

Selective Removal of Extraneous Photographs

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

Bioinformatics/Computer Science Undergraduate
Univ. of Calif. at Santa Cruz
acarlis1@ucsc.edu

James Davis

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

February 26, 2011

Abstract

In this paper, we propose a method for selecting the most important images from a set of user photographs. We rank images in relation to others in the set, and provide the user with the best images from every scene photographed. We filter out images which are technically flawed to provide the photographer with a high-quality representative subset of the imported images. The filtered set chosen by our algorithm matches the user's choices with high accuracy (XX%).

3. Retrieve: Select the retouched images most suitable to a given task.

Our research focuses on the first step of the Photographer's Process. By quantifying visual qualities which cause a photograph to be rejected, we have created an automated system which rejects over XX% of the images a human would reject, based on XX data sets of XX images tested on XX users each. We are also accurate in rating images independently of their set, achieving XX% of our no-reference ratings being accurate to one standard deviation of the users' ratings on a ten-point scale.

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, which we call the Photographer's Process:

1. Remove: Sort through the images being imported and select the ones most suited for retouching, removing the rest from consideration.
2. Retouch: Modify the raw files just selected to stylize and enhance them.

2 Related Work

Various publications have focused on improving the quality of an image (the second step of the Photographer's Process)[2][16]; more deeply researched is assessing quality for the purposes of photograph retrieval[27][24][14][8][6]. The methods for image retrieval are not as useful for image removal because they assume a finalized photograph, rather than looking at the potential to be enhanced- or, more simply, they focus on aesthetic quality over technical quality. Removal of technically flawed photographs, as opposed to aesthetically displeasing ones,

is more readily automated without consideration of personal preference. We consider color harmony a technical aspect because attempting to color adjustment is a lossy process (XX ALLISON, I DON'T KNOW WHAT I'M TALKING ABOUT, CAN YOU MAKE SOMETHING UP AND FIND A REFERENCE PERCHANCE?)

Previous works in general image quality assessment have explored single features such as blur[13], exposure[11], and color harmony[4][5]. User tags on Flickr and online search engines have provided sets of similar images to compare[1][6]. Several papers have developed methods for separating professional and amateur photographs[23][9][15][18]. Unlike our work, these assume a finalized image.

Compressed image quality assessment has been heavily researched[10][26][25], but we assume a high quality import. [3] shows how a photograph's EXIF data can be used to robustly classify images into various semantic categories. Various aesthetic-based assessments focus on exploring human visual attention and sensitivity to photograph content[21][17][19], showing how content segmentation is useful.

Our work assesses an image based on blur, exposure, and the relationship between colors. We focus on the quality of foreground, and provide a relative ranking within a set of contextually similar images. An absolute rank is also provided, providing robustness in cases where there are no similar photographs to compare to in the input set.

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 a scale which penalizes low quality images more than rewarding high quality images for this overall ranking:

$$\sqrt[4]{\sum_{i=1}^n (W_i Q_i)^4}$$

Where r is the range of ratings, Q_i represents the rating of module i , and W_i represents the weight of that module. Weights are assigned as follows: XX ARE THEY REALLY?

- $W_{exposure} = 50\%$
- $W_{blur} = 40\%$
- $W_{color} = 10\%$

XX EACH ITEM BELOW SHOULD HAVE AN IMAGE TO GO WITH IT

3.1 Content Recognition

Using NASA's Vision Workbench[12], we obtain interest points and find the most dense region. We assume this to be the primary content in the image. It is used in detecting similar images, determining exposure quality, and calculating blur levels. Although there are more accurate methods available[22], our work only requires this simpler method which provides a bounding box around the most salient foreground object.

3.2 Similar-Image Clustering

To find similar images, we perform three steps of increasing complexity for high confidence. First, we use the timestamp to obtain a similarity index between all pairs of images. Based on the algorithm explained in [20], which finds gaps in timestamps, we have derived a formula which finds temporal closeness between two images:

$$T_{i,j} = \frac{G}{A_g}$$

Where G is the log of the time gap between the two images:

$$G_{i,j} = \log(|T_i - T_j|)$$

And A_g is the average gap within a window:

$$A_g = \sum_{a=i-w}^{j+w} G_{a,i}$$

We have empirically chosen a window size of $w = 8$. We use the calculated similarity index $T_{i,j}$ as a weight for the next two steps.

