Resolution Enhancement of a Sequence of Undersampled Shifted Images

Cris L. Luengo Hendriks Lucas J. van Vliet

Pattern Recognition Group
Faculty of Applied Sciences
Delft University of Technology
Lorenzweg 1, 2628 CJ Delft, The Netherlands

cris@ph.tn.tudelft.nl

Keywords: sub-pixel image registration, interpolation, random sampling, aliasing, superresolution.

Abstract

This paper presents various methods to increase the spatial resolution of an undersampled, and thus aliased, image sequence. Random global translation between frames in the sequence provides information that can be used to remove some or all aliasing present in the frames. Translation vectors for each frame are estimated from the image content. Using these estimates, a high-resolution image is computed. The proposed methods are of moderate computational complexity and can be implemented in hardware for real-time use. They also prove to be robust under noisy circumstances. Previous approaches to the superresolution restoration problem involve iterative algorithms, which are not practical for real-time application. The proposed methods perform comparably only if enough input frames are provided. However, they give acceptable results even when these criteria are not met.

1 Introduction

The lens system in a camera limits the bandwidth of the image projected onto the detector array. The detector array can then sample this image. If the sampling satisfies the Nyquist criterion, the image projected onto the detector can be exactly reconstructed from the samples. A typical detector array used in infrared imaging has a small fill factor, due to the need to isolate each detector element separately. However, to have detector elements large enough for efficient light collection, it is necessary to limit the amount of detectors on the array. This causes the array not to be sufficiently dense to sample according to Nyquist, and the sampled images are aliased.

Some IR camera manufacturers solve this undersampling problem by mounting the detector array on a piezo-electric element that allows scanning of the image projected onto the detector plane ('microscanning'). For example, Carl-Zeiss incorporates this technique is the ProgRes digital camera.

The approach explored in this paper is to induce a random motion by vibration of the camera, and combine a series of frames to create a single image with a higher spatial resolution. Note that this can only work if there is a sub-pixel motion between those frames. This translation causes the image to be sampled at more points than provided by the detector array. Also, note that resolution cannot be improved upon if the frames are sampled correctly (according to Nyquist). The requirements that the algorithms must meet are:

- low computational cost and, more importantly, simple implementation and non-iterative algorithms (to make hardware implementation possible);
- ability to suppress noise as well as improve resolution;
- comparable image quality with respect to existing methods; and
- graceful degradation of performance for increasing amounts of noise, absence of reliable image content or lower number of input frames.

We will assume that the scene does not change during the acquisition of the image sequence, and that the motion is small and random, as if caused by vibration of the camera. Section 2 presents the imaging model.

Superresolution restoration requires knowledge about the relative shifts between the frames. This is extracted from the frames themselves using one of the registration algorithms outlined in section 3. The second step is to fuse the data present in all the frames. We have developed and tested some methods that accomplish this. They are presented in

in: M. Boasson, J.A. Kaandorp, J.F.M. Tonino, M.G. Vosselman (eds.), ASCI'99, Proc. 5th Annual Conference of the Advanced School for Computing and Imaging (Heijen, NL, June 15-17), ASCI, Delft, 1999, 95-102.

section 4. A proper evaluation requires extensive comparison with methods presented in the literature [1-5, 9, 10]. Section 5 shows the test results, as well as some results obtained on infrared images.

2 The Imaging Model

Figure 1 shows the imaging model. The object scene is slightly blurred by the lens (h_{lens}), introducing a cutoff frequency. It is then shifted (h_{shift}), integrated over the surface of a detector element (h_{pixel}), sampled ($p(x,y)$), and digitized (ADC). The output image is undersampled and degraded by noise of various sources, including readout noise, thermal noise, quantization noise and, most importantly, photon noise. The signal-to-noise ratio (SNR) of these cameras can be as high as 40 dB.

