Journal of Mathematics and the Arts
Vol. 00, No. 00, Month 200x, 1–13

submitted version

Using Complexity Estimates in Aesthetic Image Classification

Juan Romero[a]*, Penousal Machado[b], Adrian Carballal[a] and Antonino Santos[a]

[a] *Artificial Neural Networks and Adaptive Systems LAB, University of A Coruña 15071 A Coruña, Spain*
[b] *CISUC, Department of Informatics Engineering, University of Coimbra 3030 Coimbra, Portugal*

(*Received 00 Month 200x; final version received 00 Month 200x*)

In recent years the search for computational systems that classify images based on aesthetic properties has gained momentum. Such systems have a wide range of potential applications, including image search, organization, acquisition and generation. This work explores the use of complexity estimates to predict the aesthetic merit of photographs. We use a set of image metrics and two different classifiers. Our approach classifies images gathered from a photography web site, attempting to reproduce the evaluation made by a group of users. For this purpose we use complexity estimate metrics based on the encoding size and compression error of JPEG and fractal compression, which are applied to the original Value channel and to the images resulting form applying Sobel and Canny filters to this channel. By employing these estimates, in conjunction with the average and standard deviation of the value channel, i.e. 20 features, a success rate of 74.59% was attained. Using the three most influential features yields a success rate of 71.34%, which is competitive with the best results reported in literature, 71.44%, using the same dataset.

Keywords: Aesthetics, JPEG compression, fractal compression, complexity, edge-detection

AMS Subject Classification: 68T05; 68T45; 68U10;

1. Introduction

Automatic image classification based on aesthetic criteria would allow image browsers to consider the aesthetic preferences of the user, and enable online sites of artistic works, designs, or photographs to recommend works that are consistent with previous purchases. A system of this type could also be used to automatically classify artistic databases and to generate images of a given style or current aesthetic preference (by integrating the classifier as the evaluator of an image generation system) [2, 17]. We propose the use image complexity estimates and edge-detection filters to classify images based on aesthetic criteria.

Performing the aesthetic classification of a work is by no means an easy task [8]. It must be understood that such a task may possess highly subjective aspects, and that taste depends, to a great extent, on various dynamic variables, such as experience, education, culture, ideology, etc. Having a system able to simulate or reproduce the aesthetic preferences of a given individual or of a particular cultural environment would entail significant progress for image classification systems.

Evaluating the proper functioning of a classifier is also a complex task. For that reason, we will employ datasets and classification technologies adopted by other researchers; we will also study the behavior of the proposed metrics by developing classifiers based on Support Vector

*Corresponding author. Email: jj@udc.es

ISSN: 1751-3472 print/ISSN 1751-3480 online

DOI: 10.1080/1751347YYxxxxxxxx
http://www.tandfonline.com

Machines (SVMs) and on Artificial Neural Networks (ANNs). The current paper builds upon, and extends, our previous research on the same topic, namely the work presented in [21].

Historically, the relationship between aesthetics and image complexity has primarily been explored by researchers with a background in psychology or computation [3, 9, 10, 19]. In a simplified way, their work suggests that the complexity of an image is related to its entropy, and inversely related to the order.

Assuming an interpretation of complexity akin to the notion of Kolmogorov-complexity – which defines complexity as the size of the smallest program required to reproduce a given object – the complexity of an image is related with the degree of predictability of each pixel of the image. Therefore, an image with every pixel having the same color shows perfect order and has little complexity, while a white noise image can be seen as extremely complex since it is impossible to predict the value of each pixel.

Unfortunately, Kolmogorov-complexity is non-computable: while it is conceivable to calculate it for some particular objects, it is impossible to calculate it in finite time for a generic object. As such, in general, the best that can be attained are estimates of the Kolmogorov-complexity of an object. Additionally, the size of the minimum program that encodes an object depends not only on the object but also on the "machine" for which the program was built, i.e, Kolmogorov-complexity is machine dependent.

Considering the purposes of the current study, one should seek Kolmogorov-complexitiy estimates that relate well with complexity as perceived by humans. The most popular image compression schemes are lossy, the encoding of the images involves a loss of detail that is, hopefully, negligible and undetectable by the human eye. As such these compression schemes may yield complexity estimates that relate well with the human notion of complexity. The relation between image compression and perceived complexity is also supported by recent studies in the field of psychology [11]. In this study we will use two lossy compression techniques, JPEG and fractal encoding, to estimate the perceived complexity of the images.

Our purpose is *not* to demonstrate the existence of universal aesthetic principles or rules. The authors do believe, however, that some characteristics of the image – e.g. balance, symmetry, proportion, rhythm – and the way they are perceived influence the aesthetic reaction of the viewer, although different viewers may react differently to them. We also believe complexity to be one of the characteristics that influences the aesthetic reaction of the viewer.

