An Empirical Exploration of a Definition of Creative Novelty for Generative Art

Taras Kowaliw, Alan Dorin, and Jon McCormack

Centre for Electronic Media Art, Faculty of Information Technology,
Monash University, Clayton 3800, Australia
taras@kowaliw.ca, {alan.dorin,jon.mccormack}@infotech.monash.edu.au
http://www.csse.monash.edu.au/~cema/

Abstract. We explore a new definition of creativity — one which emphasizes the statistical capacity of a system to generate previously unseen patterns — and discuss motivations for this perspective in the context of machine learning. We show the definition to be computationally tractable, and apply it to the domain of generative art, utilizing a collection of features drawn from image processing. We next utilize our model of creativity in an interactive evolutionary art task, that of generating biomorphs. An individual biomorph is considered a potentially creative system by considering its capacity to generate novel children. We consider the creativity of biomorphs discovered via interactive evolution, via our creativity measure, and as a control, via totally random generation. It is shown that both the former methods find individuals deemed creative by our measure; Further, we argue that several of the discovered "creative" individuals are novel in a human-understandable way. We conclude that our creativity measure has the capacity to aid in user-guided evolutionary tasks.

1   Introduction

A recent definition of creativity recasts it as "a framework that has a relatively high probability of producing representations of patterns that can arise only with a small probability in previously existing frameworks" [2]. An interesting property of this definition is that it depends neither on notions of value nor appropriateness. These properties, of course, set it at odds with common usage of the term "creative", since the perceived creativity of a system is often culturally biased, associated with interest, or affected by context. Dorin and Korb consider the value of a value- and appropriateness-free definition of creativity extensively, and respond to some obvious criticisms. Regardless of common usage, the ability to capture even an aspect of creative novelty in an objective measure is enticing; It affords us the possibility of (a) empirically testing the consequences of the definition, and it's value to preconceived notions of creativity; and (b) suggesting a stream of new and, ideally, interesting frameworks to users. In this latter motivation, our work resembles the creation of an iterative fitness function [6], except that rather than provide exemplars of an optimal goal state, our notion of creativity supports a practically unlimited variation.

Here we explore the creativity of a *system* — something or someone that accepts an input and generates, possibly stochastically, a pattern or phenotype — rather than the creativity of a particular pattern or representation. An approach which selected for previously unseen representations or patterns would likely need to incorporate a notion of distance (since in a sufficiently rich space, any randomly-generated entity is almost certainly previously unseen), selecting for new entities "far" from previously seen entities. However, the notion of distance is problematic in many domains, due to a lack of meaningful metrics or unusual statistical distributions. Additionally, such a system would likely be biased towards statistical outliers, which could confuse an exploration/exploitation-based search technique.

We will interpret the Dorin/Korb definition as a simple boolean relation distinguishing between systems that can *reliably* generate some collection of patterns, as opposed to a system exceedingly unlikely to do so; This will allow us to side-step many of the difficulties associated with the use of genotypic or phenotypic similarity measures.

2   Formalization

In order to make this definition rigorous and practical, we need to qualify several aspects, and introduce some restraints. We will aim for systems which generate grayscale images, but generalization should be straightforward. Let us consider a space of patterns (phenotypes), $p \in P$. For simplicity, we will assume that the space of patterns is bounded, which excludes new modalities or ever-increasing scales. Since patterns are often too high-dimensional to deal with directly, we instead introduce a feature space of finite dimension, $\mathbb{F} = \{F_1, ..., F_k\}$ through which we can characterize the space of patterns: $p \rightarrow (F_1(p), ..., F_k(p)) \in \mathbb{R}^k$. We will assume only that connected regions form perceptually similar segments of phenotypic space, but assume nothing regarding volume, distribution, nor the significance of distance. Although our capacity to select a representative sample is ultimately governed by the representation space, the spread of values will be strongly affected by the number of features used, recovering the "curse of dimensionality" which plagues machine learning tasks; to prevent this, in practice, we will use only two features at a time.

