A Novel Computational Tool for Aesthetic Scoring of Digital Photography

Fabrizio Ravì, Sebastiano Battiato; Image Processing Lab (IPLab) - http://iplab.dmi.unict.it; Department of Mathematics and Computer Science, University of Catania, Italy;

Abstract

To be able to score the aesthetic and emotional appealing of digital pictures through the usage of ad-hoc computational frameworks is now affordable. It is possible to combine low-level features and composition rule to extract semantic issues devoted to isolate the degree of emotional appealing of the involved subject. We propose to assess the aesthetic quality assessment on a general set of photos focusing on consumer photos with faces. Taking into account local spatial relation between involved faces and coupling such information with simple composition rule an effective aesthetic scoring is obtained. A further contribution of the proposed solution is the novel usage of the involved facial expressions and relative pose to derive additional insights to the overall procedure. Preliminary experiments and comparisons with recent solution in the field confirm the effectiveness of the proposed tool.

1. Introduction

Computational Aesthetics applied on digital photography is becoming an interesting issue in different frameworks (e.g., photo album summarization, imaging acquisition devices) as properly reviewed in [1]. One of the main challenge in the field, is the definition of computational methods able to score in a proper way both content and appearance of semantic objects detected in a picture. Various research attempts have been done mainly to address basic understanding and solve various issues related to aesthetics, mood, and emotion inference (in pictures). Of course, despite increasing number of techniques published in the field, it is important to highlight how in general the global rating is often greatly influenced by the taste and sophistication of the viewer. Here we are interested to find some formal or mathematical explanation of aesthetics in photographs although it is widely believed and can often be experimentally demonstrated that aesthetics is mainly subjective (e.g., the same photograph can be appreciated by some viewers but not by certain others. In consumer photos (e.g., nature, people, etc.) some criterion are well understood and usually coded both in terms of overall color appearance (e.g., tonality, lighting, …) and composition. Another issue is the possibility to build a formal regression system [2, 3] to predict a score just differentiating high from low quality photos. We claim that is possible to mimic in a computational framework some well-known rule-of-thumb extracting low-level features specifically designed to capture the perceptual properties that form the aesthetic (or emotional) value of a picture. Finally, we cite some interesting attempts to collect data and resources directly from related communities [1] such as:

- Flickr [18];
- Photo.Net: A Community of Photographers [19];
- DPChallenge - A Digital Photography contest ([20]);
- Terra Galleria Photography [21];
- ACQUINE - Aesthetic Quality Inference Engine - Free Instant Impersonal Assessment of Photo Aesthetics [22].

Our interests in the field of aesthetic evaluation of digital imaging is mainly devoted to design a sort of real-time filter to be embedded on smart cameras able to drive the user to capture/retain only high quality photos. Typical imaging pipelines implemented in single-sensor cameras are designed to find a trade-off between sub-optimal solutions (devoted to solve imaging acquisition) and technological problems (e.g. color balancing, thermal noise, etc.) in the context of limited hardware resources. State-of-the-art techniques to process multichannel pictures, obtained through peculiar processing of CFA images, include demosaicing, enhancement, denoising, compression and also ad hoc matrixing and color balancing techniques devoted to preprocess input data coming from the sensor. The overall image generation pipeline (IGP) is aimed to reconstruct the final image exploiting all the information acquired by sensor to achieve the 'best' possible image. Due to the increasing computational power of image acquisition devices [4, 16], that already have some semantic engines (e.g., scene classification, face and smile detection, etc.) such methods could assist users to acquire pleasant pictures. Current imaging pipeline already include some effective mechanism to classify input scene according to semantic contents [5, 6] and properly apply some kinds of enhancement [7, 8]. Among other the method in [6] exploit a holistic representation of the scene in the discrete cosine transform domain fully compatible with the JPEG format, performing a robust classification of the scene at superordinate level of description (e.g., natural versus artificial, indoor versus outdoor) with effective performances both in terms of overall accuracy and employed computational resources.