A set of photos with different subjects, authors, etc. is not ideal for testing classification systems based on aesthetics. It can even be debated whether if makes sense to estimate the aesthetic value of photos without taking their content into account. It is reasonable to assume that the relatively low success rates attained are partially due to the way content may have influenced the perception of aesthetic value of the Photo.net users who scored the images.

In spite of these limitations, using this set of images has two important advantages: (i) Photo.net specifically asked the users to evaluate the image in terms of "aesthetics" and "originality" (ii) the set was used in previous published works.

Since the users may judge independently "aesthetics" and "originality" it is reasonable to assume that the aesthetic ratings are closer to the perceived aesthetic value of the photographs than those that would be obtained by asking for a generic evaluation of the photograph. Nevertheless, Datta et al. [7] report a high correlation between "aesthetic" and "originality" ratings, which indicates that most users do not tend to differentiate between these concepts and that aesthetics cannot be easily isolated from issues such as novelty and content.

Using a dataset employed by other researchers has the advantage of allowing the comparison of results and performing incremental research, something that is uncommon in this area.

The rest of the paper is organized in the following manner: a short summary of other approaches existing in this field, an explanation of the classifier and its composing parts, a description of the experimental setup, an analysis of the experimental results and, finally, the conclusions of this work.

2. State of the Art

There is very little literature concerning the automatic classification of images based on aesthetic criteria. Most researchers use ad-hoc metrics focusing mainly on technical and/or composition issues. Relevant examples are work carried out by Tong et al. [25], Ke et al. [14] and Datta et al. [7]. The latter two were the very first to devise various metrics expressly for the purpose of evaluating the aesthetic quality of photographs.

Tong et al. [25] use a large set of low level features to distinguish between 12,897 amateur photographs, taken by workers at Microsoft Asia, and a set of 16,643 professional photographs, obtained from Corel Image Database and Microsoft Office Online. By using different classification methods, including a Bayesian classifier, they attain a success rate above 95% on a testing set of 379 images.

Yan Ke et al. [14] propose the use of high-level features (such as spatial distribution of edges, color distribution, blur, hue count) to distinguish between "high quality professional photos" and "low quality snapshots". These categories are created based on evaluations from the "DPChallenge.com" photography portal. A set of 60,000 images gathered from the portal is sorted by average rating. The top 3,000 and bottom 3,000 images with more than 100 user evaluations are considered for the purposes of the experiments while the rest is discarded. Using the set of features proposed in [14] yields a correct classification rate of 72%. Using a combination of these metrics together with those published by Tong et al. [25], Yan Ke et al. [14] achieved a success rate of 76%.

Datta et al. [7] employed color, texture, shape and composition high level ad-hoc features and SVMs to classify images gathered from a photography portal (Photo.net). They considered two image categories: the most valued images (average aesthetic value ≥ 5.8, a total of 832 images) and least valued ones (≤ 4.2, a total of 760 images), according to the ratings given by the users of the portal. Images with intermediate valuations were discarded. The system attains 70.12% classification accuracy.

In order to create a basis for research on aesthetic classification, Datta et al. [8] proposed three types of aesthetic classification: aesthetic score prediction; aesthetic class prediction; emotion prediction. They also published a set of datasets that includes the one employed in [7], and two others extracted from Photo.net and DPChallenge.com which comprise more images and statistical information. In this work, we use one of these image datasets to allow the comparison with previous approaches. However, we apply a different set of metrics and classification techniques.

Datta et al. [6] implemented a public image rating system entitled ACQUINE (Aesthethic Quality Inference Engine), available at `acquine.alipr.com` (accessed 7 March 2012). It proposes an aesthetic score between 0 and 100 to the images uploaded by the users. Following previous work [7, 8] ACQUINE's engine is based on a two-class SVM classifier trained with a dataset comprising over 20000 photographs and ratings retrieved from Photo.net.

It is clear that the automatic aesthetic assessment of images is an unsolved problem. Nevertheless, works such as [7, 14, 21] made relevant contributions, paving the way to further research in the area and to the development of systems with such abilities.

2.1. *Complexity and Aesthetics*

The relevance of perceived image complexity is a recurring topic in the field of aesthetics [1, 3, 20]. Inspired by these theoretical works, Machado et al. proposed [16] complexity estimates which [almost] matched human competitive results in a psychological test named the ?Design Judgment Test" [12]. In [18], Machado et al. used a subset of the features proposed in this paper and an Artificial Neural Network (ANN) classifier for author identification, attaining identification rates higher than 90% across experiments. In [17], Machado et al. explored their use to assign fitness to the images evolved by an evolutionary art tool. Saunders and Gero [22] and Svangard and

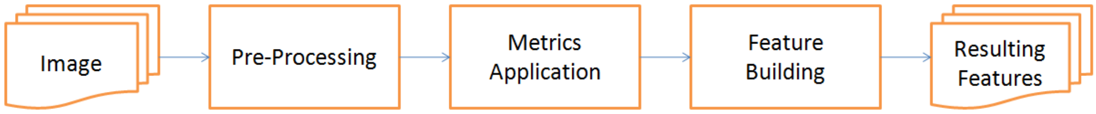


Figure 1. Feature extraction pre-processing steps.

