Super-Resolution of Moving Objects in Under-Sampled Image Sequences

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.dr.ir. J.T. Fokkema,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op vrijdag 12 juni 2009 om 10:00 uur

door

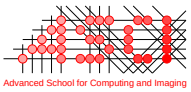
Adam Wilhelmus Maria VAN EEKEREN

elektrotechnisch ingenieur
geboren te Roosendaal en Nispen

Dit manuscript is goedgekeurd door de promotor:
Prof.dr.ir. L.J. van Vliet

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof.dr.ir. L.J. van Vliet,	Technische Universiteit Delft, promotor
Dr. K. Schutte,	TNO Defensie en Veiligheid, Den Haag
Prof.dr.ir. J. Biemond,	Technische Universiteit Delft
Prof.dr.ir. P.M. van den Berg,	Technische Universiteit Delft
Prof.dr. L.J.M. Rothkrantz,	Nederlandse Defensie Academie en Technische Universiteit Delft
Prof.dr.ir F.C.A. Groen,	Universiteit van Amsterdam
Prof.dr. P. Scheunders,	Universiteit Antwerpen, België
Prof.dr. I.T. Young,	Technische Universiteit Delft, reservelid



This work was carried out in the ASCI graduate school.
ASCI dissertation series number 176.

The printing of this thesis was financially supported by TNO Defence, Security and Safety.

ISBN 978-90-5986-317-0

Aan mijn ouders en zus

Contents

Chapter 1

Introduction

Vision is one of the five human senses. It was already studied a few hundred years BC by two major ancient Greek schools. The first school (Euclid, Ptolemy and their followers) explained vision with the 'emission theory', which states that vision occurs when rays emanate from the eyes and are intercepted by visual objects. The second school (Aristotle, Galen and their followers) advocated the 'intromission theory', which interprets vision as coming from something, a representation of the object, entering the eyes. Although they had no experimental foundation to support their theory, they were not very far from what we currently know about how our eyes work.

Ibn Al-Haytham (965 - 1039), the 'father of optics', was the first one to reconcile both schools of thought in his influential *Book of Optics* [28]. He argued that vision is due to light from objects entering the eyes. Furthermore, he pioneered the area of the psychology of visual perception, being the first scientist to argue that vision occurs in the 'brain', rather than the eyes. He pointed out that vision and perception are subjective.

Nowadays we have detailed knowledge about the Human Visual System (HVS) [41], although many aspects are still not completely understood. Light entering the eye's pupil is imaged by the lens onto the light sensitive cells (rods and cones) of the retina. Visual information from different parts of the retina is systematically ordered in the primary visual cortex and from there send to different places in our brain. There are two main pathways from the primary visual cortex: 1) the 'Where Pathway', which is associated with motion, representing object locations, and control of our eyes and arms and 2) the 'What Pathway', which is associated with shape recognition, object representations and storage in long-term memory [22].

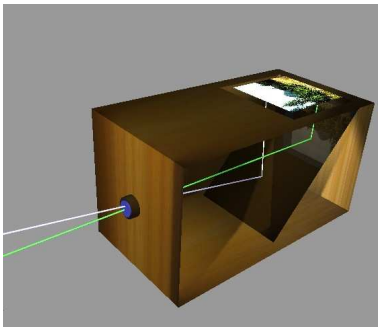


Figure 1.1: *Principle of a camera obscura.*

Al-Haytham was also the first to build a camera obscura (see Figure 1.1), i.e. a black box with a hole in one side which projects the incoming light (via a mirror) to another side of the box. In early days this projection was made on paper facilitating an artist to copy the image.

As one can see, the step from human vision to imaging and to image processing is not that big. In principle the following analogy can be made: our eyes play the role of a camera with the retina as imaging sensor and the virtual cortex as processing device. If our eyes would not be connected to our brain, we would not be able to remember the things we see and therefore we will not be able to recognize things, reason and make decisions based on the perceived visual information. This completes the sense-think-act loop. The same applies to imaging and image processing: imaging without image processing (including recording) seems to be of no use.

Although human vision is a very interesting topic, it is beyond the scope of this thesis. Here, the focus will lie on a specific area of image processing in which we aim to improve the resolution, the Signal-to-Noise Ratio (SNR), as well as the contrast between foreground (such as moving objects) and background. This enables the observer to extract more information from the image, such as small, low contrast details. The observer might detect or recognize something in the processed image which he was not able to do in the raw data. Specifically, this thesis focuses on improving the detection and recognition of moving objects in under-sampled image sequences.

