interesting characteristics such as self-organization, i.e., networks organize themselves to form useful information, as well as competitive learning, i.e., network neurons compete with each other. The winners of the competition strengthen their weights while the losers' weights are unchanged or weakened. SOMs are used in a clustering process. Input data are vectors composed of m features (elements).

In our approach, each SOM is a feedforward three-layer neural network, one layer for one color component (R, G, or B). This architecture can also be used for other color models, e.g. YUV.

The training process of SOMs can be presented as a list of the following steps:

1) **Initialization**. At the beginning, weights are initialized with random values from a given interval (w_{min}, w_{max}), i.e.:

$$map(i,j,k) = random(w_{min}, w_{max}),$$

where k determines a layer of the multilayer map, i and j are coordinates in the k-th map layer. An initial value for a learning rate α is equal to α_{top}. During the learning process, a value of α is changed from α_{top} to α_{bottom} with step α_{step}, where p is a number of epochs. The initial size of the map is equal to $s_{min} \times s_{min}$ of neurons. This size is progressively increased to $s_{max} \times s_{max}$, where $s_{max} = max(s_{min}, \sqrt{2n}+1)$.

2) **Reading input data**. Input vectors are normalized to the interval $[0,1]$. All input vectors have the same dimension, i.e., m.

3) **Iterations**.
   a) At each iteration, a random input vector is entered to the input layer $x_{current} = \mathbf{x}(random(1,n))$.
   b) A winning Kohonen neuron is determined, i.e., the neurons compete on the basis of which of them have their associated weight vectors "closest" to $x_{current}$. The winner is selected on the basis of minimization of the mean squared error, i.e.:

$$\min_{i,j} \sum_{k=0}^{m} \left(x_{current}(k) - map(i,j,k)\right)^2$$

   c) For the winner and direct neighbours only, weights are modified according to:
      - winner:

$$map(i,j,k) = \alpha\left(x_{current}(k) - map(i,j,k)\right),$$

      - direct neighbours:

$$map(i,j) = 0.6\alpha\left(x_{current} - map(i,j)\right).$$

   d) The learning rate is updated according to $\alpha_{e+1} = \alpha_{top} - e\alpha_{step}$, where e is the current epoch number indicator.
   e) The size of the map is updated:

$$s_{e+1} = s_{max}\frac{e}{p},$$

   if s_{e+1} is greater than the current size, where p is a number of epochs.
   f) If the size of the map has been changed, weights are updated according to:

$$map_{new}(i,j,k) = \\ = \sum_{neighbours\ of\ (i,j)} 0.6map_{neighbour}(i,j,k) + \\ + map(i,j,k).$$

In a standard algorithm, we can distinguish the following steps. An image is recorded in the RGB format. Next, SOM is trained on the basis of the RGB components, separately for each component. However, in each step, weights of maps are also averaged (using the weighted average) for color components. Application of SOMs causes generalization of relationships between pixels of an input image by grouping similar regions close to each other. SOMs are widely used in problems demanding reduction of input data dimension. The obtained map represents classes into which input data space can be divided. In most cases, this is the last step of grouping data. The disadvantage of such an approach is a region of tolerance too broad for neighborhoods of individual classes. Generally, for each class, only its centroid is indicated.

In this paper we propose to extend the standard algorithm:

1) An input image is split into segments constituting a grid of the size $N \times N$. The size of the individual segment depends on the size of the image.
2) Each segment is independently processed. Data for a given segment are passed to the input of SOM. As a result, we obtain maps (i.e., layers of SOM) for each component: R, G, and B. SOM is trained using the procedure described earlier.
3) On the basis of maps for components, a minimal spanning tree (MST) is created using the weighted average of color components. The spanning tree includes information about correlations of pixels of the input image:

$$tree[i][j] = \begin{cases} w_{ij} & \text{if } (i,j) \text{ belongs to MST,} \\ 0 & \text{otherwise,} \end{cases}$$

where i and j are pixel indexes, w_{ij} is a correlation factor of pixels.
4) The coefficient $C_{seg}(m,n)$ is calculated for each segment (see Algorithm 1) with coordinates m and n. In

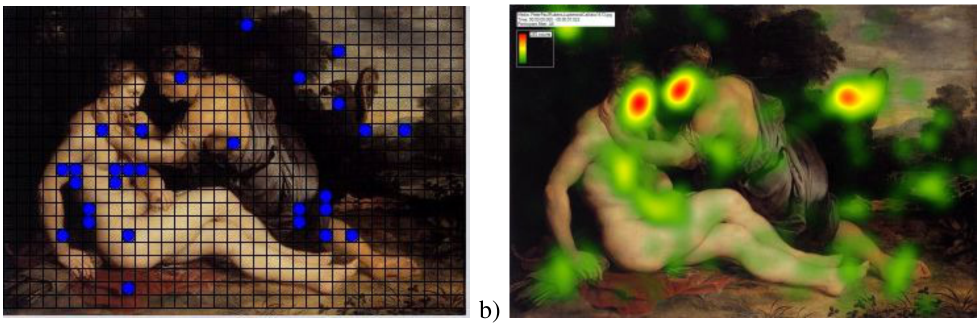
a)  b)

