2010

Super-resolution Using Adaptive Wiener Filters

Russell C. Hardie
University of Dayton, rhardie1@udayton.edu

1

Super-resolution Using Adaptive Wiener Filters

Russell C. Hardie
Department of Electrical and Computer Engineering, University of Dayton, 300 College Park, Dayton, OH, 45469-0226.

CONTENTS

1.1 Introduction

The spatial sampling rate of an imaging system is determined by the spacing of the detectors in the focal plane array (FPA). The spatial frequencies present in the image on the focal plane are band-limited by the optics. This is due to diffraction through a finite aperture. To guarantee that there will be no aliasing during image acquisiton, the Nyquist criterion dictates that the sampling rate must be greater than twice the cut-off frequency of the optics. However, optical designs involve a number of trade-offs and typical imaging systems are designed with some level of aliasing. We will refer to such systems as detector limited, as opposed to optically limited. Furthermore, with or without aliasing, imaging systems invariably suffer from diffraction blur, optical abberations, and noise.

Multiframe super-resolution (SR) processing has proven to be successful in reducing aliasing and enhancing the resolution of images from detector limited imaging systems [25]. If relative motion between the scene and camera is present, sampling diversity is provided by the multiple looks at the scene that

can be exploited to combat undersampling. Such processing can be viewed as trading temporal resolution for spatial resolution. This allows us to reduce or eliminate aliasing artifacts. Furthermore, if aliasing can be reduced to a minimal level, linear restoration techniques can be successfully applied to deconvolve the blurring effects of the system point spread function (PSF). Note that if little or no aliasing is present in the uncompensated imaging system, single frame restoration may be a more appropriate choice for many applications. If one does employ multiframe SR, it is critical that the motion include a subpixel component and be estimated accurately.

A class of computationally simple multiframe SR methods are those based on nonuniform interpolation [1, 2, 10–12, 19, 23, 26, 28, 30, 33–36]. Such methods are of particular interest for implementing SR in real-time or soft real-time. These nonuniform interpolation based methods typically begin by using image registration to position the observed pixel values from each frame onto a common high resolution (HR) grid. However, the extra samples are generally distributed nonuniformly, unless the motion is very carefully controlled. The nonuniformly sampled HR grid is of little practical use. Therefore, a nonuniform interpolation operation is used to create a uniformly sampled high resolution (HR) image with reduced aliasing. The nonuniform interpolation based SR methods then typically employ a restoration step to reduce the blurring effects of the system PSF.

Most nonuniform interpolation based SR methods use independent interpolation and restoration steps. One potential downside of this is that an independent restoration step may aggravate artifacts from an imperfect nonuniform interpolation step. For example, when the distribution of low resolution pixels on the HR grid is poor, any nonuniform interpolation step will suffer. An independent restoration step can easily exaggerate any resulting artifacts. A new approach, using an adaptive Wiener filter (AWF), combines the nonuniform interpolation and restoration into a single step [14, 15, 22]. This provides potential computational savings as well as robustness to the spatial distribution of low resolution pixels.

The AWF SR method forms a nonuniformly sampled HR grid in the traditional way. However, the AWF method produces the final output pixels with a single spatially-adaptive weighted sum operation using a finite moving window. Here each output HR pixel is formed as a weighted sum of neighboring observed pixels from the nonuniform HR grid. By designing appropriate weights, the output from this single weighted sum operation is not only on a uniform grid, but also restored from the system PSF. Note that the weights in the AWF SR approach are optimized for the specific local spatial arrangement of the neighboring pixels on the HR grid. In contrast, when independent nonuniform interpolation and restoration steps are used, the restoration step does not exploit knowledge of the original distribution of the observed pixels. Rather, it only sees a uniform HR produced the nonuniform interpolation.

Some variations of the AWF SR have been explored. In [22], the HR grid is discrete and the spatial locations of the observed pixels are quantized to

fit on the discrete grid. If more than one observed pixel falls at a particular location, those multitemporal pixel values are averaged to fill the HR grid location. Furthermore, the work in [22] uses vector quantization on the pixel intensities to select the weights used for each observation window, in addition to the specific spatial distribution of samples. Empirical models for the required correlations are used in [22] based on training data. In contrast, the work in [14] uses an unquantized HR grid and theoretical parametric correlation models. This eliminates the need for training data and reduces error due to sample position quantization. Local estimates of the signal variance are employed in [14] to allow the weights to adapt to both the spatial distribution of samples and the local signal-to-noise ratio. The vector quantization in [22] and locally adaptive SNR in [14] have been observed to be beneficial in moderate to low SNR application. In high SNRs, these extra steps may not be necessary and can be avoided to reduce computational complexity.

The central challenge with the AWF SR method lies in determining the weights. The approach for designing the filter weights for the AWF SR method is based on a finite impulse response Wiener filter. Using correlation models, weights that minimize the mean squared error (MSE) are found. The approach is adaptive in the sense that as the spatial distribution of samples in the HR grid vary with the position of the observation window, so do the weights. Furthermore, spatially varying local statistics may be used to modify the correlation model for each observation window position. Note that for pure global translational motion, the spatial distribution of samples on the HR grid is periodic. This means that the number of filter weights required is relatively small. Since computing the weights is the bulk of the computational load, such imagery can be processed very fast with the AWF SR algorithm, even with a non-quantized HR grid [14]. For non-translational motion, a potentially unique spatial pattern can be seen for each observation window. While it is possible to calculate all of these on-the-fly, this is a high computational burden. The work in [15] has proposed a modified version of the AWF SR algorithm that uses a specially selected partial observation window applied to a quantized HR grid. The partial observation window and quantized HR grid reduces the number of distinct spatial patterns observed, reducing the number of weights to be computed. With a reduced number of weight vectors, these can all be precomputed prior to processing video. This allows the AWF SR algorithm to process frames with very little computational load (given the precomputed weights), even for non-translational motion.

In this chapter, we review the AWF SR methods. We discuss the observation models used including motion models and system PSF models. Also, a number of experimental results are presented to demonstrate the performance of the AWF SR methods. The organization of the remainder of the chapter is as follows. Section 1.2 presents the relevant observation models for the AWF SR methods. The AWF SR algorithms are presented in Section 1.3. Experimental results are provided in Section 1.4 and conclusions are presented in Section 1.5.

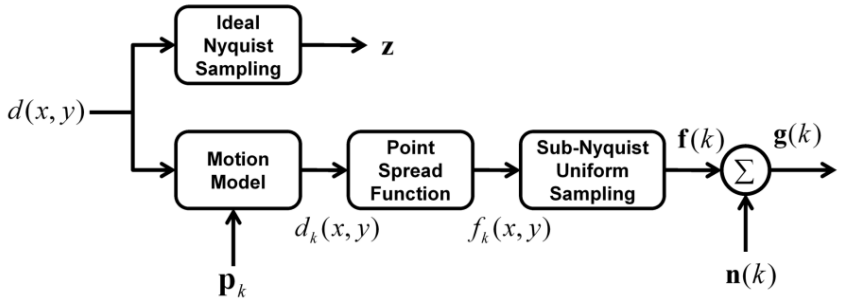


FIGURE 1.1
Observation model relating the desired 2-D continuous scene, $d(x, y)$, and the observed LR frames.

1.2 Observation Model

In this section we begin with the overall image formation model. We then discuss the motion model, registration, and finally the system PSF model.

1.2.1 Image Formation Model

The low resolution (LR) image formation model is shown in Figure 1.1. The model is used for many SR algorithms including those in [14, 22]. The model begins with a desired 2-D continuous scene, $d(x, y)$. Here this desired image is assumed to be geometrically aligned with one of the observed frames, referred to as the reference frame. A geometric transformation is used to account for any motion between acquired frames, $d_k(x, y) = T_{\mathbf{p}_k}\{d(x, y)\}$, for $k = 1, 2, ..., P$, and the reference frame. Note that the transformation depends on the motion model parameters in $\mathbf{p}_k$. Details of the motion model are presented in Section 1.2.2. Blurring from the system PSF [14, 17] occurs next in the observation model, yielding $f_k(x, y) = d_k(x, y) * h(x, y)$, where $h(x, y)$ is the system PSF. More will be said about the PSF in Section 1.2.4. The FPA in the camera serves to sample the scene for each frame yielding the vector of samples denoted $\mathbf{f}(k)$ for frame k. Here we shall assume that the detector pitch is not sufficiently small to meet the Nyquist criterion, hence the need for multiframe SR. Finally, additive noise corrupts the samples yielding $\mathbf{g}(k) = \mathbf{f}(k) + \mathbf{n}(k)$, where $\mathbf{n}(k)$ contains the additive noise samples. Note that ideal sampling of the scene would give rise to the ideal image, represented here in lexicographical form as the vector $\mathbf{z} = [z_1, z_2, \ldots, z_K]^T$. An equivalent and entirely discrete observation model can be found to relate $\mathbf{z}$ to the observed frames $\mathbf{g}(k)$ using an impulse invariant PSF and downsampling [17].

