ISAMA
The International Society of the Arts, Mathematics, and Architecture

BRIDGES
Mathematical Connections in Art, Music, and Science

Computational Aesthetics Based On Gaze Patterns

Gary R. Greenfield
Department of Mathematics & Computer Science
University of Richmond
Richmond, VA 23173, U.S.A.
ggreenfi@richmond.edu

Abstract

Generative art systems usually rely on the technique of user-guided evolution to enable "artists" to interactively search through populations of images in order to breed those images which show aesthetic promise. We consider algorithmic criteria for automating the aesthetic evaluation step. Our criteria are inspired by a recent description of techniques used for controlling the aesthetic reorganization of photorealistic imagery. Surprisingly, our approach seems to impose a rigid style on the images we evolve. A drawback to our method is that eventually it fails to clearly differentiate between non-degenerate and degenerate images. We consider how improvements might be made.

1. Introduction

It is difficult to formulate general purpose aesthetic metrics for visual imagery. The human visual system is both complex and sophisticated. Moreover, human aesthetic decision making is subject to cognitive and cultural biases. To date, there appear to be a limited number of examples where aesthetic evaluation of images by computational means has been wholly successful. At one extreme lie simple generative systems whose underlying objective is to color the squares within a relatively small rectangular grid using only a limited number of colors. In this case reasonable metrics for identifying aesthetic qualities such as balance, symmetry, and pattern have been studied [9] [16]. At the other extreme lie specialized generative systems supporting the narrowly defined "styles" of individual artists. These systems are capable of yielding high quality aesthetic imagery *i.e.* fine art. Works by Cohen, Mohr, and Knowlton (see [4] and [10]) provide well publicized examples of fine art produced by such systems. Somewhere in between these two extremes lies a *computational* realm where aesthetic evaluation of imagery has also been successful, presumably due to the high level of *mathematical* content of the underlying imagery (see, for example, the work of Sprott [15]). Indeed, one might conjecture that any generative system whose primary objective is to visualize purely mathematical objects would be a good candidate for computational aesthetics. Systems that we feel merit consideration for inclusion in this category include those of Brill [1], Krawcyzk [11], Priebe [12], and Hemert [8]. General purpose generative systems like those of Rooke [18], Sims [14], Unemi [17], and the author [6] that are capable of generating a more diverse spectrum of imagery offer a stiffer challenge in this regard. Therefore they serve as excellent test beds for proposed algorithms for measuring aesthetic fitness. A crucial difference in these systems is that they require searching for aesthetic images in spaces that are associated with *rugged* fitness landscapes, spaces where tens of thousands of aesthetically valuable images often lie cleverly concealed among tens of millions of aesthetically worthless images. Designing algorithms for assigning aesthetic fitness to imagery produced by such systems is proving to be a daunting task. It is further complicated by the fact that these systems evolve imagery over time, forcing aesthetic measures not only to identify aesthetic images but to guide their evolution.

In previous work, we have considered various computational aesthetic schemes. Our most successful scheme coevolved small sets of image "probes" in tandem with aesthetic images [6]. We also considered geometric methods for assessing image aesthetics that were based upon measurements derived from color-induced simplifications, or segmentations, of images [7]. The goal of this paper is to investigate yet another approach. It is inspired by a recently developed technique that was used to guide a sequence of aesthetic modifications that were made to a photorealistic image.

This paper is organized as follows. In section two we review this new technique. In section three we explain how we incorporated its underlying theme into the fitness evaluation step of our generative system. In section four we discuss some of the results we obtained. In section five we give our conclusions.

2. Image Analysis Based on Eye Tracking Data

DeCarlo and Santella [3] describe an impressive technique for transforming a photograph into a stylized, abstract realization of its subject matter. Their non-photrealistic transformation is accomplished in several steps. First, a human subject wearing an eye-tracker is asked to study the photograph for a brief period of time (say thirty seconds) so that a number of gaze postions can be measured and recorded. The duration of the gaze at each gaze position is also recorded and used to define a sphere of influence for the gaze position. Second, the photograph is color segmented in order to decompose it into feature regions. Third, the gaze information is used to decide how to aggregate the feature regions so that regions within the spheres of influence are preserved and others are blurred. Fourth, a smoothing operator is applied to the aggregate image. Fifth, image processing techniques are used to highlight selected feature regions with bold black lines. The resulting non-photorealistic effect can be quite stunning. Figure 1 illustrates the effect DeCarlo and Santella are able to achieve. From our point of view, the underlying premise of their technique is that the gaze information — gaze positions together with their spheres of influence – constitutes an aesthetic evaluation of the image.