2.1   Images and Image Features

We now turn our attention to systems which generate 8-bit grayscale images of a fixed size[1] of 200×200 pixels. Image similarity is a notoriously difficult problem, and these images can be considered points in a 40000-dimensional space. In order to characterize this pattern space, we select a set of features drawn from image processing in hope of drawing features perceptible to humans.

---

[1] To the reader who believes that a bounded space is insufficiently rich to generate a practically endless variation of patterns, we refer the reader to McCormack's discussion of generating simple images [7].

Following work in pattern classification, we use several statistical moments to characterize the overall form of the images (as in, for instance, [5]): specifically, we use geometric moments $M_{00}, M_{01}, M_{10}, M_{20}, M_{02}$, and M_{11}.

Following examples from content-based image retrieval, we also use histograms to characterize the space (as in, for instance, [8]). The image is transformed by an edge-detection measure (Laplacian convolution) and the normalized area is calculated. Several histogram-related measures are computed for both the original image and the edge-detected version: the maximum, mean, and standard deviation, and the entropy.

Finally, we use Grey Level Co-occurrence Matrices (GLCMs) to include texture measures into our feature set [3]. Specifically, we measure several statistical properties of the normalized co-occurrence matrix $P = \{p(i,j)\}$, chosen due to their demonstrated efficacy in content-based image retrieval [4]. These are Energy: $GE = \sum_{i,j} p^2(i,j)$; Entropy: $GI = \sum_{i,j} p(i,j) \log p(i,j)$; Contrast: $GC = \sum_{i,j} (i-j)^2 p(i,j)$; and Homogeneity: $GH = \sum_{i,j} \frac{p(i,j)}{1+|i-j|}$.

Overall, we considered a total of 18 features representing the overall form, histogram, and texture of the image and its edges. Following some initial experimentation, we selected six which support a rich characterization of phenotypic space (that is, were capable of measuring specialization during our informal experimentation of the space of images): M_{11}, M_{02}, M_{01}, *edge-area*, *histogram-entropy*, and the GLCM homogeneity measure.

2.2   Detecting Creativity-Indicating Regions

Let S_1 and S_2 be systems which map from a space of input, $x \in X$, to the space of patterns P. We wish to claim that S_2 is creative (or not) relative to S_1 on the basis of what can be reliably produced by the systems in question. We interpret this as the capacity to find a compact and connected region of feature space that can be reliably populated by system S_2 but not by system S_1. We restrict our attention to intervals since this is more likely to generate an intelligible region of space (assuming a measure of continuity in the chosen features), as opposed to the arbitrariness of general Borel sets, for example.

We now wish to formalize a notion of the capacity of a system to reliably generate a pattern, with error-tolerance τ and confidence c. We will write that an interval of feature space is $\mathbf{r} = (r_1 \pm \delta_1, ..., r_k \pm \delta_k)$, and that a point p is contained in $\mathbf{r}$, $p \in \mathbf{r}$, if it is contained within the bounds for each feature-space dimension. We aim to estimate the true probability of S_j generating a point in interval $\mathbf{r}$, $P[S_j(X) \in \mathbf{r}]$, via the frequency of sample points, written $\hat{P}$. Assuming our sample is representative, the (conservatively estimated) margin of error associated with this frequency is

$$m.e.(\hat{P}[S_j(X) \in \mathbf{r}]) = \frac{z}{\sqrt{n}} \tag{1}$$

where z is the upper critical value for confidence-level c, and n is the sample size. Let us assume that $\tau > \frac{z}{\sqrt{n}}$, i.e. that our margin of error does not dominate our error tolerance. Generally, we can achieve this by ensuring that $n > \frac{z^2}{\tau^2}$.

We will now write that $\mathbf{r}$ intersects the *reliable-support* of S_j iff

$$\hat{P}[S_j(X) \in \mathbf{r}] > \tau + \frac{z}{\sqrt{n}} \tag{2}$$

We will write that $\mathbf{r}$ is **not** in the reliable-support of S_j iff $\hat{P}[S_j(X) \in \mathbf{r}] = 0$, with no conclusions being drawn on the region in between. In the former case, we have greater than c confidence that the probability of generating samples in the region $\mathbf{r}$ is greater than τ, and in the latter, greater than c confidence that the probability is less than our margin of error.