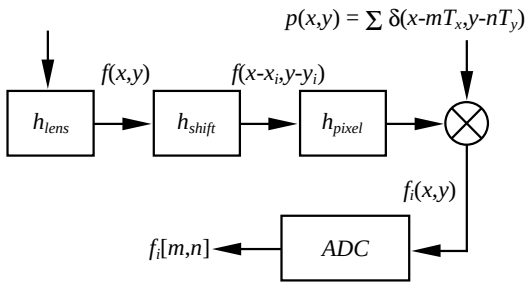


Figure 1: Model for image acquisition.

The motion blur in the individual frames is negligible because of the small exposure time (less than 5 ms in the camera we used for our experiments). The influence of the square integrating pixel elements can be ignored (due to a small fill factor) for interpolation factors up to four.

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.

2.1 How many frames do we need?

If the undersampling we are trying to undo is equal to n in both dimensions, we will create an image with n^2 times the number of pixels, and thus information, than was in one original frame. It is therefore easy to see that we will at least require n^2 frames to have enough information to fill one high-resolution image. This will, however, depend strongly on the distribution of the shifts, as well as on the signal to noise ratio.

Nyquist's sampling theorem states that a bandlimited signal with a cutoff frequency of ω_c should be sampled with a frequency $\omega_s > 2\omega_c$, so no information is lost [8]. It can, however, also be interpreted as needing N samples in each period NT, where $T = 2\pi/\omega_s$, the uniform sampling period [7]. Notice that in the last definition, the sample positions are not taken into account.

To mathematicians this is not new. For example, a polynomial of order n is completely defined by n+1 points, no matter how close together these points are. However, in mathematics data is never noisy. Here, we are dealing with noisy data, so the position can indeed be of influence. It is common sense to place the available samples as far apart as possible, to minimize the influence of noise (see figure 2).

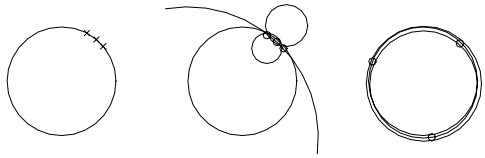


Figure 2: Any three exact samples define a circle (left). However, if the samples contain noise and are situated close to one another, almost any circle will fit (middle). Positioning these samples far apart insures a correct representation in spite of the noise.

Thus, depending on the distribution of the samples, we might need more than n^2 frames to correctly reconstruct the high-resolution image. Reconstruction of an image with too few frames is sometimes possible, but we should not expect to achieve the desired resolution. Furthermore, limits to the resolution of the output image are set by the resolution of the lens system and the fill factor of the detector array.

3 Image Registration

As stated in the previous section, we will assume that there is only a global translation over the entire image. This simplifies the registration as well as the interpolation.

For a correct detection of shifts, the image must contain some features that make it possible to match two undersampled images. Very sharp edges and small details are most affected by aliasing, so they are not reliable to be used to estimate these shifts. Uniform areas are also useless, since they are translation-invariant. The best features are slow transitions between two grayvalues (temperature ramps), which are unaffected by aliasing. Such portions of an image need not be detected, but their presence is very important for an accurate result.

We have implemented three methods for estimating the global shift between two images $f_0(x, y)$ and $f_1(x, y) = f_0(x-x_1, y-y_1)$.

1) Apply cross-correlation followed by low-pass filtering and zero-padding in the Fourier domain, and center-of-mass estimation in the spatial domain (CZP).
2) Apply cross-correlation and fit a plane through the low frequencies of the 2-D phase of the Fourier transform (CPF).
3) Minimize the squared error between a first order Taylor series expansion of f_0 and f_1 (MTS).

3.1 Cross-correlation with zero-padding (CZP)

The two images being compared are shifted realizations of the same scene. This implies that the Fourier transforms are equal in magnitude, but differ in phase [12]. The cross-correlation can be written in the Fourier domain as a multiplication,

$$\mathrm{F}\left\{\iint f_0(\eta,\xi) f_1(x+\eta, y+\xi)\, d\eta\, d\xi\right\} = F_0^*(u,v) F_1(u,v) \quad (1)$$

with u and v the spatial frequencies. Normalization yields

$$\frac{F_0^*(u,v) F_1(u,v)}{\left|F_0(u,v)\right|^2} = e^{-j(ux_1+vy_2)} \quad . \quad (2)$$

In the space domain, this is equal to a delta function positioned by the shift vector (x_1, y_1).

