Automatic Single Person Composition Analysis

Claudio Cavalcanti, Herman Gomes and Odilon Lima
Departamento de Sistemas e Computação
Universidade Federal de Campina Grande
Campina Grande, PB, Brazil
{csvcc,hmg,oflj}@dsc.ufcg.edu.br

João Carvalho and Luciana Veloso
Departamento de Engenharia Elétrica
Universidade Federal de Campina Grande
Campina Grande, PB, Brazil
{carvalho,veloso}@dee.ufcg.edu.br

ABSTRACT

This paper presents an approach to image analysis based on three photographic composition rules: Rule-of-Thirds, Zoom Rule and Integrity Rule. These rules are commonly used by experienced photographers as an important step for creating attractive photos. The proposed approach assumes there is only one person in the image and considers the use of a face detector to locate the photograph's main subject. The composition analysis computes a set of numerical measures from an input image. These measures are then combined to produce an estimate for the overall composition quality. Experiments involving a subjective evaluation by human observers have demonstrated promising results, given that some correlation has been observed between labelling by expert users and the proposed automatic analysis in up to 85% of a test set of images.

Keywords: Photographic Composition Rules, Image Analysis, Image Classification.

1 INTRODUCTION

Photography is one of the most known and accessible forms of art [Hed03a]. One consequence of the digital era is the drastic increase of the number of amateur photographers, due to the ever decreasing costs of digital equipment, storage, publishing and home printing. However, this phenomenon naturally leads to a need of selecting the better photographs within larger sets.

Automatically identifying attractive or appealing photographs in large image sets is a task that, according to our literature review, is not extensively researched. This is explained by the fact that it is not easy to objectively detect elements that allow distinguishing between attractive and non-attractive photographs. Thus, the goal of this work is defined towards the implementation of rules to assess the quality of a photograph based on composition.

Although basic photographic techniques and principles have not changed much over the years, knowledge of these techniques is not widespread. Photographic composition rules can be seen as heuristics used by experienced photographers to emphasize the subject of a photograph [Gri90a]. The subject is "what" or "who" the photographer wants to show to the viewer. Previous work show that, when evaluating the quality or appeal of a photograph, people usually consider composition as the most determinant feature [Sav00a]. Theoretical support has been provided for some composition rules. The best known is the Gestalt's Theory [Kof55a], that explains how the human brain interprets the visual patterns perceived by the eyes. Those patterns influence perception and some may be used (in the case of Photography) to draw the viewer's attention to the subject.

WSCG 2010 conference proceedings,
WSCG'2010, February 1 – February 4, 2010
Plzen, Czech Republic.

The goal of this work is to analyze images in order to assess their overall quality in an objective way, according to a set of pre-defined photographic composition rules. Three important rules are considered: Rule-of-Thirds, Zoom and Integrity. Assuming that we are dealing with images of people only, the relevant information needed can be derived from the head and eyes positions, located using existing detectors. Moreover, within the scope of this work, we only consider images with a single person as subject. Extensions of the proposed approach to deal with group photographs is left as future work.

A heterogeneous set of images was crawled from the World Wide Web to be used in the experiments, so that the rules could be tested with a large variety of faces, poses, backgrounds and photographers. Those images were labeled by volunteers as either attractive or non-attractive according to their overall feeling about the photograph. This set has been used in one of the two experimental evaluations presented in the paper. A second evaluation involved analyzing image collections belonging to specific users, experts in photography, and obtaining their opinion about the automatic classification results.

Results show that a global measure based on photographic composition and obtained using the proposed approach is aligned with human labeling of attractive photographs in up to 85% of the images used for experimentation, selected from the user's own set of images.

Next we describe some assumptions adopted in this work. First, when we refer to the quality of a photograph, this is not related to its low-level features such as compression artifacts, resolution, among others. Moreover, the images used in the experiments have been chosen so that they did not exhibit such low level quality problems. The scope of this work is centred on the analysis of the visual aspect of a photograph. The term attractiveness is related to the way one may find a photograph more appealing than others. Finally, the quality of the composition is related to the conformity to photographic composition rules, not necessarily to the more abstract concept of an attractive photograph.

