Fast super-resolution with affine motion using an adaptive Wiener filter and its application to airborne imaging

Russell C. Hardie,[1,*] Kenneth J. Barnard,[2] and Raul Ordonez[1]

[1] *Department of Electrical and Computer Engineering, University of Dayton, 300 College Park, Dayton, Ohio 45469-0232, USA*
[2] *Air Force Research Laboratory, AFRL/RYMT, Building 620, 2241 Avionics Circle, Wright-Patterson AFB, Ohio 45433, USA*

*rhardie@udayton.edu

Abstract: Fast nonuniform interpolation based super-resolution (SR) has traditionally been limited to applications with translational interframe motion. This is in part because such methods are based on an underlying assumption that the warping and blurring components in the observation model commute. For translational motion this is the case, but it is not true in general. This presents a problem for applications such as airborne imaging where translation may be insufficient. Here we present a new Fourier domain analysis to show that, for many image systems, an affine warping model with limited zoom and shear approximately commutes with the point spread function when diffraction effects are modeled. Based on this important result, we present a new fast adaptive Wiener filter (AWF) SR algorithm for non-translational motion and study its performance with affine motion. The fast AWF SR method employs a new smart observation window that allows us to precompute all the needed filter weights for any type of motion without sacrificing much of the full performance of the AWF. We evaluate the proposed algorithm using simulated data and real infrared airborne imagery that contains a thermal resolution target allowing for objective resolution analysis.

© 2011 Optical Society of America

OCIS codes: (100.6640) Superresolution; (280.4991) Passive remote sensing; (110.3080) Infrared imaging.

References and links

1. S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," IEEE Signal Processing Mag. **20**, 21–36 (2003).
2. D. P. Capel, "Image mosaicing and super-resolution," Ph.D. thesis, University of Oxford (2001).
3. S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multi-frame super-resolution," IEEE Trans. Image Processing **13**, 1327–1344 (2004).
4. G. Rochefort, F. Champagnat, G. L. Besnerais, and J. franois Giovannelli, "An improved observation model for super-resolution under affine motion," IEEE Trans. Image Processing **15**, 3325–3337 (2006).
5. R. D. Fiete, "Image quality and λ FN/ p for remote sensing systems," Optical Engineering **38**, 1229–1240 (1999).
6. 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**, 621–634 (2007).
7. R. C. Hardie, "A fast super-resolution algorithm using an adaptive wiener filter," IEEE Trans. Image Processing **16**, 2953–2964 (2007).
8. F. O. Baxley, K. J. Barnard, R. C. Hardie, and M. A. Bicknell, "Flight test results of a rapid step-stare and microscan midwave infrared sensor concept for persistent surveillance," in *Proceedings of MSS Passive Sensors*, (Orlando, FL, 2010).
9. M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," IEEE Trans. Image Processing **6**, 1646–1658 (1997).

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011

10. 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," Optical Engineering **37**, 247–260 (1998).
11. E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision* (Prentice Hall, 1998).
12. B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *International Joint Conference on Artificial Intelligence*, 674–679 (Vancouver, 1981).
13. J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," in *Proceedings of the Second European Conference on Computer Vision*, 237–252 (Springer-Verlag, 1992).
14. J. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, 1968).
15. R. A. Emmert and C. D. McGillem, "Multitemporal geometric distortion correction utilizing the affine transformation," LARS Technical Reports, Paper 114 (1973).
16. C. W. Therrian, *Discrete Random Signals and Statistical Signal Processing* (Prentice Hall, 1992).
17. J. C. Gillette, T. M. Stadtmiller, and R. C. Hardie, "Reduction of aliasing in staring infrared imagers utilizing subpixel techniques," Optical Engineering **34**, 3130–3137 (1995).
18. M. S. Alam, J. G. Bognar, R. C. Hardie, and B. J. Yasuda, "Infrared image registration using multiple translationally shifted aliased video frames," IEEE Trans. Instrum. Meas. **49**, 915–923 (2000).
19. S. Lertrattanapanich and N. K. Bose, "High resolution image formation from low resolution frames using delaunay triangulation," IEEE Trans. Image Processing **11**, 1427–1441 (2002).

1. Introduction

Multiframe super-resolution (SR) has proven to be an effective post-processing method to reduce aliasing and enhance the resolution of images from detector limited imaging systems [1]. Provided that appropriate subpixel motion is present between frames, the sampling diversity provided by fusing multiple frames helps to overcome undersampling. Consequently, this allows us to apply restoration to reduce the blurring effects of the system point spread function (PSF). The most computationally simple multiframe SR methods are the interpolation-restoration methods [1]. These methods generally employ subpixel registration to position the observed low-resolution (LR) pixel values from multiple frames onto a common high-resolution (HR) grid. This results in a nonuniformly sampled HR image with sample locations determined by the interframe motion. Nonuniform interpolation is then employed to create a uniformly sampled HR image with reduced aliasing and a restoration step is employed to deconvolve the system PSF. A key assumption that justifies the use of these simple and fast nonuniform interpolation based SR methods is that the interframe motion model and PSF blurring operations commute. For translational motion this is known to be the case [2, 3], but it is not true in general [4]. This presents a potential problem for many applications such as airborne imaging where a translational warping model may be insufficient to properly account for the imaging geometry (as we shall show in Section 4.1), but the desire for fast SR processing remains.

Here we present a new Fourier domain analysis that explores the commutation of affine motion and a PSF that includes both detector integration and diffraction effects. We show that when diffraction from a circular exit pupil in the optics is considered, the system PSF and modulation transfer function (MTF) can become very nearly circularly symmetric and smooth. Consequently, we show that the error in commuting this kind of realistic blurring function with affine motion having limited zoom and shear is very small. We specifically analyze how the commutation error is impacted by the f-number of the optics which controls the level of diffraction effects. We show that the commutation error does tend to increase as the f-number and corresponding diffraction effects are reduced. However, for the range used in typical imaging system designs [5], we show the commutation error is very small. This novel analysis opens the door for a more informed, and we believe legitimate, use of fast interpolation-restoration SR methods on data from many imaging systems with affine motion.

Based on the commutation result, we present a new fast adaptive Wiener filter (AWF) SR method for non-translational motion and study its performance with affine motion. The new method is based on the interpolation-restoration AWF SR method proposed in [6, 7]. Like most

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA

interpolation-restoration SR methods, the AWF SR method registers a set of LR frames to a common HR grid. However, unlike standard methods, the AWF then performs nonuniform interpolation and the restoration simultaneously in a single weighted sum filter step. That is, each output HR pixel is formed using a weighted sum of neighboring observed pixels on the HR grid. By combining the nonuniform interpolation and restoration into a single step, computational advantages are achieved. Perhaps more importantly, however, is that the combined interpolation-restoration approach of the AWF allows the restoration to be optimized for the specific local spatial arrangement of the pixels on the HR grid. When nonuniform interpolation and restoration are done independently, artifacts from the nonuniform interpolation step caused by a poor distribution of samples on the HR grid can be amplified by the restoration step. Without any change in tuning or modeling parameters, the AWF gives a minimum mean squared error (MSE) result for any spatial arrangement of samples, including that obtained with no motion at all. Some other nonuniform interpolation methods tend to be more sensitive to the spatial distribution of samples afforded by the specific interframe motion. This can lead to highly variable output quality and especially poor results when camera motion stops.

Previous work on the AWF SR method [7] focused on translational motion. In that case, it is readily possible to precompute all of the needed filter weights because the pattern of observed pixel values on the HR grid is periodic and the number of unique filter weights is small. For non-translational motion, however, the full set of AWF weights is generally impractical to precompute. Calculating the AWF weights on-the-fly can be effective but is a big computational burden. The fast AWF SR method presented here employs a smart partial observation window designed using a new forward sequential selection procedure based on an MSE metric. In this way, we choose the most essential subset of pixels within the observation window to use. By limiting the number of samples, we are then able to precompute all the filter weights for fast processing. We show that by using the subwindow designed with the proposed method we gain speed and lose very little performance on the data used.

We study the performance of the new AWF SR method, along with several benchmark methods that have been adapted for affine motion, using airborne mid-wave infrared (MWIR) video sequences. These datasets are particularly enlightening as they contain a specially designed thermal resolution target on the ground [8]. This allows us the rare opportunity to objectively assess the resolution enhancement of the various SR methods with real airborne imagery. As far as we are aware, this paper represents the first such experimental study of SR methods for infrared airborne imaging using a ground resolution target. Here we also use real imagery from a visible camera as well as simulated data that allow for quantitative error analysis.

This paper proceeds with Section 2 where the observation model is presented. This includes the new analysis relating to the commutation of the motion and PSF models. The fast AWF SR algorithm, including the new subwindow selection method, is presented in Section 3. Experimental results are provided in Section 4 and conclusions are presented in Section 5.

2. Observation model

In this section, we begin with the physical observation model that relates a static 2-D ideal scene image to a set of LR observed frames. This first model will be referred to as the warp-then-blur model [4]. We then focus on the motion and PSF components of the model. We end with a subsection addressing the commutation of the motion and PSF components leading to a blur-then-warp observation model. It is this latter model that is the basis for the fast interpolation-restoration SR methods including the fast AWF.

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA

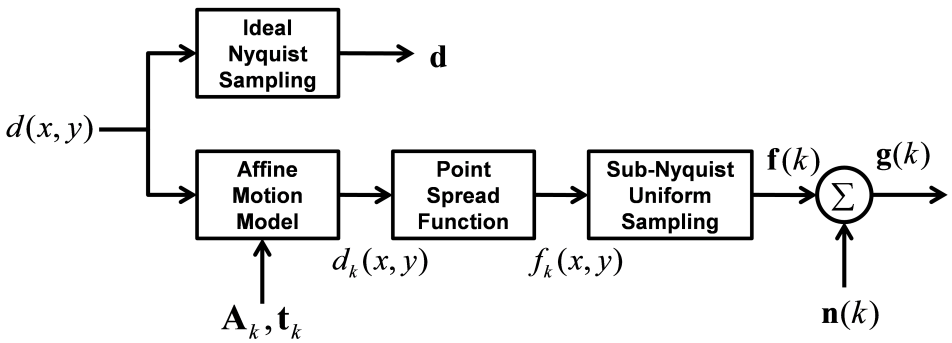


Fig. 1. Warp-then-blur observation model relating a desired 2-D continuous scene, $d(x,y)$, to a set of corresponding LR frames. This model follows the physical image acquisition process and is the basis for most of the iterative SR algorithms.

2.1. *Warp-then-blur image formation model*

The warp-then-blur observation model block diagram, similar to that presented in [7, 9], is shown in Fig. 1. This model follows the physical image acquisition process and begins with a desired 2-D continuous scene, $d(x,y)$. Affine warping is used to model the relative motion between the camera and scene, and the output is denoted $d_k(x,y)$. Next, blurring from the system PSF yields $f_k(x,y) = d_k(x,y) * h(x,y)$, where $h(x,y)$ is the system PSF and $*$ represents 2-D convolution. Here the PSF is modeled like that in [10] and details are provided in Section 2.3. Finally, the image is sampled below the Nyquist rate and corrupted with additive Gaussian noise. The result is a set of LR frames, represented in lexicographical vector notation as $\mathbf{g}(k)$, for $k = 1, 2, ..., K$. The ideally sampled image with no blur is represented using lexicographical notation as the vector $\mathbf{d}$. The model in Fig. 1 is often used as part of the more computationally demanding iterative SR methods including those in [1, 9, 10].