TABLE 1.1
Undersamping factors for two imaging systems considered here.

Imaging System	Optics	Cutoff Frequency (ρ_c)	Detector Pitch	Sampling Frequency	Under-sampling
1 Amber	F/3	83.3 cyc/mm	50 μm	20 cyc/mm	8.33×
2 L-3	F/3	83.3 cyc/mm	20 μm	50 cyc/mm	3.33×

Let is now consider undersampling in the observation model. First note that the optics serve to bandlimit the image in the focal plane. For example, the radial cut-off frequency associated with PSF from diffraction-limited optics with a circular exit pupil is [13] $\rho_c = \frac{1}{\lambda N}$, where N is the f-number of the optics and λ is the wavelength of light. To avoid aliasing, the FPA must sample at more than twice this cut-off frequency. This dictates that to avoid aliasing, the detector pitch must be less than $\frac{\lambda N}{2}$. Consider two imaging systems in Table 1.1 that will be used for experimental results in this chapter. System 1 uses a 256×256 Amber FPA with detector pitch of 50μm and the optics have an f-number of $N = 3$. System 2 uses a 20μm pitch 640×512 FPA from L-3 Cincinnati Electronics also equipped with $N = 3$ optics. Both systems produce 14 bit data. As can be seen in Table 1.1, both systems allow for the acquisition of aliased imagery with the selected optics. This is very common in imaging system design, since the desire for wide field-of-view small f-number optics often outweighs concerns over aliasing. Thus, such systems may be considered detector limited. Multiframe SR methods are a good choice for resolution enhancement for such systems, given that the pixel motion can be reliably estimated with subpixel accuracy.

Given this observation model in Fig. 1.1, one approach to SR is to treat this as an inverse problem. This generally leads to iterative image reconstruction SR algorithms that seek to find a $\mathbf{z}$ that would give rise to the observed $\mathbf{g}(k)$ when put through a discrete observation model and be consistent with prior statistical models [4, 7–9, 16–18, 21, 24, 31, 39]. However, these iterative approaches can be computationally costly. A simpler and often faster class of SR algorithms is based on nonuniform interpolation [1, 2, 10–12, 19, 23, 26, 28, 30, 33–36]. To understand these, consider the case where the motion model and the PSF degradation processes commute [11]. For such cases, consider switching the order of the motion model and PSF in the block diagram in Fig. 1.1. Now the motion model and uniform sampling blocks are back-to-back. These can equivalently be combined into a single nonuniform sampling block. That is, motion followed by sampling is equivalent to simply altering the sampling locations according to the motion. Such an alternative observation model is shown in Fig. 1.2. This figure includes a block diagram along with a representation of the imagery at various stages in the observation model.

Using registration, the LR pixels can be placed on a common HR grid

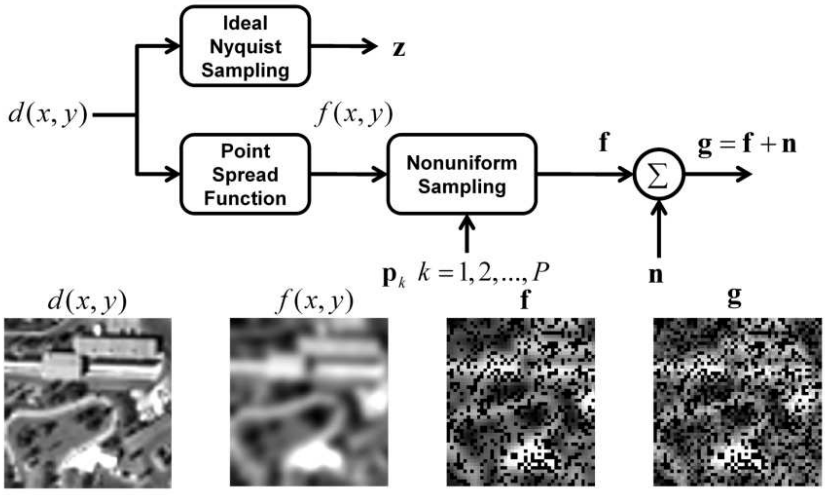


FIGURE 1.2
Alternative observation model relating a desired 2-D continuous scene, $d(x, y)$, with a set of corresponding LR frames. This model is appropriate when the motion model and PSF commute.

represented by $\mathbf{g}$ in Fig. 1.2. Nonuniform interpolation can be used to estimate a uniform grid of samples of $f(x, y)$ at or above the Nyquist rate. Finally, image restoration can be applied to reduce noise and reduce the blur caused by the system PSF, yielding an estimate of $\mathbf{z}$. Note that the motion model and PSF operations commute for translational motion due to the shift invariance property of convolution. For a circularly symmetric PSF, it can be shown that they also commute for rotation. With other types of motion, the PSF and motion do not necessarily commute. In such cases, the observation model in Fig. 1.2 does not strictly apply and hence the interpolation-restoration SR approaches which are based on Fig. 1.2 may be less effective. However, we have observed that useful results can often be obtained for other types of motion models using the interpolation-restoration approaches. In this chapter, we compare the performance of several interpolation-restoration SR approaches applied to affine motion. Note that for high levels of affine zoom or skew, one might expect some degradation in performance because such motion does not commute with the PSF blur. With a highly non-circularly symmetric PSF, we might also expect to see problems with rotational motion.

1.2.2 Image Motion Model

Consider a static 3-D scene and a moving camera acquiring video for SR. The resulting 2-D optical flow models for various scenarios are summarized in Table

TABLE 1.2
2-D optical flow models based on relative motion between a 3-D rigid scene and camera [38].

Flow Type	Model	When Applicable
Affine	$v_x(x,y) = p_1x + p_2y + p_3$ $v_y(x,y) = p_4x + p_5y + p_6$	Planar scene with orthographic projection
Quadratic (8)	$v_x(x,y) \approx p_1y + p_2xy + p_3x^2 + p_4$ $= \omega_Z y + \frac{\omega_X xy}{l} - \frac{\omega_Y x^2}{l} - \omega_Y l$ $v_y(x,y) \approx p_5x + p_6xy + p_7y^2 + p_8$ $= -\omega_Z x - \frac{\omega_Y xy}{l} + \frac{\omega_X y^2}{l} + \omega_X l$	Approx. for perspective projection projection with camera angle variation only. ω_X, ω_Y, ω_Z are camera angles in radians, l is focal length
Quadratic (10)	$v_x(x,y) \approx$ $p_1x + p_2y + p_3x^2 + p_4xy + p_5$ $v_y(x,y) \approx$ $p_6x + p_7y + p_8y^2 + p_9xy + p_{10}$	Approx. for planar scene with full perspective projection
Planar Projective	$v_x(x,y) = \frac{p_1+p_2x+p_3y}{p_7+p_8x+p_9y} - x$ $v_y(x,y) = \frac{p_4+p_5x+p_6y}{p_7+p_8x+p_9y} - y$	Exact for planar scene with full perspective projection

1.2 [38]. Note that for a planar scene and orthographic projection, the 2-D flow is affine. However, for the more realistic perspective projection case, but still with a planar scene, the flow is nonlinear and can be approximated with a quadratic flow model. For non-planar scenes and arbitrary camera motion, the 2-D motion is dependent on the specific 3-D scene geometry. In most cases the scene 3-D geometry is not known and estimating it is an extremely demanding problem in its own right. Furthermore, occlusion effects and motion parallax that can lead to discontinuities in the 2-D flow for general camera motion and non-planar scenes. These factors make accurate subpixel registration of multiple frames to a common grid for SR a very difficult task. It is interesting to note that when no translational camera motion is present (i.e., stationary camera with angular camera pointing motion only), a simple quadratic model is a good approximation, even for a non-planar scene. In this camera angle variation only scenario, there are no motion parallax or occlusion effects to contend with and the 2-D optical flow is not scene dependent. From Table 1.2, it can be seen that for relatively small image regions, the resulting 2-D flow can be approximated by a affine model for the camera angle variation only case. This is particularly true for longer focal length optics (as the nonlinear terms are divided by the focal length). Thus, perhaps one of the most favorable acquisition scenarios for video for SR involves a camera at a fixed location panning and/or rotating from that fixed position relative to a static scene.

In this chapter, we focus on an affine motion model. Although it does not fully capture all of types of motion in Table 1.2, it can often serve as a useful approximation and has a number of useful properties. In particular, it only has 6 parameters to estimate and multiple sequential affine transformations is still an affine transformation. Thus, we can register each frame to the previous frame and then accumulate these to reference all the frames to a common

frame or HR grid. Iterative and multiscale registration approaches also benefit from the accumulation property of affine flow.