In the current proposal the final scoring is obtained just evaluating and combining together some aesthetic features that consider the presence of people and visual balancing issues (e.g., rule of thirds, visual balancing, etc.). For group photos we measure also the reciprocal distance and the size of the region enclosing each face. Finally, a refinement is introduced taking into account facial expression with respect to the main emotional status (Happiness, Sadness, etc.) and face appearance (e.g., eye closed, etc.). Although in [9] some preliminary attempts to include features computed by facial characteristics has been proposed here we propose to include high-level emotional status and corresponding poses [10, 11]. The method proposed in [10] is able to detect face in input images just employing a robust method with respect to illumination changes; its recent updates [11] returns also faces with different poses and a set of information about involved facial

expressions. On-going research is devoted to find a way to properly weight different facial expressions with respect the underlying context and/or viewer preferences [1]. The proposed method has been validated by comparing the final scoring with respect to [12] but also including some subjective evaluations making use of standard MOS (Mean Opinion Score) procedures. Preliminary experiments and comparisons with existing works confirm the effectiveness of the proposed tool. A proper demo have been also provided reporting the full integration of the system in a mobile platform enclosing also further consideration about preferred (or expected) color [7] with respect to the involving semantic scene (e.g., indoor/outdoor, natural/artificial, etc.).

The paper is structured as follows. Section 2 summarizes the main steps of the proposed algorithms. Next Section reports in details the experimental setting, presenting also some brief comparisons with existing approaches while future works are briefly sketched in the conclusions.

2. Proposed Framework

The aesthetic scoring is determined by a suitable arrangement of both composition techniques (e.g., visual balancing and the rule of thirds) as well as the facial expressions of the people present inside the image. A proper pre-processing step making use of some scene analysis have to determine if there is a single face or a group of people. Both cases are very common in consumer photos. On the basis of the number of detected faces, (obtained from a proper detector [10, 11] as detailed below), the proposed algorithm proceeds as follows.

If the input image $I = M*N$ is composed by only one face of size $a*b$ the evaluation criteria of the aesthetic score is based on:

1. Visual balancing which is based on the Euclidian distance between the underlying face, just referring to the center of its bounding box (x_i, y_i) with respect to the center of the entire photo ($round(M/2)$, $round(N/2)$):

$$visualBalance = \sqrt{(x_i - \frac{M}{2})^2 + (y_i - \frac{N}{2})^2} \quad (1)$$

2. The ratio between the two regions (face area with respect to the overall area):

$$faceRespectImg = \frac{M*N}{a*b} \quad (2)$$

The score is the following:

$$score = \frac{1}{visualBalance} + \frac{1}{faceRespectImg} \quad (3)$$

Whenever more than one face is detected the overall score have to be computed taking into account a series of aesthetic criterion useful to manage the reciprocal links between the involved subjects. As in [12, 15] each picture can be associated to a linked graph where the nodes correspond to the faces while the edges consider the ratio between areas and distances of neighboring faces. Differently than [12] we propose the following schema:

1. For each face $i = 1, \ldots, F$ (e.g., the number of faces) to compute a weight $w(i)$:

$$w(i) = \sum_j c(i, j) = \sum_j \frac{area(i)}{dist(i, j)} \quad (4)$$

where, $c(i, j)$ is the cost associated to the link/edge $i \rightarrow j$, $area(i)$ is the area of the i^{th} face, $dist(i, j)$ is the distance between face i and face j; in this way both scale and closeness between the various subjects inside the scene are considered. The underlying ratio is the following: group photo where people are placed in a chaotic way (different distance from the camera) should have a low aesthetic score. The computed weights above embed such information for each involved face.

2. Let W_max the maximum computed value. The faces whose weight $w(.)$ is less than 25% of the W_max are discarded because are considered not relevant for our purposes (i.e., to remove small faces and also those far away from the main subject);

3. On the remaining faces, we update such score just considering some well-known heuristics related to image composition as the rule of thirds [13]. The rule of thirds is an imaginary "tic-tac-toe" board depicted across an image to break it into nine equal squares. The four points where these lines intersect are denoted by G_j. The rule of thirds makes use of a natural tendency of the human eye to be more strongly drawn towards certain parts of an image.In our case first, the minimum distance between the center of each face S_i and the four points of strength G_j is computed:

$$scoreRdt(i) = \frac{D(S_i, G_j)}{2\partial_1} \quad (5)$$

where $\partial_1 = 0{,}17$ (as in [13]) and $D(S_i, G_j)$ is equal to

$$D(S_i, G_j) = \min_{j=1,2,3,4} d_M(C(S_i, G_j)) \quad (6)$$

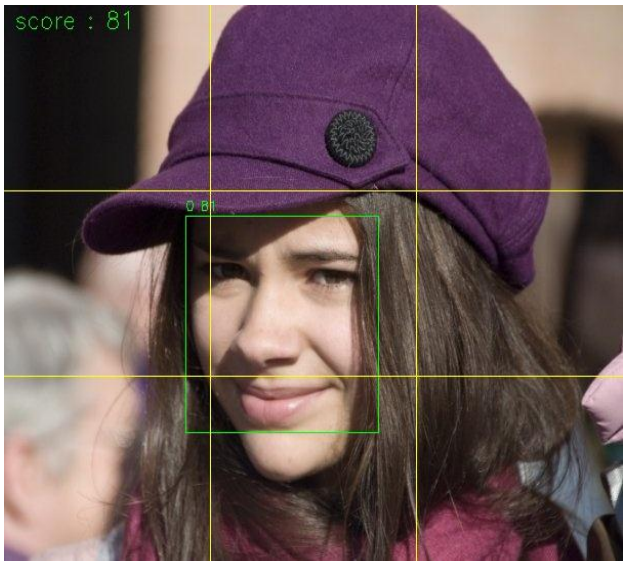
where d_M is the Manhattan distance.

4. The score for each face is then computed as:

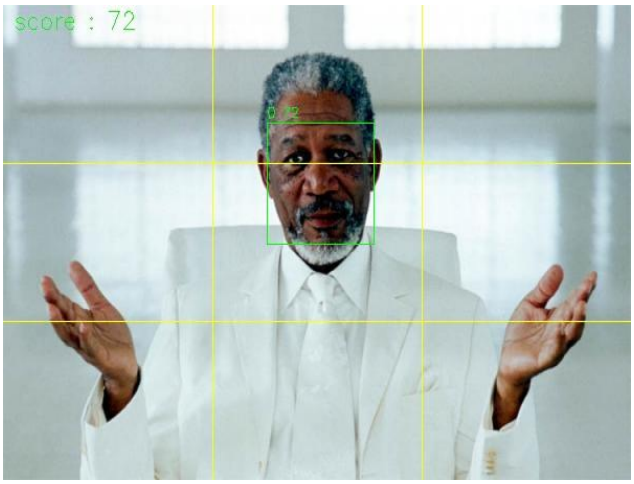
$$score(i) = \sum_i w(i) + \frac{1}{scoreRdt(i)} \quad (7)$$

where weights and distances are properly normalized in the range [1,100].

For both cases (whether single faces or group photos) we propose to include into the aesthetic evaluation the facial expression obtained just applying the method [9,10]; in particular the library SHORE locates faces (at different poses) and returns for each of them a value in the range [0,100] considering the following expressions: Happy, Angry, Sad, Surprised. These facial expression has been considered to provide positive or negative values to the score according to their common meaning. A further aesthetic criterion that consider the relative "closeness" of involved eyes has been also included.



a)



b)

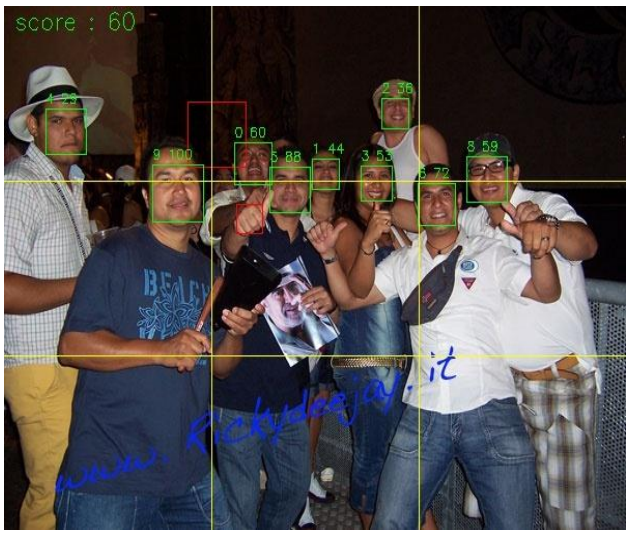
Figure 1. *Aesthetic assessment for pictures with a single face in clear foreground: a) SCORE 81% MOS 75% TOWARDS 71%. b) SCORE 72% MOS 70% TOWARDS 75%.*

Figure 2. *Aesthetic assessment for pictures with a single face in clear foreground: SCORE 73% MOS 80% TOWARDS 75%.*

Figure 3. *Example of aesthetic assessment for a picture with a single face having a low aesthetic scores: SCORE 10% MOS 43% TOWARDS 50% Edit the image in order to zoom the subject.*