2.2. *Affine motion model and registration*

The motion model is an important component of any SR algorithm and its selection is application dependent. In depth treatment of the 2-D image motion produced as a result of relative motion between a 3-D rigid scene and camera can be found in [11]. For terrestrial imaging from a tripod mounted camera with small pointing angle variations, the image motion is well treated with a 2-D translational motion model. However, for airborne imaging, translation alone is often insufficient. The need for a more complex motion model becomes especially critical with lower frame rates and larger window sizes. An affine model may be a good choice as it allows for rotation, shear, and zoom, in addition to translation, with only 6 parameters. It is also convenient because multiple sequential affine motions remains affine. Thus, we can register each frame to the previous and then accumulate these into an affine transformation to relate all the frames to a common reference. This accumulation property is also helpful if one uses an iterative and/or multiscale registration method where parameters are updated at successive iterations and/or scales. Note that the affine motion model assumed here is designed as a background model. For low frame rates it may not accurately model moving objects or motion parallax effects. These are important issues that we intend to address in future work, but shall be left outside the scope of the current paper for the sake of focus and page length.

Let us now define the affine motion model and describe the registration method employed to estimate the affine model parameters. Let the affine parameters for each frame relative to a

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA

reference frame be designated

$$\mathbf{A}_k = \begin{bmatrix} A_{1,1}(k) & A_{1,2}(k) \\ A_{2,1}(k) & A_{2,2}(k) \end{bmatrix} \tag{1}$$

and $\mathbf{t}_k = [t_x(k), t_y(k)]^T$, for $k = 1, 2, \ldots, K$. Note that $\mathbf{t}_k$ contains the translational shift parameters and $\mathbf{A}_k$ contains information on rotation, zoom and shear. Using the model and notation $\mathbf{x} = [x, y]^T$ and $\tilde{\mathbf{x}}(k) = [\tilde{x}(k), \tilde{y}(k)]^T$, we have $\tilde{\mathbf{x}}(k) = \mathbf{A}_k\mathbf{x} + \mathbf{t}_k$ such that

$$d_k(\mathbf{x}) = d(\tilde{\mathbf{x}}(k)) = d\left(\mathbf{A}_k\mathbf{x} + \mathbf{t}_k\right). \tag{2}$$

To estimate the affine parameters we use a gradient-based least-squares algorithm based on that in [12]. To describe the LS affine registration algorithm, let us begin using a truncated Taylor series representation of $d_k(\mathbf{x})$ in terms of $d(\mathbf{x})$ and its gradients, $g_x(\mathbf{x})$ and $g_y(\mathbf{x})$, as follows

$$d_k(\mathbf{x}) \approx d(\mathbf{x}) + (\tilde{x}(k) - x)\, g_x(\mathbf{x}) + (\tilde{y}(k) - y)\, g_y(\mathbf{x}). \tag{3}$$

Note that we have one such equation for each pixel used and 6 unknowns embedded in $\tilde{\mathbf{x}}(k)$. Even though we use the observed LR frames for registration, we can generally estimate the parameters to provide sufficient subpixel accuracy because we have a highly overdetermined set of equations. The set of linear equations can be put in matrix form as $\mathbf{M}_k\mathbf{a}_k = \mathbf{b}_k$, where

$$\mathbf{M}_k = \begin{bmatrix} x_1 g_x(\mathbf{x}_1) & y_1 g_x(\mathbf{x}_1) & g_x(\mathbf{x}_1) & x_1 g_y(\mathbf{x}_1) & y_1 g_y(\mathbf{x}_1) & g_y(\mathbf{x}_1) \\ x_2 g_x(\mathbf{x}_2) & y_2 g_x(\mathbf{x}_2) & g_x(\mathbf{x}_2) & x_2 g_y(\mathbf{x}_2) & y_2 g_y(\mathbf{x}_2) & g_y(\mathbf{x}_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_N g_x(\mathbf{x}_N) & y_N g_x(\mathbf{x}_N) & g_x(\mathbf{x}_N) & x_N g_y(\mathbf{x}_N) & y_N g_y(\mathbf{x}_N) & g_y(\mathbf{x}_N) \end{bmatrix}, \tag{4}$$

$\mathbf{a}_k = [A_{1,1}(k), A_{1,2}(k), t_x(k), A_{2,1}(k), A_{2,2}(k), t_y(k)]^T$, and

$$\mathbf{b}_k = \begin{bmatrix} d_k(\mathbf{x}_1) - d(\mathbf{x}_1) + x_1 g_x(\mathbf{x}_1) + y_1 g_y(\mathbf{x}_1) \\ d_k(\mathbf{x}_2) - d(\mathbf{x}_2) + x_2 g_x(\mathbf{x}_2) + y_2 g_y(\mathbf{x}_2) \\ \vdots \\ d_k(\mathbf{x}_N) - d(\mathbf{x}_N) + x_N g_x(\mathbf{x}_N) + y_N g_y(\mathbf{x}_N) \end{bmatrix}. \tag{5}$$

Finally, the least squares parameters are given by

$$\hat{\mathbf{a}}_k = \left(\mathbf{M}_k^T \mathbf{M}_k\right)^{-1} \mathbf{M}_k^T \mathbf{b}_k. \tag{6}$$

Because the truncated Taylor series approximation is only accurate for small motions, we use this method iteratively [12] and at multiple scales [13]. The gradients needed above may be estimated with a Sobel operator and a smoothing prefilter may be applied prior to registration to reduce the impact of noise and aliasing [7]. Also, the results are most numerically stable when the center of the image is defined to be $x = y = 0$. Note that homogeneous coordinates can be used to conveniently combine sequential affine transformations or compute an inverse [11].

2.3. PSF model

The blurring in the observation can come from a number of sources. The diffraction from the optics and spatial integration from the detector elements are the main ever-present components. Other sources may include optical aberrations, defocus, and atmospheric turbulence. Here we follow the approach in [10] and model diffraction and detector integration. The optical transfer function (OTF) is given by

$$H(u, v) = H_{\text{dif}}(u, v) H_{\text{det}}(u, v), \tag{7}$$

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA

where u and v are the horizontal and vertical spatial frequencies in cycles per millimeter, $H_{\text{dif}}(u,v)$ is from the diffraction-limited optics, and detector integration is included as $H_{\text{det}}(u,v)$. Diffraction-limited optics with a circular pupil function gives rise to an OTF [14] given by

$$H_{dif}(u,v) = \begin{cases} \frac{2}{\pi}\left[\cos^{-1}(\omega/\omega_c) - (\omega/\omega_c)\sqrt{1-(\omega/\omega_c)^2}\right] & \text{for } \omega < \omega_c \\ 0 & \text{else} \end{cases}, \tag{8}$$

where $\omega = \sqrt{u^2+v^2}$ and $\omega_c = 1/(\lambda\mathscr{N})$, and $\mathscr{N}$ is the f-number of the optics. Note that $H_{\text{dif}}(u,v)$ is circularly symmetric and cone shaped. The detector component of the OTF is given by the Fourier transform of the active area of a single detector on the focal plane array (FPA). Assuming a rectangular active area, $H_{\text{det}}(u,v)$ will be a sinc function. The sampling frequency is given by $1/p$, where p is the detector pitch. Note that the f-number of the optics controls the cut-off frequency of the overall OTF and p controls the spatial sampling frequency. The Nyquist sampling theorem requires $1/p > 2\omega_c = 2/(\lambda\mathscr{N})$ to avoid aliasing.

Most optical systems are designed with some aliasing in an attempt to balance undersampling with a number of factors such as diffraction blurring, noise, field of view, and FPA size [5]. A useful metric for analyzing optical systems is $Q = \lambda\mathscr{N}/p$ [5]. When $Q = 2$, the detector array is sampling the diffraction blurred image on focal plane at the Nyquist rate. As Q is decreased (usually with f-number), the system gets increasingly undersampled. The work in [5] suggests that $Q \approx 1$ or lower may be a good choice, even though such systems are $2\times$ or more undersampled. It is this design trade-off that continues to give SR a potentially important role to play in modern imaging systems, even as detector elements and corresponding pixel pitches get smaller with advancing FPA technology. Another important observation to make is that as Q is decreased, the diffraction PSF component narrows while the detector integration PSF component remains the same. Thus, the overall OTF begins to look more like the detector sinc OTF and correspondingly less circularly symmetric. This symmetry breaking plays a role in our commutation analysis in Section 2.4

The MWIR imager used in Section 4 has a spectral bandwidth of $\lambda = 3-5$ μm (with the exception of the CO_2 absorption band) and we use $\lambda = 4$ μm for our PSF model. The system uses $F/2.3$ optics and has a pixel pitch of $p = 19.5$ μm. To model $H_{\text{det}}(u,v)$ we assume 100% fill factor rectangular detectors. This system is theoretically $4.24\times$ undersampled with a $Q = 0.472$. A lower upsampling factor may be used for SR in practice, taking into consideration noise and other non-ideal sensor characteristics. A cross-section of the 2-D modulation transfer function (MTF) and its components are shown in Fig. 2(a). Note that the folding frequency, $1/(2p)$, is well below the optical cut-off frequency, allowing for significant aliasing. The continuous PSF for the same system is shown in Fig. 2(b). The discrete impulse invariant PSF can be found by sampling the continuous PSF with a sampling period of below 4.6 μm. Note that even with this low Q, the PSF is still nearly circularly symmetric. As the f-number is increased, the PSF and OTF become more circularly symmetric. We also present results in Section 4 using an Imaging Source DMK 21BU04 visible camera. This is a 640×480 8-bit grayscale camera with a Sony ICX098BL CCD sensor with 5.6μm detectors. The camera is fitted with a Computar $F/4$ lens with a focal length of 5mm. Considering a central wavelength of $\lambda = 0.55$ μm, the visible system is theoretically $5.09\times$ undersampled with $Q = 0.393$.

2.4. Blur-then-warp model (nonuniform sampling model)

To develop a justification for the interpolation-restoration SR methods, we wish to begin by commuting the motion and PSF models in Fig. 1. The motion model and uniform sampling can then be combined into a nonunifrom sampling process operating on a blurred version of the desired image. This commuted model is shown in Fig. 3. Note that if **g** contains samples

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA

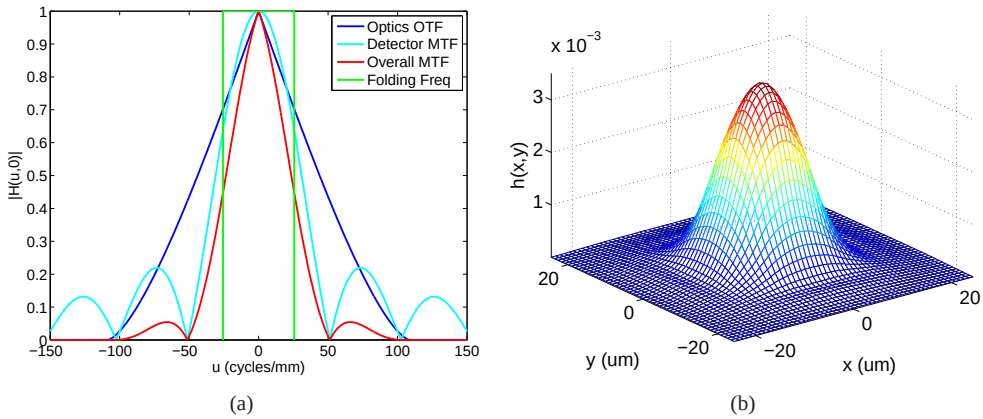


Fig. 2. (a) Cross-section of the MWIR system MTF and its components. (b) MWIR system PSF model.