Nordin [24], among others, follow a similar line of work.

3. Proposed Features

In this section we describe the feature extraction process. The feature extraction can be summarized using the following steps (see Figure 1):

(1) Pre-processing – which includes all the transformation and normalization operations applied to a given input image;
(2) Edge-Detection filters application – to identify points in images at which the image brightness has discontinuities.
(3) Metrics application – the application of certain operations and calculations based on statistical methods and image complexity estimates; methods based on statistical measurements and image complexity estimates;

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-bits 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.

The need for these normalization operations results from the characteristics of the metrics employed, which are sensitive to size and resolution. These transformations result in a loss of information that may hinder the performance of the system. However, preliminary empirical tests indicate that, for the set of features considered, performing such operations yields better results than using the original images or resized versions of the images with the same proportions as the original.

Next, the image is converted to the HSV (Hue, Saturation and Value) color space and its HSV channels are split. When we only need the image representation in black and white format, the V-channel, serving as a 1-channel grayscale image, is stored.

Previous works such as [7, 14, 15] rely, to a large extent, on color information to extract features. Ke et al. [14] states "the color palette seen in professional photos and snapshots is likely to be very different". In this paper, we rely exclusively on features obtained from a single channel interpreted as if it contained grayscale information. We want to make the system as generic as possible, and in every dataset we have there are some grayscale images. This means, later, when we want to provide results using all the color information channels we will need to recode the H and S channels and extract features from them separately.

3.1. *Edge-Detection Filter Application*

Once the grayscale image is available, two edge detection filters are applied, Canny and Sobel, which will yield two new black and white images (see Figure 2). In previous works (e.g., [14, 25]) filters such as Canny, Sobel, Gauss and Laplace have been applied.

Sobel's edge-detection method [23] is a discrete differentiation operator. It computes an approximation of the gradient of the image intensity function. The Sobel operator uses two 3x3 kernels, one for horizontal changes and one for vertical changes, which are convolved with the

(a) No Filter (b) Canny Filter (c) Sobel Filter

Figure 2. Edge-detection filters application.

original image, yielding the norm of the gradient vector for each pixel. Informally, this results in an image that highlights abrupt changes.

Canny's edge-detection algorithm [4] results from the attempt to develop an edge detector with good: *detection*, all real edges should be identified; *localization*, the marked edges should be as close as possible to the real edges; *response*, each edge should only be marked once. The algorithm includes four stages: noise reduction, which is accomplished by applying a convolution filter that "blurs" the original image; gradient calculation, which applies four convolution filters to detect horizontal, vertical and diagonal edges; non-maximum suppression, which determines the dominant direction of the edge; thresholding with hysteresis, which discards isolated gradient changes and values gradient changes that form lines.

Canny's edge-detection algorithm is more sophisticated than Sobel's and, therefore, it is likely to be more useful. On the other hand, Sobel's approach gives an indication of the strength of the edges. This additional information that may prove valuable by itself or when combined with the information provided by Canny's method.

3.2. *Metrics Application*

Most image compression schemes used are lossy, therefore there is a compression error, i.e., the compressed image will not exactly match the original. All other factors being equal, complex images will tend toward higher compression errors and simple images will tend toward lower compression errors. Additionally, complex images will tend to generate larger files than simple ones. Thus, compression error and file size are positively correlated with image complexity.

To explore these aspects, we consider three levels of detail for the JPEG and fractal compression metrics: low, medium, and high. For each compression level the process is the same, the image being analyzed, I, is encoded in JPEG or fractal format, and its complexity estimated using the following formula:

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

where $RMSE$ stands 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 here, we use a quadtree fractal image compression scheme 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 first partitioned into 64 blocks. 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,

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

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.

There are three filtering options (no filtering, Canny and Sobel); two compression methods (JPEG and fractal) and three levels of compression detail (high, medium and low). The combination of these options yields 18 different features per channel. In addition, we use two statistical metrics: Average (Avg) and Standard Deviation (StD). The average and the standard deviation are calculated using the pixel intensity values of each image. This results in a total of 20 image features per channel.

4. Assessing the merit of the proposed features

The present section studies the usefulness and relevance of each of the proposed features for an image classification task focusing on aesthetic criteria, following previous work mentioned in Section 2. We begin by describing the dataset used in these experiments.

