Aesthetic Quality Assessment of Headshots

Matija Maleš, Adam Heđi, Mislav Grgić
University of Zagreb, Faculty of Electrical Engineering and Computing
Unska 3, 10000 Zagreb, Croatia
matija.males@gmail.com

***Abstract* - An automated system that can provide feedback about aesthetic value or quality of headshot photos based on learned rules could be a very useful support in photo searching, sorting and editing. This is a challenging problem as it requires semantic understanding of photos, which is beyond the state-of-the-art in computer vision. In this paper, we present a method built on most important rules or guidelines used by professional photographers to assess aesthetic quality of headshots. Proposed method uses low-level features and face-related high-level features. We make use of popular machine learning algorithms, support vector machines and Real AdaBoost, to determine whether a headshot is aesthetically appealing or unappealing. The results of extensive experiments indicate that proposed method is valid and effective: the overall classification accuracy for binary classification is greater than 86 %. This work is difficult to compare with previous attempts to assess aesthetic quality as no other research group studied this particular field of photography before.**

***Keywords* - Aesthetic Assessment; Photography; Headshot**

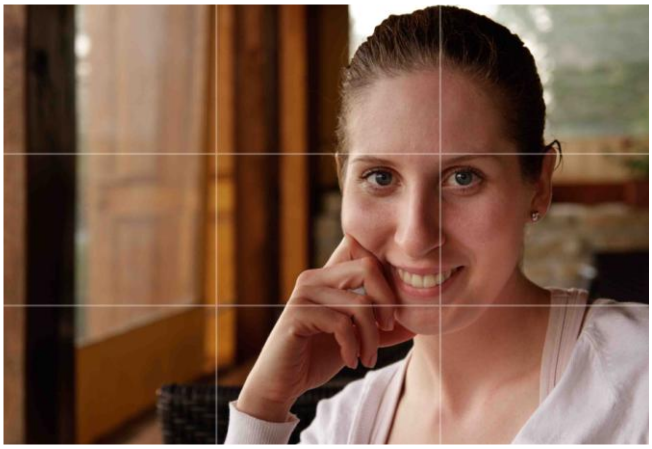

Figure 1. Example of rule of thirds being used effectively.

I. INTRODUCTION

With the current widespread use of smart phones and digital cameras anyone can take photos while the number of photos in private and public collections is increasing drastically [1]. An automated system that can provide feedback about aesthetic value or quality based on learned rules could be a very useful support in photo searching, sorting and editing. This is a challenging problem as it requires semantic understanding of photos, which is beyond the state-of-the-art in computer vision.

Headshot photography has become a popular choice and style for modern branding, such as websites, social media and even traditional business cards. A headshot in simplest terms represents a modern portrait for branding. In this paper, we briefly discuss most important rules or guidelines used by professional photographers to improve headshot aesthetic quality. Based on these rules we extract features for aesthetic quality assessment and build a classifier that can qualitatively distinguish between aesthetically appealing and aesthetically unappealing headshots. To the best of our knowledge, our work makes the first attempt to assess aesthetic quality of headshots, using minimum number of features.

This paper is organized as follows: in Section II a brief review of past work in the field of computational aesthetics is given; the basic idea of the method is described in Section III; the individual features used for aesthetic quality assessment of headshots are described in Section IV; in Section V we give a description of the experimental design and report results; finally, in Section VI we draw a conclusion and present future work.

II. RELATED WORK

Despite the challenges, various research attempts have been made and are increasingly being made to address basic understanding and solve various sub-problems under the umbrella of aesthetics, mood, and emotion inference in photos [2]. Some researchers use 'black box' approach, but this design does not provide any insight into the process of decision making itself [3, 4]. Others use a top-down, principled approach by using a number of low-level and high-level features [5, 6, 7]; such tools for automatic aesthetic assessment of photos are publicly available online [8]. The latest approach is characterized by focusing on the subject and using a set of salient features that characterize the subject and subject-background relationship, in addition to a set of low-level features [9, 10, 11]. Although there are promising results, in general, algorithmic solutions for detecting the subject and salient features remain inefficient. Therefore, recent work is being focused on more specific domains designed specifically for them, such as photographs of people and their faces.

Li et al. [12] evaluated aesthetic quality of consumer photos with faces. They conducted an online survey to collect people's opinions towards a set of consumer photos, resulting in a large database with human scores. Moreover, they extracted technical, perceptual and social relationship features to represent the aesthetic quality of a photo, by focusing on face-related regions. Defined as a multiclass categorization problem, their algorithm was able to achieve classification accuracy of 68% within one cross-category error.

Khan et al. [13] evaluated visual aesthetics in photographic portraiture. They developed a set of 7 classification features for spatial and color composition of human portraits. In their experiment they used only small part of the human photo database [12] of consumer photos with a single person. Their algorithm was able to achieve classification accuracy greater than 61% on the aforementioned database.

