Abstract Art from a Model for Cellular Morphogenesis

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

Abstract

In this paper, we modify a mathematical model for differential gene expression introduced by Eggenberger for simulating cell morphology in order to evolve aesthetic imagery from grids of cells. We focus upon investigating fitness criteria to use so that genetic learning can effectively guide the evolution of the underlying cellular processes that lead to aesthetic results. In the model, cellular processes are governed by regulatory genes and transcription factors in such a way that cells with identical genomes exhibit differences during development. By associating certain cell products with color channels, images obtained from grids consisting of only two types of interacting cells are shown to yield a rich generative framework for artistic exploration.

1. Introduction

Modeling cellular development for aesthetic purposes first received attention thanks to Fleischer's doctoral thesis [4] and his subsequent work [5] [6]. More recently, Hoar et al [9] used a model for the life cycle of a bacterium to make images visualizing the simulated evolution of bacteria colonies which they referred to as "creative bacteria patterns." Complex models such as these that are used for aesthetic pattern generation from cellular processes supersede earlier reaction-diffusion models based on Alan Turing's seminal paper [12]. Reaction-diffusion models for generating aesthetic patterns have also been considered by many other authors [14] [13] [7] [2]. In this paper we investigate an evolutionary framework for pattern formation arising from cellular processes based on cell genomes that use regulatory genes. The genomics and cellular developmental model follow [3]. After presenting our model, our principal focus is on the problem of formulating aesthetic fitness criteria needed for implementing genetic learning in such as a way that it maximizes the aesthetic potential of the evolved imagery. Due to its subjective nature, the problem of non-interactively guiding evolution for aesthetic purposes is not well-studied. Previous approaches that are of interest include those using neural nets [1][10], co-evolution [8], and statistical analysis [11].

2. Cell Genome

In this section we give a formal description of the genomes we use for our cells. The key idea is that a genome consists of structural genes and regulatory genes, and that sequences of regulatory genes affect, and are affected by, immediately adjacent sequences of structural genes. Formally, we define a *gene* to be a string $g_0 g_1 \dots g_7$ of digits. The last digit g_7 is called the *marker* of the gene. Markers may assume any of the values zero through six, but all other digits are constrained to lie in the range zero through four. For each gene, we calculate an *offset* $o = g_0 + g_1 \bmod 3$, a *diffusion coefficient* $d = (g_2 + g_3)/9$, and a *type* $t = g_4 + g_5 + g_6 \bmod 5$. A *unit* is a sequence of one or more genes whose final gene has marker five *concatenated* with a sequence of one or more genes whose

final gene has marker six. Genes within the first segment are designated *regulatory* genes while genes within the second segment are designated *structural* genes. The purpose of this definition is to make it possible to "read" any sequence of genes and identify functional units simply by scanning the sequence, segmenting on the basis of markers, and locating adjacent regulatory and structural segments. Formally, then, a gene *unit* is a sequence of genes of the form $R_1R_2 \dots R_uS_1S_2 \dots S_v$ where each R_i is a regulatory gene, and each S_j is a structural gene, but only R_u and S_v have the requisite markers. Later we will restrict our attention to genomes with a fixed number of gene units, each consisting of precisely two regulatory genes and one structural gene, but for the time being we continue to maintain full generality in order to describe how cellular processes depend on gene units.

In nature, individual cells maintain concentrations of transcription factors denoted TF's. When a structural gene is activated, or expressed, its type determines the resulting cell "products," or morphogens, that it produces. Cell products, by affecting the concentrations of TF's, lead to higher order cellular processes. Due to our incomplete understanding of the underlying effects of chemical reactions within cells, and in light of the inherent complexity of modeling chemical reactions, following [3], for the purpose of simulation we introduce a simplification by using a structural gene's type to directly initiate higher order cellular processes when it is expressed. In [3] examples of higher order processes included cell division, cell death, creation of special molecules, creation of new TF's, etc. However, in our model when a structural gene is expressed the result is a change in the concentration of one, and only one, of its cell's TF concentrations and, as will be explained shortly, in some situations a change in the concentration of that same TF in its neighboring cells.