This paper is divided in the following sections. Section 2 reviews some previous work related to automating photographic composition analysis. Section 3 describes the three rules that are part of the approach. Section 4 discusses some processes for obtaining image classification thresholds. Section 5 presents the experiments performed to validate the proposed approach and their corresponding results. Finally, Section 6 brings final considerations and proposals for future work.

2 RELATED WORK

Banerjee and Evans [Ban04a] have developed a system for automating some aspects of photography. Working on a static image, the system implements a transformation that shifts the subject to a place that agrees with the Rule-of-Thirds. They also considered embedding their method into the camera, and defined automatic methods for background blurring and merger mitigation [Ban07b].

The Robot Photographer by Byers et al. [Bye03a] is one interesting example of autonomous photography. This robot can detect subjects in a large indoor area, where there are people. The composition rules used are the Rule-of-Thirds, the empty space (requiring that faces must occupy at least a third of the image), the no-middle rule (not allowing a person at the image center) and the edges rule (subject should not cross the photograph edges). After performing composition, the photograph is taken and stored by the robot until it is transferred to a computer. This approach can be applied to standing people only. According to the authors, the acceptance rate of the photographs autonomously taken in an experiment was around 35% for a set of human viewers.

Zhang et al. [Zha05a] present an approach for *auto cropping*, with the main goal of correcting failures detected in an image. The system works with three evaluation models: the composition model, the penalty model and the conservative model. The subject is located using face detectors. It proposes and implements 14 models of composition, to which each image is compared. A voting scheme was used to analyze the final corrected images, which gives the participants 3 options: good, acceptable and bad. A set of 100 images is considered in the work. Results show that the rate of images considered better than the original is 41%. The work proposed by Zhang et al. differs from our approach in the sense that they aim to enhance existing photographs by means of image cropping whereas our approach is designed to perform image analysis or classification. Moreover, their image evaluation was performed differently in the sense it gave three options to human viewers, what might have polarized the votes to the middle option.

A recent effort by Santella et al. [San06a] also emphasizes the idea of image cropping, using the viewer gaze movement to decide what are the image main points of interest, with no need to maintain image ratio. This approach is compared with a method that uses the points obtained using visual attention mechanisms. According to a voting experiment, the acceptance rate is 58% greater when compared to the traditional method of visual attention.

Datta et al. [Dat06a] present a computational approach to analyze the aesthetics of an image, using 15 metrics, from 56 initially considered. When using SVM (Support Vector Machines) [Vap99a] to classify an image by those metrics, a rate of 70.12% of agreement with a subjective evaluation is reached. The approach of Datta et al. [Dat06a] is similar to the one proposed in this paper, since, in both cases, metrics are used to evaluate an image. However, Datta's approach applies the composition rules by analyzing image pixels, whereas our approach uses higher-level information (face position), which is more effective to images containing people.

Ke et al. [Ke06a] propose a two-class photo classification. Photographs are classified as professional and snapshot based on some semantic features, such as the spatial distribution of edges, color distribution, hue count, blur, contrast and brightness. The aim is classifying images according to its quality in low quality or high quality.

In another work, Song et al. [Son04a] propose a face appeal model based on the size and position of each face in the image. The face appeal was represented using a Gaussian Mixture Model, where each Gaussian in the mixture corresponds roughly to one type of photographic style. In the approach proposed by Song et al., the rules are extracted from a large dataset which corresponds to the most of photographer's style.

A negative aspect of the two latter approaches is that they consider the spatial distribution of the ground truth data. This might force the images look the same way and allow common mistakes being considered as a desired behavior. They may also require a meticulous construction of the data set for dealing with social and cultural issues.

3 COMPOSITION RULES

Three rules were used in this work to analyze composition of a photograph: Rule-of-Thirds, Zoom Rule and Integrity Rule. The first analyzes how close the center-of-interest, i.e. the subject of the photograph, is to any of the "points-of-thirds". The second analyzes how close to the subject the camera is. Finally, the last rule verifies if the subject's head is preserved in the photo.

Those rules are commonly used by most photographers, both amateur and professional. There are many other possible rules, but we have chosen the three which are considered the most relevant by the specialized photographic literature [Hur04a, Hed05b]. We believe that the addition of more rules into our system will progressively have a positive impact on the quality of the results, since each rule analyzes a photograph according to different aspects.

