Swarm and Evolutionary Computation 16 (2014) 52–68

Contents lists available at ScienceDirect

Swarm and Evolutionary Computation

journal homepage: www.elsevier.com/locate/swevo

Regular Paper

Investigating aesthetic measures for unsupervised evolutionary art

Eelco den Heijer [a,b,*], A.E. Eiben [b]

[a] Objectivation B.V., Amsterdam, The Netherlands
[b] VU University Amsterdam, Faculty of Sciences, de Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 19 June 2013
Received in revised form
19 December 2013
Accepted 1 January 2014
Available online 15 January 2014

Keywords:
Evolutionary computation
Genetic programming
Evolutionary art
Computational aesthetics
Multi-objective optimisation

ABSTRACT

We present an extensive study into aesthetic measures in unsupervised evolutionary art (EvoArt). In contrast to several mainstream EvoArt approaches we evolve images without human interaction, using one or more aesthetic measures as fitness functions. We perform a series of systematic experiments, comparing 7 different aesthetic measures through subjective criteria ('style') as well as by quantitative measures reflecting properties of the evolved images. Next, we investigate the correlation between aesthetic scores by aesthetic measures and calculate how aesthetic measures judge each others' image. Furthermore, we run experiments in which two aesthetic measures are acting simultaneously using a Multi-Objective Evolutionary Algorithm. Hereby we gain insights in the joint effects on the resulting images and the compatibility of different aesthetic measures.

1. Introduction

Evolutionary art is a research field that investigates the application of evolutionary computation in the creation of aesthetically pleasing images. The field of evolutionary art was instigated by 'The Blind Watchmaker' by Richard Dawkins [1], a book on biological evolution. In his book Dawkins evolved stick figures called 'biomorphs' to demonstrate the process of evolution. The idea of interactively evolving images led to the birth of evolutionary art (EvoArt), and also started interactive evolutionary computation, or IEC, as a methodology within the field of evolutionary computation.

In IEC, a human being fulfils the role of the fitness function (a function that determines the fitness of an individual in the population) and for quite some years EvoArt was closely tied to IEC, mainly because it was widely considered that aesthetic evaluation was too complex to automate. Takagi [2] provides a a good overview of IEC applied in EvoArt, evolutionary design and many other domains. Since the work of Dawkins, several researchers have successfully evolved aesthetically pleasing images [3–5] and good overviews of EvoArt are by Romero and Machado [6] and Bentley and Corne [7].

Whereas IEC has been successful in the field of EvoArt, IEC is not without its disadvantages. In a typical interactive evolutionary art system, a user is presented with a number of images, and the user has to select one or more images that may survive into the next generation. This step is repeated for a number of generations. Using this setup, a number of restrictions emerge. First of all, there is a limit of images that one could present to a user (per generation). Next, there is a limit on the number of generations that users are willing (or able) to select images. These restrictions are caused by 'user fatigue', and user fatigue is one of the fundamental 'issues' of IEC. User fatigue may lead to inconsistent evaluations by users (e.g. a user may not make the same aesthetic evaluations under similar conditions).

A natural way to circumvent the limitations in IEC is to remove the human from the loop: unsupervised evolutionary art. One of the earliest attempts at unsupervised evolutionary art was published in 1994 by Baluja et al. [8]. Baluja et al. trained a neural network to perform the aesthetic evaluation of evolved images, but the authors concluded that the results were 'unsatisfactory'. In the following years, very little work has been published on the topic of unsupervised evolutionary art, but recently the idea has been gaining traction, resulting in papers on EvoArt that use aesthetic measures as fitness functions, and on aesthetic measures in the context of Computational Aesthetics. However, many papers on aesthetic measures are not 'tested' in an EvoArt system, and many papers on unsupervised EvoArt are incomparable because they not only differ in the aesthetic measures, but also in the evolutionary algorithms, genotype representations, and statistics.

The development of unsupervised EvoArt systems may benefit from the field of 'computational aesthetics'. This research field investigates the development of functions that calculate an aesthetic value of images (and sometimes other artefacts) and are known as 'aesthetic measures'. Machado and Cardoso [5] worked

* Corresponding author at: Objectivation B.V., Amsterdam, The Netherlands.
E-mail addresses: eelcodenheijer@gmail.com (E. den Heijer), gusz@cs.vu.nl (A.E. Eiben).

2210-6502/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.swevo.2014.01.002

on their well-known system NEvAr in which they use an aesthetic measure described in Machado and Cardoso [9]. We have implemented a variation of the aesthetic measure from Machado and Cardoso [9] (see Section 2.5 for more details). Ross et al. [10] evolved aesthetically pleasing images using William Ralph′s bell curve aesthetic measure. We have re-implemented this aesthetic measure and use it in our experiments and compare the resulting images with images evolved using other aesthetic measures (see Section 2.6 for more details). Good overviews of the field are by Greenfield [11] and Hoenig [12] An extensive recent overview by Galanter [13] describes a large number of aesthetic evaluation functions from different origins (complexity, neural networks, distance to an example, etc.). Colin Johnson [14] compiled a survey on the use of fitness functions in EvoArt and evolutionary music from nine editions of the EvoMusart conference.