3. Gene Activation

Given a gene unit $R_1R_2 \dots R_uS_1S_2 \dots S_v$, in order to determine whether a structural gene S from the unit is active, consider a fixed gene R_j from the regulatory sequence of the unit and the i-th TF of the cell. Using the offset o of S, first extract the five digit string from R_j beginning at position o, perform a base five conversion, and then subtract the result from the *weight* w_i of the TF thereby obtaining the *affinity* f_i of R_j for the i-th TF. The weight w_i an environmental quantity that for us is constant across all cells. Note that since weights are allowed to be negative, f_i is a signed quantity. Now, multiply the affinity f_i by the concentration c_i of the i-th TF, and then sum over all TF's to obtain the activity level r_j for R_j. Next, sum r_j over all regulatory genes to associate to S the quantity, $a = 1/(1 + \exp(\Sigma_j r_j))$. Finally, determine the activity level γ for S by setting

$$\gamma = \begin{cases} -1.0 & \text{if } a < 0.2 \\ +1.0 & \text{if } a > 0.8 \\ 0.0 & \text{otherwise} \end{cases}.$$

To make allowances for the inherent complexity, we make a further simplification by saying that the structural gene S is expressed in an *excitory* state if $\gamma = +1$, expressed in an *inhibitory* state if $\gamma = -1$, and not expressed otherwise. This simplification makes it easier to simultaneously manage both the increases and decreases of TF concentrations within cells. With reference to [3], we should point out that it is not clear precisely how Eggenberger makes use of the three different values γ assumes, and also point out that his model is more sophisticated than ours because his calculation for r_j sums only over a "current list" of TF's and his cell products are capable of dynamically adding and removing TF's from this current list.

4. Cell Development

In this paper we use four TF's named Red, Blue, Green, and Communication. We devote one structural gene to each TF and two regulator genes to each structural gene. Thus a cell consists of a genome constructed from four gene units, with three genes per unit, together with concentrations of each of the four TF's. When expressed, a structural gene either increases or decreases the concentration of the TF it affects. The extra feature in our model is that if the concentration in the cell of the *communication* TF is sufficiently high then the concentrations in neighboring cells of the TF under consideration will also change. It is under these circumstances that the diffusion coefficient d of the structural gene comes into play. It is used to determine what proportion of the increase (respectively decrease) of the affected TF's concentration the structural gene's cell will receive, and what proportion of the increase (respectively decrease) in the affected TF's concentration all the neighboring cells will receive. More precisely, if structural gene S with diffusion coefficient d is required to change the concentration of a TF by the amount Δ, then the cell's allotment of that change will be $100(1-d)\%$ while the remaining $100d\%$ of that change will be equally distributed among the eight neighboring cells. When cells are organized in a grid, cellular development occurs over time by initializing TF concentrations for all cells and then simulating the gene activation, TF concentration update cycle for all the cells for a prescribed number of time steps.

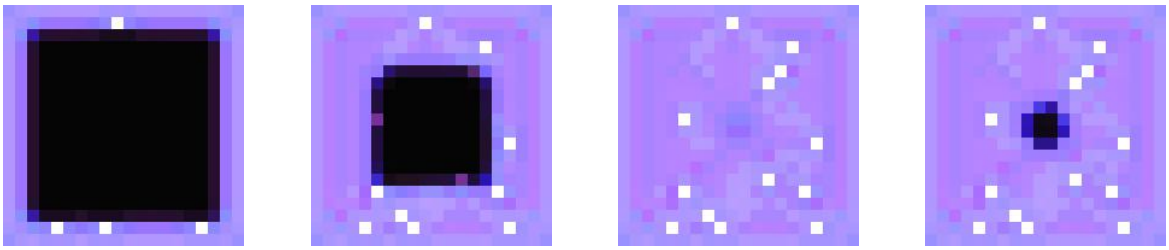

Figure 1: *An example showing the outside-in development of the cell pattern obtained after 50, 150, 250, and 350 time steps.*