Hence, provided with values for reliability τ and c, we seek to find a region $\mathbf{r}$ which intersects the reliable-support of S_2, but not the reliable-support of S_1. Finding such a region in a set of samples, we shall soon see, is also a matter requiring some interpretation.

This definition relies heavily on the abilities of a new system relative to some base system, a metaphor for the existing worldview of an audience. Of course, if one begins with a trivial base system, nearly any new system will appear creative.

2.3   Generation of Intervals Given Sample Data

Here we attempt to find intervals using our sample and feature space that exist in the reliable support of system S_2, but not in the reliable-support of system S_1. Using a confidence interval of $c = 95\%$, an error tolerance of $\tau = 0.03$, and a sample size of $n = 5000$, we need to find regions with at least $n(\tau + \frac{z}{\sqrt{n}}) = 292$ samples from system S_2, and none from system S_1. It is natural to base the minimal size of an interval on the standard deviation of the sample; Rather than include any such interval capable of supporting the mass of points required, we will instead require than a minimal length[2] of interval be $\beta = \frac{\sigma}{5}$.

The attempt is made to find intervals surrounding each sample point provided. If our sample pattern is $p = (F_1(p), ..., F_k(p))$, then we initially define our interval about p as

$$\mathbf{b}(p) = \left(F_1(p) \pm \frac{\beta}{2}\sigma_1^j, ..., F_k(p) \pm \frac{\beta}{2}\sigma_k^j \right) \tag{3}$$

where S_j is the system from whence sample p was drawn, and σ_i^j is the standard deviation of the system S_j in the i-th dimension.

For each such created interval, we ask if it does not contain points from S_1 (as we know it contains at least one point from S_2). If so, we attempt to generalize it. For each dimension in turn, we attempt to widen the width of the interval by a factor of 2. If we successfully define a new interval containing equal or greater points from S_2 and none from S_1, we retain the new interval. Once we have

---

[2] This number was chosen on the basis that during a search of possible values it was the smallest value leading to a small proportion of random genomes being declared creative; That is, using a smaller value we risked declaring regions resulting from a particularly impotent application of mutation (i.e. mutations which had no effect on phenotype) as creative.

traversed all dimensions, if our new interval is an improvement on the original, we traverse the dimensions again. This process will terminate either when a locally maximal interval is found, or when the interval covers the entirety of all dimensions. Finally, the best found interval is tested to see if it is in the reliable-support of S_2.

Since searching for our formally defined creativity is a slow process, we also introduce a multi-valued procedure for estimating relative creativity quickly. **Creativity Lite** will take a smaller sample, and return the maximum number of samples from set S_2 that can be found in some region containing no samples from S_1. We use a sample of points from set S_1 of size $n_1 = 1000$, and a sample of points from set two of size $n_2 = 100$. Intervals are constructed using the technique described above.

2.4   Model Limitations

There are several limitations to the presented model. The first, already noted, is the reliance on constructing a representative sample of pattern space, which sets practical limits on the number of features and maximal pattern sizes.

A second potential limitation is the pre-selected feature space. Given any two non-identical systems, it is likely always possible to generate some feature which separates them. Clearly, the choice of a trivial feature in feature space will allow for the discovery of a trivial form of creativity; Our choice of features introduces a notion of appropriateness alongside our originally appropriateness-free definition of creativity.

Finally, we note the reliance on several predetermined values — for certainty, minimal intervals, and interval discovery — likely to be dependent on the problem domain and data source distributions. As a result, we can never say, with measurable certainty, that S_2 is **not** creative relative to S_1.

3   Biomorphs

Biomorphs were introduced by Dawkins as a simple example of evolutionary search [1]. Here we present our own interpretation of Dawkins' original (and vaguely specified) biomorph growth process.

