REPRESENTATIVE IMAGE THUMBNAILS FOR GOOD BROWSING

Ramin Samadani, Suk Hwan Lim and Dan Tretter

HP Labs, Media Technologies Laboratory
Ramin.Samadani@hp.com

ABSTRACT

Image thumbnails are commonly used for selecting images for display, sharing or printing. Standard thumbnails, generated with current techniques, do not distinguish between high and low quality originals. Both sharp and blurry originals appear sharp in the thumbnails, and both clean and noisy originals appear clean in the thumbnails. This leads to errors and inefficiencies during image selection. In this paper, thumbnails generated using image analysis better represent the local blur and the noise of the originals. The new thumbnails provide a quick, natural way for users to identify images of good quality, while allowing the viewer's knowledge to select desired subject matter. Computer simulations with added blur and noise show the new thumbnails better represent images of differing qualities. Validation of these findings is found in a subjective evaluation reported elsewhere but summarized below.

Index Terms— image quality, blur, noise, resampling, image thumbnails

1. INTRODUCTION

Image thumbnails are pervasive. Computer operating systems and applications display image thumbnails of folders or albums. Retail photo kiosks let users review thumbnails, touch the screen at the thumbnails, and then print the selected photos. Image sharing sites display thumbnails of photo albums. Small displays on printers, cameras, cell phones, and video players let users preview images before taking actions such as viewing, mailing, printing, or deleting.

Standard thumbnail generation involves lowpass filtering and downsampling. This process results in thumbnails that do not represent the quality of the high resolution originals. None of the many sources of image blur, including unintentional misfocus and camera shake, as well as intentional depth of field local blurs are represented. Image noise, particularly prevalent in night or indoor scenes, is also not preserved.

Browsing with standard thumbnails leads to errors and inefficiencies. While browsing, one can easily select a normally appearing thumbnail to find that the original is blurred, noisy, or both. The same problem on printer or camera LCDs leads to erroneous selections for printing or deleting.

This paper describes new thumbnails that alleviate these problems by representing original image quality in addition to image composition. Figure 1 shows examples of the results for the cropped originals shown in Figure 2.

A benefit of the new thumbnails is that they are natural to interpret; there is no learning necessary to understand the blur and noise shown in the new thumbnails. The alternate approach of automatic image ranking by quality [1] is extremely difficult because the knowledge about the subjects of interest resides with the user, not with the image. For example, with the new thumbnails the user can quickly check whether the subject of interest is in focus.

2. PROBLEM FORMULATION

For simplicity of notation, images are considered column stacked vectors [2]. Given a distorted input image, $\boldsymbol{d}$, we create a thumbnail $\boldsymbol{t} = T(\boldsymbol{d})$ using a potentially non-linear process. The distortion of the input image is modeled by,

$$\boldsymbol{d} = B\boldsymbol{c} + \boldsymbol{n}, \tag{1}$$

Where the vector $\boldsymbol{c}$ represents an ideal image, the matrix B represents, in general, a space-varying blur that may correspond to unintended distortions such as camera shake, motion blur or misfocus, and $\boldsymbol{n}$ represents an additive, perhaps correlated, noise. Well taken digital photographs will not have unintended distortions. In this case, the noise $\boldsymbol{n} = 0$. But the matrix B may not be the identity, still representing the space-varying depth of field blur. In the special case of infinite depth of field, $B = I$, and therefore $\boldsymbol{d} = \boldsymbol{c}$.

3. STANDARD THUMBNAILS

We first consider the limitations of standard thumbnail generation. Commonly used thumbnails differ in the details of the filters applied [2], but they consist of a linear process, first applying an antialiasing lowpass filter, A, followed by subsampling, S. The thumbnails are thus given by

$$\boldsymbol{t}_s = T_s\boldsymbol{d} = SA\boldsymbol{d}, \tag{2}$$

where T_s represents the combination of filtering and subsampling.

