Computing Aesthetics

Penousal Machado[1], Amílcar Cardoso[2]

[1] Instituto Superior de Engenharia de Coimbra
[2] Departamento de Engenharia Informática da Universidade de Coimbra
{machado, amilcar}@dei.uc.pt

Abstract. Aesthetic judgment is one of the most valued human characteristics, being regarded as a remarkable display of intelligence. Popular knowledge tells us that it is purely cultural. In this paper, we present a brief theory according to which aesthetics depends on biological and cultural issues, namely on visual image processing. We present an implementation of this theory and the experimental results achieved. The tests were made with two purposes in mind, validating the theory and using it as a basis for the automatic generation and evaluation of images.

1 Introduction

The work presented in this paper, is part of an ongoing research project (NEvAr), whose goal is the development of computer programs capable of creating artworks [5]. The main field of study and application is the field of visual arts.

We can divide AI applications to the field of Arts in three categories: (1) "Intelligent" tools; (2) Systems performing "art understanding" tasks (e.g. musical analysis); (3) C*onstructed artists*, which "are supposed to be capable of creating aesthetically meritorious artworks on their own, with minimal human intervention" [10].

In our quest for the creation of constructed artist, we identified a set of features that such a system should exhibit [5]. One of these is the ability to make aesthetic judgements. Consequently, we tried to answer two basic questions, namely: "Can this be achieved?" and "How can it be achieved?".

The success of constructed artists is greater in music then in visual arts [4], this can be explained by the higher quantity of data required by image handling [4], but this explanation seems a little shallow. The fact is that music theory is more developed and quantitative than theory in visual arts [4]. In music we can "write down" a piece, in visual arts we have neither the alphabet nor the grammar to do so. In other words, we also lack a notation system. In music there is a clear distinction between composer and interpreter, in visual arts these two different roles are mixed.

To our knowledge, there was only one attempt to create a system (in the field of visual arts) with the ability to make aesthetic judgements [1]. In this system a Neural Network makes the evaluation of the images. In the field of music, the approach to the problem of creating a constructed artist is radically different. The generation of new musical pieces is, usually, guided by knowledge (i.e. by a musical theory), and the evaluation of the pieces, when it exists, is also based on a musical theory. There was no attempt in [1] to support the system in an underlying aesthetic theory, we consider this to be a factor that hindered the success of the system.

In this paper, we will focus on aesthetics and on how to compute aesthetic value. The paper has the following structure: The second section pertains to the origins of art and aesthetic judgment: we give a short biological explanation to the devotion of humans to art and to how natural evolution favored the appearance of art. We also make a brief outline of an aesthetic theory that serves as basis, and is part of our ongoing research. In section three, we present an implementation of this theory. Some experimental results are referred in the fourth section and, finally, in section five, we draw some conclusions and talk about further work.

2 Art & Aesthetics

If we ask someone why he/she likes a certain painting, they will usually talk about the emotions or feelings triggered by the artwork, the combination of colors, or, even more frequently, we will get the intriguing answer: "I just like it.". It is extremely difficult to find a consensual definition for what is Art.

Art and Aesthetics are different, yet highly related, fields. We can say, to a certain point, that Aesthetics is a subset of the Arts, and we can define it as the study of the form in itself, striped from its meaning.

From our point of view, the assessment of an artwork is influenced by two factors:

- The "content" of the artwork, which relates to what is represented by the artwork. If we consider Art as a form of communication, then "content" is what is communicated.
- The "visual aesthetic value" of the artwork, that is related to color combination, composition, shape, etc. We are talking about the form of the artwork, thus, how "content" is represented.

By assuming this point of view, we aren't creating a false dichotomy between "form" and "content". We are aware of the fact that these factors aren't completely independent. Consequently, the value of an artwork rests on these factors and their interactions. It is possible to have an artwork that is visually pleasing, but whose content is displeasing. In fact, many art styles rely on the mixed feelings caused by this discrepancy (e.g. many of S. Dali's paintings).

