Partially Interactive Evolutionary Artists

Penousal MACHADO and Amílcar CARDOSO

Centre for Informatics and Systems of the University of Coimbra
3030 Coimbra, Portugal

machado@dei.uc.pt, amilcar@dei.uc.pt

Juan ROMERO and Antonino SANTOS

Faculty of Computer Science, University of Coruña
15071 A Coruña, Spain

jj@udc.es, nino@udc.es

Abstract User fatigue is probably the most pressing problem in current Interactive Evolutionary Computation systems. To address it we propose the use of automatic seeding procedure, phenotype filters, and partial automation fitness assignment. We test this approaches in the visual arts domain. To further enhance interactive evolution applications in aesthetic domains, we propose the use of artificial art critics – systems that perform stylistic and aesthetic valuations of art – presenting experimental results.

Keywords Interactive Evolutionary Computation, User Fatigue, Partial Automation, Artificial Art Critics

§1 Introduction

The development of solutions to creative and artistic tasks may prove fundamental for the progress of the AI field. The desire to use computation for building artistic systems can be traced back to Ada Lovelace, who dreamt 150 years ago of the creation of a computer with musical capabilities. From that moment, artificial artistic systems have been studied using numerous computational techniques, including expert systems, Artificial Neural Networks (ANNs) and statistical and stochastic methods[12]. In the last few years, Evolutionary Computation (EC) techniques have been widely used in artistic domains[6, 14, 4, 19]. In most cases these systems are guided by the user that evaluates the individuals. This paradigm is named Interactive Evolutionary Computation (IEC), for which a thorough survey can be found in[17].

One of the main problems in this kind of systems is user fatigue, caused by the need to evaluate a large number of artistic works, which results from the lack of critical capacity of most systems. In the field of music there are several systems that perform the partial or total automation of fitness assignment (see, e.g.[19]). In the field of visual arts the attempts to include valuation capabilities are few. Baluja et al.[3] proposed the use of an ANN, trained with a set of IEC generated images, and used it to assign fitness. However, the results are, according to the authors, disappointing[3]. In previous work[8, 7, 9] we described the use of complexity estimates to automate fitness assignment presenting experimental results. Using a related approach, Sauders and Gero[15] propose the use of an Self Organizing Map ANN to evolve images with an adequate degree of novelty. Although interesting and potentially useful in the scope of IEC, the approach is limited to novelty aspects of the artworks. Also related to these research efforts, Svangaard and Nordin[16] resort to complexity estimates to model user preferences, suggesting the use of this scheme for fitness assignment. In the presented experiments, the authors use sets of two randomly generated images, and compare, for each pair, the choices made by the system with the ones made by the user. Depending on the comparison methodology, the success rates vary between 34% and 75%. Since no example of the considered images is presented, it is difficult to assess the feasibility of using this scheme in the context of IEC.

We propose several solutions to user fatigue, namely: automatic seeding procedure, phenotype filters, and partial automation fitness assignment. These approaches were implemented and tested in NEvAr[*1], an artistic IEC system that allows the evolution of populations of images.

Our approach involves the development of methods for partially automating fitness assignment, thus decreasing the amount of human made evaluations and improving the overall quality of artworks proposed to the user. As a result, there is an improvement of user satisfaction and lessening of user fa-

[*1] Examples of artworks generated with NEvAr can be found at: http://www.dei.uc.pt/ machado/NEvAr.

tigue. We also propose the use of Artificial Art Critics – systems that perform an evaluation of the generated pieces according to some criteria. In the next section, we present relevant background information for the understanding and contextualization of the proposed valuation and comparison methods.

§2 Background

Our proposal relies on the notion that the complexity of an image is an important feature for the assessment of its aesthetical proprieties. This view is supported by a variety of studies. Arnheim[1, 2] was probably the first author to point out and analyze the relations between complexity and aesthetics. Fractal dimension has been used by Taylor to authenticate and date Jackson Pollock paintings in terms of their fractal dimension[18]. For instance, he discovered two different fractal dimensions: one attributed to Pollock's dripping process, and the other attributed to his motions around the canvas. He was also able to track how Pollock refined his dripping technique; the fractal dimension increased through the years (from 1 in 1943 to 1.72 in 1952).

