Research Article

Performance Evaluation of Super-Resolution Reconstruction Methods on Real-World Data

A. W. M. van Eekeren,[1] K. Schutte,[1] O. R. Oudegeest,[2] and L. J. van Vliet[2]

[1] *Electro-Optics Group, TNO Defence, Security and Safety, P.O. Box 96864, 2509 JG The Hague, The Netherlands*
[2] *Quantitative Imaging Group, Department of Imaging Science and Technology, Faculty of Applied Sciences,*
 Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands

The performance of a super-resolution (SR) reconstruction method on real-world data is not easy to measure, especially as a ground-truth (GT) is often not available. In this paper, a quantitative performance measure is used, based on triangle orientation discrimination (TOD). The TOD measure, simulating a real-observer task, is capable of determining the performance of a specific SR reconstruction method under varying conditions of the input data. It is shown that the performance of an SR reconstruction method on real-world data can be predicted accurately by measuring its performance on simulated data. This prediction of the performance on real-world data enables the optimization of the complete chain of a vision system; from camera setup and SR reconstruction up to image detection/recognition/identification. Furthermore, different SR reconstruction methods are compared to show that the TOD method is a useful tool to select a specific SR reconstruction method according to the imaging conditions (camera's fill-factor, optical point-spread-function (PSF), signal-to-noise ratio (SNR)).

1. INTRODUCTION

During the last decade, numerous super-resolution (SR) reconstruction methods have been reported in the literature. Reviews can be found in [1, 2]. SR reconstruction is the process of combining a set of undersampled (aliased) low-resolution (LR) images to construct a high-resolution (HR) image or image sequence. A typical solution for SR reconstruction of an image sequence involves two subtasks: registration and fusion. Occasionally, an additional deblurring step is performed afterwards. First, the LR images are registered against a common reference with subpixel accuracy. During the fusion, an image at a higher resolution is constructed from the scattered input samples. Nonlinear deblurring is needed to extend the frequency spectrum beyond the cut-off limit of the imaging sensor.

Although SR reconstruction has received significant attention over the past few years, not much work has been done in the field of performance (limits) of SR. Relevant works are reported in [3, 4]. Both study the problem of SR from an algebraic point of view. Robinson and Milanfar [5] recently analyzed the performance limits from statistical first principles using Cramér-Rao inequalities. This analysis has the advantage that the performance bottlenecks can be related to the subtask level of an SR reconstruction method.

This paper discusses the performance of an SR reconstruction method under different conditions such as number of input frames and signal-to-noise ratio (SNR), for a specific vision task, using the characteristics of modern infrared (IR) imagers. This vision task is the discrimination of small objects/details in an image and is measured quantitatively using triangle orientation discrimination (TOD) [6, 7]. TOD is a task-based evaluation method, which measures the ability to discriminate the orientation of an equilateral triangle under a specific condition.

The performance of an SR reconstruction method on real-world data is especially interesting to measure, as it shows the capability of the algorithm in practice. In this paper, it is shown that with the TOD method a quantitative performance measure of an algorithm on real-world data can be obtained. Moreover, it is shown that the results of this measure can be predicted accurately by measuring the TOD performance on simulated data. This enables the optimization and selection of the algorithm in advance given a real-world camera.

The paper is organized as follows. In Section 2, the registration of the real-world and simulated data is discussed. In Section 3, the different SR reconstruction methods are discussed. In Section 4, the TOD method is explained and the setup of the measurements is given. The results are presented in Section 5 and finally conclusions will be provided in Section 6.

2. REGISTRATION

The scenes (real-world and simulated) in our experiments are static and captured with a moving camera. Therefore, the scene movement between two frames can be described with a single shift. All LR frames of an image sequence are registered to a reference frame, which is typically the first frame of the image sequence. The registration of the LR frames is performed with an iterative gradient-based shift estimator [8]. A gradient-based shift estimator [9] finds the displacement $t_{\vec{x}}$ between two shifted signals as the least squares solution of

$$\text{MSE} = \frac{1}{N} \sum_R \left(s_2(\vec{x}) - s_1(\vec{x}) - t_{\vec{x}} \frac{\partial s_1}{\partial \vec{x}} \right)^2 \qquad (1)$$

with s_2 a shifted version of s_1, $\vec{x}$ the sample positions, and N the number of samples in supported region R.

The solution of (1) is biased, which is corrected in an iterative way. In the first iteration, s_2 is shifted with the estimated subpixel displacement, which is accumulated in the next iteration with the estimated displacement between s_2' (shifted s_2) and s_1. This schema is iterated until convergence, finally resulting in a very precise ($\sigma_{\text{disp}} \approx 0.01$ pixel for noise free data) unbiased registration, which approaches the Cramér-Rao bound [10].

In our experiments, the set of registered LR frames is processed by each of the SR fusion/deblurring methods described in the following section. It is important to note that all methods use the same set of registered LR frames. This implies that differences in overall performance are not due to differences in registration.

3. SUPER-RESOLUTION FUSION/DEBLURRING METHODS

This section briefly describes the different SR reconstruction methods used in the performance evaluation. The first three methods perform only fusion, whereas the last three methods also incorporate deblurring.

3.1. Elad's shift and add method

After registration of all LR frames, Elad's [11] reconstruction method assigns each LR sample to the nearest HR grid point. When this is done for all LR samples, the mean is taken of all LR samples on each HR grid point. Note that the shift and add method is only a fusion method and does not incorporate deblurring.

3.2. Lertrattanapanich's triangulation-based method

In [12], Lertrattanapanich proposes a triangle-based surface interpolation method for irregular sampling. First, a Delaunay triangulation of all registered LR samples is performed, followed by an approximation of each triangle surface with a bicubic polynomial function. The pixel value $z(x, y)$ at a new HR grid location (x, y) is expressed as in (2):

$$z(x, y) = c_1 + c_2 x + c_3 y + c_4 x^2 + c_5 y^2 + c_6 x^3 + c_7 x^2 y + c_8 x y^2 + c_9 y^3. \qquad (2)$$

Note that the monomial xy is omitted to maintain the geometric isotropy. The nine parameters c_i can be solved with three vertices (LR samples) and their corresponding estimated gradients along x and y directions. Lertrattanapanich's triangulation-based method performs fusion only.

3.3. Kaltenbacher's least-squares method without regularization

This method [13] is based on the idea of estimating the "underlying" unaliased frequency spectrum from multiple, aliased spectra. For sake of clarity, the 1D case will be explained below. With the shift property, the Fourier transform F_i of a shifted frame i before sampling is

$$F_i(\omega) = F(\omega) e^{j \delta_i \omega}, \qquad (3)$$

where δ_i is the shift of frame i and $F(\omega)$ is the Fourier transform of the original image. After sampling by the camera the transform in (3) converts to

$$\widetilde{F}_i(n) = \frac{1}{S} \sum_{m=-\infty}^{\infty} F_i \left(\frac{2\pi}{NS} n - m \omega_s \right). \qquad (4)$$

Here, $\widetilde{F}_i(n)$ is the discrete Fourier transform of LR input frame $i = 1, \ldots, P$. S is the sampling period and $\omega_s = 2\pi/S$ is the sampling frequency, N is the amount of samples per LR frame, and $n = 1, \ldots, N$ is the sample index (here $S = 1$ and $\omega_s = 2\pi$).

