An extensive empirical evaluation of focus measures for digital photography (extended abs.)

Hashim Mir, Peter Xu, and Peter van Beek
Cheriton School of Computer Science, University of Waterloo
{vanbeek}@uwaterloo.ca

1 Problem Statement

Automatic focusing of a digital camera in liveview mode, where the camera's display screen is used as a viewfinder, is done through contrast detection. Let $f(x, y)$ be the luminance or grayscale at pixel (x, y) in an image of size $M \times N$ pixels. In focusing using contrast detection, a focus measure is used to map an image to a value that represents the degree of focus of the image. Many focus measures have been proposed and evaluated in the literature (see, e.g,. [1–3, 5–9]). In specifying a focus measure, we define a function $g(x, y)$. The value F of the focus measure is then given by,

$$F = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} g(x, y). \tag{1}$$

For example, the squared gradient focus measure, based on the square of the horizontal first-derivative, is given by,

$$g(x, y) = (f(x + 1, y) - f(x, y))^2.$$

Previous studies on focus measures have either used a small number of benchmarks images in their evaluation, been directed at microscopy and not digital cameras, or have been based on *ad hoc* evaluation criteria. In this paper, we perform an extensive empirical evaluation of focus measures for digital photography and advocate using two standard statistical measures of performance, precision and recall, as evaluation criteria (see below for definitions).

2 Experimental Methodology

The experimental methodology for evaluating the focus measures consisted of two stages. All implementations were in C++ running under Windows 7.

In the first stage, we implemented a camera remote control application whereby a camera is tethered to a computer via a USB cable and controlled by

Table 1: Notation for defining precision and recall.

	ground truth	
predicted	*tp* (true positive)	*fp* (false positive)
	fn (false negative)	*tn* (true negative)

the software running on the computer. Our remote control application makes use of the Canon SDK (Version 2.11) and can control and replicate the basic functionality of the camera such as setting the aperture, displaying the liveview stream, and controlling the focus position of the lens. Using the remote control software, we gathered 18 sets of benchmark images that covered a range of common photography settings including landscapes, closeups, interiors, still lifes, and so on. Each of the sets of benchmark images contains either 167 (when using a 50 mm lens) or 231 (when using a 200 mm lens) jpeg images, one for each focus position of the lens. The jpeg images are captured from the liveview stream once the lens is moved from one position to the next. The camera used in our experiments was a Canon EOS 550D/Rebel T2i.

In the second stage, we implemented more than 30 focus measures that we found in an extensive survey of the literature. Given a benchmark set of images and a focus measure, the focus measure was applied to each jpeg image in the benchmark (see Figure 1 for an example). The peaks in the graph of a focus measure correspond to the lens position (or image) where the focus measure predicts the corresponding object will be in maximum focus. The predictions of the focus measure were then compared to the ground truth, where the ground truth is the true lens position (or image) where the corresponding object was in maximum focus. The ground truth was determined by one of the authors viewing the images but there was little or no subjectivity involved as the image in sharpest focus was generally clear. In some cases two images were considered equally in focus and in those cases we considered both as true or correct. Given the predictions of each focus measure and the ground truth, we then calculated the precision and recall of each focus measure.

Consider the notation shown in Table 1. For example, *tp* means that the image corresponding to a peak in the focus measure was in maximal focus (i.e., the image that the focus measure predicted was in maximal focus was correct) and *fn* means that an image that was in maximal focus did not correspond to a peak in the focus measure (i.e., the focus measure failed to predict that the image was in maximal focus). Given this notation, the standard definitions of precision and recall can be stated,

$$\text{precision} = \frac{tp}{tp + fp}, \qquad \text{recall} = \frac{tp}{tp + fn}.$$

We also calculated the mean absolute prediction error, $|x_t - \hat{x}_t|$, where x_t

is the lens position of the actual image that is in maximal focus and $\hat{x}_t$ is the lens position of the image that a focus measure predicts is in maximal focus. The mean absolute prediction error was chosen over the mean squared error as the former prefers prediction errors that are occasionally large but small on average, while the latter prefers prediction errors that avoid large errors while still possibly being quite unsatisfactory overall. In other words, if focus fails, it should fail big. There is little that is more frustrating than to return from a shoot to discover that, while an image looked sharp in preview mode on the camera, one has *just* failed to capture a subject, such as the eyes in a portrait shot, in sharp focus.

