DOI: 10.22213/2658-3658-2019-33-39

Solving the Problem of Automated 2D Images Compositional Characteristics Evaluation

F. F. Gazimzyanov [1], M. A. Al Akkad [2]

Computer Science Department, Kalashnikov Izhevsk State Technical University, Izhevsk, Russian Federation
E-mail: [2] aimanakkad@yandex.ru

Received: July 21, 2019

Abstract. This paper contains an overview of the progress achieved by researchers in solving the problem of evaluating 2D images compositional characteristics. A review of methods potentially suitable for solving such a problem is given, the selected method is justified, and the adaptation of the selected method to a specific task is given. A mathematical model adapted to work with existing models using a new method called Adjusting the Structural Skeleton Coefficients is presented. The structure of the training samples and the special aspects of data collection are described. Data analysis and sorting is performed using a developed genetic algorithm, and the choice of the method is justified. The obtained results are analyzed, and visualization of the compositional parameters of simple scenes, for different groups of respondents identified during data sorting and analysis, is introduced. Finally, the overall results of the research are presented, concluding that they coincide with the suggestion of Arnheim about perception.

Keywords: Structural Skeleton, Image Analysis, Computer Vision, Perception, Genetic Algorithms

1. INTRODUCTION

A general concept of an automated system for evaluating the compositional characteristics of 2D images was developed based on the work of Rudolf Arnheim. A mathematical model [1] for the first chapter of Arnheim's book "Balance" [2] was proposed, which allows obtaining the compositional characteristics of objects located inside the "structural skeleton" of the image.

For the accurate operation of the system, *pF* (perceptive force) coefficients of the structural skeleton should be customized with the peculiarities of human perception. There are no exact values by which a graph of such function could be plotted in the book of Rudolf Arnheim except the value at which the object stands at the source of the structural skeleton. So the system should be configured with human perception data [3].

There are several methods for solving this problem: 1) *Least square method:* The developed mathematical model can be expressed as an equation, and the survey will provide statistics. The least squares method can be applied similarly to the equalization on elements given by Linnik Yu. V. [4] in chapter VIII pp. 203-211. 2) *Gradient descent:* The method is also seen to be usable, since the developed mathematical model of the structural skeleton can be presented as a complex function. The acceptable solutions set of this function can be investi-

---

gated and we can find the required minimum, at which the parameters of the equation of the structural skeleton would be adjusted in accordance with the human perception [5]. 3) *Artificial neural networks:* It is worth to note that this approach is universal, but maybe redundant in some cases. However, it can be used to solve the problem of selecting coefficients for sources of perceptive forces of the structural skeleton [6, 12]. 4) *The genetic algorithm*: Genetic algorithms and such approaches are used in the training of neural networks, and also they can be an independent optimization method. In this case, we are talking about an independent method. It is also important to note that, despite the fact that genetic algorithms work similar to the gradient descent, they do not require a strict definition of the mathematical model.

To offer machines the ability of aesthetic quality assessment of images a lot of research still need to be made. For assessing photos quality, high-level features were designed [15], the consensus on visual quality was learned for image management [16], for automated albuming and retrieval multidimensional image value assessment and rating was used [17], several visual quality assessment algorithms were investigated [18], an interesting study about aesthetic visual quality assessment of paintings was carried out [11].

2. ADJUSTMENT OF THE MATHEMATICAL MODEL COEFFICIENTS

The genetic approach was selected because it provides a convenient isolated layer of abstraction over existing methods. Moreover, it allows operating with atomic units and fairly simple concepts from genetics to control the learning process. Empirical results of efficiency show incomparably higher efficiency than random solution search [7]. However, of course, such an approach will lose speed in any analytical solution, such as LSM (Least square method) or unconstrained optimization.

2.1. Adjusting the Structural Skeleton Coefficients Method

The principle of DNA functioning, genes transfer from parents to children and directed development of genetic systems is described by Richard Dawkins in his book "The Selfish Gene" [8, pp. 60–94].

The author defines the gene as a carrier of certain genetic information, a feature that, however, has not definite physical boundaries in the sequence of the chromosome. However, in our case we operate with atomic numerical parameters, in the same way they are defined in the computer memory. Therefore, it will be more convenient to identify the concepts of a gene and a cistron, chromosomes regions bounded by certain sequences that designate the beginning and end of a cistron [8, p. 69]. Then accordingly the chromosome is a sequence of genes. A genotype is a specific set of chromosomes, and the concept of a phenotype can be reduced to a combination of genotype and environmental factors.