To measure the complexity of an image, one could resort to the notion of information content. Although Kolmogorov complexity is non-computable, it is possible to use a standard compression algorithm to estimate it. We are interested in the apparent complexity of an image according to a human viewer. This apparent complexity can be radically different from the theoretical Kolmogorov complexity. For instance, it is possible to generate a pseudo-random image using a very compact program; in spite of its high apparent complexity, this image would have a low Kolmogorov complexity. Therefore, besides being non-computable, Kolmogorov complexity can be inappropriate for our goals.

To estimate complexity we employ two well-known image compression schemes, jpeg compression and fractal based compression. Since these are lossy compression schemes, we must take into account the amount of error involved in the compression. Accordingly, for each compression scheme, we estimate complexity by computing the ratio between the compression error and the compression rate. We vary the quality of the encoding by setting limits to the maximum error per pixel, and thus the amount of detail kept, which results in three complexity estimates for each compression technique.

Our approach revealed some shortcomings when dealing with the color information. This is partially related with the circular nature of the Hue channel, but it also relates with the interdependency between Hue, Saturation and

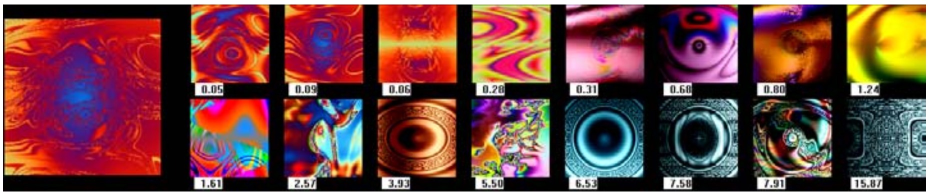

Fig. 1 Output of the retrieval mechanism.

Lightness. Our experiments indicate that, as presently calculated, the complexity of the Hue and Saturation channels is misleading, and that better results are attained considering only the lightness information. As such, the system "sees" the images as greyscale ones.

§3 Automatic Seeding

In NEvAr the user has the possibility to store highly fit individuals in a database, which can later be injected in an ongoing experiment, or used to create a non-random initial population. By resorting to this database, one can significantly decrease the amount of time necessary to create "good" images.

To use the database efficiently it is important to provide methods for searching images according to their characteristics. We want to provide a way of finding images similar to one selected by the user and include them in the current population of the IEC.

To achieve this we need a way of comparing images. After testing several measures of distance among images, we arrived to the conclusion that a pixel by pixel comparison would not be appropriate. Due to the artistic nature of our system, we propose that images should be compared according to their aesthetical properties. Taking into account the importance of complexity, in our approach two images are considered similar when their complexity estimates are similar. More precisely the distance between two images is the Euclidian distance among their complexity estimates. Svangard et al. proposed recently an alternative approach, which consists in estimating the distance between two images using the difference between the complexity of the individual images, and the complexity of their concatenation[16].

The user specifies the number of images that will be retrieved from the database, and the system retrieves the most similar ones. In Fig. 1 we present the output of the retrieval mechanism. The image on the left is the one selected by the user, the remaining ones are images from the database sorted according

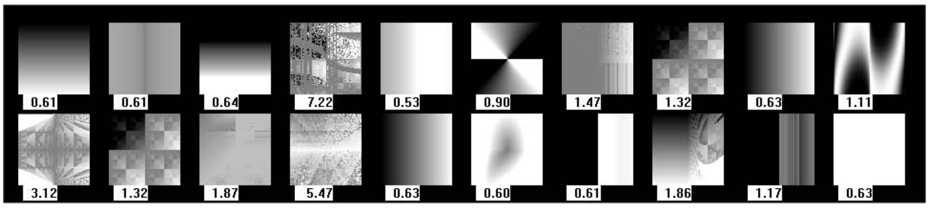

Fig. 2 An initial random population and the corresponding image complexity estimates.

to their similarity with the selected image.