Figure 4*. Example of aesthetic assessment for a picture with more than one people. Green boxes report the aesthetic score of each face: SCORE 67% MOS 73% TOWARDS 58%*

Figure 5. *Example of aesthetic assessment for a picture with more than one people. The face on the left has been discarded (thresholds too high): SCORE 58% MOS 42% TOWARDS Not see the face.*



a)

b)

Figure 6. *Example of aesthetic assessment for a picture with more than one people. False positives are correctly discarded: a) SCORE 60% MOS 64% TOWARDS 50% Eliminates the face with id 4; b) SCORE 56% MOS 76% TOWARDS 51%.*

The global aesthetic score of the input picture is then obtained summing up the contribution (for each face) of the various involved components:

$$finalScore = \frac{1}{F}\sum_i score(i) + \\ + k(Happy - Angry - LeftEyeClosed \\ - RightEyeClosed - Sad + Surprised)_i \quad (8)$$

where k=0.1. The value of k has been fixed after some empirical attempts but we think that there is space for a deeper investigation about the role of each involved aspects. Future works will be devoted to properly consider single contribution of each involved facial expressions through some machine learning engines [14].

3. Experiments and Results

To estimate the proposed aesthetic engine a database of about 100 images of different resolution (minimum 640x480) and quality has been considered. Images have been obtained by public repositories on the web and private collection.

They depict typical consumer photos involving people in different situations (holiday, party, etc.). For each picture we have applied our methodology obtaining a final aesthetic score in the range [0,100] taking into account both visual balancing and evaluation of facial characteristics as reported in Section 2. For sake of comparisons we have compared our results with [12, 15]. The system implemented in [15] returns also a suggestion for possible editing to improve the overall quality.

Also to have another subjective evaluation we compare our score with results obtained by visual assessment of 15 people through a MOS (mean opinion score) process. In all presented results we report the scoring computed by the proposed algorithm, the method in [12], and the subjective evaluation. For pictures having a single face, in a clear foreground, located in the middle of the framing all methods typically give an high value (Figure 1, Figure 2).

Conversely if a single face is less evident (small size, decentralized, etc.), a low score is typically obtained (Figure 3). Also for group photos the algorithm works in a satisfactory way (Figure 4) giving a greater score to ensemble of people (closeness is an issue) also verifying the location of the faces with respect to the four points derived by the rule of the thirds.

To avoid to pick-up some false positives as reported in the red boxes in the bottom part of Figure 5 and in Figure 6 a proper hard threshold have been considered. For our purposes is fundamental to avoid to include false positives in the pool of considered faces. In Figure 4 we report an example where our system loses one face but [12, 15] is not able to assess any score because it fails to detect people.

All images and results are available for download at the following web address:

http://iplab.dmi.unict.it/download/CGIV2012/

4. Conclusions and Future Works

In this paper we have presented some preliminary results in the field of aesthetic scoring of consumer photos involving people. Facial expressions and pose are used together with a series of heuristics devoted to encode the global spatial relation including neighborhood and size. Preliminary results and comparisons with existing works confirm the ability of the proposed method to encode aesthetic and emotional insights as expected. We plan to increase the number of images used for assessment of the overall methodology.

Future works will be also devoted to improve the overall methodology with respect the following issues:

- Deeper investigation about the role of each involved facial expression and relative pose; some subjective experiments devoted to better evaluate such aspects will be designed.

- The integration inside the model to further criterion that includes color appearance [7, 8] of involved semantic scenes [5, 6]. For still pictures of natural scenes (e.g. landscape, portrait, etc.) colors related to a few semantic classes have the most perceptive impact on the human visual system. From this point of view some basic chromatic classes are prominent (e.g., skin, vegetation, sky/sea). Although most aesthetic scoring techniques are completely blind to scene appearance, we aim to improve the overall performances for natural scene images by strongly relying on actual, and expected, image appearance.

To further assess the overall effectiveness of the proposed method during acquisition we are also working on a reference implementation of the system on a mobile platform [16, 17].

