

Influence of Signal-to-Noise Ratio and Point Spread Function on Limits of Super-Resolution

Tuan Q. Pham¹, Lucas J. van Vliet¹, Klammer Schutte²

¹Quantitative Imaging Group, Delft University of Technology, Lorentzweg 1, 2628 CJ, Delft, The Netherlands

²TNO Physics and Electronics Laboratory, P.O. Box 96864, 2509 JG, The Hague, The Netherlands

ABSTRACT

This paper presents a method to predict the limit of possible resolution enhancement given a sequence of low-resolution images. Three important parameters influence the outcome of this limit: the total Point Spread Function (PSF), the Signal-to-Noise Ratio (SNR) and the number of input images. Although a large number of input images captured by a system with a narrow PSF and a high SNR are desirable, these conditions are often not achievable simultaneously. To improve the SNR, cameras are designed with near optimal quantum efficiency and maximum fill-factor. However, the latter widens the system PSF, which puts more weight on the deblurring part of a super-resolution (SR) reconstruction algorithm. This paper analyzes the contribution of each input parameters to the SR reconstruction and predicts the best attainable SR factor for given a camera setting. The predicted SR factor agrees well with an edge sharpness measure computed from the reconstructed SR images. A sufficient number of randomly positioned input images to achieve this limit for a given scene can also be derived assuming Gaussian noise and registration errors.

Keywords: super-resolution limit, fill-factor, under-sampling, deconvolution limit, registration error.

1. INTRODUCTION

Super-resolution (SR) is a technique that uses multiple low-resolution (LR) images to produce images of a higher resolution (HR). Since its conception 20 years ago,²⁴ the topic has received a huge public interest due to its potential to increase performance of existing camera systems without the need for dedicated hardware.² Numerous algorithms have been proposed.^{3, 6, 10} For optical systems, the reported performance of SR in practical situations is rather limited, reaching four-times resolution enhancement at its best. Strange enough, recently reports on resolution improvements are typically worse than what have been reported in the past (a SR factor of 1.6 in 2004¹¹ compared to a factor of 4 in 2000¹²). This does not imply a poor research progress but rather a new challenge due to advances in sensor development that have caused significant changes in LR data characteristics. In particular, a smaller pixel size with a higher fill-factor results in a lower Signal-to-Noise Ratio (SNR) and a wider Point Spread Function (PSF) with respect to sampling pitch.

The SNR and PSF characteristics have been influenced mainly by recent advancement in micro-electronics. In the past, there were fewer photodiodes on the same surface area than what is currently allowed by state-of-the-art technology. Each photosensitive element was bigger, allowing better SNR images to be captured. There was also ample spacing between adjacent photodiodes due to electronics or charge barriers. As a result, the ratio of photosensitive area over total area of the FPA (Focal Plane Array), also known as the fill-factor, was quite small. Four-times super-resolution could easily be achieved by interleaving LR images on the same HR grid without the need for deblurring. Recent FPA technologies, however, attempt to fill the sensor surface with as many pixels as possible, and at the same time increase the fill-factor up to 100%. Although these hardware improvements significantly increase the resolution of modern camera systems, further resolution enhancement in software is much more difficult.

Corresponding author: Tuan Pham: E-mail: tuan@qi.tnw.tudelft.nl Phone: +31 15 278 6054 Fax: +31 15 278 6740

In this paper, the influences of PSF and SNR on the limit of SR are identified and estimated. The system PSF is a result of multiple blur sources along the imaging path, two of which are considered in this paper: the optical PSF and the sensor PSF (i.e. the effective photosensitive region per pixel). The influence of the optical PSF on the system's PSF depends on the sampling rate. Smaller values for the sampling rate and fill-factor produce a better SR factor without the need for deconvolution. Should deconvolution be required, then the SNR plays an important role because noise destroys information at frequencies where it overpowers the signal. A low SNR also results in a large registration error. This mis-registration leads to edge jaggedness in the HR image and hampers the reconstruction of small details. Although noise can be reduced by fusion of multiple frames, the registration error cannot be reduced below a theoretical lower bound.²¹ As a result, unless an *a priori* knowledge of the motion of the scene or the camera is available, the SR factor is ultimately limited by the registration error.

In this paper we present many technical details which are often overlooked or oversimplified by current SR algorithms such as non-uniform sensor spread function and cross-talk between adjacent sensors due to microlenses. We stress the importance of having a low fill-factor that have led to previous claims on SR limits¹¹ or capabilities.¹² The paper also gives a guideline to system designers who have to select vision systems that suit their SR needs. Since resolution enhancement and SNR improvement can never be achieved simultaneously, there are a many parameters to be considered: sensor separation for reduced cross-talk at the expense of a lower quantum efficiency, the use of microlenses or alternative technology to increase the fill-factor at the expense of a wider PSF, integration time to increase signal strength at the expense of motion blur susceptibility.

2. THREE-STEP SUPER-RESOLUTION ALGORITHM

SR algorithms are generally comprised of three subtasks: image registration, image fusion and image deblurring. Image registration refers to the task of aligning the LR input images onto a common coordinate system. Image fusion weaves these irregularly positioned samples to form a regular HR image. Image deblurring is finally applied to the HR image to obtain a SR image. Although some algorithms^{6, 10} incorporate image registration, fusion and deblurring into a single optimization procedure, we choose to treat them separately. This has a number of advantages including lower computational complexity, better optimization of each module, and the possibility of using many existing deconvolution packages.^{5, 6, 15} In addition, earlier steps can be revisited for further optimization after the completion of later steps. For example, more accurate gradient-based registration is achievable using image gradients of the non-aliased HR image,¹⁸ and better adaptive fusion¹⁹ is possible by fine-tuning the steering parameters from a first SR estimate.

