

Measuring the performance of super-resolution reconstruction algorithms

Judith Dijk, Klamer Schutte, Adam W.M. van Eekeren and Piet Bijl

TNO

P.O. Box 96864, 2509 JG The Hague, The Netherlands

Email: Judith.Dijk@tno.nl

ABSTRACT

For many military operations situational awareness is of great importance. This situational awareness and related tasks such as Target Acquisition can be acquired using cameras, of which the resolution is an important characteristic. Super resolution reconstruction algorithms can be used to improve the effective sensor resolution. In order to judge these algorithms and the conditions under which they operate best, performance evaluation methods are necessary. This evaluation, however, is not straightforward for several reasons. First of all, frequency-based evaluation techniques alone will not provide a correct answer, due to the fact that they are unable to discriminate between structure-related and noise-related effects. Secondly, most super-resolution packages perform additional image enhancement techniques such as noise reduction and edge enhancement. As these algorithms improve the results they cannot be evaluated separately. Thirdly, a single high-resolution ground truth is rarely available. Therefore, evaluation of the differences in high resolution between the estimated high resolution image and its ground truth is not that straightforward. Fourth, different artifacts can occur due to super-resolution reconstruction, which are not known on forehand and hence are difficult to evaluate.

In this paper we present a set of new evaluation techniques to assess super-resolution reconstruction algorithms. Some of these evaluation techniques are derived from processing on dedicated (synthetic) imagery. Other evaluation techniques can be evaluated on both synthetic and natural images (real camera data). The result is a balanced set of evaluation algorithms that can be used to assess the performance of super-resolution reconstruction algorithms.

Keywords: super-resolution reconstruction, signal-to-noise ratio, MTF enhancements, Nyquist frequency

1. INTRODUCTION

For many operations situational awareness is of great importance. This situational awareness can be obtained by using cameras. The resolution of the resulting images is of great importance for this evaluation. This can be improved using super-resolution reconstruction techniques. Multi-frame 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. During the last decade numerous SR methods have been reported in the literature. Reviews can for instance be found in van Eekeren,⁷ Farsiu⁹ and Park¹² There is a need for evaluating these algorithms. However, this is hard due to four different reasons. First of all, it is hard to determine the difference between noise, which should not be attenuated, and structure, which should gain higher frequencies. For this reason, only an evaluation based on frequency will not provide the entire answer. Secondly, most super-resolution methods perform other image enhancement techniques such as noise reduction and edge enhancement. This may not be the primary goal for super-resolution techniques, but as these algorithms improve the results they cannot be evaluated separately. The three effects (super-resolution reconstruction, noise reduction and edge enhancement) are not independent, which makes it somewhat harder to interpret the results. Third, there is no single high-resolution ground truth. The resulting image may be shifted half a pixel without any penalty. For this reason, evaluation of the differences in high resolution between the estimated high resolution image and its ground truth is not that simple. Last, different artefacts can occur due to super-resolution reconstruction, which are not known on forehand. The idea to evaluate these unknown artefacts is really hard.

Some of the evaluation measures are task related. An example of this is the evaluation of point target detection using super-resolution reconstruction.⁵

In this paper we present different evaluation techniques that evaluate different aspects of image quality measures for super-resolution reconstruction. Some of these evaluation techniques are derived from processing on dedicated imagery. Other evaluation techniques, such as the residue image, can be evaluated on every set of processed images. This paper is organized as follows. In section 2 methods for super-resolution reconstruction and evaluation algorithms from literature are discussed. In section 3 the tested algorithms are described, and in section 4 results for these algorithms are shown. In the last section the conclusions are given and issues for new research are discussed.

2. LITERATURE

2.1 Super-resolution reconstruction

In case multiple frames of a scene are available temporal filtering can produce frames of a higher quality. This can both be static frames or a video. One of the algorithms that can be used for this is super-resolution reconstruction. As already said in the introduction, this is the process of combining a set of undersampled (aliased) low-resolution (LR) images to construct a high-resolution (HR) image or image sequence. An enabler for this technique is accurate alignment of the image, so that the scene structure will be preserved. One of the challenges of super-resolution reconstruction is the preservation of moving elements in the scene.

In this paper we use the Dynamic Super Resolution algorithm described by Schutte et.al¹³ which can be implemented in real-time. This algorithm can also be used for noise reduction, that is, for enhancement without increasing the number of pixels.

2.2 Evaluation algorithms from literature

To measure the performance of SR reconstruction, several quantitative measures such as mean squared error (MSE) and modulation transfer function (MTF) are often used.¹¹ A simple one is to take the root mean square error (RMSE) between two images. However, it is not typical what images in this case to take, as there are high resolution, low resolution and even ground truth images. Evaluation functions based on the MTF are also commonly used. An example is presented in the next section.

van Eekeren⁸ uses the normalized mean squared error (NMSE) between the SR result of a small moving object and its ground truth as a figure-of-merit. Note that this measure considers only the foreground intensities, the background intensities are set to zero

$$NMSE = \frac{1/N \sum_{n=1}^N (z_{gt}(n) - z_{est}(n))^2}{\max(z_{gt})^2} \quad (1)$$

with N the number of HR pixels, z_{est} the estimated foreground contributions using SR and z_{gt} the ground truth. Normalization is performed with the squared maximum value of z_{gt} .

