Computational Aesthetics in Graphics, Visualization and Imaging (2005)
L. Neumann, M. Sbert, B. Gooch, W. Purgathofer (Editors)

Designing Metrics for the Purpose of Aesthetically Evaluating Images

Gary Greenfield

Mathematics and Computer Science, University of Richmond, Richmond, Virginia, USA

Abstract

The algorithmic and evolutionary art movements within computer-generated art have helped spur interest in evaluating images on the basis of their aesthetic merit. When attempting to use non-interactive techniques to address this issue, two problems arise: (1) designing metrics that have explicit computational representations, and (2) establishing that such metrics actually fulfill their intended purpose. We survey our experiences in designing metrics for non-interactively guiding image evolution to obtain aesthetic images and we propose a taxonomy for metric frameworks. We also discuss some issues relevant to validating such metrics.

Categories and Subject Descriptors (according to ACM CCS): J.5 [Computer Applications]: Arts and Humanities, I.4.7 [Image Processing and Computer Vision]: Feature Measurement

1. Introduction

The literature on perception, digital art, and art criticism provides limited guidance and even fewer suggestions for helping researchers design metrics for evaluating images on the basis of their aesthetic merits. In this paper, by restricting out attention to the problem domain of generative art, we first survey previous work on using metrics to guide aesthetics in a non-interactive evolutionary setting and then, by focusing on the generative technique known as "evolving expressions," we discuss several ways we have gone about implementing such metrics. We also consider the design of metrics for biologically inspired generative methods, and then we propose a taxonomy of design methods for metrics. Although many researchers implementing such metrics have included user testing as a part of "future work," apart from the belief that such metrics must be subject to innate social and cultural biases, few suggestions have emerged for how to perform constructive testing. We therefore include some remarks on the problems of validation and user testing.

It is important to make clear at the outset that we are making a distinction between devising metrics that *substitute* for one's artistic expression and thereby serve as *extensions* of the individual artist's themselves in the way artists such as Cohen, Knowlton, or Mohr have done [EC02], and devising general purpose metrics that can be tuned or customized to evaluate images according to a variety of different aesthetic criteria.

2. Generative Art

Perhaps because computer-generated abstract art is the least contentious problem domain, or perhaps because it is easy to develop test suites of images using generative techniques, most researchers investigating metrics for aesthetic evaluation of images have focused on non-photorealistic images obtained using generative methods.

2.1. Chronology of generative techniques

We list several well-known generative art schemes in roughly chronological order. Our list is not exhaustive, but we feel it is a useful guide to the literature. Most well-known among the generative techniques are the line drawings of Dawkins [Daw89], the abstract art of Sims [Sim91], the organic forms of Latham [TL92], and the dynamical system visualizations exhibited as fine art by Field and Golubitsky [FG92]. Less well-known, but indicative of the wide range of techniques that this domain encompasses, are the implicit surfaces evolved by Bedwell [BE98], the aesthetic textures of Ibrahim [Ibr98] and Lewis [Lew01], the aesthetic patterns of Staudek [Sta03], and the image re-colorings of Greenfield [Gre04].

2.2. Origins of metrics

The first attempt to implement a metric for aesthetically evaluating a population of images within the context of an evolutionary, generative art system was by Baluja et al [BPJ94]. These researchers attempted to train a neural net to perform this evaluation task using as training sets images that were obtained by categorizing the user rankings of images evaluated while users were running an interactive version of their generative system. Rooke, in unpublished work, evolved expression trees in such a way that the aesthetic rankings of the images within the image population made by these trees coincided almost exactly with his own rankings of the images in the same population. He then allowed his evolved population of expressions, or "critics" as he called them, to control the evolution of images in his generative system. The ability of his trees to make aesthetics rankings is explained by the fact that the underlying primitives in the nodes of the trees were able to make statistical assessments of the images. Sprott appears to have been the first digital artist to investigate the use of global complexity measures for aesthetically evaluating fractal-like images [Spr96]. Greenfield is the first to have published about the use of co-evolutionary predator-prey metrics for evolving images [Gre00a]. More will be said about this topic in the next section. Finally, it should also be mentioned here that Machado and Cardoso made use of neural nets when aesthetically evaluating images in the image populations evolved by their generative system [MC98].

3. Evolving Expressions

As a result of his now famous SIGGRAPH '91 paper [Sim91], Karl Sims helped spawn a cottage industry of computer artists who have built generative systems based on the "evolving expressions" technique for creating abstract images that he first introduced. Early practitioners of this craft include Rooke, Greenfield, Unemi, Machado, Mount, Rowbottom, and Musgrave. More recent converts include Ashmore, Kleiweg, Rowley and Ross. Much of the work of this cadre of artist-researchers is web accessible.