4.1. *Dataset*

The features presented have been tested on a collection of images previously used for aesthetic classification tasks [7, 15]. It is a large and diverse set of ranked photographs for training and testing available via `http://ritendra.weebly.com/aesthetics-datasets.html` (accessed 7 March 2012). This web 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. Its users 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 includes 3,581 images. All the images were evaluated by at least two persons. We use the average of the "aesthetic" scores assigned by the users of Photo.net to each image. Using the average may attenuate the influence of aspects not related with aesthetics in the judgments made.

The dataset includes color and grayscale images. Additionally, some of the images have frames. None of these images were eliminated or additionally processed. To allow comparison with previous works, the data used to obtain the results from our experiments, namely the image identification numbers and average ratings, was gathered from the previously mentioned web page maintained by Datta et al. The current rating of these images is likely to have changed, and presently Photo.net does not ask for separate assessments for aesthetics and originality. Unfortunately, the raw data from each image, namely number of votes, value of each vote, etc., is not available.

Following the approach of previous works [7, 8, 15], two image categories are considered: the most valued images (average aesthetic value ≥ 5.8, a total of 832 images) labeled as "favored" and the least valued ones (≤ 4.2, a total of 760 images) labeled as "unfavored", according to the ratings given by the users of the portal. Images with intermediate scores were discarded. The justification of Datta et al. 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" [7]. Some examples of images randomly chosen from the

Figure 3. Random samples from the dataset of 3,581 images gathered from Photo.net.

dataset are included in Figure 3.

When we performed our experiments, some of the images used by Datta et al. were no 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.

The difference between the number of photos in the dataset of Datta et al. [7] and the number of photos in the subset we were able to procure makes a straightforward comparison of results impossible. However, Datta et al. have provided the feature values for all the images used in their experiments, as well as the input parameters for their classifiers. This allowed us to faithfully replicate their experiments using the subset of images that are currently available. Datta et al. perform classification using the standard RBF Kernel (gamma=3.7, cost=1.0) using the LibSVM package [5] and a 5-fold cross-validation (5-CV). Their success rate using this configuration was 70.12%. In our replication of their experiments using their input data and the images currently available 71.44% of the images are correctly classified. Since the original performance was slightly lower, we will compare our approach with the results obtained in the reproduction of the experiments in order to perform an unbiased comparison.

4.2. *Binary classification based on aesthetics*

We have built classifiers using two different approaches, SVMs and ANNs, in order to compare the functioning of the metrics proposed.

SVMs represent the input data in a decision space with a dimensionality equal to the number of features considered. The input data, which is typically not-linearly separable in this feature space, is mapped by a kernel into a space of higher dimensionality. This transformation allows the definition of boundary hyperplanes that separate and group the input data according to the specified categories and that can be re-mapped to the original feature space. The SVM classifiers were built using the standard Linear Kernel configuration using the LibSVM package [5, 27].

The ANN-based classifiers employ a feed-forward ANN with one hidden layer. For training purposes, we resorted to SNNS [28] 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, i.e. 20 input neurons for classifiers that only use information gathered from one of the HSV channels and 60 input neurons for classifiers using information gathered from the three color channels. All ANNs have 12 units in the hidden layer, which are fully connected with the input and output layers. Although the number of input units varies, adjusting the number of hidden units did not provide better performance. The

Table 2. Parameters relative to the ANNs.

Parameter	Setting
Initial weight interval	[-0.1, 0.1]
Learning rate	0.15
Shuffle weights	yes
Class distribution	one-to-one
Max. tolerated error	0.3

Table 3. Classification rate achieved according to color channels and techniques used. The success rate attained using Datta et al.'s [7] approach is 71.44%. The "Favored" and "Unfavored" columns show the success rate for each of the two categories labeled under the same name. The "Overall" column shows the overall success rate with the dataset used.

		SVMs			ANNs		
Color Channel	Features	Overall	Favored	Unfavored	Overall	Favored	Unfavored
HSV	60	75.81%	71.00%	80.62%	75.37%	71.76%	78.98%
V	20	74.59%	67.07%	82.11%	73.27%	66.71%	80.05%
S	20	68.09%	54.20%	81.98%	68.90%	60.53%	77.30%
H	20	65.45%	53.12%	77.78%	65.49%	56.05%	74.96%

output layer has 2 units (one for each category). A training pattern specifying an output of (0, 1) indicates that the corresponding image belongs to the "unfavored" category. Likewise, a training pattern with an output of (1,0) indicates that the corresponding image belongs to the "favored" category. 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 when a RMSE in both the training and test sets lower than 0.01 is reached. The remaining parameters used in ANN training are presented in Table 2. The parameters used in both types of classifier were established empirically in previous experiments.

