Aesthetic Classification and Sorting Based on Image Compression

Juan Romero[1], Penousal Machado[2], Adrian Carballal[1], and Olga Osorio[3]

[1] Faculty of Computer Science, University of A Coruña, Coruña, Spain
jj@udc.es, adriancarballal@gmail.com
[2] CISUC, Department of Informatics Engineering,
University of Coimbra, 3030 Coimbra, Portugal
machado@dei.uc.pt
[3] Faculty of Communication Sciences, University of A Coruña, Coruña, Spain
olga.osorio@udc.es

Abstract. One of the problems in evolutionary art is the lack of robust fitness functions. This work explores the use of image compression estimates to predict the aesthetic merit of images. The metrics proposed estimate the complexity of an image by means of JPEG and Fractal compression. The success rate achieved is 72.43% in aesthetic classification tasks of a problem belonging to the state of the art. Finally, the behavior of the system is shown in an image sorting task based on aesthetic criteria.

1 Introduction

Having an estimate of aesthetic value, allowing the differentiation among various objects based on merely aesthetic criteria, would have a great theoretical and practical value in the field of Evolutionary Art.

This paper presents a set of 18 features, based on JPEG and Fractal compression, paying attention to the complexity of an image. Their adequacy is shown in two different aesthetic tasks: classification and sorting. First of all, we tackle the issue of image classification based on aesthetic criteria presented by Datta et al. [4]. Using both the image dataset and the features provided by them, a thorough comparison was established with those detailed in the present paper by means of Support Vector Machines (SVMs) and ANNs. A linear combination of the outputs of the neural network trained in the previous task is used to sort several image sets presented by [10]. That combination is presented as a possible aesthetic fitness function which we intend to use within an Evolutionary Art System in the future.

2 Complexity and Aesthetics

The relationship between aesthetics and image complexity has been explored in several psychology and graphic computation papers [2,5,6,16]. In a simplified

way, the complexity of an image is related to its entropy, and inversely related to the order. It is related to the minimal information (or the minimal program) required to "construct" the image. It may be said to depend on the degree of predictability of each pixel of the image [17]. Thus, a flat image with every pixel of the same color shows a perfect order, and it is less complex. A pure random image can be seen as extremely complex and the value of each pixel is impossible to predict, even taking into account the values of neighbor pixels.

The relevance of perceived image complexity is a recurring topic in the field of aesthetics [1,2,17]. According to [12], "Aesthetic value is related to the sensorial and intellectual pleasure resulting from finding a compact percept (internal representation) of a complex visual stimulus". In the same paper, two different estimates are presented: one for the Complexity of the Visual Stimulus (CV), using JPEG Compression and another for the Complexity of the Percept (CP), using Fractal Compression. Finally, the metrics are tested with psychological test: "Design Judgment Test" [8]. In [15], Machado used a subset of the features proposed in this project and an Artificial Neural Network (ANN) classifier for author identification, attaining identification rates higher than 90% across experiments. This paper presents an aesthetic fitness function based on the metrics proposed by [12].

3   Proposed Features

While several preceding works [4,10,20] use ad-hoc metrics designed for a specific problem, the present paper will use general metrics based on edge detection and complexity estimates of black and white images. The said estimates are determined from the compression error generated from the original image. The advantage posed by these metrics is their generality; they are easily estimated and can be applied only on grayscale information of the image.

Before carrying out the calculations of the different features, every image is individually subjected to a series of transformations before being analyzed. A given input image is loaded and resized to a standard width and height of 256×256 pixels, transformed into a three channel image in the RGB (red, green and blue) color space, with a depth of 8-bit per channel and all pixel values scaled to the $[0, 255]$ interval. This step ensures that all input images share the same format and dimensions.

Afterwards, the image is converted into the HSV (Hue, Saturation and Value) color space and its HSV channels are split. Only the V channel is stored as a 1-channel grayscale image, given that we just need its representation in black and white format.

Previous works such as [4,10,11] rely, to a large extent, on color information to extract features. [10] states "the color palette seen in professional photos and snapshots is likely to be very different". In this work, we rely exclusively on grayscale information. We want to make the system as generic as possible, and in every dataset we have there are some grayscale images. In the future, however, we will analyze the results by using also color information (channels HS).

Once the grayscale image is available, two edge detection filters are applied, Canny and Sobel, which will yield two new black and white images. In previous works (e.g., [18,10]) filters such as Canny, Sobel, Gauss and Laplace have been applied.

