Journal of Mathematics and the Arts
Vol. 00, No. 00, 2011, 1–21

RESEARCH ARTICLE

Modelling the underlying principles of human aesthetic preference in evolutionary art

Anikó Ekárt[a*], András Joó[ab], Divya Sharma[c] and Stayko Chalakov[a]

[a] *Aston University, Aston Triangle, Birmingham B4 7ET, UK*
[b] *Sapientia University, 540485 Targu Mures, 1C Soseaua Sighisoarei, Romania*
[c] *IIT Ropar, India*

(*Received 22 July 2011, revised 4 March 2012*)

Our understanding of creativity is limited, yet there is substantial research trying to mimic human creativity in artificial systems and in particular to produce systems that automatically evolve art appreciated by humans. We propose here to study human visual preference through observation of nearly 500 user sessions with a simple evolutionary art system. The progress of a set of aesthetic measures throughout each interactive user session is monitored and subsequently mimicked by automatic evolution in an attempt to produce an image to the liking of the human user.

Keywords: aesthetic measure, human preference modelling, evolutionary art, genetic programming, interactive vs automatic evolution

AMS Subject Classification: 68T99

1. Introduction

Ever since the invention of the first computing devices, artists and scientists alike have been thinking about actively involving them in creative endeavours [29]. Producing visually appealing images is certainly one such creative endeavour. Computer generated art is an emerging area [3], with one famous example being the software controlled AARON robot of Cohen that produced pictures (exhibited at Tate Gallery) based on knowledge in the form of rules hard-coded into it. Evolutionary art [3] is a subfield of computer generated art pioneered by Dawkins [11] and Sims [33] that involves using artificial evolution as the underlying engine to generate artworks. Over the past twenty years a lot of effort has been spent on working towards more effective evolutionary art systems that produce visually appealing images. Successful examples attracting substantial public attention include the Electric Sheep [14], the NEvAr system [25] and the Painting Fool [5, 8]. The Painting Fool is not an evolutionary art system per se, but includes evolutionary art modules, such as Elvira [7].

The majority of evolutionary art systems are either interactive (for example [37]) or automatic (for example [1, 13]). Interactive systems tend to generate more visually appealing images, as their driving force is human selection, but at the same time need a lot of effort on the part of the human, may incur user fatigue and could be inconsistent over time. Automatic systems have the advantage of built-in automatic fitness evaluation, so the human effort is reduced; however, the aesthetics of the resulting artworks may suffer because the automatic evaluation has not been perfected yet. To overcome the disadvantages and also combine the advantages of both approaches, Machado et al. propose partially interactive evolution [26], where the human user's contribution is much reduced compared to the fully interactive approach, but the human still guides the evolution.

It is conceivable that a completely random walk in the space of all possible images will eventually

*Corresponding author. Email: a.ekart@aston.ac.uk

ISSN: 1751-3472 print/ISSN 1751-3480 online

DOI: 10.1080/1751347YYxxxxxxxx
http://www.informaworld.com

lead to visually appealing images, but this is unlikely due to the huge size of the space even in the case of restricted classes of images. The quest for visually appealing images really needs to be directed. Therefore, progress toward truly successful automatic evolutionary art can only be achieved if measures reflecting human aesthetic preference are available to drive the automatic evolution process. Intuitively, we define as *aesthetic* anything that is considered worth exhibiting in an established art gallery. Although Stiny and Gips argue that evaluations of art works by figures in the art world may be regarded as suspect because someone can simply declare themselves to be a member [35], nevertheless the measure of success we strive for, and our long-term goal, is exhibition is an established art gallery.

Studying, modelling and understanding user behaviour is also essential in order to produce tools that support or mimic human creative processes. The incorporation of user aesthetic preference to create intelligent graphic design tools is studied by Colton et al. [9] They find that when designing image filters, the quality of the designs as assessed by the users can be improved by intelligently employing the user's previous aesthetic preferences in the graphic design tool.

We propose to investigate the progression of human aesthetic preference during interactive evolution and then base automatic evolution on the aesthetic measures that were found to be followed by the human. We report our findings based on an extensive web-based pilot study, involving 137 users and nearly 500 sessions using the system. In this study we consider four established aesthetic measures to model human preference. We monitor how these measures might explain what is driving the selection process as users interact with the system over several generations. We identify the best pair of aesthetic measures for each user and employ them to automatically evolve further images that attempt to reflect the user's individual aesthetic preferences.

We found that although the progress of the automatic evolution process is based on the same aesthetic measures as those seemingly followed by the users, it could rarely produce images appreciated by the users. We consider this to be evidence for the need for more accurate and more theoretically grounded aesthetic measures better reflecting human user preference.

The paper is organised as follows. The pilot study is described in Section 2 and the technical details of the system are explained in Section 3. The selected aesthetic measures are described in Section 4. The proposed model underlying human selection is introduced in Section 5. The experimental results are presented in Section 6 and discussed in Section 7. Finally, conclusions are drawn in Section 8.

2. Design of experiment

We devised an evolutionary art pilot study using a software system we developed called **evo::art** in the hope of making progress toward better understanding human aesthetic preferences. The experiment was designed with four phases:

(1) interactive evolution
(2) automatic evolution
(3) rating the outcome of automatic evolution
(4) selecting the best image generated through interactive evolution.*

For reliability of findings, a large group of users would need to be included, therefore we created a web application. As we wanted to include users with the widest possible background, the interface had to be very simple and unambiguous and the user experience had to be consistent for all users, irrespective of their knowledge of evolutionary art. So that the different user sessions are comparable, we restricted the user actions to selection only. All evolutionary parameters were preset and were the same for all participants and all sessions.†

*This is deliberately the last phase: only users who progress through automatic evolution are allowed to submit their best image. Participants who abandon the experiment before automatic evolution can begin are considered to be unsatisfied with their progress, therefore it was concluded they did not really like their interactively generated images.

†Participants were allowed to perform as many sessions as they wished.

User sessions

One user session consisted of a compulsory interactive part followed by an optional automatic evolution. The interactive evolution would always start from a random set of 16 images and the user would have to select the two best images in each generation as parents for the next generation, thus driving the evolution toward images that they like. To ensure that no previous selection is lost, the two selected images were always transferred to the next generation. Also, to have extra variation, one image in the new generation was random. The remaining 13 images were generated through crossover and mutation applied on the two selected parent images. In order to both have a sufficient sample of progressive selections of a particular user and avoid the user becoming bored or disengaged, a minimum limit of 25 interactive generations before moving to automatic evolution was imposed. We found by initial experimentation that a lower number of generations may not lead to a reliable selection model. At the same time, some users would find the experiment repetitive and lose interest if there were too many generations to go through. The loss of interest could lead to early abandonment or random clicking through the generations without judging the images. At any stage of interactive evolution, if the user found that the process did not really go in the direction they wanted, they could restart it. After the minimum 25 interactive generations elapsed, the user could carry on with further interactive evolution for as long as they wanted, or proceed to automatic evolution.

When a user selected automatic evolution, **evo::art** would calculate the aesthetic measure pair best fitting the user's selections during interactive evolution and produce an image according to the measures identified using automatic evolution. Starting from a random set of images, 16 generations of populations were constructed in the same way as in interactive evolution, with the only difference that selection was performed automatically. The user then had to rate the highest rated automatically evolved image that appeared after 16 generations on a five-point Likert scale ranging from "I do not like the image at all" to "I like the image a lot". This allowed us to judge to what extent the automatic evolution models the particular user's preference.

At the end of the session the user had the opportunity to submit their best image from the interactive portion of their session for a competition. This allowed us to build up a set of images considered visually appealing by participants and analyse their features.*

Image properties

In earlier experiments [15] we used full colour images. We found that the selection would often be difficult for the human if the population contained images where they liked the colours but not the shapes and images where they liked the shapes but not the colours. Participants would either consciously concentrate on selecting shapes, disregarding colour, or oscillate between colour and shape preference in their choices. In fact, even deciding whether they liked the shapes in an image became more difficult if the colours were not to their liking.

Thus, concurrently evolving the colours as well as the shapes in an image may not lead to a better understanding of human preference. In this experiment, similarly to the co-evolutionary approach of [17], we only use greyscale images to help ensure all participants concentrate on shape preference. Adding colour through designing a colour scheme will be the subject of further work, possibly based on the work of Moretti [28].

As opposed to earlier experiments with square images, in the **evo::art** interface the images are framed into golden rectangles. The decision for golden rectangles is supported by the general belief that they are considered visually appealing [30].

3. The evolutionary art system

evo::art (available at `http://evoart.aston.ac.uk:8080/evoart/login.jsp`, accessed 23 February 2012) uses genetic programming (GP) for the evolution of visually appealing images. A

*The images are available upon email request to the corresponding author at a.ekart@aston.ac.uk.

clear distinction between the notions of genotype and phenotype is made. By genotype we mean an expression tree encoding a mathematical function and by phenotype the rendered image derived from the genotype. The genotype-phenotype mapping is a three-step procedure: (a) the genotype is fed with a custom vectorized input, (b) the genotype is processed to produce as output an array of real numbers and (c) the output is scaled, discretized and rearranged to form the phenotype – the intensity of the pixels in a 2D image. The representation and the mapping procedure are described below.

Genetic representation

Using a traditional tree-based representation and corresponding evaluation for **evo::art** is not feasible as the expression would need to be parsed and evaluated individually for each pixel of each image in each generation. Instead, we designed and implemented a more efficient representation using an ordinary integer array for the genotype and a genotype-phenotype mapping that takes vectorized input and produces vectorized output.