Expanding Equation 2 using the image model in Equation 1 results in,

$$\boldsymbol{t}_s = S(AB\boldsymbol{c} + A\boldsymbol{n}). \tag{3}$$

Analysis of the quantity in parenthesis explains why standard thumbnails appear sharp and clean, even if the input image $\boldsymbol{d}$ has blur and noise added. First, the bandwidth of a typical blur filter B is broader than the bandwidth of the antialiasing filter A for typical subsampling factors (the ones used in our tests were between 10 and 23). Thus $AB \approx A$, in Equation 3. Considering noise next, antialiasing filter A applied to $\boldsymbol{n}$ will result in output filtered noise variance much lower than the input variance, so that $A\boldsymbol{n} \approx \boldsymbol{0}$. This is true for typical noise levels and for any practical antialiasing filter. The case of a $k \times k$ boxcar filter, for a subsampling factor k is particularly easy to analyze. If the input noise is uncorrelated, the output noise variance will be reduced by a factor of $\frac{1}{k^2}$. For antialiasing filter, A, the simulations in this section used a boxcar filter corresponding to $k = 10$, resulting in noise standard deviation of $\frac{1}{10}$ of the input standard deviation.

Fig. 1. Standard thumbnails (left column) and new thumbnails (right column) for three different images.

With these approximations, $T_s\boldsymbol{d} \approx T_s\boldsymbol{c}$. The standard thumbnail for the distorted image $\boldsymbol{d}$ will be very similar to the thumbnail for the ideal image $\boldsymbol{c}$ for typical levels of blur and noise.

These approximations are confirmed by simulations that apply differing amounts of blur and noise to input test images to generate different distorted images $\boldsymbol{d}$ in Equation 1. For the matrix B, $m \times m$ boxcar filters with $m \in \{1, 3, 5, 7, 9, 11\}$ were applied. For the noise, a moving average noise generated by filtering white Gaussian noise with a 3×3 boxcar filter (to roughly simulate the observed noise correlation in actual photographs [3]) was applied. The standard deviation of the noise was set to $\sigma_n \in \{0, 2, 4, \cdots, 20\}$. Thus, for each test image, this generates a set of images $\boldsymbol{d_{ij}}$ indexed by blur and noise, that include the original and 65 distorted versions with differing amounts of blur and noise.

Considering the original image $\boldsymbol{c}$ as a vector in a very high dimensional space, a distorted image may be expressed as the addition of two vectors to the original image, given by,

$$\boldsymbol{d} = \boldsymbol{c} + (B\boldsymbol{c} - \boldsymbol{c}) + \boldsymbol{n} = B\boldsymbol{c} + n. \tag{4}$$

This equation shows that the noise component is independent of the input image, but that the blur distortion will depend on image content. Higher spatial frequency content in the input image results in more blur distortion.

The mean square error (MSE) between a distorted image $\boldsymbol{d}_{ij}$, and undistorted image $\boldsymbol{c} \equiv \boldsymbol{d}_{00}$, is proportional to the square norm between the images interpreted as vectors, given by $\frac{1}{N} \parallel \boldsymbol{c} - \boldsymbol{d}_{ij} \parallel^2$, where N is the total number of pixels in the image.

Figure 3 plots the MSE between each distorted image and the undistorted version. Two different images, each with 1024×1280 pixels, are used in the simulations, to illustrate the changing nature of the blur, depending on image content. The image of the water plants on the left has typical spatial frequencies, but the image of the ground cover on the right has higher spatial frequencies, that are seen in the figure by observing the faster increase of MSE along the blur axis of the ground cover image. Since the blur and noise are independent, the MSEs add for images with both blur and noise. It is also seen that the noise only component of MSE is image independent, by viewing the graphs along the noise axis.

Figure 4 plots the MSE of thumbnails, of size 102×128 pixels, (subsampling factor $k = 10$) generated by our new approach as well as standard thumbnail MSEs. The mean square errors shown in each plot are between the reference thumbnail, for the input image without blur and noise, and the thumbnails for the other simulated