An analysis of this type of experiment involves some degree of subjectivity. Nevertheless there are a several indicators that show that the system is working properly. The first three images retrieved by the system are stylistically similar to the user selected one. In fact, they were all generated in the same evolutionary run and with a very small generation gap. It also clear that the last four images are the least similar ones.

One of the advantages of this approach is its low computational cost. We only need to compute the complexity estimates for the user selected image, the other complexity estimates are stored in the database. The user can also select images that weren't generated with NEvAr, e.g. famous paintings, an try to retrieve similar ones. However, the quality of the results depends on the existence of similar images in the database of images generated with NEvAr. Trying to retrieve images similar to a Van Gogh painting will give poor results, since the current database doesn't include images of that kind.

§4 Filtering

In NEvAr, the generation of images that are either too simple or too complex, e.g. completely blank or totally random, is frequent during the first populations of an evolutionary run. As the population number increases, these images become less frequent, but still occur. Therefore, a significant amount of the images presented to the user are inappropriate. To avoid this problem we developed two types of filters, one working at the phenotype level and the other at the genotype level.

The phenotype filter works as follows: we calculate the image complexity of the individuals belonging to the population, and discard images with complexities values outside a given interval. The user can specify lower and upper bounds for this interval and, therefore, adjust the filtering level.

This method proved to be extremely efficient. In Fig. 2 we present a typical initial population. The numbers bellow the images indicate the corresponding complexity estimate using jpeg compression. Typically, images with complexities outside the $[1, 5]$ interval would be eliminated by the filtering layer and not presented to the user, significantly decreasing the amount of images presented to the user during the first populations. Alternatively, new images can be generated to replace the ones eliminated by the filtering layers, which increases the average quality of the presented images.

The main drawback of this approach is its high computational cost, since the individuals must be rendered in order to estimate their complexity. To cope with this disadvantage we developed a simple genotype level filter. Basically, we verify if the variables x and y are both present (if none of them is present, the pixels of the image will all have the same value; if only one is present, the image will be composed by vertical or horizontal lines); and check if the root of the tree is an appropriate function (e.g. a noise generation operator at the root would result, unavoidably, in a random image).

The genotype filter is applied first, reducing the number of images that are analyzed by the phenotype filter and hence the computational cost.

§5 Partially Automated IEC

In this section we present ongoing research that aims at the development of a partially automated IEC system. Previous experiments with a fully automated system revealed some limitations, namely, the incapability of dealing with color information, and a tendency to premature convergence. Additionally, the quality of the images created using Interactive Evolution is clearly higher.

To overcome these deficiencies we are now developing a partially automated scheme. In each interaction the user can chose between making the evaluation of the population's images or order the system to run in automatic mode during N generations.

The evaluation procedure is a development of the one presented in previous work, which was validated with a psychological test[5)] attaining good results[8)], and tested in the context of a fully automated evolutionary system[9)]. This version was based on the relation between several complexity estimates according to a the following formula:

$$\frac{IC^a}{(PC_1 \times PC_2)^{b_1} \times (PC_1 - PC_2)^{b_2}} \tag{1}$$

Where, IC is the image complexity estimate resulting from jpeg com-

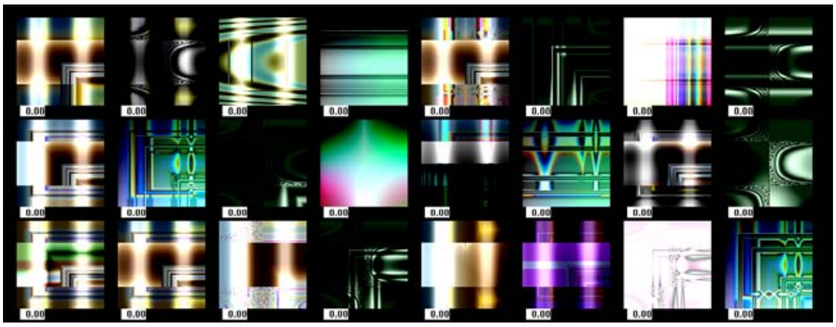

Fig. 3 Twentieth population of a partially automatic run.

pression and PC_1, PC_2 are image complexity estimates resulting from fractal image compression using different levels of detail. The exponents a, b_1 and b_2 allow to control the weights given to each of the components of the formula.