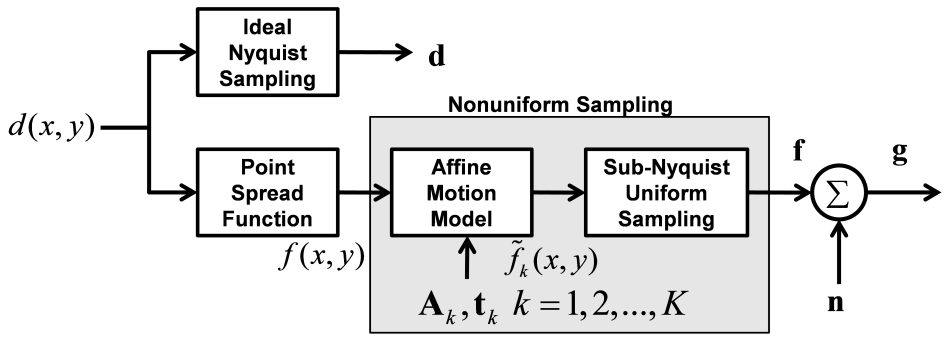


Fig. 3. Blur-then-warp observation model that performs nonuniform sampling of a single blurred image. This version of the observation model is valid when the motion model and PSF blurring commute. The fast interpolation-restoration SR methods (including the AWF) are based on this model.

of a nonuniformly sampled burred image, it is reasonable to perform nonuniform interpolation to create uniform sampling and then use restoration to deblur the resulting image. This is the reasoning behind the fast interpolation-restoration SR approaches. The AWF SR methods does this, only the nonuniform interpolation and restoration are done jointly in one weighted sum operation using minimum MSE weights.

The validity of blur-then-warp nonuniform sampling model relies on the commutation of the motion model and PSF blurring. The commutation property for translation is noted in [2, 3]. Later in [6, 7] it was noted that the commutation is also valid for rotational motion for a circularly symmetric PSF, but no analysis was provided. Here we address this important issue formally with a new Fourier domain analysis based on the affine property of the 2-D Fourier transform [15]. In particular, let us compare the 2-D Fourier transform of $f_k(\mathbf{x})$ in Fig. 1 with that of $\tilde{f}_k(\mathbf{x})$ in Fig. 3. In the warp-then-blur model in Fig. 1, we have

$$f_k(\mathbf{x}) = d_k(\mathbf{x}) * h(\mathbf{x}) = d(\mathbf{A}_k\mathbf{x} + \mathbf{t}_k) * h(\mathbf{x}). \tag{9}$$

The Fourier transform of Eq. (9) is given by $F_k(\mathbf{u}) = D_k(\mathbf{u})H(\mathbf{u})$, where $\mathbf{u} = [u, v]^T$ contains

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA

our spatial frequency variables. Using the affine property of the 2-D Fourier transform [15] we can express the Fourier transform of the affine warped image, $d_k(\mathbf{x})$, as follows

$$D_k(\mathbf{u}) = \frac{1}{|\mathbf{A}_k|} e^{j2\pi \mathbf{u}\cdot \mathbf{A}_k{}^{-1}\mathbf{t}_k} D(\mathbf{A}_k{}^{-T}\mathbf{u}), \tag{10}$$

where $D(\mathbf{u})$ is the Fourier transform of $d(\mathbf{x})$. Using Eq. (10), the Fourier transform of $f_k(\mathbf{x})$ can be expressed as

$$F_k(\mathbf{u}) = \frac{1}{|\mathbf{A}_k|} e^{j2\pi \mathbf{u}\cdot \mathbf{A}_k{}^{-1}\mathbf{t}_k} D(\mathbf{A}_k{}^{-T}\mathbf{u}) H(\mathbf{u}). \tag{11}$$

Now, working from the model in Fig. 3, we have $\tilde{f}_k(\mathbf{x}) = f(\mathbf{A}_k\mathbf{x} + \mathbf{t}_k)$. Also, note that the Fourier transform of $f(\mathbf{x})$ is given by $F(\mathbf{u}) = D(\mathbf{u})H(\mathbf{u})$. Thus, again using the affine property of the Fourier transform we obtain

$$\tilde{F}_k(\mathbf{u}) = \frac{1}{|\mathbf{A}_k|} e^{j2\pi \mathbf{u}\cdot \mathbf{A}_k{}^{-1}\mathbf{t}_k} F(\mathbf{A}_k{}^{-T}\mathbf{u}) = \frac{1}{|\mathbf{A}_k|} e^{j2\pi \mathbf{u}\cdot \mathbf{A}_k{}^{-1}\mathbf{t}_k} D(\mathbf{A}_k{}^{-T}\mathbf{u}) H(\mathbf{A}_k{}^{-T}\mathbf{u}). \tag{12}$$

The error between the original and commuted model images can be expressed in the spatial domain as $e_k(\mathbf{x}) = f_k(\mathbf{x}) - \tilde{f}_k(\mathbf{x})$, and in the spatial frequency domain as $E_k(\mathbf{u}) = F_k(\mathbf{u}) - \tilde{F}_k(\mathbf{u})$. Substituting Eqs. (11) and (12) and factoring out a common $D_k(\mathbf{u})$ term, we obtain

$$E_k(\mathbf{u}) = D_k(\mathbf{u}) \left[H(\mathbf{u}) - H(\mathbf{A}_k{}^{-T}\mathbf{u}) \right]. \tag{13}$$

The energy spectral density of the error is then given by

$$\Phi_{E_k}(\mathbf{u}) = |D_k(\mathbf{u})|^2 \left| \left[H(\mathbf{u}) - H(\mathbf{A}_k{}^{-T}\mathbf{u}) \right] \right|^2. \tag{14}$$

Finally, the error-to-signal energy spectral density ratio (ESR) is given by

$$\Gamma(\mathbf{u}) = \frac{\Phi_{E_k}(\mathbf{u})}{\Phi_{D_k}(\mathbf{u})} = \left| \left[H(\mathbf{u}) - H(\mathbf{A}_k{}^{-T}\mathbf{u}) \right] \right|^2. \tag{15}$$

An important observation about the ESR from Eq. (15) is that the commutation error depends on how much the specific OTF changes when undergoing the inverse of the non-translational component of the affine warping in question. In particular, if $H(\mathbf{u})$ has a symmetry such that $H(\mathbf{u}) = H(\mathbf{A}_k{}^{-T}\mathbf{u})$, the commutation error is zero. Since the translation parameters are not involved in Eq. (15), it is clear that translation has no impact on the commutation error energy. Also, when $H(\mathbf{u})$ is circularly symmetric and the affine transformation is rotation, we also get zero commutation error. Commutation error due to zoom and shear depend somewhat on the smoothness of $H(\mathbf{u})$. If $H(\mathbf{u})$ varies slowly with frequency, then slight zoom or shear will give rise to only a small ESR in Eq. (15). Some examples showing the nature of Eq. (15) for our MWIR imaging system are provided in Fig. 4. The 2-D overall system OTF is shown in Fig. 4(a). The ESR, $\Gamma(\mathbf{u})$, is shown for a 20 degree rotation, 10% zoom, and 10% horizontal shear in Figs. 4(b)-4(d), respectively. These surfaces illustrate the exact spatial-frequency dependent nature of the commutation error for our system. Note that the peak ESR values are relatively small and the error tends to primarily impact higher spatial frequencies towards the cut-off frequency (especially for rotation).

To better understand how the ESR varies with affine parameters, we plot the peak ESR for the MWIR imaging system for a wide range of f-numbers under various rotations, zooms, and shears in Fig. 5. These curves show how the ESR increases as the affine warping becomes more extreme. However, for a moderate range of affine parameters, the commutation errors are relatively low. It is interesting to note that as the f-number is decreased, we tend to see

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA

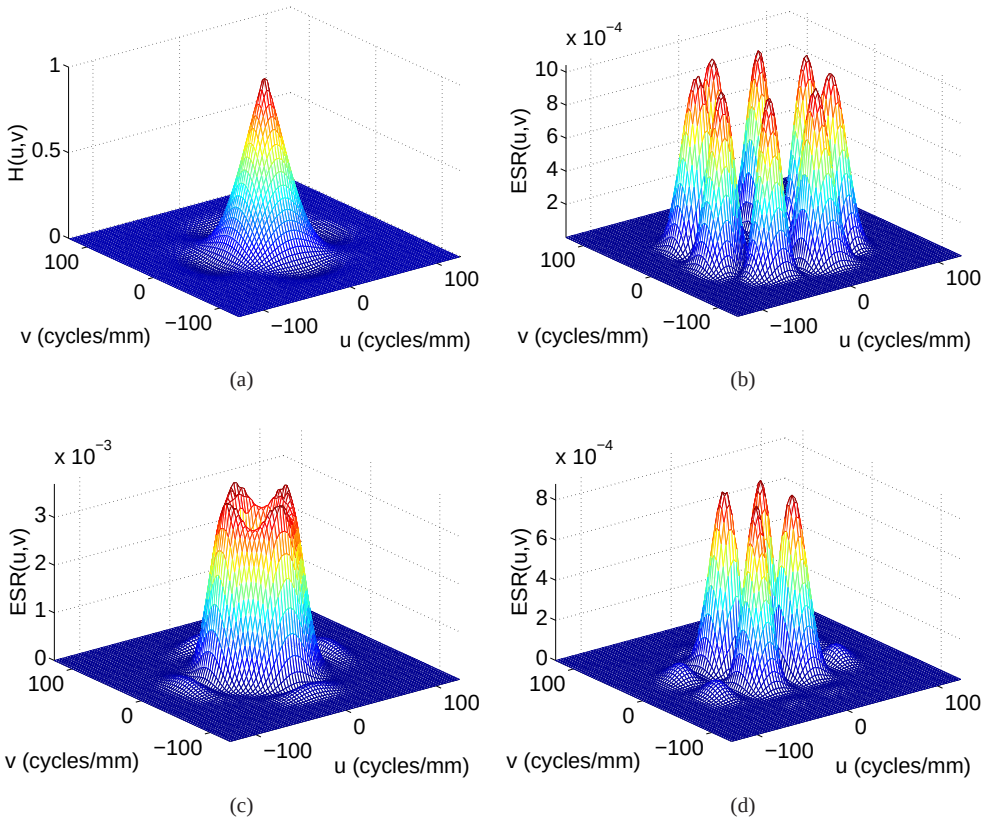


Fig. 4. Error-to-signal energy spectral density ratio analysis based on Eq. (15). (a) 2-D MWIR imaging system OTF, $H(\mathbf{u})$. $\Gamma(\mathbf{u})$ for the MWIR system with (b) a rotation of 20 degrees (c) zoom of 10%, (d) horizontal shear of 10%.

some increased sensitivity to the affine motion, especially for rotation. As mentioned earlier, this results from the OTF becoming somewhat less circularly symmetric and more sinc like. Note that the typical range of f-numbers for MWIR imaging systems is approximately $2-5$. For airborne applications, we expect to see very little zoom and shear over the set of frames used to create an SR image. Rotation is the primary non-translational component we expect and this may amount to only a few degrees in practice. For a three degree rotation, the peak ESR values for commutation error would be approximately 3×10^{-5}. This is likely to fall below the noise floor of most imaging systems. Based on this analysis, we believe the commutation of the motion and PSF for imaging systems like the one modeled here is a good approximation. This powerful conclusion opens the door for a wide variety of fast interpolation-restoration SR methods to be legitimately extended and applied to affine motion for important applications such as airborne imaging. This paves the way for the development of the proposed AWF SR method for non-translational motion, which is the focus of the remainder of this paper.