5. Aesthetic Cell Development Patterns

By strategically locating a few cells with different genomes in the grid of cells and by strategically initializing the TF concentrations of those cells, the simulation of the cell development algorithm will yield a matrix of cells that can be visualized by interpreting the Red, Green, and Blue TF concentrations as color channel values. Our objective is to control the factors responsible for creating such cell patterns in such a way as to yield aesthetic imagery. To make this task more manageable, we consider only square grids and permit only two different genomes to be assigned to the cells within the grid. The two genomes are used to distinguish between cells that are either *specialized* cells or *substrate* cells. A small, but fixed percentage of specialized cells are randomly scattered in the grid and then the grid is filled in with substrate cells. Initial TF concentrations for the cells are determined by imagining that a morphogen gradient is being applied to the grid. This means that edge cells receive small, but fixed nonzero initial concentrations for their TF's, while interior cells have all their TF concentrations set to zero. The result is that cellular development occurs slowly starting at the edges and working towards the center as shown in the time series for a 20×20 cell example in Figure 1. This further simplifies our task by only requiring us to decide what genomes to use for the two different cell types and deciding how many time steps cellular

development should be allowed to proceed for. Since a suitable number of time steps can be quickly determined experimentally once the grid dimensions are fixed, this only leaves the problem of finding appropriate cell genomes. Our approach to this problem is to use a genetic algorithm and let genetic learning participate in the solution.

6. Genetic Learning

To implement genetic learning using the simple genetic algorithm we consider a population of grids. The genotype of a grid is determined by the genomes for its two cell types, and the phenotype of a grid is the visualization of the matrix of cells following cellular development as described above. Initial populations use randomly generated genomes. At the start of each run of the genetic algorithm a randomly generated placement scheme for locating cells on a grid is fixed so that all grids in the population will have their specialized cells and substrate cells identically positioned. To mate two grids, we mate their two cell types — specialized cell genomes to specialized cell genomes and substrate cell genomes to substrate cell genomes — by invoking one-point crossover followed by point mutation on a gene by gene basis. To form breeding pairs of grids, we make random selections from a breeding pool consisting of the most fit grids, typically the top four. Due to the computational load, population sizes are small, typically 6-12, and the number of generations allotted for genetic learning to occur is short, typically 5-20.

To determine grid fitness, we make use of quantities measured during the final cycle of cell development. To help define these quantities we use the subscripts R, G, B, and C when referencing the four TF's. Let μ_i and σ_i denote the mean and standard deviation calculated over all the cells in the grid of the concentration for the TF subscripted by i. Let N_a denote the number of cells that had a change in the activation status of at least one structural gene during the final development cycle, and let N_b denote the number of cells in the grid that are dormant, meaning their R, G, and B concentrations are all below trace levels whence they appear black in the visualization.

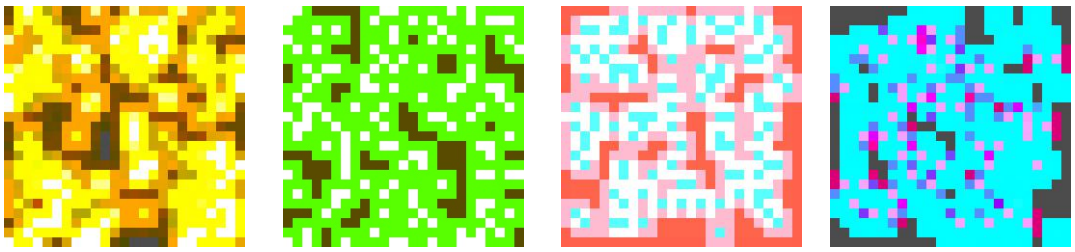

Figure 2: *Examples using fitness functions: (a) F_1 with $\delta = 1$ which rewards both communication among cells and variability within all color channels, (b) F_1 with $\delta = 0$, (c) F_2 which rewards patterns exhibiting variability within at least one color channel, and (d) F_2 giving rise to a pattern that is unusual because it has both dormant and active substrate cells on its border.*