The success rate achieved by the SVM classifiers in the validation set when using exclusively the V channel of the image was 74.59%. The results obtained with ANNs in the same circumstances are similar to those of SVMs, with a validation success rate of 73.27%.

Table 3 presents the classification rates achieved by each technique using each of the H, S, and V channels independently, and also combining them so that, with a total of 60 features available, a combined HSV evaluation could be performed. Comparing the results using the color channels independently, it becomes clear that applying our metrics and filters to the V channel is more useful than applying them to only the H channel or only the S channel. When we compare the results obtained using only the V channel to those obtained using the full HSV color information, we observe that the improvement of performance is only marginal, 1.22% for SVM classifiers and 2.1% for ANN classifiers. It should be noted that only 7.4% of the dataset images are grayscale and that the HSV classifiers use 60 features instead of the 20 features used by the classifiers that only use information gathered from the V channel.

These results indicate that either our metrics are unable to extract additional meaningful information from the color channels, or that in this particular setting color information is less relevant than previous studies [7, 14, 15] indicate. The results presented are better than those reported in literature, which appears to reinforce this second explanation. Our interpretation points towards a confluence of both factors: (i) in our dataset images that have good coloring tend to also be good when viewed in grayscale and vice-versa so the additional information available tends to be of little use; (ii) the features used by us, and by other researchers, do not appear to adequately grasp color information that is relevant for aesthetic judgement.

Figures 4 and 5 show some examples of images which the system was *unable* to classify correctly. Figure 4, presents images belonging to the favored category that have been classified by the system as belonging to the unfavored category. Figure 5, depicts the opposite situation, where images belonging to the unfavored category have been classified as belonging to the favored

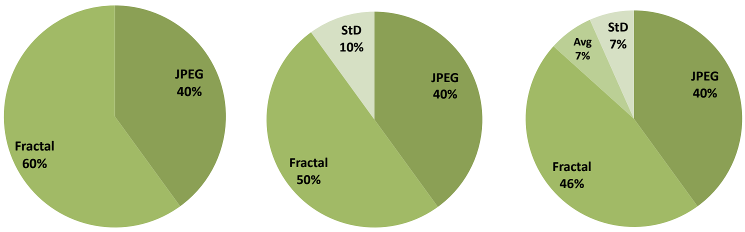
Figure 4. Examples of images belonging to the favored category which have been classified by the system as belonging to the unfavored category.

Figure 5. Examples of images belonging to the unfavored category which have been classified by the system as belonging to the favored category.

one.

Table 4. Ranking of features according to their importance as assessed by the SVMAttributeEval method.

Rank	Family	Type	Filter	Rank	Family	Type	Filter
1	JPEG	High	Sobel	11	Fractal	Medium	Canny
2	Fractal	High	No filter	12	JPEG	High	No filter
3	Fractal	Low	Canny	13	JPEG	Low	Sobel
4	JPEG	High	Canny	14	Fractal	Medium	No filter
5	Fractal	Low	No filter	15	Avg	No	filter
6	Std	No filter		16	Fractal	Low	Sobel
7	Fractal	Medium	Sobel	17	JPEG	Medium	No filter
8	Fractal	High	Canny	18	Fractal	High	Sobel
9	JPEG	Medium	Canny	19	JPEG	Low	Canny
10	JPEG	Medium	Sobel	20	JPEG	Low	No filter



(a) First 5 (b) First 10 (c) First 15

Figure 6. Prevalence of features using JPEG and fractal compression among the 5, 10, and 15 most influential features.

4.3. *Seeking the most relevant features*

For each image in our dataset we obtain 18 feature values by applying JPEG and fractal compression to three images — the grayscale image derived from the V channel of the original image, a Sobel filtering of this grayscale image, and a Canny filtering of this grayscale image (see Figure 2). Further, we calculate the average and standard deviation of the V channel, which results in two additional feature values. The present section focuses on the analysis of the relevance of these features and their attributes in our discrimination task.

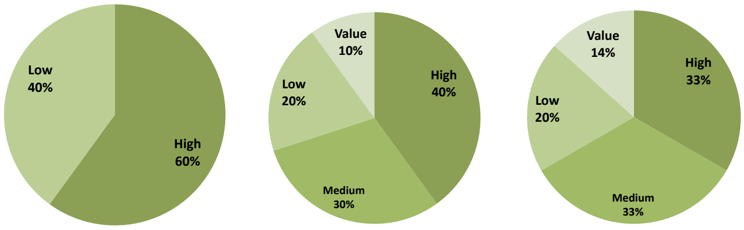
The metrics introduced have been ordered according to the "SVMAttributeEval" method provided by the WEKA tool. Given that this is a wrapper algorithm, it evaluates the attributes using the accuracy estimates provided by a particular classification algorithm. In this case, the features are classified by the squared weight assigned by the SVM.