For performance reasons **evo::art** uses prefix notation to represent expression trees. Each genome is stored as an integer array of size 64. Each gene is a 32 bit integer encoding the type `T` of the gene (3 bits) and an index in an array corresponding to `T` (29 bits). Currently four types are defined: `var_t`, `uf_t`, `bf_t` and `const_t` corresponding to variables, unary and binary functions, and constants, respectively. The end of the genome is marked by a special gene, similar to the way the `\0` character marks the end of C-style strings. An example is given in Figure 1.

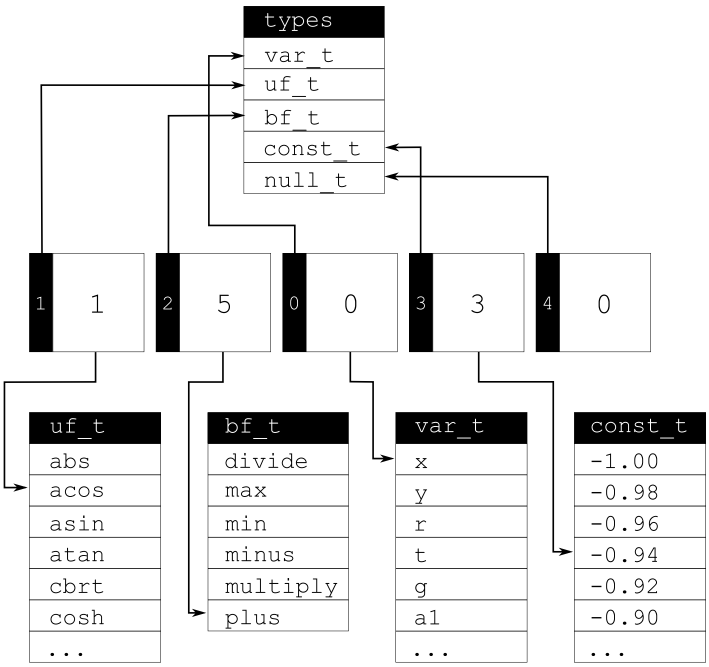
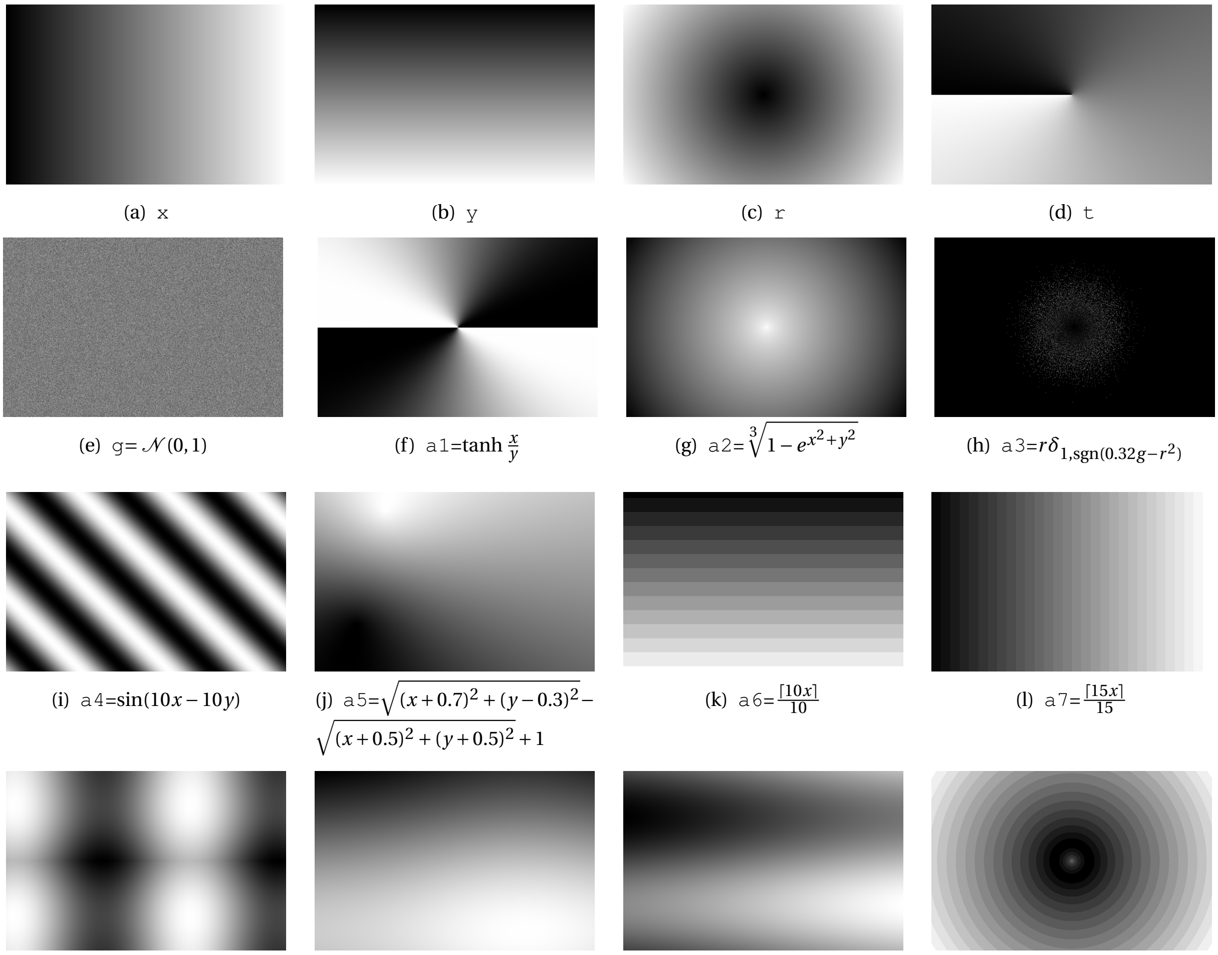


Figure 1. Internal representation of the genome `acos plus x -0.94`. The genome has four genes: the unary function `acos` followed by the binary function `plus` and its two arguments, the variable `x` and the constant `-0.94`.

In this study the array of constants consists of 101 equidistant values taken from the $[-1,1]$ interval. Besides `x`, `y`, `r` and `t` corresponding to the Cartesian and polar coordinates (introduced in [15]) we defined 12 additional variables here. The idea was to provide GP encapsulated expressions – manually defined functions (MDFs) – which are similar to automatically defined functions (ADFs), so that images are complex and/or potentially interesting starting from the first generation. The MDFs were selected based on their visual properties from preliminary experiments with **evo::art** . It was thought that allowing combinations of such MDFs will be more likely to lead to interesting images over a limited period of time. The images corresponding to the variables together with their mathematical expressions are shown in Figure 2.



(a) x (b) y (c) r (d) t

(e) g= $\mathcal{N}(0,1)$ (f) a1= $\tanh \frac{x}{y}$ (g) a2= $\sqrt[3]{1-e^{x^2+y^2}}$ (h) a3= $r\delta_{1,\mathrm{sgn}(0.32g-r^2)}$

(i) a4= $\sin(10x-10y)$ (j) a5= $\sqrt{(x+0.7)^2+(y-0.3)^2}-\sqrt{(x+0.5)^2+(y+0.5)^2}+1$ (k) a6= $\frac{\lceil 10x \rceil}{10}$ (l) a7= $\frac{\lceil 15x \rceil}{15}$

(m) a8= $\cos(5y)\log(|y|+0.5)+\sin(5x)$ (n) a9= $e^{-(x-0.5)^2}+5e^{-(y-0.5)^2}$ (o) a10= $\sin(x)+\sin(5y)$ (p) a11= $\lceil 10\left|\cos(r)-2\sqrt[3]{r}\right|\rceil/10$

Figure 2. The primitives used in **evo::art** .

evo::art employs the functions abs, acos, asin, atan, cbrt, cosh, cos, exp, log10, log, negate, sin, sinh, sqrt, tan, tanh, divide, max, min, minus, multiply, plus. As the colour resolution of the rendered images is eight bits, the precision of these functions is not crucial. Therefore, where it was reasonable the standard Java implementation was replaced with a faster version based on lookup tables. The speedup is up to one order of magnitude for certain functions (e.g. sin).

Genotype input and processing

For efficiency reasons **evo::art** genomes process vectorized input and consequently produce vectorial output. The input is given by the precalculated values of variables and constants. Precomputation of all the values needed for the 16 variables is based on the equidistant sampling of $\mathbf{D} \equiv [-1,1] \times [-0.64, 0.64]$, an origin-centered 2D domain whose width/height ratio approximates the golden ratio. The number of sampling points is given by the desired size of the image. For example, consider the image in Figure 2(a) which depicts variable x. If the target picture size is $w \times h$ pixels, then x is defined as a vector of reals whose elements are given by $x_k = 1.28\dfrac{k \mod w}{w-1} - 0.64$, $k = 0 \ldots wh-1$.

Vectorized input is fed into modified functions which process and produce arrays, (e.g. plus takes two arrays a and b and outputs c, where $c_i = a_i + b_i$, $i = 1 \ldots n$). This property opens the door for later parallelization. Genome evaluation follows the standard, stack-based algorithm [4]. As processing is vectorized, each genome is parsed only once (i.e. instead of parsing for each pixel of the image).

Discretizing and scaling the output

The output of genome processing is an array of reals. Before using it as an image, each element o of the array is linearly scaled and discretized to eight bits using $\left[255\frac{o-\lambda}{\Lambda-\lambda}\right]$, where λ and Λ are the minimum and maximum values of the output array, respectively.

Selection and fitness evaluation

During interactive evolution the user acts as evaluator by selecting the preferred images. Based on a set of aesthetic measures (discussed in Section 4), from the succession of the selected images a user profile is built. In the second part of the experiment an automatic evaluator is used. This automatic evaluator is built based on the previously created user profile (see Section 5).

