

Optimal image acquisition by auto-defocusing

Tao Ma and Stanley J. Reeves

Auburn University, 200 Broun Hall, Auburn, AL 36849, USA

ABSTRACT

This paper presents an optimal image acquisition methodology by replacing the traditional birefringent filter with slight out-of-focus blur generated by the camera lens. Since many cameras already have adjustable lenses and auto-focus systems, our method can exploit existing hardware by simply changing the focusing strategy. During the image acquisition, the optimal defocus setting is automatically adapted to the power spectrum of the scene which is evaluated by a generic autocorrelation model. A criterion to estimate reconstruction errors without the baseband knowledge of the scene is developed in the paper. This metric helps the camera to choose the optimal focus settings. An optimal Wiener filter then recovers the captured scene and yields sharper images with reduced aliasing. The numerical and visual results show that our method is superior to current acquisition methods used by most digital cameras.

Keywords: auto-focus, anti-aliasing, defocus, image acquisition

1. INTRODUCTION

A digital camera acquires images by spatially sampling continuous scenes at each pixel location of an optical sensor, such as a CCD or CMOS sensor. This sampling produces aliasing artifacts in the spatial domain during the reconstruction process, where Moiré patterns and other artifacts can be observed. Most digital cameras apply an optical anti-aliasing filter on top of the sensor using a birefringent crystal to limit these artifacts. Such a filter has a lowpass characteristic and reduces the energy of the input scene at frequencies higher than the Nyquist baseband of the optical sensors.

One commonly used anti-aliasing filter is the four-spot birefringent filter.¹ By carefully designing the thickness of the crystal plate, the input light beam can be separated into four beams then detected by four neighboring photosites of the sensor. However, the frequency response of this filter is not an ideal lowpass filter but a two-dimensional sinc function, such as in Figure 1(a). The wide transition band and the large sidelobes in the stopband of this filter make it sub-optimal. The distortion in the baseband signal results in blurry images and makes them less appealing. Although a follow-up image-enhancement step can reduce the baseband distortion, it is impossible to remove the existing aliasing artifacts without some information from the original scene. Furthermore, the thickness of the birefringent crystal is fixed by the size of the sensor cells,¹ which makes it difficult to design a compact camera.

In this paper, we consider an alternative acquisition approach without using the birefringent filter. Because of the lowpass characteristics of out-of-focus blur, one can replace the traditional anti-aliasing filter with slight out-of-focus blur generated by the camera lens. A blurred image is captured during the image acquisition with the optimal focus setting found by the camera. A Wiener filter corresponding to the optimal focus setting then recovers the captured image. Our method seeks to balance the baseband distortion and noise amplification error and the error due to aliasing artifacts during the acquisition. One equivalent anti-aliasing filter designed by our approach is shown in Figure 1(b). It is easy to see that it has a narrower transition band and the responses in the baseband and stopband are close to ideal.

This paper is organized as follows. The defocus acquisition approach is presented in Section 2. The optimal acquisition Wiener filter is defined in Section 3. The criterion to identify the optimal defocus setting is presented in Section 4. Simulation results are reported in Section 5. Finally, we draw some conclusions and discuss our future work in Section 6.

Further author information: (Send correspondence to Stanley J. Reeves)