3. EVALUATION TECHNIQUES

In this section the proposed evaluation algorithms are described. Note that the overall quality cannot be based on the outcome of one of these algorithms separately, as a high outcome of one of the values can also indicate artefacts instead of a good image.

3.1 Minimum Resolvable Temperature Difference evaluation

The idea of the Minimum Resolvable Temperature Difference (MRTD) measurement is the same as the psychophysical measurement of the MRTD, that is, see if the four bar patterns are still visible. As input an image with a bar pattern on a certain spatial frequency is used, see Figure 1. The MRTD is measured by comparing the signal (maximum-minimum) to the noise, in equations

$$CTF = \max - \min \quad (2)$$

$$MRTD = CTF / \text{noise} = CTF * SNR \quad (3)$$

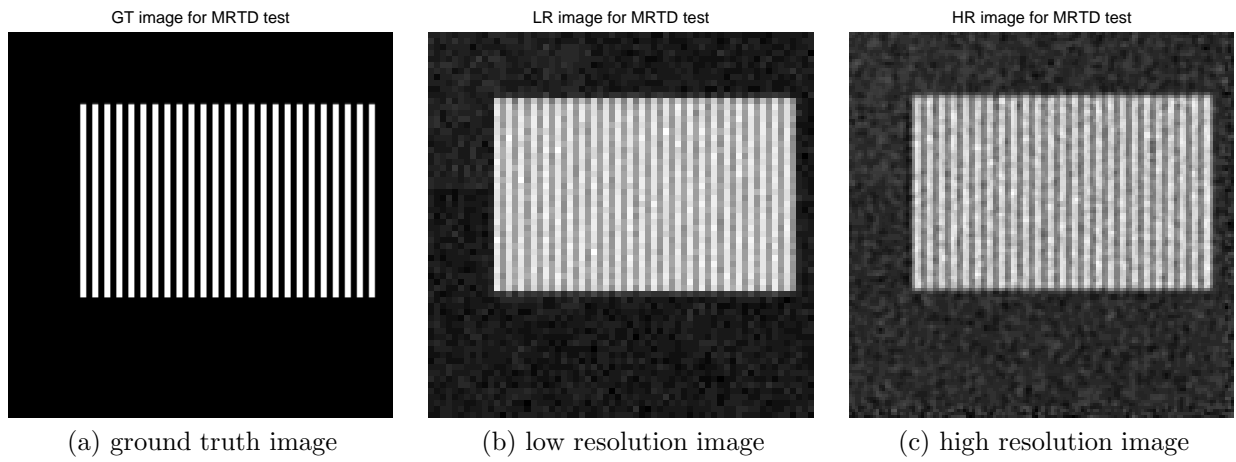


Figure 1. The MRTD test images.

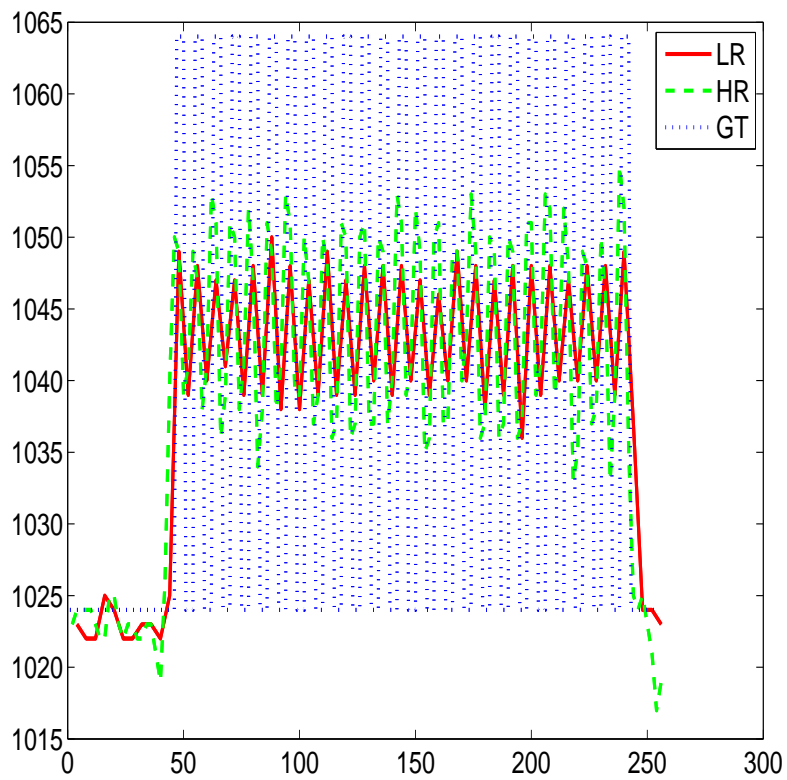


Figure 2. The MRTD cross sections.