As the higher frequencies are most affected by noise and aliasing, we multiply the frequency spectrum with a Gaussian to remove the higher frequencies (using a σ of approximately $1/6^{\text{th}}$ of the image width). This converts the peak in the spatial domain to a Gaussian. The frequency spectrum is then padded with zeros. This interpolates the peak, making it larger. The center-of-mass gives the estimated sub-pixel shift.

Using only the samples in a small rectangular window around the maximum, we make optimal use of the fact that most of the signal is in the peak itself. Points outside this area contribute very little information, but with a very low SNR. Adding these points to the estimate will lower the accuracy.

3.2 Cross-correlation with phase fitting (CPF)

A variation to the cross-correlation method uses eq. 2, and fits a plane through the argument of the exponential. This is done using least squares, and only on the lower frequency components (the same part of the spectrum that is used in section 3.1). The parameters of the plane are the shifts in the x and y directions.

This technique can be refined by using only those values of u and v for which the amplitude of eq. 2 is sufficiently close to one (between 0.9 and 1.1). Further refinement is accomplished by removing the outliers from the phase term after which the fit is repeated. We define outliers as points with an error of more than 3 times the standard deviation.

As the phase angle of a complex number is always in the range $[-\pi, \pi]$, it might wrap around for large shifts (more than one pixel). This makes the fitting impossible.

However, it is possible to estimate the integer pixel shift by doing an inverse Fourier transform of eq. 2, and locating the maximum of the peak. Once one of the images is corrected for this shift, the sub-pixel shift can be correctly obtained with this method.

3.3 First order Taylor series (MTS)

The reference image f_0 can be approximated by a first order Taylor series of f_1

$$f_0 \approx f_1 + x_1 \frac{\partial}{\partial x} f_1 + y_1 \frac{\partial}{\partial y} f_1 \quad . \quad (3)$$

Minimizing the squared difference of both sides, we obtain a sub-pixel shift estimator as in [1, 5].

The first order Taylor series requires a gradient operator. We propose the use of the Gaussian gradient. This requires both images to be convolved with the Gaussian. As a positive side effect, this removes higher frequencies, which are most affected by the aliasing. The resulting Taylor series reads

$$f_0^{(\sigma)} \approx f_1^{(\sigma)} + x_1 \frac{\partial}{\partial x} f_1^{(\sigma)} + y_1 \frac{\partial}{\partial y} f_1^{(\sigma)} \quad , \quad (4)$$

with

$$f_1^{(\sigma)} = f_1 \otimes g^{(\sigma)}$$

$$\frac{\partial}{\partial x} f_1^{(\sigma)} = f_1 \otimes \frac{\partial}{\partial x} g^{(\sigma)} \quad . \quad (5)$$

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

The estimated shifts are only accurate if they are small enough, in comparison to the width of the Gaussian. Large shifts are estimated less accurately because of the larger scale at witch the images are looked at. To overcome this, one of the images can be corrected for the integer-pixel shift first, and then the remaining shift can be determined at a smaller scale ($\sigma = 1$ pixel). This can be done in two ways:

1) Estimating the integer-pixel shift as in the CPF method.
2) Estimating the shift repeatedly using $\sigma = 1$. If the resulting shift is too large to be accurate (larger than one half of the pixel pitch), one of the images is shifted one pixel in the correct

direction, and the process is repeated. Note that this is actually very cheap since none of the convolutions needs to be repeated. It is sufficient to remove rows of pixels at the borders to correct for integer pixel shift.

3.4 Increasing accuracy

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 on their first moment), we may obtain a better estimate.

4 Image Interpolation

The data to be interpolated is randomly distributed, because of the random, sub-pixel shifts of each frame. Nearest neighbor interpolation provides an easy solution to this problem [1, 2], but it is less accurate than other methods. It also is not able to suppress noise.

