Comparing aesthetic measures for evolutionary art

E. den Heijer[1,2] and A.E. Eiben[2]

[1]Objectivation B.V., Amsterdam, The Netherlands
[2]Vrije Universiteit Amsterdam, The Netherlands
eelco@few.vu.nl, gusz@cs.vu.nl

Abstract. In this paper we investigate and compare four aesthetic measures within the context of evolutionary art. We evolve visual art with an unsupervised evolutionary art system using genetic programming and an aesthetic measure as the fitness function. We perform multiple experiments with different aesthetic measures and examine their influence on the evolved images. To this end we store the 5 fittest individuals of each run and hand-pick the best 9 images after finishing the whole series. This way we create a portfolio of evolved art for each aesthetic measure for visual inspection. Additionally, we perform a cross-evaluation by calculating the aesthetic value of images evolved by measure i according to measure j. This way we investigate the flexiblity of each aesthetic measure (i.e., whether the aesthetic measure appreciates different types of images). The results show that aesthetic measures have a rather clear "style" and that these styles can be very different. Furthermore we find that some aesthetic measures show very little flexibility and appreciate only a limited set of images.

1 Introduction

The goal of the research field of Computational Aesthetics is to investigate "computational methods that can make applicable aesthetic decisions in a similar fashion as humans can" [5]. Aesthetic measures are functions that compute the aesthetic value of an object. [2] was the first to publish on the subject of aesthetic measures, and his work has been influential in the field. Birkhoffs notion of aesthetics was based on the relation between Order and Complexity, expressed as $M = \frac{O}{C}$, where O stands for order and C for Complexity. Birkhoffs measure is now widely regarded a being mostly a measure of orderliness. Since Birkhoff, several researchers have investigated aesthetic measures from several points of view. [4] and [5] give good overviews of the field.

1.1 Research question

In this paper we investigate and compare four aesthetic measures. Each aesthetic measure is used in an evolutionary art system as a fitness function (all evolutionary parameters are kept equal for all aesthetic measures). We evolve small

Lisp like expressions that generate images, and compare the difference between the images created by the four aesthetic measures. Next, we investigate how the produced images using aesthetic measure M_N are judged by the other aesthetic measures. Hereby we obtain an indication of the neutrality of the measure.
The rest of the paper is structured as follows. First we discuss evolutionary art and the use of aesthetic measures within the context of evolutionary art (section 2). Section 3 discusses our software environment Arabitat. Next, we describe the experiments and their results in section 4.1. In section 4.2 we calculate the cross evaluation of the four aesthetic measures. Sections 5 and 6 contain conclusions and directions for future work.

2 Evolutionary art

Evolutionary art is a research field where methods from Evolutionary Computation are used to create works of art (good overviews of the field are [12] and [1]). Some evolutionary art systems use supervised fitness assignment (e.g. [15], [13]), and in recent years there has been increased activity in investigating unsupervised fitness assignment (e.g. [14]). The field of Computational Aesthetics investigates how computational methods can be used to assign aesthetic judgement to objects (see [5] and [4]). Functions that assign an aesthetic value to an object are typically called aesthetic measures. In this paper we investigate four aesthetic measures, and compare their output.

2.1 Four aesthetic measures

The four aestetic measures that we investigate in this paper have different mechanisms and backgrounds, and we will describe them briefly. For a more detailed description we refer to the original papers. We will briefly describe the aesthetic measures by Machado & Cardoso, Ross & Ralph, the Fractal Dimension measure, and the Combined Weighted sum measure.

Machado & Cardoso The aesthetic measure described in [8] builds on the relation between Image Complexity (IC) and Processing Complexity (PC). Images that are visually complex, but are processed easily have the highest aesthetic value. As an example, the authors refer to fractal images; they are visually complex, but can be described by a simple formula. The aesthetic measure M of an image I is defined as

$$M(I) = \frac{IC(I)}{PC(I)} \tag{1}$$

The Image Complexity can be regarded as the effort needed to compress an image, and is defined as

$$IC(I) = \frac{RMS(I)}{Compressionratio(I)} \tag{2}$$

where RMS refers to the difference between the original image and the compressed image, expressed as the root mean square. The compression ratio is the ratio between the original image size and the compressed image size. The authors suggest the use JPEG compression for image compression. We used a JPEG quality setting of 0.75 (medium quality). The Processing Complexity is calculated using fractal image compression; in our experiments we used images with a resolution of 300x300. The box-counting algorithm used a number of boxes between 6 and 75 and the threshold was set to 50.