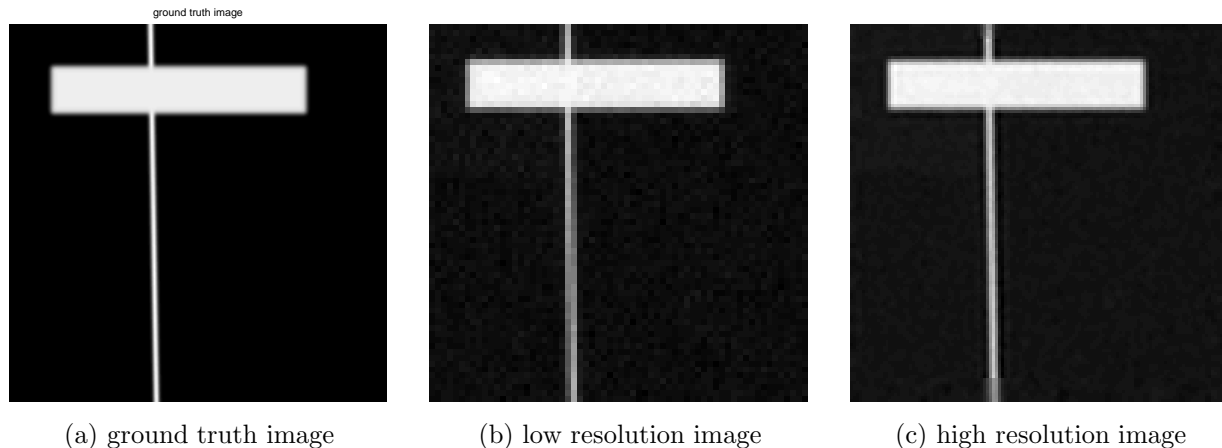


Figure 3. The MTF and SNR test images.

A larger value will give a better image for these frequencies. Note that for the SNR calculation we assume that there is no signal, so that this value only depends on the noise.

One of the benefits of this method is that it measures the improvement at a given frequency, that is the frequency of the bar pattern. We chose the Nyquist frequency of the low resolution image (LR Nyquist frequency). The main disadvantage is that the improvement is only measured for this frequency, so for a more thorough frequency evaluation this measure needs to be determined for a large number of bar patterns. The (MRTD) is proportional to the SNR and therefore inverse proportional to the temporal noise. In order to improve (increase) the MRTD the noise can be decreased and/or the CTF can be increased.

3.2 Modulation Transfer Function evaluation

For measuring the Modulation Transfer Function (MTF), an image containing an almost vertical line is used as input function. The idea is that the line is a delta function. The small slope is used to obtain different sampling of the delta function, or in other words, to have multiple aliasing configurations. To be able to also perform motion estimation on this image, a block is added. This gives structure in both the x and the y direction. This image is shown in Figure 3. In Figure 3(a) the ground truth image is shown. Using a camera model with blur (0.35), fillfactor (0.9) and temporal noise (1) a number of low resolution images are generated. One of these LR images is shown in Figure 3(b). These LR images are processed using the super-resolution reconstruction method under evaluation. In this case we used the Super-resolution reconstruction method described by Schutte et.al.¹³ The resulting high resolution image is shown in Figure 3(c).

These three images (LR, HR and ground truth) are used to estimate the MTF enhancement. It is also possible to use a number of LR, HR and GT images. First we look at the resulting lines. A cross-section of a ground-truth, low resolution and high resolution image is given in Figure 4(a). The 1-D FFT is determined for all lines, and averaged. These average FFT's are shown in Figure 4(b). After this the ratio between the LR and GT FFT's and the HR and GT FFT's is determined. These ratio's are shown in Figure 4(c).

Based on these ratio's we determine two evaluation measures:

- HR/LR at a certain frequency
- -3db point for HR and LR

For both measures a higher value indicates a better performance. In the results later we report only the second value (-3db point), which we named the MTF value. The other value is comparable to the MRTD measurement. These measures depend not only on the resolution enhancement, but are also dependent on the sensitivity enhancement. A higher sensitivity enhancement will reduce the higher frequencies in the HR image and therefore

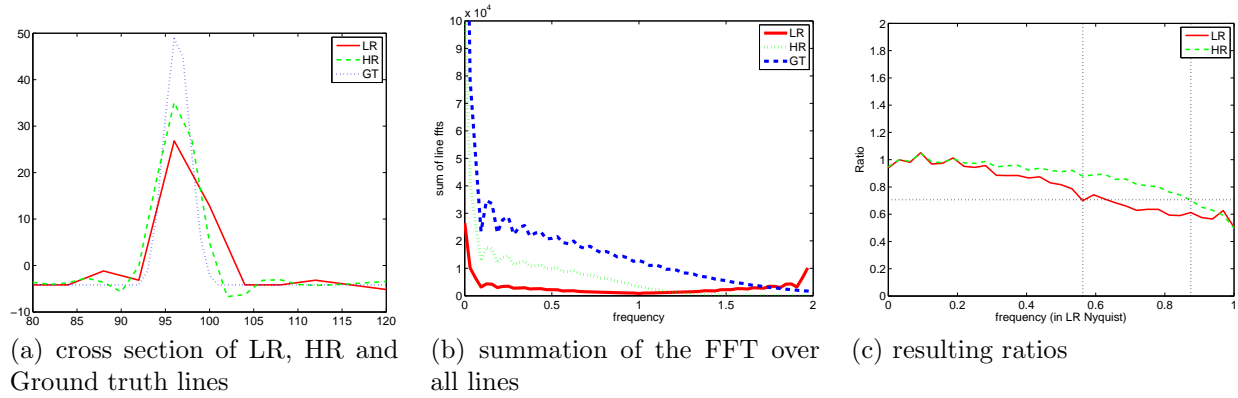


Figure 4. The MTF evaluation. In (a) the cross section of a single line is shown, for the LR, HR and ground truth image. In (b) the Fourier transformation, summed over all lines, is shown. In (c) the LR and HR FFT's are divided by the ground truth FFT. The result is normalized to 1. The -3db point is indicated by the cross section of the normalized FFT's and the line through 0.707.

the estimated MTF will be lower. For this reason we determine these values in image sequences where the LR and the HR images have the same temporal noise. This value is measured with the SNR measure. Note that we measure a processed MTF, not the pre-filter MTF.

