

Improving Spatial Resolution in Exchange of Temporal Resolution in Aliased Image Sequences

Lucas J. van Vliet and Cris L. Luengo Hendriks
Pattern Recognition Group of the Department of Applied Physics
Delft University of Technology
Lorentzweg 1, 2628 CJ Delft, The Netherlands
email: L.J.vanVliet@ph.tn.tudelft.nl
<http://www.ph.tn.tudelft.nl/~lucas>

Abstract

We present a method to improve the spatial resolution of an aliased image sequence in exchange for a lower temporal resolution. The frames are acquired by a vibrating infrared camera, which yields severely undersampled (aliased) image frames of a static scene. The vibration causes a random sub-pixel shift of the scene before sampling. The proposed method first estimates the relative position of all frames after which the periodic-random samples are interpolated on an equidistant grid with a smaller pixel pitch to improve spatial resolution and (partially) eliminate aliasing. The presented method is robust, is completely data driven and has a moderate computational complexity, which permits a real-time implementation in hardware.

1. Introduction

A camera employs a lens to produce a (de)magnified version of the scene onto an image sensor. The lens acts as a low-pass filter whose cut-off frequency depends on the lens aperture. Increasing the aperture of the lens increases the light collection efficiency and permits a higher spatial resolution. An array-based sensor in the image plane not only converts the incoming light into electrical charge, but also samples the image. The Nyquist criterion states that the sampling frequency in the image plane should be higher than twice the cut-off frequency. Satisfying this requirement yields a digital image that contains all the information of the bandlimited analog image. Exact reconstruction of this bandlimited analog image from its samples is possible [7,8].

Infrared cameras employing a 2-D array of light sensitive elements require cooling and isolation of the individual pixels to suppress thermal noise. This yields a pixel pitch that is far too large for high-resolution imaging. The individual frames are undersampled. Undersampling yields aliasing, which corrupts the image data irreversibly. However, a set of aliased realizations of the same scene may under restricted circumstances contain all the

information of the original bandlimited image. Our IR camera vibrates, which yields an image sequence in which each individual frame is shifted over a random sub-pixel vector. An image sequence of a (pseudo) static scene may contain enough information to increase the spatial resolution without being hampered by the aliasing, which corrupted the individual frames.

The challenge in this research is to develop a technique for improving the resolution of aliased image sequences which satisfies the following constraints:

- a clear perspective of a real-time implementation in hardware;
- image displacements will be estimated from the image content;
- flexible and adaptive control over improving image resolution and noise suppression;
- comparable image quality with respect to alternative methods [1,2,3,4,5,9,10];
- graceful degradation of performance for increasing amounts of noise, absence of reliable image contents or lower number of input frames.

Section 2 presents a model for image acquisition and introduces the subproblems: the registration of shifted image frames (section 3) and interpolation (image fusion) of the randomly shifted uniformly sampled image data. (section 4). In section 5 we present the test results for both subsystems on a synthetic image sequence. Section 6 presents the application of the method to some real-world IR image sequences. A discussion follows in section 7.

2. System description

A typical detector array for infrared imaging has a large pixel pitch and a small fill factor. This causes undersampling in most practical circumstances. The imaging model is depicted in Figure 1. The lens (h_{lens}) slightly blurs the image, which is then shifted (h_{shift}), integrated onto a "light" sensitive element (h_{pixel}), sampled ($p(x,y)$) and discretized (ADC). The digital output image is

undersampled and degraded by noise of various sources: photon noise, readout noise, thermal noise, and quantization noise. The signal-to-noise ratio (SNR) of these cameras can be as high as 40 dB.

The motion blur in the individual frames is negligible because of the small exposure time compared to the readout time of the camera. The influence of the square integrating pixel elements can be ignored (due to the small fill factor) for interpolation factors up to a factor four.

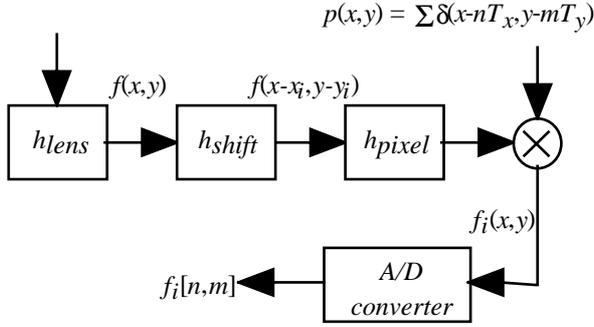


Figure 1: Model of the imaging system.

