Evolving Figurative Images Using Expression-Based Evolutionary Art

João Correia and Penousal Machado
CISUC, Department of Informatics Engineering
University of Coimbra
3030 Coimbra, Portugal
jncor@dei.uc.pt, machado@dei.uc.pt

Juan Romero and Adrian Carballal
Faculty of Computer Science
University of A Coruña
Coruña, Spain
jj@udc.es, adrian.carballal@udc.es

Abstract
The combination of a classifier system with an evolutionary image generation engine is explored. The framework is composed of an object detector and a general purpose, expression-based, genetic programming engine. Several object detectors are instantiated to detect faces, lips, breasts, and leaves. The experimental results show the ability of the system to evolve images that are classified as the corresponding objects. A subjective analysis also reveals the unexpected nature and artistic potential of the evolved images.

Introduction
Expression-based Evolutionary Art (EA) systems have, in theory, the potential to generate any image (Machado and Cardoso 2002; McCormack 2007). In practice, the evolved images depend on the representation scheme used. As a consequence, the results of expression-based EA systems tend to be abstract images. Although this does not represent a problem, there is a desire to evolve figurative images by evolutionary means since the start of EA. An early example of such an attempt can be found in the work of Steven Rooke (World 1996).

McCormack (2005; 2007) identified the problem of finding a symbolic-expression that corresponds to a known “target” image as one of the open problems of EA. More exactly, the issue is not finding a symbolic-expression, since this can be done trivially as demonstrated by Machado and Cardoso (2002), the issue is finding a compact expression that provides a good approximation of the “target” image and that takes advantage of its structure. We address this open problem by generalizing the problem – i.e., instead of trying to match a target image we evolve individuals that match a given class of images (e.g., lips).

The issue of evolving figurative images has been tackled by two main types of approach: (i) Developing tailored EA systems which resort to representations that promote the discovery of figurative images, usually of a certain kind; (ii) Using general purpose EA systems and developing fitness assignment schemes that guide the system towards figurative images. In the scope of this paper we are interested in the second approach.

Romero et al. (2003) suggest combining a general purpose evolutionary art system with an image classifier trained to recognize faces, or other types of objects, to evolve images of human faces. Machado, Correia, and Romero (2012a) presented a system that allowed the evolution of images resembling human faces by combining a general-purpose, expression-based, EA system with an off-the-shelf face detector. The results showed that it was possible to guide evolution and evolve images evocative of human faces.

Here, we demonstrate that other classes of object can be evolved, generalizing previous results. The autonomous evolution of figurative images using a general purpose EC system has rarely been accomplished. As far as we know, evolving different types of figurative images using the same expression-based EC system and the same approach has never been accomplished so far (with the exception of user-guided systems).

We show that this can be attained with off-the-shelf classifiers, which indicates that the approach is generalizable, and also with purpose-built ones, which indicates that it is relatively straightforward to customize it to specific needs. We chose a rather ad-hoc set of classifiers in an attempt to demonstrate the generality of the approach.

The remainder of this paper is structured as follows: A brief overview of the related work is made in the next section; Afterwards we describe the approach for the evolution of objects describing the framework, the Genetic Programming (GP) engine, the object detection system, and fitness assignment; Next we explain the experimental setup, the results attained and their analysis and; Finally we draw overall conclusions and indicate future research.

Related Work

The previously mentioned approaches share two common aspects: the systems have been specifically designed for the
evolution a specific type of image; the user guides evolution by assigning fitness. The work of Baker (1993) is an exception, the system can evolve other types of line drawings, however it is initialized with hand-built line drawings of human faces.

These approaches contrast with the ones where general purpose evolutionary art tools, which have not been designed for a particular type of imagery, are used to evolve figurative images. Although the images created by their systems are predominantly abstract, Steven Rooke (World 1996) and Machado and Romero (see, e.g., 2011), among others, have successfully evolved figurative images using expression-based GP systems and user guided evolution. More recently, Secretan et al. (2011) created picbreeder, a user-guided collaborative evolutionary engine. Some of the images evolved by the users are figurative, resembling objects such as cars, butterflies and flowers.