SVMAttributeEval combines a linear support vector machine (SVM) [26] and the technique of recursive feature elimination (RFE) in order to assess the relevance of attribute subsets [13]. Table 4 presents the order of importance of the 20 features according to SVMAttributeEval. Figures 6, 7 and 8 summarize the influence of the features according to their nature, grouping the 5, 10 and 15 most influential features specified in Table 4.

Perusing these results one can observe that, overall, fractal and JPEG compression have similar relevances. However, when we look at the data with respect to compression quality one may observe that higher compression quality tends to lead to higher relevance. In what concerns the use of edge detection, although the most influential feature uses Sobel's approach, overall, the most influential features are those using Canny's edge detection or the original V channel.

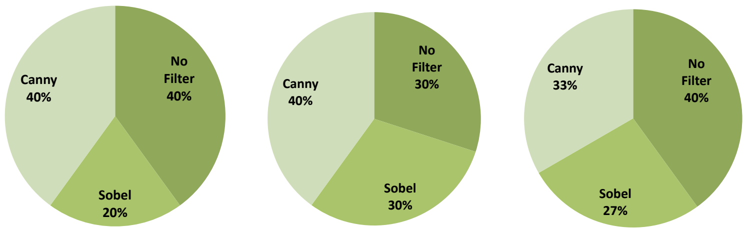
Based on the ordering presented in Table 4 we created ANN and SVM based classifiers using the x most influential features with $x \in [1, 20]$. The configuration parameters of these classifiers are equal to the ones described previously.

Table 3 and Figure 9 present the success rates attained by these classifiers. It should be noted that just using the three most influential features (see Table 5) produces results comparable to those of the state of the art (71.34% vs 71.44% obtained in the replication of the experiment) where 15 ad-hoc metrics are used. It is interesting to notice that these 3 features estimate the complexity of (i) the V channel, (ii) the V channel filtered by Sobel's method (iii) the V channel



(a) First 5 (b) First 10 (c) First 15

Figure 7. Prevalence of features using High, Medium and Low compression levels among the 5, 10, and 15 most influential features.



(a) First 5 (b) First 10 (c) First 15

Figure 8. Prevalence of features using no filtering, Sobel's edge detection and Canny's edge detection among the 5, 10, and 15 most influential features.

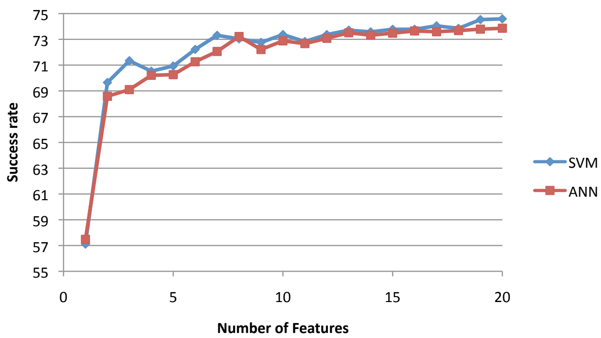


Figure 9. Success rate as the number of features increases for ANN and SVM classifiers.

filtered by Canny's method. Thus, they estimate the complexity of the image, its edges and its edge intensities.

Calculated over the full set of 656+757=1413 images, the (Pearson) correlation between the feature using the original V channel and the feature using Canny's edge detection is 0.832; between the feature using the original V channel and the feature using Sobel's edge detection is 0.912; and between the two edge detection features is 0.895.

As described in section 3.2, 18 of the 20 features can be categorized into three distinct groups according to the filtering operation performed. The average correlation among features where no edge detection operator was employed is 0.908 with a maximum deviation of 0.084. The correlation among the ones where Canny's edge detection was used is 0.987 with a maximum deviation of 0.009. Finally, the correlation among features using Sobel's edge detection is 0.953 with a maximum deviation of 0.04.

Using only the seven most influential features produces success rates of 73.31% for SVM classifiers and 72.06% for ANN classifiers, both surpassing the reported state of the art.

Table 5. Success rates attained by classifiers using the x most influential features according to SVMAtributeEval. The columns "Favored" and "Unfavored" show the success rates for each of the two corresponding image categories. The "Overall" column shows the overall success rate for the dataset used.