3. Gaze Patterns and Aesthetic Fitness Measures

A generic description of the generative system we use to generate our abstract imagery can be found in [5]. Here, it will suffice to recall that our system uses the method of evolving expressions first introduced by Sims [14]. This method generates an image from an expression tree consisting of nodes that contain functions in two variables. To assign a color to each pixel, its coordinates are passed to the expression tree as inputs and the corresponding output is scaled to an integer between 1 and 450 that is then mapped to a color according to a color lookup table. In this paper the imagery that *can* be evolved is identical with the imagery that can be evolved by the generative system we described in [7]. What differs here are the algorithms used to determine an image's aesthetic fitness.

Our objective is to make use of existing gaze data, which we think of as a set of predefined gaze

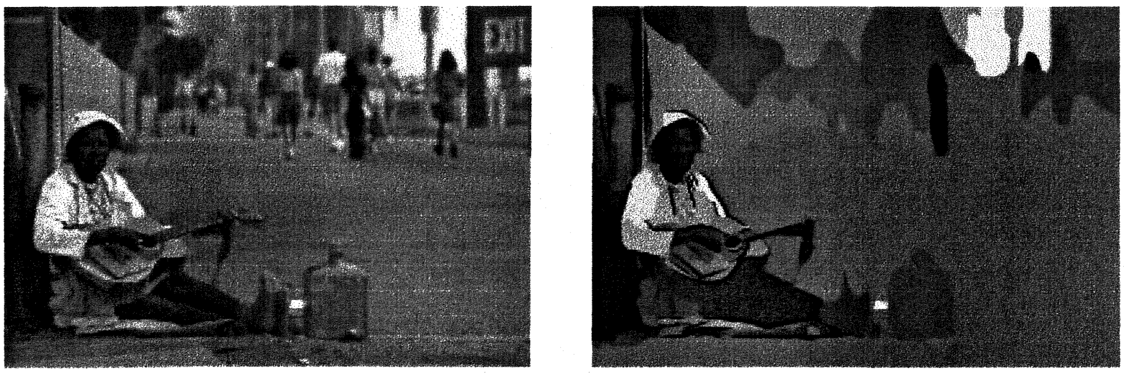

Figure 1: Sample before and after images illustrating DeCarlo and Santella's technique for creating abstractions of photographs ©2002 ACM, Inc. Reprinted by permission.

positions and predefined spheres of influence. Our test gaze data set is a first-order approximation of the gaze pattern obtained from the photograph in Figure 1. For computational convenience, however, our "spheres" of influence will always be *squares* of influence. Formally, then, a *gaze pattern* consists of a sequence of vertices associated to spheres of influence and a sequence of side lengths for the spheres *e.g.* $p_1 = (2, 8), p_2 = (4, 10), \ldots, p_{11} = (16, 24)$ and $l_1 = 8, l_2 = 4, \ldots, l_{11} = 2$.

Whereas DeCarlo and Santella used an image to obtain a gaze pattern, we are attempting to use a gaze pattern to obtain an image, or to be more precise, use a gaze pattern to determine whether or not an image should be included in the pool of images that are allowed to breed new images. To implement our approach, we color segment a thumbnail (32×32 pixel) rendering of an image. This means we use an iterative process to aggregate pixels into regions in such a way that all the pixels comprising a region have similar color characteristics. Our algorithm begins by declaring each pixel to be a region. It then successively glues together, or *merges*, the two neighboring regions that are most similar in color until the desired number of regions is obtained — a bottom-up region-merging algorithm (see [7] for details). We want the regions that are formed to have a causal connection to the regions the human visual system considers when it examines an image. To achieve this, the color similarity decisions we need to make in order to decide how to merge regions must closely match the color similarity decisions the human visual system would make. It has been argued that for natural images, specifying digital color in terms of the three color components of Lab color space allows one to make such decisions [13]. Let $v(R)$ be the Lab color vector associated with region R and let $||v(R)||$ denote its length. Given regions $R_1, R_2, \ldots, R_n$ our algorithm merges region R_i with R_j provided $i \neq j$, R_i neighbors R_j, and $||v(R_i) - v(R_j)||$ is minimal. The problem we must now confront is how to decide whether or not a color segmented image has feature regions *compatible* with the predefined spheres of influence. This requires some experimentation.