The three chosen rules require detection of the subject position. We used the Rowley-Baluja-Kanade's face detector [Row98a, Row98b] to obtain the face coordinates of the subject, from which the approximate body position can be determined (assuming the subject is not in any unusual pose).

To obtain a more precise evaluation of the Rule-of-Thirds, we also used an eye detector, as eyes are the center-of-interest of most portraits. The eyes coordinates are detected using the approach proposed by Leite et al. [Lei07a], which achieved an accuracy of 99.63% using the Caltech image database [Weba]. The employed eye detection method outperformed other well known approaches, such as the Machine Perception Toolbox [Fas05a] and the Rowley-Baluja-Kanade eye detector [Row98a].

3.1 Rule-Of-Thirds

The Rule-of-Thirds is a well known rule that is used to verify how close to the "lines-of-thirds" (or their intersection, the points-of-thirds) is to subject. The lines of thirds are two lines that divide the image in three equal parts, both horizontally and vertically. Thus, the y coordinates of the two horizontal lines of thirds are $\frac{w}{3}$ and $\frac{2\times w}{3}$, whereas the x coordinates of the two vertical lines of thirds are $\frac{h}{3}$ and $\frac{2\times h}{3}$, where w and h are the image width and height, respectively.

Photographers apply the Rule-of-Thirds to avoid centralization of the subject, which may result in a too static photograph. The term static photograph means that it does not possess features that stimulate the viewer to search the image for regions of interest, thus resulting in an non-attractive image.

In portraits, eyes are the most adopted center-of-interest for human subjects. It is accepted that the ideal positioning of the subject center-of-interest is on any of the points of thirds. Figure 1 shows a portrait where the eyes are correctly positioned, according with the Rule-of-Thirds.

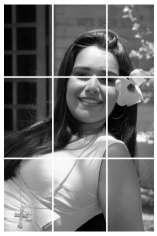

Figure 1: Rule-of-Thirds: One eye is on one of the four points of thirds, while the other is on a line-of-third.

In order to verify the Rule-of-Thirds, eight distances are calculated, forming the set $\Im$, which corresponds to the Euclidean distances of each subject eye to each of the four points-of-thirds. The smallest of those eight distances (s) is then selected as shown in Equation 1.

$$s = d \in \Im \wedge \nexists c \in \Im \mid c < d \tag{1}$$

The maximum distance (m) from any image point to the nearest point of thirds is given by Equation 2. This value corresponds to the distance from any image corner to its nearest point of thirds. In Equation 2, w is the width of the image and h is its height.

$$m = \frac{\sqrt{w^2 + h^2}}{3} \tag{2}$$

A normalized distance (called n) is defined in Equation 3, assuming values within the interval [-1..1].

$$n = 1 - (2 \times \frac{s}{m}) \tag{3}$$

If $n = 1$ in Equation 3, then a perfect match between the center-of-interest and one of the points-of-third is achieved, while $n = -1$ indicates a poor subject positioning.

3.2 Zoom Rule

The Zoom Rule analyses the relative face size within an image, with regards to the total image size. Once face position and dimensions are known, the face to image ratio can be computed.

We use Equation 4 to evaluate the parameter f which expresses the face height (k) to image height (h) ratio:

$$f = -1 + (2 \times \frac{k}{h}) \tag{4}$$

Although the face detector used in this approach returns equal values for face height and width, face height should be used for other face detectors since heads are taller than wider. Other approaches are also possible, like the ratio between face area and image area, as well as the face to image widths ratio.

The proportions between head height and head width can be obtained from anthropometric measures. Although not every measure is formally proved (since they only represent the proportions of part of the population), they are widely used by industrial designers, artists (such as painters) and anatomists to estimate the size of human body parts from known proportions. In this work, some anthropometric measures have been adopted, e.g. the ratio between the human head height and width is approximately 1.5. Once the eyes have been located, the distance between them allows obtaining the head width and, consequently, the head height.

The Zoom Rule, evaluated using Equation 4 should produce values within the interval -1 to 1. However, too much zoom may cause the face coordinates to exceed image borders, resulting in values greater than 1. In those cases the value of f is clipped to 1.