If we restrict ourselves to the field of Aesthetics, these factors gain independence because, as stated before, Aesthetics focuses in the "form". In other words, an image can have a high aesthetic value, independently from its content, and even if it is deprived of content. We don't mean that content isn't important, we just mean it is not indispensable from the Aesthetics point of view.

It seems clear that the way in which content influences the value of a given artwork depends, mainly, on cultural issues. From our point of view, visual aesthetic value, is connected to visual image perception and processing, thus being mainly: biological, hardwired, and therefore universal. We aren't saying that it is coded in our genes, we just mean that what we are, and the way our visual image perception system works, makes us favor certain images to others. The remainder of this section is dedicated to the support of the previous statement.

2.1 The Origins of Art

We will start by focusing on the origins of art and try to show how Natural Evolution could favor the appearance of art. Natural selection should favor the fittest individuals in a population, so why should a seemingly useless activity as art be favored by it?

To the vast majority of the animals, the struggle for survival takes all their time. Only in their infancy they have time to playing and games. The same happened to the primitive man. Only from a certain point in history man begun to have spare time.
The appearance of art may be explained by the necessity of using other forms of communication other than gesture or speech. Although this explanation is usually accepted, there are, other explanations that can be considered complementary. The coordination of hand movements had a great influence on human evolution. When a prehistoric man devoted himself to painting he was not doing a fruitless activity: by painting he was also training and improving his motor coordination. These two factors would give him an immediate survival advantage.

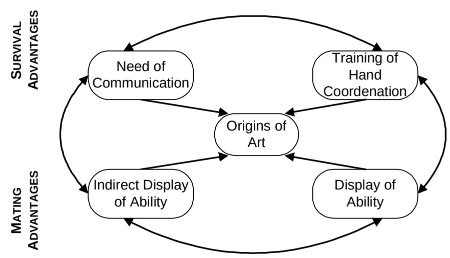

Fig. 1. Set of explanations that can justify the appearance of art.

We also have to consider the mating advantages since animals tend to choose the most fit individual they can for mating. By making an artwork, he is making a direct display of ability; furthermore he is showing that has time to spend on activities that aren't vital for his immediate survival, and thus making an indirect display of ability. In our opinion, this set of explanations gives reasonable justification for the devotion of man to art, but they don't justify why we find certain images beautiful, aesthetically pleasing or artistic.

The analysis of children's drawings allows us to observe the development of aesthetic, showing that the earliest stages of development are common to all children and that there is a universal imagery shared across cultural boundaries. Kellogg [8] recognizes five basic stages in pictorial representation: Scribbles, Diagrams, Combines, Aggregates, Pictorials. Usually, the first pictorial image is the human figure. This image is, always constructed in the same, rather puzzling way. It begins with an empty circle; in the next step bubbles are added to the inside of the circle; the bubbles are gradually transformed in eyes, mouth and nose; afterwards hair is included; Some of the hairs get longer, until they are transformed into arms and legs. Somewhere between the ages of 6 and 12, this universal imagery begins to disappear, probably, due to educational influence [8].

It seems safe to say that visual aesthetic judgment is not particular to humans. In fact, we share this ability with other species of animals. Experiments with chimpanzees show that they follow the same steps of development of human infants. The first stages of development are similar, however, Chimpanzees aren't able to go beyond the phase of the circle into the phase of the filled circle. They also never seem to be able to create a pictorial image. Nevertheless, their paintings show that the brain of a chimpanzee is capable of making simple aesthetic judgments. Morris [8] found six common principles between chimpanzee and human art: Self-Rewarding Activity, Compositional Control, Calligraphic Differentiation, Thematic Variation, Optimum Heterogeneity and Universal Imagery.