A genome consists of eleven genes, $g = (g_1, ..., g_{11})$, each a floating point number in the range $[0, 1]$. Initialization of genes is random and uniform. The generated form has a number of properties, applied recursively: A translation $(x_{trans}, y_{trans}) = (20g_1 - 10, 20g_2 - 10)$; A thickness $t = 3g_3 + 1$ and thickness variation $\Delta t = \frac{2g_4}{3}$; A magnitude $m = \frac{g_5 w}{2}$ (where w is the maximum dimension of the drawing surface) and magnitude variation $\Delta m = \frac{2g_6}{3}$; A branch angle $\alpha = \pi g_7$ with an angle variation $\Delta \alpha = 0.4g_8 + 0.8$; A branching factor $n_{kids} = \lfloor 6g_9 + 1 \rfloor$ and a branch depth variance, $\Delta n_{kids} = 0.4g_{10} + 0.8$; And a recursion depth, $n_{rec} = \lfloor 4g_{11} + 2 \rfloor$. These particular properties were chosen through informal visual search on the space of random genomes as able to generate a breadth of interesting phenotypes.

Biomorph drawing is a simple recursive procedure. Given a genome, a starting location (x_{start}, y_{start}), a current magnitude m', a current thickness t', a current number of children n'_{kids}, and a current angle α', we draw a line; This is drawn between the given start location and an end location calculated as

$$(x_{end}, y_{end}) = (x_{start} + x_{trans} + m' \cos(\alpha'), y_{start} + y_{trans} + m' \sin(\alpha')) \quad (4)$$

At the end point, if we have not exceeded n_{rec} recursive steps, we create $n''_{kids} = n'_{kids} \cdot \Delta n_{kids}$ child lines, each pointing $\alpha'' = \alpha' \cdot \Delta \alpha$ degrees from the original direction, with a magnitude $m'' = m \cdot \Delta m$, and a thickness of $t'' = \min\{1, t' \cdot \Delta t\}$. We initialize a biomorph at the central point, drawing n_{kids} lines at integer multiples of α around the $\pi/2$ axis. A selection of interesting but randomly generated biomorphs can be seen in Figure 1.
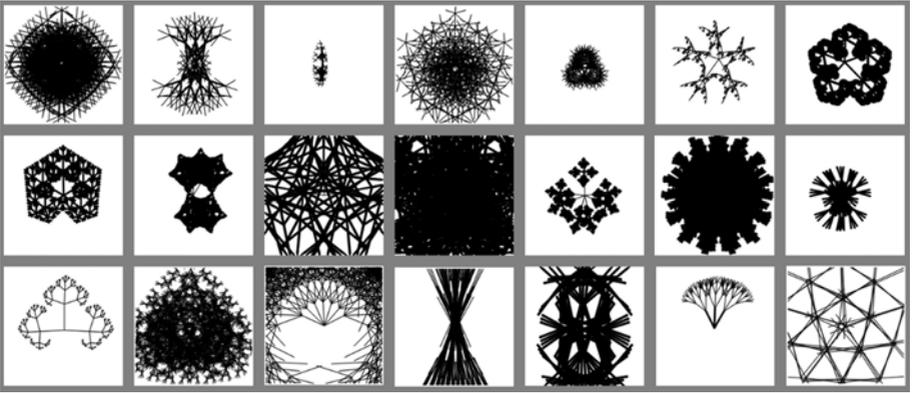

Fig. 1. A selection of randomly generated biomorphs, chosen to show some of the diverse structures possible

3.1   Discovery of Creative Biomorphs via Interactive Evolution

Biomorphs are easily evolved using an Interactive Evolutionary Algorithm (IEA). Here we use a simple $(1 + 8)$-ES, where mutation selects two randomly chosen genes and mutates both with a variance of 0.16.

Our initial system is a wide collection of randomly generated genomes. We use 14400 samples in two-dimensional feature space, effectively assuming that 120 samples is sufficient to represent each dimension. Experimentation with computed descriptive statistics over several runs convinces us that this is a sufficient choice. This pool is considered as our basic worldview (S_1 above), the yardstick from which we will measure creativity in generative systems. Note that duplicate feature values are discarded; In practice, this means an additional 20% generation time.

