SELECTIVE REMOVAL OF EXTRANEOUS PHOTOGRAPHS

K. Armin Samii, Allison Carlisle, Uliana Popov, James Davis

[ksamii,acarlisl]@ucsc.edu,[uliana,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 subjective aesthetic qualities, but instead to help a photographer filter out technically flawed photographs and focus on objectively high quality shots.

Index Terms— image processing, quality screening, exposure, meta data, color content

1. INTRODUCTION

Digitization of cameras has made photography accessable to a larger audience than ever before. Photographic collections can easily become larger than any one photographer has time to handle because of the instantaneous transition from in-camera to computer library. The first culling of photographs needs to eliminate those unworthy of human attention; the technically flawed and otherwise poor-quality images. 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.

Implementing a technical quality ranking technique will keep the photographs suited for retouching, but does not take into account 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 imported images and remove the ones least suited for retouching.
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.

We implement three modules to assess technical quality. The blur detection looks at the gradient magnitude of edges. Exposure is measured by comparing average luminosities within an image. Color harmony is based on Cohen's model[2]. Similar-images are clustered into groups using the color content and timestamps of an image. Finally, a reordering of the set is provided, with images most suited for removal at the back and high-quality images at the front. A user study has shown 82% accuracy in our absolute ranking of photographs and 78% accuracy in determining which photographs belong to the same set.

2. RELATED WORK

Previous works [7], [3], [10] assessed image quality to be able to tell whether the image was taken by a professional photographer or by an amatur user, and got very high accuracy in binarry classification of bad and good images. This allows to retrieve only high quality images from a collection of images, that were made by different users, which is extremely useful for web queries. Applying to similar data [12], [14] focused on exploring a personalized photo assessment. All those works focus on finalized images while our goal is to provide user a set which might be retouched and fixed.

Smart clustering and event classification were explored by [9], [6] while we chose the best image from each scene.

[8] described a promissing method of automatically rating and ranking images based on image content and time-metadata, but they [reword it!!!] didn't provided any datasets and their user study showed that their results are better than random. Ha-ha, we are better!!!

3. QUANTIFYING IMAGE QUALITY

First, we find a bounding box around the primary subject for use by the blur detection and exposure assessment. We then

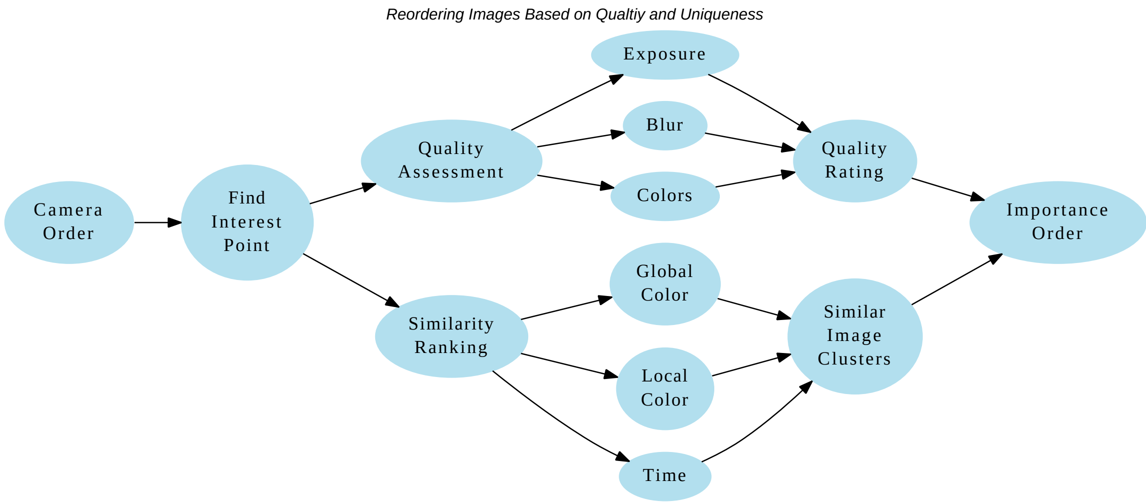

Fig. 1. Interest points are used to find the subject. Blur detection, color harmony, and exposure algorithms calculate an image's quality rating. The color distribution of the local subject and global image, coupled with the timestamp, cluster images into groups. A reordering of the input results, allowing the user to remove images at the end of this ordering.

cluster images into groups. Finally, we perform the quality assessment.