1.2.3 Image Registration

For the experimental results presented here, we use a global gradient-based least-squares algorithm for estimating the affine parameters [5, 17, 27]. To define the affine registration algorithm, consider relating a new frame $d(x, y)$ to a prior frame $\tilde{d}(x, y)$ through 2-D optical flow. Neglecting occlusion effects and noise this is given by

$$d(x, y) = \tilde{d}(\tilde{x}, \tilde{y}) = \tilde{d}\left(x + v_x(x, y), y + v_y(x, y)\right), \tag{1.1}$$

where $v_x(x, y)$ and $v_y(x, y)$ are the polynomial optical flow functions. A truncated Taylor series approximation for small motions allows us to express this as

$$d(x, y) \approx \tilde{d}(x, y) + v_x(x, y)g_x(x, y) + v_y(x, y)g_y(x, y). \tag{1.2}$$

Now that we have removed the polynomial functions $v_x(x, y)$ and $v_y(x, y)$ from the argument of $\tilde{d}(\cdot)$, we get a set of linear equations (one per pixel) that can be solved using least squares. If we assume affine flow, then $v_x(x, y) = p_1x + p_2y + p_3$ and $v_y(x, y) = p_4x + p_5y + p_6$. Because of the truncated Taylor series approximation, the least squares estimate is accurate only for small motions. To address this, an iterative approach is recommended. Here the initial registration parameters are found between two images using the method described above. Then one of the images is repositioned using interpolation according to the registration estimate and a new estimate is formed. This process repeats until the final incremental estimate is judged to be sufficiently small. The final registration estimate is formed by accumulating all of the incremental estimates. The repositioning at each iteration is always done directly from the original image (using the currently accumulated transform estimate) so as to avoid accumulating interpolation errors.

While this iterative method extends the useful range of the registration technique, it may still fail if the initial estimate moves the image in wrong direction. This can happen with very large motions between frames. So to deal with very large motions, a multiscale approach is recommended. Here the registration begins using low resolution versions of the two images and the iterative registration technique is applied. These registration parameters are used to initialize the registration at the next resolution level. This continues until registration at the full resolution is complete. Finally, for improved numerical stability, it is recommended that the x, y coordinates of the center of the image be represented as $0, 0$ when setting up the least squares equations.

The affine registration method described can be applied to the entire image. However, to deal with more complex motion, it might be beneficial to use a piecewise affine model. That is, break the image up into subimages and estimate affine parameters for each subimage. A practical way to do this and

deal with large motions is to first do a global affine registration with the entire image and then refine these estimates in subimages with local affine estimates. Note that as the size of the subimages gets smaller, fewer equations are used in the least squares estimate. Hence, the accuracy of the subpixel flow estimate can be expected to suffer. Thus, a tradeoff must be found to balance the the accuracy of the model parameter estimate (favoring larger subimages) with the accuracy of the flow model itself (favoring smaller subimages).

Note that deformable scene motion, or simply a non-static scene, greatly complicates the registration process [39]. Numerous additional registration parameters need to be estimated. Changing pose of objects within the scene as well as occlusion further complicate matters in the general motion case. This makes subpixel registration accuracy across the full image very difficult, if not impossible. Furthermore, because the image data being used for registration is aliased and noisy, highly overdetermined linear equations are generally needed to get an accurate estimate. Joint SR and registration approaches with global motion models have been proposed to improve registration in the presence of high levels of aliasing [6, 16, 29]. Handling various types of more complex scene motion is an ongoing area of research in multiframe SR.

1.2.4 System Point Spread Function

In this section, we address the modeling of the system PSF. We begin by modeling the optical transfer function (OTF) with three components as follows

$$H(u,v) = H_{\text{dif}}(u,v)H_{\text{abr}}(u,v)H_{\text{det}}(u,v), \tag{1.3}$$

where u and v are the horizontal and vertical spatial frequencies in cycles per milimeter. Diffraction limited optics contributes $H_{\text{dif}}(u,v)$, optical aberrations contribute $H_{\text{abr}}(u,v)$, and detector integration contributes $H_{\text{det}}(u,v)$. Other factors such as defocus, motion blur and atmospheric effects are not considered here, but could be included if they are deemed to be significant in a particular application. The blurring effects from diffraction limited optics with a circular pupil function are described by the following OTF [13]

$$H_{dif}(u,v) = \begin{cases} \frac{2}{\pi}\left[\cos^{-1}\left(\rho/\rho_c\right) - \left(\rho/\rho_c\right)\sqrt{1-\left(\rho/\rho_c\right)^2}\right] & \text{for } \rho < \rho_c \\ 0 & \text{else} \end{cases}, \tag{1.4}$$

where $\rho = \sqrt{u^2+v^2}$ and $\rho_c = \frac{1}{\lambda N}$. It is this function that provides the band-limiting of the continuous scene, and consequently determines the necessary detector spacing to prevent aliasing. Even very well designed optical systems are likely to have aberrations which alter this diffraction limited model. One such aberration model is given by the following OTF [32]

$$H_{abr}(u,v) = \begin{cases} 1-(W_{RMS}/0.18)^2(1-4(\rho/\rho_c-0.5))^2 & \text{for } \rho < \rho_c \\ 0 & \text{else} \end{cases}. \tag{1.5}$$

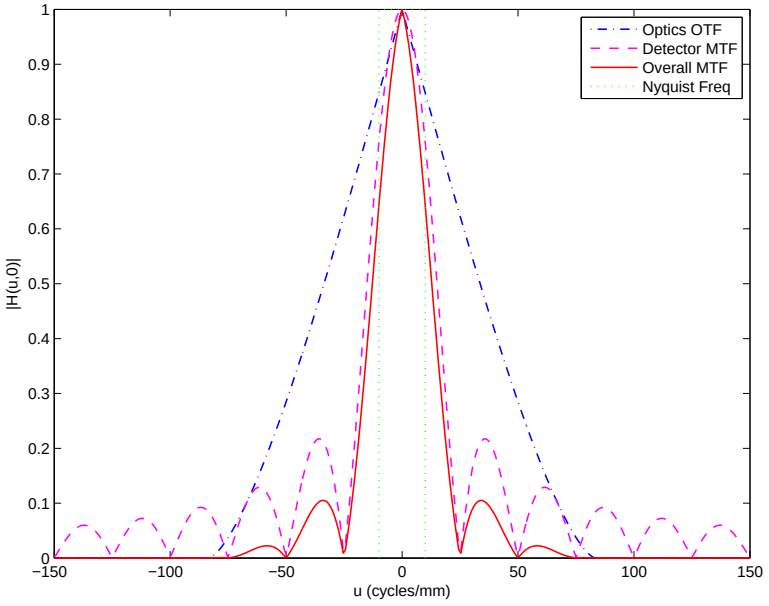
Note that for a well tuned system, $W_{RMS} = 1/14$ may be a good choice [32]. Finally, the detector component of the frequency response, $H_{\text{det}}(u, v)$, is obtained from the Fourier transform of the function describing the active area of an individual detector (assuming all detectors in the FPA have the same active area shape). Cross sections of the overall 2-D modulation transfer function (MTF) and its components are shown in Fig. 1.3 for the imaging systems in Table 1.1. Note that the MTF is simply the magnitude of the OTF. Here, we assume $\lambda = 4$ μm and $W_{RMS} = 1/14$. It can been seen that the detector MTF dominates the Amber system because of the relatively large detectors in the FPA. For the L-3 system, the optics dominates the overall MTF. Note that the Nyquist frequency in both cases is below the cut-off frequency of the optics. Any frequency content above the Nyquist frequency is "folded" into lower frequencies, creating aliasing artifacts. In addition to aliasing, another important thing to note from Fig. 1.3 is that, like most any imaging system, the MTF is not flat and high frequencies are attenuated (reducing image detail). The problem is that one cannot simply apply a high-boost filter to the imagery to restore the attenuated frequencies due to aliasing. However, if the effective sampling rate is increased by a multi-frame nonuniform interpolation process, such restoration is then possible. The system PSF can be found by taking the inverse Fourier transform of the optical transfer function. These are shown in Fig. 1.4. Note that the discrete impulse invariant PSF can be found by sampling this PSF at spacings of the detector pitch divided by the SR upsampling factor.