3. Fast AWF SR for non-translational motion

In this section we describe the fast AWF SR method that is capable of treating arbitrary interframe motion including affine. We begin with an overview of the algorithm and then focus on computing the AWF filter weights. Next we describe the optimization of the partial obser-

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA

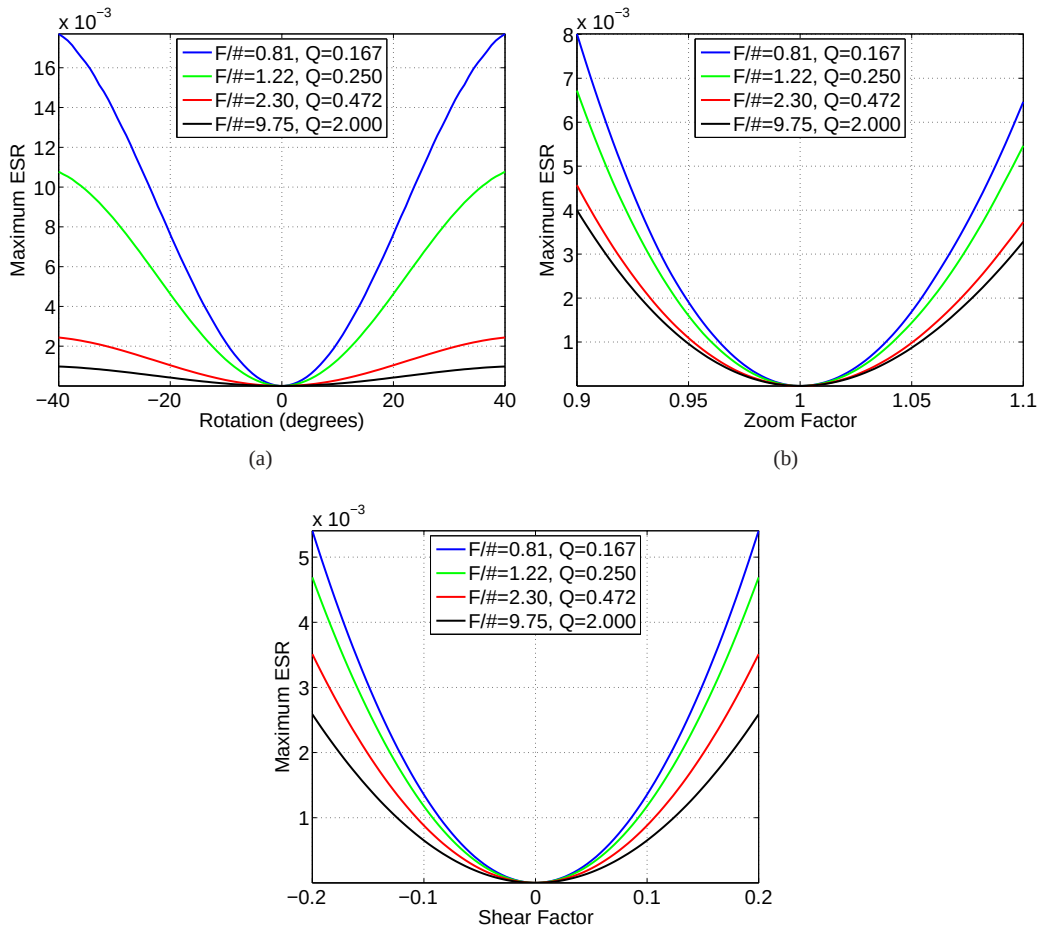


Fig. 5. Peak ESR value for the MWIR imaging system with various f-numbers (Q values) for (a) rotation, (b) zoom, and (c) shear. The actual system has an f-number of 2.30 (Q=0.472).

vation window that is the key to allowing this method to be implemented with relatively few operations. We end the section with a brief discussion on computational complexity.

3.1. Fast AWF SR algorithm overview

A block diagram showing an overview of the proposed fast AWF SR method is provided in Fig. 6. A group of K LR frames, $\mathbf{g}(k)$ for $k = 1, 2, \ldots, K$, are first registered to a reference image with subpixel accuracy. The reference frame is usually the most recent of the K LR images. The affine registration can be done incrementally frame-to-frame and then accumulated, as described in Section 2.2, to provide registration parameters relating each frame to the reference image. The LR pixels from each frame are then placed appropriately onto a common HR grid that is aligned with the reference frame. Finally, the output for each SR pixel is formed as a weighed sum of the values in the HR grid falling within the span of a finite local window of size $W_x \times W_y$ HR pixel spacings. The output SR image estimate, $\hat{\mathbf{d}} = [\hat{d}_1, \hat{d}_2, \ldots, \hat{d}_D]^T$, is computed

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA

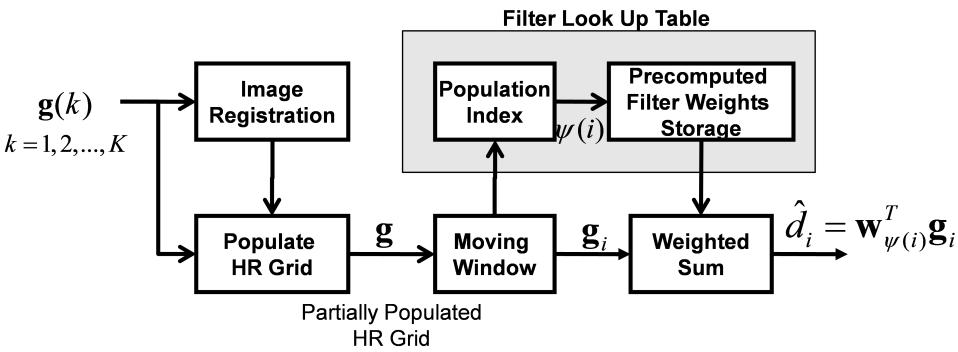


Fig. 6. Fast AWF for non-translational motion block diagram.

with L_x and L_y times the number of pixels in the horizontal and vertical directions of an LR frame, respectively.

In the case of translational motion only, addressed in [7], the LR pixels may be positioned on a continuous HR grid and the filter weights found based on the exact locations of the LR pixels. The computation of the weights can be done for just one observation window and then applied throughout the HR grid. This is because the pattern of LR pixels is periodic on the HR grid. This gives the translation AWF SR method a very low computational complexity [7]. However, for non-translational motion, the pattern of LR pixels populating the HR grid is not periodic. The minimum MSE weights can still be computed, but a unique set of weights may be needed for each observation window. Such on-the-fly processing is possible, but has a much higher computational complexity.

To address the case of non-translational motion, and maintain a low computational complexity, the novel approach presented here is to precompute all of the filter weights before SR processing begins. Computation of the filter weights is addressed in Section 3.2. To make precomputing all the weights feasible, we use a discrete rather than continuous HR grid, and we limit the number of samples that are weighted in the observation window using a new subwindow selection method. Selection of samples to be weighted in this partial observation window is addressed in Section 3.3. To populate the discrete HR grid, we begin by expanding the reference image by L_x and L_y in the horizontal and vertical dimensions, respectively, with zero filling. Subsequent frames are used to further populate the HR grid based on their alignment relative to the reference frame. This must be done with some form of interpolation. The most computationally simple method is to quantize the positions of the LR pixels provided by the registration which amounts to a nearest neighbor interpolation. However, other interpolation methods may be employed. Unless otherwise specified, we us bicubic interpolation. Regardless of the interpolation method, the maximum repositioning is only a fraction of an HR pixel size, $1/(2L_x)$ and $1/(2L_y)$. Multiple samples falling into the same HR position are averaged. Note that the way in which the HR grid is populated depends on the motion between the other frames and the reference. The HR grid can go from only the reference pixel positions populated to having the entire HR grid filled.

To complete the AWF SR processing, an observation window passes across the partially populated HR grid. Let the G_i populated position samples within the window centered at HR pixel i in the HR grid be designated $\mathbf{g}_i = [g_{i,1}, g_{i,2}, \ldots, g_{i,G_i}]^T$. The output is simply a weighted sum of these values with weights tuned for the specific spatial population for the window. As shown in Fig. 6, the output is given by $\hat{d}_i = \mathbf{w}_{\psi(i)}^T \mathbf{g}_i$, where $\hat{d}_i$ is the estimate of the i'th pixel in the desired image and $\psi(i)$ is the population index for window i. Note that the population index

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA

designates which set of precomputed weights is to be applied at this spatial location.

3.2. Optimizing the filter weights

To derive the minimum MSE weights, let us begin by expressing the MSE in terms of a $G_i \times 1$ weight vector $\mathbf{w}$ for a given observation window

$$J(\mathbf{w}|\Psi = \psi(i)) = E\left\{\left(d_i - \hat{d}_i\right)^2 | \Psi = \psi(i)\right\} = E\left\{\left(d_i - \mathbf{w}^T \mathbf{g}_i\right)^2 | \Psi = \psi(i)\right\}, \tag{16}$$

where Ψ is a random variable representing population index. Multiplying this out and distributing the expectation we obtain

$$J(\mathbf{w}|\Psi = \psi(i)) = E\{d_i^2\} - 2\mathbf{w}^T E\{d_i \mathbf{g}_i | \Psi = \psi(i)\} + \mathbf{w}^T E\{\mathbf{g}_i \mathbf{g}_i{}^T | \Psi = \psi(i)\}\mathbf{w}. \tag{17}$$

For a wide sense stationary (WSS) model where the correlations in Eq. (17) are only a function of the 2-D distance between the samples involved, all the correlations are uniquely specified as a function of the population index. Thus, let $\mathbf{R}_{\psi(i)} = E\{\mathbf{g}_i \mathbf{g}_i^T | \Psi = \psi(i)\}$ and $\mathbf{p}_{\psi(i)} = E\{d_i \mathbf{g}_i | \Psi = \psi(i)\}$, giving rise to

$$J(\mathbf{w}|\Psi = \psi(i)) = E\{d_i^2\} - 2\mathbf{w}^T \mathbf{p}_{\psi(i)} + \mathbf{w}^T \mathbf{R}_{\psi(i)} \mathbf{w}. \tag{18}$$

Taking the gradient of the MSE with respect to the weights and setting this to zero and solving for the weights leads to the well known solution to the Wiener-Hopf equations [16]

$$\mathbf{w}_{\psi(i)} = \mathbf{R}_{\psi(i)}^{-1} \mathbf{p}_{\psi(i)}. \tag{19}$$