A sphere of influence should help indicate where the image "interest" lies. One measure of interest is the color variation within the pixels that make up the sphere. Clearly, this value will be large if spheres contain only a few regions, those regions have an abundance of boundary pixels, and those region's color gradients at their region boundaries is severe. Observation suggests, however, that spheres of interest are often spheres of interest precisely because they contain *many* feature regions. Thus in order to find images where visual interest is maximized, we seek to maximize color variation within the spheres of influence while simultaneously maximizing the number of feature regions that wind up in contact with the spheres of influence. To achieve these goals, let I be an image, and let $F(I)$ be its aesthetic fitness. Let $V(c, I)$ be the *variance* in color channel c calculated over all pixels lying within spheres of influence, and let $H(I)$ be the number of regions containing pixels lying in one or more spheres of influence. Our first choice for a measure of aesthetic fitness is:

$$F_1(I) = H(I) + K \sum_{c \in C} V(c, I),$$

where C is a subset of color channels and K is a constant. We can also consider pixel activity restricted to the boundaries of spheres of influence. Let ∂V and ∂H be functions analogous to V and I but with pixel calculations taken only over the *edge* pixels of the spheres of influence. This leads us to consider the aesthetic fitness measure:

$$F_2(I) = \partial H(I) + K \sum_{c \in C} \partial V(c, I).$$

Finally, we define the function $\partial E(I)$ whose purpose is to measure how accurately spheres of influence correlate with feature boundaries. For a sphere of influence S, Let $h(S)$ be the number of region transitions that occur at its horizontal edges, and let $v(S)$ be the number of region transitions that occur at its vertical edges. We set $e(S) = |h(S) - v(S)|$ provided this number does not exceed a balance threshold B, and zero otherwise. Setting $\partial E(I) = \sum_S e(S)$ leads to our final choice for an aesthetic fitness measure:

$$F_3(I) = \partial E(I) + K \sum_{c \in C} \partial V(c, I).$$

It is important to note that in the above calculations, whenever a color component of a pixel is called for the value that is actually used is the color component extracted from the *average* color of the feature region that the pixel belongs to following color segmentation.

4. Gaze Pattern Guided Image Evolution

Our initial tests using the aesthetic fitness function F_1 gave disappointing results. In most evolved images the upper half plane was one solid color and the lower half plane was filled with horizontal stripes. We concluded this was because (1) most of the gaze positions in our test data were located within a narrowly defined horizontal band in the lower half plane, and (2) we were overzealously color segmenting images. By switching to the aesthetic fitness function F_2 in order to place more emphasis on sphere boundaries, and relaxing our segmentation parameters, we consistently evolved images similar to the one shown in Figure 2. For these images the feature regions are concentrated in the left half plane. This can be explained by the fact that for our test data most of the area spanned by the spheres of influence is concentrated there. One surprise is the *style* of these images, a style characterized by having one massive feature region with high contrast. At first we thought this might be due to anomalies associated with the L channel — the luminance channel — of Lab color space. However this style persisted even when we reverted to RGB color space. Since our aesthetic fitness measure rewards images having high contrast in one or more color channels while color segmentation constrains the number of (feature) regions that can be formed, this distinctive style arises as an artifact of the Genetic Algorithm. Images are able to exploit this

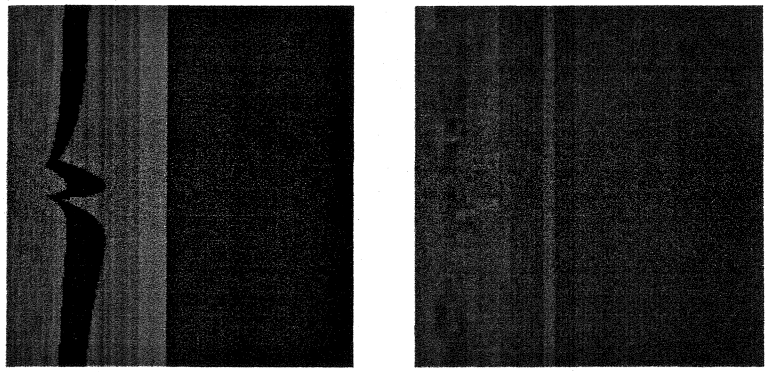

Figure 2: Left: An image appearing at generation #50. The spheres of influence are concentrated along the left hand strip. Right: An enlarged color segmented thumbnail of the image. The thumbnail has a color shift, an artifact that arose from averaging dark colors over the L channel in Lab color space.

conflict between color channel contrast and region color similarity by evolving a maximal amount of color contrast within their largest regions.