The blur, exposure, and color assessment modules 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:

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

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:

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

3.1. Subject Recognition

Using a Harris Interest Operator (utilizing NASA's Vision Workbench[4]), 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[13], our work only requires this simpler method which provides a bounding box around the most salient foreground object.

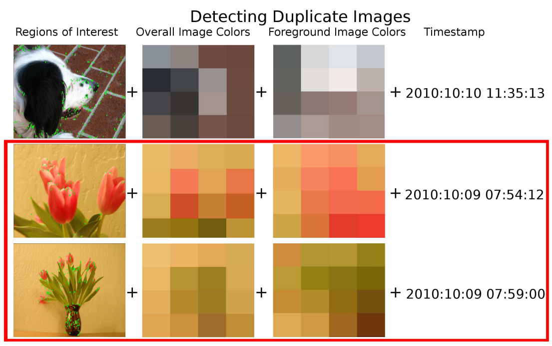

Fig. 2. We cluster images based on their temporal nearness, overall color distribution, and foreground subject color distribution.

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[11], 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:

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

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 camera 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 [5], 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.

3.4. Exposure

Image exposure is a measure of how approproate 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

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

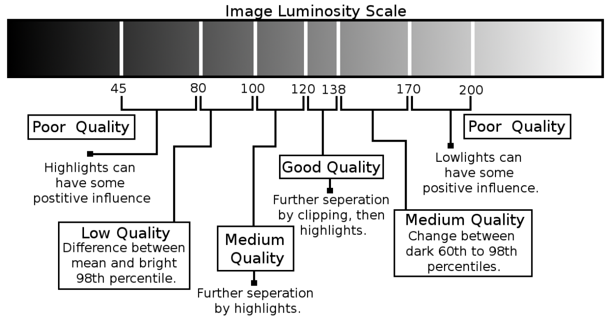

Fig. 3. The Exposure process: This figure shows the inital segmentation of images by mean (values 0 to 255), then the further segmentation of images of medium and good quality (mean ranges 100-120, and 120-138 respectively).

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

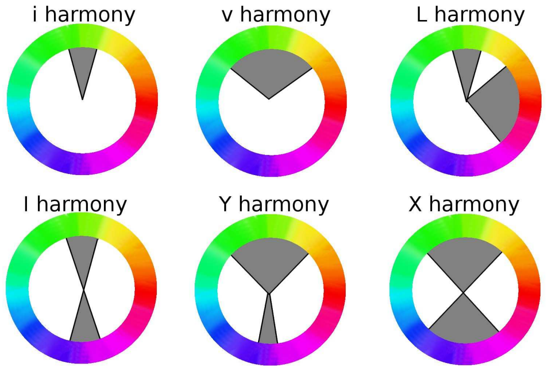

Fig. 4. The harmonies described in [1] are used to find the quality of the relationships between colors.

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 [2], 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 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 users rankings across trials.

Enumerated below are the results and comparisons to similar works.

4.1. Similar Image Detection

We asked Turk users to group together photographs within four sets of twenty images. Our similarity ranking module agreed with Turk users with 78% accuracy.

4.2. Image Rating

Separate trials were run to compare our algorithm to previously ranked images: 48 from Ke, Tang, and Jing[7], and 100 from Luo and Tang[10]. We then compared 459 of our own images to Turk users. Here we compare against an absolute rating of each image without regard to uniqueness.