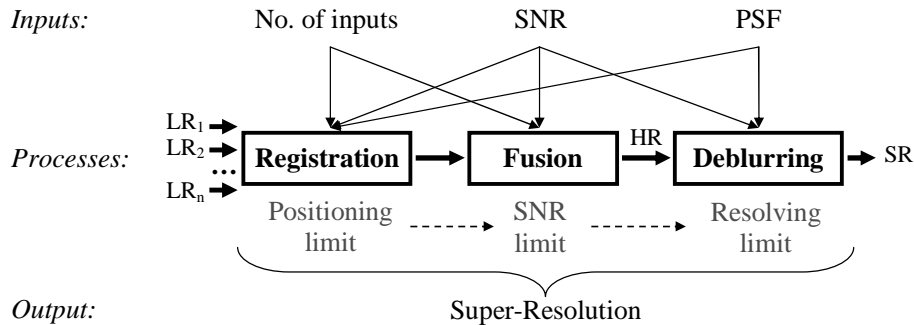


Figure 1. Limiting factors of a 3-step super-resolution algorithm

Since the SR approach we take consists of three separate modules, the limit of resolution enhancement depends on the performance limit within each module. As depicted in figure 1, the quality of the SR result depends on the registration error, the noise amplitude in the fusion result, and the capability of the deconvolution in resolving small details. The performance limit of each module in turn depends on input parameters such as the number of input images, the size of each input image, the SNR and the PSF. Without losing too much generality, we derive the limit of SR in this paper for an image sequence under translational motion, Gaussian noise and a space-invariant PSF.

3. POINT SPREAD FUNCTION

Due to the diffraction limit of finite-size lenses, a point source in the scene corresponds to an intensity blob in the imaging plane, often referred to as the Airy disk or the optical Point Spread Function (PSF). Due to charge transport and sampling, the digitized image is further degraded. The combination of all degradations can be modeled by a system blur, which is a convolution of all PSFs along the imaging path, and noise contamination. Two types of blur are present in all digital imaging systems: the optical blur which specifies the fine-resolving capability of the optics, and the sensor integration blur caused by the accumulation of incoming light over a finite-size sensor element. The optical PSF for in-focus objects can be reduced to the diffraction limit⁸ by good lenses. The sensor PSF, however, cannot be reduced arbitrarily because the sensor has to be big enough to capture sufficient light for a good SNR. The overall PSF can be measured from multiple oriented step edge response, either from a test pattern (e.g. ISO 12233) or directly from an image.¹⁹ This section reviews the formation of the PSF due to current advances in focal plane array technology.

3.1. Advances in sensor array technology

Focal Plane Array (FPA) based camera systems advance in the direction of smaller pixel size and higher fill-factor. A Scanning Electron Microscopy (SEM) image at the bottom of figure 2a shows a real pixel of the 35-mm Charge-Coupled Device (CCD) array in an 11-megapixel Kodak CCD camera.²⁰ As can be seen from the picture, a significant portion of the FPA surface is occupied by transferring gates. Since these gates should not be disturbed by incident light while transferring charges, they are shielded and separated from the photodiodes by metallic separation structures. A significant amount of photons arriving at the imaging plane therefore could not reach the photodiode if the microlenses on top are omitted. To increase the quantum efficiency, microlenses are positioned on top of the CCD sensors to gather more light onto a small photodiode below. However, even with microlenses, neither the fill-factor could reach 100%⁴ nor is the sensor response uniform across the pixel as often assumed by many SR algorithms.¹¹

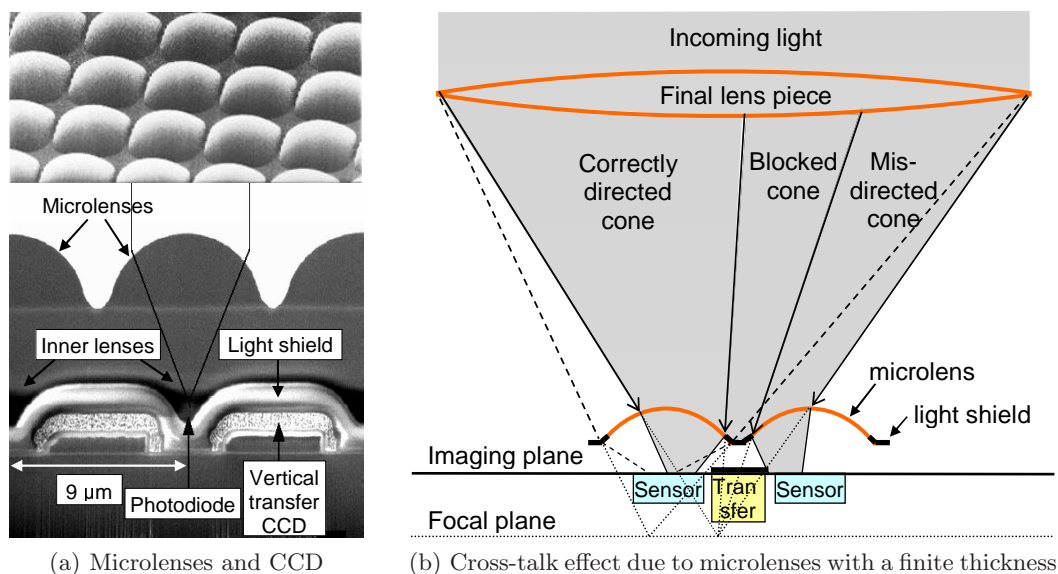


Figure 2. a-top) on-chip microlens array of Super HAD CCD (courtesy of Sony Corporation²²), a-bottom) cross-section of KAI-11000CM CCD (courtesy of Eastman Kodak²⁰), b) cross-talk effect in the presence of microlenses.

It should be noted that the ubiquitously abused 100% fill-factor term should not be taken literally. In fact, even with a very high sensitivity CCD like the Super HAD CCD from Sony,²² its on-chip microlens array reveals large areas of so-called “void space” between individual lenses (figure 2a, top image). Geometrically, the fill-factor is less than 80% for circular microlenses since the area of a unit circle is $\pi/4 = 0.7854$. Square-shaped microlenses could result in a higher photo-active area but their focusing capability is poorer than that of the

circular ones.⁴ Moreover, it is difficult to reduce the void area to zero for small pixel sizes ($\approx 3 - 10\mu m$ in diameter for visible light sensors) since the deposited islands that form the microlenses should be well-separated during thermal smoothing to avoid merging. It is therefore the shape of the microlens that dictates the overall shape of the Sensor Spread Function (SSF). As a result, neither is the SSF square nor is the fill-factor truly 100%.