3.3 SNR

Sensitivity enhancement is measured as the decrease in noise in the image. This can be both temporal as well as fixed pattern noise. The noise is measured as the standard deviation in areas without scene structure. The assumption is that, because of the lack of scene structure, all variation is due to noise, so that the standard deviation is a good measure here.

The sensitivity enhancement is given by

$$h = SNR_{HR}/SNR_{LR} = noise_{LR}/noise_{HR} \quad (4)$$

under the assumption that the signal strength remains the same. A higher value of the measure indicates better performance.

In this case we use the same images as for the MTF test, as given in Figure 3. The bottom right parts are used for the SNR evaluation. The region of interest is chosen such that the edge effects for the super-resolution reconstruction are not taken into account.

3.4 Edge Overshoot test

An edge image can be used to analyze the high-spatial frequency part of the algorithm frequency transfer function and to assess possible ringing or overshoot artifacts. The Edge Overshoot method assesses the extent of artifacts introduced by the algorithm in the vicinity of large intensity changes. The method can be best understood from Figure 5 that shows an input intensity edge and a (hypothetical) output.

From this Figure, two quality factors can be defined: the slope quality and the overshoot quality. The slope quality is defined as:

$$Q_{erf} = \frac{-2\beta}{\pi} \quad (5)$$

in which

$$\beta = \arctan(0.85(a - b)/(x_t - x_b)) \quad (6)$$

The parameters x_t and x_b are the spatial positions corresponding to the intensities $(a + b)/2 \pm 0.85(a - b)/2$ in the output pattern. The factor 0.85, chosen arbitrarily, means that x_t and x_b are located at 85% of the maximum and minimum edge values, a and b .

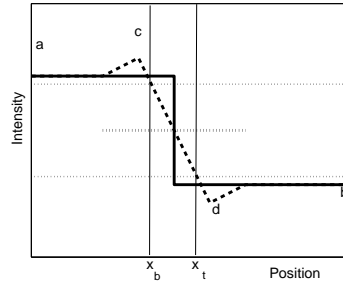


Figure 5. Intensity discontinuity for edge evaluation method. Dashed line: input edge. Solid line: hypothetical output. The dotted lines are the intensities $(a + b)/2 \pm 0.85(a - b)/2$

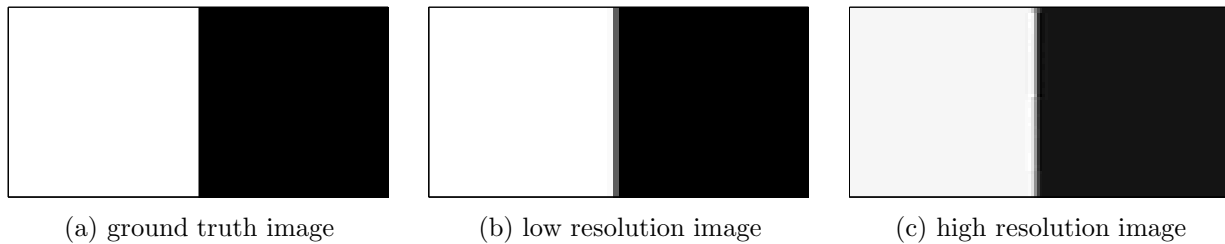


Figure 6. The ERF input images.

The overshoot quality is defined as:

$$Q_{over} = 1 - \frac{(c - a) + (b - d)}{2(a - b)} = 1 - \frac{overshoot}{edge} \quad (7)$$

With this definition, the quality factor approaches unity when the levels a and c on one hand and b and d on the other coincide. Note that these measures are applied on a processed edge and not on a pre-filter edge. The input images and cross sections of the lines are shown in Figure 6 and 7.

3.5 Residue image

The residue image can be determined during the processing itself. It determines a value estimating how well the low resolution fits the downscaled version of the estimated high resolution image. In the Hardie algorithm,¹⁰ this is for instance given by the term

$$y_m - \sum_{r=1}^H w_h z_m \quad (8)$$

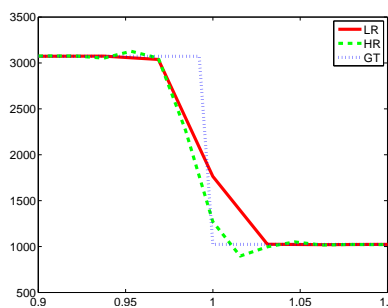


Figure 7. The ERF cross sections.

