Image Capture Modelling for High Resolution Reconstruction

Donald G Bailey

Institute of Information Sciences and Technology,
Massey University, Private Bag 11222, Palmerston North, New Zealand

Abstract: A research topic that has been receiving much interest recently is the reconstruction of a high resolution image from an ensemble of low resolution images of the same scene. Such image fusion implies that each low resolution image individually does not contain all of the available information, and therefore must be subject to aliasing. The reconstruction process involves resampling the ensemble at a higher sample rate, effectively unscrambling the aliased information. However, the reconstructed high resolution image appears blurred because of area sampling caused by the sensor, combined with the low pass response of the camera electronics. By modelling the image capture system, the point spread function may be estimated and then removed by using inverse filtering in the frequency domain. If the individual low resolution images are subject to a spatially variant point spread function (resulting from perspective or radial distortion or focus limitations), a more complex spatially variant inverse filter must be used.

Keywords: image fusion, super-resolution, aliasing, inverse filtering, camera calibration

1 Introduction

Image fusion involves the combining of data from a number of image sources in such a manner that the output image contains more information than that contained in any of the individual images. The image sources are often different in their properties or characteristics. An important use of image fusion is to trade temporal resolution for spatial resolution, by combining several low-resolution images to construct a high-resolution image [1, 2].

The pixels of an image provide a set of samples of the world. The sampling density, or the spacing between the pixels, limits the achievable resolution. However if a series of images is captured, each with the samples in slightly different locations, the combined sample density is higher that that of any single image. Constructing a high-resolution image consists of registering all of the individual low-resolution images to one another, and then resampling the ensemble using a single sampling grid.

For there to be more information in the ensemble than in a single image, a single image must not contain all of the information. If each low-resolution image was sampled at the Nyquist rate (or higher) then each image contains sufficient information to be able to recover the original data exactly by using sinc interpolation. This means than any single image would be able to provide all of the information required for reconstruction at any desired resolution. Having multiple independent images would not provide an improvement. Therefore, for multiple images to provide additional information, the sample frequency for each individual image must be below the Nyquist rate, and the images be subject to aliasing. The process of constructing a higher resolution image untangles the aliased information, so that the output image contains more information than that available from any of the individual input images.

To investigate some of the issues involved in this process, a simplified problem was considered—that of reconstructing a high-resolution bar code from a low-resolution image. While the ability to read a low-resolution bar code image is a useful problem in its own right, images of bar codes offer a number of simplifications that allow the difficulties in constructing a high-resolution image to be determined, without being confounded by difficulties inherent in the data itself. These simplifications are:

1) A bar code is really a one-dimensional image. The reduction in dimensionality simplifies both the registration and the resampling steps considerably.

2) Bar code images are well defined and have high contrast making it easier to evaluate the quality of the high resolution reconstruction.

3) If the bar code is tilted slightly with respect to the camera, a two-dimensional image of the bar code provides an ensemble of independent one-dimensional images. Therefore rather than capturing an ensemble of images, a single two-dimensional image provides all of the required input data.

2. Image Reconstruction

While this paper concentrates on modelling and removing the effects of the image capture system, it is also necessary understand image reconstruction process. This is presented in more detail in [3] but outlined here. Image reconstruction consists of two step: registration, and resampling.

2.1 Registration

To combine the low-resolution images, the individual images must first be registered to one another. For bar code images, this involves determining the horizontal offset or shift between successive rows within the two-dimensional image. While there are several approaches to registration [4], it was decided to use the phase information in the frequency domain. A spatial shift corresponds to a phase shift in the frequency domain proportional to both frequency and the size of the shift in the image. If an image is shifted by *a* pixels, the shift in the frequency domain is given by [5]

$$f(x+a) \Leftrightarrow F(\omega)e^{j\omega a} \tag{1}$$

where f(x) the pixel value as a function of position x, F(ω) is spatial frequency content of the image, and $\Leftrightarrow$ indicates a Fourier transform relationship. To make use of this relationship, a phase shift image was obtained by taking the Fourier transform of each row in the bar code image, retaining only the phase. The phase of the first row in the image was then subtracted from all subsequent rows to obtain the phase shift relative to the first row in the image. Let r be the row number and ϑ be the phase shift:

$$\vartheta = \angle F_r(\omega) - \angle F_0(\omega) = \angle e^{j\omega ra} = \omega ra \tag{2}$$