It is not easy to define adequate values of zoom. For many situations, a face close-up photograph is desirable and produces a result as good as a full body shot. Therefore, there is not, roughly speaking, an inadequate zoom value. In this work, we consider that if the subject is too far from the camera (consequently small in the photograph) he or she becomes part of the background. Excessive zoom may also be harmful to composition as it restricts the photograph to the subject face expression only. It also may show face imperfections, which is not desirable in most cases. Thus, the interval used -1 to 1 is not proportional to zoom quality, being just a measure.

3.3 Integrity Rule

The Integrity Rule is used to detect if body parts of the subject have been chopped out of the photographs. Although this chopping may be tolerated to some degree, it is usually desirable that parts like face, eyes and sometimes the arms and hands be preserved, otherwise the composition of the photograph cannot be considered good.

In the present work, we opted for analyzing the integrity of the head, which is essential in any photograph containing people. Head width (o) is evaluated by Equation 5, where e is the distance between the points that define the center-of-eyes.

$$o = 1.2 \times e \tag{5}$$

Similarly, head height (p) is defined by Equation 6:

$$p = 1.5 \times o \tag{6}$$

The concept of integrity has been expressed as the ratio between the chopped out head area (t) and the total head area (u). The parameter t is evaluated by Equation 9, using the results of Equations 7 and 8. Two types of integrity violation are considered, one that chops out the top of the head (expressed by q, Equation 7) and the other that chops out the side of the head (expressed by r, Equation 8). The coordinates (x_1, y_1) and (x_2, y_2), represent the upper left and lower right vertices of the head bounding rectangle, respectively. This evaluation is possible, since most of face detectors return coordinates out of the bounds of the image (i.e. negative values or coordinates greater than image width or height) when faces are chopped out.

$$q = \begin{cases} l \times |y_1|, & \\ \quad \text{if } y_1 < 0; & \\ o \times (y_2 - h), & \\ \quad \text{if } y_2 > h; & \\ 0, \quad \text{otherwise.} & \end{cases} \tag{7}$$

$$r = \begin{cases} n \times |x_1|, & \\ \quad \text{if } x_1 < 0; & \\ p \times (x_2 - w), & \\ \quad \text{if } x_2 > w; & \\ 0, \quad \text{otherwise.} & \end{cases} \tag{8}$$

$$t = q + r \tag{9}$$

u, the total area of the head, is defined by Equation 10.

$$u = o \times p \tag{10}$$

Finally, Equation 11 shows how integrity is evaluated. The face detector used in this work [Row98a, Row98b] does not define a maximum allowed occlusion level, however, experiments show that up to 50% of face occlusion can be tolerated in extreme situations. That is why we consider only 50% of full head area (u) in Equation 11, obtaining a value between -1 and 1 for integrity. It is possible, although not common, that the face detector correctly detects a face that is more than 50% occluded. Therefore, the maximum allowed integrity value (i) is clipped to 1, to avoid values outside the defined range.

$$i = 1 - (2 \times \frac{t}{0.5 \times u}) \tag{11}$$

Figure 2 illustrates a situation where both the top and one side of the head have been chopped out. In this case, integrity violation results from y_1 being smaller than 0 and x_2 greater than the width of the image. In the latter situation (and in other similar cases) the value of head width o must be added to the value of the chopped areas intersection t, which is subtracted twice at the equations defined above. For simplicity, this sum will be omitted.

4 THRESHOLD DETERMINATION

In our literature review, we could not find an objective method to evaluate the overall quality of photographs. This kind of assessment is usually done subjectively.

Since there are several different ways to acquire and analyze a photograph, the problem scope needs to be defined. In this work we assume photographs that have

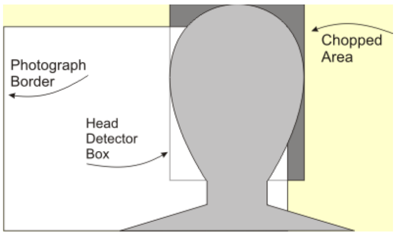

Figure 2: Integrity violation : top and side of the head chopped out.

one person only, whose face and eyes have been correctly detected by algorithms of face and eyes detection. The person is the photo's main subject. An incorrect detection happens when the face bounding box is greatly wider or narrower than the face or if any of the eyes is incorrectly detected. We did not consider paintings, drawings or post processed photographs in our experiments. Finally, only photographs with the subject at an upright position posing to the camera have been allowed in the image sets.