1.3 AWF SR Algorithms

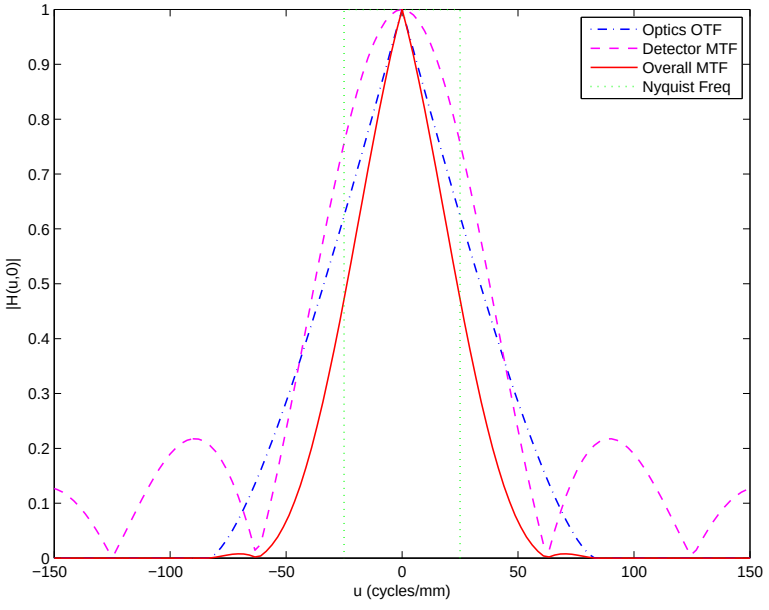
The AWF SR algorithms use $\mathbf{g}(k)$, for $k = 1, 2, \ldots, P$ to form an estimate of $\mathbf{z}$, denoted $\hat{\mathbf{z}}$. The SR algorithms can be applied to video using a temporal sliding window of frames, or it can be used to generate a single output from an input sequence. The effective sampling rate for the estimated image is defined to be L times greater than that of the observed imagery. Ideally, L would selected to meet the Nyquist criterion. However, often a lower value of L may be sufficient to provide a useful result with minimal aliasing. Inspection of the system MTF can be useful in making this selection.

The basic AWF SR algorithm is illustrated in the block diagram in Fig. 1.5. As mentioned above, registration allows us to create a nonuniform HR grid image denoted $\mathbf{g}$ in Fig. 1.5. A moving window filter spanning $W_x \times W_y$ HR pixels processes this HR grid to produce the final output. Let the pixel values spanned by the moving window at position i be denoted $\mathbf{g}_i = [g_{i,1}, g_{i,2}, \ldots, g_{i,K_i}]^T$, where K_i is the number of LR pixels within the i'th observation window.

For each observation window, an estimate of the pixel at the center of the

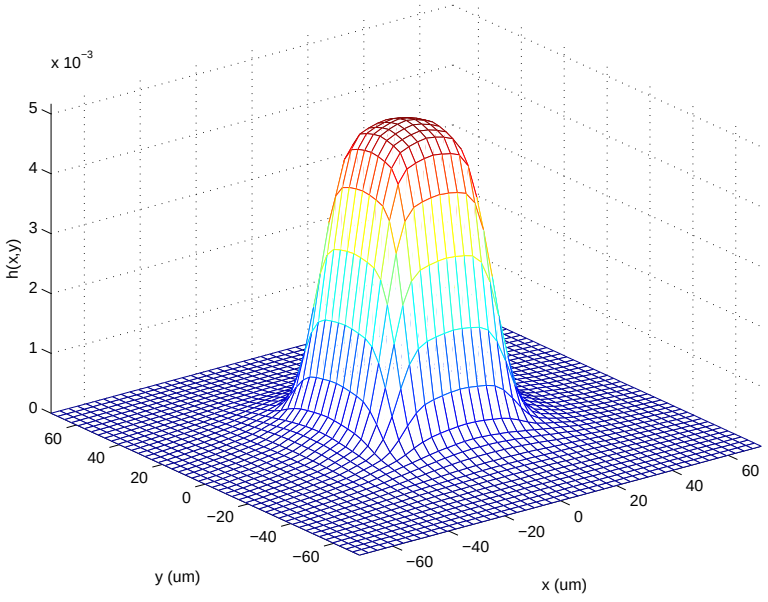


(a)

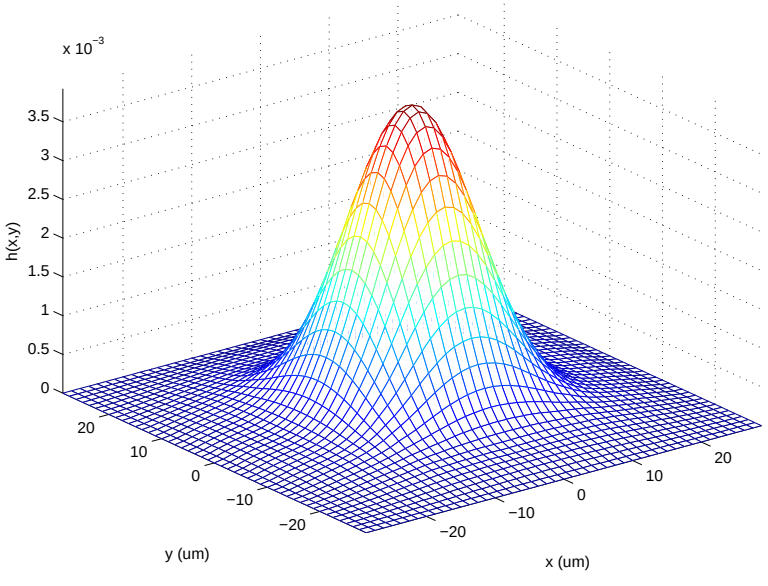


(b)

FIGURE 1.3
Cross sections of the overall theoretical 2-D MTF and its components with $\lambda = 4$ μm and $W_{RMS} = 1/14$ for (a) Imaging System 1 (Amber FPA) in Table 1.1 and (b) Imaging System 2 (L-3 FPA) in Table 1.1.



(a)



(b)

FIGURE 1.4
Theoretical PSF with $\lambda = 4$ μm and $W_{RMS} = 1/14$ for (a) Imaging System 1 (Amber FPA) in Table 1.1 and (b) Imaging System 2 (L-3 FPA) in Table 1.1.

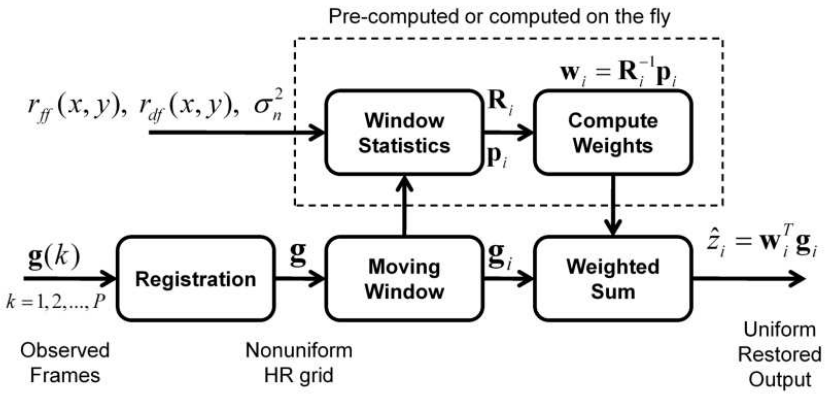


FIGURE 1.5
Overview of the proposed SR algorithm.

observation window is formed as a weighted sum. This is expressed as

$$\hat{z}_i = \mathbf{w}_i^T \mathbf{g}_i, \tag{1.6}$$

where $\hat{z}_i$ is the estimate of the i'th pixel in the desired image $\mathbf{z}$ and $\mathbf{w}_i$ is a K_i by 1 vector of weights. It is also possible to use the samples in one observation window to estimate multiple output pixels [14, 15, 22]. Consequently the observation window would move by multiple pixel positions at a time. This can have computational advantages including needing to compute or lookup fewer weight vectors. The minimum mean squared error weights are found using the well known Wiener solution

$$\mathbf{w}_i = \mathbf{R}_i^{-1}\mathbf{p}_i, \tag{1.7}$$

where $\mathbf{R}_i = E\{\mathbf{g}_i\mathbf{g}_i^T\}$ and $\mathbf{p}_i = E\{z_i\mathbf{g}_i^T\}$. The weights are normalized so that they sum to 1 to eliminate potential artifacts from variable DC response of adaptive filter.

The required statistics are found empirically in [22] based on training images. A quantized HR grid and a full autocorrelation matrix and crosscorrelation vector can be estimated from fully populated observation windows at the HR grid resolution. Then, when a partially populated observation windows are encountered, the full autocorrelation matrix and crosscorrelation vector can be subsampled as needed. This allows us to compute the weights using (1.7). The method in [22] also goes one step farther, and uses vector quantization to partition on observation space and estimates a full autocorrelation matrix and crosscorrelation vector for each partition [3, 33]. These statistics are tuned to specific structures such as edges, lines and flat regions. This is one way to deal with the non-stationarity of most image data for image restoration [3, 33]. The work in [14, 15] employs a parametric autocorrelation model

for the desired underlying image and generates all of the necessary statistics from that. Both a global and spatially varying model are considered in [14]. The spatially varying model, like the vector quantization approach, seeks to treat the non-stationary nature of the image data. Both the vector quantization approach in [22] and the spatially varying approach in [14] are beneficial with moderate and high levels of noise. Under relatively high signal-to-noise ratio conditions, this added complexity generally does not improve performance [20]. Thus, for high SNR environments, a simple wide sense stationary (WSS) auto-correlation model may be the best choice.