The most popular image compression schemes are lossy, therefore they yield a compression error, i.e., the compressed image will not exactly match the original. All other factors being equal, complex images will tend towards higher compression errors and simple images will tend towards lower compression errors. Additionally, complex images will tend to generate larger files than simple ones. Thus, the *compression error* and *file size* are positively correlated with image complexity [9]. To explore these aspects, we consider three levels of detail for the JPEG and Fractal compression metrics: *low*, *medium*, and *high*. The process is the same for each compression level; the current image in analysis is encoded in a JPEG or fractal format. We estimate each metric of image I using the following formula:

$$RMSE(I, CT(I)) \times \frac{s(CT(I))}{s(I)} \tag{1}$$

where $RMSE$ stand for the root mean square error, CT is the JPEG or fractal compression transformation, and s is the file size function.

In the experiments described herewith, we use a quad-tree fractal image compression scheme [7] with the set of parameters given in Table 1. Note that letting the minimum partition level be 3 implies that the selected region is always partitioned into 64 blocks first. Subsequently, at each step, for each block, if one finds a transformation that gives good enough pixel by pixel matches, then that transformation is stored and the image block isn't further partitioned. (Here, pixel by pixel match is with respect to the usual 0 to 255 grayscale interval encoding.) If the pixel by pixel match error is more than 8 for at least one of the pixels of the block in the partition, that image block is further partitioned into 4 sub-blocks, the level increases, and the process is repeated. When the maximum partition level is reached, the best transformation found is stored, even if the pixel by pixel match error for the block exceeds 8. The quality settings of the JPEG encoding for *low*, *medium*, and *high* level of detail were 20, 40 and 60 respectively.

Taking into account that there are 3 images available, 2 compression methods and 3 levels of detail per method, a total of 18 features are generated per image.

Table 1. Fractal image compression parameters

	low	medium	high
Image size	256×256 pixels		
Minimum partition level	2	2	3
Maximum partition level	4	5	6
Maximum error per pixel	8	8	8

4   Experiments

This section details two experiments related to aesthetics, (i) a classification one using two different approaches (ANNs and SVMs) and (ii) a sorting one. The dataset used in the first task is explained next.

4.1   Dataset

The features presented have been tested on a collection of images previously used for aesthetic classification tasks [4,11]. It is a large and diverse set of ranked photographs for training and testing available via `http://ritendra.weebly.com/aesthetics-datasets.html`. This address also provides more recent datasets, but we are not aware of any published results using them. All of these images were taken from the photography portal "photo.net". This website is an information exchange site for photography with more than 400,000 registered users. It comprises a photo gallery with millions of images taken by thousands of photographers. They can comment on the quality of the pictures by evaluating their aesthetic value and originality, assigning them a score between 1 and 7. The dataset included color and grayscale images. Additionally, some of the images have frames. None of these images was eliminated or processed. Because of the subjective nature of this problem, both classes were determined by the average user ratings.

This dataset includes 3581 images. All the images were evaluated by at least two persons. Unfortunately, the statistical information from each image, namely number of votes, value of each vote, etc., is not available. Like in the previous approaches, they considered two image categories: the most valued images (average aesthetic value $\geq$ 5.8, a total of 832 images) and the least valued ones ($\leq$ 4.2, a total of 760 images), according to the ratings given by the users of the portal. Images with intermediate scores were discarded. Datta's justification for making this division is that photographs with an intermediate value "are not likely to have any distinguishing feature, and may merely be representing the noise in the whole peer-rating process" [4]. However, when we carried out our experiment, some of the images used by Datta were not longer available at "photo.net", which means that our image set is slightly smaller. We were able to download 656 images with a rating of 4.2 or less, and 757 images with a rating of 5.8 or more. Out of the available images, about 7.4% are in grayscale.

4.2   Aesthetic Classification

The difference existing between the dataset of Datta et al. and the proposed one as regards the number of images used makes it impossible to compare the results. Having the input data of his experiment, as well as the input parameters, we have reproduced his experiment using only those images that we were able to retrieve. They perform classification using the standard RBF Kernel ($\gamma = 3.7$, $cost = 1.0$) using the LibSVM package [3] and a 5-fold cross-validation (5-CV). Their success rate using this configuration was 70.12%. On our behalf, with their

input data and the images available, 71.44% of images are classified correctly. The difference between both results shows that the task performed in this paper is less complicated than the original one. We will compare our results with the latter from now on.