Now we will apply all this to the tuning of the developed system. Let's turn to the formula:

$$pF_{scene} = \begin{cases} \left|pF_{dots}\right| \in \left[0;e\right] & \mid \; pF_{seg}\left(pF_{s1}, pF_{s2}...pF_{sN}\right)\times 2 \\ \left|pF_{dots}\right| \notin \left[0;e\right] & \mid \; pF_{seg}\left(pF_{s1}, pF_{s2}...pF_{sN}\right) + pF_{dots}\left(pF_{d1}, pF_{d2}...pF_{dM}\right) \end{cases}, \quad (1)$$

where e – the error value at which we consider the influence of point sources of structural plan is equal to zero, pF_{dots} – the influence of point sources of structural plan for the object, pF_{seg} – the influence of line sources of structural plan. While pF_{sN} and pF_{dM} are the settlement functions of the value of the impact of perceptual forces of a certain line or point source of the structural plan on object.

Each pF_s and pF_d can be represented in the context of the genetic algorithm context as a chromosome. In turn, the x parameters from formulas

$$pF_{segs} = \frac{\sum_1^N pF_{sN}\left(d_N, x_{1N}, x_{2N}, \ldots, x_{VN}\right)}{N}, \tag{2}$$

$$pF_{dots} = \frac{\sum_1^M pF_{dM}\left(d_M, x_{1M}, x_{2M}, \ldots, x_{WM}\right)}{M}, \tag{3}$$

can be defined as genes, where *d* is the distance to the center of mass of the object within the structural plan and *x* is the parameter of the source structural plan function *pF*. The function is constructed by Akima's spline interpolation [14].

The entire structural skeleton is identical to the genotype.

When we make a set of some initial "genes" or "DNA" for our mathematical model of the structural skeleton, and create "parents", we may get "successors" - models with different variations of pF functions of the structural skeleton. To determine their quality, we need a *fitness* function. We can use formula 1, as well as the polling data and then compare the results of the human estimation and pF_{scene} function. This difference will determine the quality of a certain genes combination. We can get the total difference for all the scenes data from the learning sample derived from the polls. Together with the genotype, the *fitness* function forms a phenotype.

Let's present all of this in a more formal way:

$$gene = x \tag{4}$$

$$chromosome = \left\{gene_1, gene_2, \ldots, gene_n\right\} \tag{5}$$

$$genotype = \left\{chromosome_1, chromosome_2, \ldots, chromosome_n\right\} \tag{6}$$

$$phenotype = \left\{genotype, fitness\right\} \tag{7}$$

$$fitness = \sum_{n=0}^{N} \mid pf_{scene^n}\left(genotype\right) - hE_{scene^n} \mid, \tag{8}$$

where *gene*, is the *x* values of the supporting points of the pF function graph in formulas 2 and 3. While *chromosome*, *genotype*, *phenotype*, *fitness* function, and $pf_{scene^n}\left(genotype\right)$ are the score of a particular scene in formula 1 that were used in the genes of the genotype as the parameters of the pF functions, and hE_{scene^n} is the human estimate.

2.2. Software Implementation

For the software implementation of this method the Jenetics library was selected [7, pp. 60–94]. This library is well documented, developed using the Java Stream API, which allows the full use of hyper-threading JVM, and provides a high level of compatibility with already developed systems that implemented flow control, and allows the use of popular patterns during development [13]. As a format for the declarative description, JSON was chosen[1,2].

---

[1] XML Specifications Page, http://www.xml.com

[2] JSON Specifications Page, http://www.json.org/json-ru.html

One of the minuses is that the library provides immutable entities that may confuse the user who is not familiar with the Builder and Factory patterns [9, p. 89].

3. RESPONDENTS DATA ANALYSIS METHOD

A survey was conducted for obtaining the data. People were asked to rate graphic compositions of the simple geometric objects, like in the first chapter of Rudolf Arnheim book. Amount of 76 compositions was presented in certain forms.

They were printed on A4 sheets in order to identify people's perception of the scenes as accurate as possible, the thing that different sizes and dimensions of computer screens don't offer. For the use of the obtained data, they need to be analyzed to highlight the main trends in the evaluation of certain compositions and to find (or not to find) any patterns among the responses of the questioned people. Obtaining an average estimate is uninformative, since the standard deviation in the estimates was very large. If we look at the distribution of human estimates for each scene, the spectrum of answers, it will be obvious that often in the scenes there are tendencies to different answers, for example, 2 and 5, and finding an average just does not allow to see these features.