The phase shift image contains the frequency dependent phase relative to the first row in the image. As the phase of each image is in the range ±π, the difference is in the range ±2π. Starting at the DC component of the first row, the phase is unwrapped by adding or subtracting 2π to each value to minimise the difference in phase with both the same frequency on the row above and the immediately lower frequency on the same row. This simple approach works well for the low frequencies where the amplitude is large, although at the higher frequencies where the amplitude is smaller, the phase is more easily perturbed by aliasing and noise.

The offset per row, *a*, is found by doing a least squares fit of the phase surface predicted by equation (2) to the phase image. In doing the fit, only the central 25% of the phase image is used because the higher frequency phase terms are subject to error.

$$a = \frac{\sum_{r,\omega} \vartheta(r,\omega) r\omega}{\sum_{r,\omega} r^2\omega^2} \tag{3}$$

2.2 Resampling

Since each row of the bar code image has a slight offset, by selecting and interleaving the samples from the different rows, we can increase the sampling rate. While in principle, any increase in sampling rate up to 1/*a* may be obtained in this way, it is convenient to have an integer multiple of the original sampling rate because that allows a small number of complete rows to be selected and interleaved. If α is the increase in sampling rate, then samples from rows

$$r_n = \text{round}(n / \alpha a) \quad 0 < n < \alpha - 1 \tag{4}$$

are interleaved. This effectively selects the rows with offsets nearest to the desired sample spacing. While interpolation between adjacent rows could be used, the position error in selecting the nearest row is less than $\alpha a / 2$, which is small for small offsets.

Rather than construct a single high-resolution image of the bar codes, a series of high-resolution images is constructed using the r_n as offsets. This gives additional data that may be averaged later to improve the signal to noise ratio. The slope of the bar codes within the constructed image is increased by a factor of α because of the horizontal sample rate is increased but the vertical sample rate remains unchanged.

2.3 Results

Applying this process to an actual bar code gives the results shown in figure 1. The low resolution bar code is rotated relative to the camera by about 3° to make each row of the code independent. The image is captured with a sample frequency of 80% of the Nyquist limit, and is therefore subject to severe aliasing to the extent that the thinner bars are unable to be resolved. Reconstruction yields a high resolution, but blurred image of the bar codes. The blurring results from the limitations of the image capture process.

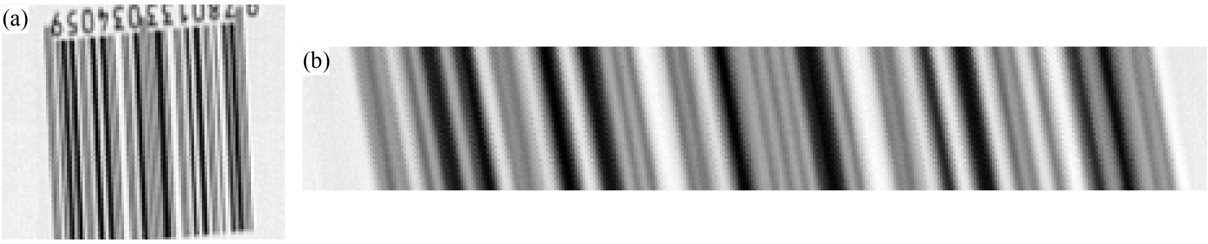

Figure 1: Applying the reconstruction to an actual image. (a) The original low resolution image. (b) The reconstructed high resolution image.