Taking into consideration that we are now interested in a partially interactive mode of execution we introduced some changes to formula 1 in order to give more control to the user. Let $A = IC$, $B_1 = PC_1 \times PC_2$ and $B_2 = PC_1 - PC_2$. In the current version the user can specify optimum values for each component of the formula, as follows:

$$\begin{aligned} A' &= max(0, optimum_A - |A - optimum_A|) \\ B'_i &= optimum_{B_i} + |B_i - optimum_{B_i}| \end{aligned} \tag{2}$$

The user can set the optimum values directly resorting to his own expertise, or indicate an image which he find suitable and let the system set the optimum values by estimating its complexity. These values can be changed at any point of the evolutionary run.

We are conducting several experiments varying the number of consecutive automatic generations and exploring different ways of controlling the automatic evaluation parameters, e.g. setting the optimum values for each component based on the most valued individual of the last population evaluated by the user and including a bias towards simpler or more complex images.

Our tests show that as long as there aren't too many generations without user evaluation, the proposed approach overcomes the deficiencies of the automatic fitness assignment scheme, and is sufficient to ensure the adequate coloring of the images.

In Fig. 3 we present the 20th population of a partially automatic run, in this experiment the user only evaluated three populations (5, 11 and 17).

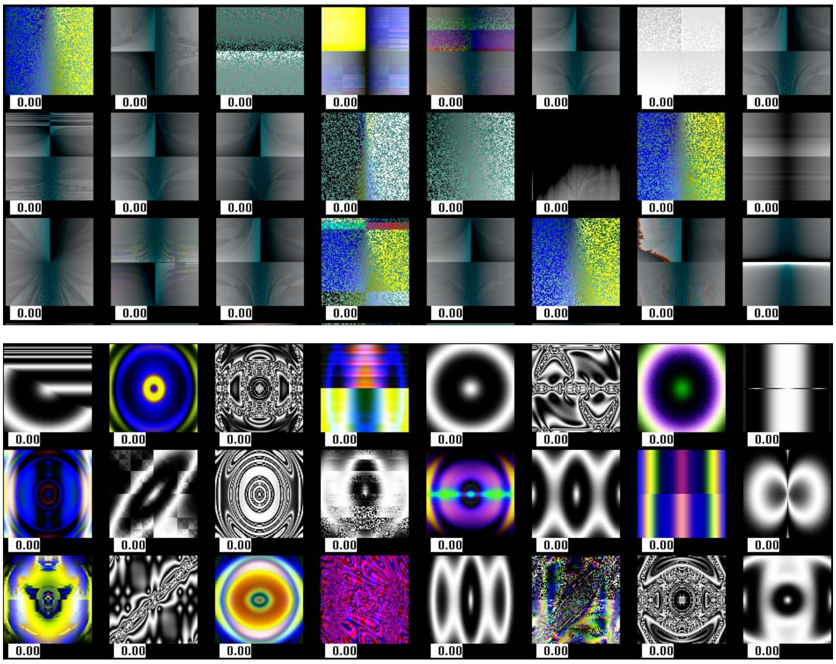

Fig. 4 On the top, the third population of a fully interactive run; On the bottom the twentieth .

The settings for the automatic evaluation formula where the following: $a = 1$, $b_1 = 0.4$, $b_2 = 0.2$, $optimum_A = 5$, $optimum_{B_1} = 0.1$, $optimum_{B_2} = 0.1$.