The modeling of the autocorrelation matrix and crosscorrelation vector is described in detail in [14, 22]. However, we repeat the key steps of the WSS model from [14] for convenience. First let $\mathbf{g}_i = \mathbf{f}_i + \mathbf{n}_i$, where $\mathbf{f}_i$ is the noise-free version of the i'th observation vector $\mathbf{g}_i$ and $\mathbf{n}_i$ is a random noise vector. Assuming a zero-mean uncorrelated noise vector with independent and identically distributed elements of variance σ_n^2, the autocorrelation matrix for the observation vector is given by

$$\mathbf{R}_i = E\{\mathbf{g}_i \mathbf{g}_i^T\} = E\{\mathbf{f}_i \mathbf{f}_i^T\} + \sigma_n^2 \mathbf{I}. \tag{1.8}$$

The cross-correlation vector is given by

$$\mathbf{p}_i = E\{z_i \mathbf{g}_i^T\} = E\{z_i \mathbf{f}_i^T\}. \tag{1.9}$$

Continuing to follow the analysis in [14], let us now assume a WSS autocorrelation function, $r_{dd}(x, y)$, for the desired image $d(x, y)$. The cross-correlation function between $d(x, y)$ and $f(x, y)$, as shown in Fig. 1.2, can be expressed in terms of $r_{dd}(x, y)$ [37] as

$$r_{df}(x, y) = r_{dd}(x, y) * h(x, y). \tag{1.10}$$

The autocorrelation of $f(x, y)$ is given by

$$r_{ff}(x, y) = r_{dd}(x, y) * h(x, y) * h(-x, -y). \tag{1.11}$$

Evaluating (1.11) based on the distances between the samples in $\mathbf{g}_i$ yields $E\{\mathbf{f}_i \mathbf{f}_i^T\}$. Incorporating the noise term to this result yields $\mathbf{R}_i$ as expressed in (1.8). Similarly, evaluating (1.10) based on the distances between the samples in $\mathbf{g}_i$ and the desired sample position gives us $\mathbf{p}_i$ from (1.9). The desired image autocorrelations, $r_{dd}(x, y)$, can be obtained empirically from statistically representative training images or defined using a parametric model. Here we use the same circularly-symmetric parametric model as that used in [14]. This model is given by

$$r_{dd}(x, y) = \sigma_d^2 \rho^{\sqrt{x^2+y^2}}, \tag{1.12}$$

where σ_d^2 is the variance of the desired image and ρ controls the decay of the autocorrelation with distance.

It is interesting to observe how the AWF SR weights vary with the spatial distribution of samples within the partial observation window. Figure 1.6

shows weights for distinct spatial distributions of samples. For these results we have selected $L = 3$, $\rho = 0.75$, $\frac{\sigma_d^2}{\sigma_n^2} = 100$ and a PSF computed for Imaging System 2 in Table 1.1. The weights are shown with the same colomap where middle gray is 0. It can be seen that these minimum MSE weights change in non-trivial ways as the spatial distribution of samples changes. Note that Fig. 1.6(a) shows the case where only the reference frame pixels are present in the observation window. This can occur when using only a single frame or when no motion in that area of the image is present. Here the AWF SR is effectively performing single frame interpolation and restoration. In Fig. 1.6(d), the observation window is fully populated. In this case, with AWF weighting is equivalent to that of a standard FIR Wiener filter operating on the HR grid.

The number of possible spatial patterns in a given observation window on the quantized HR grid is 2^b, where $b = W_x W_y - \frac{W_x W_y}{L^2}$. For the 15×15 window shown in Fig. 1.6, this amounts to 2^{200} patterns. For an unquantized HR grid, there are an infinite number of possible patterns. Thus, it is generally impractical to precompute all of the weights prior to processing video, even on a quantized HR grid. However, in [15], a partial observation window is proposed that includes the uniform reference frame samples within the $W_x \times W_y$ window plus the M closest samples to the output position. This is illustrated in Fig. 1.7 where the positions outlined form the partial observation window. The M closest samples to the output are highlighted. Note that the reference pixels are guaranteed to be present, whereas the other samples may or may not be present, depending on the motion between frames. An upper bound on the number of filter weights to be precomputed using this partial observation window is given by $L^2 2^M$ [15]. Thus, with a suitable choice of M, it is possible to precompute and store all of the weights. This makes processing frames very fast as will be seen in Section 1.4.

1.4 Experimental Results

In this section, a number of experimental results are presented where we compare the performance of the AWF SR methods to several other benchmark techniques. All of the SR methods applied are listed in Table 1.3. We use simulated data for quantitative analysis and real infrared imagery for subjective analysis in a true application.

1.4.1 SR Results for Simulated Data

The simulated LR frames are generated with different types of affine motion. We use 8 bit images with $L = 3$, the PSF for System 2 in Table 1.1, and a noise standard deviation of 2. The SR is done with $P = 9$ LR frames of

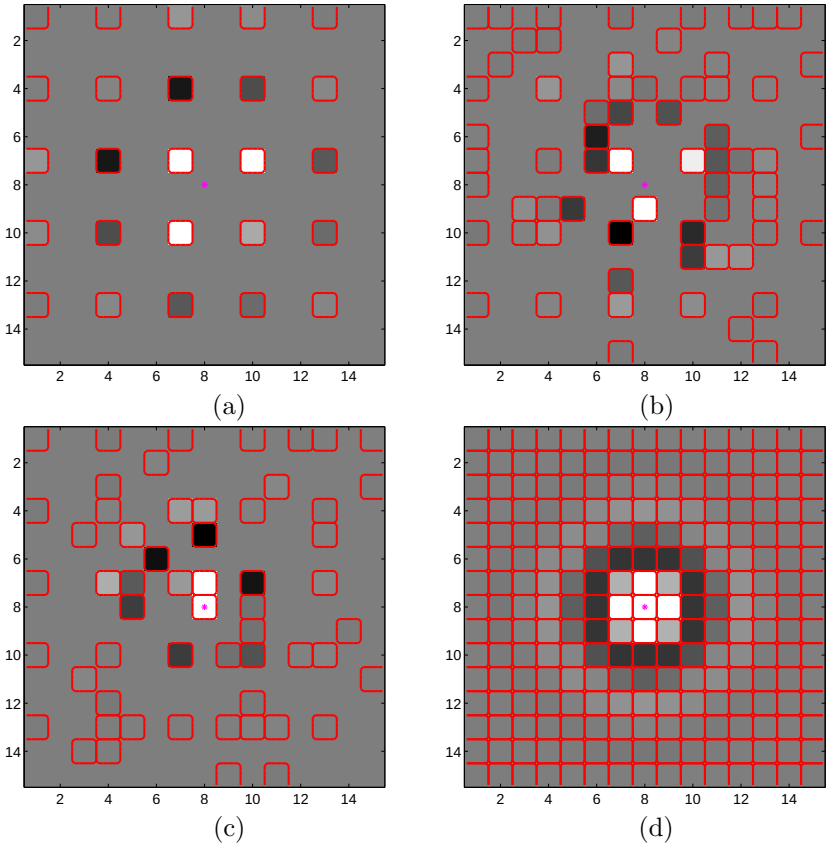


FIGURE 1.6
AWF SR filter weights for various spatial distribution of samples. Weights with only the reference frame samples (no motion) is shown in (a). Weights for a fully populated observation window are shown in (d). All weights are shown with the same colormap where middle gray is zero.

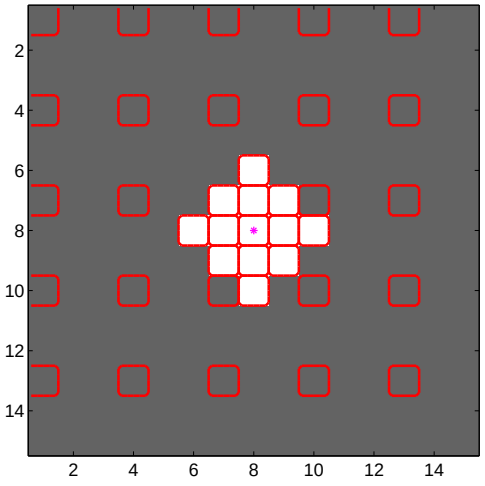


FIGURE 1.7
Partial observation window used by the Fast AWF SR algorithm in [15].

size 180×134. For the affine motion, the translation parameters are Gaussian and have a mean of zero and standard deviation of 2 LR pixel spacings. The rotation angle is Gaussian has a mean of zero and standard deviation of 5 degrees. The shear is horizontal only and the parameter is Gaussian with zero mean and a standard deviation of 0.05 and the zoom factor is Gaussian with a mean of 1 and standard deviation of 0.05. A three level multiscale affine registration is employed with 5 iterations using bicubic interpolation at each level. The image results are shown in Fig. 1.8 where the motion includes translation, rotation, shear and zoom. These image show a 250×250 region of interest (ROI) from the processed imagery. Fig. 1.8(a) shows the true HR image. The first (and reference) frame zoomed using bicubic interpolation is shown in Fig. 1.8(b). The partially populated HR grid is shown in Fig. 1.8(c). The SR outputs for all of the SR methods in Table 1.3 make up the rest of Fig. 1.8. The noise-to-signal ratios (NSRs) that provided the lowest MSE are used for each method. The AWF SR outputs use an NSR of 0.01. The WNN method uses an NSR of 0.04 and the Delaunay SR method uses an NSR of 0.02. The RLS method uses a regularization parameter of $\lambda = 0.01$ with 20 iterations [17].

