Selective Removal of Extraneous Photographs

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

Allison A. Carlisle
Bioinformatics/Computer Science Undergraduate
Univ. of Calif. at Santa Cruz
acarlisl@ucsc.edu

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

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

Abstract

In this paper, we propose a method for selecting the most representative high-quality images from a set of user photographs. To prevent redundancy arising from many similar images, we find all sets of near-duplicates. We then rank images based on three technical qualities: exposure, blur, and color harmony. Finally, we provide an ordering of the images which accounts for both quality and redundancy. Our goal is not to rank images based on aesthetic quality, but instead to help a photographer filter out technically flawed photographs and focus on high quality shots.

1. Introduction

With the ease of digital cameras, both amateur and professional photographers take more pictures than they will need. An organized user will delete excess photos and keep the most important ones.

To automate this process, an aesthetics ranking technique can be implemented, but this method would discard many images with potential to be fixed through retouching.

So, a technical quality ranking technique can be implemented, and the best images would be kept and the worst discarded. This method solves the previous problem, but does not take into account the redundancy of similar images, and the user would be left with a set which does not span their entire photo set.

To fix this, a similar image detector could be implemented giving the user the best image from every subset of similar images. This is still not ideal because it would leave the user with too many low-quality photographs that, although unique in the set, are not worth keeping.

We thus propose a technique which ranks images based on their technical qualities with a bias for unique images in the set.

To focus our research and determine which image qualities are relevant to technical image quality, 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.

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

A user study has shown 82% accuracy in our absolute ranking of photographs and 90% accuracy in determining which photographs belong to the same set.

2. Related Work

Previous works in image quality assessment have ranked image quality based on blur [10], exposure [8], color harmony [2][3], and duplicate detection [4].

Several papers have developed methods for separating professional and amateur photographs [18][6][11][13].

Compressed image quality assessment has been heavily researched[7][20][19], but we assume a high quality import. As described in [1], 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[16][12][14], demonstrating the viability of content segmentation for image assessment.

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:

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

Where n is the number of modules, Q_i represents the rating of module i, and W_i represents the weight of that module. Weights are assigned as follows:

- $W_{exposure} = 55\%$

- $W_{blur} = 35\%$

- $W_{color} = 10\%$

3.1. Content Recognition

Using a Harris Interest Operator (utilizing NASA's Vision Workbench[9]), interest points are obtained for each image. To extract a bounding box from these points, the most dense rectangle is calculated by maximizing the ratio of interest points to rectangle area. We assume this to be the primary subject. It is used in detecting similar images, determining exposure quality, and calculating blur levels. Although there are more accurate methods available[17], 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 an algorithm which segments a set based on gaps in the timestamps[15], we have derived a formula which finds temporal closeness between every pair of images. So, instead of finding gaps in sets, we rank pairs of images on a 0-9 scale to allow for nonconsecutive images to be grouped together:

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

Where G is the log of the time gap between the two images:
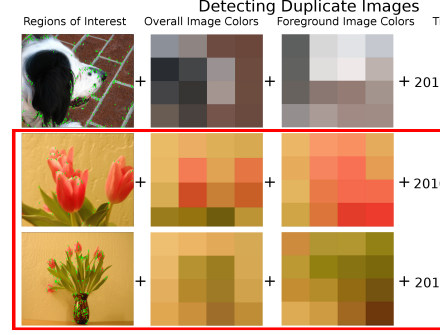
$$G_{i,j} = \log(|T_i - T_j|)$$

Figure 1. We cluster images based on their temporal nearness, overall color distribution, and foreground subject color distribution.

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 $w = 8$. The calculated similarity index $T_{i,j}$ is used as a weight for the next two steps.

Next, the distributions of the red, green, and blue channel of the image are extracted into a 4x4 grid. To compare two images, we find the minimum difference of the three channels for each square and deduct points if the most similar channel is above a difference threshold of 15%.

Because this is sensitive to subject and movement, we then take the bounding box derived in 3.1 and rerun the color channel similarity algorithm with stricter threshold of 10%.

A final similarity index is found by adding the scaled timestamp to the results of the color distribution algorithms. Similar images are clustered with a Quality-Threshold based clustering algorithm.

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 [10], 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.
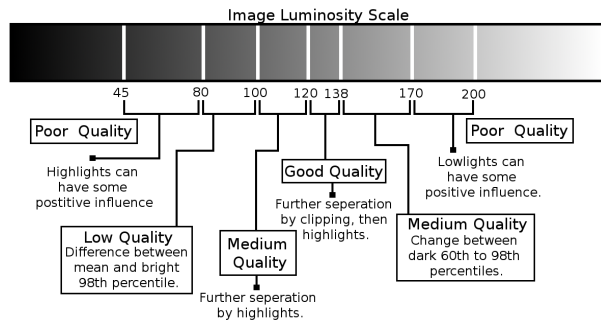
Figure 2. The Exposure process: This figure shows the inital segmentation of images by mean (values 0 to 255), then the further segmentation of images of medium and good quality (mean ranges 100-120, and 120-138 respectively).