These factors lead us to conclude that the assessment of the visual aesthetic value of an image is directly connected to the visual image perception system. This conclusion is also supported by a wide variety of experiments. One instance of such is the well-known, perfect rectangle experiment. When asked to pick the most beautiful rectangle from a list, most people will choose the rectangle with the "Golden Proportions" 1:1.618. The explanation for this is related to the range of our visual field [8]. Another example would be the difference between horizontal/vertical and diagonal lines. The diagonal line has a higher visual intensity. The periphery of the eye is very sensitive to diagonals, so it calls for complete attention from the viewer. That is why traffic signs designed to warn of hazards are diamond shaped, using diagonals. These differences must be taken into consideration, when composing an artwork, so that the desired "balance" is achieved.

Let us consider how the visual image perception system works.

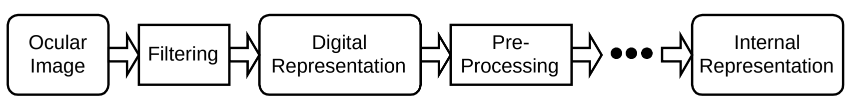

Fig. 2. Steps of visual image perception.

The process of transformation of the ocular image into its digital representation is well known. From this representation, the brain constructs internal representations, retaining only certain aspects of the image. The way in which this process works is still source of much debate. The idea that there is a "pre-processing" of the digital image (shape and contour detection, color analysis, depth and movement analysis, etc), and that "recognition" and subsequent transformation to internal representations is made based on the results of this "pre-processing", is usually accepted [11].

From the previous statement, we can say that there is a difference between image complexity, and internal representation complexity; furthermore a complex image isn't necessarily difficult to (pre)process. To clarify our previous statement, consider the following analogy: a fractal image is usually complex, and highly detailed; yet it can be compactly described by a simple mathematical formula. Therefore, we can say that there is a difference between image complexity, and internal representation complexity; furthermore a complex image isn't necessarily difficult to (pre)process.
In the book The Society of Mind, Minsky associates the concepts of fashion and style to the mental work necessary to process images: "…why do we tend to choose our furniture according to systematic styles or fashions? Because familiar styles make it easier for us to recognize and classify the things we see. If every object in a room were distracting by itself, our furniture might occupy our minds too much… It can save a lot of mental work…" [6].

If we accept this explanation, we are lead to conclude that simpler, easier to process images have higher aesthetic value than complex ones. Or, in other words, low processing complexity implies high aesthetic value. Following this idea, we come to the conclusion that a completely blank image has higher aesthetic value than any artwork, since it is certainly easier to process.[1]

Abraham Moles [7] sustains a quite different position, namely, that aesthetic value is directly connected image complexity, and bases his measure of image complexity on information's theory, e.g. image complexity is directly connected to the unpredictability of the "pixels" in the image.

Minsky's examples are related to office furniture. When we are in an office we don't want to be distracted, we want to work; when we are admiring an artwork, we want to be distracted, that's probably why we usually have, in our offices, a painting to look at when we want to distract ourselves.

In our opinion, aesthetic visual value depends on two factors: (1) Processing Complexity (the lower, the better); (2) Image Complexity (the higher, the better). This seems contradictory, but, as we said before, a complex image isn't necessarily difficult to process. Thus, images that are simultaneously visually complex and easy to process are the images that have higher aesthetic value. Our state of mind influences how we value these factors, if we are tired, we will probably give more weight to processing simplicity. Returning to the fractal example, fractal images are usually complex, however, the propriety of self-similarity makes these images easier to process, which gives an explanation to why we usually find fractal images beautiful. Another important characteristic of fractal images is that they have several levels of detail, characteristic that can also be found in many artworks, (e.g. Kandinsky's paintings). This strikes us as very important, especially if we notice that the act of "seeing" isn't instantaneous, it spans through a (sometimes-long) period of time. When we look briefly at such an artwork, we are automatically able to recognize its main shapes, if we give it more attention we will increasingly discover more detail. This makes the image easy to process, and thus less distracting when we don't want to

[1] This may not be as absurd as it seems considering paintings like "White on White" from Malevitch.

give it attention; simultaneously, when we want to give it attention, we will always find enough detail to “fill” our minds. If the image had only one level of detail, it would probably make it either difficult to process rapidly or with little complexity. This would hinder the generality of the artwork in the sense that our willingness to look at it would largely depend on our state of mind. Thus, it is important to preserve a high Image Complexity / Processing Complexity ratio through all the period of seeing.