Popular interpolation techniques like B-splines and cubic convolution are difficult to implement or cannot be used at all on this non-uniformly sampled data.

We studied two alternative interpolation techniques (LSP, NC), as well as two other reconstruction algorithms found in the literature (ER, IT).

1) Interpolation by least-squares fitting of a linear model (plane) to *n* data points (*LSP*).
2) Interpolation by normalized convolution using a Gaussian kernel (*NC*).
3) Exact image reconstruction by estimating the frequency spectrum (*ER*).
4) Iterative minimization of a cost function based on a detailed description of the imaging system (*IT*).

4.1 Least-squares plane fitting (LSP)

After selecting *n* data points in the vicinity of the new sample position, a plane is fitted through these points. The function value of this plane at the new position is used as the new sample value.

The new sample point is required to lie inside the convex hull of the *n* points. This is accomplished by selecting the first three points using Delaunay triangulation. Every new sample point has three unique neighbors that span the current Delaunay triangle. The remaining points can be selected in a variety of ways. For simplicity of implementation, we select them by distance. This, however, might not be the optimal solution in the case that the shifts are not evenly distributed (which they often are not).

The number of data points used in the fit is tuned to balance between resolution improvement and noise suppression. This parameter can be chosen once for the whole image, or differently for each pixel to be calculated (depending on the local density of the input samples).

4.2 Normalized convolution (NC)

Normalized convolution [6] can be used as an interpolator of randomly sampled data points. It is based on the convolution of the samples with a kernel, normalizing the result to account for the non-uniformity of the samples. The (continuous) interpolated image is written as

$$h(x,y) = \frac{f(x,y) \otimes k(x,y)}{s(x,y) \otimes k(x,y)} \quad , \tag{6}$$

with $f(x, y)$ the weighted sum of delta-pulses representing the input samples, $k(x, y)$ the convolution kernel, and $s(x, y)$ the sampling function.

$$f(x,y) = \sum_{i=1}^{P} \sum_{m,n} f\left(mT_x + x_i, nT_y + y_i\right) \cdot \delta\left(x - mT_x - x_i, y - nT_y - y_i\right) \quad , \tag{7}$$

$$s(x,y) = \sum_{i=1}^{P} \sum_{m,n} \delta\left(x - mT_x - x_i, y - nT_y - y_i\right) \quad . \tag{8}$$

Here we use T_x and T_y as the sampling periods, and x_i and y_i the shifts of image *i*.

We have chosen the Gaussian function as kernel. One of its advantages is that it is defined and different from 0 over the entire range $(-\infty,\infty)$, so $h(x, y)$ in eq. 6 is defined for all (x,y), no matter how sparse the samples are. Other kernels, however, might be easier to compute.

The width of the kernel function can be set to balance between resolution improvement and noise suppression, just like the number of samples in the LSP method. It is possible to have it fixed or adapted to the local sample density.

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 $\mathbf{G}_{mn}$ and the target Fourier components $\mathbf{F}_{mn}$ [5].

$$\mathbf{G}_{mn} = \Phi_{mn}\,\mathbf{F}_{mn} \qquad (9)$$

where the vector $\mathbf{G}_{mn}$ contains the $(m,n)^{th}$ Fourier component of all input images, $\mathbf{F}_{mn}$ the harmonics of (m,n) that where *aliased* onto (m,n), and Φ_{mn} the (P x $2(K+L)$) transformation matrix based upon the estimated shifts.

The solution is obtained by a matrix inversion of size PxP for each pixel in the (MxN) input image. This requires as many input frames as the undersampling factor. If more input frames are available a least-squares solution is possible. Note that matrix inversion can be numerically unstable if the shifts in the input are not evenly distributed.