3.4. Exposure

Image exposure is a measure of how appropriate the lighting is in a given image. While traditionally measured in-camera with light meter over the area, our exposure algorithm evaluates how the lighting in the result image appears to a human observer. Images are ranked according to various measures of their luminosity. 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).

Based on the idea that the the mean value of image luminosity is an approximate indication of how well-exposed an image is, images are first segmented based on their average luminosity value.

The categories of mean luminosity value (mean) correspond roughly to a parabolic mapping of exposure values, with well exposed photographs having a mean of $130 < p < 140$: 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 indicate either a positive or negative overall impact on image quality, which impact depends upon 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, relative to the area over which they are calculated.

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.

Lowlights can be caused by underexposure, but they can also be a product of poor lighting (e.g., 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.

Clipping is an indication of how much information loss there is in the image due to extreme shadows and bright spots, measured as a sum of the Highlights and Lowlights.

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, e.g., 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. After being subdivided into mean categories, they are again subdivided into further categories (see figure 3.4) before analysis.

3.5. Color Harmony

Based on the seven types of color harmonies described in [5], the "ColorCritic" module measures which types of harmony are present in a photograph. 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 it's images into several categories before further analysis. However, because people tend to like
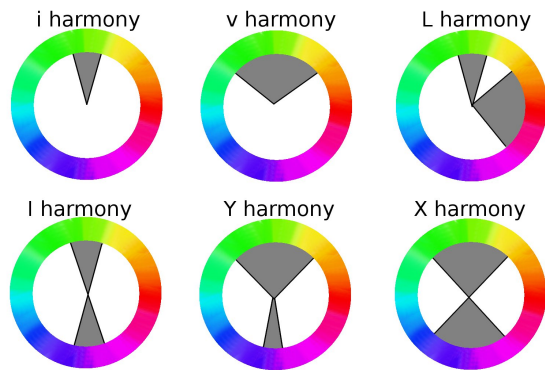
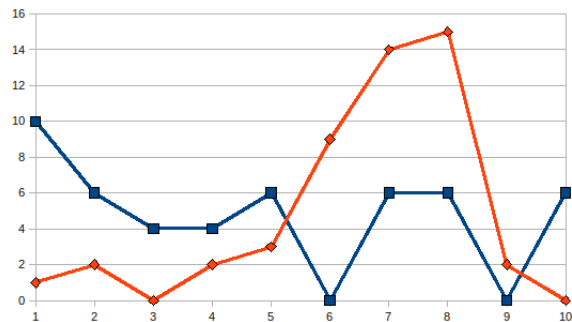Figure 3. The harmonies described in [3] are used to find the quality of the relationships between colors.

Figure 4. 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 INACCURACY ASIDE, WE SHOULD MAKE THIS WORK FOR B&W AND MAYBE LABEL AXES ON ACTUAL CHART?

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 follow a similar path, starting with analysis by the X harmony, and ending with a pass done with i harmony.

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 [11], 100 from [13], 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 [13]. 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. 4). 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.

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 [21] and Luo *et. al*[13] 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 rank-

ing 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] 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. 1

[2] Kenneth E. Burchett. Color harmony attributes. *Color Research & Application*, 16(4):275–278, 1991. 1

[3] Kenneth E. Burchett. Color harmony. *Color Research & Application*, 27(1):28–31, 2002. 1, 4

[4] 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. 1

[5] Daniel Cohen-Or, Olga Sorkine, Ran Gal, Tommer Leyvand, and Ying-Qing Xu. Color harmonization. In *ACM SIGGRAPH 2006 Papers*, SIGGRAPH '06, pages 624–630, New York, NY, USA, 2006. ACM. 3

[6] 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. 1

[7] A.M. Eskicioglu and P.S. Fisher. Image quality measures and their performance. *Communications, IEEE Transactions on*, 43(12):2959 –2965, December 1995. 1

[8] 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. 1

[9] 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. 2

[10] 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. 1, 2

[11] 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. 1, 4

[12] 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. 1

[13] 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. 1, 4

[14] 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. 1

[15] J.C. Platt, M. Czerwinski, and B.A. Field. Phototoc: 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. 2

[16] 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. 1

[17] 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. 2

[18] 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. 1

[19] 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. 1

[20] 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. 1

[21] 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. 4