3.1. Overview of the method

The details of the generative system that we will use here when considering the problem of aesthetically evaluating images may be found in [Gre00b] and [Gre02]. It is based on Sims evolving expressions method. For our purposes, it suffices to view abstract images as being generated from functions whose domain is the unit square and whose range is the unit interval, In our generative system functions are algebraic expression trees written using postfix notation. This implies an expression of the form, say `V1 U2 V0 C758 B0 B6`, defines a function of two variables, $f(V0,V1)$. The expression tree has all interior nodes labelled with `B`'s or `U`'s and all leaves labelled with `V`'s or `C`'s. The `B`'s and `U`'s are binary and unary functions respectively selected from the function library given in [Gre00b]. For the leaves the `V`'s are variables and the `C`'s are constants. A function $f(V0,V1)$ gives rise to an $N \times N$ pixel image defined with reference to a color look up table of size L whose colors vectors are $c_1,\ldots,c_L$ by coloring pixel $p_{i,j}$ with color c_k provided

$$f(i/N, j/N) \in [(k-1)/L, k/L).$$

The principal advantage of using postfix expressions is that recombination, mutation, and evaluation operators are easy to implement.

3.2. Co-evolution

In [Gre00a], we described a co-evolutionary method for evolving gray-scale images using the evolving expressions set-up we have just described. We viewed a population of images as *hosts* for *parasites* — 3×3 digital convolution filters attached at specific *sites* of the image. Parasites were able to assign a numerical aesthetic value to both host and parasite by acting as "irritants" in the following manner. The parasite's filter was convolved over the 10×10 pixel *patch* of the host determined by the coordinates of the site where the filter was attached, and then a pixel by pixel comparison of the result with the underlying image was made. The magnitude of the difference at each pixel determined whether a point was awarded to the host or to the parasite. The dynamic in force was that hosts were rewarded for increasing their "complexity" within the patch in order to ward off parasites who were rewarded for their ability to be able to "predict" the functional values of the host based solely on the values of nearby pixels. The reason why this measure of aesthetic fitness exerted evolutionary pressures that led to interesting images is because there was a local-global tension at work due to the fact that when host genomes reacted to the local irritation induced by the parasites their global structure changed.

Unlike the other generative schemes discussed below, this scheme is computationally efficient. Moreover aesthetic fitness is not an absolute quantity, but is only defined relative to the current parasite population, a population that is also mutating and evolving. For this reason premature evolutionary convergence is avoided. As Figure 1 shows, due to the nature of the fitness computation, image entropy is high, and the style of the co-evolved images is very noisy.

3.3. Fitness functions

In [Gre02], using the same generative system as before, but now with a color look-up table consisting of 450 colors, we considered the problem of designing metrics to evaluate the aesthetics of images based on the geometric characteristics of their compositions. To accomplish this we color segmented a 32×32 thumbnail of the image to yield m regions with areas $a_1,\ldots,a_m$, boundary lengths $b_1,\ldots,b_m$, and region adjacency counts $j_1,\ldots,j_m$, indexed so that $a_1 \geq a_2 \geq$

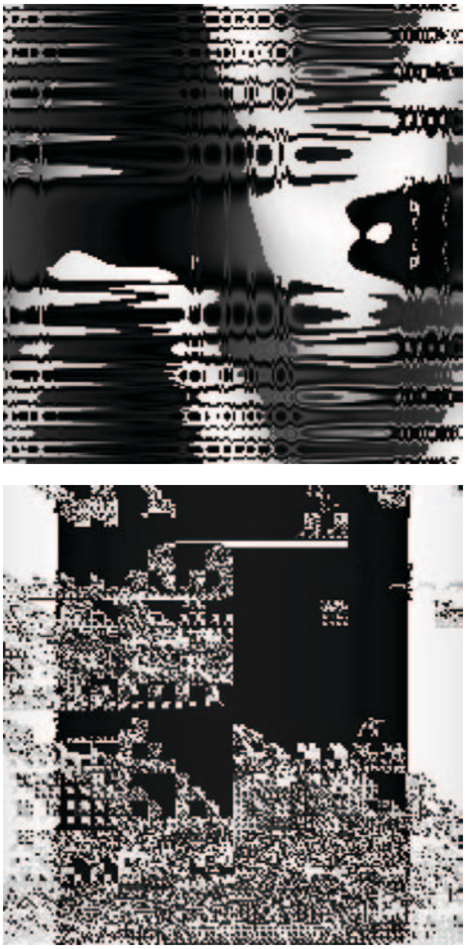

Figure 1: *Two co-evolved gray-scale images.*

$\cdots \geq a_m$ and then extracted the following measurements to help quantify the geometry of the segmentation:

$$\begin{aligned} A(s,t) &= \sum_{k=s}^{t} (k+1)a_k, \\ B(s,t) &= \sum_{k=s}^{t} b_k, \\ J(s,t) &= \sum_{k=s}^{t} j_k. \end{aligned}$$