4. Aesthetic measures

4.1. *State of the art in measuring aesthetics*

Measuring the aesthetic value of images is of interest in many fields of research. Aesthetics is considered one of the top ten unsolved problems of information visualisation [16]. It has been shown that the aesthetics of a user interface of a software system strongly influences its usability. In computer generated art and evolutionary art in particular it is crucial to have reliable aesthetic measures to drive the generation of visually appealing images. One possibility for measuring personal preference, suggested by Sims is based on the time that a person spends admiring a piece of artwork: presumably the more they like it, the longer they would admire it.

People have been interested in defining and measuring aesthetics for centuries. In the 18th century Hemsterhuis defined the “beautiful” as “that which gives the greatest number of ideas in the shortest space of time”[21]. This would be impossible to measure objectively and accurately. In 1933, Birkhoff observed how the perceived aesthetics of artworks is influenced by their order and complexity [2], providing the basis for many aesthetic measures.

The concept of algorithmic aesthetics was formalised by Stiny and Gips [35]. Because an aesthetic measure usually allows objects to be interpreted and evaluated in only one special way, a combination of such measures might seem promising. Inspired by formal theories of aesthetics, using what they called "exact aesthetics", Staudek and Machala invoked the rewriting rules from formal grammars to generate constructivist images [34]. Researchers have also considered the use of measures of unity, complexity, harmony, variety, entropy and redundancy lead to aesthetic evaluation of images. Kozárek used various machine learning methods in order to identify aesthetic characteristics for the restricted class of black and white planar periodic mosaics, where it was shown that based on a subset of such measures, image preference – human or otherwise – could be predicted with nearly 80% accuracy[22].

When designing an aesthetic measure, one must consider the aesthetic criteria to be satisfied. Greenfield describes the design of aesthetic measures as a two stage process: decision on statistics to be measured followed by decision on how to combine them into a measure [18]. Validation of any devised measure has to be done through human user acceptance.

Substantial efforts in evolutionary art research have been dedicated to studying and devising good aesthetic measures [12, 19, 24, 31, 32]. It is generally agreed that formulating a universally valid and acceptable aesthetic criterion is not within our reach. Achieving automatic evolution that produces images to the liking of the human user strongly depends on the understanding of the particular user’s aesthetic values. A recent study by Li and Hu [23] suggested using machine learning to learn the differences between visually appealing and non-appealing images, as indicated by image complexity and image order. At the same time, Colton [6] produces new rules for forming fitness functions through the use of an inference engine. As another approach, Greenfield proposes the technique of evolutionary refinement [20] to encourage aesthetic pattern formation through stages and concludes that “evolution in stages with radical changes in fitness criteria may be a profitable evolutionary exploration strategy”.

4.2. *Selected aesthetic measures*

We selected four well-known aesthetic measures to attempt modelling human preference. Measure R is based on Ross et al.[32], measure MC is based on Machado and Cardoso [24], measure BZ is based on Birkhoff and Zurek [2, 31] and finally measure S is based on Shannon entropy [31].

Aesthetic measure R

This aesthetic measure is based on the mathematical model proposed by Ross et al. [32] and adapted in a simple manner here for greyscale images. After analyzing hundreds of examples of fine art, it was found that many works consistently exhibit functions over colour gradients that conform to a normal or bell curve distribution.

The stimulus, or overall colour gradient, for each pixel is computed as:

$$S_{i,j} = \frac{\sqrt{(g_{i,j} - g_{i+1,j+1})^2 + (g_{i+1,j} - g_{i,j+1})^2}}{d}$$

where $g_{i,j}$ is the greyscale colour intensity of pixel (i, j) and $d = 0.181$ is a scaling factor. Next, the viewer's response (as detailed in [32]) to each pixel is computed as

$$R_{i,j} = \log(S_{i,j}/S_0).$$

S_0 is the detection threshold taken to be 2, as suggested by Ross et al. If $S_{i,j} = 0$ (no change in colour at a pixel), it is ignored. Afterwards, a histogram is built to approximate how close the distribution of response values is to the normal distribution, $\mathcal{N}(\mu,\sigma)$, where μ and σ denote the mean and standard deviation of $\{R_{i,j}\}$. The deviation (D) from the normal distribution is computed as:

$$D = \sum_i p_i \log\left(\frac{p_i}{q_i}\right)$$

where p_i is the observed probability in the i^{th} bin of the histogram and q_i is the expected probability assuming a normal distribution $\mathcal{N}(\mu,\sigma)$. Cases when $q_i = 0$ or $p_i = 0$ are ignored. The value

$$R = e^{-|D|}$$

will be reported as the value of the aesthetic measure.

We justify this aesthetic measure as follows:

(1) Aesthetic measure R discourages images which give rise to very high or very low response values. If a viewer gives very little response to a stimulus, it is too insignificant to be of interest. On the other hand, if a viewer gives a very large response to a stimulus, it is too disturbing or chaotic.
(2) The response value increases as the gradient increases and decreases as the gradient drops. Very low gradients give rise to single coloured monotonous areas (which do not interest a viewer) whereas very large gradients give rise to sharp lines and boundaries separating areas with huge colour differences (which is undesirable). Aesthetic measure R discourages very high and very low gradients and encourages reasonable values of gradients.

Aesthetic measure MC

This measure is based on the aesthetic theory of Machado and Cardoso [24] asserting that the aesthetic value of an artwork is directly correlated to Image Complexity (IC) and inversely correlated to Processing Complexity (PC). So, the value of the aesthetic measure is calculated as the ratio

$$\frac{IC}{PC}. \tag{1}$$

In order to compute IC, we first compress the image losslessly using JPEG compression and calculate the ratio (I) of the size of compressed image to the size of uncompressed image. We hypothesize that the IC is directly correlated to the ratio I. The inherent unpredictability, or randomness can be measured by the extent to which it is possible to compress the data [10]. Low values of I indicate a substantially compressible and low complexity image. High values of I indicate a not very compressible and therefore more complex image. That is,

$$\text{more compressible} \equiv \text{less random} \equiv \text{more predictable} \equiv \text{less complex}.$$

Hence, the less the value of ratio I (the less the size of the compressed file) is, the more compressible and hence, the less complex the image is. We substitute the ratio I for IC in Equation 1.

PC should reflect the complexity of the encoding of the image. We compress the prefix notation string representing an image (see Figure 1) and find the ratio P of the size after compression to the size before compression. We approximate PC by the ratio P. The aesthetic measure MC will be computed as

$$MC = \frac{I}{P}.$$

In theory the value of this aesthetic measure could range from zero to infinity, where infinity corresponds to an image that cannot be compressed, but whose corresponding prefix string can be compressed to the minimum. A value near zero corresponds to an image that can be compressed significantly, but with a prefix string that cannot be compressed. It is notable that the compression rate PC is applied on the actual evolved prefix strings, without any simplification.

Aesthetic measure BZ

This aesthetic measure is based on Birkhoff's measure [2] and Zurek's physical entropy [31]. We compute the value of Shannon's entropy as mentioned in [31] by creating a histogram of luminance values of pixels and computing Shannon's entropy H_p as follows:

$$H_p = -\sum_i p_i \log p_i$$

where p_i is the probability of being in the i^{th} bin of the histogram. The luminance value (L) for a pixel (i, j) will be its greyscale intensity $g_{i,j}$. Next, the Kolmogorov Complexity (K) [31] of the expression trees of the image is estimated by compressing the prefix strings and finding the length of the compressed string. The value of this aesthetic measure is given by

$$BZ = \frac{H_p}{K}.$$

This aesthetic measure discourages very high and very low luminance values because it favours high values of H_p. Very high and very low luminance values lead to low values of H_p. Here, K is used as a measure of PC.

Aesthetic measure S

As stated in [31], to analyse an image's composition, the used measures must quantify the degree of correlation or similarity between image parts. We first divide the image into four equal rectangles and compute Shannon's entropy (H_{p_i}, $i = 1, \ldots, 4$) for each of these parts. We then compute the weighted mean of these values (the weight being the area of the part). Finally, we find the ratio of the weighted mean to the Shannon entropy value of the image as a whole to obtain the value of the aesthetic measure as

$$S = \frac{H_{p_1} + H_{p_2} + H_{p_3} + H_{p_4}}{4H_p}.$$

5. Proposed model for aesthetic selection

Greenfield proposes multiobjective optimisation in order to better explore genotype space and allow for different aesthetic criteria to drive automatic evolution of artistic images [17]. We believe that better results, i.e. images that are better liked by a human user, can be expected if the multiobjective optimisation uses the objectives that the human follows in their selections.

Therefore, we are proposing a simple dynamic model of human preference based on the Pareto domination criterion and known aesthetic measures. Intuitively, the Pareto domination criterion states that a vector $\mathbf{x}$ dominates another vector $\mathbf{y}$ if $\mathbf{x}$ is better than or as good as $\mathbf{y}$ over all components and $\mathbf{x}$ is strictly better than $\mathbf{y}$ over at least one component. Usually better is quantified as higher value. We apply the criterion on pairs of aesthetic measure values of individuals in a population. For example, when considering the pair of aesthetic measures BZ-R, an individual with BZ value of 0.5 and R value of 0.9 dominates another individual with BZ value of 0.4 and R value of 0.9 (as $0.5 > 0.4$ and $0.9 = 0.9$), but does not dominate a third individual with BZ value of 0.4 and R value of 1 (as $0.5 > 0.4$, but $0.9 \ngeq 1$). On its turn, the third individual does not dominate the first individual either (as $0.4 \ngeq 0.5$).