If the sampling frequency is increased by a factor K (zoom factor) such that $K \omega_s > 2 \omega_c$ (cutoff frequency), the limits in the summation of (4) can be changed to $\lfloor -K/2 \rfloor + 1$ and $\lfloor K/2 \rfloor$. When all shifts δ_i are known and K is chosen, for each sample n a set of equations can be written:

$$\mathbf{G}_n = \mathbf{\Phi}_n \mathbf{F}_n, \qquad (5)$$

where $\mathbf{G}_n$ is a column vector with the nth Fourier component of each LR frame,

$$\mathbf{G}_n(i) = \widetilde{F}_i(n), \qquad (6)$$

and $\mathbf{\Phi}_n$ is the $(P \times K)$ transformation matrix defined by

$$\mathbf{\Phi}_n(i, k) = e^{j 2\pi \delta_i (n/N + (\lfloor K/2 \rfloor - k))}. \qquad (7)$$

$\mathbf{F}_n$ is the column vector with the K-target Fourier components dependent on n. This method needs at least $2K$ LR input frames. When more than $2K$ frames are used, a least-squares solution of the target Fourier components is obtained by the Moore-Penrose inverse of $\mathbf{\Phi}_n$:

$$\mathbf{F}_n = (\mathbf{\Phi}_n^T \mathbf{\Phi}_n)^{-1} \mathbf{\Phi}_n^T \mathbf{G}_n. \qquad (8)$$

3.4. Hardie's method using a regularized inverse observation model

Hardie et al. [14] employ a discrete observation model that relates the ideally sampled image $\mathbf{z}$ and the observed frames $\mathbf{y}$:

$$y_m = \sum_{r=1}^{H} w_{m,r} z_r + \eta_m, \qquad (9)$$

where $w_{m,r}$ represents the contribution of the rth HR pixel in $\mathbf{z}$ to the mth LR pixel in $\mathbf{y}$. This contribution depends on the frame-to-frame motion and on the blurring of the point spread function (PSF). η_m denotes additive noise.

The HR image estimate $\hat{\mathbf{z}}$ is defined as the $\mathbf{z}$ that minimizes

$$C_{\mathbf{z}} = \sum_{m=1}^{L} \left(y_m - \sum_{r=1}^{H} w_{m,r} z_r \right)^2 + \lambda \sum_{i=1}^{H} \left(\sum_{j=1}^{H} \alpha_{i,j} z_j \right)^2 \qquad (10)$$

with L the number of LR samples and H the number of HR grid points.

The cost function in (10) balances two types of errors. The left term is minimized when a candidate $\mathbf{z}$, projected through the observation model (9), matches the observed data. The right term is a regularization term, which is necessary as directly minimizing the first term is an ill posed problem. The parameters $\alpha_{i,j}$ (11) are selected to perform a Laplacian operation on $\mathbf{z}$ and ensure that the regularization term is minimized when $\mathbf{z}$ is smooth:

$$\alpha_{i,j} = \begin{cases} 1 & \text{for } i = j \\ \dfrac{-1}{4} & \text{for } j : z_j \text{ is a cardinal neighbor of } z_i. \end{cases} \qquad (11)$$

3.5. Farsiu's robust method

In comparison with Hardie's method, the reconstruction method proposed by Farsiu et al. [15] separates the fusion and deblurring processes of an SR reconstruction method: (1) the LR frames are fused with median shift and add (similar as described in Section 3.1, but now the median, rather than the mean, is taken of the samples at each HR grid point), (2) the fusion result $\mathbf{z_0}$ is deblurred using an iterative minimization method. The cost function that must be minimized to obtain the SR image $\hat{\mathbf{z}}$ from fusion result $\mathbf{z_0}$ is shown in (12):

$$C_{\mathbf{z}} = ||A(G\mathbf{z} - \mathbf{z_0})||_1 + \lambda \sum_{l=0}^{P} \sum_{m=0}^{P} \alpha^{m+l} ||\mathbf{z} - S_h^l S_v^m \mathbf{z}||_1. \qquad (12)$$

Here, matrix A is a diagonal matrix with diagonal values equal to the square root of the number of measurements that contributed to make each element of $\mathbf{z_0}$. Therefore, undefined pixels in $\mathbf{z_0}$ will have no influence on the SR estimate $\hat{\mathbf{z}}$. Matrix G is a blur matrix that models the PSF of the camera system. The regularization term on the right-hand side is based on the bilateral total variation (TV) criterion [15]. Matrices S_h^l and S_v^m shift $\mathbf{z}$ by l and m pixels in horizontal and vertical directions, respectively. The scalar weight α, $0 < \alpha < 1$, is applied to give a spatial decaying effect.

3.6. Pham's structure-adaptive and robust method

Pham et al. [16] recently proposed an SR reconstruction method using adaptive normalized convolution (NC). NC [17] is a technique for local signal modeling from projections onto a set of basis functions. Pham uses a first-order polynomial basis as shown:

$$\hat{f}(\mathbf{s}, \mathbf{s}_0) = p_0(\mathbf{s}_0) + p_1(\mathbf{s}_0)x + p_2(\mathbf{s}_0)y, \qquad (13)$$

where $\hat{f}$ is the approximated intensity value at sample $\mathbf{s}$, (x, y) are the local coordinates of $\mathbf{s}$ with respect to the center of analysis, $\mathbf{s}_0$ and p_i are the projection coefficients. In contrast with a polynomial expansion like the Haralick facet model [18], NC uses (1) an applicability function to localize the polynomial fit and (2) allows each input sample to have its own certainty value. To determine the projection coefficients at an output position $\mathbf{s}_0$, the approximation error is minimized over the extent of an applicability function a centered at $\mathbf{s}_0$:

$$\varepsilon(\mathbf{s}_0) = \int (f(\mathbf{s}) - \hat{f}(\mathbf{s}, \mathbf{s}_0))^2 c(\mathbf{s}) a(\mathbf{s} - \mathbf{s}_0) d\mathbf{s}, \qquad (14)$$

with a the applicability function and c the certainty of each sample within the extent. A schematic overview of Pham's method is depicted in Figure 1.

