

Assessing Image Quality on Import From Camera

K. Armin Samii

Computer Science Undergraduate
Univ. of Calif. at Santa Cruz
ksamii@ucsc.edu

Uliana Popov

Computer Science Graduate
Univ. of Calif. at Santa Cruz
uliana@soe.ucsc.edu

Allison Carlisle

Biomolecular Engineering Undergraduate
Univ. of Calif. at Santa Cruz
acarlisle@ucsc.edu

James Davis

Computer Science Professor
Univ. of Calif. at Santa Cruz
davis@soe.ucsc.edu

February 19, 2011

Abstract

In this paper, we propose a novel method for ranking images as they are imported from a users camera. We provide the user with the best image from every scene photographed and a ranking of these scenes. Our goal is to filter images which are technically flawed to allow the photographer to focus on the best image from each similar set. By assuming a chronological import, we rank images relative to each other, eliminating the need for no-reference image based techniques. The set chosen by our algorithm matches the top two choices of photographers in our study with astonishing accuracy (97%).

1 Introduction

To focus our research and determine which image qualities are relevant, we assume three general steps a photographer takes between shooting and using a picture.

1. Cleanse: Sort through the images being imported and select the ones most suited for retouching, flag them, and remove the rest.
2. Retouch: Modify the raw files to stylize and enhance them.

3. Find: Select the image most suitable to a given task.

Our research focuses on the first step. By quantifying qualities which cause a photograph to be rejected unanimously, we have created an automated system which rejects over SOMEPERCENT% of the images a human would reject, based on SOMENUMBEROF data sets of SOMENUMBEROF images tested on SOMENUMBEROF users each. We have created a scale for rating each image's quality on a scale from one to ten. Over SOMEPERCENT% of our ratings were within one of the average user ranking, which was tested on the same dataset of SOMENUMBEROF images and SOMENUMBEROF different users.

2 Related Work

Various publications have focused on the retouching [...refs] and retrieval[2] steps, but there has been less work on cleansing. Various approaches have been proposed to rank any set of images based on aesthetic[...] or technical quality[...] but these do not make use of the reference-based algorithms which are possible when working with images directly from a camera import. While some works assess both tech-

nical and aesthetic qualities[...refs], we believe the two are separate problems. Aesthetics are largely preference-based and an automated algorithm may not be trusted by a photographer, whereas technical quality, although dependent on properties of the Human Visual System, is more readily automated without consideration of personal preference (and thus no human input is required). Therefore, our work assesses an image based on blur, noise, exposure, and the relationship between colors. Past research has quantified these artifacts globally[...refs] and independently (no-reference algorithms) [...refs]. Our approach combines local-feature algorithms [...refs] and reference-based algorithms [...refs] to extract similar foreground content in near-duplicate photographs and compare the foreground data to find a relative ranking. By focusing on the foreground, we reduce the computational errors that may arise when looking globally. For example, background blur and underexposure are not signs of a poor photograph[...ref], but may be mistaken as such[...refs, somebody who had this mistake]. Similarly, by separating foreground and background content, we can better measure exposure balance and weight noise levels which closer match the Human Visual System (HVS)[...refs, the foreground-background noise level paper] and assess exposure levels based on foreground-background histogram comparison, similar to[...refs?].

3 Quantifying Image Quality

Each selected factor will provide a ranking between zero and nine, with larger numbers indicating higher quality. Because an image which ranks low on any part would be considered poor, we have propose an algorithm which penalizes low scores more than it rewards high scores, because our primarily goal is to weed out poor images. We use an inverse-logarithmic scale for this overall ranking:

$$\sum_{i=1}^n \frac{\log^{-1}(W_i Q_i)}{rn}$$

Where k is the range of rankings (9 in our work), Q_i represents the rating of module i , and W_i represents

the weight of that module. Weights are assigned as follows:

1. $W_{exposure} = 40\%$
2. $W_{blur} = 30\%$
3. $W_{noise} = 15\%$
4. $W_{color} = 10\%$
5. $W_{gray} = 5\%$

3.1 Content Recognition

We obtain a bounding box around the most salient foreground object, in a method similar to [#]. Rather than using the precision that [#] uses, we aim for accuracy, and observed faster and equally accurate results with this method. We achieve high accuracy because of the simplicity, though secondary objects are sometimes ignored or mistaken as foreground. We have found that this does not heavily affect the outcome of our algorithm.