Stanley J. Reeves: E-mail: reevesj@auburn.edu, Telephone: 1 334 844 1821

Tao Ma: E-mail: tzm0001@auburn.edu

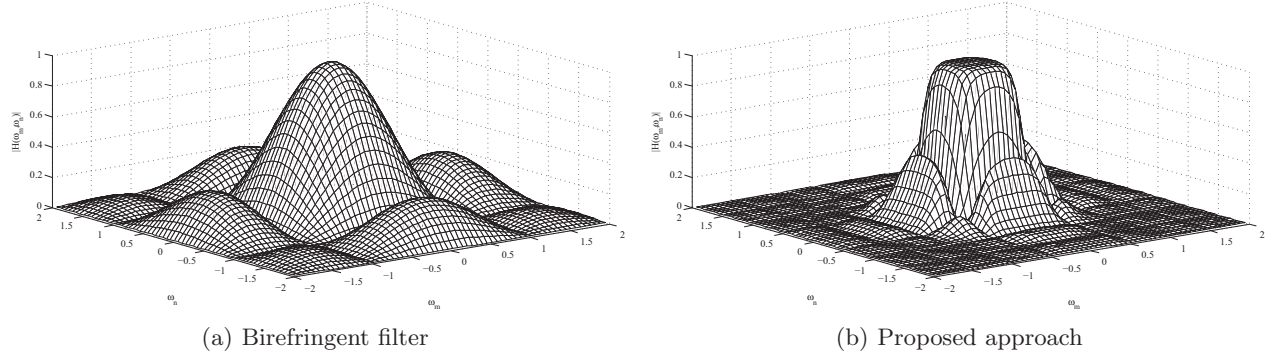


Figure 1. Comparison of the frequency responses of the birefringent anti-aliasing filter and the proposed Wiener approach. The horizontal and vertical frequencies have been normalized by the sampling frequency.

2. DEFOCUS ACQUISITION

In the frequency domain, let $F(\Omega_x, \Omega_y)$ be the spectrum of a continuous scene as viewed by the camera. If a lowpass anti-aliasing filter $H(\Omega_x, \Omega_y)$ is applied before the camera sensor, the representation of the sampled signal $F_d(\Omega_x, \Omega_y)$ is as follows,²

$$F_d(\Omega_x, \Omega_y) = \frac{1}{\Delta_x \Delta_y} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} H\left(\Omega_x - \frac{2\pi}{\Delta_x}k, \Omega_y - \frac{2\pi}{\Delta_y}l\right) F\left(\Omega_x - \frac{2\pi}{\Delta_x}k, \Omega_y - \frac{2\pi}{\Delta_y}l\right) \quad (1)$$

where Δ_x and Δ_y are sampling intervals in the horizontal and vertical directions. Let F_b be the baseband signal of the scene. Then $F_b = F$ for $|\Omega_x| < \frac{2\pi}{\Delta_x}$ and $|\Omega_y| < \frac{2\pi}{\Delta_y}$, and zero otherwise, and denote this support as I . In other word, F_b contains frequency content with frequencies smaller than the Nyquist rate. In the following context, we assume $\Delta_x = \Delta_y = 1$ for simplicity.

It is possible to recover the baseband signal F_b without distortion and aliasing using an ideal lowpass filter even if the continuous scene F is not bandlimited. For the non-ideal case, such as using a birefringent filter H_a , perfect recovery of F_b is not generally possible. Instead, we seek to reconstruct an estimate \hat{F}_b to minimize the error $\epsilon^2 = E\{\|F_b - \hat{F}_b\|^2\}$. However, the diversity of real-world scenes and the appearance of noise mean that a fixed birefringent anti-aliasing filter may not be the best choice for every scene. Consistently better results can be obtained if we adapt the response of the anti-aliasing filter according to the scene and noise level. Because of the lowpass characteristic of out-of-focus blur, a practical method is to automatically defocus a camera lens with the desired blur level to bandlimit the spectrum F instead of using a fixed birefringent filter. Non-ideal anti-aliasing filters trade off baseband distortion with aliasing. By adjusting the defocus, we seek the optimal tradeoff between these two errors.