In an effort to increase the visual interest of our evolved imagery, we balanced the test data by increasing the sizes of the spheres of influence in the lower right hand quadrant. Figure 3 shows a sequence of images obtained between generations #120 and #145 during a subsequent sample run. They illustrate the speed with which image evolution is taking place.

Figure 4 shows images from a run using the aesthetic fitness function F_3. We present them in order to reinforce the observation that during the course of nearly all our runs evolution appeared to evolve the most promising aesthetic specimens between generations #50 and #150. Soon thereafter "lethal" mutations tended to show up and the dominant images become degenerate ones, characterized by horizontal banding. Mathematical visual interest diverged from human visual interest at this point.

5. Conclusions

We have implemented a technique for automating the evaluation of the aesthetic fitness of computer generated imagery based on predefined gaze patterns. Our technique can successfully guide evolution from a "primordial soup" of images to an evolutionary niche where images possess a distinctive aesthetic style. Moreover, our technique imparts a pace to their evolution. The probability of success for our technique, as measured by the proportion of evolutionary runs where interesting images were obtained, is rather low. Our system generates abstract images. Most of our images have high frequency detail. Segmentation algorithms experience difficulty when confronted with such images. The underlying artificial genetics of our generative system cause mutations to yield a steady stream of degenerate images. Putting all these facts together, it becomes easy to see why, inevitably, a lethal mutation gives rise to a degenerate image that effectively overpowers the segmenter. It has been suggested [2] that one way to attack this problem is to filter genomes

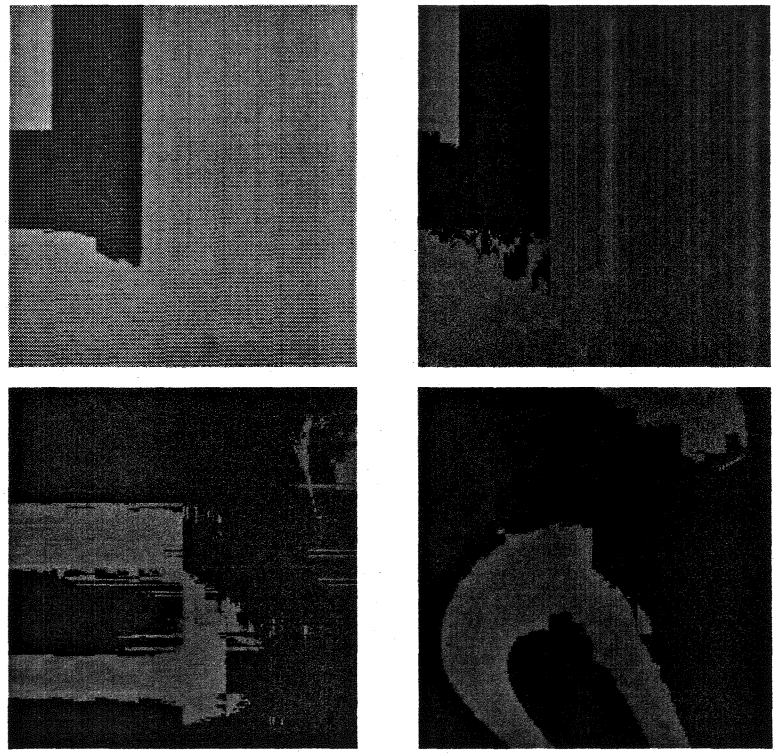

Figure 3: In row order, evolution of the most aesthetically fit images between generations #120 and #145 of a sample run.

from populations by comparing them with genomes stored in a data bank of known undesirable genomes. The current obstacle to this approach is that in most evolutionary systems based on the Genetic Algorithm the "active" components of genomes are difficult to identify whence phenotype matchings based solely on genotype comparisons turn out to be either unpredictable or unreliable.

References

[1] Brill, B., The endless wave, *2002 Bridges Conference Proceedings* (ed. R. Sarhangi), 2002, 55–66.

[2] Chao, D., Forrest, S., Generating biomorphs with an aesthetic immune system, *Artificial Life VIII Conference Proceedings* (ed. R. Standish et al), MIT Press, Cambridge, MA, 2003, 89–92.

[3] DeCarlo, D., Santella, A., Stylization and abstraction of photographs, *ACM Transactions on Graphics*, **21**, July, 2002, 769–776.