In Fig. 4 we present the 3rd and 20th populations of a typical fully interactive run. Although the aesthetical judgement of images involves some degree of subjectivity, it is clear that the overall image quality of the 3rd population is significantly lower than the one attained using the partially interactive approach, which required the same user effort. Additionally, the quality of the images of the 20th population is comparable to the ones attained by partial interaction.

In many cases using partial automation yields better results than using a purely interactive scheme, even when the same number of generations is considered.

We used NEvAr in fully interactive mode during countless hours, attempting to produce pleasing images. Like any other tool, NEvAr requires a learning period. Our first attempts where disappointing. With time, we found out that consistency in the evaluation and discipline are fundamental to attain good results. However, for us and for many other users, the temptation of valuing images due to their novelty, in relation to the remaining images of the population, is often too high. Typically, this quest for novelty leads to an unfocused search that doesn't allow the refinement of the results.

Apparently, the use of partial automation promotes a balance between consistency and search for novelty. This can be explained by several factors. The user evaluates fewer populations; therefore, it will take longer to grow tired of a certain type of imagery. The populations evaluated by the user are non-consecutive, and thus tend to be less similar than consecutive ones. The time interval between evaluations is higher, preventing the overload of the user's visual memory. Finally, the automatic evaluation scheme ensures consistency during a significant part of the run.

It is important to stress that the automatic evaluation doesn't take into account color information. It only became possible to evolve images with pleasing color schemes due to the intervention of the user. Moreover, the guidance of the user is fundamental to prevent the convergence to local optima and to compensate the shortcomings of the automatic evaluation scheme.

§6 Artificial Art Critics

We are also interested in the development of Artificial Art Critics (AACs) – systems able to perform stylistic and aesthetic valuations of artworks – for the musical and visual art domains. Once developed, these systems can be used independently to perform art classification tasks; or in the scope of an IEC system to partially automate fitness assignment. In the future, AACs used in conjunction with generation systems can (hopefully) become artificial artists with their own aesthetic criteria.

One of the main difficulties in the development of AACs is their validation. To help address this problem, we propose a multi-stage validation methodology[13]. The first stage consists of an author or style identification and allows the objective and meaningful assessment of the AACs, providing a solid basis for their development. The next stage comprise the ability to perform aesthetic and artistic judgements. In a later stage the AACs will be integrated in a dynamic society, having to adapt in order to address the changing cultural context.

The proposed AACs are composed by two modules: a feature extractor and an evaluator. The feature extractor is static and domain specific. It is responsible for the "perception" of the artwork, generating as output a set of measurements that reflect relevant characteristics of the artwork. These measurements serve as input to the evaluator, which assesses the artwork according to a specific criterion defined by the researcher. The evaluator is an adaptive system, in our case implemented by means of an ANN[13)].

The visual feature extractor is based on the previously described fitness automation methods. We compute the following measurements:

- The jpeg complexity estimates for the whole image;
- The fractal and jpeg complexity estimates for each HSV channel;
- The average value of each channel[*2]; The standard deviation; The slope of the trendline of the Zipf distribution and root mean square error associated with the calculation of the trendline.

This process yields a total of 33 metrics, three for the whole image and ten for each of the channels. To determine the variation of the considered characteristics, we partition it in five regions of the same size – the four quadrants, and an overlapping central rectangle – and compute the described measurements for each partition. This yields a total of 198 measurements (33 for the entire image, and 165 for the partitions).

From an architectural point of view, the adaptive evaluator consists of a feedforward ANN with one hidden layer. We use: standard backpropagation, with a learning rate of 0.2 and a momentum of 0, as learning function; the logistic function, as neuron activation function; and identity, as the output function. In order to build, train and test the ANN we use SNNS.