The first content-based step is simple 9-segment test. We divide the images into a 3x3 grid and compare the average color of each square. Because this is sensitive to exposure differences, subject movement, and camera reorientation, we then gather interest points using XX SOME ALGORITHM FROM NASA, OR THE AVERAGE COLOR OF THE ROI. We then compute a ranking based on the number of matched interest points.

The two content-based algorithms, weighted by the timestamp, cluster similar images with high accuracy.

3.3 Blur Detection

Using the foreground bounding box, the blur detection algorithm quantifies the contrast between the edges and background. Using the Image Gradient Model proposed in [13], we obtain a value for the sharpness. To increase accuracy and reduce the adverse affects that acceptable background blur may have, we constrain detection to the bounding box.

Next, we compute the difference in pixel values between the original image and a Gaussian blurred image. This provides little insight on no-reference images, but when looking at the bounding box of similar images, sharper edges results in a larger difference.

XX NEED EMPIRICAL EVIDENCE. COMPARE ORIGINAL IMAGE TO 5mm*1,2,3 REFOCUSED TO SHOW FOCUS BLUR, AND SOMETHING FOR MOTION BLUR (SIMULATED?).

3.4 Exposure

We compare the foreground and background exposure to determine quality. Based on the idea that the the mean value of image luminosity is an indication of how well-exposed an image is, images are segmented based on their average luminosity value. Luminosity is a measure of how bright the human visual system perceives a color. As the human eye is most sensitive to green and red ranges, those color portions are weighted more heavily to a color's overall perceived intensity. The calculation used by the

exposure analyst is

$$p = .59r + .3g + .11b$$

where p is the perceived intensity, and r , g , and b are the red, green, and blue components, respectively(XX THIS FORMULA NEEDS REFERENCE).

The categories of mean luminosity value (mean) correspond roughly to a parabolic mapping of exposure values, with well exposed photographs having a mean of $130 < p < 140$, and the further towards the extreme high and low means, the worse quality the image is. These divisions are used as the starting point of analysis. Within each category, different measures will indicate either a positive or negative overall impact on image quality. But the measurements can signify different changes in the various categories. For example, in an overall dark mean, having more extreme bright areas makes the image more balanced, while in a bright image they usually indicate that it has been overexposed.

The other measures of image quality are also based on luminosity, and were determined by taking various measurements on a deliberately chosen pool of images that had a wide variety of exposure problems. The measures are as follows: clipping, highlights, lowlights, the upper 60th and 98th percentiles, the lower 60th and 98th percentiles, and variance. Each of these measures are calculated on both the foreground and the background of the image.

Clipping is an indication of how much information loss there is in the image due to extreme shadows and bright spots. Each of these measures is taken in relation to the total number of pixels being measured.

Overexposure of an image can lead to large areas in the photo in which the human eye cannot discern any form (note that this definition can also indicate areas of a solid white value that contains absolutely no data about the form present). However, as the program is primarily concerned about how humans perceive images, the number of pixels in the highest five perceived values are used to calculate the highlights present in the image. Some highlighting can be desirable in an image, but the amount that is considered overexposed varies depending on the image mean.

Lowlights can be caused by underexposure, but they can also be a product of poor lighting (eg, photographing a shadowed object in extremely bright conditions). Again, some lowlights are desirable for contrast in an image, but too much leads to an undecipherable image.

The percentiles of the image are determined for both the bright and the dark sides of luminosity. For example, to calculate the lower 60th percentile, the value is found at which the sum of the number of pixels that are darker than the given value is equal to sixty percent of the pixels in the image. The upper percentiles are found by the summation of the pixels that are brighter than the given value. These four measures give a good indication of the spread of the luminosity, eg, how sharp the transitions between the extreme and middle values are.

The most extreme mean luminosity values are images of very poor quality, and are rated as such based solely on the mean value, although having a bright 98th high percentile can provide enough contrast to make a discernible image, though certainly nothing of quality. Means between 80 and 170 encompass nearly all images of medium through excellent exposure, and thus require the most analysis. Means below 100 are also part of the low quality batch of images, they have a fairly clear relationship with the 98th high percentile and are thus analyzed very simply.