Given any particular genome, we can consider it a generative system by considering it as a seed from which we can generate mutated children. That is, given
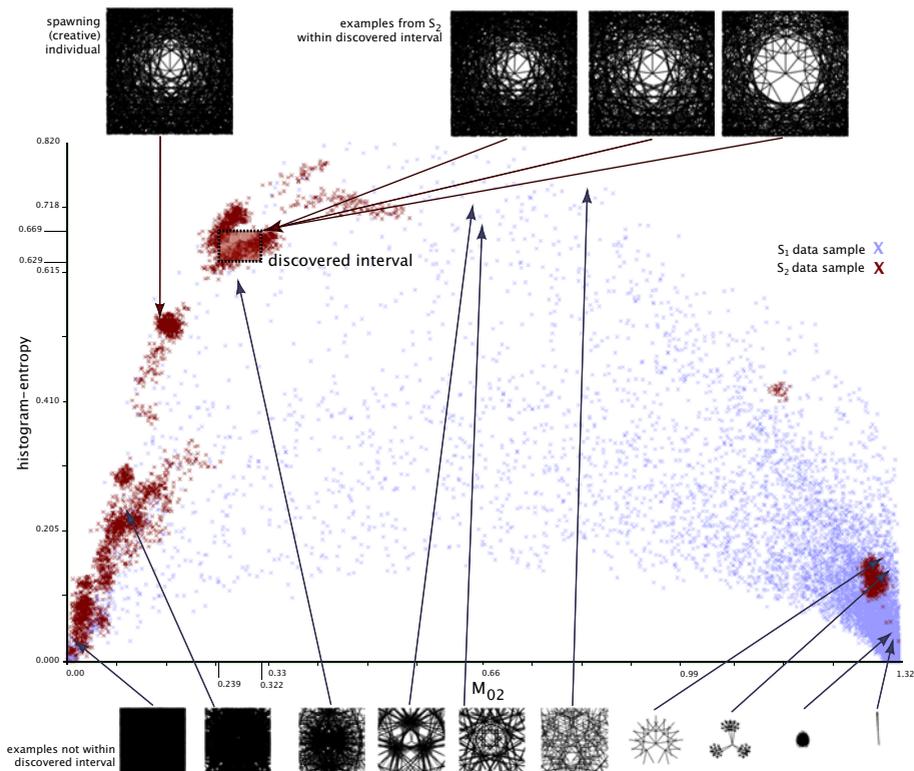
Fig. 2. *(top)* The spawning individual's phenotype and a collection of phenotypes in the discovered interval; *(middle)* a view of feature-space and sample distribution (data points have been slightly "jittered" for legibility); *(bottom)* a random sample of phenotypes not in the discovered interval

some genome g, we will generate a set of children $S_g = \{m(g)\}_k$, consisting of k mutated versions of the original genome. We will discard any children resulting from an impotent mutation (in practice, requiring an additional 10% generation time).

We use three combinations of pairs of features to evaluate the system. For each, several search strategies are evaluated on their capacity to produce creative phenotypes. The first search strategy was *totally random*, where a single random genome is evaluated. The next two search strategies are variants on evolutionary search. In each, an initial population of nine random genomes is created, and ten generations are explored by selecting a single agent and spawning an additional eight mutants to form the next population. The first evolutionary strategy is *IEA*, where the individual is selected subjectively by a human operator[3]; the

---

[3] Human operators were selected from five people at our research lab, including the authors, and asked to select for aesthetically pleasing biomorphs. Participants were not aware of creativity scores while making selections.

second is the *creativity search*, where each population member is evaluated using the creativity-lite function, and the maximum such value guides the choice of individual.