Though significantly boosting the fill-factor of camera systems, microlens array and equivalent technologies also introduce unwanted effects. One of them is cross-talk between adjacent pixels. This occurs in microlens array when incident light coming in under an angle is misdirected onto a different sensor (figure 2b). Thin film deposits on ASIC (Application Specific Integrated Circuit), another method to increase fill-factor by converting incoming photons to electrons to be collected by the electronics below, also suffers from cross-talk and stray electrons due to a minimum distance between the film and the diode underneath.¹³ Cross-talk widens the shape of the SSF, introducing more blur to the image. Although cross-talk could be reduced by thinning the microlenses or thin film layer, there are theoretical and practical limitations in doing so. Another unwanted effect is angle roll-off, defined as the decrease in response when the angle of incident light increases.²⁰ This effect causes a nonuniform response across the sensor as well as a variation in the SSF across the image. The effect is most noticeable at image borders where light arrives at the FPA at much higher angles. In conclusion, methods to increase the fill-factor of a camera system often result in a non-uniform SSF that extends outside the pixel boundary.

4. NOISE AND REGISTRATION ERROR

4.1. Signal-to-noise ratio

Noise is present in all digital imaging systems due to a number of sources such as photon shot noise, readout noise, dark current noise and quantization noise. While some noise sources can be effectively suppressed such as dark current noise by cooling and quantization noise by using more bits, others are not. Photon shot noise, for example, is unavoidable due to the particle nature of light. Readout noise increases with a higher readout rate, which is desirable in high-speed cameras. The combined effect of these noise sources is often modeled by Gaussian additive noise. The variance of this Gaussian noise can be measured from a flat region in an image⁹ or from two images of the same scene captured under the same degradation (figure 3a):

$$\sigma_I^2 = \frac{1}{2} \text{var}(I_1 - I_2) \quad \text{where } I_1 = I + n_1, I_2 = I + n_2 \quad (1)$$

The Signal-to-Noise Ratio (SNR) of an image I is then computed as $SNR = 10 \log(\text{var}(I)/\sigma_I^2)$.

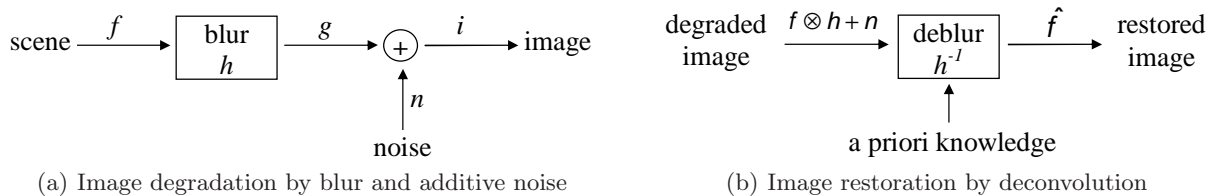


Figure 3. Image degradation and restoration models.

An image sequence with high SNR is likely to result in a good super-resolution image reconstruction. In theory, a higher SNR could be achieved by a larger light-sensitive area, a longer exposure time or a slower readout rate. However, this might not be feasible for certain SR applications due to camera motion or motion in the scene. In these cases, it is better to capture multiple images at a high frame rate, these LR images are then registered and fused together to obtain a HR image with a lower noise variance:

$$\sigma_n^2 = \frac{\mu^2}{N} \sigma_I^2 \quad (2)$$

where μ is the zoom factor between the LR and HR images, N is the number of LR images used in the fusion, and σ_I^2 is the noise variance of the LR images as previously computed in (1).

4.2. Registration error

Although better noise characteristics can be obtained by fusion of multiple frames, a better SR factor is still not guaranteed because of a fundamental lower bound on the registration error.²¹ This limit is dictated by the Cramer-Rao Lower Bound (CRLB) on the variance of *any* pair-wise shift estimator:

$$\sigma_{v_x}^2 \geq \sigma_n^2 / \sum_S I_x^2 \quad \sigma_{v_y}^2 \geq \sigma_n^2 / \sum_S I_y^2 \quad (3)$$

where $\sigma_{v_x}^2 = \text{var}(v_x)$ and $\sigma_{v_y}^2 = \text{var}(v_y)$ are variances of the estimated shift in x and y -direction, σ_n^2 is the noise variance, $I_x = \partial I / \partial x$ and $I_y = \partial I / \partial y$ are the partial derivatives in x and y -direction of a noiseless blurred image I (I corresponds to the blurred signal g in figure 3a). As expected, the minimum variance bound of the registration error is proportional to the noise variance. It is also inversely proportional to the total gradient energy in the image. Since the image I is a blurred version of the scene, the gradient energy in (3) is reduced by a wider PSF. As a result, the estimated shift is more precise for bigger images of scenes rich in texture under a narrow PSF. The CRLB also exists for more complicated motion models such as two-dimensional projective transformation and optic flow. By means of iterative optimization, we have also shown that optimal registration under these motion models are achievable because the CRLB is reached.¹⁸

The registration error can affect the quality of HR image fusion in a number of ways. If the LR images are corrupted by Gaussian noise, the registration estimates are also normally distributed.¹⁸ As a result, if an infinite number of sub-pixel shifted images is used to form a single HR image after registration, the effect of the registration error is equivalent to a Gaussian blur on top of the existing blur sources. In practice, however, the number of LR images is limited. The registration error therefore contributes to the local intensity noise. As illustrated in figure 4d, a variation in the sample position is translated to a variation in the sample signal amplitude. The noise in the reconstructed HR image now consists of two terms: one based on the intensity noise of the LR images and the other based on the registration error:

$$\sigma_n^2 = \frac{\mu^2}{N} \sigma_I^2 + \overline{\|\nabla I\|^2} \sigma_{reg}^2 \quad (4)$$

where σ_{reg}^2 is the variance of the registration error ($\sigma_{reg}^2 = \sigma_{v_x}^2 + \sigma_{v_y}^2$ for shift estimation), and $\overline{\|\nabla I\|^2} = \frac{1}{N} \sum_S \left[\left(\frac{\partial I}{\partial x} \right)^2 + \left(\frac{\partial I}{\partial y} \right)^2 \right]$ is the average gradient energy of N samples in a local neighborhood S . Note that the intensity noise in (4) can be effectively suppressed with an infinite number of LR frames ($N \rightarrow +\infty$). The registration noise, however, remains a limiting factor of the HR image's noise variance. The number of LR images can therefore be derived for any desirable SNR above this limit. As to be shown later in section 5.2, this SNR of the HR image directly controls the attainable SR factor.