4.4 Iterative reconstruction (IT)

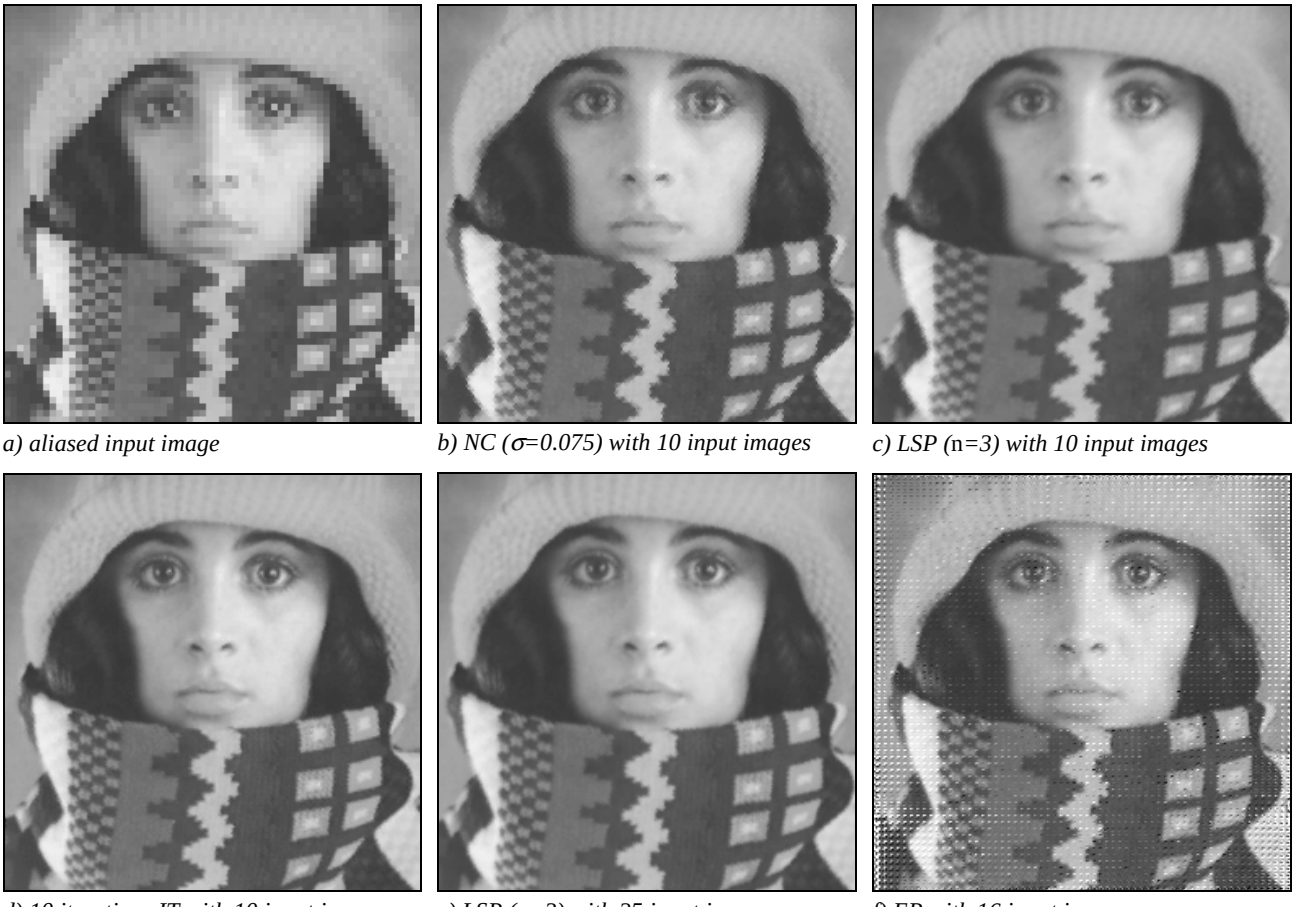
Several reconstruction methods employ iterative minimization of an error functional [3, 4, 9, 10]. We implemented the method by Hardi [4] in which the functional consists of two terms

$$\min_h \left\{ \sum_i \left(f_i - \mathrm{map}_i(h) \right)^2 + \lambda\, \mathrm{reg}(h) \right\} \quad . \qquad (10)$$

The first term represents the difference between the measured images f_i and the version generated from the high-resolution image h and the second term penalizes sharp edges and noise. The parameter λ balances resolution versus smoothness. The minimization is performed by conjugate gradients [11]. This method requires a mathematical model of the imaging process in the function $\mathrm{map}_i(h)$.