If we take the sensor-size effect h_s into account, the corresponding baseband PSF of the anti-aliasing filter h is a convolution of h_s and out-of-focus blur h_o . Our proposed imaging model can be represented by the following equation:

$$y[m, n] = h[m, n] * f_b[m, n] + f_a[m, n] + u[m, n], \quad (2)$$

where $y[m, n]$ ($0 \leq m \leq M - 1, 0 \leq n \leq N - 1$) is the sampled image captured by the camera, $f_b[m, n]$ is a baseband image without any aliasing, $f_a[m, n]$ is the aliasing component filtered by the stopband response of the blur, and u is additive noise. Note that $f_a[m, n]$ is dependent on $h[m, n]$, but the dependence is suppressed for notational simplicity. Here, “*” represents a two-dimensional convolution. In practical imaging devices, the noise $u[m, n]$ can be modeled as a mixture of independent Gaussian noise $u_0[m, n]$ and signal-dependent noise $u_1[m, n]$,³ expressed as the following equation:

$$\begin{aligned} u[m, n] &= u_0[m, n] + u_1[m, n] \\ &= k_0 \delta[m, n] + k_1 (h[m, n] * f_b[m, n] + f_a[m, n]) \delta[m, n], \end{aligned} \quad (3)$$

where k_0 and k_1 are constants and $\delta[m, n]$ follows a standard normal distribution. The frequency representation of this imaging model can be expressed as follows,

$$Y(\omega_m, \omega_n) = H(\omega_m, \omega_n)F_b(\omega_m, \omega_n) + F_a(\omega_m, \omega_n) + U(\omega_m, \omega_n). \quad (4)$$

The frequency dependency is suppressed to simplify subsequent notation. The reconstructed image $\hat{f}_b[m, n]$ will be estimated from the acquired image $y[m, n]$. It is evident that this is an ill-posed problem because of the aliasing artifact and the appearance of noise.

3. OPTIMAL ACQUISITION WIENER FILTER

To obtain the baseband estimate \hat{F}_b , a digital filter W is designed to minimize the reconstruction error $\epsilon^2 = E\{\|F_b - \hat{F}_b\|^2\}$, where $\hat{F}_b = WY$. Because k_1 is normally small and the energy of filtered aliasing signal F_a is much lower than baseband signal F_b , we can assume the noise U and the aliasing signal F_a are uncorrelated with the baseband signal F_b . As a result, W becomes a Wiener filter (linear minimum mean-square error) with the following form:

$$W = \frac{H^*}{|H|^2 + \frac{\sigma_u^2 + S_{f_a}}{S_{f_b}}}, \quad (5)$$

where S_{f_b} is the power spectrum of the baseband signal, S_{f_a} is the power spectrum of the aliasing component and σ_u^2 is the noise variance. Here, the “*” superscript is a conjugate operation. Apparently, we treat the aliasing signal as a part of the noise in the Wiener filter, which will smooth out the aliasing artifacts in the captured images. Our goal is to find an optimal focus setting frequency response H_{opt} with corresponding Wiener filter W_{opt} that minimizes the reconstructed error ϵ^2 for each specific continuous scene with a certain noise level.

It is obvious that we have no information about the baseband signal F_b and the aliasing signal F_a . To estimate the unknown power spectra S_{f_b} and S_{f_a} , an image Y_0 is initially captured with a large amount of out-of-focus blur H_0 . We then assume that the aliasing component F_a is sufficiently suppressed in this image so that it can be ignored. That is, $Y_0 \approx H_0F_b + U$. Therefore one can derive the expected power spectrum $S_{f_{b_0}}$ from the expected power spectrum S_{y_0} of the initial image as follows,

$$\begin{aligned} S_{y_0} &\approx E\{|H_0F_b + U|^2\} \\ &= E\{|H_0F_b|^2\} + \sigma_u^2 \\ &= |H_0|^2 S_{f_{b_0}} + \sigma_u^2, \end{aligned} \quad (6)$$

which when solved yields

$$S_{f_{b_0}} \approx \frac{S_{y_0} - \sigma_u^2}{|H_0|^2}. \quad (7)$$

Using the periodogram estimate $\hat{S}_{y_0} = |Y_0|^2$, we obtain:

$$\hat{S}_{f_{b_0}} = \frac{|Y_0|^2 - \sigma_u^2}{|H_0|^2}. \quad (8)$$

The generic autocorrelation function we chose in the paper is a general decay exponential model as follows,

