Aesthetic Composition Indicator Based on Image Complexity

Chapter · October 2018

DOI: 10.4018/978-1-5225-7371-5.ch009

6 authors, including:

Adrián Carballal
University of A Coruña
47 PUBLICATIONS **155** CITATIONS
SEE PROFILE

Mª Luz Castro Pena
University of A Coruña
17 PUBLICATIONS **22** CITATIONS
SEE PROFILE

Carlos Fernandez-Lozano
University of A Coruña
85 PUBLICATIONS **309** CITATIONS
SEE PROFILE

Nereida Rodriguez-Fernandez
University of A Coruña
9 PUBLICATIONS **3** CITATIONS
SEE PROFILE

Chapter 9
Aesthetic Composition Indicator Based on Image Complexity

Adrian Carballal
University of A Coruña, Spain

Luz Castro
University of A Coruña, Spain

Carlos Fernandez-Lozano
University of A Coruña, Spain

Nereida Rodríguez-Fernández
University of A Coruña, Spain

Juan Romero
University of A Coruña, Spain

Penousal Machado
University of Coimbra, Portugal

ABSTRACT

Several systems and indicators for multimedia devices have appeared in recent years, with the goal of helping the final user to achieve better results. Said indicators aim at facilitating beginner and intermediate photographers in the creation of images or videos with more professional aesthetics. The chapter describes a series of metrics related to complexity which seem to be useful for the purpose of assessing the aesthetic composition of an image. All the presented metrics are fundamental parts of the prototype "ACIC" introduced here, which allows an assessment of the aesthetics in the composition of the various frames integrating a video.

DOI: 10.4018/978-1-5225-7371-5.ch009

INTRODUCTION

From image brightness indicators to facial recognition systems, multimedia devices in the home and commercial environments have gone through a revolution from the late 90s until the present day. These systems allow access to the images intrinsic information based on different phenomena, such as contrast, for instance, showing whether there is an under or over exposure at a given time.

Most of these indicators have the task of measuring objective phenomena, given that they can be clearly identifiable and quantified. Any system, which could be capable of measuring a relevant subjective phenomenon related to taking a picture or shooting a video would possess a high added value.

This paper proposes a system allowing the evaluation of the aesthetic composition of an image or the frames in a video (Liu, Chen, Wolf, & Cohen-Or, 2010). Thus, multimedia devices could help the user to identify in real time those framings with a certain aesthetic value. This would enable users without artistic background to take pictures and shoot videos of better appearance and with a more professional look.

Numerous papers (Machado & Cardoso, 2002; Rigau, Freixas, & Sbert, 2008; Ross, Ralph, & Zong, 2006; Machado, Romero, & Manaris, 2007) have appeared in recent years evaluating different elements of the aesthetic value of images and different ways to estimate it. This chapter introduces different metrics based on those works, based on the complexity of an image, which have already proven useful in experiments related to the ordering and classification based on stylistic and aesthetic criteria (Romero, Machado, Carballal, & Osorio, 2011; Machado, Romero, Nadal, Santos, Correia, & Carballal, 2015).

First, we will make a study of those metrics and their usefulness for calculating the aesthetic composition of a landscape. An experiment of image binary classification according to their aesthetic composition will be described for this purpose. Later on, we will present the design of a prototype system indicating the aesthetic composition of the frames integrating a video: **A**esthetic **C**omposition **I**ndicator based-on **I**mage **C**omplexity (a.k.a **ACIC**). This system will be used for the purpose of differentiating professional and amateur videos. Similarly, an example of functioning will be provided based on a professional video and the comments made by an expert on the results achieved.

We understand that the resulting prototype can be used for several tasks related to aesthetic composition: identification, classification, categorization, etc.; both in real-time multimedia devices and in stand-alone applications.

Next, the present paper is structured as follows: (i) a short description of the state of the art in composition systems is included; (ii) the hypothesis of the authors about possible metrics for evaluating the aesthetic composition of an image is presented; (iii) the features to be used in the study are described; (iv) the results obtained in an experiment of image classification according to their aesthetic composition are shown; (v) the design and functioning of a prototype will be detailed by means of a real example; (vi) and, finally, the conclusions and the upcoming research lines for improving the already presented prototype will be explained.

STATE OF THE ART

Santella, Agrawala, DeCarlo, Salesin, & Cohen (2006) presented a system which records user's eye movements for a few seconds to identify important image content. The given approach is capable of generate crops of any size or aspect ratio. The main disadvantage is that the system incurs on requiring user input, so it can't be considered a fully-computational approach. Once the important area of an image

is detected, the crops are made considering three basics on photography: (i) include an entire subject and some context around, (ii) edges should pass through featureless areas whenever possible, (iii) the area of the subject matter should be maximized to increase clarity.

Santella et al. (2006) presented 50 images cropped using three different approaches: saliency-based (Suh, Ling, Bederson, & Jacobs, 2003), professional hand-crop, and gaze-crop to 8 different subjects. They obtained that their gaze-based approach was preferred to saliency-based cropping in 58.4% of trials and in 32.5% to professional cropping.