3 Implementing the Theory

On the previous section, we have outlined an aesthetics theory. We stated that: the aesthetic value of an artwork is directly connected to Image Complexity (IC) inversely connected to Processing Complexity (PC).

In order to test our theory, we must devise estimates for IC and PC. Having these estimates, we could develop a formula to evaluate the aesthetic value, e.g.:

$$\frac{IC^a}{PC^b} \qquad (1)$$

In which, *a* and *b*, can be used to change the importance given to each factor.

We have also stated that the IC/PC ratio should be maintained high trough all the period of seeing. By practical reasons, that will be explained later, we were forced to consider that IC is constant. Consequently, we must estimate this factor, based, solely, in the differences between PC at distinct moments in time. Our formula will then become:

$$\frac{IC^a}{(PC)^b * \left(\frac{PC(t_1) - PC(t_0)}{PC(t_1)}\right)^c} \qquad (2)$$

Since we now have, $PC(t_0)$ and $PC(t_1)$, we will replace PC by the multiplication of $PC(t_0)$ and $PC(t_1)$, hoping that this will give a better estimate of Processing Complexity. Therefore we arrive to the following formula:

$$\frac{IC^a}{(PC(t_0) * PC(t_1))^b * \left(\frac{PC(t_1) - PC(t_0)}{PC(t_1)}\right)^c} \qquad (3)$$

With *a*, *b* and *c* yielding the relative importance of each factor. Now, in order to test our theory we “only” have to devise estimates for IC and PC. This formula is only one instance of the ones that could be constructed based on the presented theory.

3.1 The Estimates

How can we estimate IC and PC? If we had an implementation of the visual image perception system, we could use this to estimate, at least, PC. Since we don't have this, we had to resort to other methods. One of the main steps of image perception involves transforming a digital image into some internal representation. During this transformation, some (probably lots) of detail is lost, and only certain aspects of the image are retained. This led us to establish a metaphor between this process and *lossy* image compression (e.g. *jpg* compression). In this type of compression method, the image is coded with some error, we can specify the amount of detail kept and thus, indirectly, the error resulting from the compression and the compression ratio.

Jpg compression works as follows[2]: Transformation of the image into frequencies, using the discrete cosine transform (DCT); Quantization of the resulting coefficient matrix; Lossless compression of the results (e.g. by Huffman coding). The process of quantization is the only step in which detail is lost. The process of quantization affects (mainly) the high frequencies, which can, usually, be discarded without significant loss in image quality. The amount, and quality (i.e. the error involved), of the compression achieved by this method, depends on the predictability of the pixels in the image. Considering a definition of image complexity similar to the one proposed in [7] we could use this type of compression to estimate IC. Therefore, our IC estimate will be:

$$\frac{RMS\_Error}{Compression\_Ratio} \tag{4}$$

resulting from the *jpg* compression of the image.

In the last years, several new compression methods have appeared, including fractal image compression and wavelets. Some of these methods can be compared favorably with *jpg*. Fractal image compression method has grabbed our attention, due to some of its characteristics, namely:

- Suppose that you compress an image with Fractal Image Compression and *jpg* compression. Suppose, also, that the *rms* errors resulting from this compression are equal. If you ask someone to choose the image with higher quality they will choose (almost always) the image resulting from fractal image compression [2].
- The images resulting from fractal image compression don't have a fixed size. They can be reproduced at sizes greater than the original, resulting in what has been called a "fractal zoom". The images resulting from fractal zoom have better quality than the ones produced by standard zoom techniques [2].
- The basic idea of fractal image compression is exploring the self-similarities present in an image.