6.1 Experimental Results

To check the efficiency of the AAC in distinguishing authors by means of their works we conducted some experiments (in [10)] we present a summary of the results). We use a total of 802 different artworks, belonging to six different painters: 98 from Goya, 153 from Monet, 93 from Gauguin, 122 from Van Gogh, 81 from Kandinsky, and 255 from Picasso. For each painter we include artworks of several periods. In most cases, there are significant variations of style and technique within the set of paintings of each author.

The input layer is composed by 198 neurons, one for each measurement of the feature extractor. The measurements made by the feature extractors are normalized between -1 and 1, before feeding them to the network. The hidden layer is composed by 12 neurons, since this configuration gave best overall results. The number of neurons of the output layer is equal to the number of authors, 6

[*2] Taking into account that the Hue channel is circular we compute the average angle.

in this case, the output neuron with higher value indicates the identified author.

The training sets are constructed by randomly selecting a percentage (75% or 85%) of the available pieces. The test sets comprise the remaining ones. In a set of twenty runs the success rate on the classification of test set instances varied between 93% and 96.7%, which indicates that the AAC can successfully identify relevant features in visual arts.

We made an analysis of the classification errors made by the AAC in order to determine which authors were less recognizable. Considering the entire set of performed experiments, the most recognizable painter is Gauguin, whereas Goya was the most difficult to identify. Of the errors made by the ANN 4.3% occurred while classifying Gauguin artworks; 8.5% on Van Gogh pieces; 9.4% on Monet; 10.1% on Picasso. The artworks of Kandinsky and Goya are responsible for, respectively, 30.4% and 37.3% of the total number of errors.

The relative difficulties in identification of Goya artworks can be explained by the reduced number of training instances, and by the existence of large black section in the paintings. For these black areas, the values of the hue and saturation channel is undetermined, which can obviously cause problems. The difficulties in the identification of Kandisnky's paintings can be explained by the heterogeneity of the works included in the training and test set, which include early and late works with significant stylistic differences. As before, the low number of training instances accentuates the problem. Nevertheless, overall, 80% of Goya paintings and 82% of Kandinsky works are correctly identified, so this problem isn't severe.

We also tried to prune the ANN in order to identify the most relevant features. We successively reduced the number of inputs up to 99 by taking into account the connection weights, attaining test set success rates between 93.4% and 94.2%. The reduction of the number of inputs may prove useful, specially since we intend to add other types of metrics (e.g. fractal dimension of the image, results of wavelet transformations, analysis of the color vicinity, etc.).

The experiments in the musical domain[10], using a similar approach and a feature extractor developed by B. Manaris[11], attaining similar success rates.

§7 Conclusions

One of the main shortcomings of current IEC systems is user fatigue. To address this issue we propose several alternatives, namely: the use of an automatic seeding procedure, phenotype filters, partial automation of the fitness assignment. The experimental results attained with these approaches, although subjective, are promising, indicating that these are valid approaches to decrease user fatigue without compromising the quality of the results.

In the context of an IEC, the AACs can be used in several ways to automate fitness assignment. For instance, the user can chose a particular style and let the AAC assign fitness to the individuals according their similarity to the selected style; once the IEC is able to generate individuals that fit the characteristics of the style, fitness can be assigned according to the user preferences and the AAC used as a filter. Alternatively, one can assign fitness according to both criteria, thus allowing some degree of divergence from the selected style.

Currently we are conducting a series of tests with the partially automated version of NEvAr, varying the parameters associated with the fitness assignment formula. We are also performing tests to assess the performance of the AACs in predicting the pleasantness of music scores and artworks. If successful, this will provide us another method for the partial automation of fitness. In a subsequent step AACs and IEC systems will be combined, possibly giving rise to systems that can be considered artificial artists.