Before interpolating a high-resolution image from the shifted frames we need to know the translation vectors for the entire image sequence. To allow optimal flexibility we will estimate these vectors from the image content. The camera motion (vibration) of the presented system causes translation but no significant rotation of the acquired images. For scenes with a limited depth range this yields a constant image shift over the entire image.

3. Registration of shifted frames

Image registration has received a lot of attention for a wide variety of applications. The major problem we are facing here is the aliasing that has corrupted the high frequency regions of the individual frames. The registration algorithm can reliably use the low-frequency regions such as smooth edges (temperature ramps) and should omit high-frequency textures, which manifest themselves as low-frequency patterns (aliasing) in the undersampled frames.

We have tested several methods for estimating the translation (x_i, y_i) between two images: the image f_i and the reference image f_o . $f_i(x, y) = f_o(x - x_i, y - y_i)$.

1. Apply cross-correlation with $(4x)$ zero-padding in Fourier domain followed by center-of-mass estimation in spatial domain (CZP)
2. Apply cross-correlation and fitting of a plane through the low-frequency part of the 2-D phase of the Fourier transform (CPF).
3. Minimize the squared error between a first order Taylor series expansion of f_i and f_o (MTS)

4. Average the frame positions obtained with different reference frames.

3.1 Cross-correlation with zero-padding (CZP)

The images contain shifted realizations of the same scene. The cross-correlation between f_o and f_i can be computed by a multiplication in the Fourier domain

$$F \left\{ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_o(\eta, \xi) f_i(x + \eta, y + \xi) d\eta d\xi \right\} = F_o^*(u, v) F_i(u, v) \quad (1)$$

with u and v the spatial frequencies. Normalization yields a linear phase term.

$$\frac{F_o^*(u, v) F_o(u, v) e^{-j(u x_i + v y_i)}}{|F_o(u, v)|^2} = e^{-j(u x_i + v y_i)} \equiv e^{-j\theta(u, v)} \quad (2)$$

Zero padding (n times) before inverse FFT interpolates the sinc-shaped peak. The center-of-mass yields the estimated sub-pixel shift that maps f_i onto f_o . (CZPn)

$$F^{-1} \begin{cases} e^{-j(u x_i + v y_i)} & |u|, |v| < \frac{\pi}{n} \\ 0 & \text{elsewhere} \end{cases} = \frac{\sin(\frac{\pi}{n}(x - x_i))}{\pi x} \frac{\sin(\frac{\pi}{n}(y - y_i))}{\pi y} \quad (3)$$

The phase of the high frequencies is dominated by noise. A low-pass filter suppresses the noise in exchange for a wider peak. Skipping the normalization of eq. (2) yields similar results.

3.2 Cross-correlation with phase fitting (CPF)

Compute the cross-correlation as in eq. (2) after which a 2D plane is fitted to the low-frequencies of the linear phase term $\theta(u, v)$. The angles of the plane with the u - and v -axis correspond to the estimated sub-pixel shifts in respectively the x - and y -direction.

$$\min_{\{x_i, y_i\}} (\theta(u, v) - (u x_i + v y_i))^2 \quad (4)$$

This technique can be refined by removing the outliers from the phase term after which the fit is repeated (CPF2).

3.3 First order Taylor series (MTS)

Reference image f_o can be approximated by a (first) order Taylor series of image f_i at scale σ .

$$f_o^{(\sigma)} \approx f_i^{(\sigma)} + x_i \frac{\partial}{\partial x} f_i^{(\sigma)} + y_i \frac{\partial}{\partial y} f_i^{(\sigma)} \quad (5)$$

with

$$f_i^{(\sigma)} = f_i \otimes g^{(\sigma)} \quad \frac{\partial}{\partial x} f_i^{(\sigma)} = f_i \otimes \frac{\partial}{\partial x} g^{(\sigma)} \quad (6)$$

$$g^{(\sigma)} = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2 + y^2}{2\sigma^2}}$$

