Sharpness Estimation for Document and Scene Images

Jayant Kumar
University of Maryland
College Park, USA
jayant@umiacs.umd.edu

Francine Chen
FX Palo Alto Laboratory
Palo Alto, USA
chen@fxpal.com

David Doermann
University of Maryland
College Park, USA
doermann@umiacs.umd.edu

Abstract

Images of document pages have different characteristics than images of natural scenes, and so the sharpness measures developed for natural scene images do not necessarily extend to document images primarily composed of text. We present an efficient and simple method for effectively estimating the sharpness/blurriness of document images that also performs well on natural scenes. Our method can be used to predict the sharpness in scenarios where images are blurred due to camera-motion (or hand-shake), defocus, or inherent properties of the imaging system. The proposed method outperforms the perceptually-based, no-reference sharpness work of [1] and [4], which was shown to perform better than 14 other no-reference sharpness measures on the LIVE dataset.

1. Introduction

As the number of digital photos increases, automatic assessment of the perceived quality becomes more important for search and automatic selection, as well as for feedback during capture and possible automated enhancement. For image search the sharpness (or measure of best focus) can be an important attribute in ranking the retrieved results. In automatic creation of photobooks, sharpness can be used to select the least blurry of several similar images; and the size of images can be scaled so that blurrier images are smaller.

With increasing quality of cameras on mobile devices, taking photos of document pages as an alternative to *scanning* is becoming more feasible. However, estimating the sharpness of photos of text is not well addressed by current sharpness measures. From the user perspective, it is often difficult to determine whether a photo is focused on a small mobile screen, so a real-time method for estimating sharpness directly over the entire field of view would be useful.

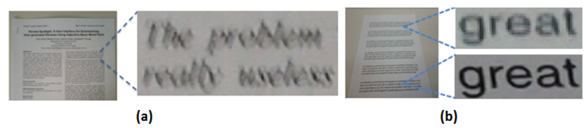

Figure 1. Motion and out-of-focus blur

When capturing an image, there are different causes of blur. Figure 1(a) illustrates blur due to the motion of camera relative to the object. When objects in a scene are at different distances from the camera lens, some parts of the image may be blurrier than others. This is especially evident when taking close-ups, as shown in Figure 1(b). Small, high-resolution cameras in smartphones are more susceptible to these distortions due to their light-weight and single-hand usage, which make them difficult to hold sufficiently steady.

Most sharpness measures have been developed and tested on images of natural scenes, and the best performing no-reference measures are based on edge-width [1]. However, these measures are somewhat coarse for images of text. Text is usually composed of sharp edges with high contrast, and many of the edges span only few (≤ 3) pixels. For smaller fonts, the text pixels may be so close together, such as between the vertical strokes in the letter 'm', that the true background is not evident. Thus, sharpness measures based on counting pixels to compute edge-width may incorrectly estimate the sharpness in these cases.

We propose a local grayscale-based method for effectively estimating the sharpness of images composed primarily of text, which also performs well on natural scenes. Like previous perceptually-based approaches [1, 3, 4], our method is based on estimating the sharpness of edges, but instead of computing edge-width, we estimate the changes in grayscale (luminance) values that are observed at an edge. We use this estimate to classify an edge-pixel as *sharp* or *blurred*. Others have noted that only the horizontal direction is adequate for computing sharpness [1, 3], but we observed bet-

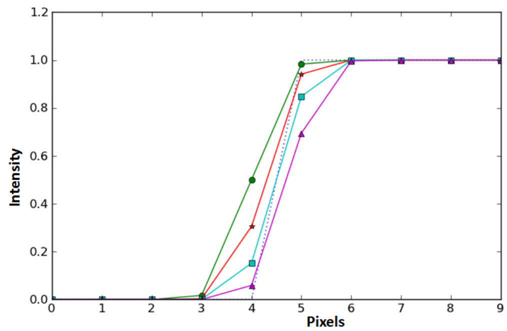

Figure 2. Blur width measured in pixels.

ter performance for document images by combining estimates in both the horizontal and vertical directions. We attribute this to the frequent horizontal and vertical strokes in text. Our experiments with a corpus of document images show that the performance of our method is better than earlier perceptually-based models. Additionally, our method is significantly faster than previous methods, which is important if the method is to run on a mobile phone.