Ross & Ralph (bell curve) A second aesthetic measure that we implemented is Ross and Ralph (Ralph's Bell Curve, [14]). This measure is based on the observation that many fine art painting exhibit functions over colour gradients that conform to a normal or bell curve distribution. The authors suggest that works of art should have a reasonable amount of changes in colour, but that the changes in colour should reflect a normal distribution (hence the name 'Bell Curve'). The computation takes several steps and we refer to [14] for details.

Fractal dimension In [16] the authors investigate the aesthetic preference of people for several types of fractals (natural, artifical and man-made). The authors found a peak in the preference for fractal images with a fractal dimension around 1.35. Images with a higher fractal dimension were considered complex, and images with a lower dimension were considered uninteresting. We use this finding to construct an aesthetic measure. For a given image I with a fractal dimension d, we define our fractal dimension aesthetic measure M as

$$M(I) = max(0, 1 - |1.35 - d(I)|) \tag{3}$$

which means that only images with a fractal dimension between 1.1 and 1.6 have a positive aesthetic measure (where images with a fractal dimension of 1.35 have an aesthetic value of 1). We calculate the fractal dimension using a technique called "box-counting" (see [16]).

Combined Weighted Sum We also wanted to investigate the usefullness of a combination of the aesthetic value by the aforementioned three aesthetic measures. We used a simple straightforward weighted sum measure were all weights were set to 1:

$$M(I) = \frac{\sum_{i=1}^{n} M_i(I)}{n} \tag{4}$$

3 Arabitat: the Art Habitat

Arabitat (Art Habitat) is our software environment in which we investigate evolutionary art. It uses genetic programming with Lisp expressions and supports both supervised and unsupervised evaluation. In this paper we only discuss unsupervised fitness evaluation using aesthetic measures. Currently we have implemented three aforementioned aesthetic fitness functions and a weighted sum

combination measure, and intend to implement more in the near future. In our system, a genotype consists of 1) a Lisp-style expression that returns a value of type double, and 2) a color lookup table. Lisp-like expressions are common within genetic programming (see [7]). Our genetic programming is type-safe and returns only results of type doubles.
The computation of a phenotype from the genotype is done as follows; for a target phenotype image with a resolution *(width, height)* we calculate the function value from the lisp expression (the genotype) for each *(x,y)* coordinate of the image. The resulting matrix of floating points is mapped onto an indexed colour table, and this results in a matrix of integers, where each integer refers to a colour index of the corresponding colour scheme. This way the colouring is independent of the double values (other approaches like [15] have functions that directly address colouring). The colour scheme is thus part of the genotype, and is also subject to mutation and crossover. A mutation in the colour scheme could result in an entirely different coloured image, even if the expression remain unaltered. The resulting image is passed to the fitness function (one of the aesthetic measures) for evaluation. See Figure 1 for a schematic overview (see http://www.few.vu.nl/˜eelco/ for more examples in colour).

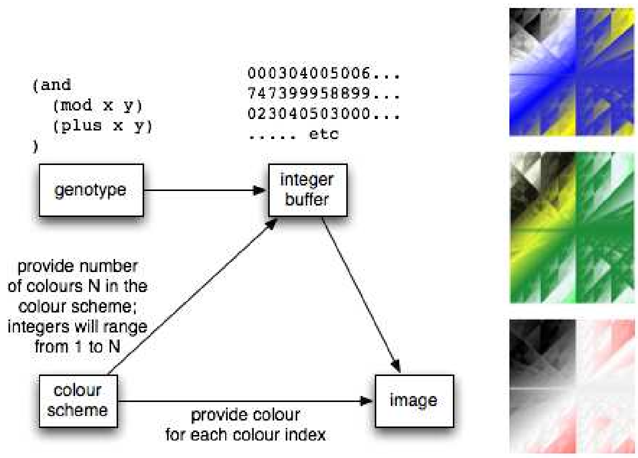

Fig. 1. A schematic overview of the expression of the genotype into the phenotype (image) for LISP expression ((and (mod x y) (plus x y))); the three images on the right are three renderings of the same expression, using three different colour schemes.