Means between 100 and 120 are further subdivided by the amount of highlighting present. XX WRITE SOME MORE STUFF

Means between 120 and 138 were found to have the highest concentration of high ratings. XX WRITE SOME MORE STUFF

The images on the most extreme end of the bright scale were not very populated in the training image set, but were hypothesized to follow a similar curve as the dark side, with the redeeming quality being the presence of dark values.

3.5 Color Harmony

Based on the seven types of color harmonies described in [7], the "ColorCritic" module measures which types of harmony are present in a photograph.

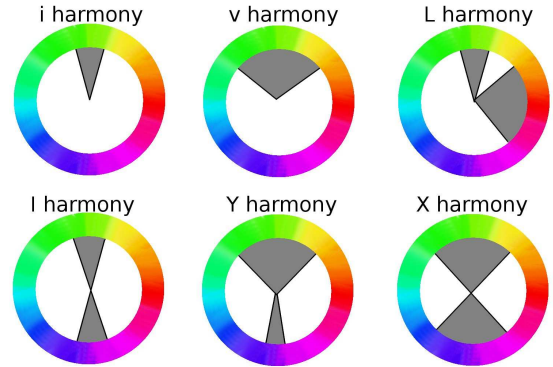


Figure 1: The harmonies described in [5] are used to find the quality of the relationships between colors.

It uses the I,v,L,I,Y, and X type harmonies (T and N are omitted).

For each harmony type, the hue is found with the most occurrences of supporting harmony (other pixels that fall in the harmonic range) as well as the percent of the image that has harmonious pixels. Similar to the Exposure Analyst, ColorCritic divides its images into several categories before further analysis. However, because people tend to like colorful, intense images (XX THIS NEEDS TO BE SUPPORTED WITH A REFERENCE- IF THE SAME AS COHEN, WE CAN JUST CITE THAT), ColorCritic divides the images by the color saturation in the image using the results from Mechanical Turk rankings to determine where the exact divisions are. The harmony types are used to differentiate between various qualities of images withing the saturation ranges.

The first harmony checked is the average and great image's I type. If there is a large amount of I harmony, the difference between I and Y harmonies is checked. Because I and Y overlap, you need to check how much of an increase if between the two. If it is very significant, Y harmony is most likely. When the difference is slight, you have a good I harmony (and thus high rating). If there is a medium difference, the I harmony is weak. Rating of the images can be determined on medium to great harmony. A similar check is done in X versus Y harmony. A last pass is done to check the L type on the great images, and a

pass checking the v and i harmonies is done on the average images.

The poor images are less well-defined but, the X harmony was the best place to start, and then a pass was done with i. XX ALLISON REWRITE THIS PARAGRAPH

4 Results

We have run our algorithm on several publicly available datasets as well as our own. We used Amazon Mechanical Turk and gathered a total of 454 (XX ADD QUALITY FILTER) users rankings across trials.

Enumerated below are the results and comparisons to similar works.

4.1 Quality Filtering

When asking a user to choose which images to keep and which to discard, our algorithm correctly discarded XX% of the images, and incorrectly discarded XX%. When limiting both the algorithm and the user to discarding a fixed number of chronological photographs (ten out of twenty), we improved these numbers to XX% accuracy with XX% false-negatives. (XX WE PROBABLY ONLY HAVE TIME TO GATHER THE SECOND NUMBER- FIXED NUMBER OF DISCARDS)

4.2 Image Rating

Separate trials were run to compare our algorithm to previously ranked images: 48 from [15], 100 from [18], and 459 from our own data (XX DIDN'T WE DECREASE THE SIZE OF THIS? WE DON'T HAVE JULIE'S, ONLY ULA'S). Here we obtain an absolute, no-reference rating of each image.

We took a random sample of 100 images from the 11,981 images provided by [18]. We asked 44 Turk users to rank 50 images on a scale from one to ten.