2. Sharpness Estimation

Figure 2 illustrates that measuring edge-width by counting pixels is only a rough estimate of sharpness. We simulate blurring in a row of pixels in a captured digital image using a step function convolved with a Gaussian window of length 5. To model the digitization of an analog signal, the values in windows of size four are averaged to produce the values shown in the figure. The step function is shown as a dotted line and the four solid lines correspond to the results when the Gaussian window is shifted with offsets of 0, 1, 2, and 3. Note that the edges are two (red, turquoise, purple) or three (green) pixels in width, and that for a given width, the sharpness of the edge varies. Thus the edge width may have a different value depending on the location of the underlying edge during digitization, and using edge width as a measure of sharpness is only a rough estimate.

2.1 △DoM

From Figure 2, we note that blurred edges are somewhat complicated. We assume an underlying edge is non-linear and can be modeled by a non-parametric model, such as the 'difference of differences' that can model changes in the direction of a line. From this model we derive a measure that captures whether the slope changes quickly, a characteristic of sharp edges. We propose to use *difference of differences* in grayscale values of a median-filtered image (△DoM) as an indicator of edge sharpness. Median filtering is used to smooth variations due to noise while preserving edges. We compute △DoM separately for horizontal and vertical directions. In the x (horizontal) direction, △DoM is:

Figure 3. Illustration of proposed △DoM.

$$\triangle DoM_x(i,j) = [I_m(i+2,j) - I_m(i,j)] - [I_m(i,j) - I_m(i-2,j)] \quad (1)$$

where $I_m(i,j)$ is the grayscale value of a pixel located at (i,j) in an image that has been median-filtered. We use differences with an offset of two which is more robust to intensity variations. The change in slope, $\triangle DoM$, will be greater for sharp edges, because sharp edges transition to and from a larger slope more quickly. The absolute value of the difference of differences is taken, since we only care whether the slope changes quickly.

To relate the *difference of differences* to edge-width, we note that edge-width is inversely proportional to slope. The change in slope, $\triangle DoM$, is a discrete version of the second derivative. Thus, by integrating the second derivative, or for our discrete model, summing $\triangle DoM$ over a window of size 2w+1, we have a measure that is inversely proportional to edge-width $w(e_i)$:

$$\sum_{i-w \le k \le i+w} |\triangle DoM_x(k,j)| \propto \frac{1}{w(e_i)} \quad (2)$$

Since the width of noticeable blur decreases as contrast increases [1], we need to normalize the quantities by the contrast at the edge. The contrast is estimated using the same window of size $2w+1$ as used for △DoM, around each identified edge at pixel (i,j):

$$C = \sum_{i-w \le k \le i+w} |I(k,j) - I(k-1,j)| \quad (3)$$

where $I(k,j)$ is the value of a pixel located at (k,j) in the image.

2.2 Estimating Overall Sharpness

Sharpness estimation is performed separately in the x and y directions. To identify edges quickly, the image is first smoothed with a 1-D filter ($[\frac{1}{2}\, 0\, \frac{-1}{2}]$) in either the x or y direction. The location of absolute values greater than a threshold are chosen to identify probable edge pixels. We use a threshold of 0.0001 after normalizing by the maximum value in the filtered image. The normalized estimate at each edge-pixel e_i is computed, and the pixel is classified as *sharp* if it is greater than a pre-defined threshold T.