Function set Many functions used are similar to the ones used in [15], [13] and [14]. Table 1 summarizes the used functions (including their required number of arguments); The terminals x and y are variables that refer to the (x, y) coordinate of a pixel. 'Width' and 'height' are variables that refer to the width

Terminals	x,y, ephem_double, ephem_int, width, height, golden_ratio, pi
Basic math	plus/2, minus/2, multiply/2, mod/2, div/2, average/2
Other math	sin/1 ,cos/1, tan/1, sinh/1, cosh/1, tanh/1, atan2/2, cuberoot/1, squareroot/1, hypot/2
Relational	minimum/1, maximum/1, if-then-else/3
Bitwise	and/2, or/2, not/1, xor/2
Noise	perlinnoise/2, smoothnoise/2, marble/2, turbulence/2, plasma/2
Fractal	mandelbrot/2, julia/2
Boolean	equals/2, lessthan/2, greaterthan/2

Table 1. Function and terminal set of our evolutionary art system

and height of the image. The use of width and height is useful because we usually perform evolutionary computation using images with low resolution (say 300x300) and want to display the end result on a higher resolution. [15], [13] and [14] contain details on the functions used in our function set.

4 Experiments

In order to investigate and compare the four different aesthetic measure we conducted a number of experiments. We performed 10 runs for each aesthetic measure and collected the images of the 5 fittest individuals of each run. Next, we calculated the aesthetic measure of those 5 individuals by the other aesthetic measures. From the 50 images of each experiment (10 runs, 5 fittest individuals) we handpicked 9 images that were typical for that image set. Besides the aesthetic measure, all evolutionary parameters were the same for each run. We did many preliminary experiments and found that populations of around 200 usually tended to converge to one or two dominant individuals and their similar offspring. Since the goal of this paper is to compare the output of evolutionary art using different aesthetic measures, we decided to perform evolutionary search for 10 generations with a population of 200. For the genetic operators we used subtree mutation (rate 0.05), subtree crossover (rate 0.85), we initialized the population using the well-known ramped half-and-half initalization method (see [7]), and used tournament selection (tournament size 3) for both parent selection and survivor selection. For survivor selection we use elitist selection (best 1).

4.1 Results

We did 10 runs with our evolutionary art system using each aesthetic measure and collected the images of the 5 fittest individuals of each run. The average fitness of the population of 200 over 10 generations is given in Figures 3 and 5. Of the collected 50 images, we hand-picked 9 images. The reason for hand-picking from the image collection instead of selecting the images with the highest fitness is that some runs ran into premature convergence and had 5 very similar images at the end. Therefore we picked the images by hand, to give an impression of the variety of the images. Since we were mostly interested in comparing aesthetic measures using the same EA parameters, we did not focus on optimizing the EA to reach an average fitness of 1.0. Our goal was exploration, not optimization. Therefore, many runs do not end in an average fitness of 1.0. In the next sections we shortly describe the characteristics of these selections.

Machado & Cardoso The images produced using the Machado & Cardoso measure are presented in Figure 2. The images tend to be simple in structure, and they have a slight preference for primary colours (although not in all images). We suspect that the use of JPEG compression could possibly favour images with primary colours. Also apparant is that the images are diverse in structure, even if they are relatively simple. Most images produced using this aesthetic measure have a 'sixties'/ pop art look and feel. The images in [9] are slightly different; we suspect that is caused by using a different function set and a different colouring.