This approach allows to obtain interesting learning samples, but if we assume that there are certain taste groups with similar inclinations in the evaluation, which may intersect, finding the mode also will not ensure the representation of all existing conformities. It should be noted that the array of data obtained from interviews can be represented as a three-dimensional plane, a two-dimensional table, where the value in each cell can be interpreted as a height. If we assume that people have common aspirations in the perception of these scenes, smoothing the plane by reversing the rows of a table can group data and allow us to see the necessary dependencies that will allow to adjust the system.

Despite the fact that there is an analytical approach to the problem of smoothing surfaces using sufficient conditions of smoothness [10, p. 71], we have already well configured the system for optimization of a genetic algorithm. Because we haven't the task to obtain the fastest algorithm, we try to sort the data with directional selection. In this case, each chromosome will contain a single gene – ID of a specific line. In addition, all chromosomes must contain a unique set of gene-identifiers, a table row must not be repeated. The Jenetics library [7] provides tools for working with such tasks by using ordered sets and selectors, which leads the enumeration process to recombination, not to crossing-over. The codec allows to obtain the set of identifiers arranged in a certain order from a genotype. We can then build a new version of the table based on this set. The fitness function can work in several ways: we can calculate a sum of absolute differences of the vertical neighboring elements or go through the table by a vertical window, calculating the sum of the average values for each neighboring element. In each case, the smaller result will mean the smoothest surface, and so the people's answers are likely to be grouped by common trends.

The software implementation of such sorting does not differ from the implementation of the optimization of a genetic algorithm mentioned above in section 2.2. "Software implementation". A special type of genes *EnumGene<Integer>* was used, and *PartiallyMatchedCrossover* was used for the crossover, which ensures that each gene will only occur on the chromosome once.

3.1. Results of Respondents Data Analysis

We have analyzed the survey data above mentioned in section 3, which was collected by Google Sheets service as shown in Figure 1, and adjusted using the Jenetics library, and we've discovered that the most scenes had a tendency to one or two estimates, and only a few did not have a tendency and the answers was evenly distributed. The map built by the modes, is depicted on Figure 3. The result is consistent with Arnheim's experiments described in his book that is shown on Figure 2: the more "stable" areas are the right-hand side of the square, as well as the lower sources of the structural skeleton. The lower right corner is the most stable.

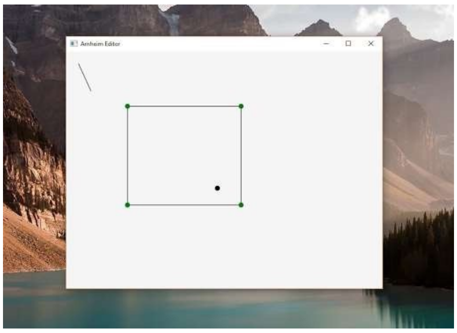

Figure 1. The software environment visual interface

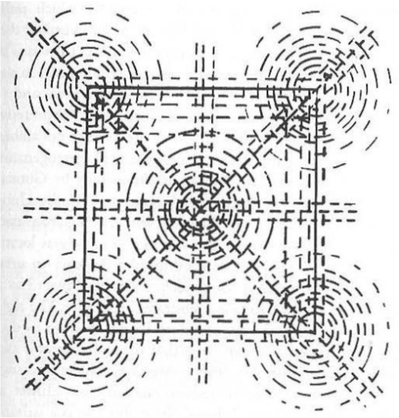

Figure 2. Arnheim's structural skeleton ([2], pp. 13)