Fig. 2. Cropped portions of the three originals

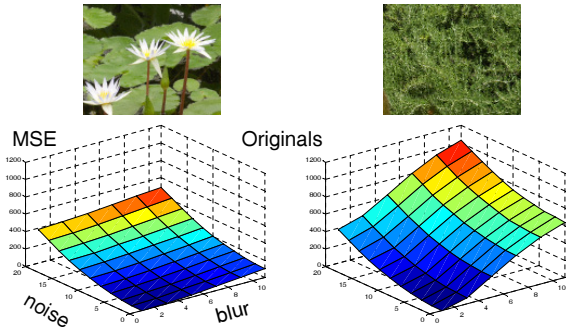

Fig. 3. Mean square errors for two different images

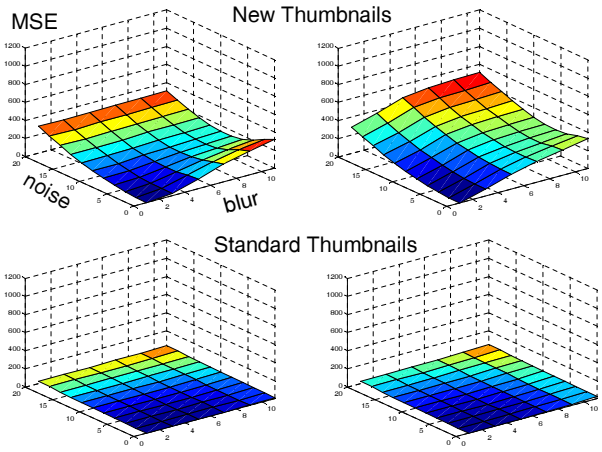

Fig. 4. Mean square errors for thumbnails

images. Relevant to the discussion in this section, the bottom plots of the figure show the surfaces for the standard thumbnails are near zero for all the different thumbnails (corresponding to the different input images $\boldsymbol{d}_{ij}$). The thumbnails are also observed to be visually very similar, showing neither blur nor noise.

4. NEW THUMBNAILS

The new thumbnails, $\boldsymbol{t_n}$, are generated by starting first with the standard thumbnail, which was shown in the previous section to be clean even for distorted input images. To this standard thumbnail, blur and noise are applied to correspond to the blur and noise in the original,

$$\boldsymbol{t}_n = \boldsymbol{t}_b + \boldsymbol{n}_t = B_t \boldsymbol{t}_s + \boldsymbol{n}_t. \tag{5}$$

Our work takes advantage of prior work on the very difficult problems of image denoising [4] and blind deconvolution [5], where the goal is to recover $\boldsymbol{c}$ in Equation 1 from $\boldsymbol{d}$. The goal for our work, however, is to generate a low resolution thumbnail $\boldsymbol{t}_n$, not the exact reconstruction of high resolution $\boldsymbol{c}$. This changes the requirements of our component algorithms. For example, our solution works well with both shake and defocus blurs, by applying an appropriate space-varying Gaussian blur. The details of the applied blur kernel is not critical to our results. Similarly for noise, we do not need an extremely accurate noise estimate, but rather a rough, fast one may be sufficient.

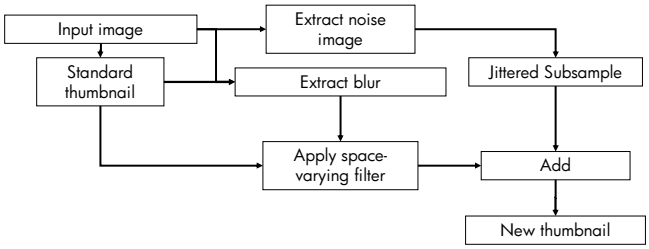

Fig. 5. To generate a new thumbnail, we start with a standard thumbnail and use image analysis to estimate and apply local blur and noise.

Figure 5 shows the processing that generates the new thumbnails. The *Extract blur* block results in a two dimensional thumbnail resolution *blur map*, $\boldsymbol{m}$, with estimates of the amount of blur at each location. The block *Apply space-varying filter* applies a filter based on the blur map. This local computation accounts for depth of field as well as undesired blurs. The blur map is determined without identification of the type of blur. The assumption is that users will not be able to distinguish between different types of blur in a thumbnail very easily.