Next we defined the fitness of an image I to be a weighted linear combination of these terms. This enabled us to explore the parameterized space of fitness functions of the form

$$F(I) = w_A A(s_1,t_1) + w_B B(s_2,t_2) + w_J J(s_3,t_3).$$

Figure 2 shows two evolved images with their corresponding segmentations. Figure 3 shows additional examples. All of these images were evolved using simple fitness functions such as $F(I) = A(2,4) + B(1,m)$ or $F(I) = A(1,1) + J(1,m)$.

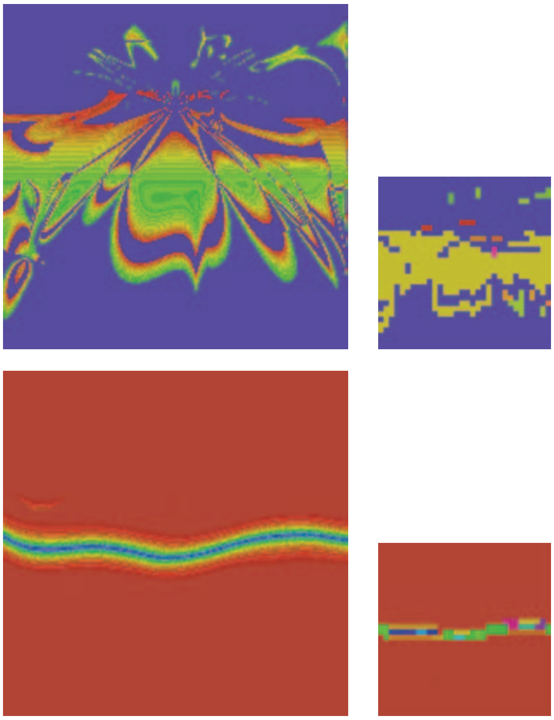

Figure 2: *Two images accompanied by their color-segmented thumbnails that were evolved from user designed fitness functions.*

The point is that our geometric assessments allowed us, as fitness function *designers*, to exert evolutionary pressure on image evolution by biasing it in favor of images with contiguous sequences of regions that were area balanced, delicately intertwined, or even densely connected. Notice however that fitness was not directly responsible for color content only image composition because the fitness functions did not use color components as arguments.

3.4. Multi-objective optimization

To overcome the premature evolutionary convergence that frequently occurred using the previous method, we next turned to multi-objective optimization [Gre03b]. Using the NSGA II algorithm of Debs as a diversity mechanism, we were able to simultaneously evolve two or more *interacting* subpopulations of images, where each population was induced according to the above scheme. As Figure 4 shows, we achieved some successes by using "round-robin" fitness function schemes such as:

$$\begin{aligned} F_1(I) &= 10J(1,25) + B(1,4), \\ F_2(I) &= B(1,4) + A(1,4)/5, \\ F_3(I) &= A(1,4)/5 + 10J(1,25). \end{aligned}$$

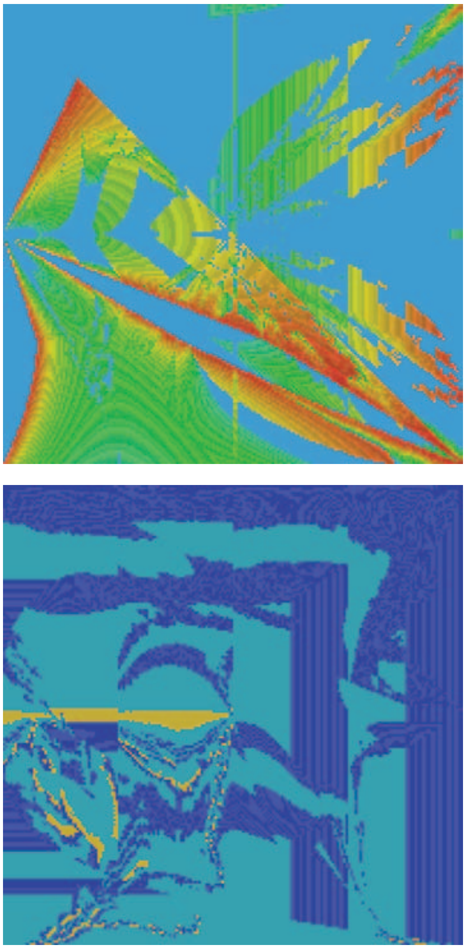

Figure 3: *Two images evolved using user designed fitness functions to influence certain composition characteristics.*

Figure 5 shows four other examples obtained using other combinations of the various elementary fitness functions at our disposal.