3 Image Capture Degradations

Capturing an image for processing is not a simple process. There are several steps involved, and each step results in a deterioration of the quality of the data. To obtain a good reconstruction, the image capture process must be modelled to enable the effects of the imaging system to be removed of compensated for. The various steps or processes involved in capturing an image are identified in figure 2.

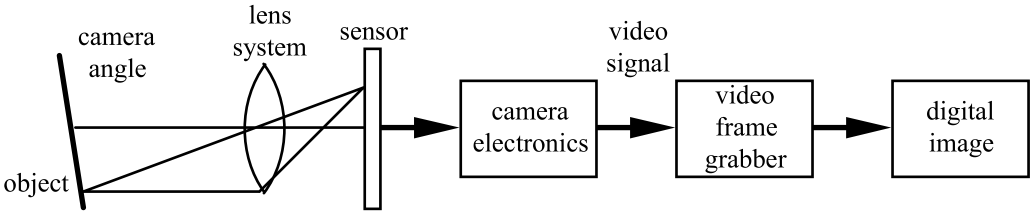

Figure 2: The steps involved in capturing an image.

1) **Camera angle:** The angle between the camera and the object can introduce perspective distortion, where an object closer to the camera occupies a larger proportion of the field of view than the same object would if it was further away. Any perspective distortion must be taken into account in the

registration and resampling processes. The camera angle also affects the lens and sensor PSFs as described in those sections.

2) **Lens system:** Lenses never focus perfectly so there is always some blur even when in best focus. This will be exacerbated if the image is slightly out of focus. If the object is not flat and perpendicular to the line of sight of the camera, not all object points will be in best focus. The effect of the lens point spread function (PSF) is to attenuate the high frequencies, effectively acting as a low pass filter. Determining the width of the PSF is not a straightforward procedure since it depends on the lens and aperture settings as well as the object distance. This is further complicated by the fact that the PSF is also spatially variant, being slightly wider on the edge of the field of view than in the centre. This is in addition to any spatial variance resulting from the camera angle. The quality of the lens is often chosen to match the sensor resolution, although for most solid state sensors, this is unlikely to be the limiting factor for moderate improvements in resolution. Any radial distortion introduced by the lens must also be taken into account in the resampling process.

3) **Image sensor:** Images are not point-sampled by the sensor, but area-sampled. The area sampling is equivalent to convolving an ideal image with a rectangular function, and then sampling the result. Convolution with the rectangular function has the effect of scaling the frequencies with a $\sin\omega / \omega$ low pass filter. The zeros will limit what can be reconstructed, since information at those frequencies is lost. The type of sensor may have significant bearing on the point spread function. Not all of the available area is active for sensing and there can be significant variations in sensor geometry between different sensor types. These factors affect the width of the PSF. Information on the sensor geometry is readily available from specification sheets, so determining the sensor PSF is straight forward. Both the lens PSF and the sensor PSF act as low pass filters, and reduce the aliasing caused by the sampling process. This will limit the extent to which the resolution may be improved. To enhance the high-resolution images, inverse filtering may be used to partially compensate for the effects of the sensor PSF. The situation becomes more complex however if there is either perspective or radial distortion because the area of the object imaged by each pixel will be different. These distortions therefore make the sensor PSF spatially variant, complicating its removal.

4) **Camera electronics:** Unless it is a digital camera, the output from the sensor is converted into an analogue video signal. This may involve a sample and hold circuit to reduce clock noise, followed by a low pass filter to interpolate the samples and produce a continuous signal. Some cameras have a high frequency emphasis on the low pass filter to reduce the blur to any sharp edges. If the frequency response of the camera system is known, it may be removed by inverse filtering. Note that when reconstructing two-dimensional images, the camera response will be one-dimensional, along the rows. If the low resolution images are subject to rotation, then in the reference frame of the reconstructed high resolution image, the one-dimensional filter PSF will have different orientations for each of the low resolution images.