Liu, Chen, Wolf, & Cohen-Or (2010) have translated several basic composition guidelines into quantitative aesthetic scores, including the rule of thirds, diagonal, visual balance, and region size. Based on which, an automatic crop-and-retarget approach to producing a maximally-aesthetic version of the input image. Their approach searches for the optimal composition result in a 4D space, which contains all cropped windows with various widths and heights.

A dataset of 900 casual images arbitrarily collected from international websites in which skilled photographers rank photographs through them was employed to evaluate their score function. To evaluate the performance of their method generated a set of 30 triplets of images; the original image, one crop using Santella's method and one using theirs. These triplets were shown to 56 subjects, males and women, between 21 and 55 years old. In 44.1% of cases, the subjects preferred the cropped images provided by their approach. In addition, 81.8% were not able to distinguish whether the image was hand-cropped or computationally optimized.

Wang & Cohen (2007) propose an algorithm for composing foreground elements onto a new background by integrating matting and compositing into a single process. The system is able to compose more efficiently and with fewer artifacts compared with previous approaches. The matte is optimized in a sense that it will minimize the visual artifacts on the final composed image, although it may not be the true matte for the foreground. They determine the size and position that minimizes the difference between a small shell around the foreground and the new background, and then run the compositional matting. The developed algorithm not always gives satisfying compositions when the new background differs significantly from the original.

Zhang, Zhang, Sun, Feng, & Ma (2005) presented an auto-cropping model to obtain an optimal cropped image using the width and height of the original image, the conservative coefficient, the faces detected and the region of interest (ROI). The model consists of three sub models: (i) a composition sub model to describe how good the composition is, (ii) a conservative sub model to prevent the photograph from being cropped too aggressively and (iii) a penalty factor to prevent faces or ROIs being cut off. They used 100 pictures randomly selected from 600 home photographs. All the images were used into two studies. The first user study evaluated the auto cropping result in different aspect ratios. They obtained that the algorithm exhibits a satisfactory score on cropping. The second user study evaluated the improvement of the picture composition after cropping, in which observed the considering of the artistic rules leads to a good score of the improvement of the picture composition.

Suh, Ling, Bederson, & Jacobs (2003) proposed a set of fully automated image cropping techniques using a visual salience model based on low-level contrast measures. According to them, the more salient a portion of image, the more informative it is; and the visual search performance is increased as much recognizable the thumbnail is. They used their feature set on recognizing objects in small thumbnails (Recognition Task) and to measure how the thumbnail generation technique affects search performance (Visual Search Task). They ran an empirical study over 20 subjects, which were college or graduate

students at the University of Maryland, and 500 filler images. In both tasks, the proposed set was capable to provide thumbnails substantially more recognizable and easier to find in the context of visual search.

HYPHOTESIS

The already described works focus on the search for metrics which show the composition quality or cropping methods which enhance the visual and aesthetic quality of a given image. Most of them use metrics related to Rule of Thirds (RoT), Region of Interest (ROI), or Saliency individually. RoT is a photographic framing technique, which divides the scene into 9 equally sized parts by means of three vertical and horizontal equidistant lines. This technique is based on placing the heaviest elements at the intersection among these lines. On the other hand, the use of ROI determines those image areas grouping the elements which attract the greatest interest. The saliency allows the differentiation of a foreground object from the background and to classify it as an interesting point.

Our hypothesis entails that the quality of aesthetic composition may be related to the visual complexity of the composition itself, as well as to the complexity derived from each of the elements represented in the same image. We assume that, inside the images, there are elements which attract the observer's attention, and their complexity must be taken into account when determining the composition aesthetics.

We propose the joint use of metrics allowing the determination of the complexity of an image as a whole, as well as of all the elements integrating it and, particularly, those which are its focus of attention. The proposed metrics are listed next.

Complexity

Machado & Cardoso (1998) based on previous works (Arnheim, 1956), proposes JPEG and Fractal Compression methods to estimate the image complexity. Forsythe et al. (2011) found a correlation between compression error and complexity of the image.

The error involved in the JPEG compression method, which affects mainly to high frequencies, depends on the variability of the pixels in the image. From this point of view, more variability involves more randomness and therefore more complexity. The fractal method tends to compress an image by filtering the self-similarities within. In this case, more self-similarities imply less variability, and therefore less complexity. Hence we considered applying JPEG and Fractal Compression methods as image complexity estimatives (Romero, Machado, Carballal, & Santos, 2012).

Subject Salience

Saliency is the quality that stands out one or multiple important objects from those that surrounds it/them. Somehow, saliency facilitates to focus the perception of the viewer on the most pertinent item or items on a scene. The saliency algorithm chosen to implement was the subject saliency algorithm also known as subject region extraction (Luo & Tang, 2008). Based on the idea that the subject in a photograph would be clearer and the background would be blurred, the algorithm extracts the clear region of an image, which theoretically holds the subject. This algorithm uses images statistics to detect 2D blurred regions in an image, based on a modification of (Levin, 2006). Subject Salience will be used to detect the foreground item/s, which should get the focus of attention.

Sobel Filter

The Sobel filter calculates the gradient of the image intensity at each point, giving the direction of the greater variation from light to dark and the amount of variation in that same direction. This gives us an idea of the variation of brightness at each point, from smooth to sharp differences. With this filter it is estimated the presence of the light–dark transitions and how they are oriented. With these light–dark variations corresponding to the intense and well-defined boundaries between objects, it is possible to obtain edge detection.