Table 1.4 shows the MSE results for the SR methods along with the average run time (excluding registration). The processing was done on a Pentium 4 CPU with a clock speed of 2.8 GHz. Note that the registration of the 9 frames took 3.31 seconds. As noted in [15] this can be sped up using fewer levels and bilinear interpolation. When processing video, only one new frame needs to be registered to the previous frame to produce an output frame. This is because the affine parameters relating each frame to the reference frame can

TABLE 1.3
SR Algorithm Table

Name	Description
Fast AWF	Adaptive Wiener filter method using quantized HR spatial grid and partial observation window with precomputed weights [15]. Simultaneous nonuniform interpolation and restoration.
Full AWF	Adaptive Wiener filter method in [14] adapted for affine motion and using a discrete (quantized) HR spatial grid. Full observation window is used and the optimum weights are computed for each window on the fly. Simultaneous nonuniform interpolation and restoration.
WNN	Weighted nearest neighbor method in [2, 12] adapted for affine motion. Nonuniform interpolation done using an inverse distance based weighting of the nearest 4 neighbors. Restoration is done with an FFT based Wiener filter.
Delaunay	Based on the method in [19] adapted for affine motion. Nonuniform interpolation done using Delaunay triangulation. Restoration is done with an FFT based Wiener filter.
RLS	Regularized least squares interative SR method in [17] adapted for affine motion. This method does not assume the PSF blurring and motion models commute.

be determined by accumulating the incremental frame-to-frame registration parameters. It should be noted, however, that with approach, registration errors may also accumulate. Notwithstanding this, we have observed that this can generally be a very effective video registration method. Note that single frame bicubic interpolation output has, by far, the highest MSE. All of the multiframe SR methods produce a much lower MSE. The RLS SR method with 20 iterations produces the lowest MSE in this experiment, but at the cost of a high run time. The Fast AWF has the lowest run time with MSEs comparable to WNN and RLS with 5 iterations.

1.4.2 SR Results for Infrared Video Data

The first infrared dataset used for evaluating the SR methods is one obtain with Imaging System 1 in Table 1.1. These results are shown in Figure 1.9. The camera is mounted on a stationary tripod and is manually panned and

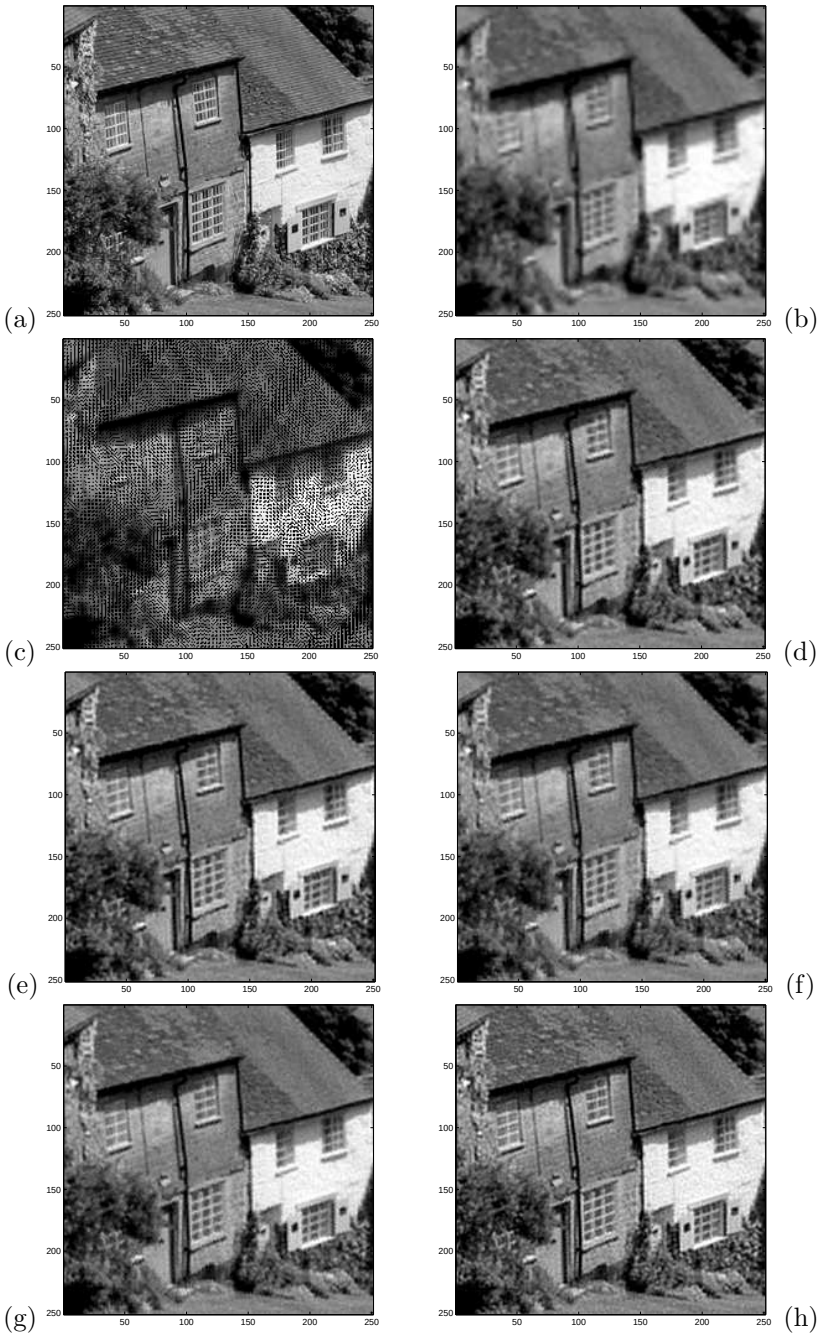


FIGURE 1.8
Output images for the simulated image sequence with $L = 3$ and $P = 9$. (a) Desired image (b) single frame bicubic interpolation, (c) partially populated high resolution grid after registration (d) output of the fast AWF method with partial observation window and precomputed weights, (e) full AWF method with weights computed on the fly (f) WNN with 4 nearest neighbors (g) Delaunay triangulation based output (h) RLS method.

TABLE 1.4
MSE for the various SR algorithms with affine motion ($L = 3$ and $P = 9$).

Algorithm	Translation	Rotation	Shear	Zoom	All	Time (s)
Bicubic	253.64	253.64	253.64	253.64	253.64	0.357
Fast AWF	141.20	157.25	185.58	158.46	152.23	0.811
Full AWF	127.36	143.65	181.50	144.51	139.26	57.238
WNN	141.93	162.82	211.37	163.95	164.85	1.784
Delaunay	121.36	137.44	191.06	141.20	144.29	21.105
RLS (20 Iterations)	107.40	118.20	167.74	120.40	114.55	275.337
RLS (5 Iterations)	141.32	151.64	182.73	150.88	149.50	67.204

rotated acquiring a 30 Hz video sequence. The SR methods use $P = 20$ frames and form output images of size 256×256 with $L = 4$. Bicubic interpolation of the first (reference) frame is shown in Fig. 1.9(a). The partially populated HR grid is shown in Fig. 1.9(b). The output for the Fast AWF, Full AWF, WNN and RLS methods are shown in Fig. 1.9(c)-(f), respectively. The NSR for the AWF methods is 0.05. The NSR for the WNN method is 0.1. The RLS method uses 10 iterations and $\lambda = 0.01$. The tuning parameters have been chosen based on subjective image quality. Because of the high level of aliasing for this imaging system, the aliasing artifacts are rather obvious in Fig. 1.9(a). Most prominently, the artifacts take on the form of jagged diagonal edges. Overall blurring of the image is also evident from the detector dominated PSF. It is clear that if single frame restoration is applied to this imagery, the aliasing artifacts would only be pronounced. The multiframe SR methods clearly reduce aliasing and sharpen the imagery. The RLS again appears to provide the best results. However the fast SR methods like the Fast AWF and WNN provide a significantly enhanced image (compared with bicubic interpolation) in a small fraction of the run time of the RLS method.