By minimizing the squared difference between both sides of eq. (5) we obtain a sub-pixel estimate of the translation between the two images [1,5]. In vector notation we solve

$$\mathbf{G}\mathbf{s} = \mathbf{v} \quad (7)$$

with \mathbf{s} the shifts to be found, \mathbf{G} the integrated squared gradients, and \mathbf{v} the gradient weighted difference image

$$\mathbf{s} = \begin{pmatrix} x_i \\ y_i \end{pmatrix} \quad \mathbf{v} = \begin{pmatrix} \sum \frac{\partial}{\partial x} f_i^{(\sigma)} (f_o^{(\sigma)} - f_i^{(\sigma)}) \\ \sum \frac{\partial}{\partial y} f_i^{(\sigma)} (f_o^{(\sigma)} - f_i^{(\sigma)}) \end{pmatrix} \quad (8)$$

$$\mathbf{G} = \begin{pmatrix} \sum \left(\frac{\partial}{\partial x} f_i^{(\sigma)} \right)^2 & \sum \left(\frac{\partial}{\partial x} f_i^{(\sigma)} \frac{\partial}{\partial y} f_i^{(\sigma)} \right) \\ \sum \left(\frac{\partial}{\partial x} f_i^{(\sigma)} \frac{\partial}{\partial y} f_i^{(\sigma)} \right) & \sum \left(\frac{\partial}{\partial y} f_i^{(\sigma)} \right)^2 \end{pmatrix} \quad (9)$$

Solving for \mathbf{s} yields the shift vector

$$\mathbf{s} = \mathbf{G}^{-1} \mathbf{v} \quad (10)$$

The size of the Gaussian (σ) allows scale selection and suppresses the high frequencies which are dominated by noise.

3.4 Complete set of reference frame

All three methods described above compute the shift between two images whereas our application requires the relative position of a set of P images. By calculating the shifts with respect to a single reference image we obtain just one realization of the relative positions. By repeating the procedure for another reference image we obtain a second estimate for the relative positions. By averaging P sets of relative positions (centered around their first moment) we may obtain a better estimate.

4. Interpolation by fusion of shifted images

Several image fusion schemes capable of interpolating randomly sampled images onto a equidistant grid have been developed and tested:

1. Least-squares fit (LSP) of a linear (plane) model to n data points ($n \geq 3$ for 2D) or to *all* data points inside a fixed-sized window.
2. Normalized convolution (NC) using either a fixed or variable (inversely proportional to the distance to the nearest neighbor) sized Gaussian kernel.
3. Exact (least-squares) image reconstruction (ER) from a series of shifted undersampled images. The number of images needs to be larger or equal to the amount of undersampling.
4. Iterative minimization of a functional which balances resolution improvement and noise suppression (IT). This method estimates the image shifts while optimizing the functional.

The first three methods require that the relative positions of the frames are known. The fourth can be used with and without this information. The computational complexity of the four methods varies enormously. LSP and NC are at

least an order of magnitude faster than ER or a single iteration of IT.

4.1 Least-squares plane fitting (LSP)

Fit a plane through the n data-points inside a window centered at the new sample position. The function value of the plane at the newly determined grid point yields the new sample value. The new sample point is required to lie inside the convex hull of the n points. The first three points are found by Delaunay triangulation. Every new sample point has three unique neighbors that span the current Delaunay triangle. The size of the window (and thus the number of data points used in the fit) is used to balance the interpolator between resolution improvement and noise suppression.

4.2 Normalized convolution (NC)

Normalized convolution [6] can be used as interpolator of randomly sampled data points

$$h(x, y) = \frac{f(x, y) \otimes c(x, y)}{s(x, y) \otimes c(x, y)} \quad (11)$$

with $f(x, y)$ the period-random weighted sum of delta-pulses

$$f(x, y) = \sum_{i=1}^P f_i \quad (12)$$

$$= \sum_{i=1}^P \sum_{n,m} f(nT_x - x_i, mT_y - y_i) \delta(x - nT_x + x_i, y - mT_y + y_i)$$

$c(x, y)$ the certainty function (a Gaussian function $g^{(\sigma)}$), $s(x, y)$ the periodic random sampling function, and T_x and T_y the pixel spacing in the input frames

$$s(x, y) = \sum_{i=1}^P \sum_{n,m} \delta(x - nT_x + x_i, y - mT_y + y_i) \quad (13)$$

The width of the Gaussian certainty function plays the role of the window size in the LSP method.