We build the user profile consisting of the most frequently followed selection pattern of aesthetic measures as follows. We consider that in a generation the user follows a particular pair of aesthetic measures if the newly selected images are not dominated by the images selected in the previous generation. We consider all pairs of aesthetic measures BZ-MC, BZ-R, BZ-S, MC-R, MC-S, R-S. We record whether the values of measures (k,l) on the interactively selected individuals in each generation were dominated by the values of the same measures on the selected individuals in the previous generation. We establish which pair of measures is most often followed by the human user by finding the least dominated pair of measures (k,l) out of all combinations BZ-MC, BZ-R, BZ-S, MC-R, MC-S, R-S during interactive evolution. The `calc_domination_value(k,l)` algorithm simply counts for each pair of measures the number of interactive generations in which the selected individuals are not dominated by the previously selected individuals (see Figure 3). The pair with the highest non-domination count is selected, so that when automatic evolution begins, after each generation only these two measures are used to determine the two images that will be selected to populate the next generation. Thus, during the automated evolution phase, we approximate a multiobjective scheme with these two measures as the two objectives. This model could be refined further by considering different stages of evolution, accounting for less visually appealing images in early stages and user fatigue in final stages.

```
 Let i be the generation number
 Let g be the maximum number of generations
 Let individual(i,j) be selected individual number j in generation i (j=1,2)
 Let k and l be a pair of measures used in interactive evolution

function dv = calc_domination_value(k,l)
    dv = 0;
    for i = 1:g
       for j = 1:2
            if individual(i,j) is not dominated by
               individual(i-1,1) or individual(i-1,2) over measures k and l
                   then dv = dv+1;

    return dv;
```

Figure 3. The pseudocode of the least domination algorithm.

Our model can cater to the tastes of individual users by selecting different aesthetic measure pairs.

The current experiment is limited by the particular aesthetic measures used and the choice of using only two measures rather than a variable size subset. As new aesthetic measures become available, they can be added to the set of measures being monitored and then automatic evolution will subsequently be provided with the best matching subset of aesthetic measures. We expect that *if the employed aesthetic measures truly reflected human preference, automatic evolution would progress in a desirable direction for the particular human user.*

6. Experimental results

137 participants (95 male and 42 female) took part in the **evo::art** pilot study, totalling 498 sessions. The age varied between 19 and 59 years, with 78% of users below the age of 35. The majority of users performed one or two sessions. There were four users who performed a substantial number of sessions (39, 53, 69 and 86, respectively). The histogram showing the number of sessions performed by different users is presented in Figure 4(a). The total number of generations performed by users is shown in Figure 4(b). The majority of users performed a small number of generations and four users performed an extremely large number of generations. The number of sessions and the number of generations performed by a user are very strongly correlated, with a coefficient of 0.98. The histogram in Figure 4(c) indicates that the average number of generations per session was usually around 25, the minimum number required to proceed to automatic evolution. Lower values correspond to abandoned or restarted sessions. The time spent between two image selections shown in Figure 4(d) is usually below 50s, with few exceptions. The longer times are thought to correspond to situations where the user did indeed spend more time judging the images or to situations where they took a break (away from the computer) during the interactive evolution process. There is no correlation between the average time spent between two image selections and the number of sessions or total number of generations.

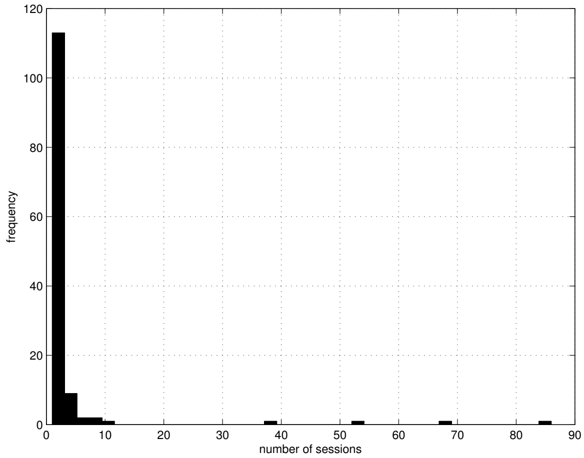
Some of the images users selected as their best ones during their interactive sessions, and thus entered into our competition, are presented in Figure 5. The winning image of the competition, judged by professional art critic Júlia Németh is shown in Figure 6. The image was described as “approaching the limits of figurative, a representative composition. Harmonious play with shapes, light effects, mutually complementary, supportive and coordinating formal elements, suggesting strong spatial effects. Indisputable compositional virtues.” (Júlia Németh)

Our rationale for using partial automatic evolution [26, 36] is that we hypothesize that by analyzing the human’s previous selections and modelling them by the best fitting set of measures, we can allow automatic evolution to imitate human selection. We believe that this increases the chances that images to the particular human user’s liking are produced by automatic evolution.

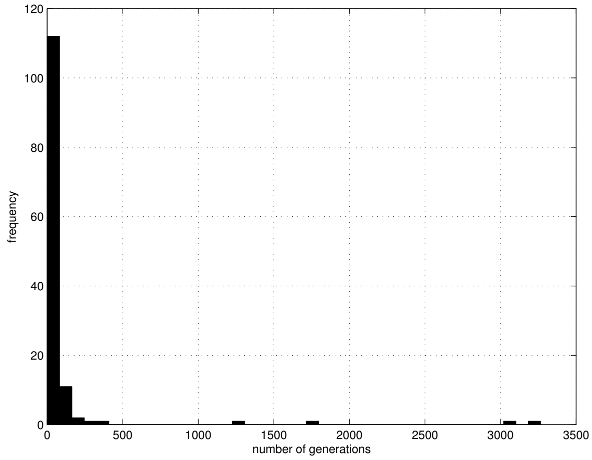
In all sessions, the users rated the outcome of automatic evolution on a Likert scale of five values corresponding to the range from “I do not like it at all” to “I like it a lot” and could also provide a free-text comment. The ratings of the 498 sessions are shown in Figure 7. Very few people were really satisfied with the outcome of automatic evolution. Quite often, the images generated by automatic evolution were perceived by the users as very different from what they were trying to achieve, or visually different from their interactively evolved image. This was reflected in the free-text comments.

We selected for further analysis two sessions where the human users were happy with the image created by automatic evolution (A and B) and two sessions where they were not (C and D). These individual sessions were analyzed to better understand the driving force of interactive evolution (human selection), and to see whether subsequent automatic evolution followed similar principles. In Table 1 we present the non-domination values calculated for the selected sessions. A non-domination value entry in the table indicates in how many generations the selected images were not dominated by the previously selected ones, according to the two aesthetic measures being monitored. It can be seen that there is no clear winner for user A, the pair BZ-R is preferred by users B and C, and the pair BZ-MC is preferred by user D. The winning pair of measures is followed by the users at least 79.17% of the time.

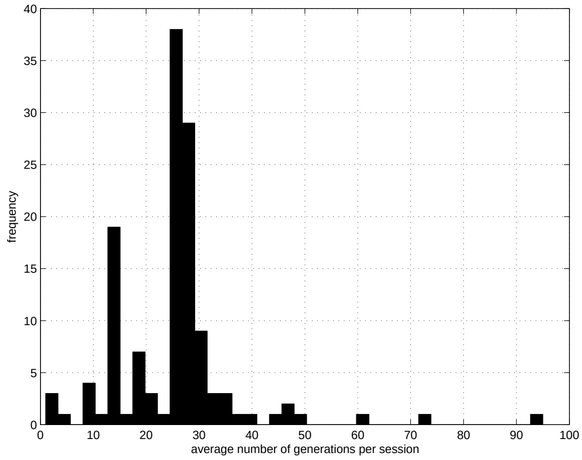
In Figures 8–11 we compare interactive evolution and automatic evolution in the four selected ses-



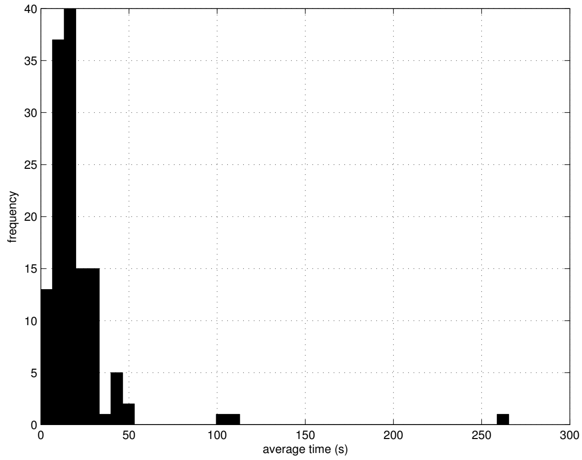
(a) Number of sessions performed by the users.



(b) Total number of generations performed by the users.



(c) Average generations/session performed by the users.



(d) Average time spent between two image selections.

Figure 4. Characteristics of **evo::art** usage by pilot study participants.

sions. We present the resulting images together with the evolution over time of the four aesthetic measures leading to them. In the first two cases, by simple visual inspection we can find traces of features in the interactively generated image also present in the automatically evolved image. In the latter two cases this is not possible.

For user B, all four measures follow similar trends in interactive and automatic evolution:

- BZ has a peak at the beginning and then oscillates over a smaller range
- MC oscillates mostly between 0.1 and 0.5, but slowly increases
- R oscillates around a smaller range and then has a peak
- S varies between 0.75 and 1, increasing.

For user A such close similarities in the progress of all individual measures are not visible. However, the ranges for the aesthetic measures are very similar and the best images for interactive and automatic evolution have similar aesthetic values. For user C there is a high peak in measure BZ during

Table 1. Non-domination values during interactive selection for selected sessions. Winning pairs highlighted. Users A and B liked the automatically evolved image, whereas users C and D did not.