Substituting the optimum weights into Eq. (18), we get the minimum MSE as

$$J(\mathbf{w}_{\psi(i)}|\Psi = \psi(i)) = E\{d_i^2\} - \mathbf{p}_{\psi(i)}^T \mathbf{R}_{\psi(i)}{}^{-T} \mathbf{p}_{\psi(i)}. \tag{20}$$

To obtain the weights for each population index using Eq. (19), we require $\mathbf{R}_{\psi(i)}$ and $\mathbf{p}_{\psi(i)}$. These correlations can come from an analytical correlation model as in [7] or can be estimated empirically from training data as in [6]. Here we shall use the WSS analytical model in [7] which assumes the underlying correlation model for the desired signal is given by

$$r_{dd}(x,y) = \sigma_d^2 \rho^{\sqrt{x^2+y^2}}, \tag{21}$$

where x and y are conveniently measured in HR pixel spacings, σ_d^2 is the variance of the desired signal, and ρ is the one HR pixel step correlation value. The cross-correlation function between $d(x,y)$ and $f(x,y)$, as shown in Fig. 3, can be expressed in terms of $r_{dd}(x,y)$ [7,16] as

$$r_{df}(x,y) = r_{dd}(x,y) * h(x,y). \tag{22}$$

Again following the analysis in [7], the autocorrelation of $f(x,y)$ is given by

$$r_{ff}(x,y) = r_{dd}(x,y) * h(x,y) * h(-x,-y). \tag{23}$$

Sampling Eq. (23) at x,y positions corresponding to the displacement between samples in $\mathbf{g}_i$ yields $E\{\mathbf{f}_i \mathbf{f}_i^T | \Psi = \psi(i)\}$, where $\mathbf{f}_i$ is the noise-free version of $\mathbf{g}_i$. Assuming independent additive white Gaussian noise of variance σ_n^2, the autocorrelation matrix for the corresponding population index is given by $\mathbf{R}_{\psi(i)} = E\{\mathbf{f}_i \mathbf{f}_i^T | \Psi = \psi(i)\} + \sigma_n^2 \mathbf{I}$. Similarly, evaluating Eq. (22) based on the displacements between the samples in $\mathbf{g}_i$ and the desired sample position yields $\mathbf{p}_{\psi(i)}$. A spatially varying statistical model is also presented and applied in [7]. However, we

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA

have found that this is mainly beneficial in low signal-to-noise ratio environments. Here we focus on high signal-to-noise ratios and do not use the spatially varying statistical model in order to reduce computational complexity.

The essential feature of this analysis to note is that the AWF SR weights vary with the spatial distribution of the available samples in each observation window. For each population index, we have a unique spatial sample distribution requiring a weight vector designed for that pattern. To illustrate how the weights vary with the spatial distribution of samples Fig. 7 (Media 1) shows four different sets of computed weights. These weights are for the MWIR imaging system model with $L_x = L_y = 3$, $\rho = 0.7$, $\sigma_d^2/\sigma_n^2 = 100$. Only the samples in the green boxes are assumed to be available and have a corresponding weight shown as a grayscale, where the background gray corresponds to a weight of zero. Note how the weights change as the spatial distribution of available samples varies to provide the minimum MSE estimate for each observation window. The challenge now is to find a way to limit the number of weight vectors that need to be computed and stored so that we may be able to practically apply the system in Fig. 6 without excessive memory storage and access requirements.

3.3. *Optimizing the partial observation window*

An observation window that is an integer multiple of L_x and L_y in size is guaranteed to be populated by $(W_xW_y)/(L_xL_y)$ pixels from the reference image. However, the remaining positions within the window may or may not be populated depending on the interframe motion. The number of unique ways that the window could be populated is given by 2^W, where $W = W_xW_y - \frac{W_xW_y}{L_xL_y}$. A unique set of weights would ideally be precomputed for each of these population patterns to allow for fast processing. However, as the window size grows, the exponential growth in population patterns for the full window quickly approaches an unmanageable size. We wish to limit the number of filter weights while obtaining a high level of performance. To do so, we propose using a specially selected subset of the full observation window that is made up of the reference pixel positions and M additional positions, where $0 \leq M \leq W$. This gives rise to a total $L_xL_y2^M$ filter weight vectors that are precomputed and stored. The factor of L_xL_y comes into play because we have this many unique window positions relative to the reference grid, and each requires its own weights. The population index for a given position relative to the reference grid is found by simply converting the populations into an M-bit binary number, with a one for a present sample and zero for a missing sample. Converting this to decimal provides a unique and convenient population index.

The question now is how to find the most salient M positions to add to the reference samples to generate the desired subwindow. Our selection method is based on the expected MSE for a candidate subwindow given by

$$J = E\{J(\mathbf{w}_\psi|\Psi)\} = \sum_{\psi=1}^{2^M} J(\mathbf{w}_\psi|\Psi = \psi)Pr\{\Psi = \psi\}, \tag{24}$$

where $Pr\{\Psi = \psi\}$ is the probability that we observe population index ψ and $J(\mathbf{w}_\psi|\Psi = \psi)$ is the minimum MSE for this population index, as defined in Eq. (20). The goal is to find the subwindow with M samples plus the reference samples that minimizes J in Eq. (24). For a typical window size, an exhaustive search is impractical since it would involve searching among $\binom{W}{M} = \frac{W!}{M!(W-M)!}$ subwindows. Thus, we propose a novel application of a forward sequential selection procedure to choose a highly salient subwindow. Using this method, the partial observation window begins with just the fixed reference samples. Next, the one non-reference pixel position that minimizes J in Eq. (24) is found by exhaustive search. We continue in the fashion, adding one sample at a time until the desired window size is reached. While this method does

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA

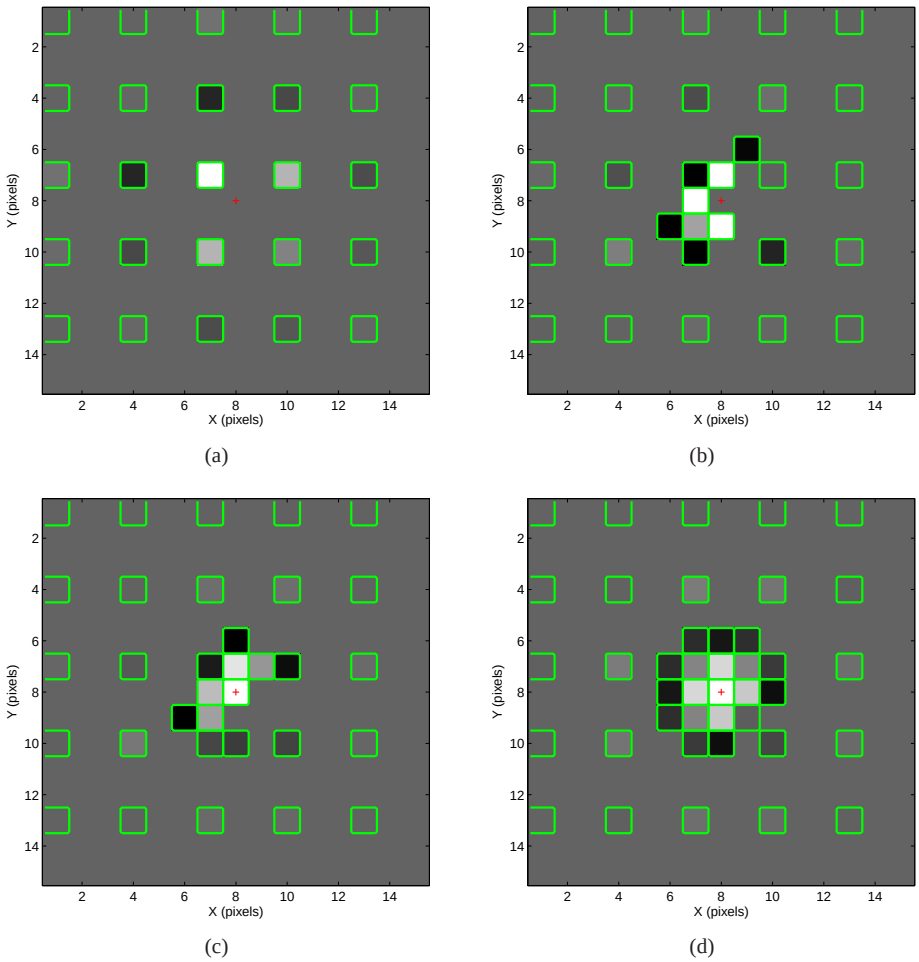


Fig. 7. Minimum MSE weights for estimating the position shown with the red plus sign for four different population patterns (Media 1). Only samples in the green boxes are assumed to be available and the grayscale value represents the weight with background gray corresponding to zero. The pattern in (a) is the case of no motion or a single frame (the reference frame).

not guarantee a global minimum expected MSE, we believe it yields a useful and computable local minimum solution.

To compute the probability for each population index, let us assume the probability that any non-reference HR grid position is filled is independent and denoted p_1. The probability a non-reference HR grid position is empty is given by $p_0 = 1 - p_1$. The probability of a given population pattern occurring within a candidate subwindow is a function of how many of the subwindow positions are filled and how many are empty. For a candidate subwindow with M non-reference positions, the probability of any patten with M_f filled and $M_e = M - M_f$ empty positions is given by $p_1^{M_f} p_0^{M_e}$. Let us now further assume that each frame provides on average one sample for each $L_x \times L_y$ superpixel on the HR grid and these are distributed uniformly. Thus, we are neglecting the cases where no pixels or multiple pixels from one frame populate the same

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA

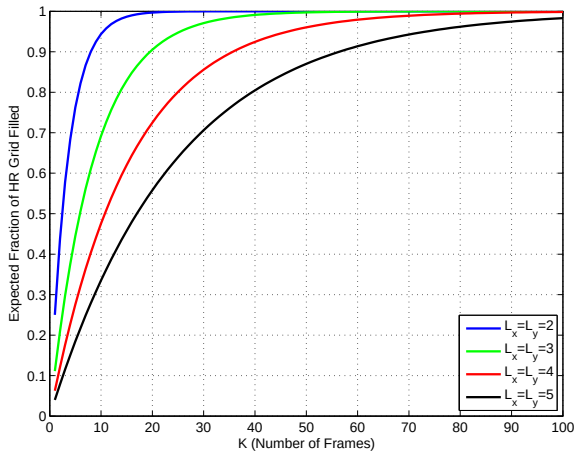


Fig. 8. Expected fraction of the HR grid populated as a function of the number of frames with uniform motion.

$L_x \times L_y$ superpixel on the HR grid. These cases are only likely with high levels of zoom. Based on these assumptions it can be shown that the probability that a non-reference sample is empty after populating the HR grid with K frames is given by $p_0 = ((L_xL_y - 1)/(L_xL_y))^{K-1}$. The number of non-reference grid positions populated in each superpixel is described by a binomial probability mass function with an expected value of $(L_xL_y - 1)p_1$. Therefore, the expected fraction of the HR grid that is populated is $((L_xL_y - 1)p_1 + 1)/(L_xL_y)$. This fraction is plotted as a function of the number of frames K for different values of $L_x = L_y$ in Fig. 8. This plot shows that K impacts the likely population density when uniform motion is present. It is interesting to note that a significant number of frames is needed to get close to a fully populated grid, even with uniform motion. This highlights the potential benefit of the AWF approach that adapts to the spatial distribution of the available and missing samples on the HR grid.