The evolution of figurative images using hardwired fitness functions has also been attempted. The works of by Ventrella (2010) and DiPaola and Gabora (2009) are akin to a classical symbolic regression problem in the sense that a target image exists and the similarity between the evolved images and the target image is used to assign fitness. In addition to similarity, DiPaola and Gabora (2009) also consider expressiveness when assigning fitness. This approach results in images with artistic potential, which was the primary goal of these approaches, but that would hardly be classified as human faces. As far as we know, the difficulty to evolve a specific target image, using symbolic regression inspired approaches, is common to all “classical” expression-based GP systems.

The concept of using a classifier system to assign fitness is also a researched topic: in the seminal work of Baluja, Pomerlau, and Todd (1994) an Artificial Neural Network trained to replicate the aesthetic assessments is used; Saunders and Gero (2001) employ a Kohonen Self-Organizing network to determine novelty; Machado, Romero, and Manners (2007) use a bootstrapping approach, relying on a neural network, to promote style changes among evolutionary runs; Norton, Darrell, and Ventura (2010) train Artificial Neural Networks to learn to associate low-level image features to synsets that function as image descriptors and use the networks to assign fitness.

**Overview of the Approach**

Figure 1 depicts an overview of the framework, which is composed of two main modules, an evolutionary engine and a classifier.

The approach can be summarized as follows:

1. Random initialization of the population;
2. Rendering of the individuals, i.e., genotype-phenotype mapping;
3. Apply the classifier to each phenotype;
4. Use the results of the classification to assign fitness; This may require assessing internal values and intermediate results of the classification;
5. Select progenitors; Apply genetic operators, create descendants; Use the replacement operator to update the current population;
6. Repeat from 2 until some stopping criterion is met.

The framework was instantiated with a general-purpose GP-based image generation engine and with a Haar Cascade Classifier. To create a fitness function able to guide evolution it is necessary to convert the binary output of the detector to one that can provide suitable fitness landscape. This is attained by accessing internal results of the classification task that give an indication of the degree of certainty in the classification. In the following sections we explain the components of the framework, namely, the evolutionary engine, the classifier and the fitness function.

**Genetic Programming Engine**

The EC engine used in these experiments is inspired by the works of Sims (1991). It is a general purpose, expression-based, GP image generation engine that allows the evolution of populations of images. The genotypes are trees composed of a lexicon of functions and terminals. The function set is composed of simple functions such as arithmetic, trigonometric and logic operations. The terminal set is composed of two variables, \(x\) and \(y\), and randomly initialized constants. The phenotypes are images that are rendered by evaluating the expression-trees for different values of \(x\) and \(y\), which serve both as terminal values and image coordinates. In other words, to determine the value of the pixel in the \((0,0)\) coordinates one assigns zero to \(x\) and \(y\) and evaluates the expression-tree (see figure 2). A thorough description of the GP engine can be found in (Machado and Cardoso 2002).

Figure 3 displays typical imagery produced via interactive evolution using this EC system.

**Object Detection**

For classification purposes we use Haar Cascade classifiers (Viola and Jones 2001). The classifier assumes the form of a cascade of small and simple classifiers that use a set of Haar features (Papageorgiou, Oren, and Poggio 1998) in combination with a variant of the Adaboost (Freund and Schapire 1995), and is able to attain efficient classifiers. This classification approach was chosen due to its state of the art relevance and for its fast classification. Both code and executable files are integrated in the OpenCV API\(^1\).

The face detection process can be summarized as follows:

\(^1\)OpenCV — [http://opencv.org/](http://opencv.org/)

---

Figure 1: Overview of the system.
1. Define a window of size $w$ (e.g. $20 \times 20$).