	SVMs			ANNs		
# Features	Overall	Favored	Unfavored	Overall	Favored	Unfavored
1	57.11%	73.44%	40.79%	57.48%	58.05%	56.90%
2	69.65%	65.31%	73.98%	68.58%	62.58%	74.60%
3	71.34%	63.01%	79.67%	69.10%	61.96%	76.27%
4	70.53%	58.27%	82.79%	70.21%	57.59%	82.87%
5	70.93%	58.67%	83.20%	70.26%	56.93%	83.64%
6	72.22%	67.21%	77.24%	71.26%	63.14%	79.41%
7	73.31%	67.89%	78.73%	72.06%	62.29%	81.87%
8	73.04%	66.12%	79.95%	73.22%	64.93%	81.54%
9	72.76%	65.99%	79.54%	72.21%	62.69%	81.76%
10	73.37%	65.72%	81.03%	72.88%	64.46%	81.32%
11	72.83%	63.82%	81.84%	72.67%	63.76%	81.61%
12	73.37%	66.53%	80.22%	73.08%	64.15%	82.04%
13	73.71%	66.94%	80.49%	73.51%	64.92%	82.13%
14	73.58%	66.80%	80.35%	73.32%	65.88%	80.79%
15	73.78%	67.75%	79.81%	73.48%	66.79%	80.19%
16	73.78%	67.62%	79.95%	73.65%	66.71%	80.61%
17	74.05%	67.48%	80.62%	73.59%	64.90%	82.31%
18	73.85%	67.34%	80.35%	73.68%	66.45%	80.93%
19	74.53%	67.07%	81.98%	73.79%	64.27%	83.35%
20	74.59%	67.07%	82.11%	73.86%	67.23%	80.52%

5. Conclusions and Future Work

It has been shown how a set of 20 low level features based on two widespread compression methods, two edge detection filters and two statistical metrics, can be used for image classification based on aesthetic criteria. The experimental results obtained from the aesthetic classification of photographs using these features produced success rates above 74.5%, which compares well with the results obtained from the same dataset by state of the art approaches using high level metrics specially developed for this task, which attain a maximum success rate of 71.44%.

Nevertheless, a 74.5% success rate in a binary classification task is far from optimal. The relatively low success rate can be explained by a wide variety of factors, including: the lack of controls in the evaluations made through a website; the subjectivity and diversity of opinions regarding aesthetics; the relatively low number of user evaluations per image. Additionally, when considering the variety of the images it becomes clear that the size of the dataset is insufficient for providing a representative enough sample to promote effective learning.

In spite of these shortcomings, using a dataset which was adopted in previous works by other researchers allows the comparison of the results. Further experiments using image datasets evaluated by viewers under a controlled experimental environment are already taking place. These experiments will also enable a statistical analysis of the viewers' evaluations and, therefore, comparison of the results of the classifier systems with those of the typical viewer.

Although in this paper we focused exclusively on aesthetic assessment, we are also interested in image generation. In that scope, the classifiers developed may help guide the generation process, not only by providing estimates of perceived aesthetic value but also by providing estimates of novelty of the generated images.

We have shown that the use of metrics based on complexity estimates yields results comparable with the state of the art, suggesting that our complexity metrics may capture some image characteristics that might be relevant for aesthetic judgment. As such, this work contributes to the development of systems that can automate or assist tasks of analysis and categorization of visual contents based on aesthetic properties.

Acknowledgments

This research is partially funded by: the Portuguese Foundation for Science and Technology, research project PTDC/EIAEIA/115667/2009; the Spanish Ministry for Science and Technology, research project TIN200806562/TIN; Xunta de Galicia, research project XUGAPGIDIT10TIC105008PR.