$$\Phi_f(x, y) = \beta e^{-\sqrt{ax^2 + by^2 + cxy}} + \bar{f}^2, \quad (9)$$

where β is a scale factor and \bar{f} is the mean of the continuous scene f . The parameters a , b and c obey the conditions $a > 0$, $b > 0$ and $|c| \leq \sqrt{ab}$. To generate the power spectra of the baseband signal and the aliasing signal using this generic model, we first obtain the observed autocorrelation function of the baseband signal by taking an inverse Fourier transform $\Phi_b[m, n] = \mathcal{F}^{-1}\{\hat{S}_{f_{b_0}}(\omega_m, \omega_n)\}$. Assume that there exists a positive integer

R which satisfies $\Omega_{f_{max}} < \min\{R\frac{\pi}{\Delta_x}, R\frac{\pi}{\Delta_y}\}$, where $\Omega_{f_{max}}$ is the highest frequency content in the continuous scene. In other words, if the Nyquist rate is increased by a factor of R , no aliasing artifacts are introduced by the sampling. Let Φ_{f_R} be the autocorrelation function of the observed image with the high sampling rate. We fit the observed autocorrelation function $\Phi_b[m, n]$ at spatial location $[mR, nR]$ to the autocorrelation function Φ_{f_R} .

To simplify the fitting process, we only select data inside a spatial support Ψ from the autocorrelation function to fit the model. The spatial region Ψ is limited to $[-q, q] \times [-q, q]$. The parameters a , b and c can be estimated using a closed-form solution.⁴ The power spectrum of the continuous scene $S_{f_R}(\omega_m, \omega_n)$ can be acquired by taking a Fourier transform of $\Phi_{f_R}[m, n]$. The power spectra of the baseband signal f_b and the aliasing component f_a can be estimated according to the ratio of the sampling rate R . More precisely, the estimated power spectrum of the baseband signal $S_{f_b}(\omega_m, \omega_n) = S_{f_R}(\omega_m, \omega_n)$ for $|\omega_m| < \frac{\pi}{R}$ and $|\omega_n| < \frac{\pi}{R}$, while the estimated power spectrum of the aliasing signal S_{f_a} is the remaining part of S_{f_R} multiplied by the squared frequency response of the out-of-focus blur, then wrapped according to the ratio R .

4. FINDING THE OPTIMAL FOCUS SETTING

As discussed in Section 2, the out-of-focus blur that minimizes the reconstruction error $\epsilon^2 = E \left\{ \|F_b - \hat{F}_b\|^2 \right\}$ is taken to be the optimal focus setting H_{opt} . The Wiener filter corresponding to H_{opt} is the optimal reconstruction Wiener filter W_{opt} . However, since the baseband signal F_b is unknown, to calculate the reconstruction error directly is impossible. Furthermore, it is impractical for a digital camera to capture a set of images with different focus settings for the same scene while tracking the reconstruction error. An alternative approach to approximate the reconstruction error ϵ is preferred.

We propose a criterion to monitor the true ϵ^2 using an approximation $\hat{\epsilon}^2$. Consider a captured image Y with out-of-focus blur H . The expected power spectrum of the estimated baseband image \hat{F}_b can be represented by:

$$S_{\hat{f}_b} = E \left\{ |\hat{F}_b|^2 \right\} = E \left\{ |WY|^2 \right\} \approx |WY|^2. \quad (10)$$

Using the imaging model formulated in (4) in equation above and assuming that the aliased component and noise are uncorrelated with the baseband signal, we can express $S_{\hat{f}_b}$ as

$$\begin{aligned} S_{\hat{f}_b} &= E \left\{ |W(HF_b + F_a + U)|^2 \right\} \\ &\approx E \left\{ |WHF_b|^2 \right\} + E \left\{ |WF_a|^2 \right\} + E \left\{ |WU|^2 \right\} \\ &= |WH|^2 S_{f_b} + E \left\{ |WF_a|^2 \right\} + E \left\{ |WU|^2 \right\} \end{aligned} \quad (11)$$