Ross & Ralph (bell curve) The images produced using the aesthetic measure of Ross & Ralph are presented in Figure 2. It is immediately apparant that these

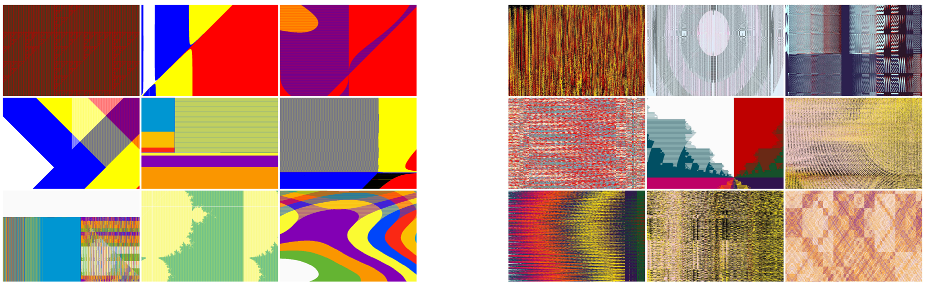

Fig. 2. Summary of images evolved using the aesthetic measure of Machado & Cardoso (left) and Ralph & Ross (right)

images are very different from the ones produced using the Machado & Cardoso aesthetic measure. Most images are very abstract and have a very distinct colour progression within the images. Many images resemble textures that are used in computer graphics, and that is similar to what the original authors found in their evolutionary art system (see [14]).

Fractal dimension The image produced using our fractal dimension aesthetic measure are presented in Figure 4.

What is apparant from these images is that the style is again different from the previous two aesthetic measures. Next, we see that there is a tendency to use the fractal functions *mandelbrot* and *julia* (which generates Julia set figures)

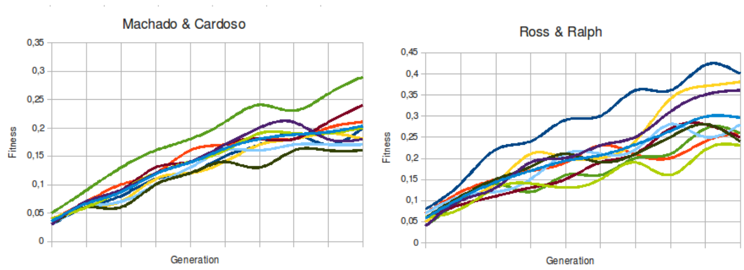

Fig. 3. Fitness progression of 10 different runs using the aesthetic measure by Machado & Cardoso (left) and Ralph & Ross (right); both ran 10 generations

and the binary function *xor* and *and*. As far as we know we are the first to use the fractal dimension in an evolutionary art context, so we can not compare our images with other generated images.

Combined Weighted Sum The images produced using the combined weighted sum aesthetic measure is given in Figure 4.

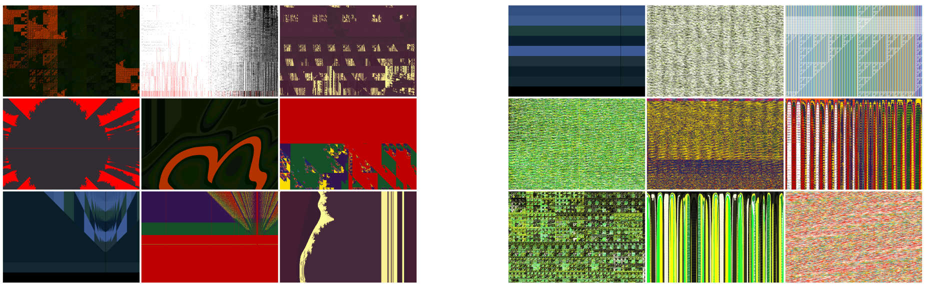

Fig. 4. Summary of images evolved using the Fractal Dimension aesthetic measure (left) and Combined weighted sum (right)

Ideally these images would be a combination of features of the previous images of the other aesthetic measures. We see that features of the aesthetic measure of Ralph & Ross seems to be dominant in the images, and that can be explained by the fact that the the average fitness of the runs using the Ross & Ralph measure is around 0.5, and the other two measures have an average fitness of around 0.2. Using weighted sum combination, the Ralph & Ross aesthetic measure thus has more 'weight' than the other two. In future implementations of combinations of aesthetic measures, we will use techniques from [3] to combine aesthetic measures in a more neutral fashion.