After registration of the LR samples, the first step of the fusion process consists of estimating an initial polynomial expansion (using a flat model at a locally weighted median level), which results in I_{HR_0}. Next, NC using a robust certainty (15) is performed, which results in a better estimate I_{HR_1} and two corresponding derivatives I_{HR_x} and I_{HR_y},

$$c(\mathbf{s}, \mathbf{s}_0) = \exp\left(- \frac{|f(\mathbf{s}) - \hat{f}(\mathbf{s}, \mathbf{s}_0)|^2}{2\sigma_r^2} \right). \qquad (15)$$

Here, the photometric spread σ_r defines an acceptable range of the residual error $|f - \hat{f}|$. The derivatives are used in the last fusion step to construct anisotropic applicability functions for adaptive NC. Such an applicability function is an anisotropic Gaussian function whose main axis is rotated to align with the local dominant orientation. Deblurring is done with bilateral TV regularization (as in Farsiu's method).

4. PERFORMANCE EVALUATION EXPERIMENTS

To measure the performance of SR reconstruction, several quantitative measures such as mean squared error (MSE) and modulation transfer function (MTF) are often used. However, we use the triangle orientation discrimination (TOD) measure as proposed in [6]. The TOD method determines the smallest triangle size in an image of which the orientation can be discriminated. This evaluation method is preferred
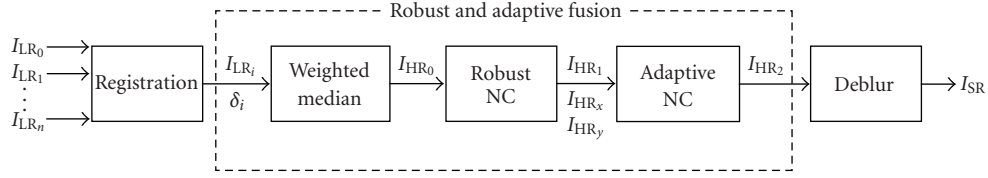
FIGURE 1: Flow diagram of Pham's structure-adaptive and robust SR reconstruction method.
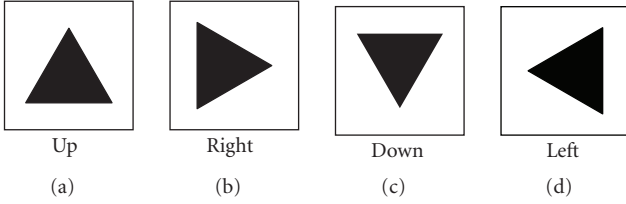

FIGURE 2: The four different stimuli used in the TOD method.

over methods like MSE and MTF because (1) the measurement is done in the spatial domain and is well localized, and (2) it employs a specific vision task. This vision task is directly related to the acquisition of real targets, which was first shown by Johnson [19]. Such a relationship is relevant for determining the limitations of your camera system including the image processing for recognition purposes. The MSE and MTF are neither localized nor task related. The MTF method is also not suited for evaluating nonlinear algorithms, which most SR reconstruction methods are.

4.1. TOD method

The TOD method is an evaluation method designed for system performance of a broad range of imaging systems. It is based on the observer task to discriminate four different oriented equilateral triangles (see Figure 2).

The observer task is a four-alternative forced choice, in which the observer has to indicate which of the four orientations is perceived, even when he is not sure. In the experiments, an automatic observer is used which makes its choice $\hat{\theta}$ based on the minimum MSE between the triangle in the SR result I_{HR} and a triangle model M:

$$\hat{\theta} = \min_{\theta, s} \left\{ \frac{1}{N} \sum_{\vec{x}} \left(I_{HR}(\vec{x}; \theta_f, s_f) - M(\vec{x}; \theta, s) \right)^2 \right\}. \quad (16)$$

Here, θ indicates the orientation, s indicates the size of the triangle, $\vec{x}$ are the sample positions, and N is the number of samples. Note that θ is limited to the four different orientations and s is quantized in steps of 4/17th of the LR pixel pitch. The subscript f denotes one member of these sets. Although (16) is minimized for θ and s, only the estimated orientation $\hat{\theta}$ is used as a result. Note that triangle model M can also incorporate a gain and offset parameter.

The probability of a correct observer response increases with the triangle size. In [6] it is shown that this increase can be described with a Weibull distribution:

$$p_c(x) = 0.25 + \frac{0.75}{1.5^{(\alpha/x)^{\beta}}}, \quad (17)$$

where α is x at 0.75 probability correct and β defines the steepness of the transition. Such a Weibull distribution can be fitted to a number of observations for different triangle sizes as depicted in Figure 3. From this fit the triangle size that corresponds with an 0.75 probability correct response (T_{75}) is determined. T_{75} (in LR pixels) is a performance measure, where a smaller T_{75} indicates a better performance. When for different conditions, for example, SNR, T_{75}s are determined, a performance curve can be plotted. Such curves will be used in Section 5 to show the results.

4.2. Real-world data experiment

In this experiment the performance of an SR reconstruction method on real-world data is measured.

4.2.1. Setup

The setup of the experiment (including TOD) is depicted in Figure 4. The LR data I_{LR} comes from a real-world thermal IR camera (FLIR SC2000) with a rotating mirror in front of the lens. In the scene a thermal camera acuity tester (T-CAT [20]) is present as depicted in the left-hand side of Figure 4. This apparatus contains an aluminium plate with 5 rows of 4 equilateral triangle shaped cutouts. A black body plate is placed 3 cm behind this plate. Between the plates several temperature differences can be created. By controlling the temperature difference, different contrast levels (SNRs) are obtained. Although the triangle shaped cutouts on the plate vary in size, more size variation can be obtained by changing the distance from the apparatus to the camera. Real-world data sequences (40 frames) are processed with three different SR reconstruction methods with optimized parameter settings: Elad's method, Hardie's method, and Pham's method.

From both the I_{LR} data and the reconstructed I_{HR} data the orientation of the triangles is determined. This is done using (16) with gain and offset estimation in triangle model M. The triangle model M is implemented with shifted, blurred, and downsampled triangles in the triangle database. The triangle database contains equilateral triangles with sides $12, 16, \ldots, 280$ pixels. In our evaluation each triangle is equidistantly shifted, blurred ($\sigma = 0.9 \times S$), and
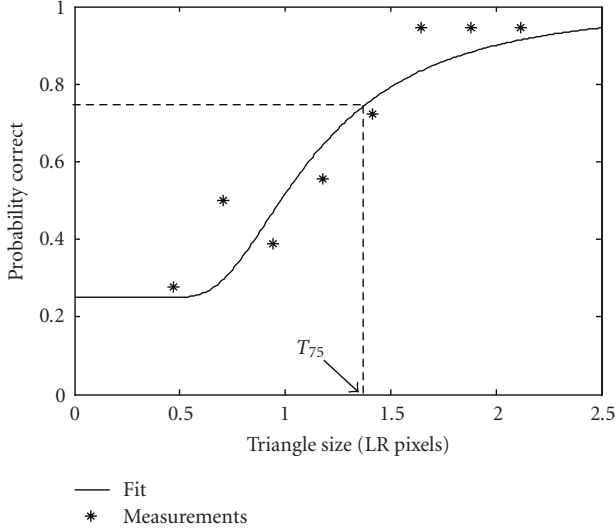
FIGURE 3: Example of a possible Weibull distribution of probability correct observer response.
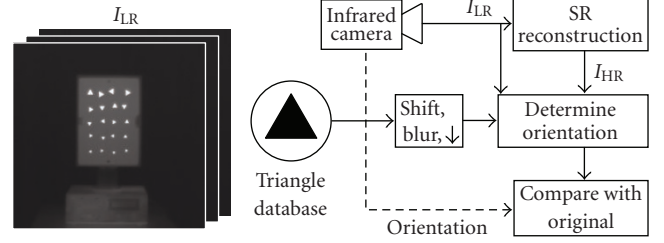

FIGURE 4: Left: example of real-world data I_{LR}. Right: flow diagram of the real-world data experiment.