4.3 Exact reconstruction (ER)

If the shifts (x_i, y_i) and the amount of undersampling is known we can write in a series of linear equations the relation between the P instances of the Fourier components \mathbf{G}_{nm} and the target Fourier components \mathbf{F}_{nm}

$$\mathbf{G}_{nm} = \Phi_{nm} \mathbf{F}_{nm} \quad (14)$$

where the vector \mathbf{G}_{nm} contains the $(n, m)^{\text{th}}$ Fourier component of all input images, \mathbf{F}_{nm} the harmonics of (n, m) that where *aliased* onto (n, m) , and Φ_{nm} the $(P \times 2(K+L))$ transformation matrix

$$\Phi_{nm}(i, k + \ell) = \frac{1}{T} e^{jx_i \left(\frac{2\pi}{nT_x} n + (k-1-K)u_s \right)} \frac{1}{T} e^{jy_i \left(\frac{2\pi}{mT_y} m + (\ell-1-L)v_s \right)},$$

with i the frame index $i=0, \dots, P-1$, T_x and T_y the original sample spacing, u_s and v_s the original sampling frequencies, and Ku_s and Lv_s the new sampling frequencies. The solution is obtained by

$$\mathbf{F}_{nm} = \Phi_{nm}^{-1} \mathbf{G}_{nm}$$

This method [5,7] needs at least as many input frames as the undersampling factor ($P \geq 2(K+L)$). More images yield a least-squares estimation of the target Fourier spectrum.

$$\mathbf{F}_{nm} = (\Phi_{nm}^T \Phi_{nm})^{-1} \Phi_{nm}^T \mathbf{G}_{nm} \quad (15)$$

Note that we need to perform a matrix inversion of size $P \times P$ for each pixel in the $(N \times M)$ input image.

4.4 Iterative reconstruction (IT)

Several reconstruction methods employ iterative minimization of an error functional [3,4,9,10]. We implemented the method by Hardi [3] in which the functional consists of two terms

$$\min_h \left\{ \sum_i (f_i - \text{map}_i(h))^2 + \lambda \text{reg}(h) \right\}$$

The first term represents the difference between the measured image f_i and a version generated from the high-resolution image h and the second term penalizes sharp edges and noise. The parameter λ balances resolution vs smoothness. The minimization is performed by conjugate gradient [11]. This method requires a mathematical model