Battiato et al. [14] proposed aesthetic scoring function of photographic portraits and group photos, taking into account faces aspect ratio, composition, colors and face expressions. In their experiment they used their own database of 100 photos, subjectively rated by a group of ten people. Although they draw some conclusions, no classification accuracy was reported.

III. Features for Aesthetic Quality Assessment

Extracting features to assess the aesthetic quality of headshots is a crucial part of this work. With knowledge and experience in photography, we selected 10 features based on rules and methodology in photography and intuitive assumptions on human vision and psychology. Features described below are especially helpful when assessing aesthetic quality of headshots and proved efficient through various experiments.

A. Sharpness

When taking a photo, the area that we want to draw the viewer's eyes to should be in sharpest focus. For portrait/headshot it is fairly widely agreed that focus should be on frontal features of the face and the eyes. Contrary, if the foreground subject does not have sharp focus, it adds confusion to the photo, since we naturally see things up close sharper than far away objects. In our work, we examined sharpness of the detected face inside the face bounding box. Our sharpness measure draws on the concept proposed by Vu et al. [15] and is based on two factors: (1) a spectral measure based on the slope of the local magnitude spectrum and (2) a spatial measure based on local maximum total variation. The sharpness feature f_1 is defined as a single scalar value which denotes overall perceived sharpness.

B. Low depth of field

In optics, particularly as it relates to photography, depth of field (DoF) is the distance between the nearest and farthest object in a scene that appears acceptably sharp in a photo. High quality headshots stand out because the subject (especially his/her eyes) is in sharp focus while the background is blurred. A background that is in sharp focus gives the viewer too many details. We propose a novel method to detect shallow depth of field in headshots. Our DoF feature f_2 is computed as ratio of scalar values denoting perceived sharpness of object inside the face bounding box to sharpness of area outside of the face bounding box. As mentioned before, object of interest (a face in our case) should be in sharp focus while the background should be blurred. This essentially means that a large value of the DoF feature tends to indicate a good quality headshot. Conversely, small value of the DoF feature tends to indicate a low quality headshot.

C. Composition

In the visual arts, in particular painting and photography, composition is the placement or arrangement of visual elements in a work of art. Although good composition is highly subjective, there are number of established composition guidelines which can be applied in almost any situation to enhance the aesthetic appeal of a photo. The rule of thirds is the most common compositional guideline, commonly used by professional photographers today. It is one of the most frequently used ways of directing the viewer's eyes to the centre of interest in a photo. The idea is to imagine the frame split into nine equal sections or rectangles by two horizontal lines and two vertical lines. By placing important compositional elements (especially his/her eyes) near one of the two upper intersection points, photographer will lead viewer's eyes through the photo and create a more balanced composition. Besides, the photo will be more aesthetically pleasing. In our work three power points are defined: two upper intersection points and a center point (as illustrated in Figure 1). Features f_3, f_4 and f_5 are defined as distances between center of the face bounding box and three fixed power points, clearly determining the face position in a photo.

D. Contrast

Contrast is a measure of the difference in brightness between the brightest and darkest parts of a photo. The contrast feature f_6 is calculated using the root mean-square (RMS) contrast and does not depend on the spatial frequency content or the spatial distribution of contrast in the photo. RMS contrast is defined as standard deviation of the pixel values of V (value) component in HSV color space and has been proven useful when comparing contrast of different photos.

E. Lightness

As we already mentioned, headshot is not much more than a photo used to promote a particular person's general look. While headshot photos must be technically excellent, they must not be too "creative", by which we mean that the lighting is usually rather flat, with minimal shadows and highlights [16]. The lighting should support the person in the shot and not steal focus. The spark that drives the headshot should come from the model; presumably that's what will make a positive impact on other people. Our lightness feature f_7 is calculated inside the face bounding box as a mean pixel value of L (lightness) component in HSL color space.

F. Clipping and blown-out highlights

In digital photography, clipping is a result of capturing or processing a photo where the intensity in a certain area falls outside the minimum and maximum intensity which can be represented, due to a limited dynamic range. The clipped area of the photo will typically appear as a uniform area of the minimum or maximum brightness, losing any detail. The extent of the clipped area affects the degree to which the clipping is visually noticeable or undesirable in the resulting photo. Bright areas due to overexposure are sometimes called blown-out highlights. The clipped area will typically be completely white and this is something to avoid in every circumstance when taking headshots. Our blown-out highlights feature f_8 is calculated on the face bounding box as a percentage of total area of the grayscale photo having maximum intensity value pixels.