2. Define a scale factor $s$ greater than 1. For instance 1.2 means that the window will be enlarged by 20%.

3. Define $W$ and $H$ has the size of the input image.

4. From $(0,0)$ to $(W,H)$ define a sub-window with a starting size of $w$ for calculation.

5. For each sub-window apply the cascade classifier. The cascade has a group of stage classifiers, as represented in figure 4. Each stage is composed, at its lower level, of a group of Haar features 5. Apply each feature of each corresponding stage to the sub-window. If the resulting value is lower than the stage threshold the sub-window is classified as a non-object and the search terminates for the sub-window. If it is higher continue to next stage. If all cascade stages are passed, the sub-window is classified as containing an object.

6. Apply the scale factor $s$ to the window size $w$ and repeat 5 until window size exceeds the image in at least one dimension.

**Fitness Assignment**

The process of fitness assignment is crucial from an evolutionary point of view, and therefore it holds a large importance for the success of the described system. The goal is to evolve images that the object detector classifies as an object of the positive class. However, the binary output of the detector is inappropriate to guide evolution. A binary function gives no information of how close an individual is to being a valid solution to the problem and, as such, the EA would be performing, essentially, a random search. It is necessary to extract additional information from the classification detection process in order to build a suitable fitness function.

This is attained by accessing internal results of the classification task that give an indication of the degree of certainty in the classification. Based on results of past experiments (Machado, Correia, and Romero 2012a; 2012b) we employ the following fitness function:

$$\text{fitness}(x) = \sum_{i} \text{stagedif}_x(i) \times i + nstages_x \times 10$$  \hspace{1cm} (1)

The underlying rational is the following: images that go through several classification stages, and closer to be classified as an object, have higher fitness than those rejected in early stages. Variables $nstages_x$ and $\text{stagedif}_x(i)$...
positive rate ensures that almost every positive example is come to be considered a positive example. The high true rate is only one of the several possible ones. That is, an image that passes several stages is likely to be closer of being recognized as having a object than one that passes fewer stages. In other words, passing several stages is a pre-condition to be classified as having the object. Variable \( n_{\text{stages}} \), holds the number of stages that image, \( x \), has successfully passed. That is, an image that passes several stages is likely to be closer of being recognized as having a object than one that passes fewer stages. In other words, passing several stages is a pre-condition to be classified as having the object. Variable \( \text{stageddiff}_x(i) \) holds the maximum difference between the threshold necessary to overcome stage \( i \) and the value attained by the image at the \( i^{th} \) stage. Images that are clearly above the thresholds are preferred over ones that are only slightly above them. Obviously, this fitness function is only one of the several possible ones.

### Experimentation

Within the scope of this paper we intend to evolve the following objects: faces, lips, breasts and leaves. For the first two we use off-the-shelf classifiers that were already trained and used by other researchers in different lines of investigation (Lienhart and Maydt 2002; Lienhart, Kuranov, and Pisarevsky 2003; Santana et al. 2008). For the last two we created our own classifiers, by choosing suitable datasets and training the respective object classifier.

In order to construct an object classifier we need to construct two datasets: (i) positive – examples of images that contain the object we want to detect; (ii) negative – images that do not contain the object. Furthermore, for the positive examples, we must identify the location of the object in the images (see figure 6) in order to build the ground truth file that will be used for training.

For these experiments, the negative dataset was attained by picking images from a random search using image search engines, and from the Caltech-256 Object Category dataset (Griffin, Holub, and Perona 2007). Figure 7 depicts some of the images used as negative instances. In what concerns the positive datasets: the breast object detector was built by searching images on the web; the leaf dataset was obtained from the Caltech-256 Object Category dataset and from web searches. As previously mentioned, the face and lip detector are off-the-shelf classifiers. Besides choosing datasets we must also define the training parameters. Table 1 presents the parameters used for training of the cascade classifier.

The success of the approach is related to the performance of the classifier itself. By defining a high number of stages we are creating several stages that the images must overcome to be considered a positive example. The high true positive rate ensures that almost every positive example is learned per stage. The max false positive rate creates some margin for error, allowing the training to achieve the minimum true positive rate per stage and a low positive rate at the end of the cascade. Similar parameters were used and discussed in (Lienhart, Kuranov, and Pisarevsky 2003).