Finally, results obtained with data from Imaging System 2 in Table 1.1 are shown in Figure 1.10. This camera is also mounted on a stationary tripod and is manually panned and rotated acquiring a 30 Hz video sequence. The SR methods use $P = 9$ frames and form output images of size 230×220 with $L = 3$. Bicubic interpolation of the first (reference) frame is shown in Fig. 1.10(a). The partially populated HR grid is shown in Fig. 1.10(b). The output for the Fast AWF, Full AWF, WNN and RLS methods are shown in Fig. 1.10(c)-(f), respectively. The tuning parameters used on the previous dataset are also applied here. While the aliasing here is more subtle, the results appear to be consistent with those for the previous dataset.

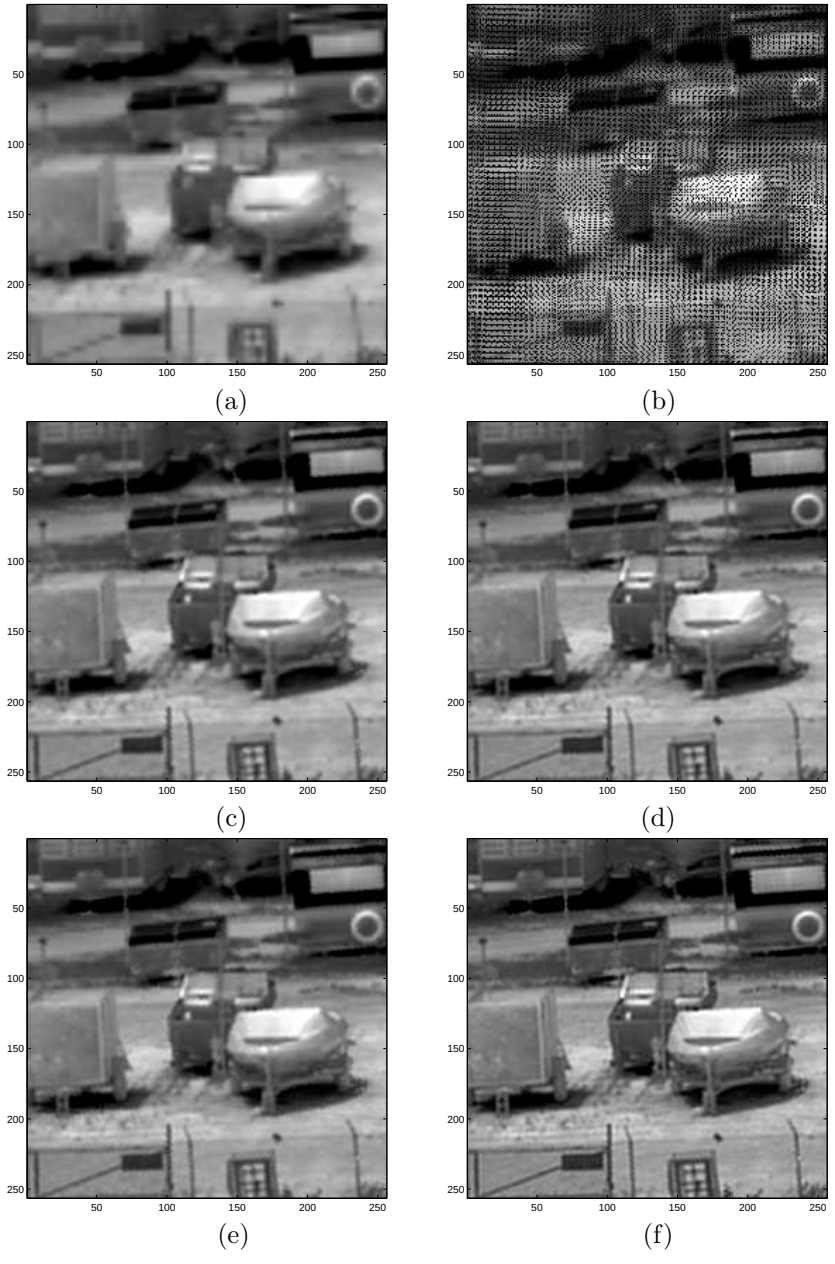


FIGURE 1.9
Outputs for image sequence from the tripod mounted Amber imager with $L = 4$ and $P = 20$. (a) Single frame bicubic interpolation, (b) partially populated high resolution grid (c) Fast AWF with partial observation window and precomputed weights, (d) Full AWF with weights computed on the fly, (e) WNN with 4 nearest neighbors, (f) RLS method.

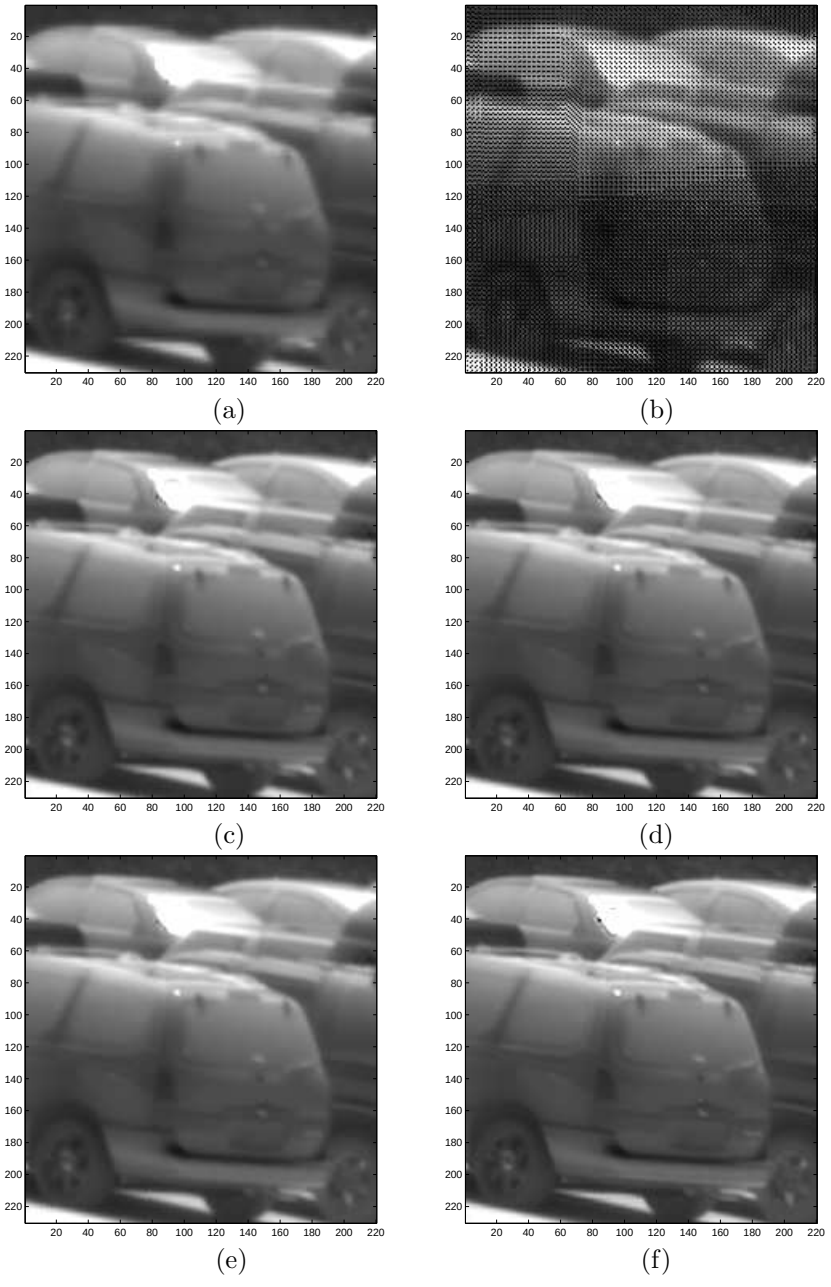


FIGURE 1.10
Outputs for image sequence from the tripod mounted L-3 imager with $L = 3$ and $P = 9$. (a) Single frame bicubic interpolation, (b) partially populated high resolution grid (c) Fast AWF with partial observation window and pre-computed weights, (d) Full AWF with weights computed on the fly, (e) WNN with 4 nearest neighbors, (f) RLS method.

1.5 Conclusions

The AWF SR methods are a type of nonuniform interpolation based SR algorithm. A distinctive feature of the AWF SR methods, however, is that the nonuniform interpolation and restoration are done simultaneously in a single weighted sum operation. The weights are determined based on FIR Wiener filter theory and they adapt to the specific spatial distribution of LR samples in the nonuniformly populated HR grid. Various methods have been explored in the literature for modeling the correlation statistics needed to determine the weights. Spatially varying statistics based on vector quantization [22] and local variance [14] have been used. The spatially varying statistics provide the most benefit in low signal-to-noise environments. In high signal-to-noise environments, like those considered here, a global statistical model is very effective and simpler to implement.

One of the main benefits of the AWF SR method is the potential for fast processing. Furthermore, the method is robust to the spatial distribution of LR pixels on the HR grid and it degrades gracefully towards the case where no motion is present (or only one frame is used). For translational motion, the number of distinct weight vectors needed for the Full AWF is small, and this allows for fast processing. The Fast AWF SR algorithm precomputes all of the weights for any motion model using a partial observation window [15].