User	BZ-MC	BZ-R	BZ-S	MC-R	MC -S	R-S	No. of generations
A	19	**21**	18	19	19	**21**	26
B	27	**30**	27	26	24	29	35
C	38	**39**	36	31	30	33	48
D	**24**	20	20	16	18	18	29

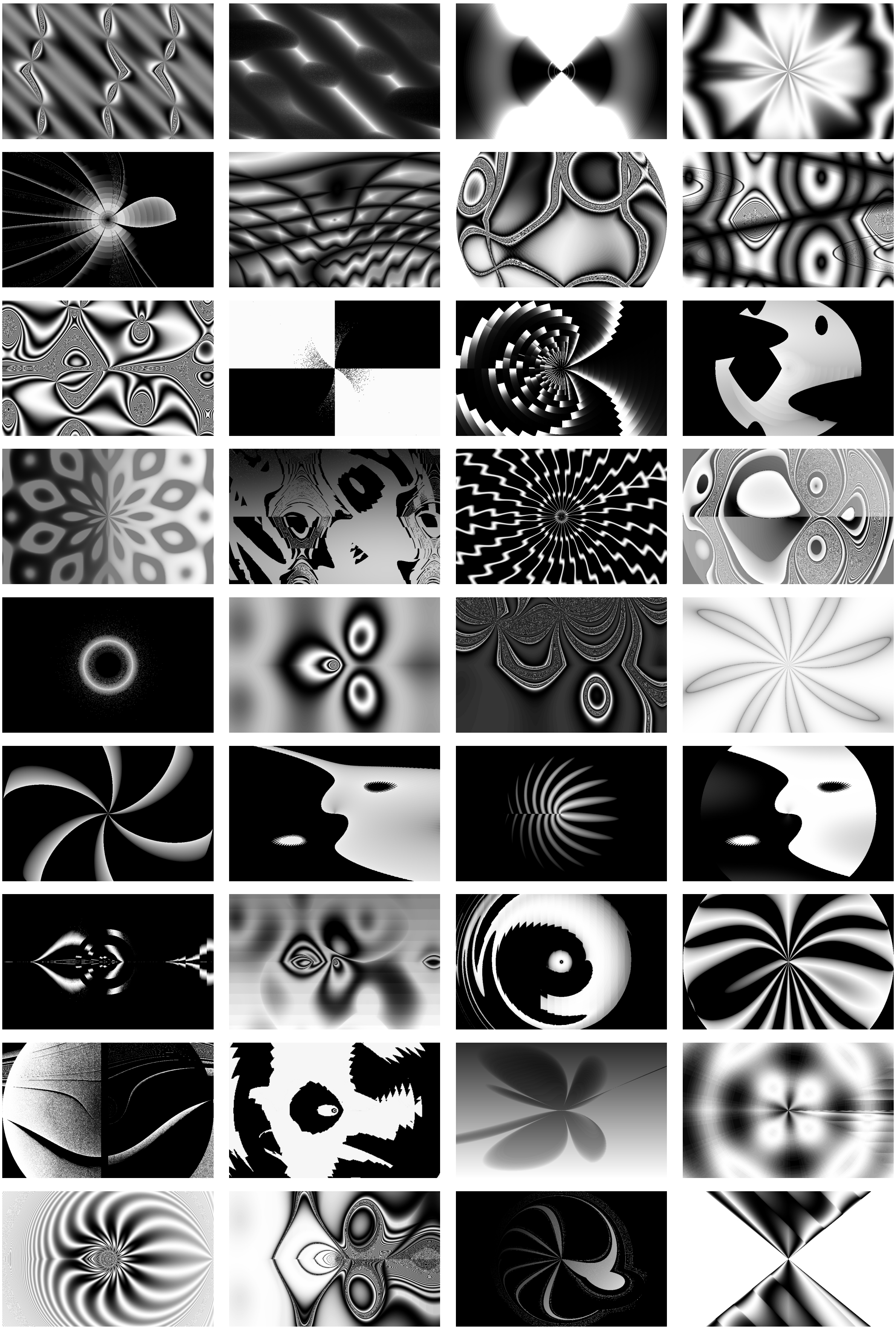


Figure 5. Example images generated by interactive evolution.

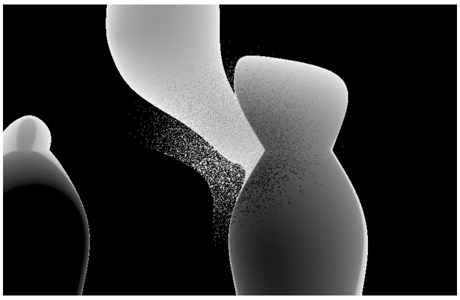


Figure 6. Winning competition image created by Jort van Mourik.

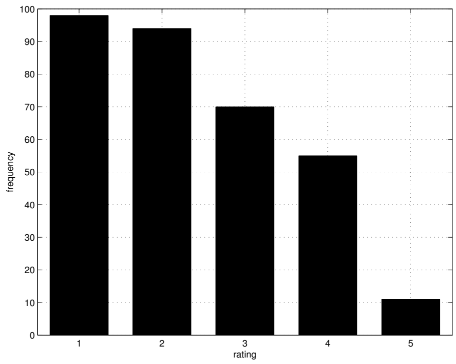


Figure 7. The user rating of images produced by automatic evolution. A higher rating corresponds to better preference.

interactive evolution, which is not repeated in automatic evolution. Measure R oscillates between low and high values in both phases of the experiment. According to the values of measures BZ and R, the automatically evolved image dominates the interactively evolved image. We consider this evidence for the fact that the aesthetic measures used can not always represent user intention. On the other hand, for user D, the range of values for measure BZ during automatic evolution is smaller than for interactive evolution. The behaviour of measure MC is very similar during interactive and automatic evolution. According to the values of measures BZ and MC, the interactively evolved image dominates the automatically evolved image. This captures the fact that user D prefers the interactively evolved image to the automatically evolved image.

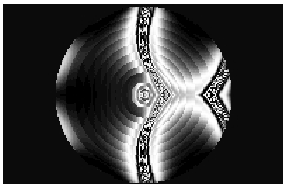
7. Discussion

At first, we found it quite puzzling that the vast majority of the users were not satisfied with the automatically generated image. We attribute this to two reasons:

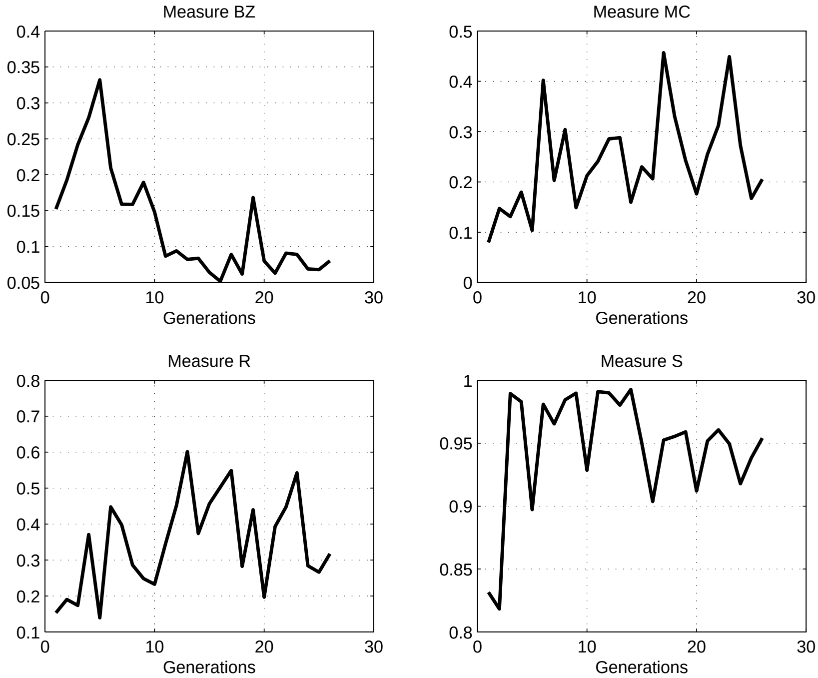
(1) Gaussian noise appeared too often in the automatically evolved images;
(2) the aesthetic measures used do not fully describe the particular user's selections.

We investigated whether there are any particular genes that are preferred more than others and whether there are any major differences in gene usage between interactively and automatically evolved images. We analysed the gene frequency in the 498 images submitted for the competition* as well as in the subsequently automatically evolved images. The number of occurrences of the 40 most frequent genes are shown in Figure 12 for both interactive and automatic evolution. The frequencies are ordered in decreasing order for interactive evolution. Although the values of the frequencies are somewhat different, the trend is similar for interactive and automatic evolution: a substantial drop is experienced after the first 5 or 6 genes, which is then followed by gradual reduction to values below 10. There is remarkable similarity between the sets of most frequent genes, `plus, minus,`

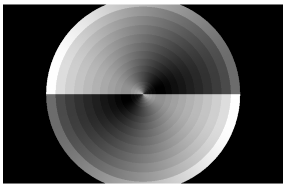
*If someone submitted an image for the competition, that was seen as an indication that they considered that image visually appealing.