The Sobel Filter will give a simple representation of all the elements standing on the image by identifying their silhouettes.

PRESENTED FEATURES

The proposed metrics have already been listed. This section will show the features used in the experiments which are related to each of those metrics.

Before entering into a detailed explanation of the features used, we must explain the way in which they will be obtained. Four auxiliary images are generated from every image. Three of those images are obtained by separating the color channels following the HSV model. The fourth image stems from an attempt to solve the existing problems of the HSV color model for the extreme values of the H and V channels. For instance, a totally black pixel ($V = 0$) can be represented with any value of S and H. This new image is determined by multiplying pixel by pixel the S and V channels within the range [0, 255]. It will be referred to from now on as CS or Colorfulness (Correia, Machado, Romero, & Carballal, 2013).

Control Features

We have chosen to use a set of basic features related to the statistical variability of the pixels integrating an image. Said features calculate: (i) the mean (ii) and the typical deviation of the pixels with regard to the adjacent pixels in each channel.

Since the Hue channel is circular, the mean and standard deviation are calculated based on the angle values of Hue and its norm. In addition, it is performed the multiplication of the Hue angle by the pixel intensity values of CS, and a new value of the norm is calculated using values from H and CS. Splitting the image in mentioned color channels and applying the metrics to each of the resulting images yields a total of 12 features per image, 7 related with Average and 5 related with Standard Deviation.

Complexity Features

As already explained in the “Hypothesis” section, the use of the JPEG and FRACTAL compression are used as estimates of an image complexity; while the Subject Saliency and the Sobel Filter are used for the identification of the main elements in the scene, as well as all the items appearing in it.

In the case of the compression methods, since they are both lossy compression schemes, there might be a compression error, i.e., the compressed image will not exactly match the original. In our case, three levels of detail for the JPEG and fractal compression metrics are considered: low, medium and high. For

each compression level the process is the same. The image is encoded in JPEG or FRACTAL format, and its complexity is estimated as following:

$$Complexity\left(\text{Image}\right) = RMSE\left(\text{Image}, CT(\left(\text{Image}\right)\right) \times \frac{File_{Size}\left(CT\left(\text{Image}\right)\right)}{File_{Size}\left(\text{Image}\right)}$$

where RMSE represents the root mean square error and CT is the JPEG or fractal compression transformation.

$$Complexity\left(I\right) = \varepsilon\left(I, I_{\gamma}\right) \times \frac{\theta\left(I_{\gamma}\right)}{\theta\left(I\right)}$$

The quality settings of the JPEG encoding for low, medium, and high level of detail were 20, 40, and 60, respectively. Nonce, a quadtree fractal image compression scheme was used to calculate de PC of the image. More info available at (Romero et al., 2012).

Splitting the image in mentioned color channels and applying the complexity metrics to each of the resulting images gives a total of 32 features.

It must be noted that these 32 features will be calculated based on the original image, having applied the subject saliency and the Sobel Filter again. Therefore, we will achieve a total number of 96 features.

EXPERIMENTS ON AESTHETIC COMPOSITION

The previous section has identified all the features to be used in the experiment shown next. A total of 1961 landscape images of high aesthetic quality in their composition have been compiled for carrying out this experiment, most of them wallpapers in landscape format. All of them have a resolution higher than 1024x1024 pixels. Their visual topics vary a great deal: night, day, mountain, beach, etc. From this initial dataset, a random algorithm was created which will provide sub-images with a width/height ratio equal to the original image (see functioning in Algorithm 1). Said algorithm has been used on every image, thus providing a second set of images of the same sampling size.

A photography expert has identified those images which, because of the random cropping, generated a new image which was better than the original one as regards framing. All these images have been discarded, achieving a final dataset integrated by two sets of 1757 images each.

Figure 1 shows a simple subset of images of both sets. The left side shows images of the original set, while the right side shows the same subset once the algorithm has been applied.

Algorithm 1. Random Image Cropping

Per each image

1. A random height is established (between 400 and 1/2 of the height of the original image).
2. A width is established according to the ratio of the original image
3. Random loc_x and loc_y are created (>0, <original size)
4. If the cropped image does not exceed the original on the right or at the bottom:

cropping

otherwise

return to 1

Figure 1. Images of both sets (images of the original set on the left and the cropped version on the right) (© 2018, A. Carballal. Image may be subject to copyright.)

RESULTS

Both images to be used in this experimental part and the features which will characterize them individually have been presented so far. The present section explains the experiment carried out in order to try to validate the initially proposed hypothesis.

As already explained in the Control Features section, we have a set of 12 basic features related to the statistic variability of the pixels integrating the image (Avg and STD) calculated on the different color channels. This set will be referred to as BASE from now on.

Besides, we have a second set which will be integrated by those 12 features and by another 96 features related to the complexity of the whole image, to the main attention element and to the boundaries of the items integrating that image. This set of 108 features will be referred to as COMPLEX from now on.

The classification model chosen has been the SNNS (Stuttgart Neural Network Simulator). In particular, a backpropagation MLP is used with 3-layer architecture: an input layer with 108 neurons, a single hidden layer of 15 and an output layer with 1 neuron. This configuration has been established based on previous experiments and experiences of the research team in tasks of the same field (Machado, Romero, Nadal, Santos, Correia, & Carballal, 2015).