We have used two different approaches in order to compare the functioning of the metrics proposed. One of them is based on Support Vector Machines (SVMs), while the other one is based on Artificial Neural Networks (ANNs). In the case of SVMs, we have decided to use the standard Linear Kernel configuration using the LibSVM package [19][3]. The success rate achieved in that case was 72.43%.

The other classifier is composed of a feed-forward ANN with one hidden layer. For training purposes, we resorted to SNNS [21] and standard back-propagation. The values that result from the feature extractor are normalized between 0 and 1. The results presented in this paper concern ANNs with one input unit per feature, 12 units in the hidden layer, and 2 units in the output layer (one for each category). A training pattern specifying an output of $(0, 1)$ indicates that the corresponding image belongs to the "low quality" set. Likewise, a training pattern with an output of $(1, 0)$ indicates that the corresponding image belongs to the "high quality" set. For each experiment we perform 50 independent repetitions of the training stage so as to obtain statistically significant results. For each of these repetitions we randomly create training, test, and validation sets with respectively 80%, 5%, and 15% of the patterns. The training of the ANNs is halted at 400 training cycles, or an RMSE in both the training and test sets lower than 0.01 is reached. Some other parameters used are shown in table 2. The results obtained with ANNs are very similar to those of SVMs, with a validation success rate of 71.16%.

Table 2. Parameters relative to the ANNs

Parameter	Setting
Init. of weights	random, $[-0.1, 0.1]$
Learning rate	0.15
Shuffle weights	yes
Class distribution	one-to-one
Max. tolerated error	0.3

4.3   Image Ranking

We will try to show the aptness of our metrics visually by showing the sorting capacity of the images obtained from a web search application and previously used by Ke et al. [10]. They used Google and Flickr to search for six image sets, labeled "apple", "bmw", "cow", "rose", "Statue of Liberty", and "violin". The retrieved images were then ranked by their quality assessment algorithm with a success rate of 72% obtained with a dataset of 12,000 images coming from the photography portal "DPChallenge.com".

The advantage of using a neural network lies in achieving two continuous outputs with values that can be used for another purpose, for instance, as fitness

function determining the aesthetic quality of a particular image. In our case, we will use both neural network outputs in order to create the formula 2 which will be used as sorting criterion, having been used by [13]:

$$\frac{(O_1 - O_2) + 1}{2} \tag{2}$$

In this case, O_1 and O_2 will correspond to the ANN outputs. In case the first one has a high value, the ranking value obtained will be close to 1, which indicates, in our case, a high aesthetic quality. However, in case the value of the second output is higher, then the ranking value will be close to 0, indicating a low aesthetic quality. When $O_1 = O_2$ the ranking value will be 0.5.

Following the approach of Ke et al. [10], in Figure 1 displays ends of the sorting, that is, the three best and the three worst. It is also important to observe what happens in the intermediate area of the ranking. In Figure 2 we present the entire list of images from the gallery retrieved by the search word "rose" sorted accordingly to formula 2. The full sorted lists of each of the 6 image sets are available on the Internet at `http://193.147.35.124/papers/evomusart2011`.

Taking into account the network outputs and the formula proposed, the values given to each image should be distributed in a space with range [0,1]. Due to the training model proposed for the ANN, the interval [0, 0.3] equals 0 and the interval [0.7, 1] equals 1. Thanks to that, the network output models can have a more linear approach, thus allowing the exploration of the ends, as done by [14]. In that particular case, the end values seen in Figure 1 are located within the range [0.85, 0.25]

In the subjective perspective of authors, the sorting achieved is far from perfect but quite successful from the point of view of aesthetics, particularly in what concerns the "best" and "worst" images of each set, albeit some isolated exceptions. One of these exceptions is "Statue11", which we consider as one of the best images of the subset.

By analyzing the sorted lists produced by the proposed approach one can try understand how the rankings are being determined. The results indicate that the best valued images tend to be those where the difference between the figure and the background is more evident, as well as those that have high contrast. It seems that two of the most determining elements are: the simplicity of the background (either due to flat elements or due to a low depth of field leading to an unfocused background); the existence of a significant difference between the background and the figure in the foreground. The image contrast can be also a decisive element, together with the existence of pure white and deep black, and a well-balanced distribution of both. For instance, image "Cow33" in Figure 1 has deviation similar to the deviations of the best valued high-contrast images, however, unlike them, it is underexposed, which causes a lack of information in the highlights and a trimming in the shadows, making it harder to differentiate between the background and the figure.