3 Experimental Results and Discussion

Table 2 shows a summary of the results of our experiments. Several focus measures standout from the rest. In terms of precision, the squared gradient and the similar Brenner measure are perfectly precise. In terms of recall, the squared gradient and Brenner still perform well, but the Laplacian of Gaussian and the 5×5 Laplacian operator have perfect recall. Our results contrast with previous experimental studies.

In the full version of the paper, we will include:

- concise descriptions or definitions of the focus measures evaluated in our experiments,
- a thorough review of previous empirical studies,
- a detailed comparison of our results contrasted with those in previous empirical studies, and
- the results on more benchmarks to ensure statistical significance.

In current work we are using machine learning to automatically learn a focusing algorithm that uses the best focus measure to control the camera and achieve focus. We have achieved preliminary results that are promising in comparison to, for example, the hand-crafted rule-based algorithm proposed by Kehtarnavaz and Oh [4]. We hope to include these results in the final version of the paper.

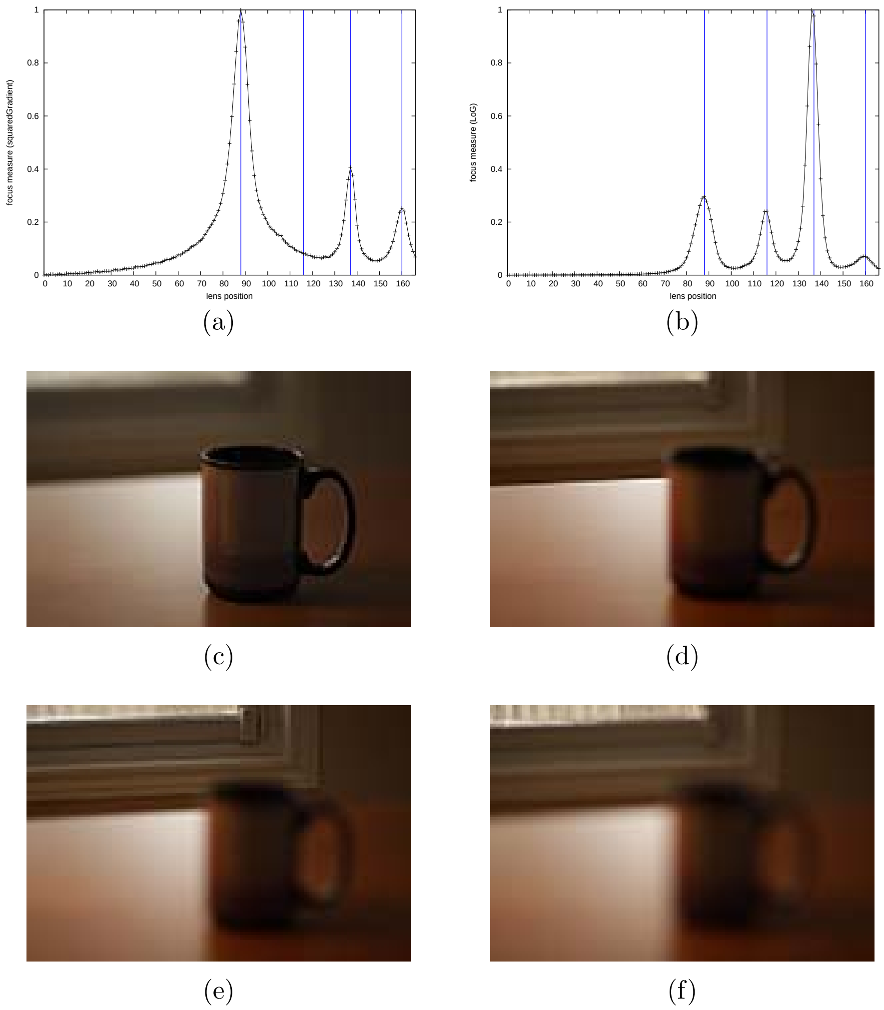

Figure 1: Focus measures of images at each of the 167 lens positions (50 mm lens) for an example scene using (a) the squared gradient focus measure, and (b) the Laplacian of Gaussian focus measure. The four vertical bars refer to the four images that have objects that are in maximal focus: (c) cup, (d) edge of desk, (e) window sill, (f) and fence in focus. On this benchmark, the squared gradient has precision 3/3 and recall 3/4, and the Laplacian of Gaussian has precision 4/4 and recall 4/4.