5 Results

The modules "registration" and "interpolation" have been tested separately. Synthetic image sequences were created by subsampling a properly sampled target image ("trui" as in figure 3a) with random offsets, using cubic interpolation. The subsampling factor was four, and the resulting frames have 64^2 pixels.



a) aliased input image *b) NC (σ=0.075) with 10 input images* *c) LSP (n=3) with 10 input images*

d) 10 iterations IT with 10 input images *e) LSP (n=3) with 25 input images* *f) ER with 16 input images*

Figure 3: 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 (f) for uneven distribution of samples.

5.1 Image registration

The performance of the registration algorithms was measured by estimating the known shift between two synthetic images. The mean error over 490 realizations is shown in table 1. The same set of images was used for all algorithms. The CZP algorithm employs zero padding to obtain a grid four times as large in both dimensions. CPF_2 is the same as CPF, using the removal of outliers as an improvement. The MTS algorithm is both the cheapest in computational complexity and the most accurate.

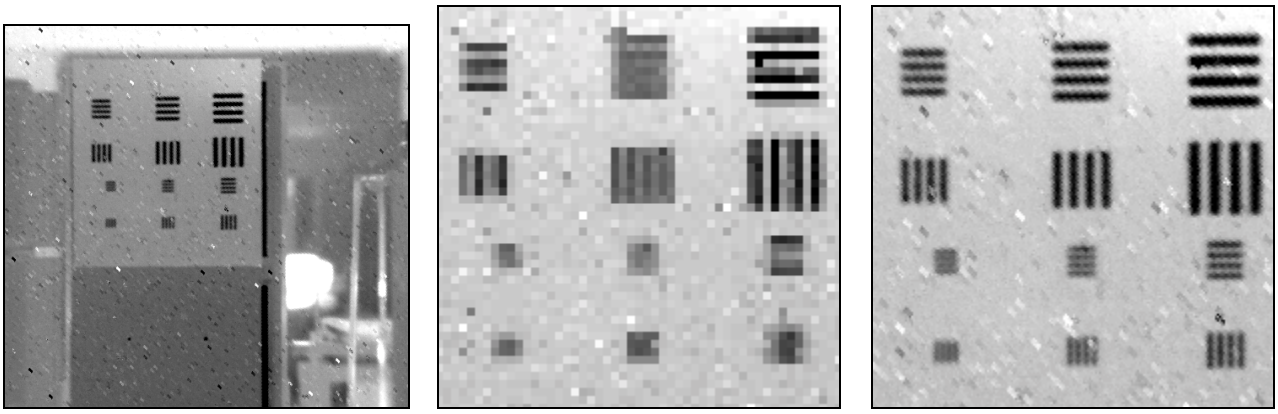
The MTS algorithm has also been extended as in section 3.4. We tested it by estimating the shifts of groups of 10 images. The mean error decreased, but not significantly.

Table 1: Errors in the estimated shift for the test image sub-sampled by a factor four. Means and standard deviations are over 490 realizations.

	x shift		y shift	
	mean	**std.dev.**	**mean**	**std.dev.**
CZP	0.019	0.015	0.040	0.030
CPF	0.017	0.014	0.021	0.017
CPF_2	0.013	0.011	0.019	0.015
MTS	0.009	0.007	0.012	0.010