FIGURE 5: Camera model used in the experiments.

downsampled ($S = 17$) resulting in 25 realizations for each triangle. Here the blurring with $\sigma = 0.9 \times S$ is chosen such that these reference triangles will have a right balance between residual aliasing and high-frequency content [21]. The orientation of the triangle obtained from the triangle database that results in the smallest mean-square error with the triangle in the data is selected. In the final step of the experiment setup the obtained orientation in the previous step is compared with the known ground-truth (GT) orientation of the triangle in the original real-world data.

4.2.2. Measurements on real-world data

To validate the performance on real-world data of the SR reconstruction methods with simulations, some measurements are needed of the real-world data: (1) SNR, (2) point-spread-function (PSF) of the lens, and (3) fill factor (ff), which is the percentage of photo-sensitive area of the pixels on the focal plane array sensor.

The real-world data was recorded with three different temperature differences of the T-CAT, which results in three SNRs. Here, the SNR dB is defined as

$$\text{SNR} = 20 \log_{10} \left(\frac{|I_{TR} - I_{BG}|}{\sigma_{BG}} \right), \tag{18}$$

with I_{TR} is the triangle intensity, I_{BG} the background intensity on the T-CAT plate, and σ_{BG} the standard deviation of I_{BG}. Our measurements resulted in SNRs 7 dB, 30 dB, and 48 dB.

The parameters of the camera (PSF and ff) are obtained by estimating the overall blur (LR pixels), σ_{tot}, in the real-world data by fitting an erf model to several edges in the data (with highest SNR). Measurements on edges of large triangles resulted in an overall blur of $\sigma_{tot} \approx 0.7$, whereas on medium-sized triangles an overall blur of $\sigma_{tot} \approx 0.5$ was measured. When comparing these measurements with the specifications of the camera (FLIR SC2000), the smallest overall
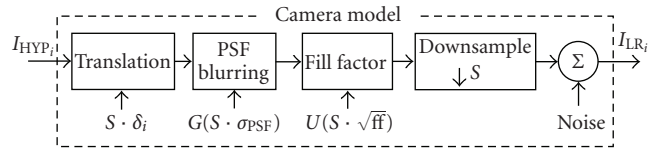
blur seems more likely. Given the camera model as depicted in Figure 5, the PSF blur can be determined from the overall blur for a certain fill factor. In modern infrared cameras a realistic fill factor is approximately 80% [22, page 101]. Given a $\sigma_{tot} = 0.5$ the blurring of the lens is $\sigma_{PSF} = 0.4$.

4.3. Simulated data experiment 1

Based on the estimates of the camera's parameters, simulated data sets have been generated. After processing the simulated data sets with the same SR reconstruction methods as in the previous experiment an indication can be obtained of the predictability of the real-world performance of these algorithms.

4.3.1. Camera model

A data set is simulated with a camera model as depicted in Figure 5, where I_{HYP_i} is a discrete representation of a scene sampled at the Nyquist rate with an $S\times$ smaller sampling distance than the observed frames I_{LR_i}. δ_i represents the translation of the camera, the PSF of the lens is modeled with a 2D Gaussian function G with standard deviation $S \cdot \sigma_{PSF}$ and the fill factor are modeled with a uniform filter U with width $S \cdot \sqrt{ff}$. The overall noise in the camera model is assumed to be Gaussian distributed.

In this experiment two simulated data sets I_{LR} are generated: (1) $\sigma_{PSF} = 0.3$, ff $= 0.8$, which results in a less-blurred data set as derived in Section 4.2.2 and (2) $\sigma_{PSF} = 0.55$, ff $= 0.8$, which results in a more-blurred data set. The downsampling factor is chosen as $S = 17$. The shift vectors $S \cdot \delta_i$ are random integer shifts ([0,S] pixels in the hyper-resolution (HY) domain) such that this results in subpixel shifts in the simulated data. Different amounts of Gaussian noise are added, resulting in a SNR varying from 12 dB to 42 dB.
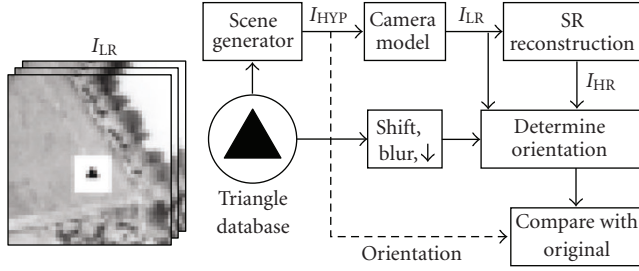
FIGURE 6: Left: example of simulated data I_{LR}. Right: flow diagram of the simulated data experiment.

4.3.2. Setup

The setup of the experiment on simulated data is depicted in Figure 6. The scene generator produces HY scenes I_{HYP} containing different triangle sizes and orientations from the triangle database. The camera model converts the I_{HYP} data to I_{LR} data in such a way that for each triangle size 16 realizations are present in the data set. Note that the number of realizations determines the statistical validity of the experiment. The I_{LR} data, of which an example is shown in the left-hand side of Figure 6, is the input for the SR reconstruction methods. Note that the settings of these methods are the same as for processing the real-world data. From both the I_{LR} data and the reconstructed I_{HR} data the triangle orientation is determined using (16). Note that for this experiment no gain and offset estimation is used in the triangle model M.

4.4. Simulated data experiment 2

This experiment is done to show that the TOD method is a useful tool to select a specific SR reconstruction method according to the imaging conditions (camera's fill factor, optical PSF, SNR). Here, camera model parameters ($\sigma_{\text{PSF}} = 0.2$, ff $= 1$) that result in a more-aliased data set than the previous simulated data sets are chosen. These parameters are chosen to enhance the differences between the SR reconstruction methods. To measure the performance of each method, the same setup is used as in "simulated data experiment 1" (see Figure 6). The performance of the SR reconstruction methods is measured for the following conditions

(1) Different number of frames.
(2) Different SNRs.
(3) Different zoom factors.