3.5. Image re-coloring

To re-color our images, in [Gre04] we evolved color look-up tables of the form $(t_1, \ldots, t_L)$ where the t_i's were not necessarily distinct colors drawn from our fixed set of $L = 450$ HSV color vectors. Again we used color segmentation and multi-objective optimization, but now we included the color assessment measure $T(i,h)$ to force region i to be a color whose hue component was approximately h, and $C(i,j)$ to force regions i and j to have complementary hues, in our fitness functions. With these enhancements, image re-colorings such as those shown in Figure 6 were obtained by using fitness schemes such as:

$$F_1(I) \;=\; A(2,6) \cdot J(13,25) + C(4,5)$$

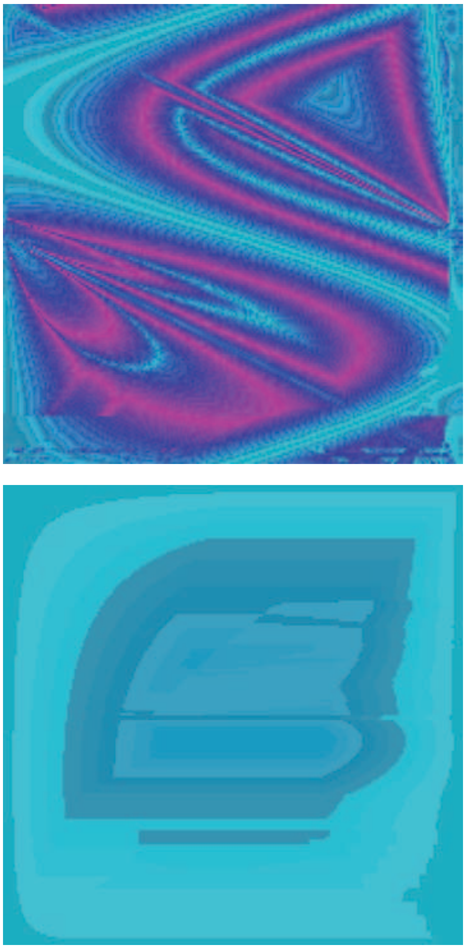

Figure 4: *Two images evolved during the same run using evolutionary multi-objective optimization.*

$$F_2(I) \;=\; \min(T(1,4.2), T(2,3.7)) \cdot B(1,4).$$

4. Biologically Inspired Examples

In this section we survey our efforts to design aesthetics metrics for generative art that is loosely based on biologically inspired processes.

4.1. Ant colony optimization

Following Monmarché et al [ABM*03], in [Gre05] we considered an ant colony optimization simulation where a small number of virtual ants are allowed to roam on an $N \times N$ pixel grid seeking and depositing color. By evaluating the individual ants *behavior* during the "painting" phase, we were able to remove ants and breed replacement ants for the population in such a way that when the underlying grid was re-set to white, the ants could improve the aesthetic quality of their