Rearranging the equation above, we have:

$$E \left\{ |WF_a|^2 \right\} + E \left\{ |WU|^2 \right\} \approx |WY|^2 - |WH|^2 S_{f_b}. \quad (12)$$

Using the imaging model defined in Section 2, the true reconstruction error ϵ^2 can be expressed as follows:

$$\begin{aligned} \epsilon^2 &= E \left\{ \|F_b - \hat{F}_b\|^2 \right\} \\ &= E \left\{ \|(1 - WH)F_b - WF_a - WU\|^2 \right\}. \end{aligned} \quad (13)$$

We must approximate the true reconstruction error ϵ^2 with computable quantities. Applying (12) and again assuming the baseband signal F_b is uncorrelated with the aliasing signal F_a and noise U , one can approximate

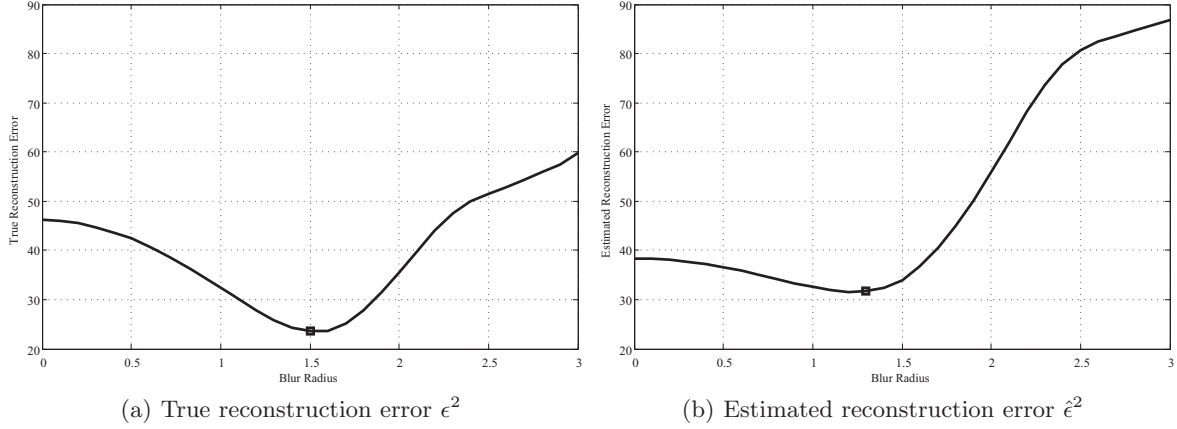


Figure 2. Comparison of the true and estimated reconstruction error of one test scene.

the above equation as follows:

$$\begin{aligned}
\epsilon^2 &\approx E \left\{ \|(1 - WH) F_b\|^2 \right\} + E \left\{ \|WF_a\|^2 \right\} + E \left\{ \|WU\|^2 \right\} \\
&\approx \sum_{\omega_m \in I} \sum_{\omega_n \in I} \left[|1 - WH|^2 S_{f_b} + |WY|^2 - |WH|^2 S_{f_b} \right] \\
&= \sum_{\omega_m \in I} \sum_{\omega_n \in I} \left[(|1 - WH|^2 - |WH|^2) S_{f_b} + |WY|^2 \right] \\
&= \sum_{\omega_m \in I} \sum_{\omega_n \in I} \left[(1 - 2\Re\{WH\}) S_{f_b} + |WY|^2 \right],
\end{aligned} \tag{14}$$

where $\Re\{\cdot\}$ denotes the real part of a complex number. Approximating both S_{f_b} and $|WY|^2$ with the estimated power spectrum $\hat{S}_{f_{b0}}$ of the initial image derived in (8), we have the estimate reconstruction error $\hat{\epsilon}^2$:

$$\hat{\epsilon}^2 \approx \sum_{\omega_m \in I} \sum_{\omega_n \in I} \left[(2 - 2\Re\{WH\}) \hat{S}_{f_{b0}} \right]. \tag{15}$$