4.1 Building the Image Set

To the best of our knowledge, there is no publicly available image database labeled considering photographic composition rules. Therefore, in order to conduct our experiments an image set had to be acquired and labeled.

Firstly, images were downloaded from the World Wide Web, using a web crawler configured to download only photographs containing people. Several web sites of free photograph publishing were used as source of professional and consumer photographs, such as Flickr and DPChallenge.com. However, since part of the constructed image set is under copyright protection, it was not possible to provide a permanent link to it.

After downloading more than 2000 photographs, a face detector was used to filter those with only one face. Since the face detector sometimes fails (false detection, rejection or imprecision in location), the filtered images had to be manually examined and selected. The goal was to decouple the performance of the whole system from that of the face detector. Photographs which presented false detections or rejections were both discarded. The images that did not meet the scope requirements have also been discarded.

The decision to work with single person only has the benefit of considerably simplifying the problem, given the composition rules we are using are not well defined for groups in the photography literature. On the other hand, this imposes a restriction to the image set. We are currently developing additional rules for dealing with images containing more than one person.

In total, 417 images were used in the labeling process. Two photography experts participated in the labeling processes. A total of 65 out of 417 images were classified as attractive and the remaining 352 were considered non-attractive. The resulting image sets were carefully analyzed by a third expert.

4.2 Obtaining Thresholds

After the labeling process, two methods were used for obtaining classification thresholds: the first method is an automatic approach, whereas the second method needs user human interaction.

First Method This method uses a Neural Network to learn thresholds so that the test set images could be correctly classified as good or bad. A MLP Neural Network with the standard Back-Propagation algorithm was used to this goal.

A set of 126 images (half labeled as good and half labeled as bad) was used for the training The number of neurons in the hidden layer and the number of cycles has been systematically varied. For each set of parameters, the network has been trained 10 times (with different weight initializations) - the best network was then chosen from this set. The average correct classification rate for this best network was 95% with standard deviation of 4.11. This network had 64 hidden neurons, GL5 [1], 10,000 cycles as stop criteria and 0.001 as the learning step.

Second Method This approach consisted of classifying the photographs using the values obtained from the composition rules, according to all possible combinations of threshold values. All thresholds were varied from -1 to 1 in steps of 0.5. The step of 0.5 was chosen for minimizing the time needed for each user to evaluate an image set. In total, there are 250 possible combinations, when varying thresholds for all three composition rules and considering that the Zoom Rule requires two thresholds (lower and upper bounds). All 250 sets of images were analyzed, in order to choose the best one. The strategy chosen to ease this task was to analyze groups of ten sets of images at a time, choose the best group and finally select the best set from that group.

The resulting sets were analyzed to decide which of them contained the best images. The set chosen as the best indicates the thresholds that are most adequate to a specific user. The threshold values used were too restrictive in most cases, resulting in 191 sets with all images labeled as non-attractive, 12 groups where all images were classified as attractive, other 4 sets are extremly unbalanced. The first 207 sets are discarded resulting in 43 sets. Further analysis indicated the best image set and, consequently, the best set of thresholds. Some specific results are discussed in Section 5.2. In this setup, one set of thresholds was eventually determined as the best, but a possible approach is to select a

[1] Criterion to stop training when the objective function for a given set of weights produces error greater than 5% comparing to the best set of weights obtained so far.

small amount of acceptable sets from which the thresholds can be extracted. Table 1 shows the thresholds obtained for this method and the previous one. An image is labeled as good when obtained values for each rule are greater or equal the indicated threshold.

Method	Thirds	Zoom	Integrity
1 - Neural Network	0.70	-0.73	0.50
2 - Human Evaluation	0.00	-0.50	0.00

Table 1: Thresholds obtained.

4.3 Threshold analysis

The defined thresholds were analyzed by a group of volunteers to verify if the image set classified with those thresholds are representative. The first experiment presented 126 photographs to the participants without identifying which ones were classified as good composition aspect according to the defined threshold. Sixty people contributed to this experiment. The participants agreed with 60% of the image set labels. The labels were obtained by combining both methods using a logical AND.

This experiment considered an image set for which the contributors were not necessarily familiarized with. This fact might interfere with the capability of performing good judgment of some viewers. For instance, one may find a particular face, place or circumstance non-attractive and unconsciously vote the composition based on this feeling.

