Performance study on point target detection using super-resolution reconstruction

Judith Dijk[a], Adam W.M. van Eekeren[ab], Klamer Schutte[a]
Dirk-Jan J. de Lange[a], Lucas J. van Vliet[b]

[a]Electro Optics Group
TNO Defence, Security and Safety
P.O. Box 96864, 2509 JG, The Hague, the Netherlands;
[b] Quantitative Imaging Group
Faculty of Applied Sciences
Delft University of Technology
Lorentzweg 1, 2628 CJ Delft, the Netherlands

ABSTRACT

When bright moving objects are viewed with an electro-optical system at very long range, they will appear as small slightly blurred moving points in the recorded image sequence. Detection of point targets is seriously hampered by structure in the background, temporal noise and aliasing artifacts due to undersampling by the infrared (IR) sensor.

Usually, the first step of point target detection is to suppress the clutter of the stationary background in the image. This clutter suppression step should remove the information of the static background while preserving the target signal energy. Recently we proposed to use super-resolution reconstruction (SR) in the background suppression step. This has three advantages: a better prediction of the aliasing contribution allows a better clutter reduction, the resulting temporal noise is lower and the point target energy is better preserved.

In this paper the performance of the point target detection based on super-resolution reconstruction (SR) is evaluated. We compare the use of robust versus non robust SR reconstruction and evaluate the effect of regularization. Both of these effects are influenced by the number of frames used for the SR reconstruction and the apparent motion of the point target. We found that SR improves the detection efficiency, that robust SR outperforms non-robust SR, and that regularization decreases the detection performance. Therefore, for point target detection one can best use a robust SR algorithm with little or no regularization.

Keywords: Super-resolution reconstruction, robust, regularization, evaluation

1. INTRODUCTION

In surveillance applications moving targets need to be detected early. Electro-optical surveillance systems observe distant missiles or other incoming threats as moving point targets. At maximum detection range these point targets will have a low signal-to-noise ratio with respect to the background. Detection is further hampered by high contrast structure (clutter) in the background. In addition to this, the sensor further complicates detection by undersampling of the signal and adding temporal noise.

Usually, the first step of point target detection is to suppress the clutter of the stationary background in the image. A clutter suppression step should remove the information of the static background while preserving the target signal energy. A standard background suppression technique is to align and subtract subsequent frames (AS).

Recently[1] we proposed to use super-resolution reconstruction (SR) in the background suppression step. This has the advantage that 1) a better prediction of the aliasing contribution substantially reduces the clutter related error in the difference image, 2) the temporal noise in the difference image is reduced, and 3) the

Further author information: Send correspondence to Judith Dijk (judith.dijk@tno.nl)

Automatic Target Recognition XIX, edited by Firooz A. Sadjadi, Abhijit Mahalanobis,
Proc. of SPIE Vol. 7335, 73350M · © 2009 SPIE · CCC code: 0277-786X/09/$18 · doi: 10.1117/12.819093

point target amplitude in the difference image is higher. These advantages are explained in more detail in the next section. In this paper we evaluate the effect of different super-resolution reconstruction techniques on the performance of point target detection. SR can be done in a robust or non-robust way. We expect that robust SR performs better for data sets with less frames. This is caused by the fact that the point target is more likely to be seen as an outlier, and will therefore be removed from the background using a robust method. SR reconstruction techniques often employ regularization. This will produce a smoother background image, and therefore reduces artefacts. However, it also removes frequency content and therefore fails to suppress the aliasing noise in the current frame by background subtraction.

This paper is organized as follows. In section 2 the theory of the point target detection using SR reconstruction algorithms is described. In section 3 experiments are described in which the effects of the different settings are shown. Finally, conclusions are presented.

2. THEORY

Super-Resolution (SR) reconstruction is a well-known technique to increase the resolution of a sequence of aliased Low-Resolution (LR) images. The Zoom Factor (ZF) of a SR reconstruction method is the ratio of the size of the resulting High-Resolution (HR) image with respect to the size of the LR images. Note that the resolution gain is usually less than ZF.[2] Generally, SR reconstruction can be split up in two parts:[2] 1) registration, 2) fusion and deblurring. The fusion and deblurring is described in more detail in the next subsection.