We computed the proportion of individuals termed creative for a sample of runs for each strategy. Each strategy was run forty times (thirty for the IEA runs), using one of three pairs of features. These proportions were:

search type	M_{02}, *edge-area*	M_{01}, *GLCM-H*	M_{11}, *histogram-entropy*
totally random	0.067	0.025	0.000
IEA	0.433	0.233	0.700
creativity search	0.800	0.475	0.733

The probability of a totally random individual being termed creative was very small, which we consider further evidence of the statistical sample being representative. Via creative-search, relatively high proportions of evolved genomes are termed creative. It appears that creativity-lite does indeed serve as an approximation of our more formal notion of creative novelty, allowing for a faster approximate search. The proportion of individuals termed creative discovered via IEA lay consistently between the proportions for the totally random and creative search techniques. This corresponds to the general intuition that humans tend to seek out novelty, but also that aesthetic pleasure and novelty are not the same thing. It is also evident that it is easier to discover creative regions using some combinations of features than others; Using the M_{01}, *GLCM-H* combination, only two rough patterns of creative regions were discovered — namely, regions comprised of homogeneous and top-heavy trees, and regions of very high non-homogeneity — while for the other two explored feature combinations many different creative regions were discovered.

An example of an easily understood discovered creative individual is illustrated in Figure 2. This individual was discovered using creative search and the M_{02} and *histogram-entropy* features. The interval discovered is characterized by a high *histogram-entropy* value, and a low M_{02} value. The high *histogram-entropy* value is difficult to discover in a line drawing, since most pixel values are either black or white; The only means of obtaining a high value is to include many spaced-out diagonal lines on a variety of angles, and rely on the anti-aliasing of the line drawing routine. Simultaneously, we require a low M_{02}, meaning that both the top and the bottom of the image must be mostly black in colour. Satisfying both simultaneously is non-trivial — since it is difficult to space out black lines on a white background and minimize the white content of the image — and thus a rarity; This individual satisfies both by creating patterns with a blank hole in the centre, surrounded by an increasing density of angled lines.

Generally speaking, the examples of individuals termed creative by the system ranged from (in the authors' opinions) trivial to interesting. Several additional interesting generators are illustrated in Table 1; Trivial examples include minor variations on these themes (such as selection for slightly smaller or larger circles, or a slightly larger or smaller number of diagonal lines), or examples of individuals that have discovered regions of trivial mutation (for instance, during the M_{02}, *edge-area* runs, some regions consisting of a very large number of near-identical

Table 1. Examples of discovered "creative" individuals, including the phenotype of the *(left)* spawning individual; *(middle)* an averaged image of all children in the discovered interval; and *(right four)* some examples of those children. The discovery technique and the reason for the individual being termed creative below.
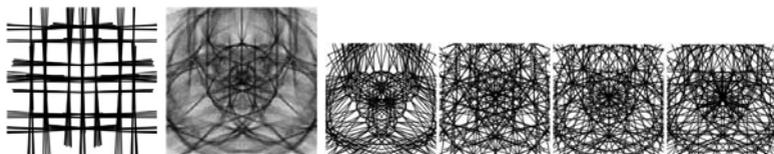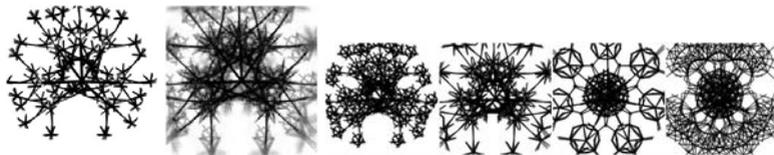

Discovered via IEA using features M_{02}, *edge-area*. Slightly bottom-heavy (difficult in a system which automatically begins with branches pointing upwards), high number of edges (i.e. spaced out lines).


Discovered via creativity-search using features M_{11}, *histogram-entropy*. A collection of dark shapes, solid black in the centre with a fuzzy boundary. The fuzzy shapes create a small but non-zero histogram entropy by having a small portion of the image devoted to many different angles of line, but a consistent mass via area of shape
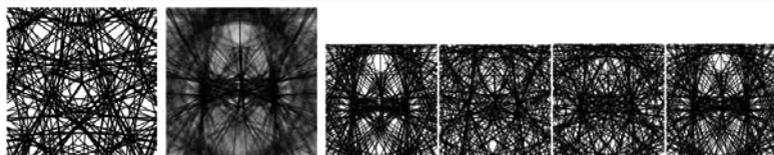
Discovered via creativity-search using features M_{01}, *GLCM-H*. Selection for near-perfect symmetrical mass over the x-axis and a very large number of colour transitions.