5 EXPERIMENTS AND RESULTS

The experiments described in this section were performed to analyze composition considering an image set known by the human participant. The goal is to validate both thresholds obtained in the previous section and evaluate the overall performance of the proposed approach.

5.1 Evaluating on Specific Image Sets

In order to evaluate the composition analysis in a different way and minimize the problem described in Section 4.3, some restrictions were imposed: (i) each contributor analyzed their own photographs, (ii) photographs are of a single event, (iii) photographs must contain exactly one person, and (iv) the face detector performed correctly.

By using only participants own photographs, it is possible to minimize the influence of subjective factors such as photogeny of unknown people, although a professional photographer is capable of well distinguishing among those (and many other) concepts. Another consequence of using experts photographs is the reduced number of non-attractive photographs. That allows our system to analyze, in a group of good photographs, which are the best ones.

By using photographs of a single event - a vacation, one weekend, or any well-defined period of time - photographs will look similar, thus minimizing the difficulties to analyze photographs from very different scenarios.

By using photographs containing only one person, a more acurate analysis is possible, once composition rules are not well-defined for group photography. In order to avoid an inaccurate analysis, images whose faces were not correctly detected by face detector detector did not performed correctly were also discarded.

Finally, by using only photographs without false face detections, the performance of the proposed analysis algorithm is decoupled from the face detection problem.

Although all those restrictions dramatically reduce the number of photographs to be analyzed, it better reflects the actual process of selecting photographs, where a photographer obtains dozens of photographs and only a few are analyzed and considered to be developed.

This experiment considered 55 images. Four observers, who declared themselves as specialists in photography, contributed with this experiment. The reduced number of participants is due to lack of experts in photography available for the experiment (since it is a time demanding experiment and those professionals are usually unable to stop their work for much time).

Photographs are then analyzed and automatically classified by our algorithm. After this analysis is performed, photographs are presented in a web form. A color scheme has been used to help users identify the automatic classification. In this scheme, images whose composition aspect was considered as good or bad are bounded by a green or a red rectangle respectively. It was not informed to the participants what was the automatic selection criteria. Images are also sorted with respect to the global measure. This scheme is illustrated in Figure 3.

Figure 3: Voting scheme : Images that user would develop should be in green (represented by dark grey) while photos which user has no intention of developing must be in red (represented in light grey).

Each contributor is then invited to change the status of the photograph in the presented form until it correctly reflects their opinion about the following question: "Which photographs should be developed and which should not". In our point of view, an image labeled as "to be developed" was considered attractive once it was selected among others, while the remaining are considered "less attractive" but not necessarily "non-attractive" nor "bad" images.

When a observer clicks in a photograph, the bounding rectangle switches the color between green and red, representing user's opinion about images attractiviness. There is no time requirement for the participant conclude their selection. At the end of the selection process, for each disagreement between automatic and manual classification (either good being classified as bad or bad being classificed as good), it was also asked to the participant the reason for their decision. The possible options were "Facial Expression", "Brightness, Color of image", "Photographic Composition" and "Beauty of the person".

The result analysis has been done in two steps. The first step considered the images labeled as attractive. We verified that the users agree with the decision of the algorithm in 85% of the images. This is a promising result. Moreover, 75% of the 15% disagreement set were rejected by other factors not related to composition. The second step analyzed the images labeled as non-attractive. The participants agreed with 75% of these images.

5.2 Results Discussion

Despite some misclassified images, the obtained results look very promising. In the following paragraphs we analyze some photographs processed by our approach to substantiate this conclusion. Four images are shown in Figure 4. They were all classified as a good, according to the composition rules. In addition, they were classified as attractive by their owners. In another example, shown in Figure 5, images were considered bad according to their composition values. However, they were considered attractive by their owners, resulting in a disagreement.

Table 2 shows the values obtained for each of the example images (in Figure 4) for all photographic composition rules. For those images, there was agreement between manual and automatic classifications. Table 3 shows the values of the photographic composition rules obtained for images (in Figure 5) which presented divergence between manual and automatic classifications. In our experiment, the image was considered good by the algorithm when values are lower than the thresholds of any of the methods.