Subwindows for the MWIR imaging system designed with the forward sequential method are shown in Fig. 9. These are designed for the case where $L_x = L_y = 3$, $\rho = 0.7$, $\sigma_d^2/\sigma_n^2 = 100$, $K = 10$ and $M = 16$. Four of the nine different estimation positions relative to the reference grid are shown. The estimation position is shown with a red plus sign and the reference grid samples are shown in green. The $M = 16$ selected positions are shown with blue numbers, indicating the order in which they were selected. The grayscale value is proportional to the expected MSE obtained when adding the corresponding sample. Note that the selection sometimes follows a non-obvious path, influenced by the locations of the reference grid samples and prior selections. The specifics of the PSF, correlation models, and HR grid density (determined by K) all play a role in which samples are selected. Note that the weights shown in Fig. 7 (Media 1) are for the subwindow in Fig. 9(d).

The theoretical expected MSE values from Eq. (24) are shown in Fig. 10 as a function of M for the subwindows in Figs. 9(a), 9(b), and 9(d). The values are normalized by σ_d^2 to provide a type of error-to-signal ratio. The expected MSEs clearly decline as we increase the size of the subwindow. The "knee" in the curve occurs early, suggesting that a relatively small M can be effective. The error for the Position 1 starts low, because for that position, the sampled being estimated is on the reference grid giving this position a big advantage. On the other hand, for Position 5, the guaranteed reference grid sample is farthest away from the estimation position, putting it initially at a disadvantage. However, it is this position that benefits most from adding samples to the subwindow to compensate for its distance to the reference grid.

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA

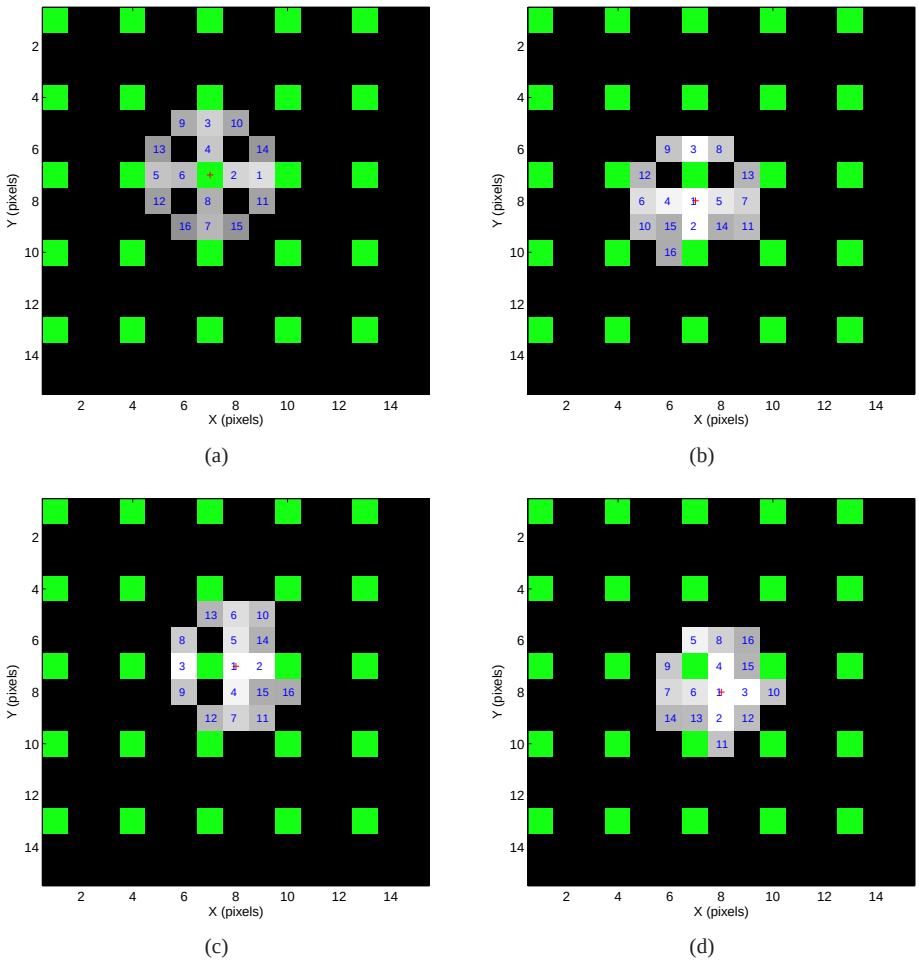


Fig. 9. Subwindows designed using the forward sequential selection method for $L_x = L_y = 3$, $\rho = 0.7$, $\sigma_d^2/\sigma_n^2 = 100$, $K = 10$ and $M = 16$. The estimation position is shown with a red plus sign and the reference grid samples are shown in green. The selected positions are shown with blue numbers in order of selection.

3.4. *Computational complexity*

Let us briefly consider the computational complexity of the proposed AWF SR method with subwindow processing. With the weights precomputed, the entire filtering operation consists of computing the weighted sum $\hat{d}_i = \mathbf{w}_{\psi(i)}^T \mathbf{g}_i$. This means that the number of floating point operations (flops) for output i is $G_i \leq M + W_x W_y/(L_x L_y)$. Note that a flop represents one floating point multiply and add operation. If the weights were not precomputed and we could not count on translational motion, it may be necessary to also solve $\mathbf{R}_{\psi(i)}\mathbf{w}_{\psi(i)} = \mathbf{p}_{\psi(i)}$ for each observation window. The weights can be computed using Cholesky factorization which requires $G_i^3/3$ flops to perform LU decomposition on the autocorrelation matrix. Solving for the weights using forward and backward substitution requires an additional $2G_i^2$ flops [6,7]. Note that in [6,7], some computational saving is achieved by estimating several output pixels for each observation window. This reduces the number of LU decompositions needed. However, precomputing all the

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA

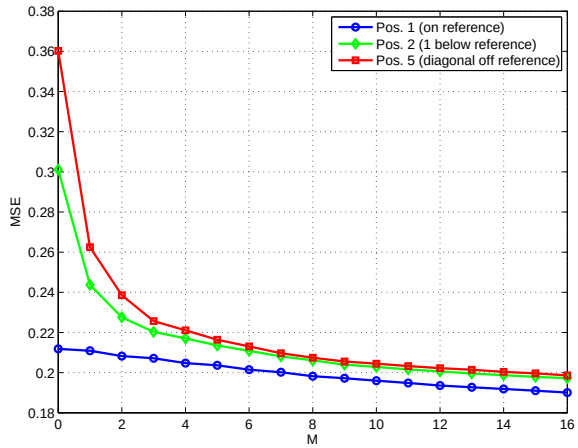


Fig. 10. Theoretical expected MSE values normalized by σ_d^2 as a function of M for the subwindows in Figs. 9(a), 9(b), and 9(d).

weights is clearly the most significant way to reduce runtime computations. Other operations that should be considered include registration and population of the HR grid. Using incremental registration, only one frame-to-frame registration is done per output frame. Populating the HR grid by simple spatial quantization has a very low computational cost. If other interpolation methods are used, the number of flops per output increases accordingly. The memory requirements are another important consideration with the proposed method. The number of non-zero precomputed filter weight values that must be stored is $(L_x L_y 2^M) \times (W_x W_y/(L_x L_y) + M/2)$.

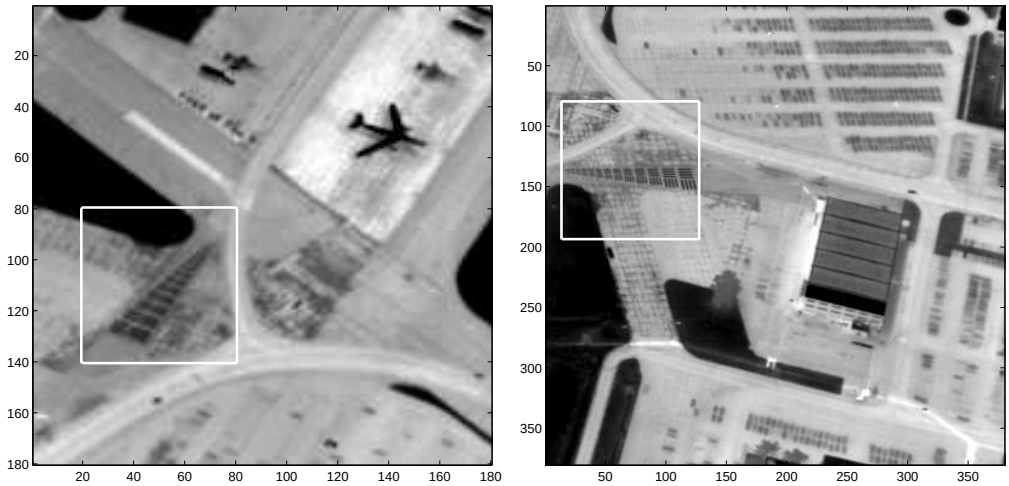
4. Experimental results

We present experimental results using MWIR flight data, video from a visible camera, and simulated data generated from a still frame aerial image degraded according to the model in Fig. 1. Results using several SR methods are presented. Fast AWF SR refers to the proposed method using precomputed weights and partial observation window. The weighted nearest neighbor (WNN) method is presented in [17, 18]. Nonuniform interpolation is done using an inverse distance based weighting of the nearest 4 neighbors and restoration is done with an FFT based Wiener filter with a constant noise-to-signal ratio (NSR) tuning parameter. The Delaunay triangulation method is based on the method in [19] and also uses a Wiener filter for restoration. The regularized least squares (RLS) iterative method is described in [10]. Note that the RLS method does not assume the PSF blurring and motion models commute. All of these benchmark methods have been extended here for affine motion.

4.1. Infrared flight data

The details of the MWIR imaging system used to acquire the fight data are provided in Section 2.3 and further details about the data collection can be found in [8]. The scene imaged includes a thermal resolution target composed of 13 pairs of 4-bar groups orthogonally oriented. The bars have a 7:1 aspect ratio and range in width (same as spacing) from 1.0m to 0.25m. The scaling factor between bar groups is designed to be $2^{(1/6)}$ [8]. We process two image sequences of flight data here. The first has a frame rate of 50Hz. In this mode, a group of 10 frames is acquired and then a step-stare mirror repositions the field of view. The second sequence is acquired at 16Hz with no step-stare. The region of interest (ROI) that is processed for each sequence is shown in

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA



(a) (b)

Fig. 11. Flight data regions of interest processed. (a) 50Hz sequence frame (b) 16Hz sequence frame. The thermal resolution target area is boxed in each image.

Fig. 11.