3.2 Similar-Image Clustering

To find near-duplicates, we first look at the time the photo was taken. The algorithm explained in [#] provides a good estimation of whether or not images were taken sequentially. We use this as a weight for further near-duplicate detection: The first step is to perform a fast 9-segment test similar to [...refs]. We divide the images into a 3x3 grid and compare the average color of each square. Because this is sensitive to exposure differences and movement, we perform a second content-based test if the results are inconclusive. Using a SOME ALGORITHM FROM NASA, we gather a ranking based on the number of matched interest points. We then calculate a weight from the timestamp using the formula ENTER FORMULA, with DESCRIBE SENSITIVITY CONSTANT. Weighting the two content-based algorithms with the rank from the timestamp, we can cluster similar images with high accuracy.

3.3 Blur Detection

Using the bounding box, the blur-detection algorithm quantifies the contrast between the edges and background. It ignores the rest of the image. In a method similar to [...refs], it obtains a value for the sharpness. This algorithm does not perform well when ranking diverse images, due to differences in scale and levels of acceptable background blur, so constraining it to the bounding box increases accuracy. Furthermore, we compare these rankings to the near-duplicates to remove any photograph-specific artifacts which we may not have accounted for. This increases the accuracy of the algorithm with just one photographs, and does even better when there is a sequence of similar images.

3.4 Noise Detection

We use a binary-weight scheme to weigh the noise, in a weak combination of local and global quality assessment. Noise within the bounding box has more negative weight than noise in the background. Our algorithm is similar to [...refs], which has results good enough to support itself.

3.5 Exposure

Because there is no universally well-performing model known for well-exposed histograms, the content-separated results provide a more accurate way of examining the photograph. As described in [...refs], we can accurately rate well exposed images bysome cool method....

3.6 Color Harmony

LA DI DA DI DA

4 Results

We have run our algorithm on several publicly-available datasets as well as our own. To obtain ground truth, we ALLISON DO STUFF HERE X USERS IN TURK AND WHATEV. Enumerated below are the results and comparisons to similar works.

1. Binary Classification: (fake)With a dataset of 4,000 users obtained through Amazon Mechanical Turk, we have found that our algorithm can distinguish between professional and non-professional images well. Barsky [2] and Luo *et. al*[1] obtained 93% accuracy when classifying images into the two categories. When classifying "non-professional" as ratings below five and "professional" as ratings above five (ratings of exactly five are ignored), our rating system matches a user's with 96% accuracy.
2. Quality Filtering: When asking a user to choose which images to keep and which to discard, our algorithm correctly discarded 82% of the images, and incorrectly discarded 18%. When limiting both the algorithm and the user to discarding a fixed number of images (ten), we improved these numbers to 92% accuracy with 6% false-negatives. These results rank favorably with [...ref].
3. Image Ranking Because this was not the goal of our work, we cannot compare to the state of the art[...ref] which received XX% accuracy, but we did manage a no-reference accuracy of within 30% of the user's rating. Further, when we introduced a single set of twenty near-duplicates to a user and asked them to rank from best to worst, their top three and bottom three matched our results with 97% accuracy.
4. Comparison with Ke[...cite]: When comparing Ke's data set to our ground truth, we see that their ground truth is based on different factors. Whereas our work matches Amazon Mechanical Turk users' ratings with an accuracy of 80%, Ke's only matches it with an accuracy of 45%. (Here, accuracy is defined as being within one standard deviation of the users' rankings.) CHARTS AND GRAPHS TO FINALLY MAKE IT CLEAR.

5 Future Work

Honestly, there's nothing left to be done. We beat this game. Maybe we can combine our work and Barsky and some photobooth folk in India together to make a superawesome combination of the three steps above. That takes some cutting and pasting, then bam, this conference is over. Great success!

References

- [1] Yiwen Luo and Xiaoou Tang. Photo and video quality evaluation: Focusing on the subject. In David Forsyth, Philip Torr, and Andrew Zisserman, editors, *Computer Vision – ECCV 2008*, volume 5304 of *Lecture Notes in Computer Science*, pages 386–399. Springer Berlin / Heidelberg, 2008. 10.1007/978-3-540-88690-7_29.
- [2] Che-Hua Yeh, Yuan-Chen Ho, Brian A. Barsky, and Ming Ouhyoung. Personalized photograph ranking and selection system. In *Proceedings of the international conference on Multimedia*, MM '10, pages 211–220, New York, NY, USA, 2010. ACM.