2.1 Registration

Registration aims to align all LR frames that are used for SR reconstruction. A variety of image registration techniques have been reported in the literature.[3] In our experiments the registration is done with a precise iterative gradient-based shift estimator.[4] This gradient-based shift estimator[5] finds the displacement (dx_{k1}, dy_{k1}) between two shifted images, $I_{k-1}(x, y)$ and $I_k(x, y)$, as a least-squares solution:

$$\text{it.1} : \min_{dx_{k1}, dy_{k1}} \frac{1}{P} \sum_{x,y} \left(I_k - I_{k-1} - dx_{k1} \frac{\partial I_{k-1}}{\partial x} - dy_{k1} \frac{\partial I_{k-1}}{\partial y} \right)^2 . \tag{1}$$

Here, image I_k is approximated with a Taylor expansion of image I_{k-1}, (x, y) are the pixel positions and P is the number of pixels in image I_k. The solution of equation 1, (dx_{k1}, dy_{k1}), is biased by a fraction of the displacement. Due to the relative aspect, the bias can be corrected in an iterative way:

$$\text{it.}n : \min_{dx_{kn}, dy_{kn}} \frac{1}{P} \sum_{x,y} \left(I_k(x + dx_{k(n-1)}, y + dy_{k(n-1)}) - I_{k-1} - dx_{kn} \frac{\partial I_{k-1}}{\partial x} - dy_{kn} \frac{\partial I_{k-1}}{\partial y} \right)^2 . \tag{2}$$

In iteration n $(n > 1)$, I_k is translated by interpolation with the estimated subpixel displacement $(dx_{k(n-1)}, dy_{k(n-1)})$ with respect to the original image I_k estimated in the previous iteration. This schema is iterated until convergence and results in a very precise ($\sigma_{disp} \approx 0.01$ pixel for noise free data) unbiased registration.[4]

Note that this registration method, due to its iterative character, can also cope with multiple-pixel image shifts. In such a case, the registration will not be accurate after the first iteration because the Taylor expansion is not accurate for large shifts. However, after a few iterations the remaining shift will be small and hence the Taylor expansion becomes accurate. This will only work if the result is the global minimum instead of a local minimum. Therefore, in the first steps often a coarse scale is applied, whereas in the last steps a finer scale is used.

2.2 Robust versus non-robust fusion and deblurring

The fusion and deblurring steps of SR reconstruction algorithms are used to fuse the aligned data and deblur the data on a HR grid. The resulting HR model of the background is free of aliasing. Super-resolution reconstruction can be done using fusion and deblurring methods that are robust or not robust with respect to outliers. If enough frames are available and the apparent motion of the point target w.r.t. the background is significant, a robust SR algorithm will see the point target as an outlier. A non-robust method will average the point target with the background and will therefore always place some target energy into the background.

A well performing non-robust method[6] is the method proposed by Hardie et al.[7] Zomet proposed a robust version of this algorithm.[8] In this paper we compare these two methods for point target detection.

Like many other SR reconstruction methods Hardie and Zomet use an observation model of the form:

$$I_k(x,y) = H_k Z + \theta_k(x,y) \tag{3}$$

where I_k is the k^{th} LR frame, Z is the HR image scene and θ_k denotes additive noise. The transfer matrix H_k describes 1) the model of the camera, 2) the estimated motion between Z and I_k and 3) the Zoom Factor (ZF). For simplicity, H_k represents the relation between Z and LR frame I_k in one matrix.

To obtain the solution, the total squared error L between the LR frames and the results of the resampling of the estimate of the HR scene $\tilde{Z}$ is minimized:

$$L(\tilde{Z}) = \frac{1}{2}\sum_{k=1}^{N}\left(I_k - H_k\tilde{Z}\right)^2 \tag{4}$$

with N the total number of LR frames. Taking the derivative of L with respect to $\tilde{Z}$ results in:

$$\nabla L(\tilde{Z}) = \sum_{k=1}^{N} H_k^T(H_k\tilde{Z} - I_k) = \sum_{k=1}^{N} G_k \tag{5}$$

with H_k^T the transposed of H_k. A new estimate of the background $\tilde{Z}$ is obtained by

$$\tilde{Z}_{n+1} = \tilde{Z}_n + \epsilon\nabla L(\tilde{Z}) \tag{6}$$

with ϵ the step size in the direction of the gradient. The robustness of Zomet's method is introduced by replacing the sum of back-projected images G_k in equation 5 by a scaled pixel-wise median :

$$\nabla L(\tilde{Z}) \approx N \cdot \mathrm{median}\,(G_k)_{k=1}^{N}\,. \tag{7}$$

2.3 Regularization

Adding a regularization term to the functional of equation 4 allows incorporation of a priori knowledge about the solution. To avoid artifacts in the background due to noise enhancement we generally add a term that favours a smooth solution as shown below.