The SR results using the 50Hz sequence are shown in Fig. 12 for $L_x = L_y = 3$ and $K = 10$ frames. Although an upsampling factor of 3 is below that theoretically required to ensure no aliasing for this sensor, we have found essentially no noticeable improvement for these data using a higher value and the lower value reduces processing time. The output using bicubic interpolation for the ROI immediately around the resolution target is shown in Fig. 12(a). Note that aliasing makes all of the more horizontally oriented bars appear unresolvable. In fact, the lowest resolution 4-bar group incorrectly appears like three diagonal bars. The more vertical bar groups can be resolved only up to Group 2 (0.89m bar width). The partially populated HR grid used for the AWF SR method is shown in Fig. 12(b). The fraction of populated pixels is 0.686 and that predicted using the analysis in Section 3.3 is 0.692. The fast AWF SR output is shown in Fig. 12(c) for $W_x = W_y = 15$, $M = 16$, $\rho = 0.7$, and $\sigma_d^2/\sigma_n^2 = 100$. Here the more horizontally oriented bar groups are resolvable up to and including Group 3 (0.79m bar width), while the more vertically oriented ones are resolvable up to Group 5 (0.63m bar width). For comparison the outputs for WNN (NSR=0.04), Delaunay (NSR=0.02), and RLS (with 20 iterations and regularization parameter of 0.1 [10]) are shown in Figs. 12(d)-12(f), respectively. The selection of the tuning parameters is based on subjective evaluation of the results and an analysis of the quantitative simulation results. Note that the maximum estimated rotation over the 10 frames is 0.49 degrees and the maximum translational shift is 0.97 LR pixels. The maximum zoom and shear parameters are estimated to be 1.00068 and 0.0038, respectively.

To illustrate the robustness of the various algorithms to interframe motion, outputs are shown in Fig. 13 for the case of no interframe motion. Here one frame is repeated to simulate lack of camera motion. This represents a worst case of sampling diversity that SR methods may encounter. The tuning parameters for all the algorithms are kept as they were in Fig. 12. It is interesting to note that AWF SR method handles this case gracefully, as does the more computationally demanding RLS method. However, the the WNN and Delaunay SR methods that perform the nonuniform interpolation and restoration steps independently have significant artifacts. These artifacts result from the fixed Wiener filter, optimized for diverse motion, amplify-

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA

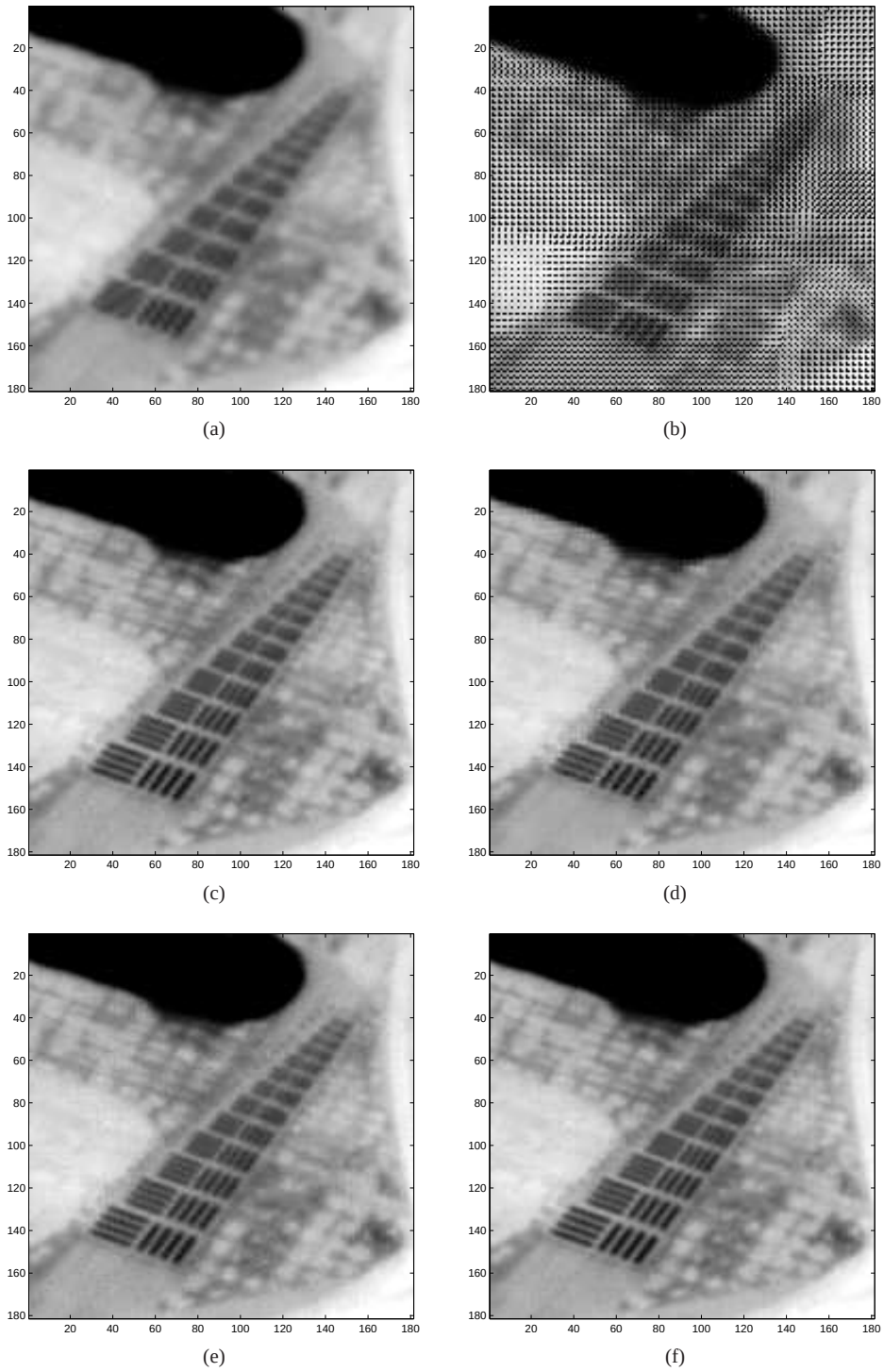


Fig. 12. SR results for 50Hz MWIR flight sequence for $L_x = L_y = 3$ and $K = 10$. (a) Bicubic interpolation, (b) partially populated HR grid, (c) fast AWF SR method (d) WNN, (e) Delaunay, (f) RLS (20 interations).

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA

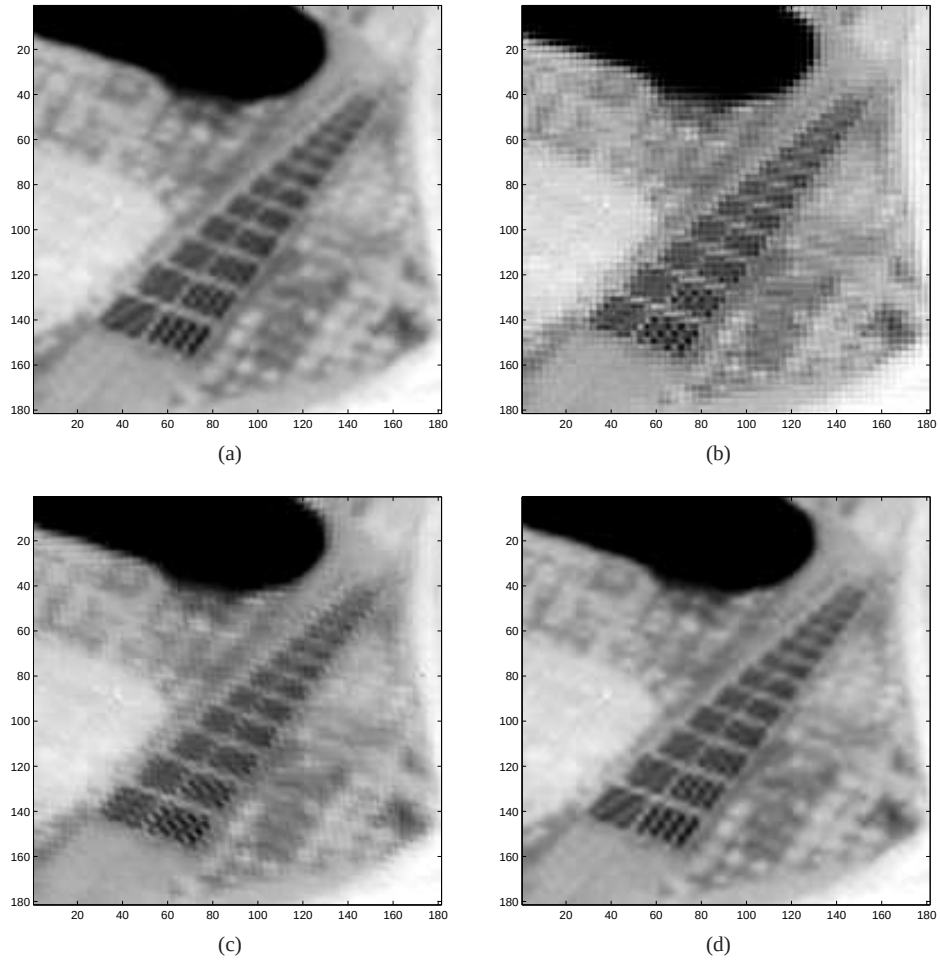


Fig. 13. SR results for 50Hz MWIR data simulating no camera motion by repeating a single frame $K = 10$ times. Algorithm tuning parameters are the same as in Fig. 12. The results are for (a) fast AWF SR method, (b) WNN, (c) Delaunay, (d) RLS (20 interations).

ing the artifacts from a poor nonuniform interpolation step. This result highlights the potential importance of jointly performing nonuniform interpolation and restoration, which is a key asset of the AWF approach.

The final MWIR flight data results are shown in Fig. 14 (Media 2). Here we evaluate the performance of the fast AWF SR method using $K = 10$ frames of the 16Hz sequence and investigate the impact of different motion models. The reference frame with $L_x = L_y = 3$ bicubic interpolation is shown in Fig. 14(a) (Media 2, left). It appears that the horizontal bar groups are resolvable to Group 2 (0.89m) and the vertical to perhaps Group 6 (0.56m). The SR result using the fast AWF SR method with an affine motion model is shown in Fig. 14(b) (Media 2, right), using the same parameters as before. Here the horizontal bars are resolvable to perhaps Group 7 (0.50m), and the vertical bars to Group 10 (0.35m). The result using the same method but with a rotation and translation motion model (no zoom or shear) is shown in Fig. 14(c). The result using a translational only motion model is shown in Fig. 14(d). Note that translation alone is