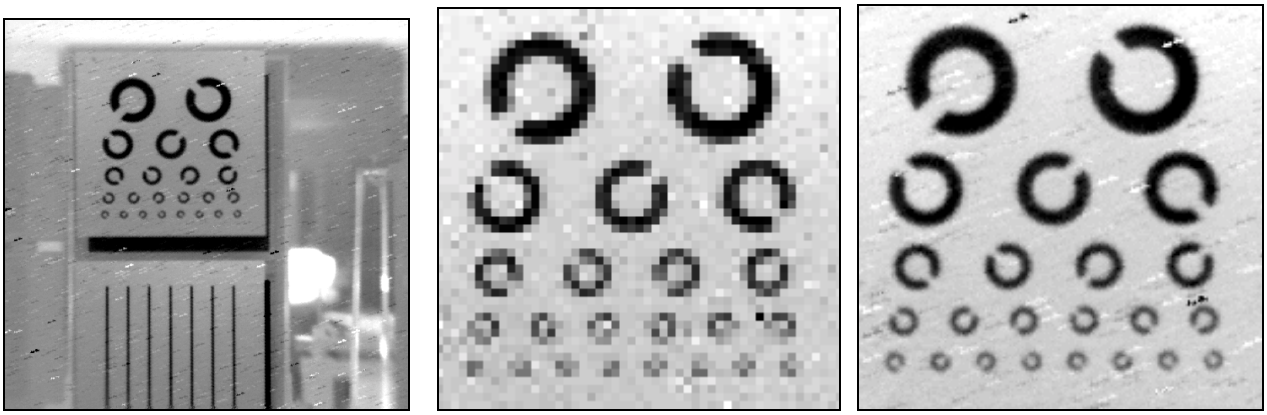
5.2 Image interpolation

The algorithms from section 4 were tested by reconstructing a high-resolution image using different numbers of input frames *p*, and calculating the RMS difference between the result and the original image. The relative shifts are known exactly. The results are shown in table 2. Some of the images obtained are shown in figure 3.



a) high resolution output image *b) ROI of aliased input image* *c) ROI of high resolution*

*Figure 4: 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 (σ = 1) and LSP (*n = 3 samples) with 25 frames. All images are contrast stretched.*



a) high resolution output image *b) ROI of aliased input image* *c) ROI of high resolution*

Figure 5: 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 (σ = 1) and NC (σ = 0.075) with 25 frames. All images are contrast stretched.

Table 2: RMS difference between the reconstructed images and the original one. As a comparison measure, the RMS of the bicubic interpolated image is $5.1 \cdot 10^{-2}$. Grayvalues of the images are in the interval [0,1].

	p = 10	**p = 16**	**p = 25**
LSP, n = 3	$1.8 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$	$8.9 \cdot 10^{-3}$
LSP, n = 6	$2.2 \cdot 10^{-2}$	$1.6 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$
NC, σ = 0.3	$2.3 \cdot 10^{-2}$	$1.6 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$
NC, σ = 0.6	$2.2 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$	$1.3 \cdot 10^{-2}$
ER	*no solution*	$8.0 \cdot 10^{-2}$	$1.8 \cdot 10^{-2}$
IT (10 iter.)	$9.1 \cdot 10^{-3}$	$5.6 \cdot 10^{-3}$	$3.0 \cdot 10^{-3}$

These results show that the ER method does not work well for random shifts, especially if only the minimum amount of input frames is available. Both the LSP and the NC methods are much faster than the IT algorithm (about 3 times faster than a single iteration). Although the accuracy of these interpolation algorithms is not as good as the iterative scheme, the results are satisfactory.

To examine the tolerance of the proposed interpolation algorithms to errors in the shift estimates and noise, we tested a one-dimensional variant of both LSP (not shown) and NC (figure 6), adding noise to both the samples and the (known) shifts. The figure shows the RMS error of the result compared to the original (ideal) 1D image. We can see that this algorithm is approximately equally sensitive to both kinds of noise. Furthermore, it is apparent that the accuracy needed in the shift estimates is in the order of $1/20^{th}$ of the pixel pitch (for upsampling of a factor four). This accuracy is achieved with both the MTS and the CPF methods.