Initial tests revealed that the fundamental obstruction to achieving aesthetic imagery using the genetic algorithm was the presence within the population of grids with either too many dormant cells or with monochrome colorings. Denoting by $F(P)$ the fitness of grid P, the first reliable fitness function we discovered calculated fitness using

$$F_1(P) = {\sigma_C}^{\delta} \cdot \min(\sigma_R, \sigma_G, \sigma_B),$$

where δ is zero or one. This fitness function rewards grids that exhibit various communication behaviors between neighboring cells and that possess highly variable color channels. It was used to evolve the 25×25 cell patterns in Figure 2a and Figure 2b. In fact, patterns such as those found in Figure 2c and Figure 2d, obtained by changing the fitness function to

$$F_2(P) = \max(\sigma_R, \sigma_G, \sigma_B),$$

showed that by ignoring the communication TF and rewarding those grids where *at least* one color TF showed significant variability we could obtain promising results.

The principal drawback to the two fitness functions introduced so far is that although they consistently produce interesting results, they do not exert sufficient evolutionary pressure on initial populations of grids where a preponderance of grids have large numbers of dormant cells. This is explained by the fact that a grid with a large number of black cells (indicating no TF activity) or white cells (indicating maximal TF activity) that are set off against a background color produced by active substrate cells can lead to misleading standard deviation measurements. Various fitness functions we designed incorporating color channel averages and the "activity" measure N_a failed to alleviate this problem. To overcome this difficulty we used the dormancy measure N_b to introduce a term for penalizing such patterns. This led us to formulate our two most successful fitness functions

$$F_3(P) = \frac{N_a \cdot \min(\sigma_R, \sigma_G, \sigma_B)}{1 + N_b},$$

and

$$F_4(P) = \frac{\sigma_C \cdot N_a \cdot \min(\sigma_R, \sigma_G, \sigma_B)}{1 + N_b}.$$

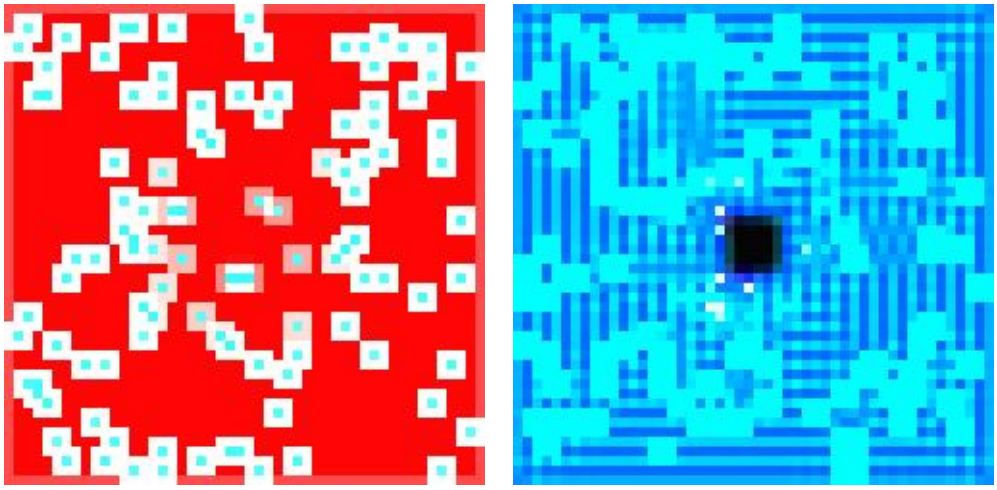

Figure 3: *(a) An example using fitness function F_3, which incorporates a penalty term for black, dormant cells, and (b) a high resolution example using fitness function F_4, which also incorporated a penalty term for black, dormant cells.*