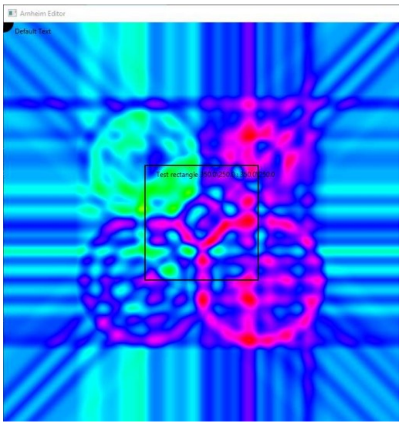

Figure 3. *pF* map for the modes by all estimates

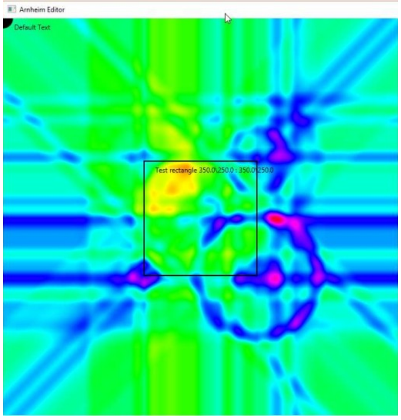

Figure 4. *pF* map for modes on negative evaluations

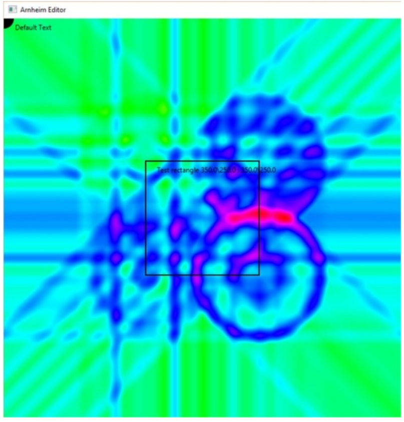

Figure 5. *pF* map for modes on neutral evaluations

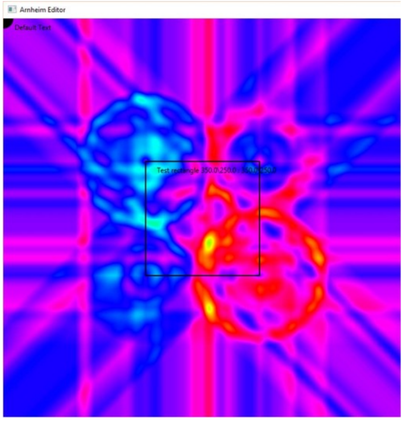

Figure 6. *pF* map for modes on positive evaluations

Also, we can see the predominance in the right and lower parts inside the structural skeleton relatively to the vertical line of symmetry. The visual assumption of the area for the visual center is also shifted to the right lower part. Also results of sorting of all answers are given, with the help of the vertical sum of differences (Figures 4, 5, 6). In the sorted data visually it is possible to allocate groups of people who are inclined to negative (adolescence female artists) and to positive (no clear pattern in this group) estimates.

Also middle-aged men without artistic skills have tendency to more moderate, average estimates. The sorting by the genetic algorithm based on the average value of the estimate in the groups is not informative because of the small number of surveys. Perhaps, more large-scale surveys and additional information about respondents will make it possible to identify the dependencies on average, but the identification of the tendencies seems to be more effective.

If we hold more accurate multivariate analysis, which would allow us to group the results of surveys in a different plane, it would be possible to identify certain “tastes” among people, according to which the system can be configured. This will solve the problem of subjectivity of such systems and methodologies.

4. THE OBTAINED RESULTS AND CONCLUSIONS

We have chosen genetic algorithm as an optimization method and it is an appropriate choice for the research. Besides the overall simplicity of the approach, it turned out to be universal for adjusting the coefficients of the mathematical model, and for sorting data. Analysis of the theoretical values of the mathematical model confirmed that it was a correct representation for the methodology described in Arnheim’s book. There is an obvious similarity between the visualized results in his book and the obtained map. Data analysis showed that operating on the mean values in this subject area was not informative.

Finding the mode of the survey’s results showed a better representation of the distribution of estimates. Analysis of the responses made it possible to conclude that there were certain trends in people’s responses, so the results were sorted by the developed method. System setting according to the survey data and visualization of that field showed that the principles described by Arnheim, manifested even in the analysis of small groups. Created system is ready for more serious data analysis and extension of mathematical models and software tools that will provide more accurate research results.

ACKNOWLEDGMENT

The authors thank Maria Aleksandrovna Nesterova, Ekaterina Alekseevna Belekhova and Anatoly Sergeevich Zagniy for their help in studying the subject domain and its historical part; Regina Shemyakina for help in translation, and Andrei Ilyich Murynov, without whom this research would not have started.

This research is funded by Kalashnikov Izhevsk State Technical University grant 27.06.01/18BCB.