The rankings produced cannot be fully explained by these factors alone and the exact sorting method of the system is far from being understood.
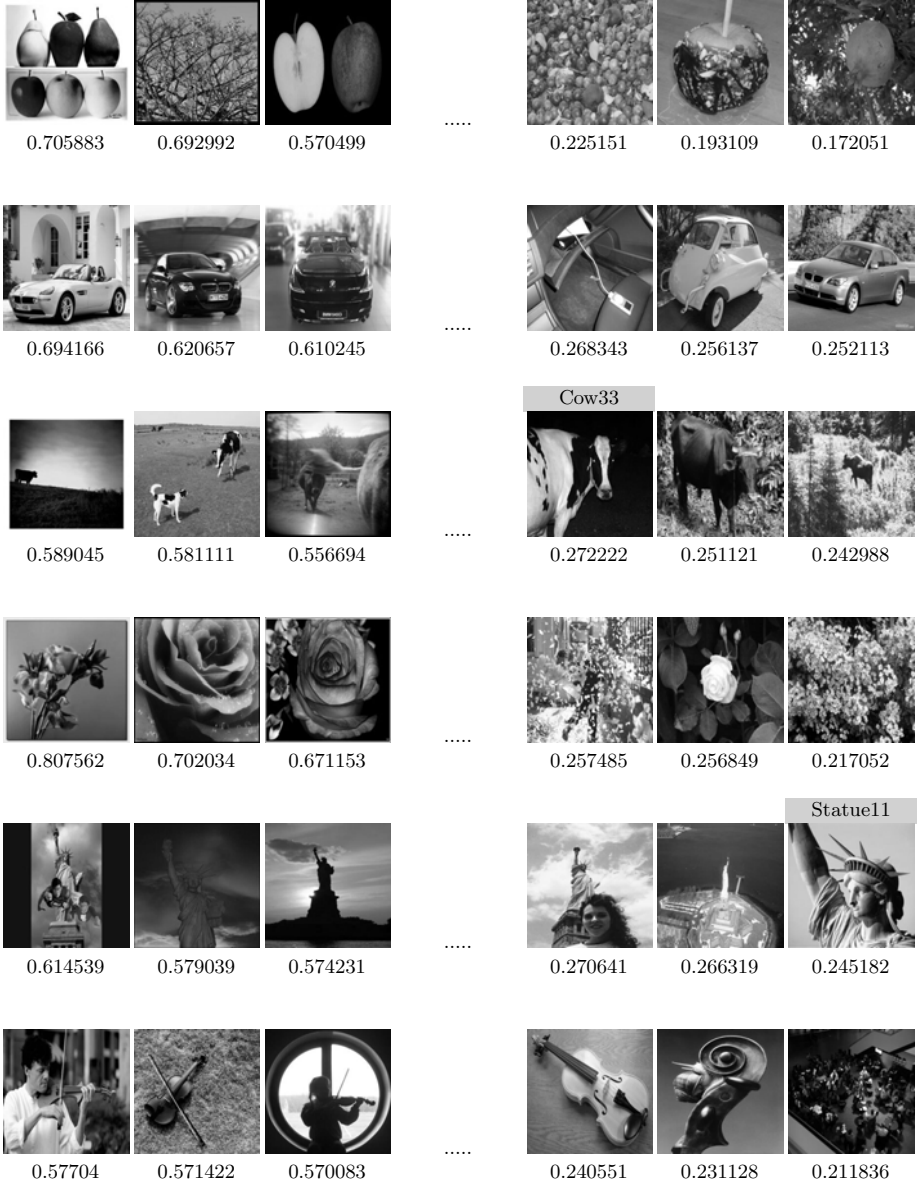
Fig. 1. End images of each gallery with its associated aesthetic value. Each set is shown in a row, with the three "best" images on the left and the three "worst" on the right.

0.807562	0.702034	0.671153	0.627365	0.515773	0.484149	0.481490
0.472965	0.471704	0.469493	0.466644	0.463819	0.455546	0.449485
0.449094	0.448771	0.448690	0.446400	0.426468	0.417754	0.407865
0.403056	0.400600	0.400134	0.396309	0.380985	0.376629	0.375872
0.374920	0.365691	0.348017	0.341926	0.341267	0.334960	0.326114
0.319226	0.314488	0.314461	0.310659	0.308311	0.301810	0.296129
0.291962	0.279273	0.276291	0.257485	0.256849	0.217052	

Fig. 2. Whole sorting list of the image gallery "rose"

Therefore, we can state that among the worst classified images most of them have brightness levels tending towards a concentration at the medium values of the image, together with over and underexposed ones.