We took a random sample of 100 images from Luo's and Tang's 11,981 images [10]. We asked 44 Turk users to rank 50 images each on a scale from one to ten.

Using the 48 image data set from Ke, Tang, and Jing[7], and Amazon Turk ratings as a ground truth, we achieve an accuracy of 80%, and the authors achieve an accuracy of 45%. We assume ”accuracy” is a rating that is within one standard deviation of the distribution of users ratings per image.

Similarly, with Luo and Tang[10], we achieve an accuracy of XX%, and the authors an accuracy of XX%, showing we are comparable.

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 [14] and Luo *et. al*[10] obtained 96% accuracy when classifying images into the two categories. When classifying ”non-professional” as Turk ratings below XX/9 and ”professional” above XX/9, 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 focusing on obtaining an ordering which is representative of all photographs taken, we obtain a diverse set of high quality images similar to what a user would have chosen manually. We derive a novel algorithm for analyzing exposure quality. We improve upon previous algorithms which find temporal gaps between images to obtain a metric for temporal nearness. The final ordering depends on both the quality ranking and the number of similar images which have already appeared.

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.

6. REFERENCES

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

[2] D. Cohen-Or, O. Sorkine, R. Gal, T. Leyvand, and Y. Xu. Color harmonization. In *Siggraph*. ACM, 2006.

[3] R. Datta, D. Joshi, J. Li, and J. Wang. Studying aesthetics in photographic images using a comput. approach. In *ECCV*, volume 3953 of *Lec. Notes in Comp. Sci.* Springer Berlin / Heidelberg, 2006.

[4] M.D. Hancher, M.J. Broxton, and L.J. Edwards. *A User's Guide to the NASA Vision Workbench.* NASA Ames, Intelligent Systems Division, 2006.

[5] P. Hsu and B. Chen. Blurred image detection and classification. In *Adv. in Mult. Model.*, volume 4903 of *Lec. Notes in Comp. Sci.* Springer Berlin / Heidelberg, 2008.

[6] Chul-Jin Jang, Ji-Yeon Lee, Jeong-Won Lee, and Hwan-Gue Cho. Smart management system for digital photographs using temporal and spatial features with exif metadata. In *Digital Information Management, 2007. ICDIM '07. 2nd International Conference on*, volume 1, pages 110 –115, October 2007.

[7] Yan Ke, Xiaoou Tang, and Feng Jing. The design of high-level features for photo quality assessment. In *CVPR 2006*, volume 1, pages 419 – 426, June 2006.

[8] Daniel Kormann, Peter Dunker, and Ronny Paduschek. Automatic rating and selection of digital photographs. In *Semantic Multimedia*, volume 5887 of *Lecture Notes in Computer Science*, pages 192–195. Springer Berlin / Heidelberg, 2009. 10.1007/978-3-642-10543-2_23.

[9] A.C. Loui and A. Savakis. Automated event clustering and quality screening of consumer pictures for digital albuming. *Multimedia, IEEE Transactions on*, 5(3):390 – 402, September 2003.

[10] Y. Luo and X. Tang. Photo and video quality evaluation: Focusing on the subject. In *ECCV 2008*, volume 5304, pages 386–399. Springer Berlin / Heidelberg, 2008.

[11] J.C. Platt, M. Czerwinski, and B.A. Field. Phototoc: automatic clustering for browsing personal photographs. In *ICICS and PCM*, volume 1, 2003.

[12] X. Sun, H. Yao, R. Ji, and S. Liu. Photo assessment based on computational visual attention model. In *Proc. ACM Multimedia*, MM '09. ACM, 2009.

[13] Z. Tang, Z. Miao, Y. Wan, and J. Li. Automatic foreground extraction for images and videos. In *ICIP*, 2010.

[14] C.H. Yeh, Y.C. Ho, B.A. Barsky, and M. Ouhyoung. Personalized photograph ranking and selection system. In *Proc. of the Int. Conf. on MM.*, MM '10. ACM, 2010.