Not only did these two functions lead to an increase in the number of interesting patterns that were evolved, but they allowed us to increase the dimensions of our images while simultaneously decreasing the percentage of specialized cells we used. This, in turn, led to a better understanding of how the underlying cellular processes that were occurring within our cells functioned. The images in Figure 2 used 20% specialized cells. Figure 3a shows an example of a 40×40 pattern with only

5% specialized cells after only 300 developmental time steps that was evolved using the fitness function F_3, while Figure 3b shows a 50×50 pattern, also using only 5% specialized cells, after only 400 developmental time steps that was evolved using the fitness function F_4.

7. Conclusions

We have adapted a model for simulating cellular morphogenesis in order to generate aesthetic cell patterns and we have investigated automating the search for such patterns using a genetic algorithm. Initial results are encouraging. Given that random placement of only two different types of cells within a grid was used to initiate the pattern formation process, future work will consider the use of "template" placement schemes using additional cell types and the use of other grid geometries so that cell interactions between tissue types can be simulated and visualized.

References

[1] Baluja, S., Pomerleau, D., and Jochem, T., Towards Automated Artificial Evolution for Computer-Generated Images, *Connection Science*, **6**, 1994, 25–354.

[2] Behravan, R., and Carlisle, R., Interactive organic art, *Proceedings of the Seventh Interantional Conference on Generative Art*, AleaDesign, 2004, Vol. 1, 239–246.

[3] Eggenberger, P., Evolving Morphologies of Simulated 3d Organisms Based on Differential Gene Expression, *Proceedings of the Fourth European Conference on Artificial Life (ECAL97)*, Springer Verlag, 1997, 205–213.

[4] Fleischer, K.W., *A Multiple-Mechanism Developmental Model for Defining Self-Organizing Structures*, Ph.D. Dissertation, Caltech, Department of Computation and Neural Systems, 1995.

[5] Fleischer, K.W., Laidlaw, D.H., Currin, B.L. and Barr, A.H., Cellular Texture Generation, *SIGGRAPH 95 Conference Proceedings August 6-11, 1995*, ACM Press, 239–248.

[6] Fleischer, K.W., Investigations With a Multicellular Developmental Model, *Artificial Life V Conference Proceedings*, MIT Press, 1996, 229–236.

[7] Greenfield, G.R., Case Study: A Sculptor - Programmer Collaboration, Technical Report TR-94-03, Dept. of Mathematics & Computer Science, University of Richmond, 1993.

[8] Greenfield, G.R., Art and Artificial Life — A Coevolutionary Approach, *Artificial Life VII Conference Proceedings*, eds. M. Bedau et al, MIT Press, 2000, 529–536.

[9] Hoar, R.M., Penner, J.K., and Jacob, C., Transcription and Evolution of a Virtual Bacteria Culture, *2003 Congress on Evolutionary Computation Proceedings*, IEEE Press, 2003,54–61.

[10] Machado, P. and Cardoso, A., Computing Aesthetics, *Proceedings XIV-th Brazilian Symposium on Artificial Intelligence SBIA'98, Porto Allegre, Brazil*, ed. F. Oleiveira, Springer-Verlag, 1998, 219–229.

[11] Staudek, T., Computer-Aided Aesthetic Evaluation of Visual Patterns, *ISAMA/Bridges 2003 Conference Proceedings*, eds. J. Barrallo et al, University of Granada, 2003, 143–150.

[12] Turing, A.M., The Chemical Basis of Morphogenesis, *Philosophical Transactions of the Royal Society of London Series B*, **237**, 1952, 37–72.

[13] Witkin, A. and Kass, M., Reaction-Diffusion Textures, *SIGGRAPH 91 Conference Proceedings*, ACM Press, 1991, 299–308.

[14] Young, D.A., A Local Activator-Inhibitor Model of Vertebrate Skin Patterns, *Theory & Application of Cellular Automata*, ed. S. Wolfram, World Scientific, 1986, 320–327.