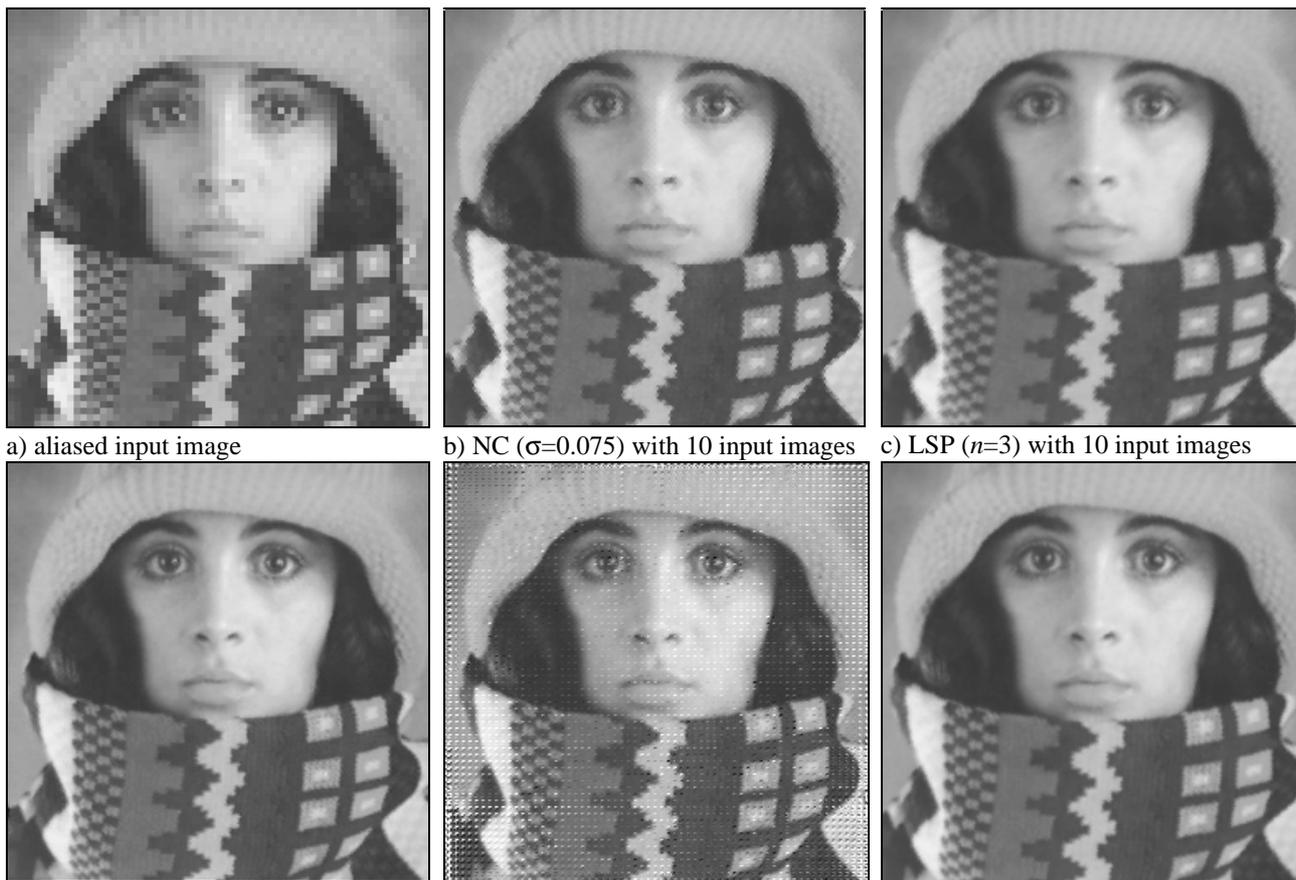
of the imaging process in the function $\text{map}()$. Estimation of the imaging parameters as occurs in “blind” deconvolution adds uncertainty and yields a lower performance.

5. Test results

The “registration” and “interpolation” modules have been tested separately. Synthetic image sequences were created by subsampling a target image (“trui” as in Figure 2) with random offsets. The subsampling of a factor four was performed by cubic interpolation of a properly sampled target image. Each image sequence consisted of 25 frames.

5.1 Results: Image registration

Several image registration methods have been tested on the synthetic image sequences. The errors in the estimated shifts are listed in Table 1. The MTS method will be used in practice because it is fast, accurate and precise. The MTS-all method did only use 10 of the 25 input frames as reference frame.



a) aliased input image b) NC ($\sigma=0.075$) with 10 input images c) LSP ($n=3$) with 10 input images
d) 10 iterations IT with 10 input images e) ER with 16 input images f) LSP ($n=3$) with 25 input images

Figure 2: Interpolation results. Note that the ER method may perform as well as IT and LSP (reasonably “uniform” distribution of samples) or as poor as shown in (e) for uneven distribution of samples.

Table 1: Performance of the registration algorithms in estimating the image shift between four-times subsampled (in both directions) versions of “Trui”.

	x- shift		y- shift	
	mean	std.dev.	mean	std.dev.
MTS	0.008	0.007	0.012	0.011
MTS-all	0.007	0.005	0.010	0.008
CZP4	0.10	0.063	0.096	0.057
CPF	0.019	0.018	0.020	0.016
CPF2	0.015	0.013	0.016	0.013

5.2 Results: Interpolation

The performance of the reconstruction was measured by the rms error between the reconstructed and the target image. The rms errors for all four methods using a varying number of frames are listed in Table 2. The result of ER depends heavily on the spatial distribution of image shifts. The performance degrades abruptly for unevenly distributed image shifts, which occurred in a few of the tested sequences. Such a result is shown in Figure 2e. It is a periodic distortion which is caused by stability problems associated with the inversion of ill-posed matrices. The iterative method (IT) performs best at the expense of an enormous computational load. The LSP interpolator is the second best choice at a small fraction of the computational cost.

Table 2: rms error for various interpolation methods and number of input frames. The grey-values of the images are in the interval [0,1].