The local amount of blur is computed by noting that the image edge profiles differ between sharp and blurry images. At an edge, for example, the profile of a blurry high resolution image wil be more gradual than its corresponding low resolution standard thumbnail, $\boldsymbol{t}_s$, whose profile will be steeper [6]. Applying successively larger blurs to $\boldsymbol{t}_s$ will cause its edge profile to become more gradual, and to correspond better to the blurry original. To have the system work with various image features, and not just edges, the computation is based on *pixel range* (difference between maximum and minimum pixel values in a spatial neighborhood) to determine the local image profiles.

During the building of the *blur map*, a set of Gaussian filtered low-pass versions of the standard thumbnail, $\boldsymbol{l}_\sigma$ are created,

$$\boldsymbol{l}_\sigma = g_\sigma * \boldsymbol{t}_s. \tag{6}$$

Eleven $\boldsymbol{l}_\sigma$, with $\sigma_j \in \{0, 0.5000, 0.7111, 0.9222, 1.1333, \cdots, 2.4\}$, are generated. The image $\boldsymbol{l}_0$ represents the unblurred thumbnail $\boldsymbol{t}_s$. The remaining images correspond to increasing blurs, starting with $\sigma = 0.5$, ending with $\sigma = 2.4$ and with increment 0.2111.

From the original image, $\boldsymbol{d}$, a low resolution *range map*, is computed. First, the maximum absolute difference for each center pixel and its eight neighbors is calculated. This high resolution range map is reduced to thumbnail size by taking the maximum in a high resolution neighborhood of the same size as the resampling factor k. The low resolution original range map is called $\boldsymbol{r}_o$.

Similarly, from each of the images $\boldsymbol{l}_\sigma$, a low resolution range map, $\boldsymbol{r}_\sigma$, is generated. The blur map value at each pixel index, i, is then computed using

$$\boldsymbol{m}(i) = \min_j \{j \mid \boldsymbol{r}_{\sigma_j}(i) \leq \rho \boldsymbol{r}_o(i)\}, \tag{7}$$

where ρ is a constant that sets the amount of blur added. Equation 7 implements the idea, described earlier, of reflecting in the thumbnails the local pixel range found in the high resolution original. Using this blur map, at each pixel, i, a space-varying blurred thumbnail $\boldsymbol{t}_b$, the first term in Equation 5, is created by selecting values from the appropriate blurred thumbnail,

$$\boldsymbol{t}_b(i) = \boldsymbol{l}_{\sigma_m}(i). \tag{8}$$

More accurate blur maps and space-varying filter implementations are possible [7], but this simple approach has worked well with the tested images.

For the noise component, $\boldsymbol{n}_t$, a simple, modified wavelet based soft thresholding [8] (known as *VisuShrink*), was used. Although superceded for traditional denoising applications, this simple approach is suitable for our application.

The noise residual is based on a high-pass filtered original, $\boldsymbol{h} = \boldsymbol{d} - g_1 * \boldsymbol{d}$, where g_1 is a Gaussian filter with $\sigma = 1$. Following Donoho, the high resolution noise, $\hat{\boldsymbol{n}}$ at each pixel i, is estimated using

$$\hat{\boldsymbol{n}}(i) = \boldsymbol{h}(i) - sgn(\boldsymbol{h}(i))(|\boldsymbol{h}(i)| - \lambda)_+, \tag{9}$$

with $x_+ = x$ if $x \geq 0$ and $x_+ = 0$ if $x < 0$.

The threshold λ is determined by first estimating noise standard deviation, $\hat{\sigma}_n = h_m/.6745$, where h_m is the median of absolute values of the pixels $|\boldsymbol{h}(i)|$. Then, threshold $\lambda = 1.6\hat{\sigma}_n log(N)$ is used, where N is the number of pixels in the original. The factor 1.6 is an empirical gain adjustment to account for noise correlation typically seen in digital photographs [3]. The same fixed factor was used for all of the images.

From $\hat{\boldsymbol{n}}$, the low resolution thumbnail noise $\boldsymbol{n}_t$ in Equation 5 is generated by subsampling. The subsampling of the noise component is justified by considering the autocorrelation of a *discrete time*, stationary Gaussian noise process after subsampling. In particular, the noise or variance will remain unchanged after subsampling on a regular sampling grid. The noise generation process used is not perfect, however, allowing some high spatial frequency signal to appear in the noise image. Jittered sampling [9] was used to breakup potential Moire (observed in only two of the images) from any residual image textures that appear in the noise image.