G. Hue count

When taking a headshot, it is always important to make the right decision about prevailing color scheme. It is a good idea to limit the number of hues to just a few or at least work them into a grouping of related colors. That way the subject will look more important than the surrounding environment. The layout will have clean uncluttered professional look. The hue count feature f_9 of a photo is calculated in HSV color space as follows. Only pixels with saturation values > 0.2 and brightness values in the range 0.15 - 0.95 were considered since outside this ranges the color aims to be white, gray or black to human eyes, no matter what the hue value is. A 20-bin histogram was computed on the pixels with good hue values. With a number of pixels in each bin known, hue count feature is calculated as the standard deviation of the pixel number in each bin.

H. Face size

Wikipedia defines a headshot as: "A headshot is a photographic technique where the focus of the photograph is a person's face". Accordingly, the viewer's eyes will be drawn to the face and their attention will be held there. Typically, headshots focus on the head and shoulders because people normally want to see your face as opposed to your body. The more of your body showing in the photo means your face gets smaller in the frame. If a face size is below a certain threshold, we are approaching to the torso and full body portrait. The face size feature f_{10} is calculated as a ratio of face bounding box area to the total photo area.

IV. EXPERIMENTS AND RESULTS

A. Database

To the best of our knowledge, there is no publicly available database with human ratings focusing on this specific type of portraits. Therefore we collected a set of 380 photos with a single face, using both photos from private collections and photos uploaded on Flickr.com photo sharing website under the Attribution – NonCommercial License. Our set of photos contains a positive set of 258 RGB photos that were marked as aesthetically appealing and a negative set of 122 RGB photos marked as aesthetically unappealing (as illustrated in Figure 2). The photos in our database are at different resolutions and aspect ratios (min. 200 x 200, max. 1000 x 667), acquired with both consumer and professional digital cameras. Classification of photos as being positive or negative was performed by a group of five people, including authors of this article. To test the classifiers on our photos, we separated photos in two subsets: the first one was the training set and the second one was the evaluation set. We chose to take at random 75 % of the photos as a training set, and the rest 25 % of the photos as an evaluation set.

Figure 2. Example of bad headshots (top row) and good headshots (bottom row) from our database.

B. Face detection

In our work, we made use of Viola-Jones face detection method [17], using the implementation included in the OpenCV [18], Intel's free open-source computer vision library. This method is quite effective at detecting faces in photos and outputs an accurate position and size of bounding box containing the detected face. Position and size of each bounding box are described with 4 values: first pair of values is the pair of coordinates (x, y) of the top-left corner, while the second pair of values is height and width of the face bounding box. Nevertheless, the method has some drawbacks in its application which can lead to possible multiple detections and miss detections. In case of multiple detections around the same face, the largest face bounding box was used. In case no face was detected (32 out of 380 photos), the face bounding box was specified manually.

C. Traditional feature set

To compare the performance of our feature set with previous work in terms of classification accuracy, 66 traditional features described in [13] were computed for each headshot.

D. Experiments and Results

In this paper we propose a method that uses low-level features and face-related high-level features to assess aesthetic quality of headshots. Low-level features are calculated over the entire photo (e.g. contrast and hue count), while high-level features are calculated inside the bounding box containing the detected face (e.g. sharpness and blown-out highlights). Consequently, constituent part of our method is face detection method by which the position and size of each face is determined (described in Section IV.B).

For every photo in our database (described in Section IV.A) 10 low-level and high-level features are calculated; using this features normalized to unit length, feature vector for each headshot is built. These new feature vectors of the training set are used to train classifier by using popular classification techniques. The idea is to deduce classification rules from the training set, in order to allow one to classify a new headshot, i.e. to assign it to one of the classes. There are only two possible classes of headshots: one class that is aesthetically appealing and the other class that is not. This kind of classification is also called binary classification.

We used Support Vector Machines (SVM), a group of supervised learning models first introduced by Cortes et al. [19] that can be applied for binary classification problem. In our work we used the sigma-tuned Radial Basis Function kernel model with $\gamma = 3.7$ and cost = 1.0. We also used Real AdaBoost, the generalization of a basic AdaBoost algorithm first introduced by Schapire et al. [20]. Using newly created classification rules we made prediction on the evaluation set and finally compared the success rates of different classifiers. Table 1 shows results obtained using different classification techniques. The final accuracy presented is an average over 100 repeated measurements. In general, SVM achieved higher accuracy than the Real AdaBoost classifier. The former method has achieved most significant result with classification accuracy of 86.33 % and can assess aesthetic quality in fast and efficient manner. On the other hand, similar classification accuracy was achieved with traditional feature set, roughly six times larger than our feature set, thus significantly increasing the total training and classification time.

TABLE 1 Results obtained using different classification techniques.