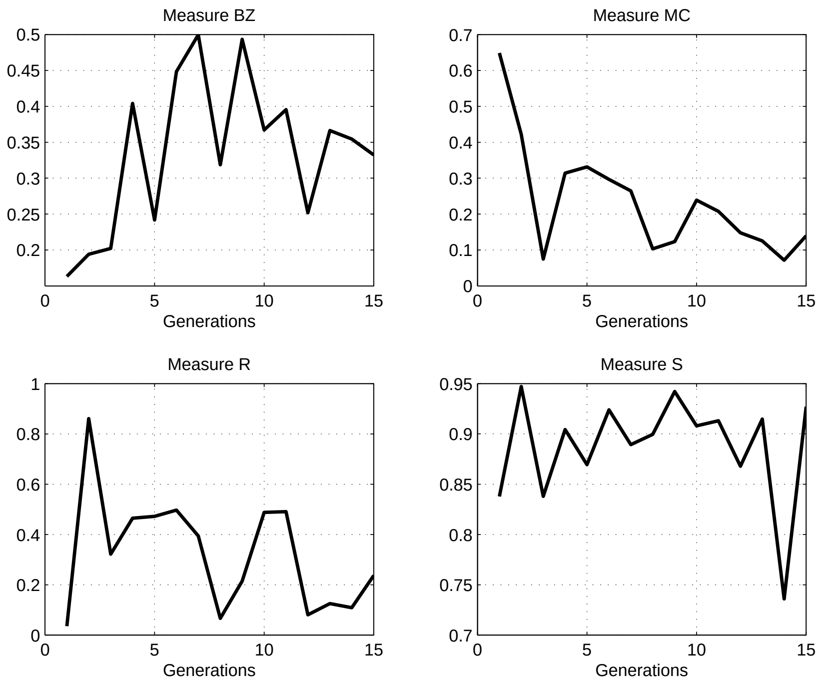
(a) Interactive evolution best image



(b) Progress of aesthetic measures during interactive evolution

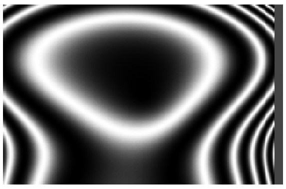


(c) Automatic evolution best image

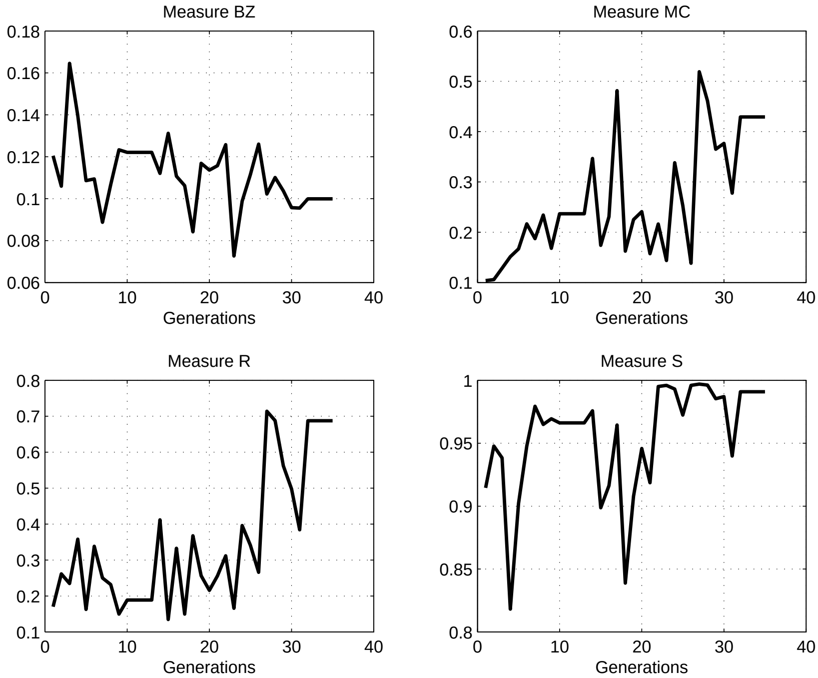


(d) Progress of aesthetic measures during automatic evolution

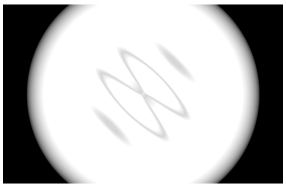
Figure 8. User A's image created through interactive evolution and the subsequently created image by automatic evolution. The pairs BZ-R and R-S were followed equally in interactive evolution. User A liked the automatically evolved image.



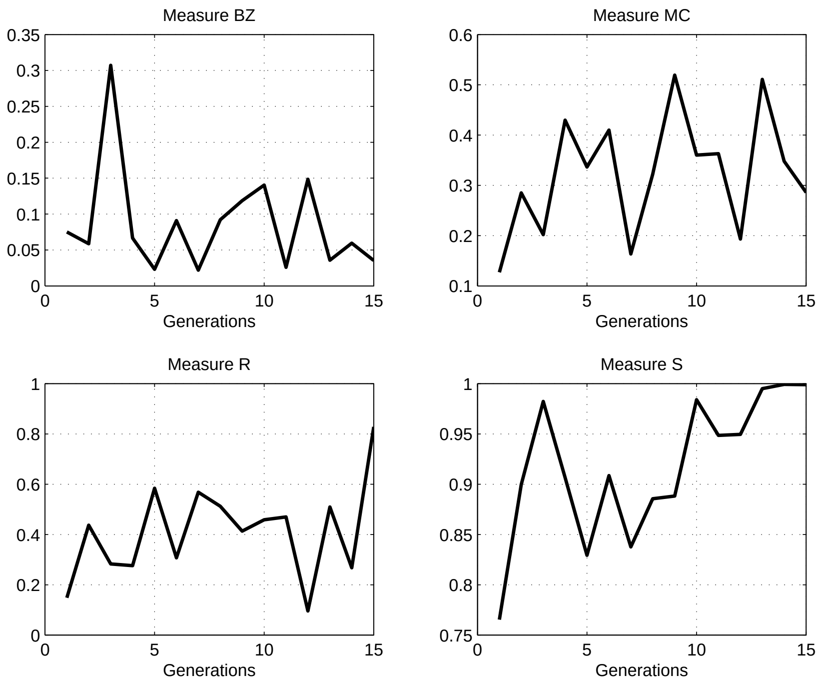
(a) Interactive evolution best image



(b) Progress of aesthetic measures during interactive evolution

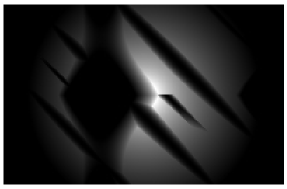


(c) Automatic evolution best image

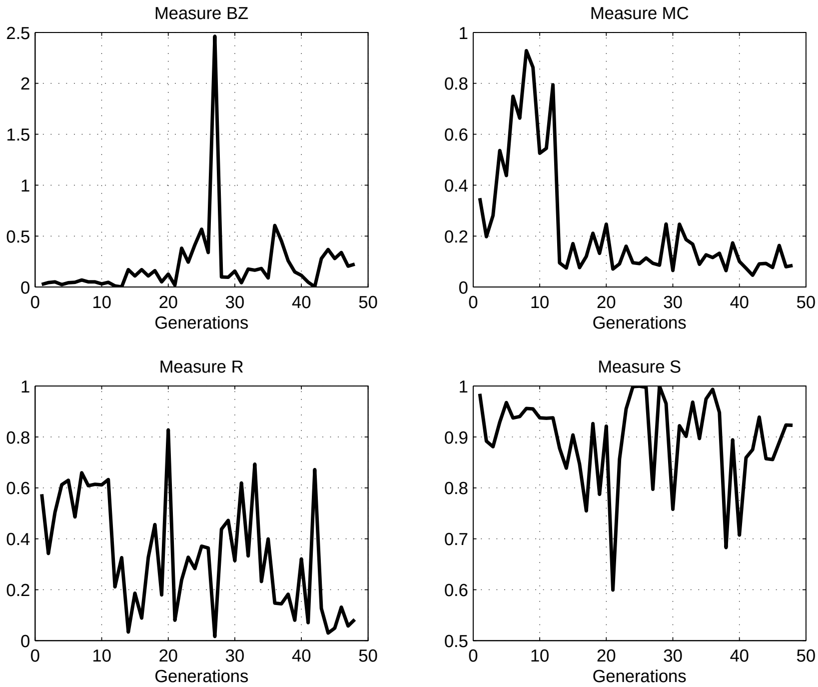


(d) Progress of aesthetic measures during automatic evolution

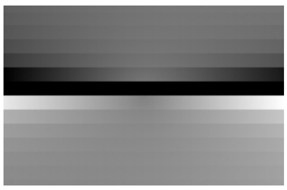
Figure 9. User B's image created through interactive evolution and the subsequently created image by automatic evolution. The pair BZ-R was followed in interactive evolution. User B liked the automatically evolved image.