Once the classifiers are obtained, they are used to assign fitness in the course of the evolutionary runs in an attempt to find images that are recognized as faces, lips, breasts and leaves. We performed 30 independent evolutionary runs for each of these classes. In summary we have 4 classifiers, with 30 independent EC runs, totaling 120 EC runs.

The settings of the GP engine, presented in table 2, are similar to those used in previous experimentation in different problem domains. Since the classifiers used only deal with greyscale information, the GP engine was also limited to the generation of greyscale images. The population size used in this experiments 100 while in previous experiments we used a population size of 50 (Machado, Correia, and Romero 2012a). This allows us to sample a larger portion of the search space, contributing to the discovery of images that fit the positive class.

In all evolutionary runs the GP engine was able to evolve images classified as the respective objects. Similarly to the behavior reported by Machado, Correia, and Romero (2012a; 2012a), the GP engine was able to exploit weaknesses of the classifier, that is, the evolved images are classified as the object but, from a human perspective, they often fail to resemble the object. In figure 8 we present examples of such failures. As it can be observed, it is hard to recognize breasts, faces, leaves or lips in the presented images. It is important to notice that these weaknesses are not a byproduct of the fitness assignment scheme, as such they cannot be solved by using a different fitness function, nor particular to the classifiers used. Although different classi-

### Table 1: Haar Training parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of stages</td>
<td>30</td>
</tr>
<tr>
<td>Min True Positive rate per stage</td>
<td>99.9%</td>
</tr>
<tr>
<td>Max False Positive rate per stage</td>
<td>50%</td>
</tr>
<tr>
<td>Object Width</td>
<td>20 or 40(breasts, leaf)</td>
</tr>
<tr>
<td>Object Height</td>
<td>20 or 40(leaf)</td>
</tr>
<tr>
<td>Haar Features</td>
<td>ALL</td>
</tr>
<tr>
<td>Number of splits</td>
<td>1</td>
</tr>
<tr>
<td>Adaboost Algorithm</td>
<td>Gentle Adaboost</td>
</tr>
</tbody>
</table>

![Figure 6: Examples of images used to train a cascade classifier for leaf detection. On the top row the original image, on the bottom row the cropped example used for training.](image-url)
Figure 7: Examples of images belonging to the negative dataset used for training the cascade classifiers.

Table 2: Parameters of the GP engine. See (Machado and Cardoso 2002) for a detailed description.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population size</td>
<td>100</td>
</tr>
<tr>
<td>Number of generations</td>
<td>100</td>
</tr>
<tr>
<td>Crossover probability</td>
<td>0.8 (per individual)</td>
</tr>
<tr>
<td>Mutation probability</td>
<td>0.05 (per node)</td>
</tr>
<tr>
<td>Mutation operators</td>
<td>sub-tree swap, sub-tree replacement, node insertion, node deletion, node mutation</td>
</tr>
<tr>
<td>Initialization method</td>
<td>ramped half-and-half</td>
</tr>
<tr>
<td>Initial maximum depth</td>
<td>5</td>
</tr>
<tr>
<td>Mutation max tree depth</td>
<td>3</td>
</tr>
<tr>
<td>Function set</td>
<td>+, −, ×, ÷, min, max, abs, neg, warp, sign, sqrt, pow, mdist, sin, cos, if</td>
</tr>
<tr>
<td>Terminal set</td>
<td>x, y, random constants</td>
</tr>
</tbody>
</table>

These results have opened a series of possibilities, including the use of this approach to assess the robustness of object detection systems, and also the use of evolved images as part of the training set of these classifiers in order to overcome some of their shortcomings. Although we already are pursuing that line of research and promising results have been obtained (Machado, Correia, and Romero 2012b), it is beyond the scope of the current paper.