Our research software generated new thumbnails on a 2 Gigahertz Pentium IV laptop in around .14 second per image.

The top plots of Figure 4 show the MSEs for the new thumbnails, comparing the thumbnail of each distorted input image, $\boldsymbol{d_{ij}}$, with the new thumbnail for the image without blur and noise, $\boldsymbol{c}$. These plots show that the new thumbnails discriminate much better than the plots for the standard thumbnails shown in the bottom of the figure. The slight dip observed in the plots at high blur levels show that the blur estimation is somewhat sensitive to noise. The MSEs, however, show that significant blur is still present in the thumbnails. Noise resistant blur estimation [6] may provide improvements to the plots. On the other hand, careful visual study of the interaction of blur and noise may show that for noisy images, correct blur estimation is not critical to image quality determination.

5. SUBJECTIVE EVALUATION AND CONCLUSION

Figures 1 and 2 show thumbnail comparisons and originals for three examples that are best viewed in the original PDF document. The top image shows an example where the new thumbnails and standard thumbnails are indistinguishible for a good quality image. In the middle image, the hands, yellow flowers and red butterfly in the middle image are misfocused, as is seen in the new thumbnail, but not the standard thumbnail. The bottom image is noisy, as seen in the new thumbnail, but not the standard thumbnail. The originals shown in Figure 2 are cropped to save space while showing the thumbnails and originals at the correct relative scales.

A subjective comparison of the thumbnails using four input image categories, *Blurry*, *Clean*, *Noisy* and *Textured*, and twenty judges is reported in an externally accessible report [10]. The judges were asked to determine which thumbnail version of a pair best represented the original full resolution image.

The results of the evaluation, even without algorithm parameter tuning, are encouraging. The study found that the new thumbnails are more representative of their originals for *blurry* images. In addition, there are no significant differences between the results of the new thumbnails and the standard thumbnails for *clean* images. The noise component improves the results for *noisy* images, but degrades the results for *textured* images. The blur component of the new thumbnails may always be used. The decision to use the noise component of the new thumbnails, $\boldsymbol{n}_t$ in Equation 5, should be based on testing with the particular image mix expected for the application. Future work may develop a noise component that better separates between noise and texture, allowing the noise component to always be used without degrading textured images.

Thumbnail generation, with its less exacting requirements, may be well suited for application of techniques developed for more traditional deblurring and denoising applications.

6. REFERENCES

[1] Yan Ke, Xiaoou Tang, and Feng Jing, "The design of high-level features for photo quality assessment," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, vol. 1, pp. 419–426.

[2] Anil K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, Inc., 1989.

[3] Suk Hwan Lim, "Characterization of noise in digital photographs for image processing," in *Proceedings of SPIE*, January 2006, vol. 6069.

[4] Alan S. Wilsky, "Multiresolution markov models for signal and image processing," *Proceedings of the IEEE*, vol. 90, no. 8, pp. 1396–1458, August 2002.

[5] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing magazine*, pp. 43–64, May 1996.

[6] L.J. Ferzli, R. Karam, "No-reference objective wavelet based noise immune image sharpness metric," in *IEEE ICIP 2005 Proceedings, Vol. 1*, Sept. 2005, pp. 405–408.

[7] Javier Portilla and Rafael Navarro, "Efficient method for space-variant low-pass filtering," in *VII National Symposium on Pattern Recognition and Image Analysis*, Barcelona, Spain, 1997, vol. 1, pp. 287–292.

[8] David L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. on Inf. Theory*, pp. 613–627, Dec. 1995.

[9] Robert L. Cook, "Stochastic sampling in computer graphics," *ACM Transactions on Graphics*, pp. 51–72, Jan. 1986.

[10] R. Samadani, T. Mauer, D. Berfanger, J. Clark, S Lim, and D. Tretter, "Honest image thumbnails: Algorithm and subjective evaluation," Tech. Rep. HPL-2007-88, HP Labs, June 2007.