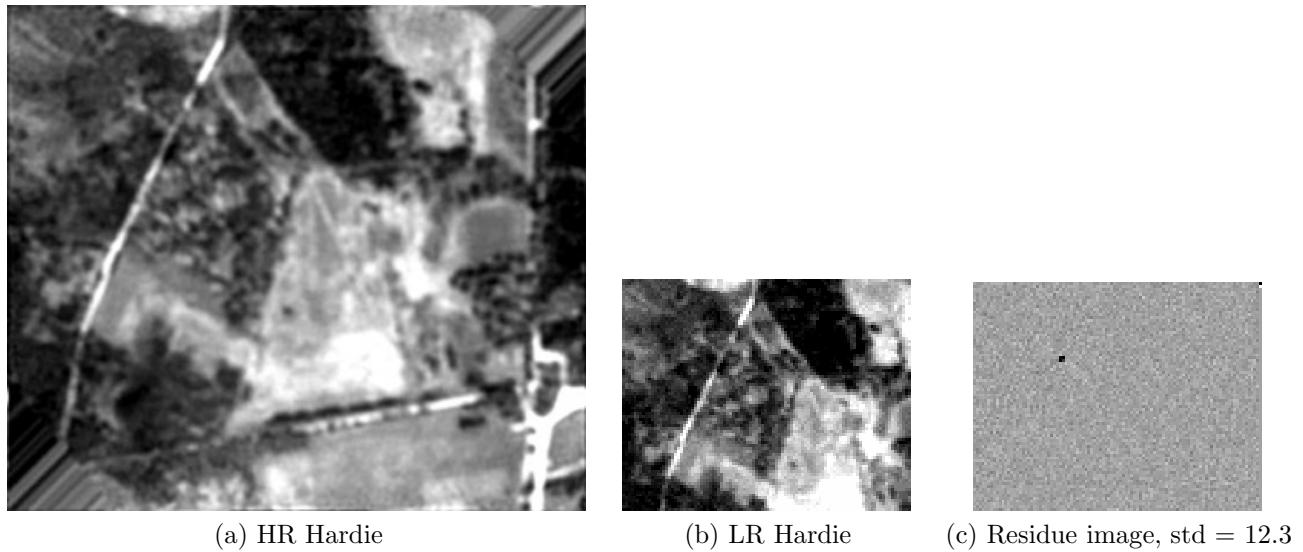


Figure 8. Obtaining the residue image for a resolution enhancement factor of 2.

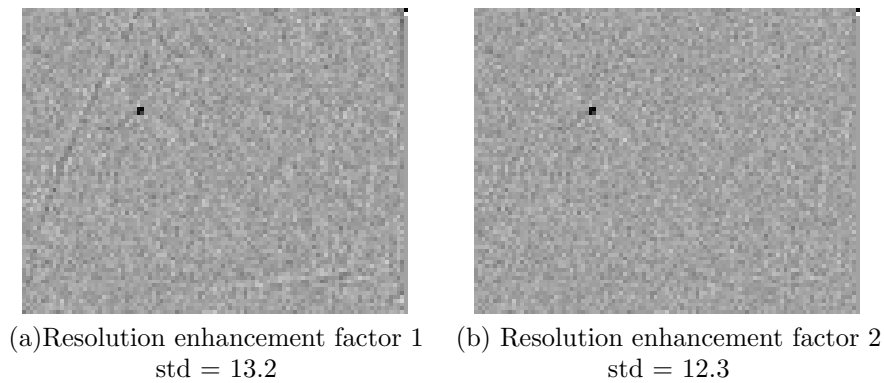


Figure 9. Visual results for residue image

in which y_m is the low resolution image, z_m is the high resolution image and w_h indicates the camera model. This value can be shown as an image and evaluated visually, as can be seen in Figure 8. For non-optimal performing algorithms a significant part of the image can be seen in the residue. This is shown in Figure 9 a. These results are taken from Dijk et.a.⁵ For good performing super-resolution reconstruction algorithms the residue image will be containing mostly temporal noise, and no structure from the input images, see Figure 9 b. Note that the dark spot is a moving object, which is visible in the LR image and not in the HR image, and will therefore give a large absolute residue value.

The residue can be represented by its standard deviation. Due to the noise reduction step of the algorithm, this will with perfect performance be the same as the noise in the input image. Without the noise reduction, the resulting noise will be similar to $\sqrt{2}$ * input noise, assuming that they are not correlated. Values much larger than the standard deviation of the input image indicate that the algorithm is not performing well.

3.6 TOD

To evaluate the performance also the triangle orientation discrimination (TOD) method can be used. In this test the orientation of equilateral triangles is measured. Originally this method was developed for subject experiments with the human-in-the-loop.¹ Van Eekeren⁷ also used this method using a automatic scoring algorithm.

Table 1. Results for tests on synthetic data and natural data. std is the standard deviation of the residue image.