When one builds a face detector, for instance, one is typically interested in building one that recognizes faces of all types, sizes, colors, sexes, in different lighting conditions, against clear and cluttered backgrounds, etc. Although the inclusion of all these examples may lead to a robust classifier that is able to detect all faces present in an image, it will also mean that this classifier will be prone to recognize faces even when only relatively few features are present. In contrast, when building classifiers for the purpose described in this paper, one may select for positive examples clear and iconic images. Such classifiers would probably fail to identify a large portion of real-world images containing the object. However, they are would be extremely selective and, as such, the evolutionary runs would tend to converge to images that clearly match the desired object. Thus, although this was not explored, building a selective classifier can significantly reduce the number of runs that converge to atypical images such as the ones depicted in figure 8.

According to our subjective assessment, some runs were able to find images that actually resemble the object that we are trying to evolve. These add up to 6 runs from the face detector, 5 for the lip detector, 4 for the breast detector and 4 for the leaf detector.

In figures 9, 10, 11 and 12 we show, according to our subjective assessment, some of the most interesting images evolved. These results allow us to state that, at least in some instances, the GP engine was able to create figurative images evocative of the objects that the object detector was design to recognize as belonging to the positive class.

By looking at the faces, figure 9, we can observe the presence of at least 3 facial features per image (such as eyes, lips, nose and head contour). The images from the first row have been identified by users as resembling wolverine. The
ones of the second row, particularly the one on the left, have been identified as masks (more specifically african masks). In what concerns the images from the last row, we believe that their resemblance “ghost-like” cartoons is striking.

In what concerns the images resulting from the runs where a lip detector was used to assign fitness, we consider that their resemblance with lips, caricatures of lips, or lip logos, is self evident. The iconic nature of the images from the last row is particularly appealing to us.

The results obtained with the breast detector reveal images with well-defined or exaggerated features. We found little variety in these runs, with changes occurring mostly at the pixel intensity and contrast level. As previously mentioned, most of these runs resulted in unrecognizable images (see figure 8), which is surprising since the nature of the function set would lead us to believe that it should be relatively easy to evolve such images. Nevertheless, the successful runs present images that are clearly evocative of breasts.

Finally the images from the leaf detector, vary in type and shape. They share however a common feature they tend to be minimalist, resembling logos. In each of the images of the first row the detector identified two leaf shapes. On the
The results from 30 independent runs per each classifier shown that is possible to evolve images that are detected as the corresponding objects and that also resemble that object from a human perspective. The images tend to depict an exaggeration of the key features of the associated object, allowing the exploration of these images in design and artistic contexts.

The paper makes 3 main contributions, addressing: (i) A well-known open problem in evolutionary art; (ii) The evolution of figural images using a general-purpose expression based EC system; (iii) The generalization of previous results.

The open problem of finding a compact symbolic expression that matches a target image is addressed by generalization: instead of trying to match a target image we evolve individuals that match a given class. Previous results (see (Machado, Correia, and Romero 2012a)) concerned only the evolution of faces. Here we demonstrate that other classes of objects can be evolved. As far as we know, this is the first autonomous system that proved able to evolve different types of figural images. Furthermore the experimental results show that this is attainable with off-the-shelf and purpose build classifiers, demonstrating that the approach is both generalizable and customizable.

Currently, we are performing additional tests with different object detectors in order to expand the types of imagery produced.

The next steps will comprise the following: combine, refine and explore the evolved images, using them in user-guided evolution and automatic fitness assignment schemes; combine multiple object detectors to help refine the evolved images (for instance use a face detector first and an eye or a lip detector next); use the evolved examples that are seen as shortcomings of the classifier to refine the training set and boost the existing detectors.

Acknowledgements

This research is partially funded by: the Portuguese Foundation for Science and Technology in the scope of project SBIRC (PTDC/EIA–EIA/115667/2009) and of the iCIS project (CENTRO-07-ST24-FEDER-002003), which is co-financed by QREN, in the scope of the Mais Centro Program and European Union’s FEDER; Xunta de Galicia Project XUGA?PGIDIT10/TIC105008PR.

References


Freund, Y., and Schapire, R. E. 1995. A decision-theoretic generalization of on-line learning and an application to...


Machado, P., and Cardoso, A. 2002. All the truth about NEvAr. Applied Intelligence, Special Issue on Creative Systems 16(2):101–119.