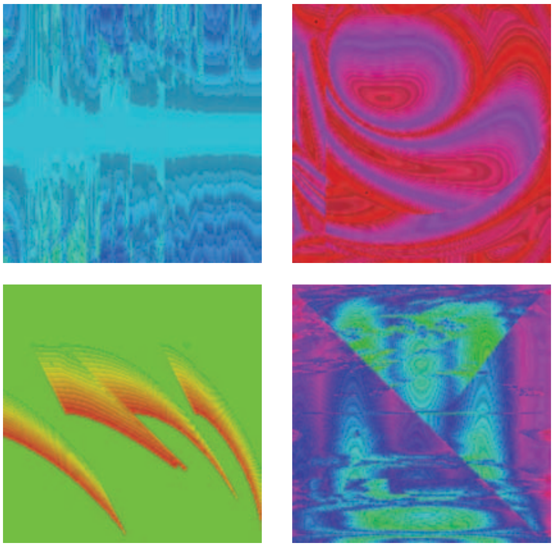

Figure 5: *Four images evolved using evolutionary multi-objective optimization.*

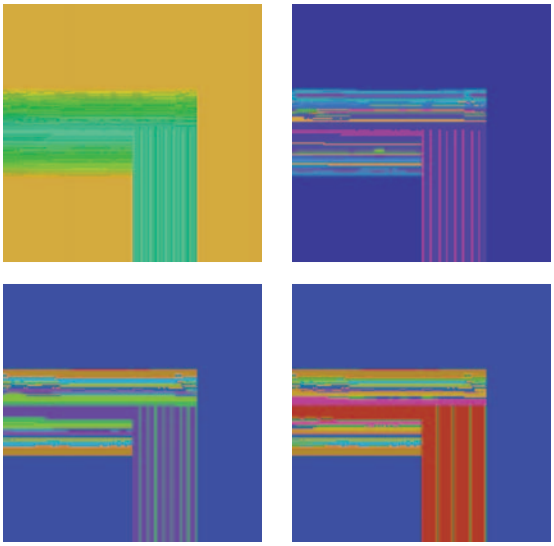

Figure 6: *Original image at upper left together with three re-colorings evolved using evolutionary multi-objective optimization.*

paintings. Since we were indirectly controlling the composition of the ant painting by evolving ant behaviors, for each ant we measured n_v, the number of distinct cells the ant visited during the period allotted for painting, and n_f, the number of times scent following occurred. Figure 7 shows two ant paintings that were evolved in under twenty generations using ant populations of size twelve, grids of size 200×200, allotting 2400 time cycles for ants to complete

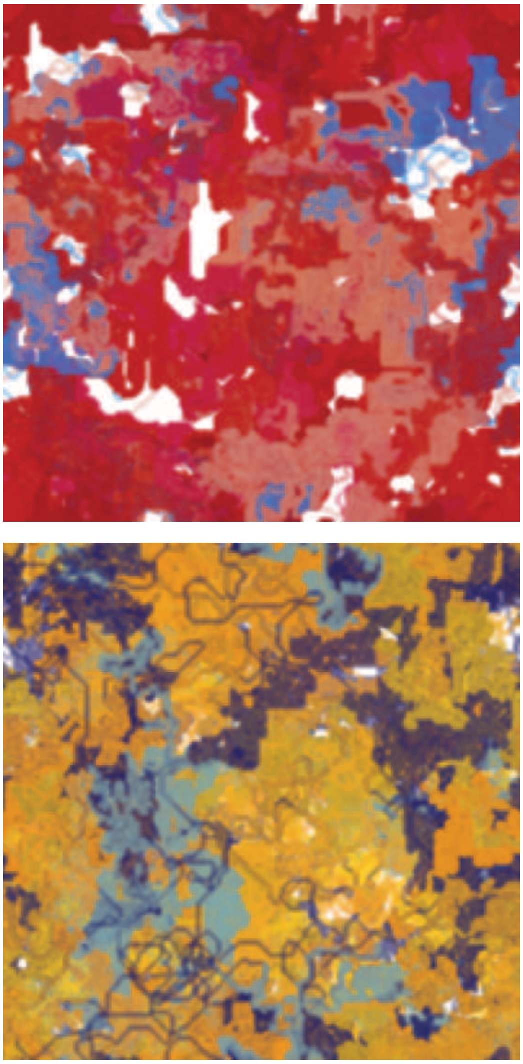

Figure 7: *Two ant paintings evolved using ant fitness functions $F(A) = n_v + n_f$ and $F(A) = n_v \cdot n_f$ respectively.*

their painting. Once again, simple fitness functions such as $F(A) = n_v + n_f$ and $F(A) = n_v \cdot n_f$ were used. Note also that the color schemes for the ant paintings were not specified explicitly but were evolved in response to evolutionary pressures exhibited on initial populations of ants whose pseudo-randomly generated genomes coded for the colors to deposit and seek.

4.2. Cellular morphogenesis

Following Eggenberger [Egg97], in [Gre03c] we considered the evolution of aesthetically pleasing visualizations of cellular morphogenesis processes of conglomerates of cells where cell activities were governed by a regulatory gene structure. To oversimplify, an $N \times N$ substrate was filled with two types of cells. Each cell contained four products, or morphogens, whose concentrations affected the regulatory genes which in turn affected the production and diffusion of additional morphogens. By initializing cells with trace amounts of morphogens and applying a morphogen gradient to the external cell boundary, over time an outside-in cell activation pattern developed as shown in Figure 8. If we label the morphogens (R)ed, (B)lue, (G)reen, and (C)ommunication, and we let $sigma_X$ denote the standard deviation of morphogen X, n_d denote the number of cells that are dormant after the prescribed number of developmental cycles has occurred, and n_a denote the number of cells that altered their morphogen production behavior during the last developmental cycle, then

by letting the fitness function be

$$F(I) = \frac{\sigma_C \cdot n_a \cdot \min(\sigma_R, \sigma_G, \sigma_B)}{1 + n_d},$$

we were able to evolve visualizations such as those shown in Figure 9. This fitness function uses the σ_C term to ensure diffusion of cell products, penalizes cell patterns with too many dormant cells, and thanks to the presence of the n_a term ensures that cellular activity has not reached a steady state. Moreover, since it only requires at least one color morphogen be present in varying concentration levels, it too does not directly control for color.