$$L(\tilde{Z}) = \frac{1}{2}\sum_{k=1}^{N}\left(I_k - H_k\tilde{Z}\right)^2 + \lambda\sum_{i=1}^{H}||\nabla^2\tilde{Z}||^2 \tag{8}$$

with H the number of HR grid points, and

$$\nabla^2\tilde{Z} = \begin{pmatrix} 0 & -1/4 & 0 \\ -1/4 & 1 & -1/4 \\ 0 & -1/4 & 0 \end{pmatrix} * \tilde{Z} \tag{9}$$

2.4 Point target detection

After background subtraction, the point targets need to be detected in the difference image. The difference images show the amplitude difference between an image frame containing a moving target and background and an estimation of the background. A standard method to estimate the background is to use another frame as background estimate. With this so-called Align-and-Subtract (AS) method, the difference image is based on two aligned frames:

$$\Delta D_k^{AS}((x,y) = I_k(x,y) - I_{k-1,aligned}(x,y) \tag{10}$$

In $I_k(x,y)$ and $I_{k-1,aligned}(x,y)$ noise and aliasing artefacts are present, as well as the target.

For the super-resolution case, the difference image D_k^{SR} of frame k is calculated based on by:

$$\Delta D_k^{SR}(x,y) = I_k(x,y) - H_k\tilde{Z}, \tag{11}$$

in which $H_k\tilde{Z}$ is the background estimate based on the super-reconstruction result.

Using SR reconstruction for background suppression has the following advantages. The improvement depends on the apparent motion of the point target, the SR reconstruction algorithm, the amount of regularization and the number of frames used.

Preservation of the point target energy
For point targets with small apparent motion w.r.t. the background, subtracting subsequent frames will suppress the point target in ΔD_k^{AS}, which hampers detection. Ideally ΔD_k^{SR}, i.e. with SR reconstruction for background estimation, the point target is fully preserved. However, in practice, some of the point target's energy will be present in the background image, which causes a slight reduction of the point target in the difference image.

Reduction of aliasing artifacts
By applying SR reconstruction a better (aliasing free) estimate $\tilde{Z}$ of the background can be obtained. This HR estimate can be used to obtain an LR image of the background with the same aliasing artifacts as the original LR images. Therefore, the difference image will contain less aliasing artifacts. In other words, ΔD_k^{SR} is free from clutter due to aliasing, because $H_k\tilde{Z}$ now suffers from aliasing in exactly the same way as I_k,

Suppresion of temporal noise
As $H_k\tilde{Z}$ is based on a number of N recorded frames, the noise in this image will be lower than in the original camera image. This means that the noise level in ΔD_k^{SR} will also be lower than the noise level in ΔD_k^{AS} .

In the difference images the point targets will be detected. A simple detection technique is to threshold the magnitude of the difference image. All pixels with a value above a certain threshold value are detected as targets.

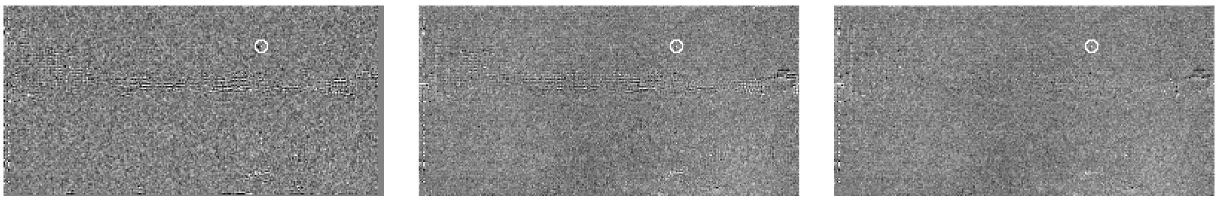
3. EXPERIMENTS

For the experiments camera images were simulated using a camera model. A super-scale input image is downscaled a factor 15 in both directions using a camera model, in which the subpixel motion, the lens blur (σ_{psf} = 0.35 LR pixel) and fill-factor (81% area) of the camera are modeled. For the SR reconstruction methods, a camera model with only lens blur ($\sigma_{psf} = 0.4$) is taken. The amplitude of an inserted point target is defined as the average maximum intensity of the point target in the LR images.