5) **Video frame grabber:** The analogue video signal is then sampled at the video frame grabber, after possibly passing through a low pass filter to reduce aliasing. The sample rate of the video frame grabber will almost certainly be different from that at the sensor. The multiple sampling will introduce another set of aliasing artifacts because the replications resulting from the sensor sampling will not have been completely eliminated by the low pass filter.

The combined effects of the imaging may be seen in figure 1(b) for a reconstructions of a bar code image. The overall response of the image capture process is to strongly attenuate the high frequency content of the image. This has the consequence that the reconstruction process yields a blurred image of the high-resolution bar codes.

4 Image Capture Modelling

This paper does not address camera calibration issues relating to perspective distortion and lens distortion. These are covered in much detail in other papers (see for example [6]). The individual low resolution images need to be corrected for distortion prior to, or as part of the image registration and resampling steps. For the bar code images, the line of sight of the camera was perpendicular to the bar code to eliminate perspective distortion, and only the central portion of the two dimensional image was used to minimise radial distortion.

Many of the degradations to each of the low resolution images cannot easily be removed before reconstructing the high resolution image because of the scrambling of the data caused by multiple aliasing. This aliasing is unavoidable, and is in fact essential for reconstructing a high resolution image. The image reconstruction process effectively unscrambles the aliasing, but does not compensate for the various PSFs inherent in the image capture system.

While the system PSF (including a different spatially variant PSF for each low resolution image) may be removed as part of the reconstruction using the methods described in [1] and [2], the simplified reconstruction method described in section 2 requires a separate post-reconstruction filter. By definition, a spatially variant PSF cannot be removed using a linear filter, and requires a spatially variant compensation filter. With the bar code images used here, this is not a problem because the camera setup was arranged to minimise any spatial variance of the PSF (by minimising distortion, and only using the central part of the captured image). This allowed a linear compensation filter to be used in this instance.

The frequency response of the system was determined empirically by comparing the reconstruction of an ideal image with that of the actual image. The estimated response is shown in figure 3.

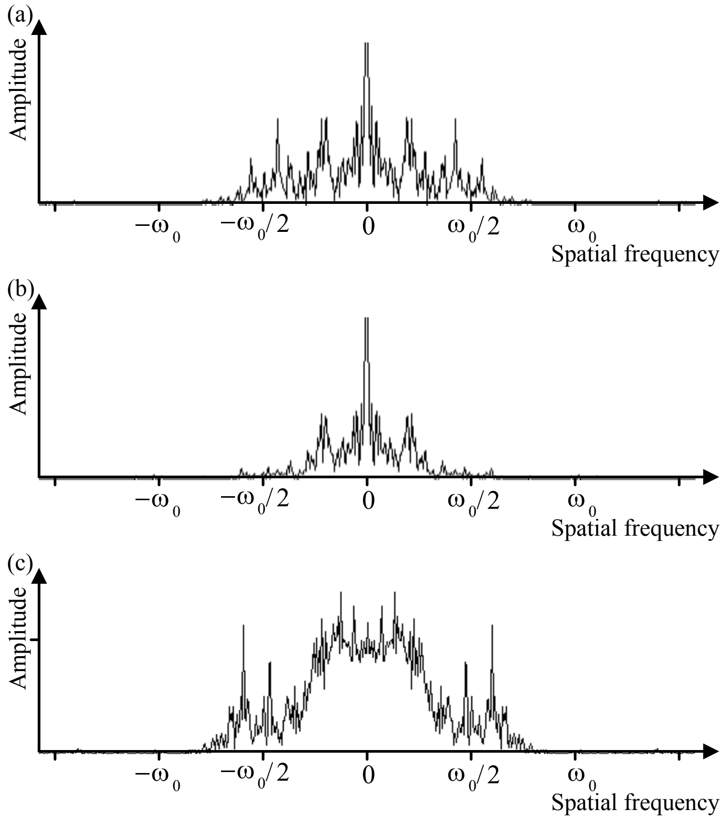

Figure 3: Determining the frequency response of the imaging system:
(a) The frequency content of a reconstructed ideal bar code image.
(b) The frequency content of the actual bar code image from figure 1(b).
(c) The estimated frequency response of the imaging system calculated by dividing (b) by (a).