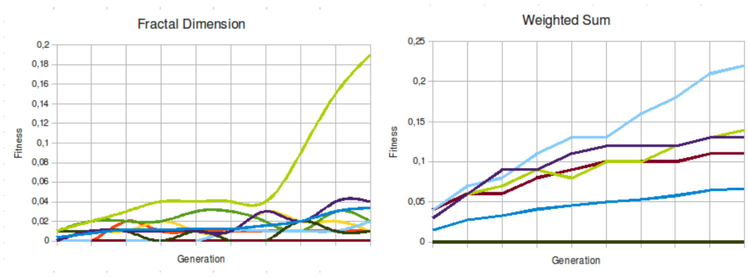

Fig. 5. Fitness progression of 10 different runs using the fractal dimension aesthetic measure (left) and the combined weighted sum measure (right); both ran 10 generations

4.2 Cross evaluation

After we had done the experiments with the four aesthetic measures, we wanted to know how the aesthetic measures would evaluate 'each others' work. The evaluation of the work of measure M_n of images produced using aesthetic measure M_m might give us an indication of the scope of the aesthetic measure. If an aesthetic measure only appreciates images that were generated using its own measure, then we could assume that its scope were fairly limited. On the other hand, if a measure also appreciates images that were created using another aesthetic measure, we could conclude that it is applicable to a broader scope of images. In the following table we have gathered the average fitness (and standard deviation) of the fifty fittest individuals that were collected for each experiment. The producing aesthetic measure is presented horizontally and the

		Evaluated by			
		Machado& Cardoso	Ross & Ralph	Fractal Dim.	Combined Weighted Sum
	Mach.& Card.	0.096 (0.054)	0.246 (0.363)	0 (0)	0.114 (0.124)
Produced	Ross & Ralph	0.035 (0.023)	0.562 (0.476)	0 (0)	0.199 (0.161)
By	Fract. Dim.	0.03 (0.009)	0.061 (0.194)	0.136 (0.305)	0.076 (0.115)
	Comb. Wei. Sum.	0.049 (0.031)	0.194 (0.337)	0 (0)	0.081 (0.115)

Table 2. The cross evaluation of the aesthetic value of each others images. We present the average asethetic value and the standard deviation in parentheses

evaluation by all aesthetic measures is presented in the columns. From this table we can conclude a number of findings. First, all aesthetic measures like their own work best (except for the combined weighted sum measure). Next, we can clearly see that the fractal dimension aesthetic measure does not appreciate of images produced by other aesthetic measures; the average score is 0.0, which means that all images not produced using the fractal dimension aesthetic measure have a fractal dimension outside the range [1.1,..,1.6]. This basically means that the fractal dimension aesthetic measure is not widely applicable as a aesthetic measure; many people like fractal properties in images, but in reality, not many images actually have fractal properties (i.e. a fractal dimension within the range [1.1,..,1.6]). Next, we see that the Ralph & Ross aesthetic measure appreciates of its own work (which is not surprising) but also appreciates of the works produced using the Machado & Cardoso aesthetic measure.