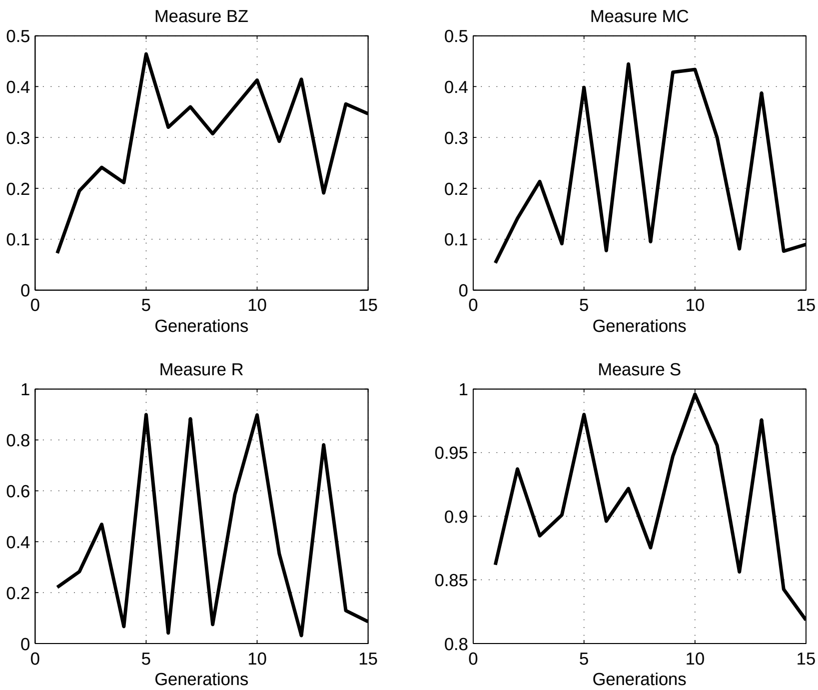
(a) Interactive evolution best image



(b) Progress of aesthetic measures during interactive evolution

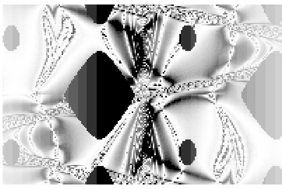


(c) Automatic evolution best image

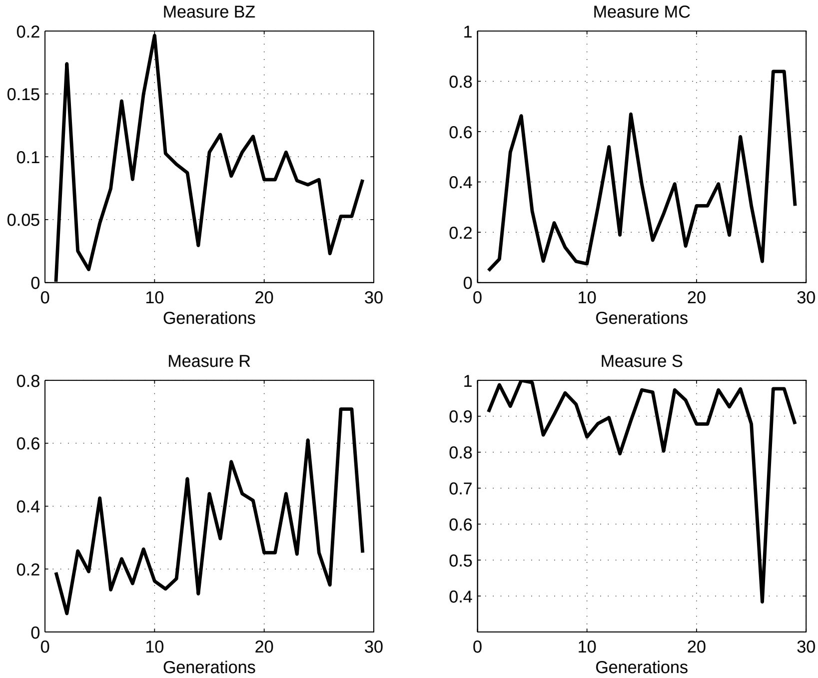


(d) Progress of aesthetic measures during automatic evolution

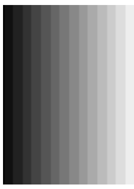
Figure 10. User C's image created through interactive evolution and the subsequently created image by automatic evolution. The pair BZ-R was followed in interactive evolution. User C did not like the automatically evolved image.



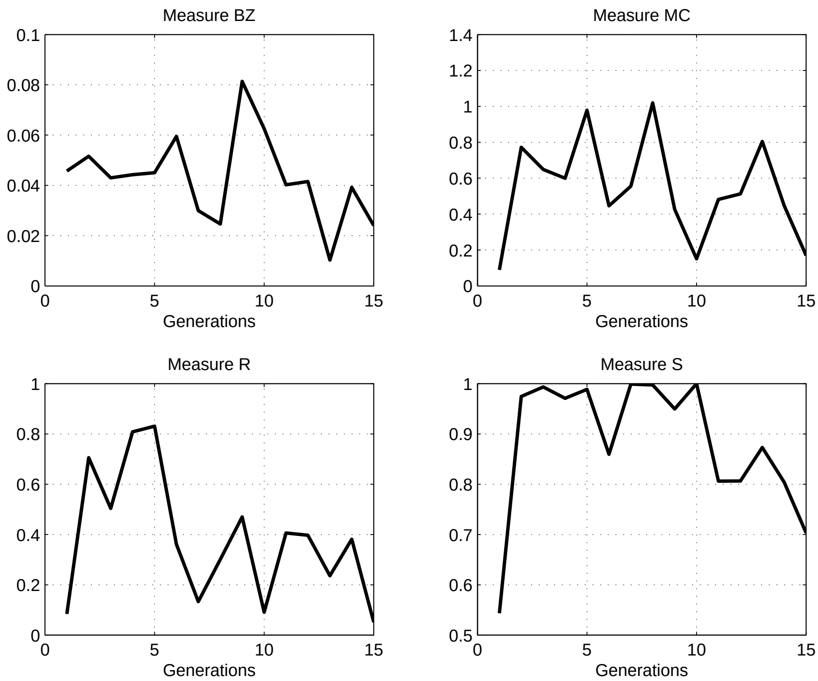
(a) Interactive evolution best image



(b) Progress of aesthetic measures during interactive evolution



(c) Automatic evolution best image



(d) Progress of aesthetic measures during automatic evolution

Figure 11. User D's image created through interactive evolution and the subsequently created image by automatic evolution. The pair BZ-MC was followed in interactive evolution. User D did not like the automatically evolved image.

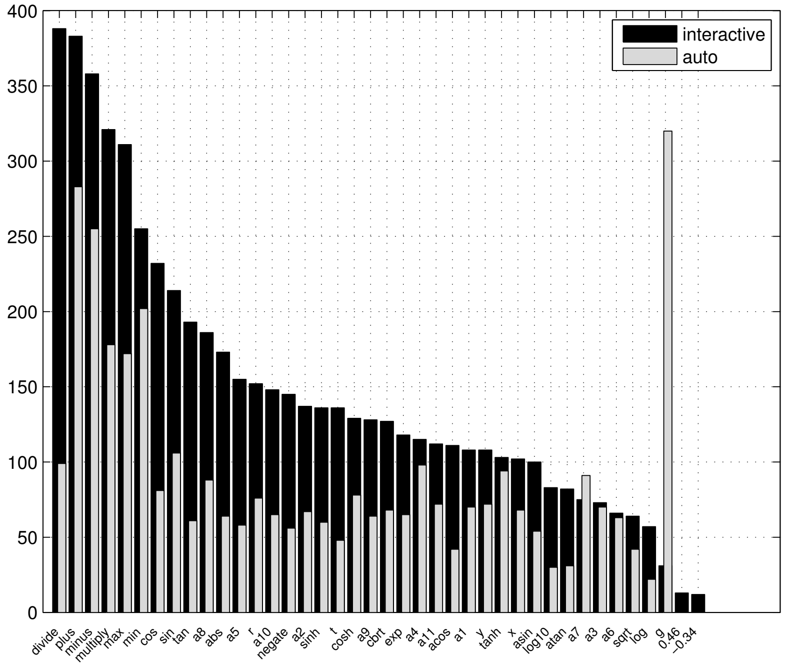


Figure 12. Occurrences of the most frequent genes during interactive and automatic evolution.

`multiply`, `max`, `min` are five out of the six most frequent genes for both interactive and automatic evolution. The major difference is in the most frequent gene: `divide` for interactive evolution and `g` for automatic evolution. More precisely, `g` only shows in 40 interactively evolved images, but occurs 320 times in the automatically generated images. In fact `g` dominates 64.5% of the automatically evolved images not liked by the users.

The values of the four aesthetic measures for all 16 primitives shown in Figure 2 are presented in Table 2. It can be seen that `g` has very high values according to all the measures used, in fact the highest of all on measures S, R and MC and fourth largest value on measure BZ. Therefore, it is highly likely that in automatic evolution, whenever present, `g` is non-dominated by other genotypes, so can become the best solution. At the same time, human users could ignore images containing `g` and still be perceived as following a consistent selection pattern.

We conducted another smaller scale experiment where `g` was removed from the set of variables. From 24 completed sessions, we found that the user rating of the automatically evolved image followed a bell curve distribution, with one user not liking the image at all, 5 users not liking the image much, 10 users giving a neutral rating, 7 users somewhat liking the image and one user liking the image a lot. This confirms our expectation that the very large proportion of users in the larger exper-

Table 2. Aesthetic measure values for each of the primitives. Higher values are expected for visually more appealing images.

Measure/Variable	x	y	r	t	g	a1	a2	a3
S	0.87	0.86	1.00	0.75	1.00	0.85	1.00	0.99
R	0.01	0.03	0.05	0.10	0.90	0.20	0.09	0.41
MC	0.06	0.08	0.08	0.06	0.59	0.07	0.07	0.37
BZ	7.64	7.00	7.74	7.88	7.07	3.30	3.79	0.85

Measure/Variable	a4	a5	a6	a7	a8	a9	a10	a11
S	0.89	0.85	0.73	0.80	0.99	0.86	0.86	1.00
R	0.13	0.05	0.02	0.02	0.15	0.03	0.05	0.14
MC	0.16	0.07	0.10	0.09	0.09	0.07	0.08	0.13
BZ	3.72	3.93	1.88	2.46	3.93	3.91	2.63	1.31

iment not liking the image were mostly caused by the occurrence of `g`.

A different possibility for rectifying this deficiency is to try and guarantee a better spread of variables in automatic evolution that follows the spread of variables experienced during interactive evolution. This could be achieved by enriching the user profile with a set of highly preferred primitives, for example, and forcing automatic evolution to include similar proportions of such primitives in the genotypes. The exploration of this aspect in a similar web-based experiment can form the basis of future research.