Fig. 2. Example 1: a) the clustering result; b) the heat map

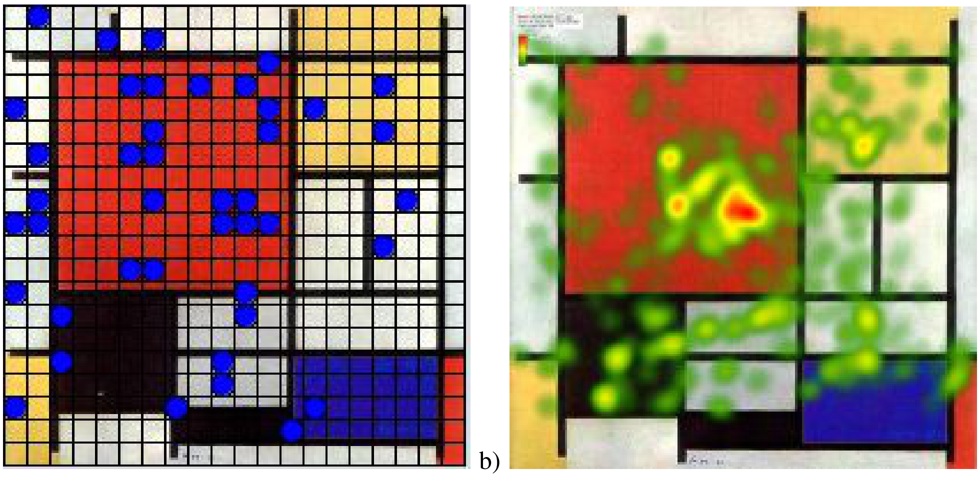
a)  b)

Fig. 3. Example 2: a) the clustering result; b) the heat map

Algorithm 1: Algorithm for calculating the coefficient $C_{seg}(m, n)$

$C_{seg}(m, n) \leftarrow 0$;
for *each pixel* i **do**
  for *each pixel* j **do**
    if $tree[i][j] > CORR_{min}$ *and* $dist(i, j) < DIST$ **then**
      $C_{seg}(m, n) \leftarrow C_{seg}(m, n) + 1$;
    end
  end
end

this algorithm, pairs of pixels are compared. Moreover, $CORR_{min}$ is a threshold correlation between pixels and $DIST$ is a threshold distance between pixels (for example, the Manhattan distance).
This step enables us to omit pixels which potentially are not important for visual perception. Moreover, correlation between pixels which are far apart is not taken into consideration. Thresholds $CORR_{min}$ and $DIST$ have been determined experimentally, and they are equal to:

$$CORR_{min} \approx 2,$$

$$DIST = \frac{D_{segm}}{2},$$

where D_{segm} is the width of the segment.
The greater value of the coefficient $C_{seg}(m, n)$ indicates a more complex structure of the segment, e.g. gradients, colors, shapes, etc.

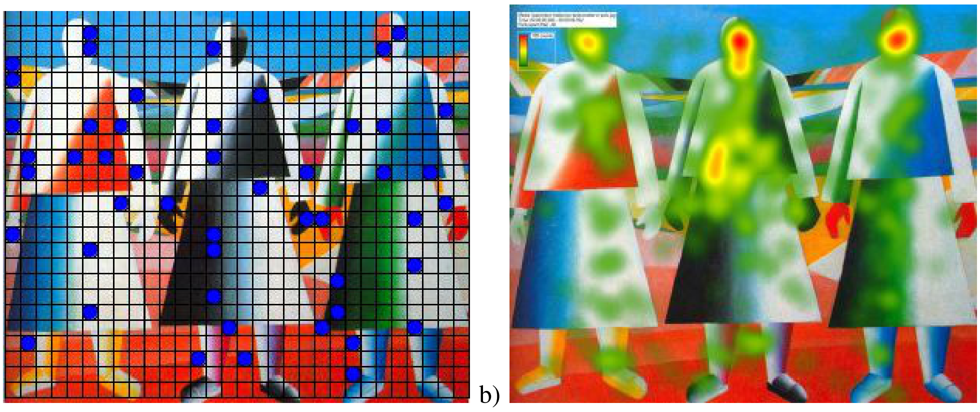
5) We search for numbers of segments with the same coefficient $C_{seg}(m, n)$. Let $VAL_{C_{seg}}$ be a set of all values of coefficients $C_{seg}(m, n)$ calculated in the previous step. For each value v in $VAL_{C_{seg}}$, we calculate its number $OCC_{C_{seg}}(v)$ of occurrences.
6) We calculate the entropy for the analyzed image:

$$H = -\sum_{v=1}^{k} p(v) \ln p(v),$$

where:

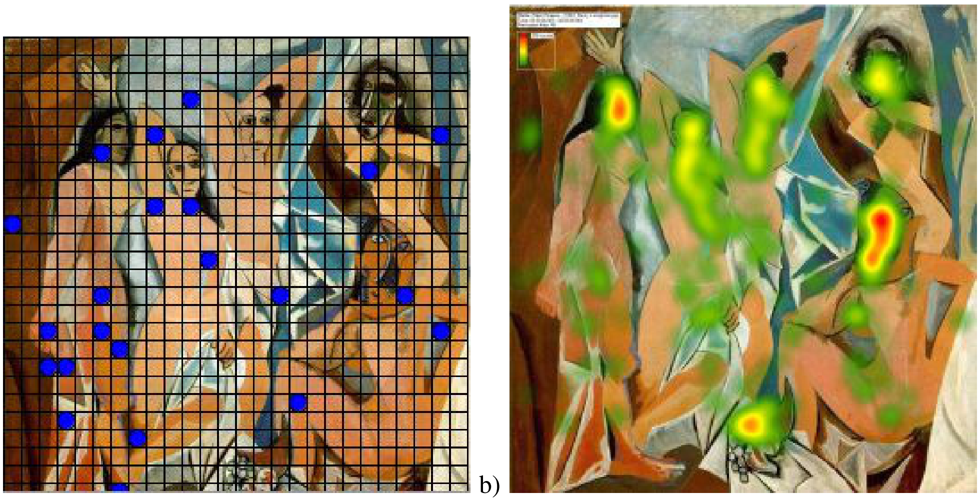
- k is a cardinality of $VAL_{C_{seg}}$,
- $p(v) = \frac{OCC_{C_{seg}}(v)}{N}$,
- $N = \sum_{v=1}^{k} OCC_{C_{seg}}(v)$.

The greater entropy indicates that the complexity of the image is greater, i.e., a number of segments with different coefficients $C_{seg}(m, n)$ is greater.



a) b)

Fig. 4. Example 3: a) the clustering result; b) the heat map



a) b)

Fig. 5. Example 4: a) the clustering result; b) the heat map

IV. Experiments

Experiments have been carried out using some well known paintings. Figures 2 - 5 present exemplary comparisons of the clustering results and the heat maps obtained in the eye-tracking process.

V. Conclusions and Further Works

In the paper, we have shown the computer tool based on Self-Organizing Feature Maps enabling us to automatically determine complexity of visual stimuli used in syntactic analysis. Such a method is necessary in diagnosis and therapy of some neuropsychological and emotional disorders using eye-tracking. This problem is the main task for our future work.

References

[1] C. A. Malchiodi, Ed., *Handbook of Art Therapy*. New York: The Guilford Press, 2003.

[2] J. Rollins, J. Sonke, C. R., A. Boles, and J. Li, Eds., *State of the Field Report: Arts in Healthcare*. Washington, DC: Society for Arts in Healthcare, 2009.

[3] R.S. Ulrich, "Effects of interior design on wellness: theory and recent scientific research," *Journal of Health Care Interior Design*, vol. 3, pp. 97–109, 1991.

[4] A. A. A. Salah and A. A. Salah, "Technoscience art: A bridge between neuroesthetics and art history?" *Review of General Psychology*, vol. 12, no. 2, pp. 147–158, 2008.

[5] V.-M. Saarinen, M. Laine-Hernandez, and H. Saarelma, "The influence of image content levels and looking type on eye movements," *Graphic Arts in Finland*, vol. 35, no. 2, pp. 1–12, 2006.

[6] A. Jaimes and S. fu Chang, "A conceptual framework for indexing visual information at multiple levels," in *Proceedings of SPIE INTERNET IMAGING 2000*, 2000, pp. 2–15.

[7] C. Jörgensen, "Indexing Images: Testing an Image Description Template," in *ASIS Annual Conference Proceedings*, 1996.

[8] D. Melcher and P. Cavanagh, "Pictorial cues in art and in visual perception," in *Art and the senses*, F. Bacci and D. Melcher, Eds. Oxford University Press, 2011, pp. 359–394.

[9] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, no. 1, pp. 59–69, 1982.

[10] B. Jaskuła and K. Pancerz, "Toward interactive computer systems based on eye-tracking technology modernizing didactics of visual art perception," *CyberEmpathy: Visual Communication and New Media in Art, Science, Humanities, Design and Technology*, vol. 1, 2012.

[11] B. Jaskuła, J. Szkoła, and K. Pancerz, "SOM Based Segmentation of Visual Stimuli in Diagnosis and Therapy of Neuropsychological Disorders," in *Proceedings of the International Conference on Man-Machine Interactions*, 2013.

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