In Figure 1 point targets are shown which are added to real camera images. Adding the point target to the background instead of replacing the background introduces an error. In this simulation the error is small because of two reasons. First, the point target is placed in a superscale image and downscaled as described above, instead of placing it directly into the low resolution image. In this way, the point target will suffer from

(a) frame 1 (b) frame 24 (c) frame 48

Figure 1. Three frames (256×128) of a constructed point target sequence. The position of the point target can be seen in Figure 2. The amplitude of the point target is 56 grey values. The point target is moving with an apparent velocity of 2 LR pixel/frame w.r.t. the background.



(a) AS (b) Zomet 1 (c) Zomet 2

Figure 2. Difference images for the different background suppression methods displayed with intensity range [-6, 6]. The positions of the point target are indicated with a circle. The difference image is shown for the 24^{th} frame in the sequence. The amplitude of the point target is 12 grey values and its apparent motion w.r.t. the background is 2 LR pixel/frame.

aliasing in a similar way as the background. Second, the target is a point target and has therefore a small footprint. It can be shown[1] that the maximum error that is made is 100 times smaller than the temporal noise.

In Figure 1 it is hard to see the location of the point targets. In Figure 2 the different images for the AS and Zomet background suppression methods are shown. Without the background the point targets are detected more easily. It can be seen that the difference image resulting after background suppression with Zomet's SR reconstruction method with zoom factor 2 contains much less background contributions than the other methods. This effect is best seen in the center part of the image where the structure of the buildings is hardly visible in comparison with the other two difference images. Furthermore it can be seen that both difference images based on Zomet's method contain less noise than the difference image based on AS.

3.1 Point target intensity

Ideally the point target is not present in the projection of the HR background image, i.e. the point target intensity in the difference image is the same as the amplitude of the point target. In practice, the point target can be visible in the projection of the HR image. We evaluate the point target intensity in the difference image for the (almost) noise-free case. For these experiments, point targets with amplitude 1 were inserted in a constant background with very little noise ($\sigma_n = 0.002$). The registration was assumed to be perfect.

The maximum of the absolute difference image is used as measure for the point target intensity. The point target intensity as a function of the point target velocity for 16 and 80 frames are shown in Figure 3. As expected, the robust Zomet method performs better for SR reconstruction with a smaller number of frames and for point targets with a smaller velocity. The number of frames needed to obtain the original point target amplitude for slow point targets is higher than for fast point targets. There is no significant effect between the different Zoom Factors. For different regularization factors no effects were found. These results are not shown here.

3.2 Temporal noise

We assume that temporal noise in the different camera frames is independent. For the Hardie method without regularization the estimated LR frame $H_k\tilde{Z}$ is based on N recorded frames, which reduces the noise standard

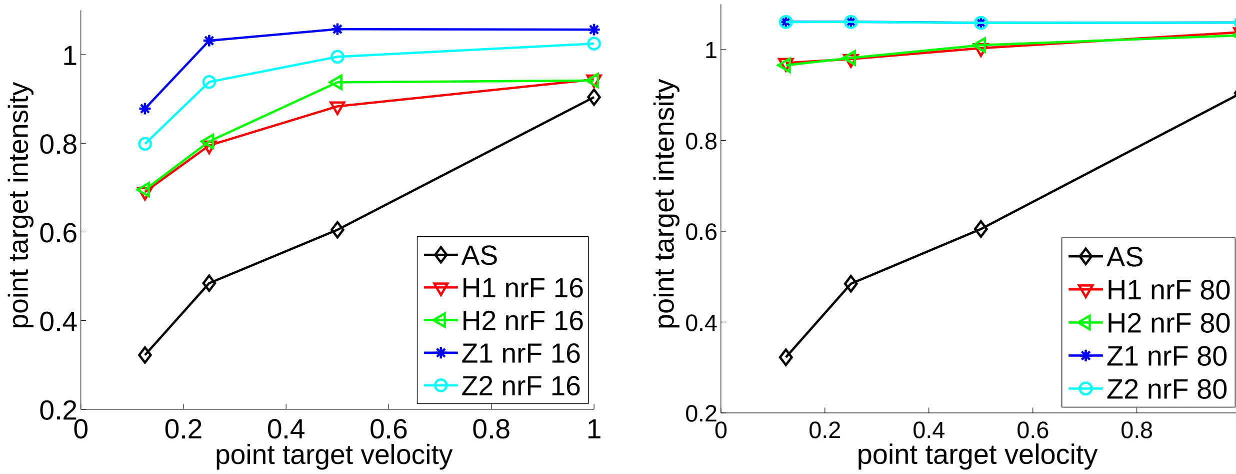


Figure 3. The maximum point target energy in the difference image for the Align-and- (AS) and SR reconstruction with the Hardie method (H) and the Zomet method (Z), for Zoom Factor 1 and 2, with 16 frames (left) and 80 frames (right).