Acknowledgment We take this opportunity to express our gratitude towards Bill Manaris who provided invaluable help and made much of this research possible. We also wish to thank María Luisa Santos for her input and for the help in the revision and correction of the paper. This work was partiaally supported by Xunta de Galicia under research project PGIDIT04TIC105012PR)

References

1) Arnheim R., *Art and Visual Perception: A Psychology of the Creative Eye*, University of California Press, 1954.
2) Arnheim R., *Entropy and Art*, University of California Press, 1971.
3) Baluja, S., Pomerleau, D. and Jochem, T., "Towards Automated Artificial Evolution for Computer-generated Images", *Connection Science, 6, 2-3*, pp. 325-354, 1994.
4) Burton A. R., Vladimirova T., "Applications of Genetic Techniques to Musical Composition", *Computer Music Journal, 23, 4*, 1999.
5) Graves, M., *Design Judgement Test*, The Psychological Corporation, New York, 1948.
6) Johnson C., Romero J., "Genetic Algorithms in Visual Art and Music", *Leonardo, 35, 2*, pp. 175-184, 2002.
7) Machado P., Cardoso A., "NEvAr - The Assessment of an Evolutionary Art Tool", in *Proceedings of the AISB'00 Symposium on Creative & Cultural Aspects*

and Applications of AI & Cognitive Science 2000(Wiggins G., ed.), Birmingham, UK, 2000.

8) Machado P., Cardoso A., "Computing Aesthetics", in *Proceedings of the XIVth Brazilian Symposium on Artificial Intelligence SBIA'98*(Oliveira F., ed), Springer-Verlag, LNAI Series, pp. 219-229, Porto Alegre, Brazil, 1998.
9) Machado P., Cardoso A., "All the truth about NEvAr", *Applied Intelligence, Special issue on Creative Systems, 16, 2*,(Bentley P., Corne D., eds.), Kluwer Academic Publishers, pp.101-119, 2002.
10) Machado P., Romero J., Santos M.L., Cardoso A., Manaris B., "Adaptive Critics for Evolutionary Artists", *Lecture Notes in Computer Science, Applications of Evolutionary Computing*, LNCS 3005, Springer-Verlag, pp. 437-446, 2004.
11) Manaris B., Purewal T., McCormick C. "Progress Towards Recognizing and Classifying Beautiful Music with Computers", in *Proceedings of EEE SoutheastCon* pp. 52-57, Columbia, SC, USA, 2002.
12) Papadopoulos G., Wiggins G., "AI Methods for Algorithmic Composition: A Survey, a Critical View and Future Prospects", in *AISB'99 Symposium on Musical Creativity* pp. 110-117, Edinburgh, UK, 1999.
13) Romero J., Machado P., Santos A., Cardoso A., "On the Development of Critics in Evolutionary Computation Systems", *Lecture Notes in Computer Science, Applications of Evolutionary Computing*, LNCS 2611, Springer-Verlag, pp.559-569, 2003.
14) Romero J., Santos A., Dorado J., Arcay B., Rodriguez J., "Evolutionary Computation System for Musical Composition", *Mathematics and Computers in Modern Science*, World Scientific and Engineering Society Press, pp 97-102, 2000.
15) Saunders r., Gero J.S., "The Digital Clockwork Muse: A Computational Model Of Aesthetic Evolution", in *AISB'01 Symposium on Artificial Intelligence and Creativity in Arts and Science* York, UK, 2001.
16) Svangaard N., Nordin P. "Automated Aesthetic Selection of Evolutionary Art by Distance Based Classification of Genomes and Phenomes using the Universal Similarity Metric", *Lecture Notes in Computer Science, Applications of Evolutionary Computing*, LNAI Series, Springer-Verlag, pp. 447-456, 2004.
17) Takagi H., "Interactive Evolutionary Computation: Fusion of the Capabilities of EC Optimization and Human Evaluation", in *Proceedings of the IEEE, 89, 9* pp. 1275-1296, 2001.
18) Taylor R. P., Micolich A. P., Jonas D., "Fractal Analysis of Pollock's Drip Paintings", *Nature*, pp. 399-422, 1999.
19) Todd P. M., Werner G. M., "Frankestenian Methods for Evolutionary Music Composition", *Musical Networks: Parallel distributed perception and performance*,(Griffith N., Todd P. M., eds.), MIT Press, Cambridge MA, 1998.