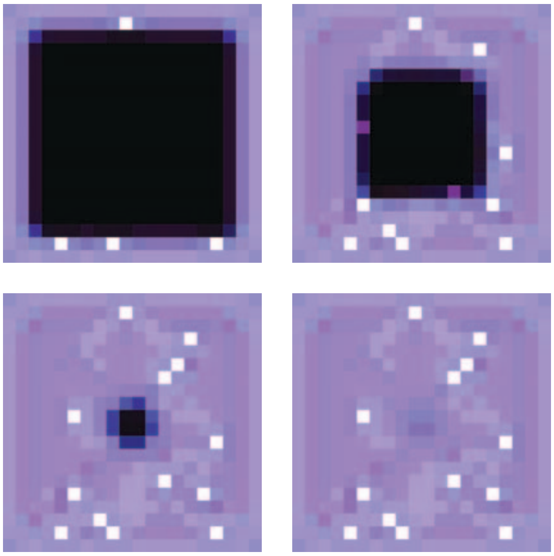

Figure 8: *A time series showing the outside-in development of a* 20×20 *cell pattern after 50, 150, 250, and 350 time steps.*

5. Metric Design

Based on our experiences, we conclude that metric design for aesthetic purposes is a two-stage process. First, one must decide what statistical measurements should be acquired from the images themselves. Second, one must decide how to combine those measurements into an aesthetic evaluation *tool*. In other words, we do not feel a metric is simply a function $F(I)$, where I is an image, but rather we feel it is an assessment *framework* derived from functions of the form

$$F(m_1(I), m_2(I), \ldots, m_r(I)),$$

where each $m_i(I)$ is a carefully chosen image assessment *parameter*. In our view metric design is *motivated* by considering cognitive, perceptual, or other psychological factors that help suggest useful parameters that can be acquired from images as well as ways to organize them, and then *implemented* using a practice-based approach that refines fitness calculation formulas until they meet either subjective or objective criteria.

6. Metric Taxonomy

We propose a taxonomy for the metric framework we formulated above. We include examples from the literature to show the kinds of metrics we wish to include in each category. Our own work listed under the "learning" category is not discussed here because its AI implications are beyond our scope. It is interesting to note that the only examples that we feel qualify for the "negative feedback" category arise from biologically inspired artificial life simulations.

- Positive Feedback
  - e.g. simulated co-evolution [Gre00a]
  - e.g. neural nets [MC98]
- Negative Feedback
  - e.g. simulated immune systems [RSMS05]
  - e.g. simulated diseases [Dor05]
- Direct Control
  - e.g. families of fitness functions [Gre02]
- Indirect Control
  - e.g. multi-objective optimization [Gre03b]
  - e.g. ant colony optimization [Gre05]
- Learning
  - e.g. image analogies [HJO*01]
  - e.g. simulated gaze data [Gre03a]

7. Validating Metrics

Concerning the problem of validating our metrics, we observe that it is confounded by the fact that any image identified using a metric is still subject to final user acceptance. This means validation must be considered in both qualitative and quantitative terms: Did a metric successfully identify images meeting the aesthetic criteria? How often did it succeed in doing so? A frequently heard suggestion is to compare automated evaluation of an image population with *artist* rankings. This means dividing testers into two groups: those with, and those without, artistic training. Although there is some evidence that the viewing *behaviors* differ for these two groups, and the viewing *preferences* differ for these two groups, there is no evidence that their aesthetic judgments differ because judging criteria are rarely specified. Moreover, in an evolutionary setting, it is easily argued that user-assigned aesthetic fitness is non-objective because image rankings are neither reproducible nor constant over time due to such factors as fatigue and boredom. We propose that metric validation by user testing cannot occur until validation of “users” occurs. In this vein, a recent experiment by Linkov and Staudek [LS04] organizing testers into aesthetic groups based on their preferences, and then analyzing the characteristics of those groups is of interest. It could form the basis for an approach that first identifies whether a testing group *should* be able to determine if a proposed metric is capable of successfully selecting images on the basis of, say, “complexity” or “symmetry.”