This set of features led us to conclude that fractal image compression is, somehow, closer to the way in which humans process images than other types of compression (e.g. *jpg*). Accordingly, we will use fractal image compression for estimating PC.

[2] For a detailed description of the *jpg* standard see, e.g. [9].

As we have seen before, we need to take these estimates at different moments in time. The idea is that the amount of detail perceived by a person when looking at an image increases over time. The amount of detail kept can be specified in both compression methods. Therefore, we can vary this amount to simulate estimates for IC and PC at different points in time.

At this point, we encountered some implementation problems. We wanted to use high compression ratios, since we think that the visual image perception system must also perform "high compressions". Furthermore, we wanted to use similar compression ratios for our IC and PC estimates. Unfortunately, although fractal image compression works well at high compression ratios, *jpg* compression doesn't, and the images get too degraded. To deal with this problem, we decided to use only one IC estimate, i.e. we only vary PC over time. Fractal image compression is a slow process, so, we decided to measure it, only, in two points in time (t_0 and t_1). The amount of detail kept in $PC(t_1)$ is significantly larger than in $PC(t_0)$. We can consider, to some degree, that the image resulting from the compression in $PC(t_1)$ can be produced as follows: divide the original image in four equal parts; enlarge each of the resulting blocks so it reaches the same size as the original image; perform fractal image compression in each of these blocks, using for each the same parameters that in $PC(t_0)$. This can be seen as a person concentrating at each of the quadrants of the image.

In the IC estimate, we used a compression ratio close to the one used in PC(t1).

Before describing our experimental results, we want to stress some points to avoid misinterpretations. We don't intend to say that the visual image perception system works by fractal image encoding, or that image complexity is the IC estimate. We merely think that we can use the results of jpg and fractal image compression as rough estimates of IC and PC. In other words, when image complexity increases, jpg compression tends to perform worse; Fractal image compression takes advantage of some of the characteristics of the images (i.e. self-similarities). The visual image perception system also takes advantages of these (and possibly many others) characteristics. Thus, the results from fractal image compression are closer to the ones of visual image perception than the results of jpg compression.

4 Experimental results

The next logical step was to test our formula, for which we used the TDA test (Test of Drawing Appreciation) [3]. This is a psychological test that can be applied to individuals or groups. Its objective is to assess the aptitude to evaluate artistic structures. The test tries to estimate the level in which an individual recognizes and reacts to basic principles of aesthetic order: unity, dominance, variety, balance, continuity, rhythm, symmetry and proportion [3]. The test is composed by 90 items. Each item consists in a pair or triad of similar drawings in which one of the drawings corresponds to the fundamental principles while the other violate one or more of these principles. The task of the individual being evaluated is to chose, for each item, the drawing that she/he likes more. The average scores in this test are 45.680 (for professionals of several activities) and 55.6897 (for Fine Arts graduates) correct answers [3]. Answering randomly to the test would result in an average of 43.47. Notice that the difference between this result and the ones achieved by a random population is very small, 2.21, which gives an idea of the toughness of this test.

Each of the images in the TDA test was scanned, and our program computed the IC, PC(t_0) and PC(t_1) estimates. Using these estimates, the program calculated the results of the previously stated formula. For each test item, the program "chose", as answer, the image yielding the higher result.