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA

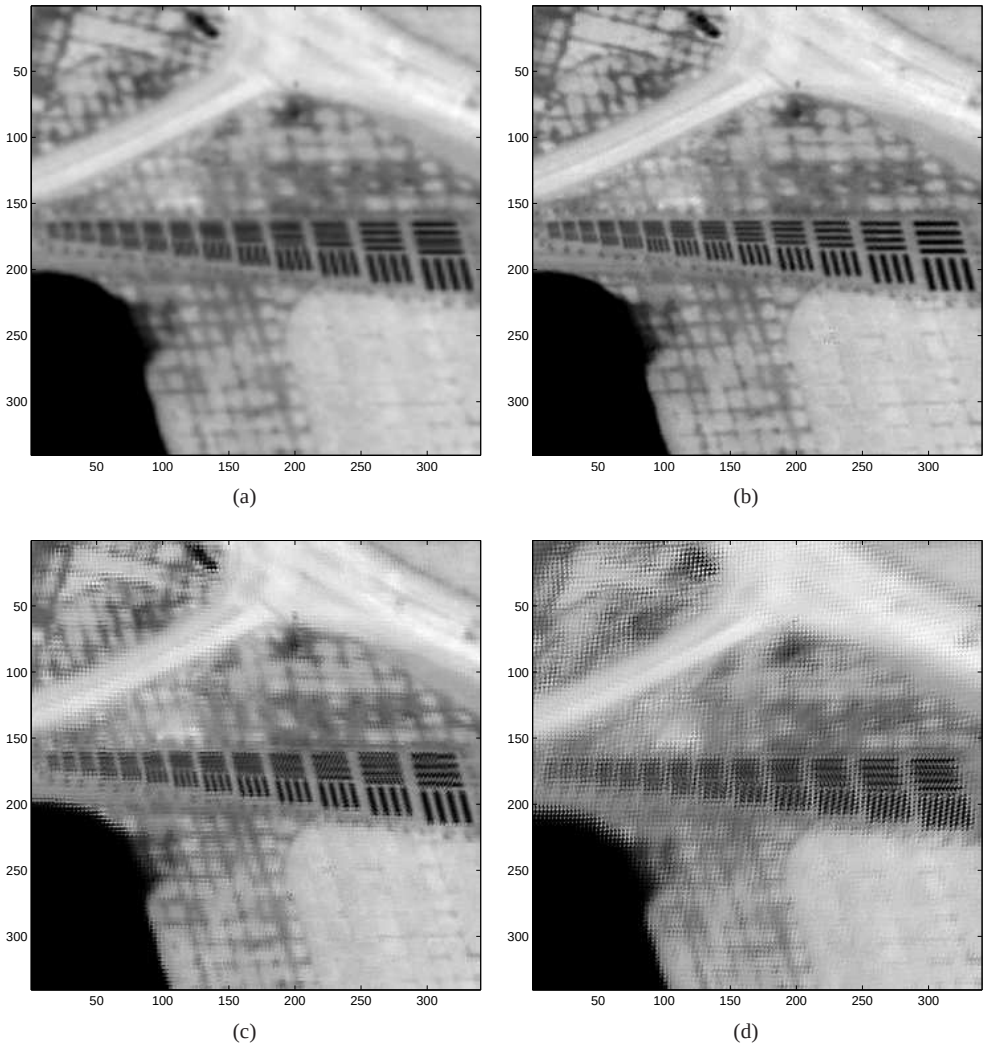
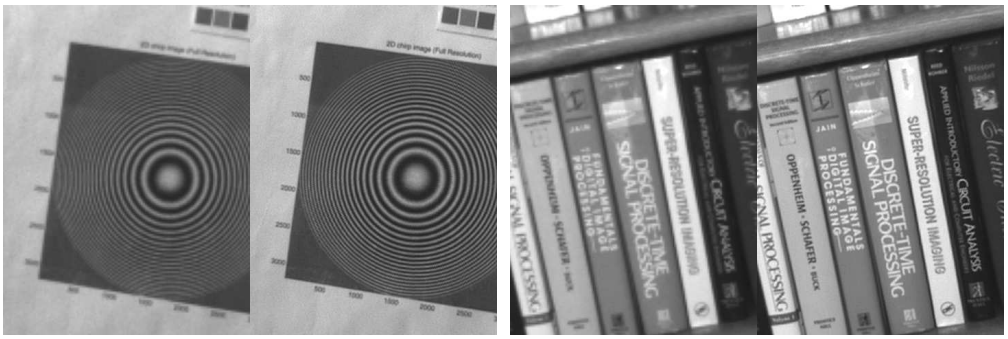


Fig. 14. SR results for $K = 10$ frames of the 16Hz MWIR sequence using the fast AWF SR method with different motion models. Algorithm parameters are the same as in Fig. 12. The results are for (a) bicubic interpolation (Media 2, left), (b) the fast AWF SR with affine motion model (Media 2, right), (c) rotation and translation, (d) translation only.

clearly not able to accurately represent the true motion. Some artifacts are still present when allowing for rotation, but they are greatly reduced. The affine motion model clearly produces the best results here, illustrating the potential importance of using this motion model for SR in airborne imaging applications. The maximum rotation over the 10 frames in this set is 1.79 degrees, and the maximum translational shift is 2.28 LR pixels. The maximum zoom and shear parameters are estimated to be 1.0064 and 0.0097, respectively. While Fig. 14 makes it clear that these non-translational affine values are important from a sample placement perspective, the analysis in Section 2.4 shows that such a small departure from translation produces only negligible levels of blur/motion commutation error.

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA



(a) (b)

Fig. 15. Single-frame excerpts from video results using the proposed fast AWF SR method with data from the visible grayscale camera described in Section 2.3. (a) 2-D chirp sequence (Media 3). (b) bookshelf sequence (Media 4).

4.2. *Visible camera video*

To further illustrate the performance of the proposed fast AWF SR method, we show SR results using for data from the visible grayscale camera described in Section 2.3. The frames are acquired at 30 frames per second and the camera is moved by hand to produce significant amounts of rotation and translation, with some small amounts of zoom and shear possible as well. The imagery are from an indoor scene with a 2-D chirp pattern shown in Fig. 15(a) (Media 3) and a bookshelf in Fig. 15(b) (Media 4). The left side of each displayed frame shows $L_x = L_y = 3$ bicubic interpolation and the right side shows the $K = 10$ fast AWF SR output using the same algorithm parameters as before with the appropriate camera PSF model. Note that like the flight data, the input imagery is taken directly from the camera with no artificial degradation. Like the bar target in the flight data, the chirp pattern clearly shows the aliasing when single frame interpolation is used. The corrected concentric circles are fully visible in the SR result in Fig. 15(a) (Media 3). Similarly, the writing on the book spines is far more legible after SR processing in Fig. 15(b) (Media 4).

4.3. *Simulated data error analysis*

The final set of results are using a simulated image sequence derived from the still 8-bit aerial image of size 434×491, shown in Fig. 9(b) in [7]. The image is put through the observation model in Fig. 1 to generate a set of $K = 10$ LR frames with affine motion. The PSF for the MWIR imaging system is used with a noise variance of $\sigma_n^2 = 4$. This simulation is done to allow for quantitative error analysis, since we have knowledge of the true desired image. The MSE results and average run times are shown in Table 1. Note that that method labeled Fast AWF (Q) populates the HR grid using the simple quantization, as compared with the standard method which uses bicubic interpolation. Full AWF SR refers to computing weights for the full observation window on-the-fly (also using a discrete HR grid). The run times were generated using MATLAB on an Intel Core i7 64 bit CPU with a clock speed of 3.07 GHz. The multiscale affine registration uses 3 levels with 5 iterations at each level and employs bicubic interpolation. The run time for the registration of all 10 frames is 0.76 seconds. The affine motion considered here includes translation, rotation, shear, zoom, a combination of all of these, and the no motion case. The translation parameters have a normal distribution of $N(0, 2^2)$ LR pixel spacings, the rotation angle is $N(0, 10^2)$ degrees, the shear is horizontal with parameter

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA

Table 1. MSE for SR with affine motion for aerial image ($L_x = L_y = 3$ and $K = 10$).

Method	No Motion	Trans.	Rot.	Shear	Zoom	All	Time (s)
Bicubic	470.03	470.03	470.03	470.03	470.03	470.03	0.088
Fast AWF (Q)	384.24	185.13	222.49	303.12	224.32	218.39	0.165
Fast AWF	384.24	168.98	195.05	299.42	199.35	191.64	0.284
Full AWF	384.24	164.41	184.92	295.68	190.66	183.95	14.862
WNN	711.59	203.18	256.85	481.97	267.84	270.62	0.390
Delaunay	514.74	166.33	198.44	336.94	215.00	232.07	3.711
RLS (20 Its.)	389.73	132.60	156.67	282.62	157.30	160.76	64.066
RLS (5 Its.)	391.55	191.78	217.62	309.44	219.20	219.90	16.114

distribution $N(0, 0.1^2)$, and the zoom parameter is $N(1, 0.1^2)$. The NSRs that provide the lowest MSE in the "All" column are used for each method. The AWF SR outputs use an NSR of 0.005. The NSR for the WNN and Delaunay SR methods are 0.04 and 0.02, respectively. The RLS uses a regularization parameter of 0.01 [10]. The other processing parameters are the same as before.

Note that the 20 iteration RLS generally produces the best results but has a much longer run time than the other methods. The fast AWF SR method has the shortest run time and MSE values that are generally better than all but the 20 iteration RLS and full AWF. The MSE results show that the full AWF does outperform the fast AWF, however by using the subwindow selection method described in Section 3.3, the difference is relatively small. Also, one can see in Table 1 that populating the HR grid with simple quantization is faster than using bicubic interpolation, but the MSEs are consistently higher. This comparison highlights one of the implementation tradeoffs for the AWF SR method. Simple quantization provides a shorter run time but with a modest increase in MSE. Another interesting thing to note is how the methods perform in the no motion case. In a manner consistent with what we see in Section 4.1, the SR methods that do independent restoration are most exposed here, actually increasing the error compared with bicubic interpolation. The AWF performs relatively well with all motions because of its ability to adapt to the sampling distribution. Finally, note that the error for shear motion is the highest for all the methods because horizontal shear alone provides no vertical sampling diversity.

5. Conclusions

In this paper, we have presented a novel analysis showing that for many imaging systems, the commutation error for the PSF blur and modest amounts of affine motion is relatively small. This opens the way for the use of fast interpolation-restoration SR methods, that are based on this commutation, for applications such as airborne imaging where an affine motion model may be needed. Building on this finding, we have proposed a new fast version of the AWF SR method that can accommodate affine motion using entirely precomputed filter weights. The weights are designed to jointly perform nonuniform interpolation and restoration for minimum MSE. These weights vary with the spatial pattern of the observed LR pixels which depends on the particular motion in the video sequence. Precomputing the weights for all possible populations of a full observation window is impractical. However, we have developed a new forward sequential selection procedure to pick a highly salient subset of pixels from the full observation window for which weights are computed. This brings the number of weights that are computed and stored down to a manageable size, without much loss in performance. Because the weights are computed prior to video processing, the computational complexity during processing is rel-

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA

atively low. With suitable hardware, the proposed AWF SR method could operate in real-time.

Using MWIR flight data with a thermal resolution target we have demonstrated objective resolution enhancement and aliasing reduction using the new method. The flight data also clearly show the potential need for a non-translational motion model, such as affine, when using an airborne platform. This is especially true for lower frame rate data like the 16Hz flight data. Here we also demonstrate the benefit of the AWF in adapting to different sample distributions, most dramatically shown in the worst case when there is no motion at all. In the quantitative error analysis, the Fast AWF method has the lowest run time of the SR methods tested and the MSE values generated are generally better than all but the 20 iteration RLS and full AWF, which are significantly more computationally demanding.

Acknowledgements

The MWIR flight data have been collected under Air Force Research Laboratory contract number FA8650-06-D-1078-0004. The authors would like to thank Mark A. Bicknell for his support of the project and Dr. Frank O. Baxley for leading the flight data acquisition effort and providing the MWIR data used here. Engineers from General Dynamics Advanced Information Systems and the University of Dayton Research Institute, including Patrick C. Hytla, Joseph C. French, and Phil Detweiler, configured the MWIR system and NASA Glenn Research Center provided flight support.

#153532 - $15.00 USD Received 29 Aug 2011; revised 18 Oct 2011; accepted 20 Oct 2011; published 8 Dec 2011
(C) 2011 OSA