8. Conclusions

We surveyed some of our work on designing metrics to automatically evaluate the aesthetic merits of images belonging to populations evolved using evolutionary, generative art methods. We proposed a taxonomy for metric design. We briefly considered the problem of metric validation and user testing. Clearly this work is only in its beginning stages.

References

[ABM*03] AUPETIT S., BORDEAU V., MONMANRCHÉ N., SLIMANE M., VENTURINI G.: Interactive evolution of ant paintings. In *Congress on Evolutionary Computation* (2003), pp. 1376–1383. (Proc. CEC ’03).

[BE98] BEDWELL E., EBERT D.: Artificial evolution for implicit surfaces. In *SIGGRAPH Conference Abstracts and Applications* (1998), p. 261. (Computer Graphics Annual Conference Series 1998).

[BPJ94] BALUJA S., POMERLEAU D., JOCHEM T.: Towards automated artificial evolution for computer-generated images. *Connection Science 6* (1994), 325–354.

[Daw89] DAWKINS R.: The evolution of evolvability. In *Artificial Life* (1989), Langton C., (Ed.), pp. 201–220. (Proc. Alife I).

[Dor05] DORIN A.: Artificial life, death and epidemics in evolutionary, generative electronic art. In *EvoWorkshops 2005 Proceedings* (2005), pp. 448–457. (Proc. EvoMUSART ’05).

[EC02] EDMONDS E., CANDY L.: Creativity, art practive and knowledge. *Comm. of the ACM 10*, 10 (2002), 91–95.

[Egg97] EGGENBERGER P.: Evolving morphologies of simulated 3d organisms based on differential gene expression. In *Proceedings of the Fourth European Conference on Artificial Life* (1997), pp. 205–213. (Proc. ECAL ’97).

[FG92] FIELD M., GOLUBITSKY M.: *Symmetry in Chaos*. Oxford University Press, 1992.

[Gre00a] GREENFIELD G.: Art and artificial life — a coevolutionary approach. In *Artificial Life VII Conference Proceedings* (2000), pp. 529–536.

[Gre00b] GREENFIELD G.: Mathematical building blocks for evolving expressions. In *BRIDGES 2000 Conference Proceedings* (2000), pp. 61–70.

[Gre02] GREENFIELD G.: Color dependent computational aesthetics for evolving expressions. In *BRIDGES 2002 Conference Proceedings* (2002), pp. 9–16.

[Gre03a] GREENFIELD G.: Computational aesthetics and gaze patterns. In *ISAMA-BRIDGES 2003 Conference Proceedings* (2003), pp. 85–92.

[Gre03b] GREENFIELD G.: Evolving aesthetic images using multiobjective optimization. In *Congress on Evolutionary Computation* (2003), pp. 1903–1909. (Proc. CEC ’03).

[Gre03c] GREENFIELD G.: The void series — generative art using regulatory genes. In *Generative Art 2004 Conference Proceedings* (2003), pp. 297–304. (Proc. GA ’04).

[Gre04] GREENFIELD G.: Automated recoloring of evolved designs using evolved palettes. *Journal Mathematics & Design 4*, 1 (2004), 47–54. (Proc. MD ’04).

[Gre05] GREENFIELD G.: Evolutionary methods for ant colony paintings. In *EvoWorkshops 2005 Proceedings* (2005), pp. 478–487. (Proc. EvoMUSART ’05).

[HJO*01] HERTZMANN A., JACOBS C., OLIVER N., CURLESS B., SALESIN D.: Image analogies. In *Proc. SIGGRAPH ’01* (2001), pp. 327–340.

[Ibr98] IBRAHIM A.: *GenShade*. PhD thesis, Texas A&M University, 1998.

[Lew01] LEWIS M.: *Creating continuous design spaces using interactive genetic algorithms with layered, correlated, pattern functions*. PhD thesis, Ohio State University, 2001.

[LS04] LINKOV V., STAUDEK T.: Personality characteristics and aesthetic preference for chaotic curves. *Journal Mathematics & Design 4*, 1 (2004), 297–303. (Proc. MD ’04).

[MC98] MACHADO P., CARDOSO A.: Computing aesthetics. In *Proceedings XIV-th Brazilian Symposium on Artificial Intelligence* (1998), pp. 219–229. (Proc. SBIA ’98).

[RSMS05] ROMERO J., SANMARTIN E., MACHADO P., SANTOS A.: Artificial immune system based art. In *Proceedings 18-th International FLAIRS Conference* (2005). (Proc. FLAIRS ’05).