Table 2: Precision, recall, and mean absolute prediction error of various focus measures on the benchmarks.

type	focus measure	precision	recall	abs error
first derivative	squaredGradient	100.00	98.61	0.00
	Brenner	100.00	98.61	0.00
	firstDerivGaussian	98.61	95.83	0.01
	Sobel3x3	97.22	97.22	0.03
	Scharr3x3	97.22	97.22	0.03
	Roberts3x3	97.22	97.22	0.03
	Prewitt3x3	97.22	97.22	0.03
	firstorder3x3	97.22	97.22	0.03
	Sobel5x5	87.50	87.50	0.12
	thresholdGradient	73.99	86.11	16.45
second derivative	LoG	97.22	100.00	0.44
	Sobel5x5so	97.22	98.61	0.44
	Laplacian5x5	95.37	100.00	0.64
	Sobel5x5soCross	92.13	97.22	0.86
	Sobel3x3soCross	91.20	97.22	1.58
	Sobel3x3so	89.44	98.61	5.20
	Laplacian3x3	82.87	90.28	1.74
image histogram	entropyHistogram	14.10	29.17	51.56
	mgHistogram	7.59	15.28	58.04
	mmHistogram	0.93	2.78	34.96
	rangeHistogram	nan	20.83	nan
image statistics	variance	nan	27.78	nan
	normalized variance	nan	50.00	nan
	power	0.00	0.00	39.90
	threshold cont.	nan	6.94	nan
	num. pixels	nan	12.50	nan
correlation	autoCorrelation	79.26	98.61	25.98
	Vollath4	63.16	86.11	22.58
	Vollath5	nan	27.78	nan
compression	bzip2	60.62	70.83	30.13
	gzip	42.56	51.39	27.79
	jpeg	65.10	70.59	2.43

References

[1] L. Firestone, K. Cook, K. Culp, N. Talsania, and K. Preston. Comparison of autofocus methods for automated microscopy. *Cytometry*, 12:195–206, 1991.

[2] F.C.A. Groen, I.T. Young, and G. Ligthart. A comparison of different focus functions for use in autofocus algorithms. *Cytometry*, 6:81–91, 1985.

[3] J. Jeon, J. Lee, and J. Paik. Robust focus measure for unsupervised auto-focusing based on optimum discrete cosine transform coefficients. *IEEE Transactions on Consumer Electronics*, 57:1–5, 2011.

[4] N. Kehtarnavaz and H.-J. Oh. Development and real-time implementation of a rule-based auto-focus algorithm. *Real-Time Imaging*, 9:197–203, 2003.

[5] A. Santos, C. Ortiz de Solórzano, J. J. Vaquero, J. M. Peña, N. Malpica, and F. del Pozo. Evaluation of autofocus functions in molecular cytogenetic analysis. *Journal of Microscopy*, 188:264–272, 1997.

[6] M. Subbarao and J.-K. Tyan. Selecting the optimal focus measure for auto-focusing and depth-from-focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:864–870, 1998.

[7] D. C. Tsai and H. H. Chen. Reciprocal focus profile. *IEEE Transactions on Image Processing*, 21:459–468, 2012.

[8] G. Yang and B.J. Nelson. Wavelet-based autofocusing and unsupervised segmentation of microscopic images. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 3, pages 2143–2148, 2003.

[9] S. Yousefi, M. Rahman, and N. Kehtarnavaz. A new auto-focus sharpness function for digital and smart-phone cameras. *IEEE Transactions on Consumer Electronics*, 57:1003–1009, 2011.