In the previous work we describe the use of aesthetic measures in unsupervised evolutionary art [15,16], and the use of a combination of aesthetic measures using multi-objective optimisation [17]. This paper is a rewritten and extended version of these 3 papers; we performed experiments in which we compare 7 aesthetic measures under the same conditions, using larger populations and more evaluations. Furthermore, we performed more runs and measured more observables than in the original papers, and added the symmetry aesthetic measure to the comparison. This structured and detailed comparison of 7 aesthetic measures in an unsupervised EvoArt system is the first main contribution of this paper. The second main contribution of this paper is the description of the use of a number of combinations of aesthetic measures in a Multi-Objective Optimisation setup. We address the following research questions:

1. What is the effect of different aesthetic measures on the resulting images?
2. Are there correlations between the scores calculated by different aesthetic measures?
3. How do the aesthetic measures judge each others visual output?
4. How do aesthetic measures differ in terms of evolutionary search speed? In other words, which aesthetic measures lead to rapid convergence and which ones lead to long exploratory phases?
5. How do aesthetic measures differ in the appearance of bloat? (We use a representation with variable chromosome size.)
6. What combinations of two aesthetic measures (in a multi-objective EA) result in images that merge the visual effects of both of them?

With regards to the first research question, we expect that each aesthetic measure will direct the search process into ‘its own part’ of the search space, resulting in an own ‘style’ for each aesthetic measure. We verify this by calculating a range of image features for image evolved by the different aesthetic measures, and compare the image statistics of each aesthetic measure. The second research question concerns similarities between aesthetic measures; we calculate the correlation between the aesthetic scores produced by two aesthetic measures, and present the correlation between all 7 aesthetic measures. Furthermore, we calculate the ‘aesthetic appeal’ of the images evolved by a certain aesthetic measure; we calculate the aesthetic score for aesthetic measure AM_i with aesthetic measure AM_j. We are interested to find how the images that were evolved with an aesthetic measure (as the fitness function) are ‘liked’ by its peer aesthetic measures. An aesthetic measure has high ‘aesthetic appeal’ if its images are appreciated by its peer aesthetic measures. Research question 4 concerns the evolutionary search speed of an aesthetic measure; previous experiments have suggested that some aesthetic measures are ‘easier’ to satisfy than others. This results in convergence after 5–10 generations with some aesthetic measures and with exploration search behaviour after 20 generations with other aesthetic measures. We measure the progress in fitness for the aesthetic measures, and compare the normalised fitness values (per generation) for all aesthetic measures. In order to answer research question 5 on the development of bloat, we measure the average sizes of the colour schemes and the average tree depth using different aesthetic measures as the fitness function, and compare the results. In order to answer research question 6, we use a MOEA setup with a number of combinations of aesthetic measures. We show a portfolio of images using each combination and show Pareto fronts of each combination.

This paper is organised as follows: the aesthetic measures that we used are described in Section 2. Our evolutionary art system is described in Section 3, and our experiments with single aesthetic measures and their results are described in Section 4. Next, we investigated the combination of multiple aesthetic measures and we describe this in Section 5. We end this paper with our conclusions and directions for future work in Section 6.

2. Aesthetic measures

In this section we will describe the aesthetic measures that we used in our experiments. All aesthetic measures were used in the first series of experiments using a single aesthetic measure (Section 4) and some were also used in the series of experiments using multi-objective optimisation (Section 5). The aesthetic measures are (in alphabetical order) Benford′s Law [18], Fractal Dimension [19], Global Contrast Factor [20], Information Theory [21], Machado and Cardoso [9], Ross et al. [10], and Reflectional Symmetry [22]. In the next subsections we will give a brief description of each aesthetic measures. Full details can be found in the original papers.

2.1. Benford′s law

The first aesthetic measure that we describe is based on Benford′s Law [23,18]; Benford′s Law (or first-digit law) states that a list of numbers obtained from real life (i.e. not created by man) are distributed in a specific, non-uniform way. The leading digit occurs one-third of the time, the second digit occurs 17.6%, etc. (see Fig. 1).

We use Benford′s law to measure the distribution of (light) intensity of pixels. For an image we calculate the intensity histogram using 9 bins. Next we calculate the difference between the actual histogram and the Benford histogram

$$M_{bl}(I) = \frac{d_{max} - d_{total}}{d_{max}} \quad (1)$$

where

$$d_{total} = \sum_{i=1}^{9} (H_{image}(i) - H_{benford}(i))^p \quad (2)$$

where $H_{image}(i)$ is the number of entries in the intensity histogram bin number i. $H_{benford}(i)$ is the value from the Benford distribution (see Fig. 1). The maximal difference d_{max} for $p=3$ is $(1-0.301)^3 + (0.176)^3 \cdots + (0.046)^3 = 0.3511$. Lower values for p (we experimented with $p=3$, $p=2$ and $p=1$) result in a higher penalty for differences in intensity distribution. For our experiments we used $p=1$.