Note that the first two conditions are determined by the simulated data and the last one (ratio between resulting HR grid and original LR grid) is determined by the algorithm. Only Hardie's, Farsiu's, and Pham's methods are tuned to perform optimally under the varying conditions. For all three methods the parameter λ is tuned. The tuning criterium is to obtain a smallest T_{75} triangle size under the condition at hand. Note that the parameter λ in Hardie's method has a slightly different meaning than in the other two methods. The parameter σ, which is the standard deviation of a Gaussian

function and represents both the PSF due to the optics and the sensor blur due to the fill factor, is chosen in such a way that it fitted best to the blurring of our used camera model.

The results of all experiments are discussed in the following section.

4.5. TOD versus MSE

An alternative measure to TOD is the MSE:

$$\text{MSE} = \frac{1}{N} \sum_{\vec{x}} \left(I_{\text{HR}}(\vec{x}; \theta_f, s_f) - M(\vec{x}; \theta_f, s_f) \right)^2. \tag{19}$$

To show the difference between both measures, the following experiment is performed. Simulated LR data (varying SNR) is processed with the Hardie SR reconstruction method with different settings (varying λ and number of frames).

The resulting images are first scored with the TOD method and subsequently the MSE is calculated between the SR results and a triangle model M of size s_f closest to the triangle threshold (T_{75}) found. Contour plots of both measures are depicted in Figure 7.

It is clear from Figure 7 that the profiles of the TOD measure differ from the corresponding MSE profiles. Analyzing the profiles for a fixed frame number shows that the "optimal" λ resulting in the lowest T_{75} is significantly smaller than the "optimal" λ resulting in the lowest MSE: 10^{-2} and 1, respectively. The corresponding SR results (not depicted in this paper) show that a small λ result in steep edges with some ringing at the boundary of the triangles. Note that TOD and thereby correct identification does not solely depend on the lowest MSE found, but rather on the separability (= expected difference in MSE between the observation and the correct assignment and the MSE between the observation and an incorrect assignment divided by the variance of the MSE). Hence, the ringing imposes a positive influence on this measure of separability.

5. RESULTS

All results of the experiments can be found at the end of this paper. Note that the vertical axis in the plots indicate the triangle threshold size at 75% probability correct. A smaller triangle threshold size (T_{75}) corresponds with a better performance, hence the lower the curve, the better the performance.

5.1. Results of real-world data and simulated data experiment 1

The results of the "real-world data experiment" and the "simulated data experiment 1" can be seen in Figure 8. These graphs show that the performance on real-world data can be approximated by the performance of a simulated data set. The depicted performance of the two simulated data sets form a *performance* lower bound ($\sigma_{\text{PSF}} = 0.55$ and ff $= 0.8$, resulting in an "overall" $\sigma_{\text{tot}} \approx 0.6$) and a *performance* upper bound ($\sigma_{\text{PSF}} = 0.3$ and ff $= 0.8$, resulting in $\sigma_{\text{tot}} \approx 0.4$) on the real-world performance. Note that in Figure 8 the *performance* upper bound is visually a lower bound and the
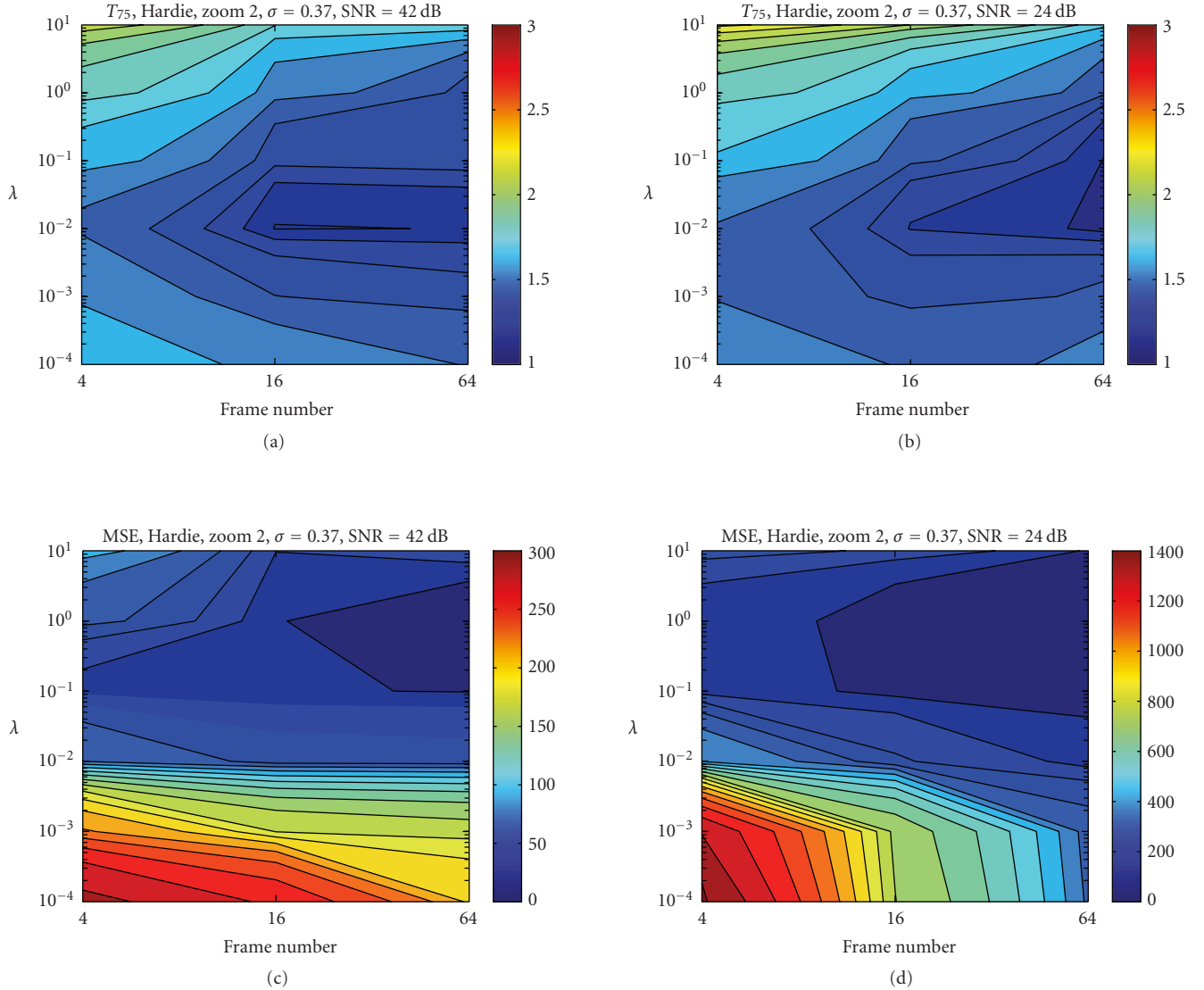
FIGURE 7: (a) Contour plot T_{75}, SNR = 42 dB, (b) contour plot T_{75}, SNR = 24 dB, (c) contour plot MSE, SNR=42 dB, (d) contour plot MSE, SNR = 24 dB.

performance lower bound is visually an upper bound. Elad's method shows that for all SNRs the performance on the real-world data is close to the *performance* upper bound. For Hardie's method we see the opposite for high SNRs: here the real-world performance is equal to the *performance* lower bound. Furthermore, it can be seen that the performance on real-world data of the three algorithms is similar for low and medium SNR, whereas for high SNR Pham's and Hardie's methods perform slightly better.