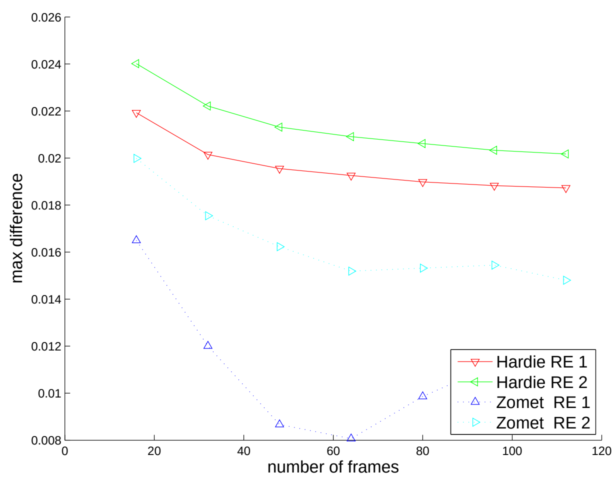


Figure 4. The maximum error in the difference images plotted against the number of frames used. Four methods are used: super-resolution reconstruction with the Hardie method and the Zomet method (lower row), for ZF 1 and 2, respectively. The input image contains only noise.

deviation with factor $\sqrt{N}/2F$, where F is the zoom factor. Therefore, the resulting noise in a difference image after SR reconstruction is: $\sigma_n^{\Delta hardie} = \sqrt{\frac{N+(2F)^2}{N}}\sigma_n$, which is for a large number of frames only slightly higher than the temporal noise of one frame σ_n. This is low compared to the Align-and-Subtract method, where the noise in the difference image is $\sqrt{2}\sigma_n$.

For the Zomet methods such a formula is not straightforward. Here, we also assume that the noise in the difference image is lower than for Align-and-Subtract.

These effects of the SR methods are tested with a simulation. The test images contain only temporal noise with $\sigma_n = 1$. The resulting noise for all SR reconstruction algorithms is around 1. The resulting noise of the difference image is plotted as function of the number of frames for the Hardie and Zomet method in figure 4. The noise is higher for higher Zoom Factors. The resulting noise is somewhat higher for Zomet than for Hardie, which shows that the mean is more effective in suppressing the noise than the median.

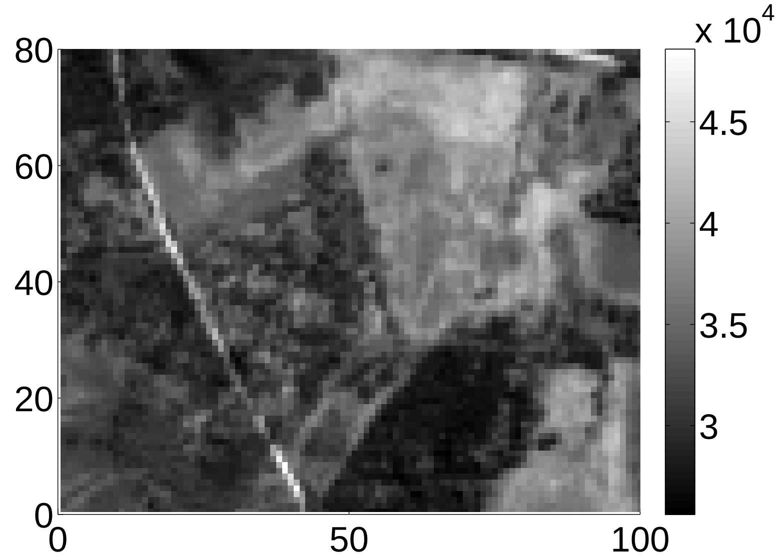


Figure 5. One of the input images for the aliasing experiments.

3.3 Aliasing noise

If the SR reconstruction algorithms are tuned correctly, the resolved aliasing contribution and noise should be similar for the robust and the non-robust case. A higher Zoom Factor should give a better aliasing prediction and therefore smaller values in the difference images. Regularization will increase the aliasing noise in the resulting difference image, as it favors smooth solutions.

To test these assumptions, experiments were done in which SR reconstruction was performed on sequences of an image containing a natural scene (see Figure 5). These images were constructed from an image from the Ikonos database, and contain no point target. The regularization parameters were varied in this simulation.