The out-of-focus blur \hat{H}_{opt} and the corresponding Wiener filter \hat{W}_{opt} which result in the minimum $\hat{\epsilon}^2$ value are the optimal focus setting we desire. An optimal acquisition can be accomplished by capturing the scene using the out-of-focus blur \hat{H}_{opt} and recovering it by \hat{W}_{opt} . The curves in Figure 2 depict the true and estimated reconstruction errors vs. the PSF radius for one test scene. It is evident that an optimal focus setting exists and our criterion tracks it as expected.

This criterion is automatically adapted to the power spectrum of the scene, which is estimated via the initially captured image. Without mechanically moving the actual lens, a digital camera can evaluate the reconstruction error for a specific scene using a look-up table which defines the PSF of the lens as a function of radius. Because only the initial image y_0 needs to be captured to find the optimal defocus setting, this criterion is appealing for real-time applications.

5. SIMULATION AND RESULTS

In this section, we present some experimental results to verify the defocusing acquisition and evaluate the performance of the criterion we proposed in Section 4. We chose 100 gray-scale images with size 512×512 and considered them as continuous scenes. The intensity values of these natural images are between 0 and 255. Each image was subsampled by a factor of R in both directions to model the sampling processing. The resulting image acquired by the proposed approach was evaluated by comparing it to the baseband image using mean square error (MSE).

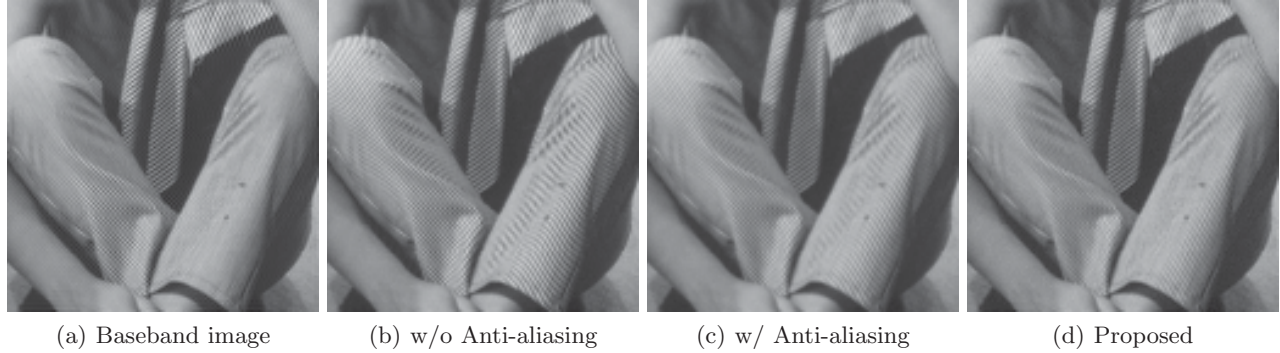


Figure 3. The cropped *Barbara* images acquired by different methods. The additive noise parameters are $(k_0, k_1) = (2, 0.1)$ and the downsampling rate $R = 2$.

In particular, we conducted simulation experiments as follows. The downsampling rate R was set to 2 and 4. A 2×2 averaging filter was added in the process as the sensor-size effect for the $R = 2$ case. A 4×4 averaging filter was chosen for the same purpose in the other case. The same filter as the sensor-size effect was used to simulate the four-spot birefringent filter in different downsampling cases. The baseband image for each scene was calculated using an ideal lowpass filter. Since the PSF of out-of-focus blur is close to circular,⁵ we implemented defocusing using a circular blur with different radii r . The spatial region Ψ defined in Section 3 was set to $[-4, 4] \times [-4, 4]$ to match the autocorrelation model. The initial image y_0 was acquired using blur with $r = 1.5$ pixels. The blur radius at the minimum point of $\hat{\epsilon}^2$ is the estimated optimal blur radius \hat{r}_{opt} .