REFERENCES

1. Al Akkad, M. A. & Gazimzyanov, F. (2019). Avtomatizirovannaya sistema otsenki kompozitsionnykh kharakteristik 2D-izobrazheniya: matematicheskaya model' [Automated system for evaluating 2D-image compositional characteristics: mathematical model]. *Intellektual'nye sistemy v proizvodstve [Intelligent Systems in Manufacturing],* 15(2), 105–108. doi: 10.22213/2410-9304-2017-2-105-108 (in Russian).

2. Arnheim, R. (2012). *Iskusstvo i vizual'noe vospriyatie [Art and visual perception].* Moscow, Russia : Arkhitektura-S, 392 pp. (in Russian).
3. Al Akkad, M. A., & Gazimzyanov, F. F. (2019). Avtomatizirovannaya sistema otsenki kompozitsionnykh kharakteristik 2D-izobrazheniya: nastroyka matematicheskoy modeli [Automated system for evaluating 2D-Image compositional characteristics: configuring the mathematical model]. *Intellektual'nye sistemy v proizvodstve [Intelligent Systems in Manufacturing],* 17(1), 26–33. doi: 10.22213/2410-9304-2019-1-26-33. (in Russian).
4. Linnik, Yu. V. (1962). *Metod naimen'shikh kvadratov i osnovy matematiko-statisticheskoi teorii obrabotki nablyudenii [Ordinary least squares and mathematical statistical theory of observational processing basics. 2nd edition].* Moscow, Russia : Gos. izd-vo fiz.-mat. lit. [State publishing house of physical and mathematical literature], 349 pp. (in Russian).
5. Akulich, I. L. (1986*). Matematicheskoe programmirovanie v primerakh i zadachakh [Mathematical programming in examples and exercises].* Moscow, Russia : Vysshaya Shkola [Higher School Publ. House], 319 pp. (in Russian).
6. Korobeinikov, A. V. (2013). *Programmirovanie neironnykh setei [Neural network programming].* Izhevsk, Russia : Publ. House of Izhevsk State Technical University, 44 pp. (in Russian).
7. Wilhelmstötter, F. (2017). *Jenetics library user's manual.* Jenetics. Retrieved from: http://jenetics.io/manual/manual-3.8.0.pdf.
8. Dawkins, R. (2013). *Egoistichnyi gen* [Egoistic gene]. Moscow, Russia : Corpus Publ., 512 pp. (in Russian).
9. Gamma, E., Khelm, R., Dzhonson, R., & Vlissides, D. (2016). *Priemy ob"ektno-orientirovannogo proektirovaniya. Patterny proektirovaniya [Object-oriented design techniques. Design Patterns].* St. Petersburg, Russia : Piter Publ., 366 pp. (in Russian).
10. Poznyak, E. G., & Shishkin, E. V. (1990). *Differentsial'naya geometriya: pervoe znakomstvo [Differential geometry: first acquaintance].* Moscow, Russia: Publ. house of Moscow State University, 384 pp. (in Russian).
11. Li, C., & Chen, T. (2009). Aesthetic visual quality assessment of paintings. *IEEE Journal of Selected Topics in Signal Processing, 3*(2), 236-252. doi: 10.1109/JSTSP.2009.2015077.
12. Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In: *European Conference on Computer Vision,* pp. 818-833. arXiv: 1311.2901.
13. Eckel, B. (2006). *Thinking in Java.* Prentice Hall, 4th edition. 1150 pp.
14. Akima, H. (1970). A new method of interpolation and smooth curve fitting based on local procedures. *Journal of the ACM, 17,* 589–602. doi: 10.1145/321607.321609.
15. Ke, Y., Tang, X., & Jing F. (2006). The design of high-level features for photo quality assessment. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06),* 1, 419–426. doi: 10.1109/CVPR.2006.303.
16. Datta, R., Li, J., & Wang, J. Z. (2007). Learning the consensus on visual quality for next-generation image management. In: *Proceedings of the ACM International Conference on Multimedia* (pp. 533–536). doi: 10.1145/1291233.1291364.
17. Loui, A., Wood, M. D., Scalise, A., & Birkelund, J. (2008). Multidimensional image value assessment and rating for automated albuming and retrieval. In: *15th IEEE International Conference on Image Processing* (pp. 97–100). doi: 10.1109/ICIP.2008.4711700.
18. Moorthy, A. K. & Bovik, A. C. (2011). Visual quality assessment algorithms: what does the future hold? *Multimedia Tools and Applications, 51,* 675–696. doi: 10.1007/s11042-010-0640-x.