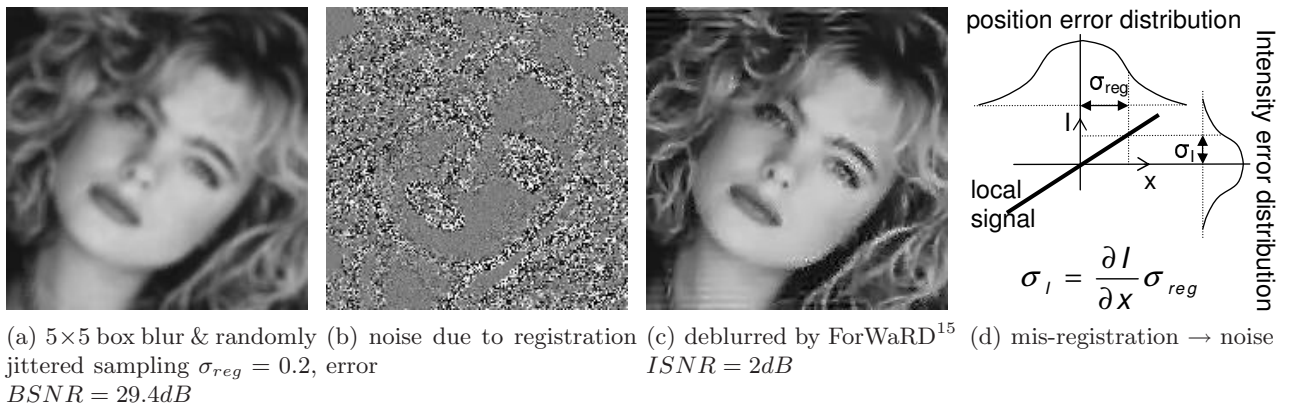


Figure 4. Sample position error reduces local SNR and increases edge jaggedness of the deblurred image.

Noise in the LR image sequence limits the registration precision, which on its turn causes amplitude noise in the HR image. The latter limits the deconvolution step in textured regions. Figure 4a shows a simulation of HR image reconstructed from multiple LR images under mis-registration. First, the original portrait is blurred with a 5×5 box blur. Random noise is then added to the positions of all pixels ($\sigma_{reg} = 0.2$ HR pitch). The image in figure 4a is finally reconstructed from these irregularly positioned samples by two-dimensional spline interpolation. Although noise due to registration error is not visible in the reconstructed HR image, it is visible in the difference image (figure 4b) of the HR image and the 5×5 box-blurred image before position error addition. This space-invariant noise results in poorer resolution near high-gradient details, which are often the targets of SR. As can be seen from the deblurred image in figure 4c, ringing and noise artifacts around edges reduce the detail resolving capability of deconvolution. BSNR in the caption of figure 4a stands for a Blurred Signal-to-Noise Ratio of the corrupted image, whereas ISNR in the caption figure 4c stands for an Improvement in Signal-to-Noise Ratio (ISNR) of the restored image over the corrupted image:

$$BSNR = 10 \log_{10} \left(\frac{\text{var}(g)}{\sigma_n^2} \right) \quad ISNR = 10 \log_{10} \left(\frac{\|f-g\|}{\|f-\hat{f}\|} \right) \quad (5)$$

where g, \hat{f}, f denote the corrupted image, restored image and original image, respectively; and $\|\cdot\|$ is a symbol for quadratic norm.

5. LIMITS OF SUPER-RESOLUTION

Given our previous analysis of the PSF, SNR and registration error, in this section we propose a method to compute practical limits of the super-resolution factor from the input images. Without deconvolution, the maximum SR factor depends on the sampling factor and fill-factor. The sampling factor is defined as the ratio of half the sampling frequency over the cutoff frequency of the optics ($f_s/2f_c$). If real-time SR is not required, deconvolution can result in further resolution enhancement. The limit of SR then depends on the signal and noise spectrum of the HR image after fusion.

5.1. Limit of super-resolution without deconvolution

Neglecting pre-optics blur sources such as atmospheric turbulence, the system transfer function is a product of the Optical Transfer Function (OTF) and Sensor Transfer Function (STF). If the lens system is diffraction-limited, the OTF for a circular aperture is given by:

$$\text{OTF}(f) = \frac{2}{\pi} \left(\arccos \left(\frac{f}{f_c} \right) - \frac{f}{f_c} \sqrt{1 - \left(\frac{f}{f_c} \right)^2} \right) \quad \text{for } f < f_c \quad \text{and } 0 \text{ otherwise} \quad (6)$$

where $f = \sqrt{f_x^2 + f_y^2}$ is the radial frequency in a two-dimensional Fourier space $\{f_x, f_y\}$, and $f_c = 2\text{NA}/\lambda$ is the cutoff frequency of the lens (NA denotes the numerical aperture of the lens and λ is the wavelength of incoming light).

Although it has been noted previously in section 3 that a sensor response might not be uniform over the whole pixel due to cross-talk and angle roll-off effects, the sensor spread function can still be approximated by a square box function after applying a small Gaussian blur. Neglecting this small Gaussian blur, the STF due to the square sensor shape is:

$$\text{STF}(f_x, f_y) = \text{sinc} \left(\frac{w}{2} \frac{f_x}{f_s} \right) \times \text{sinc} \left(\frac{w}{2} \frac{f_y}{f_s} \right) \quad (7)$$

where $\text{sinc}(x) = \sin(\pi x)/(\pi x)$, f_s is the sampling frequency and w^2 is the fill-factor of a square pixel whose width is w -times the pixel pitch ($w \leq 1$). Although the pixel size has been reduced substantially over the years, a general-purpose imaging system rarely satisfies the Nyquist sampling criterion. That is to say, the sampling frequency as dictated by the pixel pitch is often smaller than twice the cutoff frequency of the optics ($f_s < 2f_c$). This under-sampling is the key to SR image reconstruction from multiple LR images.