	$n=10$	$n=16$	$n=25$
LSP ($n=3$)	$1.8 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$	$8.8 \cdot 10^{-3}$
LSP ($n=6$)	$2.2 \cdot 10^{-2}$	$1.6 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$
NC ($\sigma = 0.075$)	$2.3 \cdot 10^{-2}$	$1.6 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$
NC ($\sigma = 0.15$)	$2.2 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$	$1.3 \cdot 10^{-2}$
ER	no sol.	$8.0 \cdot 10^{-2}$ *	$1.8 \cdot 10^{-2}$
IT (10 iter.)	$9.1 \cdot 10^{-3}$	$5.6 \cdot 10^{-3}$	$3.0 \cdot 10^{-3}$

*) See Figure 2e.

5.3 Robustness of LSP interpolator

The performance of the LSP interpolator was tested to indicate its robustness in the presence of noise with a variable number of input frames. As interpolation should carefully balance between resolution and SNR we measured the rms error also as a function of window size (Figure 3) or the number of points used in least-squares plane fit (Figure 4). Figure 3 and Figure 4 show the performance of the 1D LSP interpolator applied to a severely undersampled Gaussian function $\exp(-x^2/2\sigma^2)$ (sampling distance $T=4\sigma$). The error behavior of the NC interpolator (not shown) is similar to the LSP results.

It is clear that the rms error slowly increases for decreasing number of frames. The optimal window size is denoted by the lowest point in the error landscape and depends on the SNR. Figure 5 shows that the noise on the sample values and the noise on the estimated shifts have a similar effect on the performance.

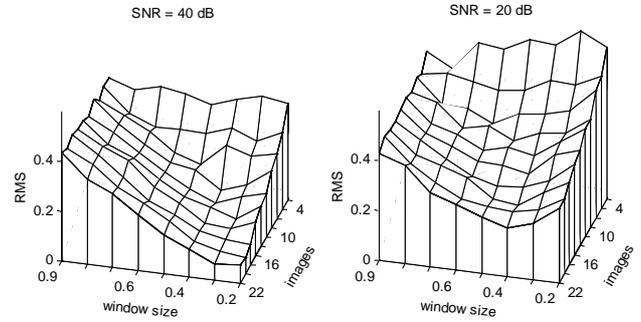


Figure 3: Rms error of the 1D LSP interpolator as a function of the number of frames and a fixed window size (variable number of points). The window size is a fraction of the original pixel spacing.

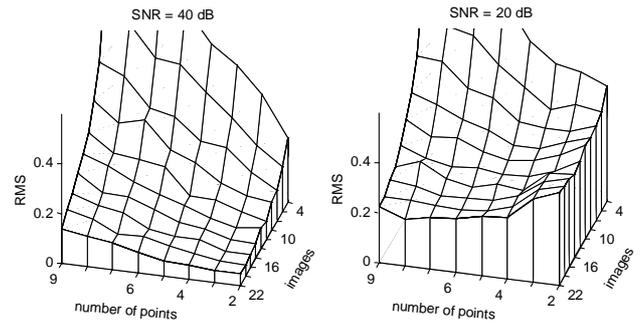


Figure 4: Rms error of the 1D LSP interpolator as a function of the number of frames and the number of points (variable window size).

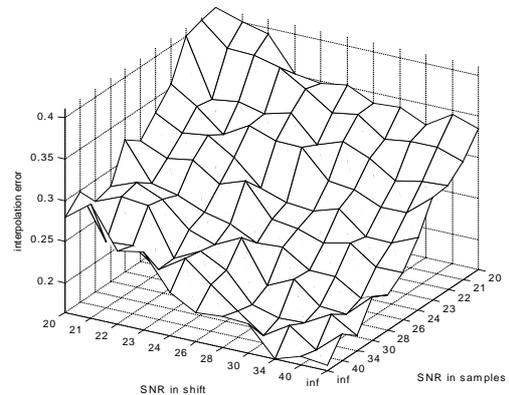


Figure 5: Rms error of the LSP estimator as a function of the shift SNR and the amplitude or sample SNR.

6. Application in Infrared imaging

The test sequences have been replaced by real-world image sequence acquired by a vibrating infrared camera in the long wavelength range (7-10 μm). The sensor has an array of size 128x128 pixels, a pixel pitch of 40 μm and a fill factor of 0.64. The output images (Figure 6a and Figure 7a) have a sampling density that is four times higher in both directions (16x as many pixels). The shifts are estimated by MTS and the interpolation is performed by LSP or NC.