5 Conclusions

In this paper we have investigated and compared four aesthetic measures in an evolutionary art system. After our experiments we can conclude that the use of different aesthetic measures clearly results in different 'styles' of evolutionary art. Since all evolutionary parameters were kept equal in all experiments, we can conclude that all differences in artistic style are directly related to the aesthetic measures. Next, we can conclude that there are also differences in variety of the output of the four aesthetic measures. The measures of Machado & Cardoso and of Ross & Ralph have varied output. The fractal dimension aesthetic measure produces less varied output and seems less suitable as a universal aesthetic measure. Next we investigated how well the aesthetic measures like each others work. We found that the aesthetic measures by Machado & Cardoso and by Ross & Ralph appreciated work by others. The fractal dimension aesthetic measure however, did not appreciate the output by the other measures, and seems less suitable as a universal aesthetic measure. We think that the fractal dimension aesthetic measure can be useful in cooperation with other aesthetic measures in a multi-objective optimization setup. The output of the combination weighted sum measure resembles the output of the measure by Ralph & Ross, mainly because the average fitness of the Ralph & Ross measure was higher than the average fitness of the other two. In future implementations we could normalize the fitness values per aesthetic measure, in order to avoid unnecessary bias due to differences in maximum fitness. Finally, it is interesting to note that aesthetic measures used to have a passive role in computing the aesthetic value of an object, but seem to have a far more active role in creating art when applied in an evolutionary art system.

6 Future work

In this paper we chose three aesthetic measures as input for experiments with evolutionary art. Machado & Cardoso continued to develop their aesthetic measure in later research; we intend to include these changes and improvements in our implementation. Furthermore, there exist more aesthetic measures in literature. We will implement the Pattern Measure of [6], and an aesthetic measure based on information theory described in [11]. Furthermore, we would like to further explore the combination of multiple aesthetic measures into a combined aesthetic measure using techniques from multi-objective optimization (see [3]). In our experiments we have hand-picked the output from the fittest individuals;

in future research we would like to investigate the use of techniques from digital image processing to extract features from images. This way, it might be possible to investigate the output per aesthetic measure in a more systematic way.

References

1. P. J. Bentley and D. W. Corne, editors. *Creative Evolutionary Systems*. Morgan Kaufmann, San Mateo, California, 2001.
2. George D. Birkhoff. *Aesthetic Measure*. Harvard University Press, 1933.
3. K Deb. *Multi-objective Optimization using Evolutionary Algorithms*. Wiley, Chichester, UK, 2001.
4. Gary Greenfield. On the origins of the term "computational aesthetics". In Neumann et al. [10], pages 9–12.
5. Florian Hoenig. Defining computational aesthetics. In Neumann et al. [10], pages 13–18.
6. Allen Klinger and Nikos A. Salingaros. A pattern measure. *Environment and Planning B: Planning and Design*, 27:537–547, 2000.
7. John R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, 1992.
8. Penousal Machado and Amílcar Cardoso. Computing aesthetics. In *Proceedings of the Brazilian Symposium on Artificial Intelligence, SBIA-98*, pages 219–229. Springer-Verlag, 1998.
9. Penousal Machado and Amílcar Cardoso. All the truth about nevar. *Applied Intelligence*, 16(2):101–118, 2002.
10. László Neumann, Mateu Sbert, Bruce Gooch, and Werner Purgathofer, editors. *Computational Aesthetics 2005: Eurographics Workshop on Computational Aesthetics in Graphics, Visualization and Imaging 2005, Girona, Spain, May 18-20, 2005*. Eurographics Association, 2005.
11. Jaume Rigau, Miquel Feixas, and Mateu Sbert. Informational aesthetics measures. *IEEE Computer Graphics and Applications*, 28(2):24–34, 2008.
12. Juan Romero and Penousal Machado, editors. *The Art of Artificial Evolution: A Handbook on Evolutionary Art and Music*. Natural Computing Series. Springer Berlin Heidelberg, November 2007.
13. Steven Rooke. Eons of genetically evolved algorithmic images. In Bentley and Corne [1], pages 339–365.
14. B. Ross, W. Ralph, and H. Zong. Evolutionary image synthesis using a model of aesthetics. In *IEEE Congress on Evolutionary Computation (CEC) 2006*, pages 1087–1094, 2006.
15. Karl Sims. Artificial evolution for computer graphics. In *SIGGRAPH '91: Proceedings of the 18th annual conference on Computer graphics and interactive techniques*, volume 25, pages 319–328. ACM Press, July 1991.
16. Branka Spehar, Colin W. G. Clifford, Ben R. Newell, and Richard P. Taylor. Universal aesthetic of fractals. *Computers & Graphics*, 27(5):813–820, 2003.