It is possible to verify in Table 2, that in all images Rule-of-Thirds and Integrity Rule were obeyed in both methods as can be seen in Table 1. Although there was a great variation in zoom values, they are still above the thresholds shown in Table 1.

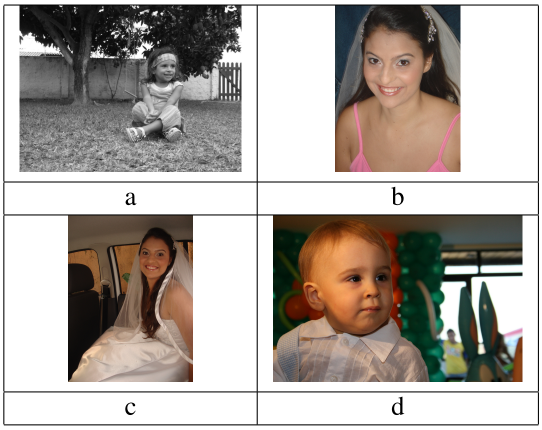

Figure 4: In these images all three composition rules were obeyed and they were classified as good according to the photographic composition aspect. They were also considered attractive by owners.

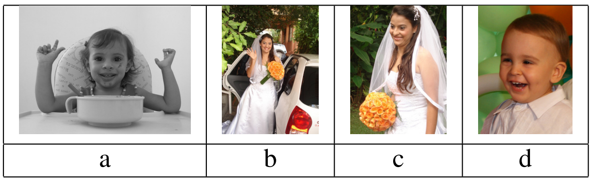

Figure 5: In these images, the analysis produced poor scores for the Rule-of-Thirds but good scores to other two rules, nevertheless, the photo was classified as bad according to the overall photographic composition aspect.

Image	Thirds	Zoom	Integrity
Figure 4a	0.86	-0.44	1.00
Figure 4b	0.90	0.00	1.00
Figure 4c	0.89	-0.42	1.00
Figure 4d	0.82	0.22	1.00

Table 2: The output values for each rule for images correctly classified. Images were considered good by both participants and automatic classification.

In the Table 3, the composition values were insuficient to result in an automatic labeling as good according to composition. However, the participants considered all images as attractive images. When asked them about their reasons, they reported factors different from composition. We consider this an expected result once an important composition rule is to allow "breaking the rules" in some occasions. If performed consciently, breaking the rules might improve the attractiveness of a photograph.

Although it was not possible to make the image set publicly available due to copyright reasons, we believe the experiments can be reproduced using other image sets and human subjects to label the images. Assuming these subjects have some knowledge in photography, the labeling process can be straightforward.

Image	Thirds	Zoom	Integrity
Figure 5a	-0.32	-0.32	1.00
Figure 5b	-0.19	-0.81	1.00
Figure 5c	-0.40	-0.55	1.00
Figure 5d	0.18	0.00	1.00

Table 3: The output values for each rule for images which there was divergence between automatic and manual labelings. Images were considered good by participantes whereas they were considered bad by automatic classification.

This work shows that there is a correlation between the proposed composition analysis and the observer's opinion with regards to the attractiveness of a photograph. That corroborates with the work developed by Savakis et al. [Sav00a] in which composition was considered by participants of a ground truth study the most important feature to be analyzed in a photograph. Our analysis shows that in a set containing mostly good photographs, the considered best was the one which the composition rules were correctly applied. It does not mean that a photograph with good composition aspect is attractive in a broad sense nor that images with bad composition aspect are unattractive, but if one has two photographs both attractive, the more attractive usually has a good composition aspect.

6 FINAL CONSIDERATIONS AND FUTURE WORK

After executing the described algorithms in 417 images, it is possible to state that composition rules play an important role in image quality. Compliance to those rules, based on the image database used, resulted in a significant number of photographs being correctly classified as attractive. It is also possible to say that not obeying to some or all rules does not necessarily result in non-attractive photographs. Moreover, in a final experiment, it was possible to verify that in up to 85% of the photographs the choice of the algorithm is aligned with the human preference. We believe this rate may be increased if other image analysis measures are included.

The next steps are to refine and extend the composition rules presented in this paper, e.g. define rules for more than one person, and add new rules to improve the quality of the analysis. It is also desirable to expand the image set and its labels. Other low-level image feature analysis, i.e. brightness, edges, etc., are also under investigation for producing a more precise classification.