Settings	Synthetic data						
	std	CTF	SNR	MRTD	MTF	Q_{erf}	Q_{over}
Standard	1.40	1.60	1.66	2.65	1.11	0.13	0.999
Smaller SE factor	1.35	2.27	1.17	2.67	1.39	0.13	0.999
Large blur	4.73	3.16	1.70	5.39	1.33	0.25	0.881
Small blur	1.35	1.34	1.92	2.57	0.89	0.13	0.999
Settings	Natural data						
	std	CTF	SNR	MRTD	MTF	Q_{erf}	Q_{over}
Standard	13.23	2.44	0.76	1.86		0.99	0.93
Smaller SE factor	11.14	2.43	0.77	1.87		0.97	0.93
Large blur	33.52	3.11	0.64	2.01		0.97	0.65
Small blur	13.98	2.38	0.75	1.79		0.99	0.93

The TOD method quantitatively measures the SR performance on the image background and on moving objects. From experiments it can be concluded that under proper conditions, the SR performance on moving objects is similar as the SR performance on the background. The TOD method determines the smallest triangle size in an image of which the orientation (apex up, down, left, right) can be discriminated. Independent variables are the triangle size S and the contrast ΔT with respect to the background. A detailed description of the assessment is given by Bijl.² The TOD results can also be related to object classification tasks.

4. RESULTS

We show the results for the different algorithms for different settings of the TNO signal conditioning suite.¹³ This real-time image enhancement software can be used with different settings, such as the resolution enhancement factor and the sensitivity enhancement factor. Also the variables of the camera model can be changed. We perform both the tests on the synthetic data, as on the natural images recorded by Weiss et.al.¹⁴ for the purpose of evaluating super-resolution reconstruction algorithms. In these natural images the CTF is calculated on the left bottom pattern on the right panel. The ERF and SNR measures are calculated using the black body panel below the truck. The MTF is not calculated as there is no isolated line pattern in the image. The TOD method could be evaluated using the TOD board; these results are not reported here.

In table 1 results for all tests are shown for four different settings.

- **Standard settings** the standard settings of the TNO super-resolution algorithm: Resolution Enhancement Factor 2, Camera blur 0.3, fillfactor 0.9, SE factor 2.78. For all metrics except for the ERF test an improvement is shown (indicated by a value larger than one), whereas the Q_{over} remains close to 1.
- **Smaller SE factor** the sensitivity enhancement factor is set to 2, which is smaller than for the standard settings, which results in a higher CTF and MTF and a smaller SNR. This is indeed shown in the results.
- **Blur of camera model too large** the blur of the camera model is over estimated, that is 0.9 instead of 0.3, which results in more sharpening and also in ringing. This can be seen by a smaller ERF factor and a higher standard deviation of the residue image. The ringing also gives a higher CTF and MRTD value. This shows that evaluating the test results separately will not give a good indication of the performance of the system.
- **Blur of camera model too small** the blur of the camera model is under estimated, that is 0.1 instead of 0.3. In this case the resolution enhancement is smaller than could be expected, resulting in a CTF and MTF that is smaller. This is indeed shown in the results.

Enlargements of results with these settings are shown in Figure 11. Here the results that are indicated by the tests can be seen visually.

The optimal standard deviation of the residue image depends on the noise in the input image. In the synthetic imagery a noise of 1 was applied. The noise in the natural images is, given the results, about a factor 10 higher.

In Figure 10 and 11 we show the residue images and standard deviation for different values of the resolution enhancement factor. The results show an improvement for larger resolution enhancement factors, both in the imagery and the standard deviation of the residue images.

The main conclusions we can base on these results is that for the settings with resolution enhancement factor 2, standard settings and settings with a smaller SE factor perform equally well. The choice for this settings would be if a smoother image is preferred, or an image with more high resolutions. A blur estimate which is too large results in ringing, which can be seen in a large performance drop of the Q_{over} and std tests. It can also be seen, that the overhoot generated “good” results for the MRTD evaluation, but in this case, this does not indicate a good performance. A blur estimation which is too low results in an underperformance of the resolution enhancement, which can be seen by a smaller value for the CTF and MTF.

A resolution enhancement factor of 3 gives a better result than a smaller resolution enhancement factor. This can be seen in the residue image, both visually as numerically.

5. CONCLUSIONS AND DISCUSSION

In this paper we present a set of evaluation techniques to assess super-resolution reconstruction algorithms. The evaluation techniques need specific structures in the image, which can be evaluated in the high resolution image only (the ERF measures), in both the low resolution and high resolution images (MRTD and SNR) or even in the LR, HR and ground truth images. The results show that different settings for the algorithm will give different results

The residue can be determined on both synthetic and natural images (real camera data). The residue image can be used as a “sanity check”, that is, if (many) structures from the LR images are still visible the algorithm is less good. From the residue image also values can be determined such as the standard deviation, which gives a numerical indication if the algorithm performs well. Using special natural images, such as recorded by Weiss et.al,¹⁴ the synthetic values can also be determined from these natural images. It is shown that the trend in the values is the same for the synthetic as for the natural images. The values themselves are different due to different input conditions.

All evaluation techniques combined form a balanced set of evaluation algorithms that can be used to assess the performance of super-resolution reconstruction algorithms .