The network training will finish when a maximum number of 1000 cycles is reached. The initial network weights are determined at random within the range [-0.1, 0.1]. A maximum error tolerance of 0.3 has been used.

The 10-fold Cross-Validation (10-fold CV) model has been used for the generation of the training data sets so that their results are statistically relevant. Each of these runs has a different training and validation set which have been randomly generated. The results shown correspond to the average results obtained in these 10 runs.

Given that the neural network provides a number value within the range of 0 and 1, a dichotomic system has been used for cataloguing the images. Those images which have a network output of less than 0.5, once they have been presented to the system, will be catalogued as having a low aesthetic composition.

According to the data, it may be seen that the image classification when using the BASE set seems to achieve relatively satisfactory results. It should be noted that the problem itself contributes to the achievement of such high results. Let's imagine that there is a landscape photography similar to the one in Figure 2A. Having applied the random cropping, the new image may result as the one seen in Figure 2C. In this case, as is usually the case in this kind of images, the cropped landscapes usually have an extreme pixel variability compared to the original image. That is, the mean and the typical deviation of the pixels integrating the resulting image either increases or decreases considerably. Anyhow, no element of the real content of any of the two images is taken into account for the classification. Similarly, it may also be seen that both the accuracy and the recall for that feature set are clearly outbalanced (particularly in the case of the recall, with a 14% difference).

Table 1. Precision and recall using ANNs

	Precision			Recall		
	ORIGINAL	**CROPPED**	**GLOBAL**	**ORIGINAL**	**CROPPED**	**GLOBAL**
BASE	71.9%	79.2%	74.5%	82.2%	67.9%	75.0%
COMPLEX	82,6%	85.8%	84.2%	86.5%	81.7%	84.1%

Figure 2. Cropping example (© 2018, A. Carballal. Image may be subject to copyright)

A) Original Image B) Cropping selection C) Resulting image

As regards the second metrics set, we may observe an increase in accuracy over 9%. It must be noted that a great part of the improvement corresponds to the increase in the capacity to detect cropped images (from 67.9% to 81.7%). Similarly, both the individual recall and accuracy seem to be better offset with the global one.

DISCUSSION

The expert was presented with the number data and the images which the BASE and COMPLEX sets were not capable of identifying correctly, without having any kind of information about the classification system or the metrics used.

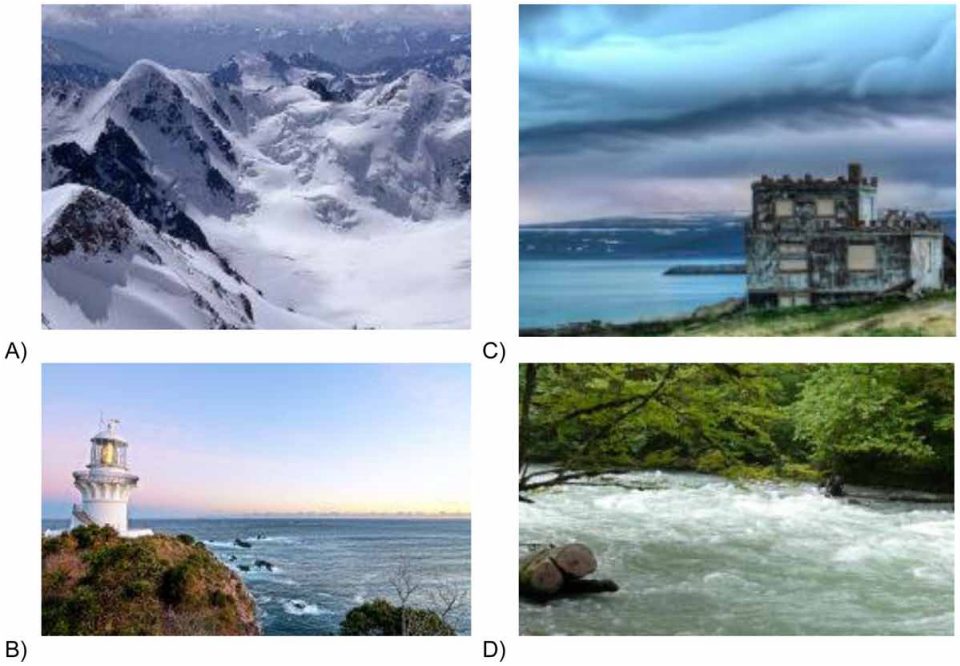
According to his criterion, the BASE features tend to classify non-cropped images incorrectly when there are minimal hue or texture differences among the composition elements (Figure 3A). Even in those images where there is some element of brightness, light or where the differentiating element is relatively small with regard to the image (Figure 3B). On the contrary, in the case of cropped images, it tends to classify erroneously those images whose content bears a great symmetry, regardless of their content or originality (Figure 3C), or those where the differentiating element is in the foreground, while the background is out of focus and homogeneous (Figure 3D).

With regard to the COMPLEX features, one of the most frequent mistakes happens in images with framing based on the horizon line (Figure 4A). It seems that the system cannot find any differentiating element in the image, understanding that it is made of two similar parts. In that case, where we are almost faced with two textures, it is possible that it interprets it as a cropped image. It is also classified as a cropped image when the differentiating element is placed on one of the far ends of the image, partially complying with the three thirds framing, but leaving the other half practically empty (Figure 4B).