The cropped *Barbara* images and a portion of one test image shown in Figure 3 and 4 depict different acquisition methods. Both input images were downsampled by 2 and corrupted by signal-dependent noise with parameters $(k_0, k_1) = (2, 0.1)$. It is obvious that visual results are also noticeably improved compared to the image acquired without any anti-aliasing filter, especially in regions with high-frequency patterns, such as the pants in the *Barbara* image. Although the aliasing artifacts in the image acquired by the four-spot birefringent filter have been largely reduced, the edges are blurrier than the image acquired by the proposed method. Figure 5 shows that similar performance also can be achieved when using a higher downsampling rate, such as $R = 4$.

Table 1 reports the performance of different acquisition approaches for different noise levels with $k_0 \in \{1, 2, 4, 8\}$ and $k_1 \in \{0, 0.2\}$. The average MSE over 100 test scenes shows that the proposed acquisition method outperforms traditional methods (with or without an anti-aliasing filter) at different noise levels. Images acquired without an anti-aliasing filter have lower MSE values than images acquired with an anti-aliasing filter. The reason is that the distortion produced by the anti-aliasing filter is increased more than the aliasing artifact is reduced in terms of MSE.

To verify the accuracy of the criterion derived in Section 4, we compared the optimal radius \hat{r}_{opt} estimated using this criterion with the true r_{opt} for different scenes and noise levels. The mean absolute difference $|r_{opt} - \hat{r}_{opt}|$

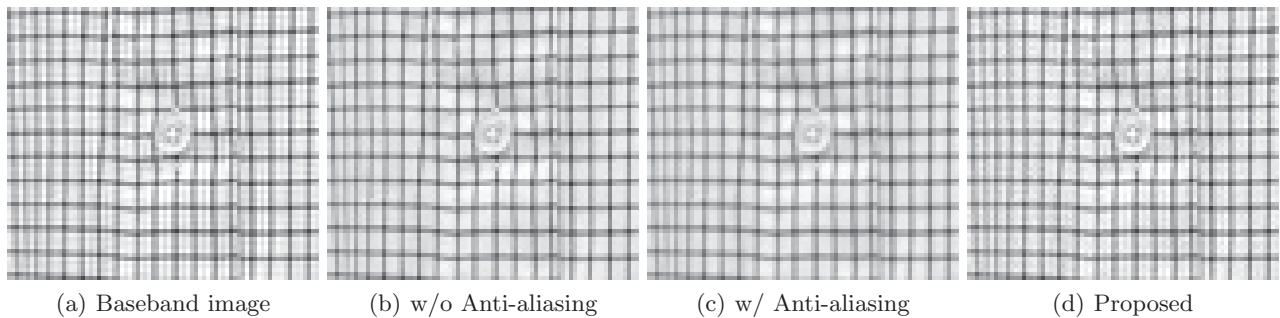


Figure 4. Portion of an image captured by different methods. The additive noise parameters are $(k_0, k_1) = (2, 0.1)$ and the downsampling rate $R = 2$.



Figure 5. The comparison of captured *Babara* image with downsampling rate $R = 4$. The additive noise parameters are $(k_0, k_1) = (2, 0.1)$.

Table 1. Image acquisition methods compared using average MSE over 100 test scenes with different noise levels. The downsampling rate are $R = 2, 4$ and the additive noise settings are $(k_0, k_1) = (8, 0), (8, 0.2), (4, 0), (4, 0.2), (2, 0), (2, 0.2), (1, 0), (1, 0.2)$.