It is not surprising that two-times SR is always possible with image sequences captured by a severely under-sampled digital camera ($f_s < f_c$). The jitter camera in,² for example, achieves two-times SR out of standard optics and low-resolution CCD sensors. By routinely shifting the CCD by half a pixel back and forth in horizontal and vertical directions, the jitter camera can construct an image two-times larger than its four constituent LR shifted images. The resulting HR image recovers smaller details than the LR images because its spectrum at high frequencies is reconstructed from the aliased spectra of the LR images. Theoretically, the HR spectrum can be extended up to the cutoff frequency by unwinding the aliased LR spectrum, giving a maximum zooming factor of f_c/f_s . However, due to an attenuation of the spectrum at high frequencies, full resolution at that zoom is not achievable. Deconvolution is therefore necessary to enhance the high-frequency part of the spectrum.

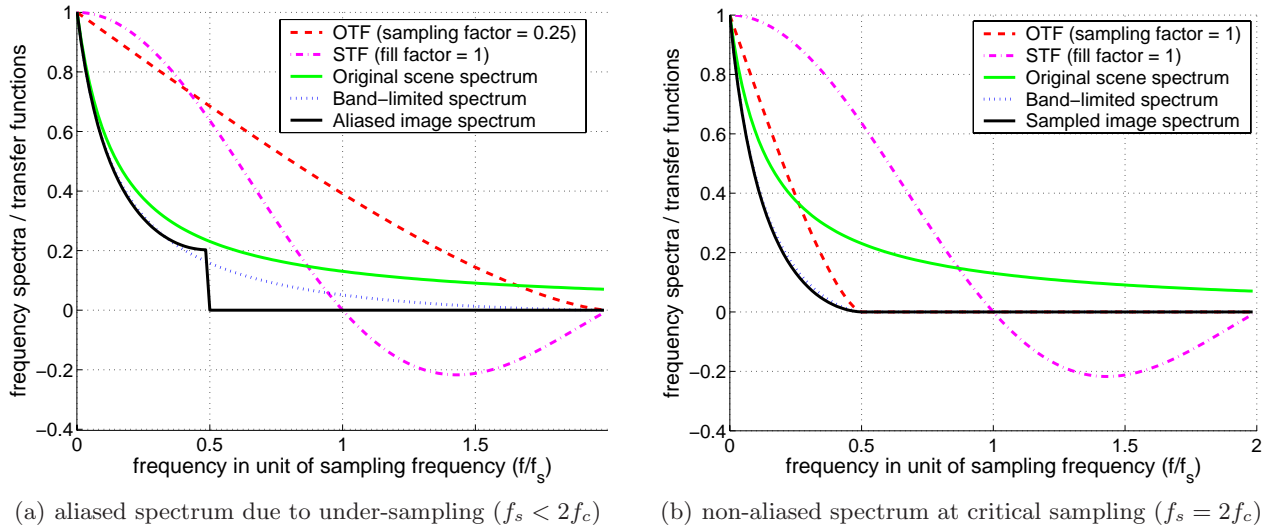


Figure 5. Frequency spectra of an image sampled at different sampling factors

The aliasing caused by an under-sampled imaging system is depicted in figure 5a. In this figure, a typical value for the sampling factor (1/4) and the fill-factor (100%) are used to plot the OTF (dash line) and STF (dash-dot line). As can be seen, the diffraction-limited OTF reduces to zero after two times the sampling frequency, whereas the STF of a square sensor with 100% fill-factor has its first zero-crossing at the sampling frequency. If the frequency spectrum of an original scene is depicted as a gradually decaying continuous line, the spectrum of a bandlimited signal due to the optics is the dotted line. After sampling onto a LR grid, the maximum frequency representable by the discrete signal is $\frac{1}{2}f_s$. As a result, the frequency spectrum of the LR image (dark continuous line) is aliased, i.e. it is *folded* at multiples of $\frac{1}{2}f_s$ due to an overlapping of adjacent copies of the bandlimited spectrum. By doubling the sampling rate, part of the aliased spectrum unfolds and extends to a $[0, f_s]$ range, resulting in two-times super-resolution. On the contrary, the spectrum of an image sampled at critical frequency ($f_s = 2f_c$) is plotted in figure 5b. Since the bandlimited signal reaching the sensors is cutoff at $\frac{1}{2}f_s$, the sampled image is not aliased. Super-resolution by spectrum expansion is therefore also not possible.

5.2. Limit of super-resolution by deconvolution

Super-resolution without deconvolution as presented in the previous subsection only achieves a significant resolution enhancement factor if both the sampling factor and the fill-factor are small. For systems that satisfies the Nyquist sampling criterion, super-resolution reconstruction is not possible without a priori knowledge. For systems with 100%-filled square sensors, the sensor integration results in a box blur with an area of $\mu \times \mu$ pixels in the HR frame (where μ is the zoom factor between the LR and HR images). This significant amount of degradation must be deconvolved to recover details smaller than μ pixels in the HR frame.

Although many deconvolution algorithms have been presented in the literature (regularized inverse filtering,⁷ Wiener filtering,⁷ Maximum Likelihood (ML) deconvolution,¹⁰ Maximum A posteriori Probability (MAP)

deconvolution,⁶ Fourier-wavelet regularized deconvolution,¹⁵ ...), they all degrade rapidly as the input noise increases. This is understandable if deconvolution is considered as a spectrum restoration problem in the Fourier domain. Since the PSF of an imaging system is often real and positive, the transfer function is also real and it only modifies the amplitude but not the phase of the signal spectrum (although phase reversal does occur due to a negative value of the transfer function). Since deconvolution aims to reverse this spectrum attenuation it rarely changes the phase information. The preservation of phase information in the blurred image therefore determines the quality of the deconvolved image. Since phase varies over the whole spectrum in the presence of noise, we can only estimate its average deviation under a certain SNR.