On the contrary, in the case of those cropped images which are classified as original ones, sometimes the mistake is perfectly justified: cropped image pieces partially comply with the principles of framing; they structure a differentiating element at the centre while their environment goes totally unnoticed as a uniform background (Figure 4C). Color contrasts may also cause confusion; for instance, if we are faced with a horizon or, simply, with areas of a well differentiated color (such as water foam, a specific color in a bouquet of flowers or tree leaves) which is an element in itself. The fact that these occupy a significantly bigger part than the true differentiating elements may lead to error (Figure 4D).

Figure 3. Examples of wrongly classified images using the BASE set (Carballal, 2018)

Figure 4. Examples of wrongly classified images using the COMPLEX set (Carballal, 2018)



The expert concluded that the mistakes generated by the COMPLEX set were not trivial and were sometimes understandable.

ACIC PROTOTYPE

In the previous experiment, we have seen the capacity of the COMPLEX set for classifying images according to their aesthetic composition. A prototype has been developed from that set of metrics with the purpose of determining the composition aesthetics on the images integrating a video. The present section will explain in further detail the functioning of the ACIC.

Design and Implementation

ACIC embeds a light indicator on a video, identifying those frames with a high aesthetic composition (green light) or those with a low aesthetic composition (red light). The prototype is based on the use of different modules, which allow the performance of specific independent tasks in an automated way. The Algorithm 2 and Figure 5 show how the identification process for video frames is carried out.

The breaking down and remaking of the video in the corresponding frames is made under the support of the FFMpeg API. The frames modifications for adding the indicators are made by using, in this instance, the API of ImageMagick. The features extraction is made by means of an automated serial process, following the same extraction method seen in (Romero et al., 2012), but changing the edge detection filter used by the Sobel filter and adding the Saliency Subject.

In our case, since it is a prototype, only 1 out of 10 frames are used, due to the fact that the extraction system is not optimized yet for using parallel processing on every frame.

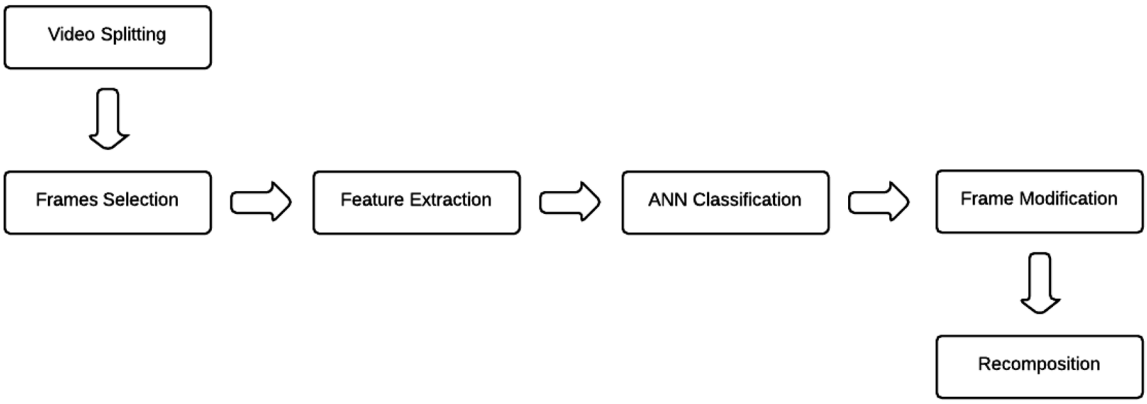
Testing

A series of tests have been made with videos containing different types of landscapes. In particular, the expert was asked to search for 3 videos considered as having a high aesthetic composition (well framed compositions, stable camera movements and professional preparation), and 3 more with a low aesthetic

Algorithm 2. ACIC Workflow

```
1. The video is broken down into its key frames with a frame rate of 25fps.
2. The first in a group of 10 frames is selected as the representative frame
in the set.
a. Its 108 corresponding features are extracted.
b. These values are presented as inputs to the ANN classifier.
c. If the classifier output value is >=0.5, then the green light is added to
the 10 frames, otherwise, the red light is added.
3. The video is remade from the frames resulting from section 2.
```

Figure 5. ACIC Workflow (Carballal, 2018, used with permission.)



composition (wrong framings, totally out of focus, inconsistent movements, something considered as amateurish). YouTube was the platform chosen for the compilation of the study videos.

Given that these were videos compiled from a multimedia portal whose main feature is the great variety of themes and contents, some modifications had to be made on the videos before presenting them to the ACIC. They were downloaded in 360p format using the H.264 codec. Those initial or final frames containing any kind of subtitles were discarded, given that they could be treated as an integral part of the image, and, probably, the system would determine that the text is the main subject, thus biasing the results achieved by the framing classifier.

Having chosen those videos and applied the ACIC system, we studied the capacity of the system for differentiating high aesthetic composition videos qualified by the expert as professional ones from those with a low composition classified as amateur (indexing), as well as the classification quality of the frames inside a video (quality).