The results shown in Figure 6b-c and Figure 7b-c contain the same region-of-interest (ROI) before and after four-times (in each direction) resolution improvement. The method used 25 input frames, MTS uses a Gaussian (smoother and gradient) of $\sigma=1$, NC uses a Gaussian certainty function of size $\sigma=0.075$ of the original pixel pitch, LSP uses a fixed number of samples ($n=3$).

7. Discussion

This paper presents a system for significantly improving the resolution of undersampled (aliased) image sequences in which the frames are shifted over a random subpixel distance by random motion of the camera (vibration). The system uses the MTS shift estimator and the LSP or NC interpolator. The MTS shift estimator is both fast and yields the best performance in our comparative study. The LSP as well as the NC interpolator are sub-optimal choices, but both methods have a computational complexity that is much lower than the best performing iterative method (IT). The system performs well over a range of SNR's that occur in practice. The performance degrades "gracefully" when the noise increases or the number of input frames decreases. The system meets all requirements and yields high quality output image in infrared imaging.

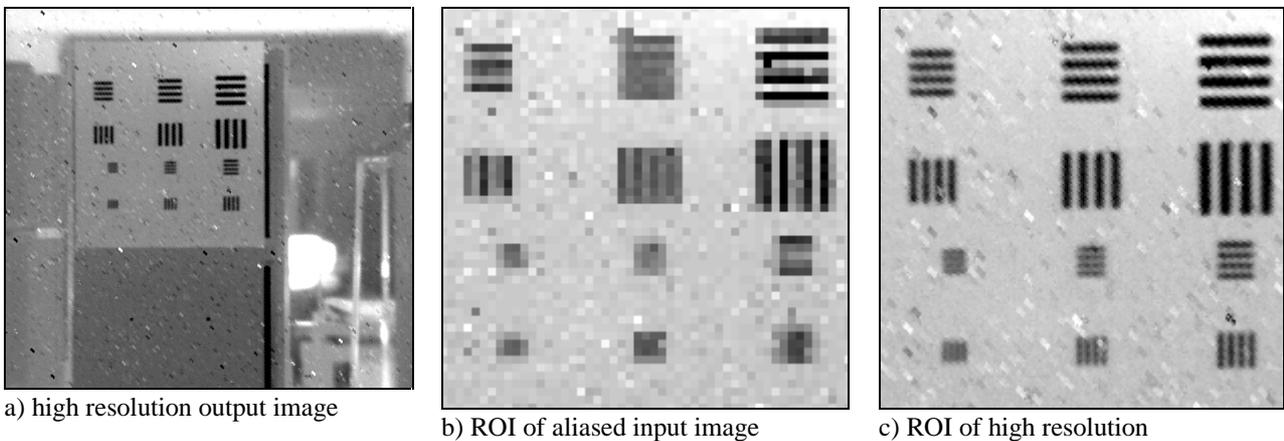


Figure 6: Image of a camera test chart with bar patterns of increasing frequency. The output images (a) and (c) have a four times higher sampling rate in both directions using MTS ($\sigma = 1$) and LSP ($n = 3$ samples) with 25 frames. All images are contrast stretched.