We propose a limit of deconvolution as the ratio between a noise-limiting frequency over half the sampling frequency ($2f_N/f_s$). The noise-limiting frequency here refers to a frequency at which the power of a noiseless blurred signal equals that of the noise. Since the measured signal equals the noiseless blurred signal plus the uncorrelated noise ($i = g + n$, see figure 3a), f_N is the frequency at which the power of the blurred and noisy signal doubles that of the noise:

$$P_I(f_N) = P_G(f_N) + P_N(f_N) = 2P_N(f_N) \quad (8)$$

where P_I , P_G and P_N are the power spectrum of the measured signal i , the noiseless blurred signal g and the noise n , respectively. Since the signal power spectrum decays with a $1/f^2$ profile¹⁶ whereas the noise power is roughly constant, this noise-limiting frequency exists and it can be located from the spectrum of the corrupted image. In super-resolution, this corrupted image is the HR image after fusion. Although the limit of deconvolution can also be computed from a LR image, its higher spectrum due to aliasing may cause a false reading of f_N .

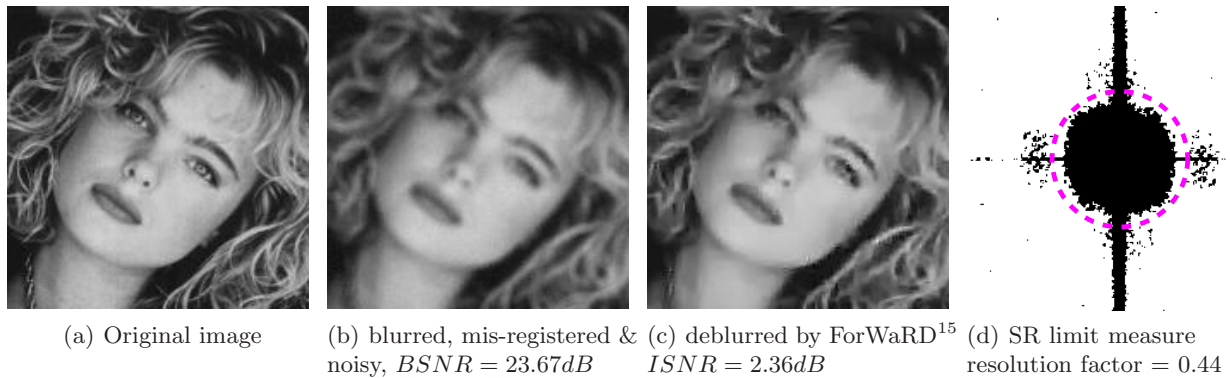


Figure 6. Prediction for limit of deconvolution directly from image: the area of the dashed square in figure 6d equals to the total number of pixels whose corrupted signal power is greater than four-times the noise power (black pixels)

A procedure for estimating the limit of deconvolution of a blurred and noisy image is described with an example as follows. Figure 6b shows a simulation of a HR image reconstructed from multiple noisy LR images under mis-registration. The original portrait in figure 6a is first blurred with a 5×5 box blur. Gaussian noise is then added to both the amplitude ($\sigma_I = 3$, $BSNR = 23.67dB$) and position ($\sigma_{reg} = 0.2$ HR pitch) of every pixel in the noiseless blurred image. The simulated HR image is finally reconstructed from these irregularly positioned noisy samples by a surface interpolation. As previously shown in section 4.2, the HR image is corrupted by both a space-invariant intensity noise and a space-variant noise due to registration error. Since we are interested in the reconstruction of textured regions in the image, $\|\nabla I\|^2$ in equation (4) is set to a value at 90-percentile of the HR image's gradient energy histogram. Due to Parseval's theorem,¹⁷ the average noise power in the Fourier domain equals to the variance of the noise in the spatial domain, which in turn can be computed from equation (4). Assuming the power spectrum of the noise is flat, we can estimate the noise-limiting frequency by a threshold of the power spectrum of the HR image at two-times the average noise power. Figure 6d shows the result of this threshold on a smoothed version of the power spectrum (2×2 box blur to get the local average power), in which black pixels denote regions where the power spectrum is greater than two-times the noise power. Due to variations in the power spectrum, the threshold result is not confined in a single blob. However, a rough size of

the black mask can be represented by the dashed circle whose area equals to the total number of pixels in the mask. The ratio between the diameter of the circle and the image size is a measure for limit of deconvolution. A ratio of 0.44 in figure 6d means that the maximum resolution contained in the blurred and noisy image in figure 6b is only 0.44 of its full size. As predicted, the deblurred image using the Fourier-Wavelet Regularized Deconvolution (ForWaRD) algorithm¹⁵ in figure 6c cannot recover many small details in the original image. The deblurred image in figure 6c is also of lower resolution than the deblurred image in figure 4c, where the equivalent noise power of the HR image in figure 4b is smaller due to zero intensity noise ($\sigma_I = 0$ in the similar experiment in section 4.2).

6. EXPERIMENTAL RESULTS

This section describes a synthetic experiment to show that SR from under-sampled LR images is possible without deconvolution. Although a higher SR factor is achievable by deconvolution, a lower SNR reduces this additional enhancement. In addition, the obtainable SR factor at a given SNR can be predicted by the limit of deconvolution introduced in this paper.

In this experiment, multiple LR images (64×64) are generated from a very high resolution Pentagon image (1028×1028) by simulating the diffraction-limited optics and the square box sampling. The very HR spectrum is first attenuated by the corresponding OTF and STF in the Fourier domain. The very HR blurred image is then randomly shifted, downsampled and corrupted by Gaussian noise to form multiple LR images. The original Pentagon image has 100% fill-factor as measured from its edge spread function. Its OTF is assumed to be 100% over the whole spectrum because a ground truth is not known. The sampling factor for the generated LR images ranges from eight-times under-sampling to two-times over-sampling ($f_s/2f_c = [0.125, 0.25, 0.5, 1, 2]$). The simulated fill-factor also ranges from a very low to a full fill ($w^2 = [0.0625, 0.125, 0.25, 0.5, 1]$). For each combination of sampling factor and fill-factor, multiple 64×64 LR images are fused together to form a 256×256 HR image.¹⁹ The number of LR images used in fusion is varied to obtain three levels of BSNRs for the HR image: high ($40dB$), medium ($20dB$) and low ($10dB$). SR images are finally obtained from the corresponding HR images by ForWaRD deconvolution¹⁵ using the known PSF from an diffraction-limited lens and a square box sampling model (equations (6), (7)). Figure 7 shows an example of a LR image generated with a sampling factor of 0.25 and a fill factor of 1, its reconstructed HR image at $20dB$ and the resulting SR image after deconvolution.