[4] Edmonds, E., Candy, L., Creativity, art practice, and knowledge, *Communications of the ACM*, **10**, October 2002, 91–95.

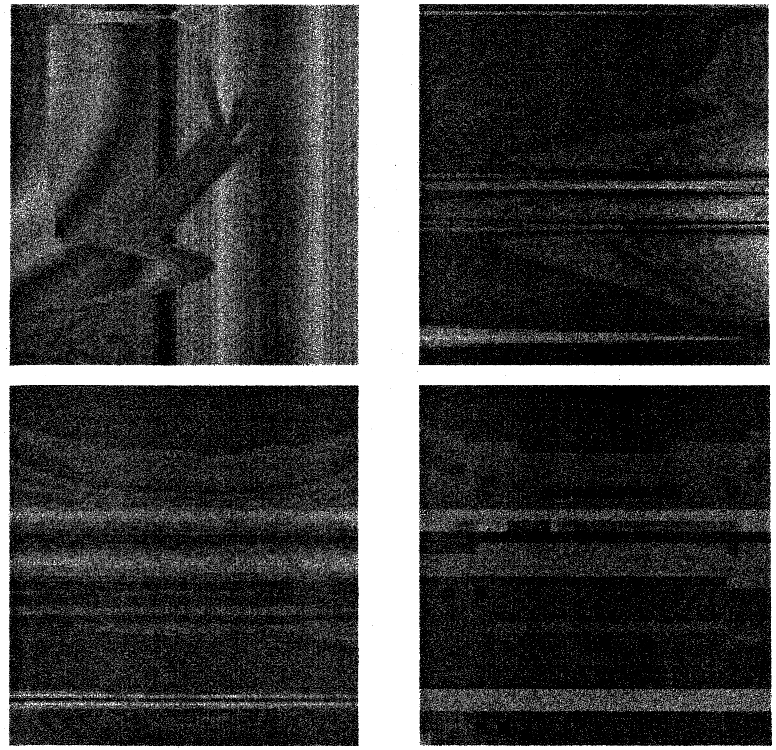

Figure 4: In row order, the most aesthetically fit images from generations #30, #105, and #525 of a sample run. Typically, the window of evolutionary interest occurred around generation #100. The bottom right enlarged thumbnail of the image from generation #525 reveals why our system judges this image to be visually interesting.

[5] Greenfield, G., Mathematical building blocks for evolving expresssions, *2000 Bridges Conference Proceedings* (ed. R. Sarhangi), 2000, 61–70.

[6] Greenfield, G., Art and artificial life — a coevolutionary approach, *Artificial Life VII Conference Proceedings* (ed. M. Bedau et al), MIT Press, Cambridge, MA, 2000, 529–536.

[7] Greenfield, G., Color dependent computational aesthetics for evolving expressions, *2002 Bridges Conference Proceedings* (ed. R. Sarhangi), 2002, 9–16.

[8] van Hemert, J., Jansen, M., An engineering approach to evolutionary art, Technical Report TR-01-01, Leiden University, 2001.

[9] Klinger, A., Salingaros, N., A pattern measure, *Environment and Planning B: Planning and Design*, **27** (2000) 537–547.

[10] Knowlton, K., `http://www.artists-nh.com/knowlton.htm`.

[11] Krawczyk, R., The art of spirolaterals, *The Millennial Open Symposium on the Arts and Interdisciplinary Computing* (eds. D. Salesin and C. Sequin), 2000, 127–136.

[12] Priebe, N., Towards a characterization of self-similar tilings via derived Voronoi tesselations, *Geometriae Dedicata*, **79** (2000) 239–265.

[13] Ruderman, D., Cronin, T., Chiao, C., Statistics of cone responses to natural images: implications for visual coding, *J. Opt. Soc. of America*, **15** (2001) 2036–2045.

[14] Sims, K., Artificial evolution for computer graphics, *Computer Graphics*, **25** (1991) 319–328.

[15] Sprott, J., The computer artist and art critic, *Fractal Horizons* (ed. C. Pickover), 1996, 77–115.

[16] Staudek, T., Machala, P., Exact aesthetics of visual patterns. Research sketch SIGGRAPH 2002, San Antonio, July 2002. (Working papers at `http://www.fi.muni.cz/~toms/publications.html`.)

[17] Unemi, T., SBART 2.4: an IEC tool for creating two-dimensional images, movies, and collages, *Leonardo*, **35** (2002) 189–192.

[18] Voss, D., Sex is best, *WIRED*, December, 1995, 156–157.