The maximum of the absolute value of the difference images is plotted in Figure 6 for the Align-and-Subtract method, the Hardie method and the Zomet method. The SR reconstruction methods are plotted with Zoom Factor 1 and 2, and with (λ=0.01) and without (λ=0) regularization. The largest improvement is found for Zoom Factor 2 instead of 1. Hardie and Zomet perform both better than Align-and-Subtract. Regularization will degrade the result. Note that this degradation is not seen for λ smaller than 0.001. Registration errors (not shown here) also increase the aliasing noise error.

4. CONCLUSIONS AND DISCUSSION

Super-resolution reconstruction techniques can be used to improve the detection of point targets in imagery. In this paper we show that for this purpose one can best use a robust SR algorithm instead of a non-robust algorithm. Also the effects of regularization are shown.

In experiments, we show that the remaining point target in the background image is smaller for the robust SR algorithm than for the non-robust algorithm, which improves the detection efficiency. This effect is highest for objects with a small velocity with respect to the scene and for SR reconstruction with a small number of frames. The robust and non-robust algorithm perform similar in suppressing temporal noise and aliasing artifacts in the difference image. The effect of regularization is highest for the aliasing noise. Here, a higher regularization factor will increase the aliasing noise and hence decrease the detection performance. The effect of regularization on point target amplitude and temporal noise is not significant. Therefore, for point target detection one can best use a robust SR algorithm with little or no regularization.

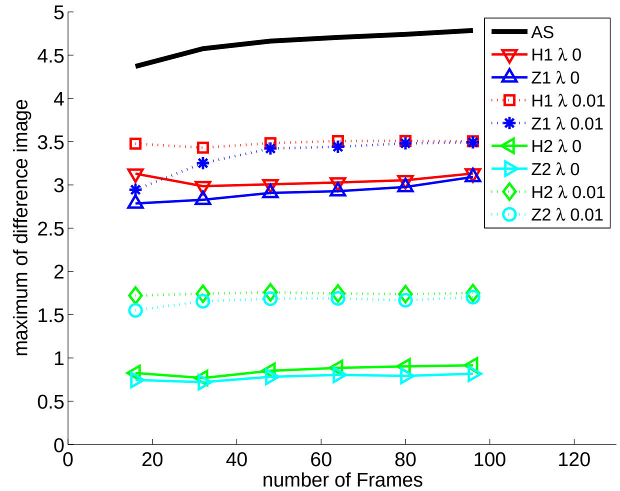


Figure 6. The maximum error in the difference images plotted against the number of frames used. The methods evaluated are Align + Subtract (AS), Hardie (H) and Zomet (Z), for Zoom Factor 1 and 2. The settings for regularization are: $\lambda = 0$ and $\lambda = 0.01$.

REFERENCES

[1] Dijk, J., van Eekeren, A. W. M., Schutte, K., de Lange, D.-J. J., and van Vliet, L. J., "Superresolution reconstruction for moving point target detection," *Optical Engineering* **49**(9) (2009). Article ID 096401-1.

[2] Pham, T., van Vliet, L., and Schutte, K., "Influence of signal-to-noise ratio and point spread function on limits of super-resolution," in [*Image Processing: Algorithms and Systems IV*], **5672**, 169–180, SPIE (2005).

[3] Zitová, B. and Flusser, J., "Image registration methods: a survey," *Image Vision Computing* **21**(11), 977–1000 (2003).

[4] Pham, T., Bezuijen, M., van Vliet, L., Schutte, K., and Hendriks, C. L., "Performance of optimal registration estimators," in [*Visual Information Processing XIV*], **5817**, 133–144, SPIE Defense and Security (2005).

[5] Lucas, B. and Kanade, T., "An iterative image registration technique with an application to stereo vision," in [*Imaging Understanding Workshop*], 121–130 (1981).

[6] van Eekeren, A. W. M., Schutte, K., Oudegeest, O. R., and van Vliet, L. J., "Performance evaluation of super-resolution reconstruction methods on real-world data," *EURASIP Journal on Advances in Signal Processing* , 1–11 (2007). Article ID 43953.

[7] Hardie, R., Barnard, K., Bognar, J., Armstrong, E., and Watson, E., "High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system," *Optical Engineering* **37**(1), 247–260 (1998).

[8] Zomet, A., Rav-Acha, A., and Peleg, S., "Robust super-resolution," *Proceedings of IEEE Computer Society Conference on Computer Vision and Patter Recognition (CVPR '01)* **1**, 645–650 (2001).