References

[1] Dhiraj Joshi, Ritendra Datta, Quang-Tuan Luong, Elena Fedorovskaya, James Z.Wang, Jia Li and Jiebo Luo. Aesthetics and Emotions in Images: a Computational Perspective - IEEE Signal Processing Magazine, Vol. 28, no. 5, pp. 94-115, September 2011;

[2] Ritendra. Datta, Dhirai. Joshi, Jia Li, and James. Z. Wang. Studying Aesthetics in Photographic Images Using a Computational Approach - In Proceedings of ECCV 2006, pp. 288–301;

[3] Y. Ke, X. Tang, and F. Jing. The Design of High-Level Features for Photo Quality Assessment - In Proceedings of CVPR 2006, pp. 419–426;

[4] Image Processing for Embedded Devices - Eds. S. Battiato, A.R. Bruna, G. Messina, G. Puglisi - ISSN: 1879-7458 - Applied Digital Imaging ebook series, ISBN: 978-1-60805-170-0, Bentham Science Publisher, 2010;

[5] S. Battiato, G.M. Farinella, G. Gallo, D. Ravì - Exploiting Textons Distributions on Spatial Hierarchy for Scene Classification – EURASIP Journal on Image and Video Processing – Special issue on Multimedia Modeling, 2010;

[6] G.M. Farinella, S. Battiato - Scene Classification in Compressed and Costrained Domain - IET Computer Vision - Vol. 5, Issue 5, pp. 320-334– 2011;

[7] F. Naccari, S. Battiato, A. Bruna, A. Capra, A. Castorina - Natural Scene Classification for Color Enhancement - IEEE Transactions on Consumer Electronics - Vol. 51, Issue 1, pp.234-239, February 2005;

[8] S. Battiato, A. Bosco, A. Castorina, G. Messina – Automatic Image Enhancement by Content Dependent Exposure Correction – EURASIP Journal on Applied Signal Processing – Vol.12 - pp.1849-1860, 2004;

[9] Congcong Li, Andrew Gallagher, Alexander C. Loui, Tsuhan Chen - Aesthetic Visual Quality Assessment of Consumer Photos with Faces - In IEEE International Conference on Image Processing (ICIP 2010);

[10] C. Küblbeck, A.Ernst. Face Detection and Tracking in Video Sequences Using the Modified Census Transformation – Image and Vision Computing, Vol.24, No. 6, Issue 1, June 2006, pp. 564-572;

[11] SHORE™ - Sophisticated High-speed Object Recognition Engine - http://www.iis.fraunhofer.de/en/bf/bsy/produkte/shore/, 2011;

[12] C. Li, A. C. Loui and T. Chen. Towards Aesthetics: a Photo Quality Assessment and Photo Selection System - In Proceedings of ACM Multimedia(2010), 827-830;

[13] Ligang Liu, Renjie Chen, Lior Wolf, Daniel Cohen-Or. Optimizing Photo Composition. Computer Graphics Forum, Vol.29, Issue 2, 2010, 469-478;

[14] J. Machajdik and A. Hanbury. Affective image Classification Using Features Inspired by Psychology and Art Theory - In Proceedings of ACM Multimedia 2010, pp. 83–92;

[15] TOWARDS: http://chenlab.ece.cornell.edu/projects/aesthetics/;

[16] S. Battiato, G. M. Farinella, E. Messina, G. Puglisi, D. Ravì, A. Capra, V. Tomaselli - On the Performances of Computer Vision Algorithms on Mobile Platforms – In Proceedings of Electronic Imaging 2012 - Digital Photography VIII - Vol. 8299 - 2012;

[17] S. Battiato, G. M. Farinella, N. Grippaldi, G. Puglisi – Content-based Image Resizing on Mobile Devices – In Proceedings of VISAPP 2012 – International Conference on Computer Vision Theory and Applications –2012;

[18] Flickr - http://www.flickr.com – 2012;

[19] Photo.Net: A Community of Photographers - http://photo.net – 2012;

[20] DPChallenge: A Digital Photography Contest - http://www.dpchallenge.com, 2012;

[21] Terra Galleria Photography - http://www.terragalleria.com - 2012;

[22] ACQUINE - Aesthetic Quality Inference Engine - Free Instant Impersonal Assessment of Photo Aesthetics - http://alipr.com – 2012.

Author Biography

Fabrizio Ravì has received his BS in Computer Science from the University of Catania (2011). His research interests include image and video enhancement for camera technology, image coding and multimedia forensics.

Sebastiano Battiato (battiato@dmi.unict.it.) is an associate professor in the Department of Mathematics and Computer Science at the University of Catania, Italy. He is also the director (and cofounder) of the International Computer Vision Summer School (ICVSS) in Sicily, Italy. His research interests include image enhancement and processing, image coding, camera imaging technology, and multimedia forensics. Prof. Battiato has a PhD in computer science and applied mathematics from the University of Naples. He is a Senior member of IEEE.