Sampling rate	Method	(8, 0)	(8, 0.2)	(4, 0)	(4, 0.2)	(2, 0)	(2, 0.2)	(1, 0)	(1, 0.2)
$R = 2$	w/o Anti-aliasing	100.8	102.2	52.7	53.5	40.7	41.1	37.7	37.9
	w/ Anti-aliasing	103.4	105.0	55.4	56.2	43.4	43.8	40.4	40.6
	Proposed	74.9	75.6	39.2	40.1	22.8	23.6	14.7	15.5
$R = 4$	w/o Anti-aliasing	102.7	104.3	54.6	55.4	42.6	43.0	42.6	39.8
	w/ Anti-aliasing	109.7	111.4	61.8	62.5	49.8	50.2	48.6	47.0
	Proposed	83.8	84.9	44.1	45.0	25.6	26.6	16.9	17.3

in Table 2 for the downsampling-by-2 experiment is less than 0.35 pixels and the corresponding MSE loss Δ_{MSE} is less than 1.7, which demonstrate that the algorithm we have proposed is robust with respect to various noise levels. When a higher downsampling rate is used ($R = 4$), the differences $|r_{\text{opt}} - \hat{r}_{\text{opt}}|$ increase when $(k_0, k_1) = (4, 0)$ and $(4, 0.2)$. But the MSE losses are still small compared to the MSE improvement in Table 1.

6. CONCLUSION

In this paper, an optimal image acquisition approach is proposed. A new imaging model of defocusing acquisition is defined which considers both aliasing and noise. A criterion to estimate the optimal focus setting for a specific scene is derived. The model-based Wiener filter is used to minimize the end-to-end reconstruction errors during the acquisition. Both the Wiener filter and the criterion are adapted to the power spectrum of the input scene. Numerical and visual results show that the proposed approach outperforms traditional acquisition methods with or without a fixed anti-aliasing filter. The proposed approach is robust with respect to various noise levels, and it is practical in some cases to replace the commonly used four-spot birefringent filter by with this method. Our future work will focus on color image defocusing acquisition, especially for single-sensor cameras.

Table 2. Evaluations of the criteria in Section 4 over 100 test scenes with different downsampling rate $R = 2, 4$ and noise levels $(k_0, k_1) = (8, 0), (8, 0.2), (4, 0), (4, 0.2), (2, 0), (2, 0.2), (1, 0), (1, 0.2)$. The unit of mean absolute difference $|r_{\text{opt}} - \hat{r}_{\text{opt}}|$ is pixels.

Sampling rate	Metric	(8, 0)	(8, 0.2)	(4, 0)	(4, 0.2)	(2, 0)	(2, 0.2)	(1, 0)	(1, 0.2)
$R = 2$	$ r_{\text{opt}} - \hat{r}_{\text{opt}} $	0.15	0.14	0.35	0.33	0.29	0.30	0.18	0.19
	Δ_{MSE}	1.47	1.34	1.67	1.67	1.23	1.21	0.69	0.78
$R = 4$	$ r_{\text{opt}} - \hat{r}_{\text{opt}} $	0.15	0.15	0.44	0.44	0.31	0.31	0.18	0.19
	Δ_{MSE}	0.95	0.92	2.31	2.31	1.86	1.91	1.22	1.16

REFERENCES

- [1] Kessler, D., Nutt, A. C. G., and Palum, R. J., "Anti-Aliasing Low-Pass Blur Filter for Reducing Artifacts in Imaging Apparatus," (2005).
- [2] Oppenheim, A. V., Willsky, A. S., and Nawab, S. H., [*Signals and Systems (2nd Edition)*], Prentice Hall (1997).
- [3] Hirakawa, K. and Parks, T., "Image Denoising Using Total Least Squares," *IEEE Transactions on Image Processing* **15**, 2730–2742 (Sept. 2006).
- [4] Shi, J. and Reichenbach, S., "Spatially Constrained Wiener Filter with Markov Autocorrelation Modeling for Image Resolution Enhancement," in [*Image Processing, 2006 IEEE International Conference on*], 2681–2684 (2006).
- [5] Sakano, M., Suetake, N., and Uchino, E., "A Robust Point Spread Function Estimation for Out-of-Focus Blurred and Noisy Images Based on a Distribution of Gradient Vectors on the Polar Plane," *Optical Review* **14**, 297–303 (2007).