Our results of aggressive sharpening within Super Resolution Reconstruction show a balance between actual sharper details and ringing related artefacts. For any actual application this balance shall be adjusted, and thus the proper balanced set of metrics for each application shall be different

REFERENCES

- [1] Bijl, P. and Valeton, J.M., “*TOD, the alternative to MRTD and MRC*,” Optical Engineering, 37(7), 1976-1983 (1998).
- [2] Bijl, P. and Valeton, J.M., “*Guidelines for accurate TOD measurements*,” Proc. SPIE 3701, 14-25 (1999)
- [3] Bijl, P., Schutte, K. and Hogervorst, M. A. “*Applicability of TOD, MTDP, MRT and DMRT for dynamic image enhancement techniques*,” Proc. SPIE 6207, 62070E (2006).
- [4] Dijk, J., “*In search of an Objective Measure for the Perceptual Quality of Printed Images*” PhD thesis, Delft University of Technology (2004).
- [5] 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*”, Opt. Eng. 47, 096401 (2008).
- [6] Dijk, J. van Eekeren, A.W.M., Schutte, K., de Lange D.J.J. and van Vliet, L.J., “*Performance study on point target detection using super-resolution reconstruction*,” SPIE 7335, 7335 (2009)
- [7] 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, 111, Article ID 43953 (2007).

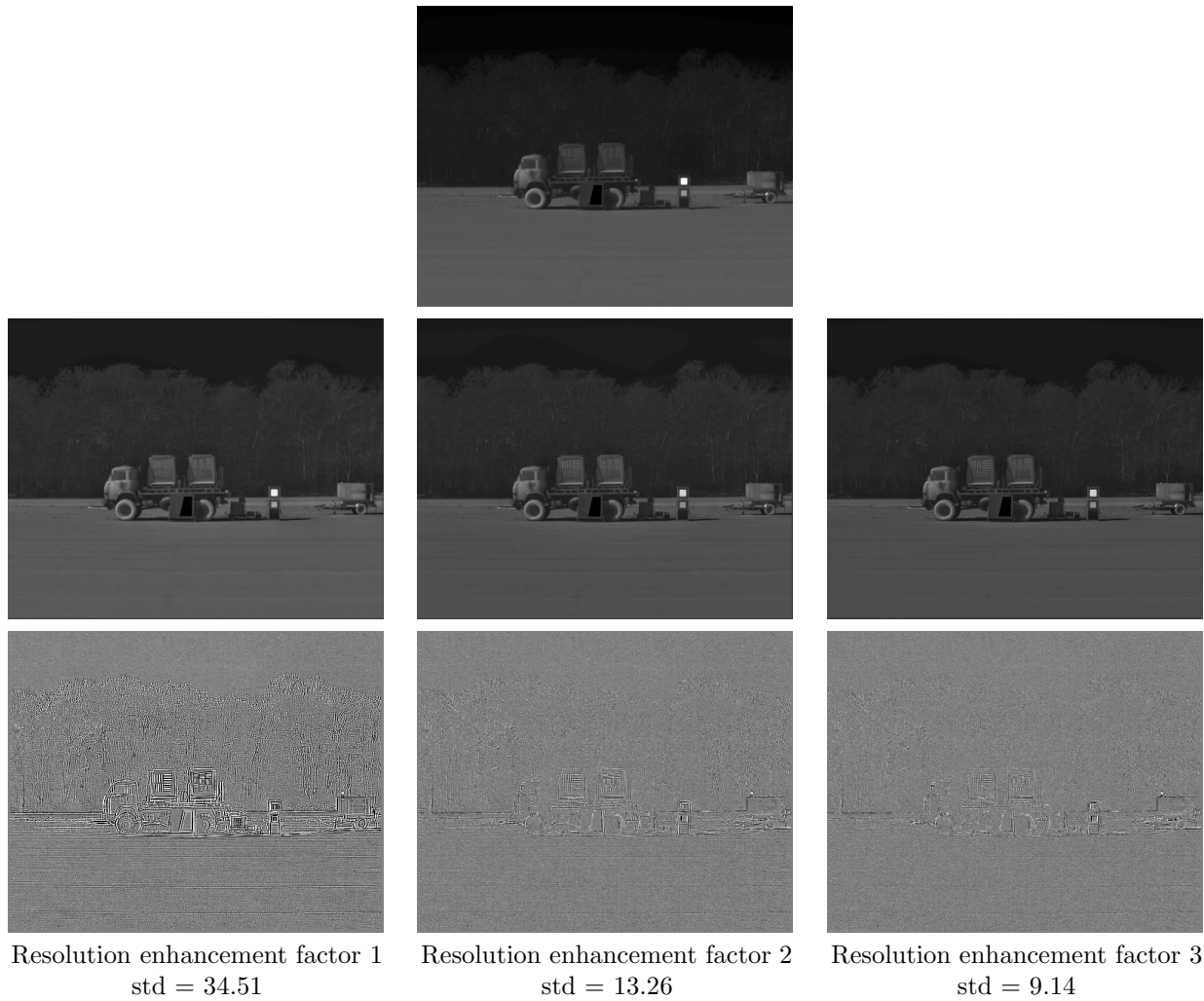


Figure 10. Visual results for super-resolution reconstruction with different resolution enhancement factors. In the top row the input image is shown. The center row shows the resulting super-resolution image. In the bottom row the residue image is shown, with the same gray scale. It can be seen that the variance in the higher resolution images is much smaller. This is also shown in the value of the standard deviation (std). Enlargements of the images are shown in Figure 11