Indexing Performance

The purpose is to analyze if the system can detect a greater number of frames correctly placed in "professional" videos than in "amateur" ones. In order to test this hypothesis, the outputs of the classifier were tested for each of the frames obtained in the sampling process for the six compiled videos. In the case of the 3 videos catalogued as professional by the expert, 74.4% out of the 1576 sampled frames were classified as having a high aesthetic composition. On the contrary, in the case of the 3 videos catalogued as amateur, only 16.8% out of the 1958 sampled frames were classified as having a high aesthetic composition (see Table 2).

Quality Performance

As previously mentioned, the ACIC prototype was used with each one of the chosen videos, and the expert was asked to evaluate its functioning. This section will explain concisely the expert's conclusions for one of the professional videos which is available at http://youtu.be/yJGXlZHtuJY.

Table 2. Accuracy of frames marked as "Professional" for the six compiled videos

Video Examples	Frames Marked as "Professional"	Frames Marked as "Amateur"	Accuracy of Frames Marked as "Professional"
Professional 1	309	46	87.04%
Professional 2	710	221	76.26%
Professional 3	153	137	52.76%
Amateur 1	185	404	31.41%
Amateur 2	123	612	16.73%
Amateur 3	21	613	3.31%

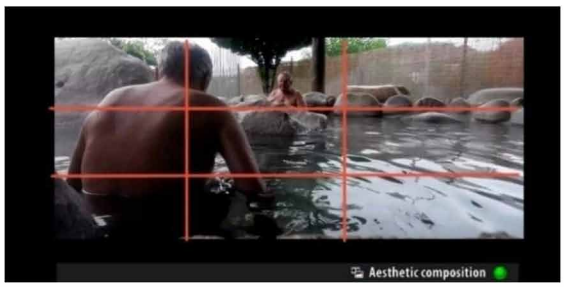
The system seems to interpret the currently most frequently used types of framings, such as the *Horizon's law*, *Vanishing Point* or even the *Rule of Thirds* (Figure 6A), differentiating and assigning a weight to each of the elements in the image. Even in those times when there is a clear intention to comply with the *Rule of Thirds* and this is not achieved, the system will label it as wrong (Figure 6B).

It also interprets *vanishing points* and *central point composition*, although they have a smaller size and scarce contrast against the background, thus positioning itself as a detailed and thorough system which is capable of recognizing elements with an objective visual weight (Figure 7).

At given times, the images possess a well-delimited and highly contrasted area with a very saturated color or an extreme brightness level with regard to the rest of the surrounding environment. In that case, the system seems to determine that the visual weight of the image is exactly located at that point, and so it is correctly classified as out of focus (see Figure 8).

The system is capable of recognizing the selective focus on the foreground. This is achieved by means of focusing on the object at the foreground while leaving the background out of focus. Although the object framing the composition may be placed in an area which is out of focus and scarcely contrasted, the system is capable of acknowledging the image as a correct one (at given times, the images possess a well delimited and highly contrasted area with a very saturated color or an extreme brightness level with regard to the rest of the surrounding environment. In that case, the system seems to determine that the visual weight of the image is exactly located at that point, and so it is correctly classified as out of focus, see Figures 8A and 8B).

Figure 6. Examples of well and badly framed images, according to the system, using the Rule of Thirds as criterion (Carballal, 2018)



A) 00:37

B) 00:10

Figure 7. Example of well framed image according to ACIC related to the Vanishing Point (Carballal, 2018)