The sharpness $S_x(i,j)$ in the x-direction at a pixel located at (i,j) in an image is computed as the ratio of Equations 2 and 3:

$$S_x(i,j) = \frac{\sum_{i-w \le k \le i+w} |\triangle DoM_x(k,j)|}{\sum_{i-w \le k \le i+w} |I(k,j) - I(k-1,j)|} \quad (4)$$

Figure 3 illustrates the value of $S_x(i,j)$ at a fixed location in a reference image and one of its blurred version from LIVE dataset. The ratio exhibits a high value at sharp locations and low values at blurred locations. A similar computation is done to obtain $S_y(i,j)$ in the y-direction. In order to compute the sharpness of an image, the sharpness estimates at each edge-pixel need to be combined. The sharpness in one direction (i.e., x or y) for either a region or the entire page is computed as:

$$R_x = \frac{\#sharpPixels_x}{\#edgePixels_x},\ R_y = \frac{\#sharpPixels_y}{\#edgePixels_y} \quad (5)$$

The sharpness in the x and y directions are combined using a $Frobenius$-norm to obtain the sharpness for image:

$$S_I = \sqrt{R_x^2 + R_y^2} \quad (6)$$

The maximum value of S_I is $\sqrt{2}$ when all the edge-pixels in both directions are detected as sharp.

3. Experimental Results

Datasets and Evaluation: To test how our proposed method performs on scene images as well as document page images, evaluations were performed using two freely available image quality datasets, LIVE [5] and CSIQ [2], and a dataset of document page images (DocSharp). We evaluated on the Gaussian-blurred subset of LIVE (174 images) and CSIQ (150 images) datasets. We used the *MOS* (Mean Opinion Score) provided for the LIVE dataset by [1] for each image and the *DMOS* (Difference in Mean Opinion Score) provided by the CSIQ dataset for each image. The MOS and DMOS scores indicate the average of user estimates of image blurriness.

Table 1. Correlation and monotonicity

	JNB	CPBD	Q	△DoM
Spearman				
DocSharp	0.43	0.31	0.65	0.63
LIVE	0.84	0.94	0.59	0.89
CSIQ	0.77	0.89	0.72	0.84
Pearson				
LIVE	0.83	0.91	0.64	0.87
CSIQ	0.81	0.88	0.77	0.86

To construct the *DocSharp* corpus, we asked smartphone users to submit five different shots of a document page. The page images were taken from magazines and technical articles varying in viewing angle and distance from the document. A total of 27×5 = 135 images, primarily from *iPhone* and *Android*, were collected. We asked the workers at Amazon's *Mechanical Turk* to identify the sharper (most in focus) image of a pair of images. We used standard controls to gauge accuracy and obtained $21-25$ judgements per image-pair with an average agreement of 85.7% among workers.

We evaluated and compared our method against the top-performing (perceptually-based) methods of *JNB*[1], *CPBD*[4] and against the gradient-based method of *Q*[6]. We used MATLAB implementation provided by each of the authors for comparing against our MATLAB implementation. Three evaluation measures were used: (1) the Spearman correlation to measure how well the rank assigned by the sharpness measures correlate with the ranked MOS/DMOS values (2) the Pearson correlation to measure how well the sharpness measures correlate to MOS/DMOS values and (3) average accuracy in predicting the sharper image of a pair of images. In evaluating the pair-wise accuracy of △DoM, cross-validation was used where the images of one page were held out for testing and the parameter values w and T were set using the rest of the images. The most frequent optimized values for w and T were used to compute the scores for all images in a dataset when evaluating correlation performance.

Results and Discussions: Table 1 shows the results of Spearman rank order correlation and Pearson correlation for the three datasets. A higher correlation value represents the method's ability to score/rank images across *different* pages or scenes. A good correlation score is needed for applications such as determining whether a captured image is sharp enough to keep or should be retaken, or which photos should be reduced in size during automatic photobook creation because they are not sharp. The Q scores do not correlate well with

Table 2. Pair-wise Accuracy (%)

	JNB	CPBD	Q	$\triangle$DoM
DocSharp	73.0	51.1	88.1	82.6
LIVE	99.7	99.3	99.5	97.0
CSIQ	96.0	96.3	99.6	92.3

Table 3. Computation time in seconds

	JNB	CPBD	Q	$\triangle$DoM
DocSharp	33.63	55.46	12.36	3.91
LIVE	2.25	1.05	0.88	0.27
CSIQ	1.71	0.68	0.66	0.26

human judgements on scene images, indicating that the relative Q scores are highly dependent on the content of image and cannot be used across images of different scenes. The JNB and CPBD method performed poorly on the DocSharp dataset indicating that measuring edge width in pixels may be too coarse for the relatively sharp text edges.