ACKNOWLEDGEMENTS

This result was achieved in cooperation with Hewlett-Packard Brasil Ltda. using incentives of Brazilian Informatics Law (Law n^{o} 8.2.48 of 1991). This work was also supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq - project 141329/2009-2).

REFERENCES

[Bye03a] Byers, Z., Dixon, M., Goodier, K., Grimm, C.M. and Smart, W.D. An autonomous robot photographer. In Conf.proc. of IROS, Location, pp. 2636-2641, v. 3, 2003.

[Ban04a] Banerjee, S. and Evans, B. Unsupervised Automation of Photographic Composition Rules in Digital Still Cameras. In Conf.proc. of IS&T SPIE Sensors, Color, Cameras, and Systems for Digital Photography, pp. 364-373, v. 5301, 2004.

[Ban07b] Banerjee, S. and Evans, B. In-Camera Automation of Photographic Composition Rules. IEEE Trans. on Image Processing, pp. 1807-1820, n. 7, v. 16, 2007.

[Dat06a] Datta, R., Joshi, D., Li, J. and Wang, J. Studying Aesthetics in Photographic Images Using a Computational Approach. Conf.Proc. of the ECCV, pp. 288-301, v. 3953, 2006.

[Fas05a] Fasel, I., Fortenberry, B., and Movellan, J.R. A Generative Framework for Real Time Object Detection and Classification. Journal of Computer Vision and Image Understanding, pp. 182-210, v. 98, 2005.

[Gri90a] Grill, T. and Scanlon, T. Photographic Composition. Amphoto Books. Watson-Guptill Publications, New York, 1990.

[Hed03a] Hedgecoe, J. The new manual of photography. D.K., London, 2003.

[Hed05b] Hedgecoe, J. The Book of Photography. D.K., London, 2005.

[Hur04a] Hurter, B. The Portrait Photographer's Guide to Posing. Amherst Media, New York, 2004.

[Kof55a] Koffka, K. Principles of Gestalt Psychology. Routledge and Kegan Paul Ltd., London, 1955.

[Ke06a] Ke, Y., Tang, X. and Jing, F. The Design of High-Level Features for Photo Quality Assessment. Journal of CVPR, pp. 419-426, v. 1, 2006.

[Lei07a] Leite, B., Pereira, E., Gomes, H., Veloso, L., Santos, C. and Carvalho, J. A Learning-based Eye Detector Coupled with Eye Candidate Filtering and PCA Features. In Conf.Proc. of XX SIBGRAPI, pp. 187-194, 2007.

[Row98a] Rowley, H. A., Baluja, S. and Kanade, T. Neural Network-Based Face Detection. IEEE Trans. on Pattern Analysis and Machine Intelligence, pp. 23-28, v. 20, 1998.

[Row98b] Rowley, H. A., Baluja, S. and Kanade, T. Rotation Invariant Neural Network-Based Face Detection. Journal on Computer Vision and Pattern Recognition, pp. 38-44, v. 20, 1998.

[San06a] Santella, A., Agrawala, M., DeCarlo, D. Salesin, D. and Cohen, M. Gaze Based Interaction for Semi Automatic Photo Cropping. In Conf.proc. of the SIGCHI Human Factors in computing systems, pp. 771-780, v. 5301, 2006.

[Sav00a] Savakis, A. E., Etz, S. P. and Loui, A. Evaluation of image appeal in consumer photography. In Conf.proc. of SPIE Human Vision and Electronic Imaging V, pp. 111-120, v. 3959, 2000.

[Son04a] Song, B., Li, M., Li, Z., Zhang, H., Liu, Z. Face Appeal Model Based on Statistics. , pp. 344-351, v. 3331, 2004.

[Vap99a] Vapnik, V. The Nature of Statistical Learning Theory, 2nd Ed.. Springer, , 1999.

[Weba] Weber, M. Caltech Frontal Image Database. http://www.vision.caltech.edu/html-files/archive.html, 1999.

[Zha05a] Zhang, M., Zhang, L., Sun, Y., Feng, L. and Ma, W. Auto cropping for digital photographs. In Conf.proc. IEEE ICME, pp. 4-8, 2005.