(a) 64×64 LR, $f_s/2f_c = 0.25$, $w^2 = 1$ (b) 256×256 HR, $BSNR = 20dB$ (c) 256×256 SR, SR factor=3.4

Figure 7. The LR, HR and SR images generated with a sampling factor of 0.25 and a fill-factor of 1 at $20dB$ BSNR

To quantify the sharpness of the SR images, the width of an edge transition in the Pentagon image is computed (marked with a light rectangle in figure 7c). The edge width is the standard deviation of a Gaussian smooth edge model fitted to the straight step edge¹⁹. The SR factor is then computed as the ratio of edge widths between the SR image and a 256×256 reference image sampled at Nyquist rate ($f_s = 2f_c$) with 100% fill-factor. Since all SR images are deconvolved with a known PSF, the reference image is also deconvolved for a fair comparison.

The edge width of the reference image after deconvolution is found to be 0.724 pixel. Edge widths are estimated for all pairs of HR/SR images. Not surprisingly, the SR images have sharper edges than their corresponding HR images.

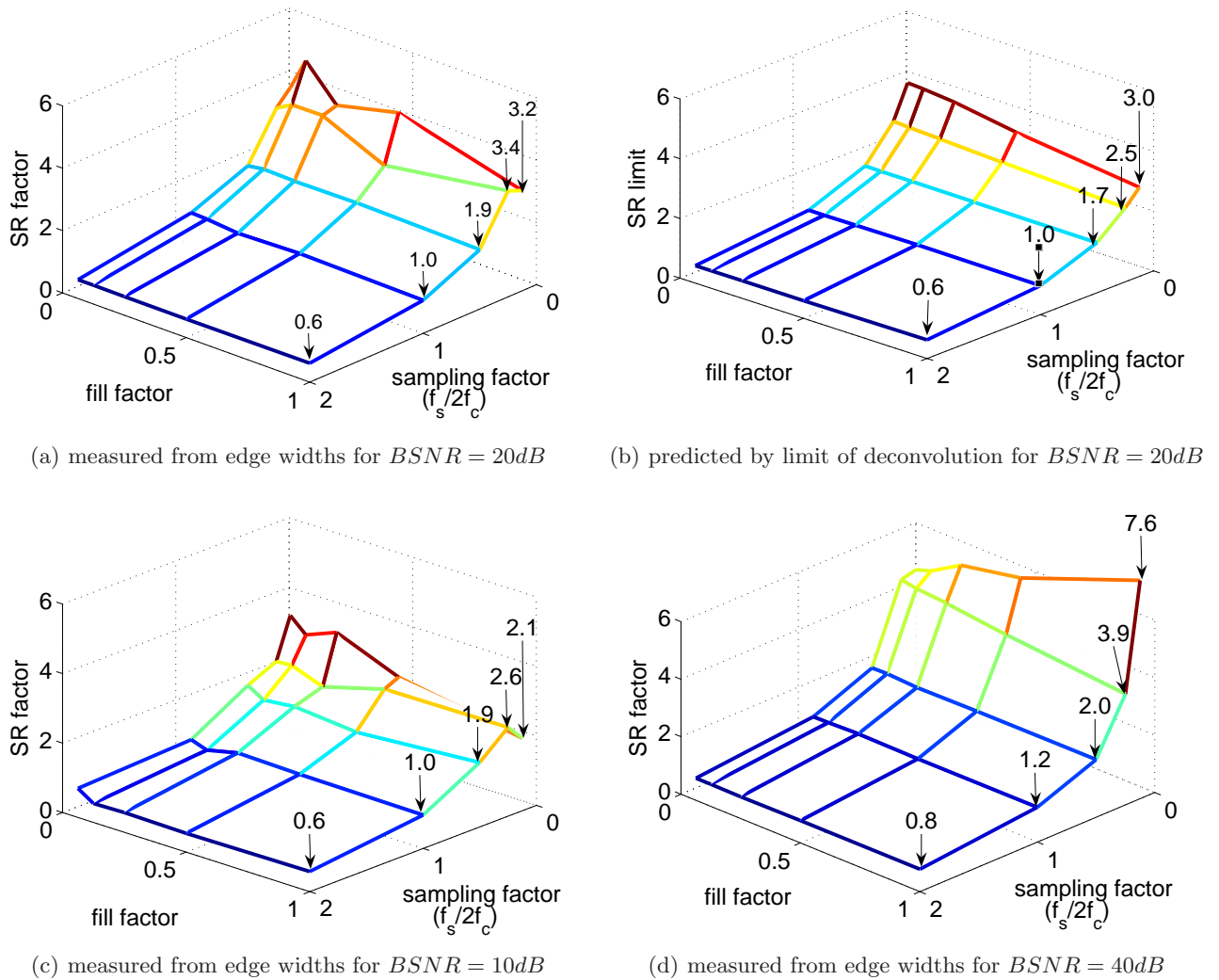


Figure 8. Super-resolution factors for different sampling factors and fill-factors

Figure 8 plots the measured SR factors of the reconstructed SR image alongside with the SR limit predictions from section 5.2. Along the sampling factor axis, the SR factor increases as the LR images are more severely under-sampled. Over-sampling, however, results in resolution reduction because the spectra of the LR images are cutoff well before its maximum representable frequency ($\frac{1}{2}f_s$). The predicted SR limits for images at 20dB (figure 8b) generally agrees with the SR factors computed from the edge width measure (figure 8a). The predicted SR factor is slightly underestimated for very low sampling factors because even then the SR images are aliased, which prevent an accurate estimation of the noise-limiting frequency (section 5.2). As can be seen from figure 8a, for a medium level of BSNR ($\approx 20dB$) and a low sampling factor (≤ 0.25), three-times SR is possible for any values of the fill-factor. SR factors are also measured for the SR images constructed from HR images at 10dB and 40dB (figure 8c and 8d, respectively). A lower mesh plot for a lower BSNR confirms that noise strongly degrades the level of attainable SR. With the same setting for the sampling factor (≤ 0.25) and fill-factor (100%), 10dB HR images result in 2.5-times SR whereas more than 4-times SR is achievable with HR images at 40dB.