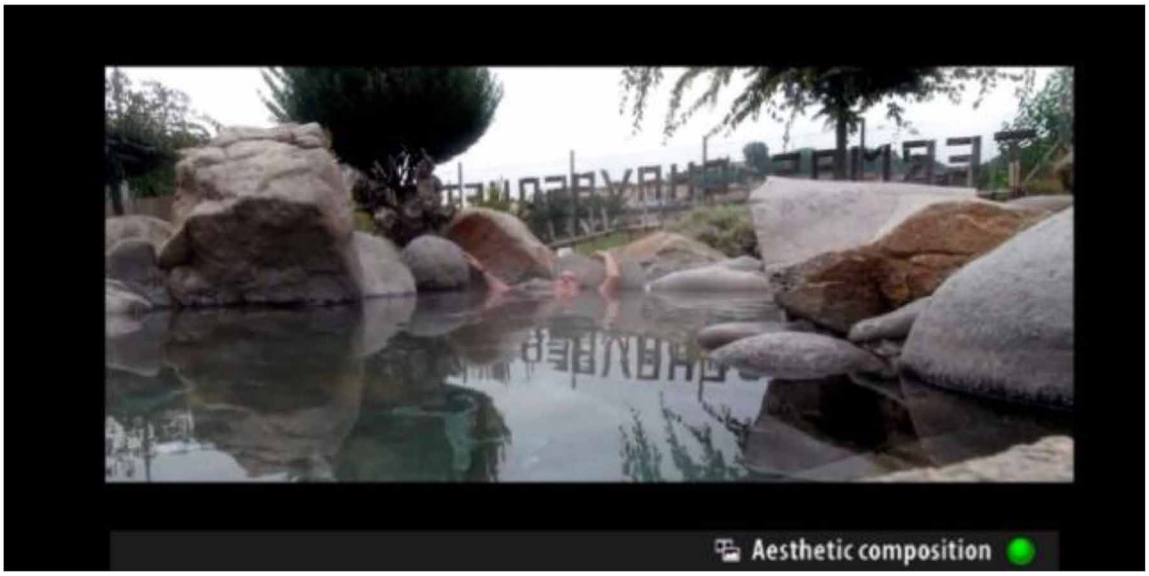


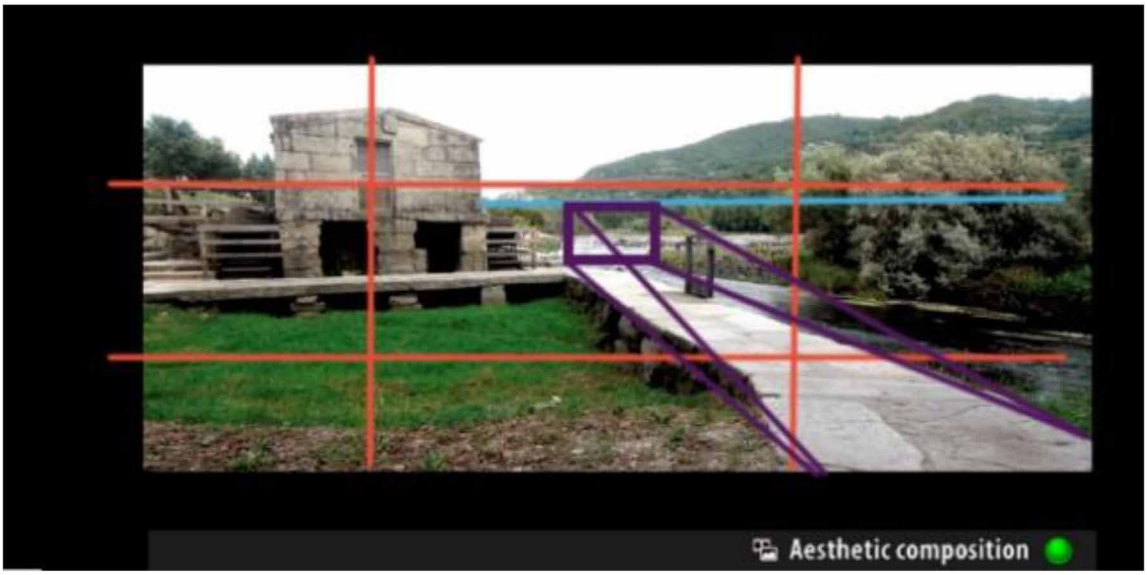
Figure 8. Examples where the system seems to work by determining the visual weight (A) and the selective focus on the foreground (B) (Carballal, 2018)

A) 00:26 B) 00:02

Sometimes we find overlaying framings in professional photography compositions. We may expect the system to fail in that case, since it would not be able to judge which of the contrasted elements has the highest weight, although this is not the case. It is capable of classifying an image with more than 3 different types of added framings as correct. For instance, Figure 9 shows a *Horizon's law* framing (blue), a *Rule of Thirds* one (red) and a *Vanishing Point* one (purple).

However, the system is not unerring. When a composition is framed based on a central element occupying an excessively big proportion, then the system makes some mistakes when classifying the framing. This happens when there are elements such as fog, sea foam, smoke, etc. which are overlaying the central element, and generating a contrast against the background. Thus, the system will classify it as a priority object and label it as out of focus (Figure 10A).

Figure 9. Example of image where several types of framings are observed: Horizon's law, Rule of thirds, and Vanishing point. (Carballal, 2018)



When the element with the visual weight in an image, for instance, the man in the photograph following the Rule of Thirds, is about to exit the framing and his position is doubtful, perhaps there is a variation and the system will recognize it as framed most of the time and out of frame at the next second, almost without variation (Figure 10B).

Another problem consists of the fact that it may recognize vegetation as an element with a high visual weight in order to analyse framings. Vegetation is usually unnoticed by the human eye, while the system takes it as a reference for classification. This is the case in Figure 11A, where the system determines the bush branches as a main element and thus, the image as well framed.

Besides, excessively cropped foreground elements are often classified as correct even if they are not, such as in the following case, where the carpet is established as a foreground element (Figure 11B).

CONCLUSION

This paper has presented a set of metrics based on complexity which seem to be useful for judging the aesthetic composition in landscape images as well as a prototype known as ACIC which would allow the final user of a multimedia device whether the image captured could be labelled as having a "high aesthetic composition".

A neural network has been used as a binary classification using the presented features as inputs, achieving accuracy and precision results of more than 84%. The trained network integrates the main axis of the ACIC, which shows by means of a green or red light if the aesthetic composition of the shown image is of high or low quality, respectively.

Figure 10. Examples of false negatives achieved by the ACIC (Carballal, 2018)

A) 00:05 B) 00:38

Figure 11. Examples of false positives achieved by the ACIC (Carballal, 2018)

A) 00:41 B) 01:25

ACIC has been tested on professional and amateur videos and seems to be capable of differentiating them based on the percentage of frames classified as well framed. Moreover, the individual classifications of the frames obtained in the simple videos, in spite of not being perfect, seem to achieve satisfactory results.