	Classification Accuracy
SVM	86.33 %
Real AdaBoost	84.66 %

V. CONCLUSION AND FUTURE WORK

In this paper, we proposed framework to assess the aesthetic quality of headshots, a specific type of portraits. This is a challenging problem as it requires semantic understanding of photos, which is beyond the state-of-the-art in computer vision. Our method is a rule-based method which makes use of popular classification techniques to determine whether a photo is aesthetically appealing or unappealing. Extensive experiments on our database demonstrated the effectiveness of the proposed method. Although our work suffers from the fact that there is no publicly available benchmark database and that true evaluation could not be performed, the performance of our method in terms of binary classification accuracy is comparable to the best performance to date. In the future, we plan to build a realistic headshot database with human ratings and make it publicly available; that would allow different research groups to compare their methods. Moreover, it is desirable to develop an improved face region extraction and eye detection algorithm. Finally, it may be beneficial to incorporate an algorithm that could calculate the color distortion due to camera white balance issues. We hope that our work will motivate other research groups in this new and practically important and challenging research direction.

REFERENCES

[1] http://royal.pingdom.com/2013/01/16/internet-2012-in-numbers/, Pingdom, Internet 2012 in numbers, 2013.

[2] D. Joshi, R. Datta, Q. Luong, E. Fedorovskaya, J. Z. Wang, J. Li and J. Luo, "Aesthetics and Emotions in Images: A Computational Perspective", IEEE Signal Processing Magazine, 28(5), 2011, pp.94-115

[3] H. Tong, M. Li, H.-J. Zhang, J. He and C. Zhang, "Classification of digital photos taken by photographers or home users", Proceedings of the 5th Pacific RIM Conference on Multimedia, Japan, 2004, pp.198-205

[4] L. Marchesotti, F. Perronnin, D. Larlus and G. Csurka, "Assessing the aesthetic quality of photographs using generic image descriptors", Proceedings of the ICCV 2011, Spain, 2011, pp.1784-1791

[5] R. Datta, D. Joshi, J. James and Z. Wang, "Studying Aesthetics in Photographic Images Using a Computational Approach", ECCV 2006, Graz, Austria, 2006, pp.288-301

[6] Y. Ke, X. Tang and F. Jing, "The Design of High-Level Features for Photo Quality Assessment", Proceedings of the IEEE CVPR 2006, New York, NY, 2006, pp.419–426

[7] L. Yao, P. Suryanarayan, M. Qiao, J. Z. Wang and J. Li, "OSCAR: On-Site Composition and Aesthetics Feedback through Exemplars for Photographers", International Journal of Computer Vision, 96(3), 2012, pp.353-383

[8] R. Datta and J. Z. Wang, "ACQUINE: Aesthetic Quality Inference Engine - Real-time Automatic Rating of Photo Aesthetics", Proceedings of the ACM ICMR, Pennsylvania, USA, 2010, pp.421-424

[9] Y. Luo and X. Tang, "Photo and Video Quality Evaluation: Focusing on the Subject", ECCV 2008, Marseille, 2008, pp.386-399

[10] C. Li, A. C. Loui and T. Chen, "Towards Aesthetics: A Photo Quality Assessment and Photo Selection System", ACM Multimedia Conference, 2010, pp.827-830

[11] S. Dhar, V. Ordonez and T. Berg, "High Level Describable Attributes for Predicting Aesthetics and Interestingness", IEEE CVPR 2011, Colorado Springs, USA, 2011, pp.1657 – 1664

[12] C. Li, A Gallagher, A. C. Loui and T. Chen, "Aesthetic quality assessment of consumer photos with faces", Proceedings of the IEEE ICIP 2010, Hong Kong, 2010, pp.3221-3224

[13] S. S. Khan and D. Vogel, "Evaluating visual aesthetics in photographic portraiture", Proceedings of the CAe, France, 2012, pp.55-62

[14] S. Battiato, M. Moltisanti, F. Ravi, A. R. Bruna and F. Naccari, "Aesthetic scoring of digital portraits for consumer appliacations", Proceedings of SPIE Digital Photography IX, Volume 8660, California, USA, 2013

[15] C. T. Vu, T. D. Phan and D. M. Chandler, "S3:A Spectral and Spatial Measure of Local Perceived Sharpness in Natural Images", IEEE Transaction on Image Processing, 21 (3), 2011, pp.934-945

[16] C. Grey, "Master lighting guide for portrait photographers", Buffalo, NY, Amherst Media, 2004

[17] P. Viola and M. Jones, "Robust real-time face detection", International journal of computer vision, 57(2), 2004, pp.137-154

[18] G. Bradski and A. Kaehler, "Learning OpenCV: Computer Vision with the OpenCV Library", Cambridge, MA, O'Reilly Media, Inc., 2008

[19] C. Cortes and V. Vapnik, "Support-vector networks", Machine Learning, 20(3), 1995, pp.273–297

[20] R. E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions", Machine Learning, 37(3), 1999, pp.297-336