5.2. Results of simulated data experiment 2

In **Figure 9** the performance of all SR reconstruction methods with zoom factor 2 for different number of LR input frames is compared. Here the black line indicates the per-

formance on "raw" unprocessed LR input data and therefore should be taken as baseline reference. From these plots it is clear that the performance of all SR reconstruction methods improves when processing more frames. For high SNRs this improvement is only marginal, but for low SNRs it is significant. Kaltenbacher's method performs poorly when processing only 4 LR frames. This can be explained by the fact that the shifted LR frames are nonevenly spread, which results in an unstable solution. When 64 LR frames are processed, Lertrattanapanich's method performs worst for low SNRs. For high SNRs the performance of Elad's method performs worst. The best performing SR reconstruction methods (when many LR frames are available) are Kaltenbacher's method and Hardie's method, closely followed by the method of Pham.
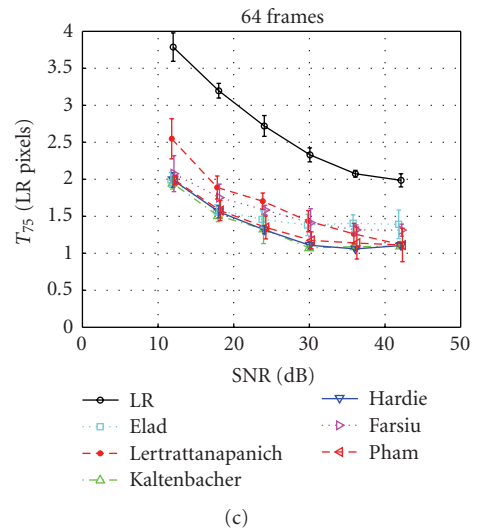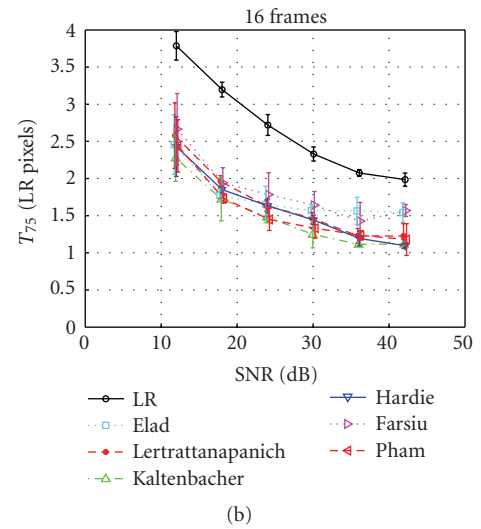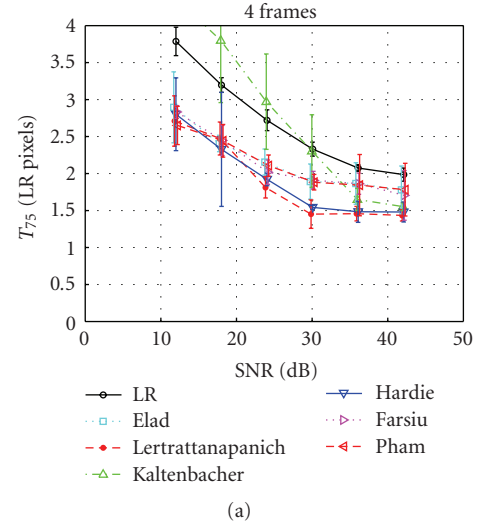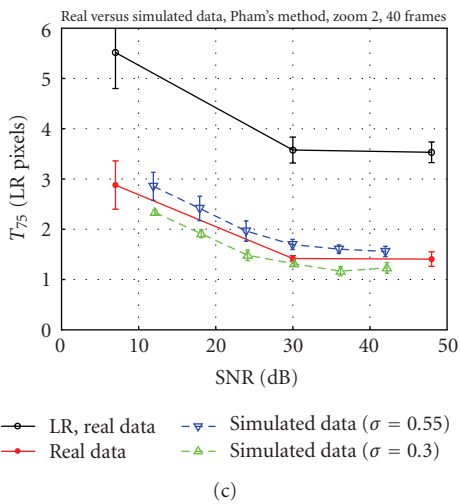
(a)

(b)

(c)
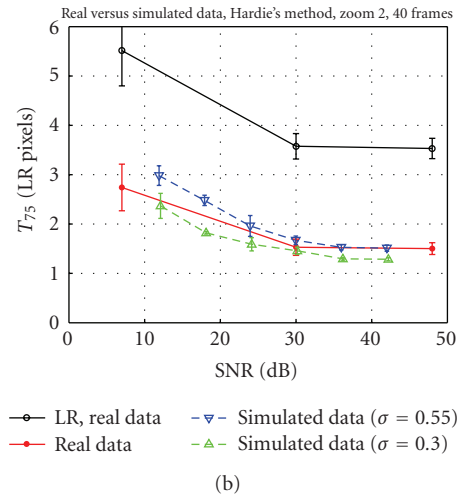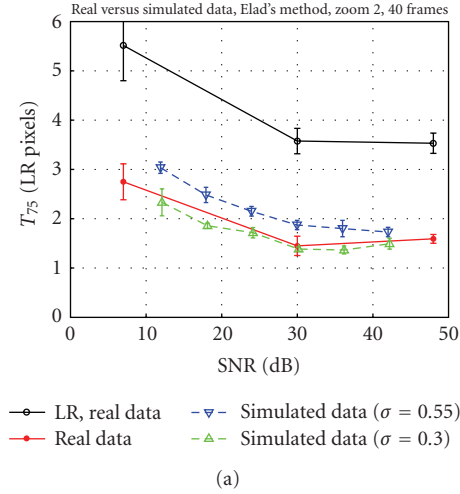
FIGURE 8: Performance measurements on real-world and simulated data (40 frames). Blue line: simulated data created with $\sigma_{PSF} = 0.55$ and ff = 80%, green line: simulated data created with $\sigma_{PSF} = 0.3$ and ff = 80%. (a) Elad, (b) Hardie ($\sigma = 0.55, \lambda = 0.01$), (c) Pham ($\sigma = 1, \lambda = 10^{-3}, \beta = 10$). All data is processed with zoom factor 2.