The experimental results show that iterative SR methods, like the RLS, generally provide the best results. However these have a high computational complexity. The Fast AWF SR method is observed to have the shortest run time of any of the SR methods tested, with performance comparable to that of WNN and the RLS with 5 iterations. Delaunay SR was a notable performer in that the MSE was comparable to that of the Full AWF and it had a shorter run time. However, the run time for Delaunay SR was still much longer than that of the Fast AWF.

This chapter has also explored motion models for multiframe SR. It is noted that perhaps the most favorable conditions for SR are for a camera that is panning and tilting from a stationary location with a static scene. In that case, the 2-D optical flow is approximately quadratic, but can be effectively modeled as affine in small regions (or piecewise affine for large areas). With the stationary camera and static scene, no motion parallax, occulusion, or object pose variations are present and the optical flow is independent of the 3-D scene. Using the affine or piecewise affine model for this scenario, it is usually possible to form a highly overdetermined set of linear equations to solve for the motion parameters. With many equations and few motion parameters, it is often possible to get sufficiently accurate estimates, even in the presence of aliasing and noise.

1.6 Acknowledgments

The author would like to thank Dr. Doug Droege at L-3 Cincinnati Electronics for providing the L-3 infrared imagery used here. The Amber infrared images have been provided courtesy of the Air Force Research Laboratory at WPAFB in Dayton, OH.

Bibliography

[1] K. Aizawa, T. Komatsu, and T. Saito. Acquisition of very high resolution images using stereo camera. In *Proc. SPIE Visual Communication and Image Processing*, volume 1, pages 318–328, Boston, MA, Nov 1991.

[2] M. S. Alam, J. G. Bognar, R. C. Hardie, and B. J. Yasuda. Infrared image registration using multiple translationally shifted aliased video frames. *IEEE Instrum. Meas. Mag.*, 49(5), Oct 2000.

[3] K. E. Barner, A. M. Sarhan, and R. C. Hardie. Partition-based weighted sum filters for image restoration. *IEEE Trans. Image Processing*, 8(5):740–745, May 1999.

[4] B. Bascle, A. Blake, and A. Zisserman. Motion deblurring and super resolution from an image sequence. *EECV*, 2:573–581, Apr 1996.

[5] James R. Bergen, P. An, Th J. Hanna, and Rajesh Hingorani. Hierarchical model-based motion estimation. In *Proceedings of the Second European Conference on Computer Vision*, pages 237–252. Springer-Verlag, 1992.

[6] S. Cain, R. C. Hardie, and E. E. Armstrong. Restoration of aliased video sequences via a maximum-likelihood approach. In *Infrared Information Symposium (IRIS) on Passive Sensors, Monterey, CA*, volume 1, pages 377–390, March 1996.

[7] R. Chan, T. Chan, M.Ng, W. Tang, and C. Wong. Preconditioned iterative methods for high-resolution image reconstruction from multisensors. *Adv. Signal Process. Algorithms, Architectures, Implementations*, 3461:348–357, 1998.

[8] P. Cheeseman, B. Kanefsky, R. Kraft, J. Stutz, and R. Hanson. Super-resolved surface reconstruction from multiple images. NASA Tech. Rep. FIA-94-12, Dec 1994.

[9] T. Connolly and R. Lane. Gradient methods for superresolution. In *Proc. Int. Conf. Image Processing*, volume 1, pages 917–920, 1997.

[10] M. Elad and Y. Hel-Or. A fast super-resolution reconstruction algorithm for pure translational motion and common space invariant blur. *IEEE Trans. Image Processing*, 10(8):1187–1193, Aug 2001.

[11] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super-resolution. *IEEE Trans. Image Processing*, 13(10):1327–1344, Oct 2004.

[12] J. C. Gillette, T. M. Stadtmiller, and R. C. Hardie. Reduction of aliasing in staring infrared imagers utilizing subpixel techniques. *Opt. Eng.*, 34(11):3130–3137, Nov 1995.

[13] J.W. Goodman. *Introduction to Fourier Optics*. McGraw-Hill, 1968.

[14] R. C. Hardie. A fast super-resolution algorithm using an adaptive wiener filter. *IEEE Trans. Image Processing*, 16:2953–2964, Dec 2007.

[15] R. C. Hardie and K. J. Barnard. Fast adaptive wiener filter based super-resolution applied to affine motion. *Submitted to IEEE Transactions on Image Processing*, September 2009.

[16] 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. Image Processing*, 6(12):1621–1633, Dec 1997.

[17] R. C. Hardie, K. J. Barnard, J. G. Bognar, E. E Armstrong, and E. A. Watson. High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system. *Opt. Eng.*, 37(1):247–260, Jan 1998.

[18] M. Irani and S. Peleg. Improving resolution by image registration. *CHIP: Graph. Models Image Process.*, 53(3):231–239, May 1991.

[19] S. Lertrattanapanich and N. K. Bose. High resolution image formation from low resolution frames using delaunay triangulation. *IEEE Trans. Image Processing*, 11(12):1427–1441, Dec 2002.

[20] Y. Lin, R. C. Hardie, Q. Sheng, M. Shao, and K. E. Barner. Improved optimization of soft-partition-weighted sum filters and their application to image restoration. *Applied Optics*, 45(12):2697–2706, April 2006.

[21] S. Mann and R. Picard. Virtual bellows: construction of high quality stills from video. In *Proc. IEEE Int. Conf. Image Processing*, Austin, TX, Nov 1994.

[22] B. Narayanan, R. C. Hardie, K. E. Barner, and M. Shao. A computationally efficient super-resolution algorithm for video processing using partition filters. *IEEE Trans. Circuits Syst. Video Technol.*, 17(5):621–634, May 2007.

[23] N. Nguyen and P. Milanfar. A wavelet-based interpolation restoration method for superresolution. *Circuits, Syst., Signal Process.*, 19(4):321–338, Aug 2000.

[24] N. Nguyen, P. Milanfar, and G. Golub. A computationally efficient superresolution image reconstruction algorithm. *IEEE Trans. Image Processing*, 10(4):573–583, Apr 2001.

[25] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: A technical overview. *IEEE Signal Processing Mag.*, pages 21–36, May 2003.

[26] A. Patti, M.Sezan, and A. Teklap. Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time. *IEEE Trans. Image Processing*, 6(8):1064–1076, Aug 1997.

[27] S. Periaswamy and H. Farid. Elastic registration in the presence of intensity variations. *IEEE Transactions on Medical Imaging*, 22(7):865–874, 2003.

[28] T. Q. Pham, L. J. van Vliet, and K. Schutte. Robust fusion of irregularly sampled data using adaptive normalized convolution. *EURASIP Journal on Applied Signal Processing*, 2006(Article ID 83268):1–12, 2006.

[29] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman. Overcoming registration uncertainty in image super-resolution: Maximize or marginalize? *EURASIP Journal on Advances in Signal Processing*, 2007:Article ID 23565, 14 pages, 2007.

[30] K. Sauer and J. Allebach. Iterative reconstruction of band-limited images from non-uniformly spaced samples. *IEEE Trans. Circuits Syst.*, CAS-34:1497–1505, 1987.

[31] R. R. Schultz and R. L. Stevenson. Extraction of high-resolution frames from video sequences. *IEEE Trans. Image Processing*, 5:996–1011, June 1996.

[32] R. R. Shannon. Aberrations and their effect on images. In *Proceedings SPIE, Geometric Optics, Critical Review of Technology*, volume 531, pages 27–37.

[33] M. Shao, K. E. Barner, and R. C.Hardie. Partition-based interpolation for image demosaicking and super-resolution reconstruction. *Opt. Eng.*, 44:107003–1–107003–14, Oct 2005.

[34] H. Shekarforoush and R. Chellappa. Data-driven multi-channel superresolution with application to video sequences. *J. Opt. Soc. Amer. A*, 16(3):481–492, Mar 1999.

[35] J. Shi, S. E. Reichenbach, and J. D. Howe. Small-kernel superresolution methods for microscanning imaging systems. *Applied Optics*, 45(6):1203–1214, Feb 2006.

[36] A. Tekalp, M. Ozkan, and M. Sezan. High resolution image reconstruction from lower-resolution image sequences and space-varying image restoration. In *Proc. ICASSP '92*, volume 3, pages 169–172, San Fransisco, CA, Mar 1992.

[37] C. W. Therrian. *Discrete Random Signals and Statistical Signal Processing.* Prentice Hall, 1992.

[38] E. Trucco and A. Verri. *Introductory Techniques for 3-D Computer Vision.* Prentice Hall, 1998.

[39] T. R. Tuinstra and R. C. Hardie. High resolution image reconstruction from digital video by exploitation on non-global motion. *Opt. Eng.*, 38(5), May 1999.