This transfer function is not a simple product of the individual transfer functions of the imaging system components because of aliasing. The central lobe of figure 3(c) corresponds to a low pass filter with a cut frequency of approximately 4 MHz, which is typical for a video camera [7]. The secondary lobes result from the higher frequency information being aliased into the pass band of the electronic low pass filter by the sampling at the sensor. This information is later unfolded by the reconstruction process, giving humps in the frequency response outside the main pass band. This makes accurate modelling of the system transfer function more difficult.

The system transfer function from figure 3(c) is used to construct an inverse filter [8] which is applied to the reconstructed high resolution image. The result (figure 4) considerably reduces the blur, but also amplifies any noise present in the image. This noise may be reduced if necessary by averaging along the length of the bars.

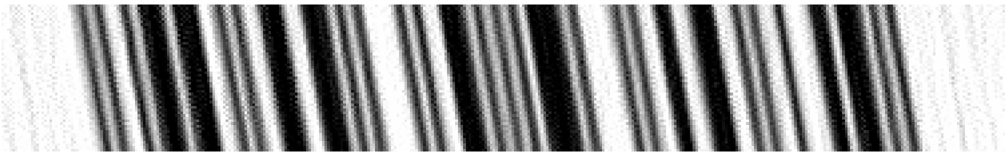

Figure 4: Reconstructed high resolution bar code image after correcting for the system response.

5 Conclusions

Using image fusion to achieve super-resolution requires that the captured low resolution images be aliased. The reconstruction process effectively unscrambles the aliased high frequency information giving an improvement in resolution. If the low resolution images is not aliased, each low resolution image contains all of the information and no higher frequency information is available to recover.

To achieve super-resolution, it is necessary to compensate for the limitations of the image capture system when reconstructing the high-resolution image. If this is not done, the imaging system limits the achievable gains in resolution. The process involves being able to accurately model the system transfer function. The image capture system has a strong low pass response with multiple aliasing caused by multiple sampling. If the point spread function of the degradation is not spatially variant, the low pass responses may be removed by inverse linear filtering. This is possible because the high frequency terms go to zero only slowly (apart from at specific frequencies).

References

[1] R.C. Hardie, T.R. Tuinstra, J. Bognar, K.J. Bernard, and E. Armstrong: High resolution image reconstruction from digital video with global and non-global scene motion. *Proc. IEEE International Conference On Image Processing*, Vol I (October 1997) 153-156.

[2] M.C. Hong, M.G. Kang, A.K. Katsaggelos: An iterative weighted regularized algorithm for improving the resolution of video sequences. *Proc. IEEE International Conference On Image Processing*, Vol II (October 1997) 474-477.

[3] D.G. Bailey: Super-resolution of bar codes. *Proc SPIE Vol 3521, Machine Vision Systems for Inspection and Metrology VII* (November 1998).

[4] A.K. Jain: *Fundamentals of Image Processing*. Prentice Hall, Englewood Cliffs, New Jersey (1989) 400-407.

[5] E.O. Brigham: *The Fast Fourier Transform*. Prentice Hall, Englewood Cliffs, New Jersey (1974).

[6] R.J. Valkenburg: Classification of camera calibration techniques. *Proc Image and Vision Computing NZ* (August 1995) 43-48.

[7] K.G. Jackson, G.B. Townsend: *TV and Video Engineers Reference Book.* Butterworth-Heinemann, Oxford (1991).

[8] K.R. Castleman: *Digital Image Processing*. Prentice Hall, Englewood Cliffs, New Jersey (1979).