(a)

(b)

(c)

FIGURE 9: Performance measurements on simulated LR data ($\sigma_{PSF} = 0.2$, ff = 100%) processed with different SR reconstruction methods (zoom factor 2) with optimized settings, (a) 4 frames, (b) 16 frames, (c) 64 frames.
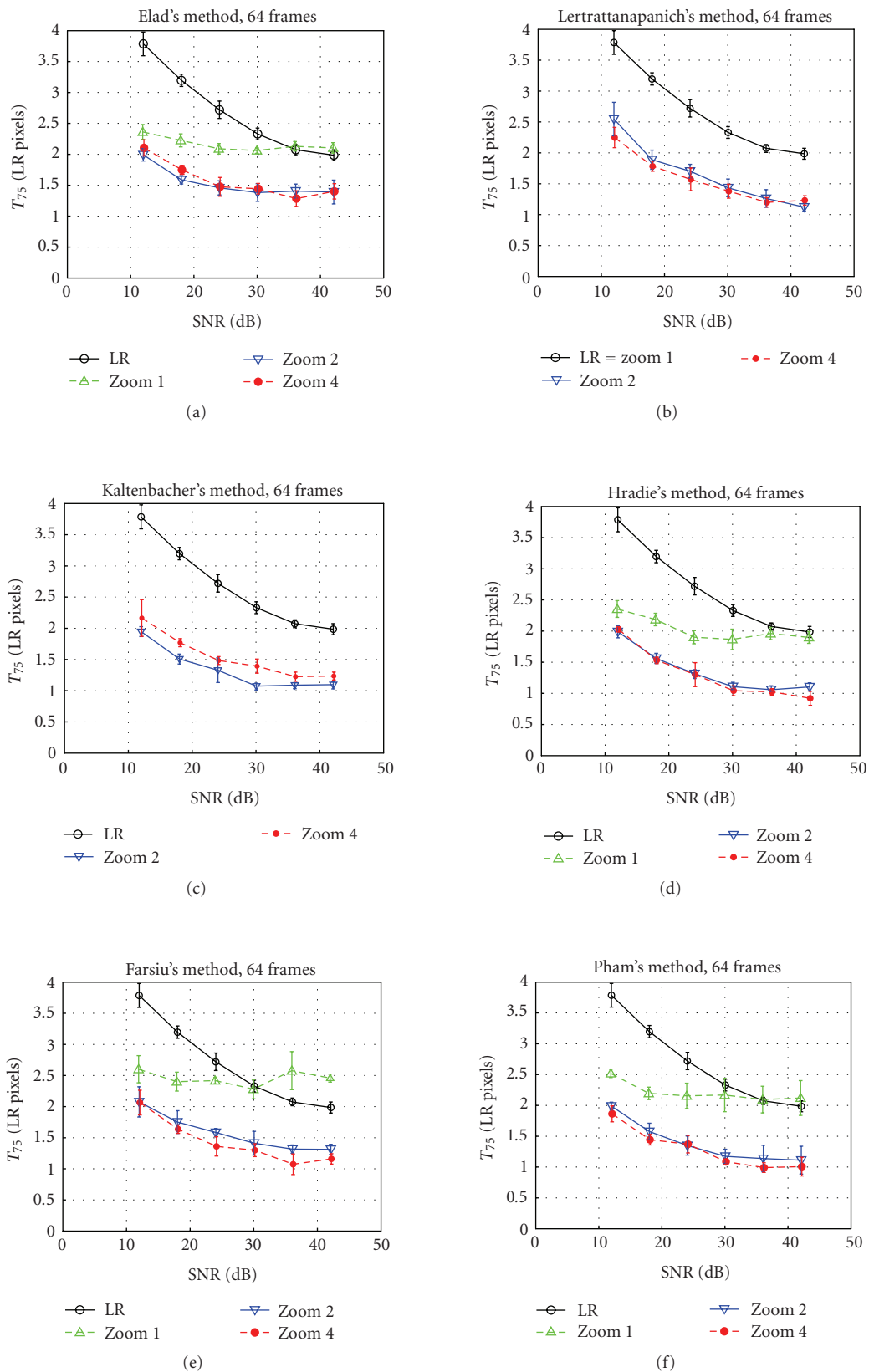
FIGURE 10: Performance measurements on simulated LR data (σ_{PSF} = 0.2, ff = 100%, 64 frames), processed with different methods with optimized settings for zoom factors 1, 2, and 4. (a) Elad, (b) Lertrattanapanich, (c) Kaltenbacher (no zoom factor 1 results could be obtained with our implementation), (d) Hardie, (e) Farsiu, (f) Pham.

To illustrate the effect of an increasing zoom factor, Figure 10 shows performance curves of all SR reconstruction methods for zoom factors 1, 2, and 4. All methods processed the same 64 LR frames ($\sigma_{PSF} = 0.2$ and ff = 100%). From Figure 10 it is clear that the performance of zoom factors 2 and 4 for most methods (except for Kaltenbacher's method and Farsiu's method) is comparable. For low SNRs the performance of each method (for all zoom factors) is significantly better compared to LR performance. Here, the temporal noise reduction is visible. For high SNRs the results show an improvement of a factor 2, which approximately equals the amount of aliasing in the LR data. This explains why zoom factor 4 does not yield a significant better performance. Note that the bad performance of Kaltenbacher with zoom factor 4 compared with zoom factor 2 can be explained by the fact that this method has no regularization and hence becomes ill posed. Furthermore, an improvement by a factor 2 (between zoom factor 1 and zoom factors 2 and 4) is not obtained for low SNRs. Here, the temporal noise reduction is more relevant than the antialiasing. The performance of some SR reconstruction methods, when processed with zoom factor 1 under high SNR, is slightly worse compared to baseline LR performance. This could be explained by blurring in the fusion process and/or blurring as a result of registration errors.

6. CONCLUSIONS

From the results in the previous section, the following conclusions can be derived.

(1) From the results of the real-world data experiment it can be concluded that the performance of different SR reconstruction methods on real-world data can be predicted accurately by measuring the performance on simulated data, if a proper estimate of the parameters of the real-world camera system is available.

(2) With the ability to predict the performance of an SR reconstruction method on real-world data, it is possible to optimize the complete chain of a vision system. The parameters of the camera and the algorithm must be chosen such that the performance of the vision task is optimized.

(3) It is shown that with the TOD method the performance of SR reconstruction methods can be compared for a specific condition of the LR input data. Considering the imaging conditions (camera's fill factor, optical PSF, SNR) the TOD method enables an objective choice on which SR reconstruction method to use.