Discovered via IEA using features M_{02}, *edge-area*. Selection for specific amounts of edges, mass.
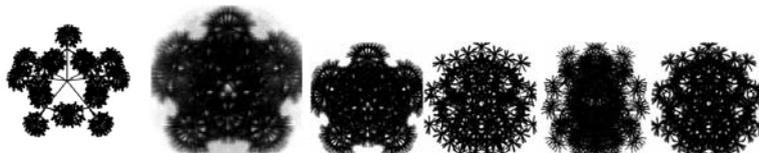
Discovered via IEA using features M_{11}, *histogram-entropy*. Selection for both a mid-level entropy (i.e. a fair number of diagonal lines at different angles) and a low but non-zero moment calculation. The moment calculation implies near-perfect symmetry in both the x- and y-axis. Children achieve this by having perfect symmetry in one axis, and near-symmetry in the other, some using the x and others the y.

patterns were found, relying on the invariance of both features to translation in the x-axis). Regardless, the tool has demonstrated its ability to find genuinely improbable and variant regions of space which occasionally correspond to human interest; Hence, it may be sufficient to remove some of the burden of search from human operators in an IEA.

4   Conclusions

In this paper we have demonstrated that the Dorin / Korb definition of creativity can be interpreted in such a way that makes it both tractable generally, and suitable for generated images. We have further shown that the definition can be integrated into an interactive evolutionary algorithm: Firstly, by treating individual genomes, along with their evolutionary operators, as pattern generators; And secondly, by considering an approximate and fast version of our creative-novelty measure, "creativity lite". It is shown that maximization of creativity lite generally leads to a system which is creative by our definition.

Several genomes have been discovered which do, indeed, generate regions of space which are highly unlikely through random generation. While some of these regions seem trivial to human operators, others are human-recognizably distinct, which we interpret as evidence that the system can detect human-recognizably novel generators; In conjunction with the work above, we believe that this measure can be used to suggest new and interesting directions for users of an IEA.

References

1. Dawkins, R.: The Blind Watchmaker. Longman Scientific & Technical (1986)
2. Dorin, A., Korb, K.: Improbable creativity. In: McCormack, J., Boden, M., d'Inverno, M. (eds.) Proceedings of the Dagstuhl International Seminar on Computational Creativity. Springer, Heidelberg (2009)
3. Haralick, R.: Statistical and structural approaches to texture. Proceedings of the IEEE 67, 786–804 (1979)
4. Howarth, P., Rüger, S.: Evaluation of texture features for content-based image retrieval. In: Enser, P.G.B., et al. (eds.) CIVR 2004. LNCS, vol. 3115, pp. 326–334. Springer, Heidelberg (2004)
5. Kowaliw, T., Banzhaf, W., Kharma, N., Harding, S.: Evolving novel image features using genetic programming-based image transforms. In: IEEE CEC 2009 (2009)
6. Machado, P., Romero, J., Manaris, B.: Experiments in computational aesthetics: An iterative approach to stylistic change in evolutionary art. In: Romero, J., Machado, P. (eds.) The Art of Artificial Evolution: A Handbook on Evolutionary Art and Music, pp. 381–415 (2008)
7. McCormack, J.: Facing the future: Evolutionary possibilities for human-machine creativity. In: Romero, J., Machado, P. (eds.) The Art of Artificial Evolution: A Handbook on Evolutionary Art and Music, pp. 417–453 (2008)
8. Vailaya, A., Figueiredo, M.A.T., Jain, A.K., Zhang, H.-J.: Image classification for content-based indexing. IEEE Transactions on Image Processing 10(1), 117–130 (2001)