Another potential difficulty in modelling selection is that users may change their mind during sessions. We examined the frequency of the followed selection pattern defined according to a pair of aesthetic measures followed by a user in a selection step. We found that the users followed their preferred pattern over 79% of the time, which we considered a consistent pattern. The users were obliged to complete only 25 interactive evolution steps and very few users completed more than 40 steps (see Figure 4(c)). We believe that longer observations are needed to observe clear changes in selection patterns. Also, more sophisticated aesthetic measures would be needed to better capture the visually noticeable changes in selection patterns that are unnoticed by the current aesthetic measures.

Automatic multiobjective optimisation has been previously found not to consistently lead to visually appealing images [17]. In this context, Greenfield proposes the meta-rule "Design criteria for enhancing creativity must be compatible." We'd like to add that what seems compatible for certain users, may not be so for all users.

Finally, it is worth mentioning that when the aesthetic measures did capture the user's intention, as illustrated by Figures 8 and 9, the result of automatic evolution was appreciated by the user.

8. Conclusion

We proposed modelling human user preference by a set of aesthetic measures monitored through observation of human selection in an interactive evolutionary art system. Although our evolutionary art system is very simple and is only capable of generating images with obvious visual limitations, it provides a suitable environment for studying human aesthetic judgment. The same principles could be applied using an extended set of aesthetic measures on more sophisticated evolutionary art systems in which case different subsets of aesthetic measures may be found to best model individual users.

McCormack [27] criticises aesthetic selection itself and proposes as an open problem "to devise formalized fitness functions that are capable of measuring human aesthetic properties of phenotypes". We feel that an important step toward solving McCormack's open problem is to *model and measure* human aesthetic properties by the available means.

We believe that there is a need for more theoretically grounded aesthetic measures. We find it difficult to believe that one aesthetic measure fits all. Therefore we propose to concentrate on identifying user dependent aesthetic measures. We argue that once a combination of measures that model increasing human preference during interactive evolution is identified, automatic evolution can be provided with a suitable fitness evaluation method.

We attempt to model human preference by recording and analysing actual human selections, without considering the processes leading to the particular selections. We use a simple model based on a combination of known aesthetic measures. We demonstrate that such simple models coupled with general aesthetic measures rarely lead to automatic evolution satisfying the visual expectations of humans. The next step will be to build a more accurate model for human selection. We are formulating two approaches as follows. The first approach will employ machine learning techniques to find potentially more accurate functions driving human aesthetic judgment and to subsequently apply these functions for evaluation and selection in automatic evolution. The second approach will build a probabilistic Bayes model for the selections performed by each user to be then used in automatic evolution. Additionally, studies such as ARtSENSE (available at `http://www.artsense.eu/project/project-overview/` accessed 13 February 2012) can provide more fine grained measurements of intermediary states leading to the particular selections which could be exploited to fine-tune the probabilistic models.

To make evo::art more capable and unbiased and also to increase the power of the user to guide the evolution, we are considering extending the user interface to allow the user to create their own MDFs and maintain a pool of interesting genomes which could be reused in future images. We expect to improve the user experience by introducing this facility at the price of increasing the difficulty of model construction.

Acknowledgments

The authors would like to express their gratitude to the special issue editors Gary Greenfield and Penousal Machado for their encouragement and support. The authors are grateful to the numerous enthusiastic participants in the pilot study. The help of art critic Júlia Németh and digital artist Mira Marincas in judging the submitted images is much appreciated. The authors acknowledge the support of Aston University's summer internship scheme which made it possible for Divya Sharma to take part in the reported work.

References

[1] S. Baluja, D. Pomerleau, and T. Jochem, *Towards Automated Artificial Evolution for Computer-generated Images*, Connection Science 6 (1994), pp. 325–354.
[2] G.D. Birkhoff, *Aesthetic measure*, Harvard University Press 1933.
[3] M.A. Boden and E.A. Edmonds, *What is generative art?*, Digital Creativity 20 (2009), pp. 21–46.
[4] A.W. Burks, D.W. Warren, and J.B. Wright, *An analysis of a logical machine using parenthesis-free notation*, Mathematical Tables and Other Aids to Computation 8 (1954), pp. 53–57.
[5] S. Colton, *The Painting Fool*, www.thepaintingfool.com (2007).
[6] S. Colton, *Automatic Invention of Fitness Functions with Applications to Scene Generation*, in *EvoApplications 2008*, LNCS 4974, Springer, 2008, pp. 381–391.
[7] S. Colton, M. Cook, and A. Raad, *Ludic Considerations of Tablet-Based Evo-Art*, in *Applications of Evolutionary Computation: EvoApplications 2011: Evolutionary Music and Art*, LNCS 6625, Springer, 2011.
[8] S. Colton, M. Valstar, and M. Pantic, *Emotionally Aware Automated Portrait Painting*, in *Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts (DIMEA)*, 2008.
[9] S. Colton, J. Gow, P. Torres, and P. Cairns, *Experiments in Objet Trouvé Browsing*, in *Conference on Computational Creativity*, 2010, pp. 238–247.
[10] S. Dasgupta, C. Papadimitriou, and U. Vazirani, *Algorithms*, Tata McGraw-Hill Publishing Company Limited 2008.
[11] R. Dawkins, *The Blind Watchmaker*, Longman 1986.
[12] E. den Heier and A.E. Eiben, *Comparing Aesthetic Measures for Evolutionary Art*, in *EvoApplications 2010, Part II*, LNCS 6025, Springer, 2010, pp. 311–320.
[13] S. DiPaola and L. Gabora, *Incorporating Characteristics of Human Creativity into an Evolutionary Art System*, Genetic Programming and Evolvable Machines 2 (2009).
[14] S. Draves, *Electric Sheep*, electricsheep.org (1999).
[15] A. Ekárt, D. Sharma, and S. Chalakov, *Modelling human preference in evolutionary art*, in *Applications of Evolutionary Computation: EvoApplications 2011: Evolutionary Music and Art*, LNCS 6625, Springer, 2011, pp. 303–312.
[16] D. Filonik and D. Baur, *Measuring aesthetics for information visualization*, in *13th International Conference Information Visualisation*, 2009, pp. 579–584.
[17] G.R. Greenfield, *Computational aesthetics as a tool for creativity*, in *Creativity and Cognition*, 2005, pp. 232–235.
[18] G.R. Greenfield, *Designing Metrics for the Purpose of Aesthetically Evaluating*, in *Computational Aesthetics*, 2005, pp. 151–158.
[19] G.R. Greenfield, *On the Origins of the Term 'Computational Aesthetics'*, in *Computational Aesthetics*, 2005, pp. 9–12.
[20] G.R. Greenfield, *Generative Art and Evolutionary Refinement*, in *EvoApplications 2010, Part II*, LNCS 6025, Springer, 2010, pp. 291–300.
[21] F. Hemsterhuis, *Lettre sur la sculpture*, Quoted in Encyclopedia Britannica,vol. 11 (1946) (1769).
[22] P. Kozárek, *Aesthetic preference research for planar periodic mosaics*, Diploma thesis, Masaryk University (2006).
[23] Y. Li and C.J. Hu, *Aesthetic Learning in an Interctive Evolutionary Art System*, in *EvoApplications 2010, Part II*, LNCS 6025, Springer, 2010, pp. 301–310.
[24] P. Machado and A. Cardoso, *Computing Aesthetics*, in *SBIA'98*, LNCS 1515, Springer, 1998, pp. 219–228.
[25] P. Machado and A. Cardoso, *All the truth about NEvAr*, Applied Intelligence, Special issue on Creative Systems 16 (2002), pp. 101–119.
[26] P. Machado, J. Romero, A. Cardoso, and A. Santos, *Partially interactive evolutionary artists*, New Generation Computing 23 (2005), pp. 143–155.
[27] J. McCormack, *Open Problems in Evolutionary Music and Art*, in *EvoWorkshops 2005*, LNCS 3449, Springer, 2005, pp. 428–436.
[28] G.S. Moretti, *A calculation of colours: towards the automatic creation of graphical user interface colour schemes*, PhD dissertation, Massey University, New Zealand (2010).
[29] A.M. Noll, *The beginnings of computer art in the United States: A memoir*, Leonardo 27 (1994), pp. 39–44.
[30] L. Pacioli, *De divina proportione*, Venice 1509.
[31] J. Rigau, M. Feixas, and M. Sbert, *Informational Aesthetics Measures*, IEEE Computer Graphics and Applications (2008), pp. 24–34.
[32] B.J. Ross, W. Ralph, and H. Zong, *Evolutionary Image Synthesis Using a Model of Aesthetics*, in *IEEE Congress on Evolutionary Computation*, 2006, pp. 1087–1094.
[33] K. Sims, *Artificial Evolution for Computer Graphics*, in *SIGGRAPH 1991*, Vol. 25, 1991, pp. 318–328.

[34] T. Staudek and P. Machala, *Recent Exact Aesthetics Applications*, in *SIGGRAPH*, 2002, p. 241.

[35] G. Stiny and J. Gips, *Algorithmic aesthetics: Computer models for criticism and design in the arts*, University of California Press 1978.

[36] S. Todd and W. Latham, *Evolutionary art and computers*, Academic Press 1992.

[37] J.J. Ventrella, *Evolving the Mandelbrot set to imitate figurative art*, in *Design by Evolution*, P.F. Hingston, C.L. Barone, and Z. Michalewicz, eds., 2008, pp. 145–167.

[38] S. Wannarumon, E.L.J. Bohez, and K. Annanon, *Aesthetic Evolutionary algorithm for Fractal-based User-centered Jewelry Design*, Artificial Intelligence for Engineering Design, Analysis and Manufacturing 22 (2008), pp. 19–39.