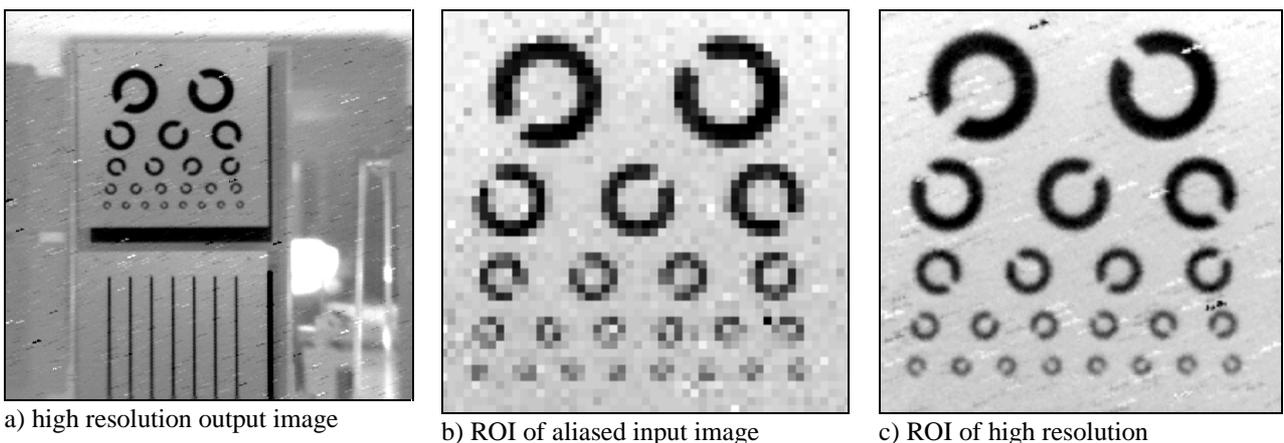


Figure 7: Image of a camera test chart: rings with opening of decreasing size. The output images (a) and (c) have a four times higher sampling rate in both directions using MTS ($\sigma = 1$) and NC ($\sigma = 0.075$) with 25 frames. All images are contrast stretched.

8. Acknowledgments

We are grateful to dr. K. Schutte and dr. A.L.D. Beckers from TNO-FEL in The Netherlands for providing the IR data sets and the helpful discussion during the course of the project. This work was partially supported by the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Rolling Grants program of the Foundation for Fundamental Research in Matter (FOM).

9. References

- [1] Alam, M.S., et al., “*High Resolution Infrared Image Reconstruction Using Multiple, Randomly Shifted, Low Resolution, Aliased Frames*”, SPIE proc. vol. 3063-9, 1997.
- [2] Gillette, C., T.M. Stadtmiller and R.C. Hardie, “*Aliasing Reduction in Staring Infrared imagers Utilizing Subpixel Techniques*”, Optical Eng., vol. 34:11, 3130-3137, 1995.
- [3] Hardie, R.C., et al., “*High Resolution Image Reconstruction From a Sequence of Rotated and Translated Infrared Images*”, SPIE Proceedings vol. 3063-10, 1997.
- [4] Hardie, R.C., K.J. Barnard and E.E. Armstrong, “*Joint MAP Registration and High Resolution Image Estimation Using a Sequence of Undersampled Images*”, IEEE Trans on Image Processing, vol. 6, no. 12, pp. 1621-1633, 1997.
- [5] Kaltenbacher, E.A., and R.C. Hardie, “*High Resolution Infrared Image Reconstruction Using Multiple, Low Resolution, Aliased Frames*”, SPIE proc., vol. 2751, pp. 142-152, 1996.
- [6] Knutsson, H., and C.F. Westin, “*Normalized Convolution: A Technique for Filtering Incomplete and Uncertain Data*”, Proc 8th SCIA, vol. 2, pp. 997-1006, 1993.
- [7] Marvasti, F., “*Nonuniform Sampling Theorems for Bandpass Signals at or Below the Nyquist Density*”, IEEE Trans on Signal Processing, vol. 44:3, pp. 572-576, 1996.
- [8] Oppenheim, A.V., A.S. Willsky and I.T. Young, “*Signals and Systems*”, Prentice-Hall International, Inc., 1983.
- [9] Patti, A.J., M. Ibrahim Sezan and A. Murat Tekalp, “*Superresolution Video Reconstruction with Arbitrary Sampling Lattices and Nonzero Aperture Time*”, IEEE Trans. on Image Processing, vol. 6:8, pp. 1064-1076, 1997.
- [10] Schultz, R.R., and R.L. Stevenson, “*Extraction of High Resolution Frames from Video Sequences*”, IEEE Trans on Image Processing, vol. 5, no. 6, pp. 996-1011, 1996.
- [11] Shewchuk, J.R., “*An Introduction to the Conjugate Gradient Method Without the Agonizing Pain*”, <ftp://warp.cs.cmu.edu/quake-papers/painless-conjugate-gradient.ps>, 1994.
- [12] Srinivasa Reddy, B., and B.N. Chatterji, “*An FFT-Based Technique for Translation, Rotation, and Scale-Invariant Image Registration*”, IEEE Transactions on Image Processing, vol. 5, no. 8, pp. 1266-1271, 1996.