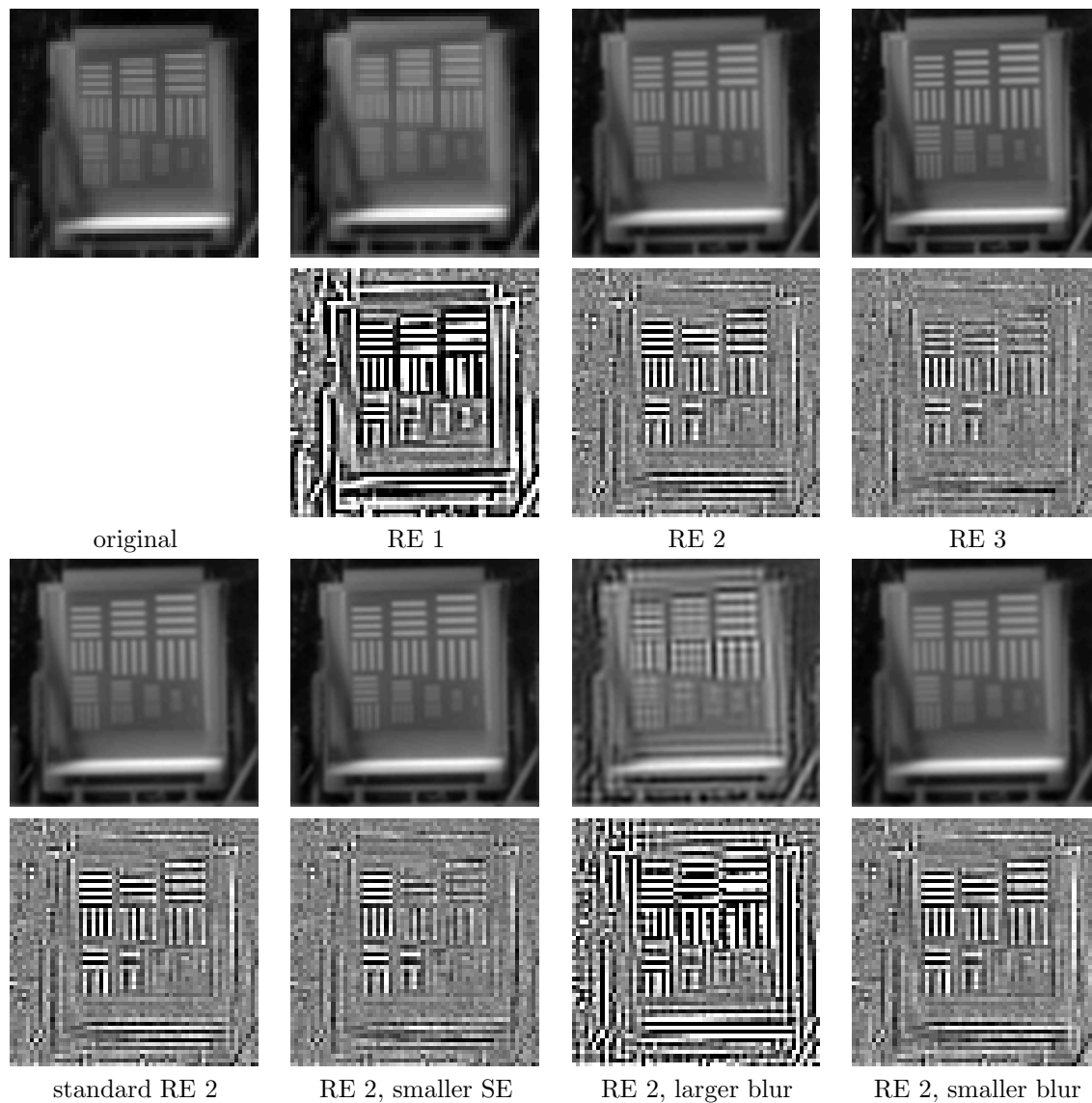


Figure 11. Visual results for the results and residue images, for a region of interest of the images. In the top the results for different resolution enhancement factors are shown. It can be seen that for a larger RE factor the bar pattern is better visible. In the bottom rows the results for different settings of the algorithm for RE factor 2 are shown. These results are also used to calculate the test results in table 1.

- [8] van Eekeren, A.W. M., Schutte, and van Vliet, L.J., "*Multiframe Super-Resolution Reconstruction of Small Moving Objects*," IEEE transactions on image processing 19(11), 2901-2912 (2010).
- [9] Farsiu,S., Robinson, D. Elad, M. and Milanfar, P., "*Advances and Challenges in Super-Resolution*," International Journal of Imaging Systems and Technology 14, 47-57 (2004)
- [10] Hardie, R.C., Barnard, K.J., Bognar, G.J. Armstrong, E. E., and E. A. Watson, "*High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system*," Optical Engineering 37(1), 247-260 (1998)
- [11] Holst, G.C. *Testing and evaluation of infrared imaging systems*, JCD Publishing, ISBN 978-0-8194-7247-2 (2008)
- [12] Sung Cheol Park, Park, S.C., Park, M.K. and Kang, M.G., "*Super- Resolution Image Reconstruction: A Technical Overview*," IEEE Signal Processing Magazine 20(3), 21-36 (2003).
- [13] Schutte,K. de Lange, D.J.J. and van den Broek, S. "*Signal conditioning algorithms for enhanced tactical sensor imagery*," Proc SPIE 5076, 92-100 (2003).
- [14] Weiss,R. Adomeit, U., Chevalier, P. and Landeau, S., "*A standard data set for performance analysis of advanced IR image processing techniques*", Proc SPIE 8855, 8855 (2012).