(4) Comparing the performance of the unregularized Kaltenbacher's method with the regularized methods of Hardie, Farsiu, and Pham (see Figure 9), it can be concluded that in general regularization is not required for good performance when many input frames are available.

(5) The relative performance of the various methods change a little as a function of SNR.

(6) The results presented in Figure 10 show that a larger zoom factor does not yield a better performance. This can be explained by the fact that sensors with high fill factors exert an amount of blurring on the LR input frames and therefore limit the resolution gain and hence the maximum achievable resolution gain. For high SNRs the resolution gain is approximately equal to the amount of aliasing in the LR data and for low SNRs the resolution gain is minor compared with the temporal noise reduction.

REFERENCES

[1] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, 2003.

[2] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution," *International Journal of Imaging Systems and Technology*, vol. 14, no. 2, pp. 47–57, 2004.

[3] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, 2002.

[4] Z. Lin and H.-Y. Shum, "Fundamental limits of reconstruction-based superresolution algorithms under local translation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 83–97, 2004.

[5] M. D. Robinson and P. Milanfar, "Statistical performance analysis of super-resolution," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1413–1428, 2006.

[6] P. Bijl and J. M. Valeton, "Triangle orientation discrimination: the alternative to minimum resolvable temperature difference and minimum resolvable contrast," *Optical Engineering*, vol. 37, no. 7, pp. 1976–1983, 1998.

[7] P. Bijl, K. Schutte, and M. A. Hogervorst, "Applicability of TOD, MTDP, MRT and DMRT for dynamic image enhancement techniques," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XVII*, vol. 6207 of *Proceedings of SPIE*, pp. 1–12, Kissimmee, Fla, USA, April 2006.

[8] T. Q. Pham, M. Bezuijen, L. J. van Vliet, K. Schutte, and C. L. Luengo Hendriks, "Performance of optimal registration estimators," in *Visual Information Processing XIV*, vol. 5817 of *Proceedings of SPIE*, pp. 133–144, Orlando, Fla, USA, March 2005.

[9] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the DARPA Image Understanding Workshop*, pp. 121–130, Washington, DC, USA, April 1981.

[10] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Upper Saddle River, NJ, USA, 1993.

[11] M. Elad and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space-

invariant blur," *IEEE Transactions on Image Processing*, vol. 10, no. 8, pp. 1187–1193, 2001.

[12] S. Lertrattanapanich and N. K. Bose, "High resolution image formation from low resolution frames using Delaunay triangulation," *IEEE Transactions on Image Processing*, vol. 11, no. 12, pp. 1427–1441, 2002.

[13] E. Kaltenbacher and R. C. Hardie, "High resolution infrared image reconstruction using multiple, low resolution, aliased frames," in *Proceedings of IEEE National Aerospace and Electronics Conference (NAECON '96)*, vol. 2, pp. 702–709, Dayton, Ky, USA, May 1996.

[14] 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*, vol. 37, no. 1, pp. 247–260, 1998.

[15] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[16] T. Q. Pham, L. J. van Vliet, and K. Schutte, "Robust fusion of irregularly sampled data using adaptive normalized convolution," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 83268, 12 pages, 2006.

[17] H. Knutsson and C.-F. Westin, "Normalized and differential convolution," in *Proceedings of IEEE Society Conference on Computer Vision and Pattern Recognition (CVPR '93)*, pp. 515–523, New York, NY, USA, June 1993.

[18] R. M. Haralick and L. Watson, "A facet model for image data," *Computer Graphics and Image Processing*, vol. 15, no. 2, pp. 113–129, 1981.

[19] J. Johnson, "Analysis of image forming systems," in *Proceedings of Image Intensifier Symposium*, pp. 249–273, Fort Belvoir, Va, USA, October 1958.

[20] J. M. Valeton, P. Bijl, E. Agterhuis, and S. Kriekaard, "T-CAT, a new thermal camera acuity tester," in *Infrared Imaging Systems: Design, Analysis, Modelling, and Testing XI*, vol. 4030 of *Proceedings of SPIE*, pp. 232–238, Orlando, Fla, USA, April 2000.

[21] L. J. van Vliet and P. W. Verbeek, "Better geometric measurements based on photometric information," in *Proceedings of IEEE Instrumentation and Measurement Technology Conference (IMTC '94)*, vol. 3, pp. 1357–1360, Hamamatsu, Japan, May 1994.

[22] T. Q. Pham, *Spatiotonal adaptivity in super-resolution of undersampled image sequences*, Ph.D. thesis, Quantitative Imaging Group, TU Delft, Delft, The Netherlands, 2006.

A. W. M. van Eekeren was born in 1977. He received his M.S. degree in 2002 from the department of Electrical Engineering at the Eindhoven University of Technology. He did his graduation project within Philips Medical Systems on the topic of image enhancement. Subsequently he worked one year at the Philips Research Laboratory on image segmentation using level sets. In 2004 he started his Ph.D. project entitled "super-resolution on small moving objects" at the Electro-Optics group within TNO in collaboration with the Quantitative Imaging group at the Delft University of Technology. His research interests include image restoration, super resolution, image quality assessment, and object detection.

K. Schutte received his M.S. degree in Physics in 1989 from University of Amsterdam and received his Ph.D. degree in 1994 from University of Twente on his thesis "knowledge-based recognition of man-made objects." Subsequently he had a postdoctoral position with the Delft University of Technology's Pattern Recognition (now Quantitative Imaging) group. Since 1996 he is employed by TNO, currently as Senior Research Scientist Electro-Optics within the Business Unit Observation Systems. Within TNO he has actively led multiple projects in areas of signal and image processing. Recently he has led many projects including super-resolution reconstruction for both international industries and governments, resulting in super-resolution reconstruction-based products in active service. His research interests include pattern recognition, sensor fusion, image analysis and image restoration. He is Secretary of the NVBHPV, the Netherlands branch of the IAPR.

O. R. Oudegeest received his B.S. degree in applied physics at Delft University of Technology in 2004. His B.S. thesis was titled: "alternatives for CT scanning in the diagnosis of endovascular aneurysm stent-graft migration." In 2007 he received his M.S. degree in applied physics at Delft University of Technology on the subject of "super-resolution on and classification of small moving objects." His research interests include super resolution, tracking, and pattern recognition.

L. J. van Vliet is a Full Professor in multidimensional image processing and analysis at Delft University of Technology. He studied applied physics at Delft University of Technology and received his Ph.D. degree cum laude in 1993. His thesis entitled "grey-scale measurements in multidimensional digitized images" presents novel methods for sampling-error-free measurements of geometric object features. He has worked on various sensor, image restoration, and image measurement problems in quantitative microscopy and medical imaging. In 1996 he was awarded a fellowship of the Royal Netherlands Academy of Arts and Sciences (KNAW). He was a Visiting Scientist at LLNL (1987), UCSF (1988), Amoco ATC (1989–1990), Monash University (1996), and LBNL (1996).