When comparing Ke's data set, we see that their ground truth is based on different factors (Fig. 2). Whereas our work matches Amazon Mechanical Turk users' ratings with an accuracy of 80%, Ke's only

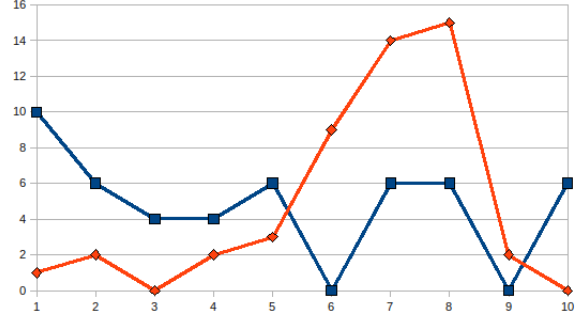


Figure 2: X-axis: ten-point scaled rating. Y-axis: number of images with this rating. We see that Ke's ratings do not match our ground truth. XX IN-ACCURACY ASIDE, WE SHOULD MAKE THIS WORK FOR B&W AND MAYBE LABEL AXES ON ACTUAL CHART?

matches it with an accuracy of 45%. Here, accuracy is defined as being within one standard deviation of the users' rankings.

XX INSERT DATA FOR OTHER DATA SET

4.3 Binary Classification

Using the same data as the Image Rating trials, we have found that our algorithm can distinguish between professional and non-professional images well. Barsky [27] and Luo *et. al*[18] obtained 96% accuracy when classifying images into the two categories. When classifying "non-professional" as ratings below XX and "professional" as ratings above XX, our rating system matches a user's with XX% accuracy.

5 Conclusion and Future Work

We propose a method of applying current research to automate another step of the Photographer's Process. By ranking an image's quality in relation to other images in the set, rather than an absolute scale, we obtain an accurate sorting of images. We improve upon past research which provides an absolute ranking by applying it to a relative scale. From this, we derive novel reference-based algorithms for exposure analysis, temporal closeness, and blur detection. To

ensure robustness when there are no similar sets (and thus no references), images are sorted both relatively within the set of similar images and globally within the set of all input images.

Our work focuses on a small portion of the Photographer's Process. In the future, we would like to see the idea of relative rankings applied to the second Retouching step. (It has already been extensively applied to the third Retrieval step.) Retouching can use relative processing to increase creativity between similar images or combine data from multiple images. With this, we would be able to automate the Photographic Process.

References

- [1] Tamara Lee Berg and David Forsyth. Automatic ranking of iconic images. Technical Report UCB/EECS-2007-13, EECS Department, University of California, Berkeley, Jan 2007.
- [2] Subhabrata Bhattacahrya, Rahul Sukthankar, and Mubarak Shah. A framework for photo-quality assessment and enhancement based on visual aesthetics. In *Proceedings of the international conference on Multimedia*, MM '10, pages 271–280, New York, NY, USA, 2010. ACM.
- [3] M. Boutell and J. Luo. Bayesian fusion of camera metadata cues in semantic scene classification. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–623 – II–630 Vol.2, 2004.
- [4] Kenneth E. Burchett. Color harmony attributes. *Color Research & Application*, 16(4):275–278, 1991.
- [5] Kenneth E. Burchett. Color harmony. *Color Research & Application*, 27(1):28–31, 2002.
- [6] Wei-Ta Chu and Chia-Hung Lin. Consumer photo management and browsing facilitated by near-duplicate detection with feature filtering. *Journal of Visual Communication and Image Representation*, 21(3):256 – 268, 2010.
- [7] Daniel Cohen-Or, Olga Sorkine, Ran Gal, Tomer Leyvand, and Ying-Qing Xu. Color harmonization. In *ACM SIGGRAPH 2006 Papers*, SIGGRAPH '06, pages 624–630, New York, NY, USA, 2006. ACM.
- [8] Jingyu Cui, Fang Wen, and Xiaoou Tang. Real time google and live image search re-ranking. In *Proceeding of the 16th ACM international conference on Multimedia*, MM '08, pages 729–732, New York, NY, USA, 2008. ACM.
- [9] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Wang. Studying aesthetics in photographic images using a computational approach. In Ale Leonardis, Horst Bischof, and Axel Pinz, editors, *Computer Vision ECCV 2006*, volume 3953 of *Lecture Notes in Computer Science*, pages 288–301. Springer Berlin / Heidelberg, 2006. 10.1007/11744078_23.
- [10] A.M. Eskicioglu and P.S. Fisher. Image quality measures and their performance. *Communications, IEEE Transactions on*, 43(12):2959–2965, December 1995.
- [11] Dong Guo, Yuan Cheng, Shaojie Zhuo, and T. Sim. Correcting over-exposure in photographs. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 515–521, June 2010.
- [12] M.D. Hancher, M.J. Broxton, and L.J. Edwards. *A User's Guide to the NASA Vision Workbench*. NASA Ames Research Center, Intelligent Systems Division, 2006.
- [13] Ping Hsu and Bing-Yu Chen. Blurred image detection and classification. In Shinichi Satoh, Frank Nack, and Minoru Etoh, editors, *Advances in Multimedia Modeling*, volume 4903 of *Lecture Notes in Computer Science*, pages 277–286. Springer Berlin / Heidelberg, 2008. 10.1007/978-3-540-77409-9_26.
- [14] Shao-Hang Kao, Wei-Yen Day, and Pu-Jen Cheng. An aesthetic-based approach to re-ranking web images. In Pu-Jen Cheng, Min-Yen Kan, Wai Lam, and Preslav Nakov, editors,