References

[1] R. Arnheim, *Art and visual perception, a psychology of the creative eye*, Faber and Faber, London 1956.
[2] S. Baluja, D. Pomerleau, and T. Jochem, *Towards automated artificial evolution for computer-generated images*, Connection Science 6 (1994), pp. 325–354.
[3] G.D. Birkhoff, *Aesthetic measure*, Harvard University Press 1933.
[4] J. Canny, *A computational approach to edge detection*, IEEE Trans. Pattern Anal. Mach. Intell. 8 (1986), pp. 679–698.
[5] C.C. Chang and C.J. Lin (2001), software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[6] R. Datta and J.Z. Wang, *ACQUINE: aesthetic quality inference engine - real-time automatic rating of photo aesthetics.*, in *Multimedia Information Retrieval*, URL http://dblp.uni-trier.de/db/conf/mir/mir2010.html#DattaW10, ACM, 2010, pp. 421–424.
[7] R. Datta, D. Joshi, J. Li, and J.Z. Wang, *Studying Aesthetics in Photographic Images Using a Computational Approach.*, in *Computer Vision – ECCV 2006, 9th European Conference on Computer Vision, Part III*, LNCS, Springer, Graz, Austria, 2006, pp. 288–301.
[8] R. Datta, D. Joshi, J. Li, and J.Z. Wang, *Image retrieval: Ideas, influences, and trends of the new age*, ACM Comput. Surv. 40 (2008), pp. 5:1–5:60, URL http://doi.acm.org/10.1145/1348246.1348248.
[9] H.J. Eysenck, *The empirical determination of an aesthetic formula*, Psychological Review 48 (1941), pp. 83–92.
[10] H.J. Eysenck, *The experimental study of the 'Good Gestalt' - A new approach*, Psychological Review 49 (1942), pp. 344–363.
[11] A. Forsythe, M. Nadal, N. Sheehy, C.J. Cela-Conde, and M. Sawey, *Predicting beauty: Fractal dimension and visual complexity in art*, British Journal of Psychology 102 (2011), pp. 49–70, URL http://dx.doi.org/10.1348/000712610X498958.
[12] M. Graves, *Design judgement test, manual*, The Psychological Corporation, New York 1948.
[13] I. Guyon and A. Elisseeff, *An introduction to variable and feature selection*, J. Mach. Learn. Res. 3 (2003), pp. 1157–1182, URL http://dl.acm.org/citation.cfm?id=944919.944968.
[14] Y. Ke, X. Tang, and F. Jing, *The Design of High-Level Features for Photo Quality Assessment*, Computer Vision and Pattern Recognition, IEEE Computer Society Conference on 1 (2006), pp. 419–426.
[15] Y. Luo and X. Tang, *Photo and Video Quality Evaluation: Focusing on the Subject*, in *ECCV (3)*, Lecture Notes in Computer Science, Vol. 5304, Springer, 2008, pp. 386–399.
[16] P. Machado and A. Cardoso, *Computing Aesthetics*, in *Proceedings of the XIVth Brazilian Symposium on Artificial Intelligence: Advances in Artificial Intelligence*, LNCS, Vol. 1515, Springer, Porto Alegre, Brazil, 1998, pp. 219–229.
[17] P. Machado, J. Romero, and B. Manaris, *Experiments in computational aesthetics: An iterative approach to stylistic change in evolutionary art*, in *The Art of Artificial Evolution: A Handbook on Evolutionary Art and Music*, J. Romero and P. Machado, eds., Springer Berlin Heidelberg, 2007, pp. 381–415.
[18] P. Machado, J. Romero, A. Santos, A. Cardoso, and B. Manaris, *Adaptive Critics for Evolutionary Artists*, in *Applications of Evolutionary Computing, EvoWorkshops 2004: EvoBIO, EvoCOMNET, EvoHOT, EvoIASP, EvoMUSART, EvoSTOC*, LNCS, Vol. 3005, 5-7 Apr., Springer, Coimbra, Portugal, 2004, pp. 435–444.
[19] N.C. Meier, *Art in human affairs*, McGraw-Hill 1942.
[20] A. Moles, *Thorie de l'information et perception esthtique*, Denoel 1958.
[21] J. Romero, P. Machado, A. Carballal, and O. Osorio, *Aesthetic Classification and Sorting Based on Image Compression*, in *EvoApplications (2)*, Lecture Notes in Computer Science, Vol. 6625, Springer, 2011, pp. 394–403.
[22] R. Saunders and J.S. Gero, *Curious agents and situated design evaluations*, AI EDAM 18 (2004), pp. 153–161.
[23] I. Sobel, *An isotropic 3 x 3 image gradient operator*, Machine Vision for Three-Dimensional Scenes (1990), pp. 376–379.
[24] N. Svangård and P. Nordin, *Automated Aesthetic Selection of Evolutionary Art by Distance Based Classification of Genomes and Phenomes using the Universal Similarity Metric*, in *Applications of Evolutionary Computing, EvoWorkshops 2004: EvoBIO, EvoCOMNET, EvoHOT, EvoIASP, EvoMUSART, EvoSTOC*, LNCS, Vol. 3005, Springer, Coimbra, Portugal, 2004, pp. 445–454.
[25] H. Tong, M. Li, H. Zhang, J. He, and C. Zhang, *Classification of Digital Photos Taken by Photographers or Home Users*, in *PCM (1)*, Lecture Notes in Computer Science, Vol. 3332, Springer, 2004, pp. 198–205.
[26] V. Vapnik, *An overview of statistical learning theory*, IEEE Transactions on Neural Networks 10 (1999), pp. 988–999.
[27] I.H. Witten and E. Frank, *Data mining: practical machine learning tools and techniques with java implementations*, SIGMOD Rec. 31 (2002), pp. 76–77.
[28] A. Zell, G. Mamier, M. Vogt, N. Mache, R. Hübner, S. Döring, K.U. Herrmann, T. Soyez, M. Schmalzl, T. Sommer, *et al.*, *SNNS: Stuttgart Neural Network Simulator User Manual, version 4.2*, Tech. Rep. 3/92, University of Stuttgart, Stuttgart, 2003.