We also compared the accuracy of the methods in predicting the sharper image of a pair. For DocSharp, the pairs were formed using different images of same physical page; and for LIVE and CSIQ, the pairs were formed using the same photo with different amounts of blur. Hence, a higher pairwise accuracy is indicative of the ability of the method to choose the best image from multiple shots of *same* page/scene. The pairwise accuracy results obtained on the three datasets are shown in Table 2. To compare the performance of Q and $\triangle$DoM in predicting sharpness, we performed a paired t-test comparing the accuracy on the 10 image-pairs for each of the 27 pages in DocSharp. At the 0.05 level of significance, the results were not significant (p-value = 0.08303). Thus, all systems performed well in predicting the sharper image for pairs of scene images, while both Q and $\triangle$DoM performed better than JNB and CPBD on the DocSharp dataset and similarly to each other. A slightly lower accuracy of $\triangle$DoM on scene images may be due to the fact that the pooling of sharpness scores across whole image is not done perceptually like JNB and CPBD. Also, the $\triangle$DoM score is more sensitive to edges with small widths as observed in text regions.

We compared the average time it took to process one image in each dataset using the different methods. A 64-bit 2.83 GHz Intel Core2 Quad Windows 7 machine with 4 GB of memory was used for all computations. Each method and dataset combination was run three times and the median of the average time (in seconds) per image is shown in Table 3. Note that $\triangle$DoM is significantly faster than Q, JNB and CPBD. Unlike the JNB and CPBD method, our method does not require computing multiple time-consuming exponentials, Canny edge-detection, or counting pixels for edge-width computation. The SVD coherence computation in Q is computationally expensive making the method slower as compared to our method.

Of the four tested sharpness measures, $\triangle$DoM was the most computationally efficient and the only method that performed well on correlation measures for both document and natural scenes, which makes it the best measure for use in document capture applications that require speed and accurate ranking, such as determining instantly whether a captured image on a mobile device is sharp enough. In contrast, JNB and CPBD performed poorly on DocSharp, Q performed poorly on LIVE and CSIQ for the correlation measures, and all three measures were noticeably slower.

4. Conclusion

In this paper, we presented a method for estimating the sharpness of images containing text which also performs well on day-to-day photos of scenes. Our method takes into account several characteristics of text which are not captured by existing methods developed for scene images. Our results on the DocSharp dataset noticeably outperformed the existing perceptual-based methods, and performed competitively on scene images. Furthermore, our sharpness measure is easy to implement and computationally efficient.

References

[1] R. Ferzli and L. Karam. A no-reference objective image sharpness metric based on the notion of just noticeable blur (jnb). *IEEE Tran. on Image Processing*, 18:717–728, 2009.

[2] E. C. Larson and D. M. Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 19(1), 2010.

[3] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi. A no-reference perceptual blur metric. In *ICIP*, volume 3, pages 57 – 60, 2002.

[4] N. Narvekar and L. Karam. A no-reference image blur metric based on the cumulative probability of blur detection (cpbd). *IEEE Tran. on Image Processing*, 20(9):2678 –2683, 2011.

[5] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. Live image quality assessment database release 2. In *http://live.ece.utexas.edu/research/quality*, 2006.

[6] X. Zhu and P. Milanfar. Automatic parameter selection for denoising algorithms using a no-reference measure of image content. *IEEE Transactions on Image Processing*, 19(12):3116 –3132, 2010.