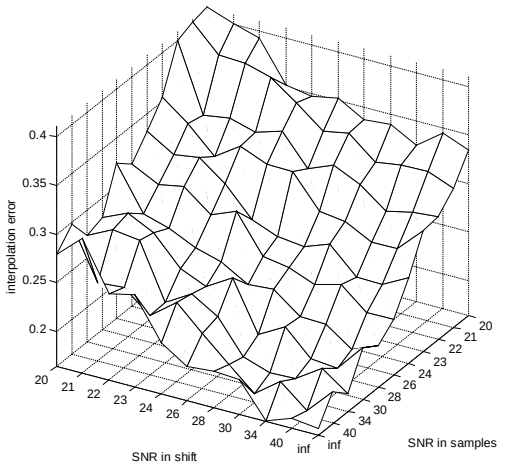


Figure 6: RMS error of the LSP estimator as a function of the shift SNR and the amplitude or sample SNR, on 1D images.

5.3 Application to infrared imaging

We have acquired two sequences of 25 test targets using a vibrating infrared camera in the long wavelength range (7 to 10 μm). The detector array has 128x128 pixels, a pixel pitch of 40 μm, and a fill factor of 0.64. The output has a pixel density four times as high in both directions (16 times as many pixels). Shifts were estimated using MTS (σ = 1), and the interpolation was done using LSP (n = 3) and NC (σ = 0.075). Some results are displayed in figure 4 and figure 5.

Note that the hot and cold pixels produced by the detector array during acquisition were not removed prior to the application of these algorithms.

6 Conclusion

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

7 Acknowledgment

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).

References

[1] M. S. Alam, J. G. Bognar, R. C. Hardie, and B. Yasuda, J,, "High Resolution Infrared Image Reconstruction Using Multiple, Randomly Shifted, Low Resolution, Aliased Frames," *SPIE Proceedings*, vol. 3063, 1997.

[2] C. Gillette, T. M. Stadtmiller, and R. C. Hardie, “Aliasing Reduction in Staring Infrared Imagers Utilizing Subpixel Techniques,” *Optical Engineering*, vol. 34, pp. 3130-3137, 1995.

[3] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, “Joint MAP Registration and High Resolution Image Estimation Using a Sequence of Undersampled Images,” *IEEE Transactions on Image Processing*, vol. 6, pp. 1621-1633, 1997.

[4] R. C. Hardie, S. Cain, K. J. Barnard, J. Bognar, E. Armstrong, and E. A. Watson, “High Resolution Image Reconstruction From a Sequence of Rotated and Translated Infrared Images,” *SPIE Proceedings*, vol. 3063, 1997.

[5] E. A. Kaltenbacher and R. C. Hardie, “High Resolution Infrared Image Reconstruction Using Multiple, Low Resolution, Aliased Frames,” *SPIE Proceedings*, vol. 2751, pp. 142-152, 1996.

[6] H. Knutsson and C. F. Westin, “Normalized Convolution: A Technique for Filtering Incomplete and Uncertain Data,” presented at SCIA'93, 1993.

[7] F. Marvasti, “Nonuniform Sampling Theorems for Bandpass Signals at or Below the Nyquist Density,” *IEEE Transactions on Signal Processing*, vol. 44, pp. 572-576, 1996.

[8] A. V. Oppenheim, A. S. Willsky, and I. T. Young, *Signals and Systems*: Prentice-Hall International, Inc., 1983.

[9] A. J. Patti, M. Ibrahim Sezan, and A. Murat Tekalp, “Superresolution Video Reconstruction with Arbitrary Sampling Lattices and Nonzero Aperture Time,” *IEEE Transactions on Image Processing*, vol. 6, pp. 1064-1076, 1997.

[10] R. R. Schultz and R. L. Stevenson, “Extraction of High Resolution Frames from Video Sequences,” *IEEE Transactions on Image Processing*, vol. 5, pp. 996-1011, 1996.

[11] J. R. Shewchuk, “An Introduction to the Conjugate Gradient Method Without the Agonizing Pain,” ftp://warp.cs.cmu.edu/quake-papers/painless-conjugate-gradient.ps, 1994.

[12] B. Srinivasa Reddy and B. N. Chatterji, “An FFT-Based Technique for Translation, Rotation, and Scale-Invariant Image Registration,” *IEEE Transactions on Image Processing*, vol. 5, pp. 1266-1271, 1996.