7. CONCLUSION

In conclusion, we have pointed out the influence of point spread function and signal-to-noise ratio on the limit of super-resolution. Without deconvolution, the limit of SR depends on the sampling factor and fill-factor. Lower values of these factors produce higher SR factors. In a typical camera setting where the sampling factor is less than 25% and the fill-factor is close to 100%, up to two-times SR is achievable without deconvolution. With deconvolution, the limit of SR depends on the signal and noise content of the HR image after fusion. The ratio between the noise-limiting frequency and half the sampling frequency is the limit of resolution enhancement by deconvolution. This noise-limiting frequency is the maximum frequency whose spectrum can be reconstructed reliably in the presence of noise. This frequency can be located in the spectrum plot where the corrupted signal power is twice the noise power.

The paper offers a method for measuring the limit of SR directly from the data. Experiments on real data with an ideal imaging model show that at SNRs of 10dB, 20dB and 40dB, the maximum attainable SR factors are 2.5, 3 and 4-times respectively. These numbers agree well with the limit of four-time SR often found in the literature. The limit of SR presented in this paper is applicable to general super-resolution reconstruction algorithms only. Higher SR factors can be achieved for certain types of image content such as texts³ or faces¹ because a priori knowledge of the scene can be incorporated into the SR algorithms. However, special care should be taken with this *recogstruction* (recognition + reconstruction) approach because a slight model error can lead to drastically different and therefore erroneous results.

REFERENCES

1. S. Baker and T. Kanade. Limits on Super-Resolution and How to Break Them. *PAMI*, 24:1167-1183, 2002.
2. M. Ben-Ezra, A. Zomet and S.K. Nayar. Jitter camera: high resolution video from a low resolution detector. In *Proc. of CVPR*, Washington DC, 2004.
3. D.P. Capel and A. Zisserman. Super-resolution enhancement of text image sequences. In *Proc. of ICPR*, Barcelona, 2000. pp 600–605.
4. D. Daly. *Microlens Arrays*. Taylor & Francis, London, 2001.
5. S. Farsiu, D. Robinson, M. Elad and P. Milanfar. Fast and robust multi-frame super-resolution. *IEEE Trans. on Image Processing*, 13(10):1327-1344, 2004.
6. R.C. Hardie, K.J. Barnard and E.E. Armstrong. Joint map registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Trans. on Image Processing*, 6(12):1621–1633, 1997.
7. R.C. Gonzalez and P. Wintz. *Digital Image Processing*. Addison-Wesley Publishing Company, 1997.
8. J.W. Goodman. *Introduction to Fourier optics*. McGraw-Hill, 1996.
9. J. Immerkaer. Fast noise variance estimation. *CVIU*, 64(2):300–302, 1996.
10. M. Irani and S. Peleg. Improving resolution by image registration. *CVGIP: Graphical Models and Image Processing*, 53(3):231–239, 1991.
11. Z. Lin and H-Y. Shum. Fundamental limits of reconstruction-based superresolution algorithms under local translation. *PAMI*, 26(1):83–97, 2004.
12. C.L. Luengo Hendriks and L.J. van Vliet. Improving resolution to reduce aliasing in an undersampled image sequence. In *Proc. SPIE vol. 3965A-23*, p. S4, San Jose, CA, 2000.
13. T. Lule, S. Benthien, H. Keler, F. Mutze, P. Rieve, K. Seibel and M. Sommer. Sensitivity of CMOS based imagers and scaling perspectives. *IEEE Trans. on Electron Devices*, 47(11):2110–2122, 2000.
14. J.C. Mullikin, L.J. van Vliet, H. Netten, F.R. Boddeke, G.W. van der Feltz and I.T. Young. Methods for CCD camera characterization, in: *Proc. of SPIE vol.2173*, pp. 73–84. 1994.
15. R. Neelamani, H. Choi and R. Baraniuk. ForWaRD: Fourier-wavelet regularized deconvolution for ill-conditioned systems. In *IEEE Trans. Signal Processing*, 52(2):418–433, 2004.
16. B.A. Olshausen and D.J. Field. Natural image statistics and efficient coding. In *Network*, 7:333–339, 1996.
17. A.V. Oppenheim, A.S. Willsky and I.T. Young. *Signals and Systems*. Prentice-Hall, 1983.
18. T.Q. Pham, M. Bezuijen, L.J. van Vliet, K. Schutte and C.L. Luengo Hendriks. Performance of optimal registration estimators. In *Proc. of Visual Information Processing XIV*, SPIE Defense and Security, 2005.

19. T.Q. Pham, L.J. van Vliet and K. Schutte. Robust fusion of irregularly sampled data using adaptive normalized convolution. *submitted*, 2004.
20. G. Putnam, S. Kelly, S. Wang, W. Davis, E. Nelson and D. Carpenter. Photography with an 11-megapixel, 35-mm format CCD. In *Proc. of SPIE 5017*, 2003.
21. M.D. Robinson and P. Milanfar. Fundamental performance limits in image registration. In *IEEE Trans. Image Processing*, 13(9):1185–1199, 2004.
22. Sony Corporation. *Super HAD CCD*. Document available online at: http://www.sony.net/Products/SC-HP/sys/ccd/sensor/super_had.html.
23. A.J.P. Theuwissen. *Solid-State Imaging with Charge-Coupled Devices*. Kluwer academic publishers, 1995.
24. R.Y. Tsai and T.S. Huang. Multiframe image restoration and registration, chapter 7 in: *Advances in Computer Vision and Image Processing*, pp. 317–339. 1984.