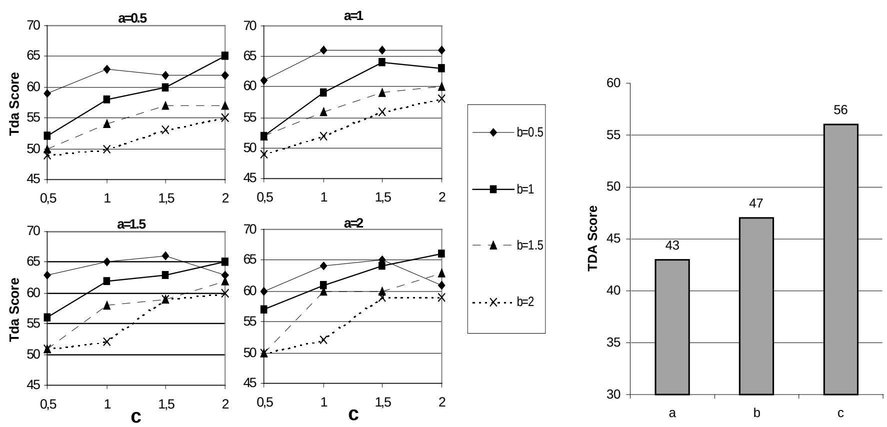

Fig. 3. Results from TDA test, using all the factors in the formula (left) and using only one factor (right).

Fig.3 shows the results achieved by our program, in the TDA test. For values of a, b and c ranging from 0.5 to 2. Our program's TDA score varies from 49 to 66 and the average score is 58.6. Thus, our minimum score is larger than the average achieved by humans (45.680), and our average score is larger than the one achieved by fine arts graduates (55.6897). Using only one factor of the formula (i.e. with the corresponding parameter different from zero and the other two equal to zero) we also get interesting results. Using the IC or PC factors by themselves gives low TDA scores (43 and 47). Using the (PC(t0)-PC(t1)) factor alone gives a TDA score of 56. These results are consistent with the presented theory, i.e. neither Image Complexity nor Processing Complexity, by themselves, are insufficient to assess Aesthetic Value.

5 Conclusions and Further Work

In this paper, we presented a brief theory of aesthetics. This theory considers the biological roots of Art and Aesthetics, stating that the assessment of an artwork depends on two factors: content and form. Furthermore, we claim that aesthetic value is connected to the image-processing task of the brain. We relate aesthetic value with image complexity and with processing complexity. We devised a formula and estimates, in order to test our theory. The experimental results achieved in the TDA test were surprisingly good, especially if we consider the roughness of the estimates. The presented formula is based on the theory, which doesn't mean that there aren't other formulas/methods that could express our theory equally well (or better).

We believe that our theory is an approximation to the truth. We also believe that basing the development of a constructed artist on a consistent aesthetic theory will increase the quality of the results.

In spite of the results achieved on the TDA, it is important to say that we can't, directly, use our formula to guide the generation process of a constructed artist. The images in each test item are quite similar, both in style and complexity. Our formula showed to be effective discriminating between these images. However, it falls short when comparing images with significant style/complexity differences. The problem is attached with the $PC(t_0)$-$PC(t_1)$ factor that can take values very close to zero, and thus, increase the result value enormously. We are currently developing a test set with images of dissimilar styles. It is our belief that with some "minor" changes the formula will be able to pass the test.

Acknowledgments

We would like to thank CISUC (Centro de Informática e Sistemas da Universidade de Combra) for supporting our research.

References

1. Baluja, S., Pomerlau, D., Todd, J., *Towards Automated Artificial Evolution for Computer-Generated Images*. Connection Science, 6(2), 1994, pp. 325-354.
2. Fisher, Y. (Ed.), *Fractal Image Compression - Theory and Application*, Springer-Verlag, 1994.
3. Graves, M., *Test of Drawing Appreciation*, The Psychological Corporation, 1977.
4. Kurzweil, R. *The Age of the Intelligent Machines*, 1990.
5. Machado, P., Cardoso, A., *Model Proposal for a Constructed Artist*, World Multi-Conference on Systemics Cybernetics and Informatics - SCI'97.
6. Minsky, M., *The Society of Mind*. Simon & Schuster, 1986.
7. Moles, A., *Art et ordinateur*, Casterman, 1971.
8. Morris, D., *The Biology of Art*, 1962.
9. Nelson, M, Gailly, J.-L., *The Data Compression Book*. M&T Books. 95.
10. Spector, L., Alpern, A., *Criticism, Culture, and the Automatic Generation of Artworks*, AAAI-94, MIT Press.
11. Vignaux, G., *Les Sciences Cognitives*, Éditions La Décoverte, 1992.