- Information Retrieval Technology*, volume 6458 of *Lecture Notes in Computer Science*, pages 610–623. Springer Berlin / Heidelberg, 2010. 10.1007/978-3-642-17187-1_57.
- [15] Yan Ke, Xiaoou Tang, and Feng Jing. The design of high-level features for photo quality assessment. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 419 – 426, June 2006.
 - [16] Johannes Kopf, Boris Neubert, Billy Chen, Michael Cohen, Daniel Cohen-Or, Oliver Deussen, Matt Uyttendaele, and Dani Lischinski. Deep photo: model-based photograph enhancement and viewing. *ACM Trans. Graph.*, 27:116:1–116:10, December 2008.
 - [17] Zhongkang Lu, W. Lin, X. Yang, EePing Ong, and Susu Yao. Modeling visual attention’s modulatory aftereffects on visual sensitivity and quality evaluation. *Image Processing, IEEE Transactions on*, 14(11):1928 –1942, November 2005.
 - [18] 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.
 - [19] Vitaly Pimenov. Fast image matching with visual attention and surf descriptors. In *Proceedings of the 19th International Conference on Computer Graphics and Vision*, pages 49–56, 2009.
 - [20] J.C. Platt, M. Czerwinski, and B.A. Field. Photoc: automatic clustering for browsing personal photographs. In *Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on*, volume 1, pages 6 – 10 Vol.1, December 2003.
 - [21] Xiaoshuai Sun, Hongxun Yao, Rongrong Ji, and Shaohui Liu. Photo assessment based on computational visual attention model. In *Proceedings of the seventeen ACM international conference on Multimedia, MM ’09*, pages 541–544, New York, NY, USA, 2009. ACM.
 - [22] Zhen Tang, Zhenjiang Miao, Yanli Wan, and Jia Li. Automatic foreground extraction for images and videos. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 2993 –2996, sept 2010.
 - [23] Hanghang Tong, Mingjing Li, Hong-Jiang Zhang, Jingrui He, and Changshui Zhang. Classification of digital photos taken by photographers or home users. In Kiyoharu Aizawa, Yuichi Nakamura, and Shinichi Satoh, editors, *Advances in Multimedia Information Processing - PCM 2004*, volume 3331 of *Lecture Notes in Computer Science*, pages 198–205. Springer Berlin / Heidelberg, 2005. 10.1007/978-3-540-30541-5_25.
 - [24] Roelof van Zwol, Börkur Sigurbjornsson, Ramu Adapala, Lluís Garcia Pueyo, Abhinav Katiyar, Kaushal Kurapati, Mridul Muralidharan, Sudar Muthu, Vanessa Murdock, Polly Ng, Anand Ramani, Anuj Sahai, Sriram Thiru Sathish, Hari Vasudev, and Upendra Vuyyuru. Faceted exploration of image search results. In *Proceedings of the 19th international conference on World wide web, WWW ’10*, pages 961–970, New York, NY, USA, 2010. ACM.
 - [25] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600 –612, April 2004.
 - [26] Zhou Wang, H.R. Sheikh, and A.C. Bovik. No-reference perceptual quality assessment of jpeg compressed images. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 1, pages I–477 – I–480 vol.1, 2002.

- [27] 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.