5   Conclusions and Future Work

It has been shown how a set of 18 metrics based on two widespread compression methods can be used for image classification and sorting tasks. An experiment of aesthetic classification of images was carried out achieving similar results to other ad-hoc metrics specifically developed for that purpose, using two different approaches: one of them based on SVMs and the other one based on ANNs. A sorting function based on the output of the ANN used in the classification experiment was proposed and its functioning when sorting particular image sets based on aesthetic criteria presented and discussed.

In the future, the research will be expanded to cover other metrics related to complexity in both tasks. The purpose is using a large set of metrics to develop a fitness function within our own evolutionary engine.

Acknowledgments

References

1. Arnheim, R.: Art and Visual Perception, a psychology of the creative eye. Faber and Faber, London (1956)
2. Birkhoff, G.D.: Aesthetic Measure. Harvard University Press, Cambridge (1932)
3. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines (2001), software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm
4. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Studying aesthetics in photographic images using a computational approach. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 288–301. Springer, Heidelberg (2006)
5. Eysenck, H.J.: The empirical determination of an aesthetic formula. Psychological Review 48, 83–92 (1941)
6. Eysenck, H.J.: The experimental study of the 'Good Gestalt' - A new approach. Psychological Review 49, 344–363 (1942)
7. Fisher, Y. (ed.): Fractal Image Compression: Theory and Application. Springer, London (1995)
8. Graves, M.: Design Judgment Test. The Psychological Corporation, New York (1948)

9. Greenfield, G., Machado, P.: Simulating artist and critic dynamics - an agent-based application of an evolutionary art system. In: Dourado, A., Rosa, A.C., Madani, K. (eds.) IJCCI, pp. 190–197. INSTICC Press (2009)
10. Ke, Y., Tang, X., Jing, F.: The Design of High-Level Features for Photo Quality Assessment. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 419–426 (2006)
11. Luo, Y., Tang, X.: Photo and video quality evaluation: Focusing on the subject. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 386–399. Springer, Heidelberg (2008)
12. Machado, P., Cardoso, A.: Computing aesthetics. In: de Oliveira, F.M. (ed.) SBIA 1998. LNCS (LNAI), vol. 1515, pp. 219–229. Springer, Heidelberg (1998)
13. Machado, P., Romero, J., Manaris, B.: Experiments in Computational Aesthetics. In: The Art of Artificicial Evolution. Springer, Heidelberg (2007)
14. 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. Springer, Heidelberg (2007)
15. Machado, P., Romero, J., Santos, A., Cardoso, A., Manaris, B.: Adaptive critics for evolutionary artists. In: Raidl, G.R., Cagnoni, S., Branke, J., Corne, D.W., Drechsler, R., Jin, Y., Johnson, C.G., Machado, P., Marchiori, E., Rothlauf, F., Smith, G.D., Squillero, G. (eds.) EvoWorkshops 2004. LNCS, vol. 3005, pp. 435–444. Springer, Heidelberg (2004)
16. Meier, N.C.: Art in human affairs. McGraw-Hill, New York (1942)
17. Moles, A.: Theorie de l'information et perception esthetique, Denoel (1958)
18. Tong, H., Li, M., Zhang, H., He, J., Zhang, C.: Classification of Digital Photos Taken by Photographers or Home Users. In: Aizawa, K., Nakamura, Y., Satoh, S. (eds.) PCM (1). LNCS, vol. 3332, pp. 198–205. Springer, Heidelberg (2004)
19. Witten, I.H., Frank, E.: Data mining: practical machine learning tools and techniques with java implementations. SIGMOD Rec. 31(1), 76–77 (2002)
20. Wong, L., Low, K.: Saliency-enhanced image aesthetics class prediction. In: ICIP 2009, pp. 997–1000. IEEE, Los Alamitos (2009)
21. Zell, A., Mamier, G., Vogt, M., Mache, N., Hübner, R., Döring, S., Herrmann, K.U., Soyez, T., Schmalzl, M., Sommer, T., et al.: SNNS: Stuttgart Neural Network Simulator User Manual, version 4.2. Tech. Rep. 3/92, University of Stuttgart, Stuttgart (2003)