Among the most immediate enhancements, we may mention above all the elimination of all those cases identified by the expert where the classifier fails, both in the case of false positives and negatives. For this purpose, we intend to search for another set of metrics which can help the already existing one with that task, and even to find alternatives for the Sobel and Saliency Subjects, so that their detection problems do not have a direct impact on the prototype.

Another problem of use stems from the need to improve the classification times to be used in real time, so that the used images have a bigger sampling size. We intend to modify the classification system so that asynchronous tasks can be performed by means of parallel programming, thus reducing the time of the task of extracting metrics from each image, which currently entails the biggest bottleneck.

ACKNOWLEDGMENT

This work was supported in part by Xunta de Galicia, project XUGA–PGIDIT10TIC105008PR; the Portuguese Foundation for Science and Technology, project PTDC/EIA–EIA/115667/2009, the General Directorate of Culture, Education and University Management of Xunta de Galicia (Ref. GRC2014/049), and the Juan de la Cierva fellowship program by the Spanish Ministry 350 of Economy and Competitiveness (Carlos Fernandez-Lozano, Ref. FJCI-2015-26071).

REFERENCES

Arnheim, R. (1956). *Art and Visual Perception*. London: Faber and Faber.

Correia, J., Machado, P., Romero, J., & Carballal, A. (2013). *Feature Selection and Novelty in Computational Aesthetics. In Evolutionary And Biologicaly Inspired Music, Sound, Art* (pp. 133–144). Vienna, Austria: Springer.

Forsythe, A., Nadal, M., Sheehy, N., Cela-Conde, C., & Sawey, M. (2011). Predicting beauty: Fractal dimension and visual complexity in art. *British Journal of Psychology*, *102*(1), 49–70. doi:10.1348/000712610X498958 PMID:21241285

Levin, A. (2006). Blind motion deblurring using image statistics. In *Proceedings of the 19th International Conference on Neural Information Processing Systems* (pp. 841-848). MIT Press.

Liu, L., Chen, R., Wolf, L., & Cohen-Or, D. (2010). Optimizing Photo Composition. *Computer Graphic Forum*, 469-478.

Luo, Y., & Tang, X. (2008). Photo and Video Quality Evaluation: Focusing on the Subject. In *Proceedings of the 10th European Conference on Computer Vision: Part III* (pp. 386-399). Marseille, France: Springer-Verlang. 10.1007/978-3-540-88690-7_29

Machado, P., & Cardoso, A. (1998). Computing Aesthetics. In *Brazilian Symposium of Artificial Ingelligence* (pp. 219-228). Springer.

Machado, P., Romero, J., & Manaris, B. (2007). Experiments in Computational Aesthetics. In J. Romero & P. Machado (Eds.), The Art of Artificial Evolution (pp. 381-415). Springer.

Machado, P., Romero, J., Nadal, M., Santos, A., Correia, J., & Carballal, A. (2015). Computerized measures of visual complexity. *Acta Psychologica*, *160*, 43–57. doi:10.1016/j.actpsy.2015.06.005 PMID:26164647

Rigau, J., Freixas, M., & Sbert, M. (2008). Informational Aesthetics Measures. IEEE Computer Graphics and Applications, 28(2), 24-34.

Romero, J., Machado, P., Carballal, A., & Osorio, O. (2011). Aesthetic Classification and Sorting Based on Image Compression. In EvoApplications (pp. 394-403). Torino, Italy: Springer.

Romero, J., Machado, P., Carballal, A., & Santos, A. (2012). Using complexity estimates in aesthetic image classification. *Journal of Mathematics and the Arts*, *6*(2-3), 125–136. doi:10.1080/17513472.2012.679514

Ross, B. J., Ralph, W., & Zong, H. (2006). Evolutionary Image Systhesis Using a Model of Aesthetics. In *Proceedings of the IEEE Congress on Evolutionary Computation* (pp. 3832-3839). Vancouver: IEEE Press.

Santella, A., Agrawala, M., DeCarlo, D., Salesin, D., & Cohen, M. (2006). Gaze-based interaction for semi-automatic photo cropping. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 771-780). Montreal, Canada: ACM.

Suh, B., Ling, H., Bederson, B. B., & Jacobs, D. W. (2003). Automatic thumbnail cropping and its effectiveness. In *Proceedings of the 16th annual ACM symposion on User interface software and technology* (pp. 95-104). Vancouver, Canada: ACM.

Wang, J., & Cohen, M. F. (2007). Simultaneous Matting and Composition. In *IEEE Conference on Computer Vision and Pattern Recognition*. Minneapolis, MN: IEEE Press.

Zhang, M., Zhang, L., Sun, Y., Feng, L., & Ma, W. (2005). Auto cropping for digital photographs. In *IEEE International Conference on Multimedia and Expo*. Amsterdam: IEEE Press.

KEY TERMS AND DEFINITIONS

Aesthetic Criteria: Standards upon which judgements are made about the artistic merit of a work of art.

Dichotomy: A division of the members of a population, or sample, into two groups.

Lossy Compression: The decompressed data will not be identical to the original uncompressed data.

Subjective Phenomenon: As distinguished from "objective," is a classification for mental phenomena that are not capable of objective validation, as in the case of physical phenomena.

Visual Artifacts: Are anomalies apparent during visual representation as in photography.

Visual Complexity: The level of detail or intricacy contained within an image.