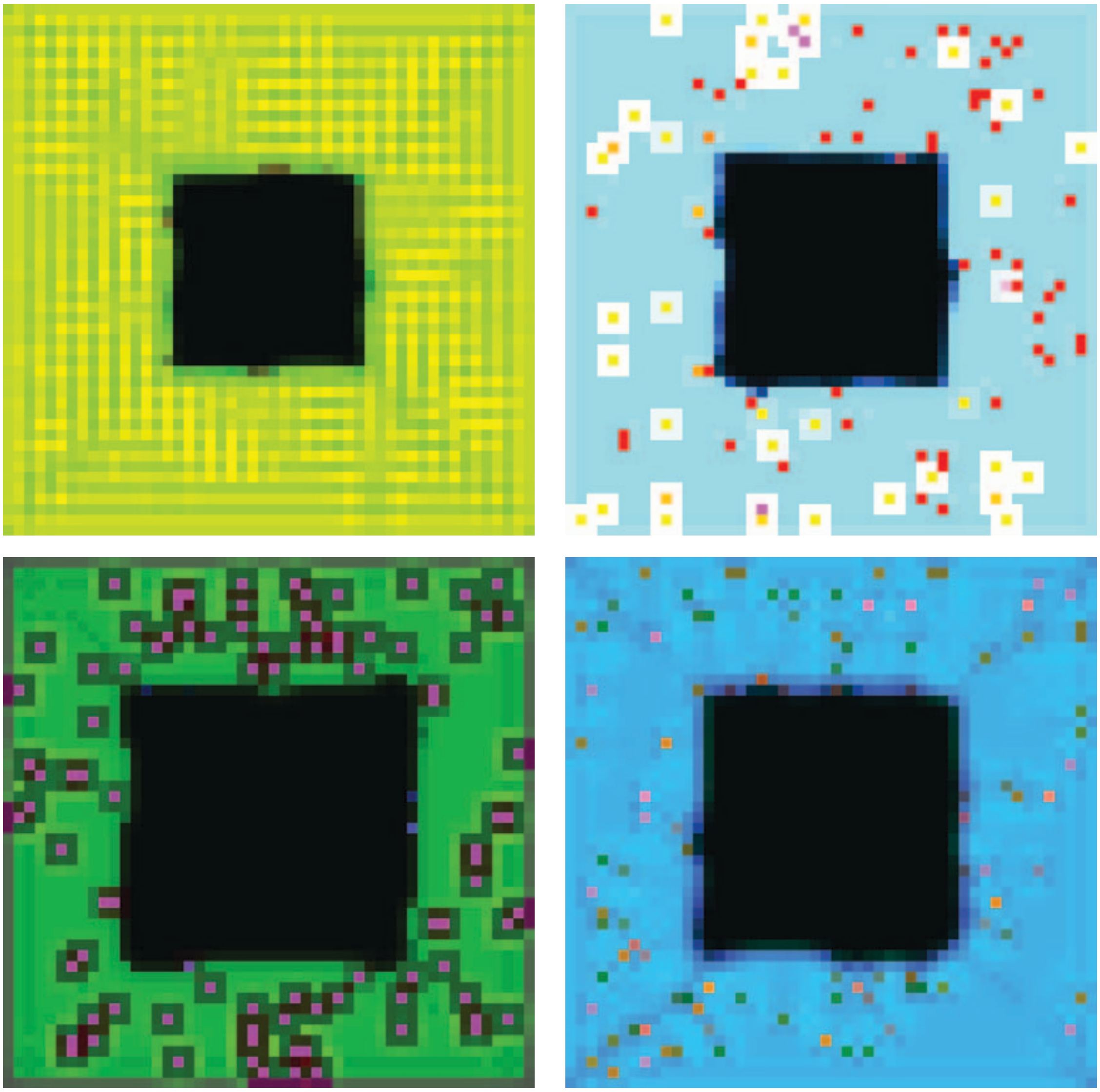

Figure 9: *Cellular morphogenesis visualizations from "The Void Series" that were evolved to satisfy certain aesthetic criteria.*

[Sim91] SIMS K.: Artificial evolution for computer graphics. In *Proc. SIGGRAPH '91* (1991), vol. 25, pp. 319–328.

[Spr96] SPROTT J.: The computer artist and art critic. In *Fractal Horizons* (1996), Pickover C., (Ed.), pp. 77–115.

[Sta03] STAUDEK T.: Computer-aided aesthetic evaluation of visual patterns. In *ISAMA/BRIFGES 2003 Conference Proceedings* (2003), pp. 143–150.

[TL92] TODD S., LATHAM W.: *Evolutionary Art and Computers*. Academic Press, 1992.