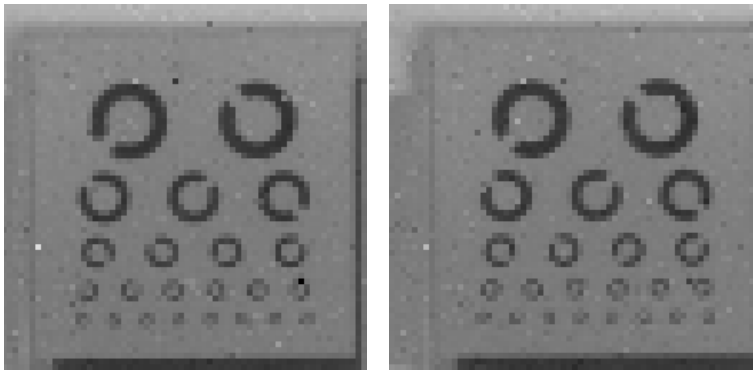


Figure 1.2: *Two LR frames (64×64 pixels) taken at different time instances showing artifacts caused by under-sampling.*

1.1 Problem description

The resolution of an optical system determines the finest detail that can be resolved and is proportional to $f\lambda/D$, where f is the focal length of the lens, λ the wavelength of the incoming light and D the diameter of the lens aperture. An optical system with a larger aperture or a smaller focal length is therefore capable of capturing finer details. However, the perception of the imaged scene is also determined by the sampling density of the sensor. To be able to reconstruct a bandlimited signal, the sampling frequency must be at least two times the highest frequency of the signal. When the sampling frequency is smaller, the Nyquist-Shannon sampling requirement is not met [56]: the signal is said to be *under-sampled*.

Spatial under-sampling hampers reconstruction of the scene after optical imaging. One experiences strong staircase effects around edges, thin lines will appear interrupted, small details will be missed or severely corrupted, and repetitive patterns may appear at a different spatial frequency. Some of these artifacts are visible in Figure 1.2.

Tackling under-sampling by decreasing the pixel pitch of the detector has the disadvantage that it reduces the SNR. Here we assume that the percentage of photosensitive area (fill-factor) of the detector stays the same. Using a lens with a larger focal length will capture details larger on the image plane, but has some disadvantages as well: they are expensive, tend to be large and provide less overview due to a smaller viewing angle. However, computerized postprocessing is a way to overcome the artifacts caused by under-sampling. Multi-frame Super-

Resolution (SR) reconstruction is the postprocessing method used in this thesis to reconstruct the scene from an under-sampled image sequence. It aims at reconstructing an image of the underlying scene, free of sampling artifacts and noise, by using a model of the camera (optics and sensor) characteristics and some prior knowledge that applies to virtually all scenes.

In many applications, the most interesting events are related to changes occurring in the scene: e.g. moving persons or moving objects. Especially on these occurrences an observer wishes to see a lot of detail. This makes resolution improvement useful for changing/moving objects in the scene. The main goal of this thesis is therefore to improve the resolution, SNR and contrast of moving objects in under-sampled image sequences by means of multi-frame SR reconstruction[1]. These improvements will help to increase the detection, and recognition rate of moving objects.

1.2 Optical imaging

In comparison to the pinhole imaging of a camera obscura, modern cameras make use of lenses (optics). This permits the collection of more light/radiation while keeping the scene in focus. Nowadays, there exist a wide range of different optical imaging devices: high-resolution megapixel cameras for digital photography, infrared cameras for night-vision, microscopes for biological research, etc.

Application of SR reconstruction is especially effective for imaging devices which have a coarse sampling grid and therefore tend to under-sample the data. Infrared cameras belong to this class due to the relatively large wavelength of infrared light in comparison to visible light. Although infrared data is used a lot in this thesis, the proposed algorithms are not limited to this type of data.

1.2.1 Camera model

If the world is observed by an electro-optical camera system, the recorded data $f[k,l]$ depends on the various steps depicted in Figure 1.3.

3D to 2D projection: Let us assume that we observe a 3D scene with 3D objects. The optics of the camera system projects the 3D scene onto a 2D image plane.

Blurring: The optical Point-Spread-Function (PSF), together with the sensor PSF, will cause a blurring of the scene at the 2D plane. In this thesis, the optical blur is modeled by a Gaussian function with standard deviation σ_{psf}. It is considered independent of the depth of the scene, i.e. space-invariant. The sensor blur is modeled by a uniform rectangular function representing the fill-factor of

[1] In the remainder of this thesis 'SR reconstruction' refers to 'multi-frame SR reconstruction'.

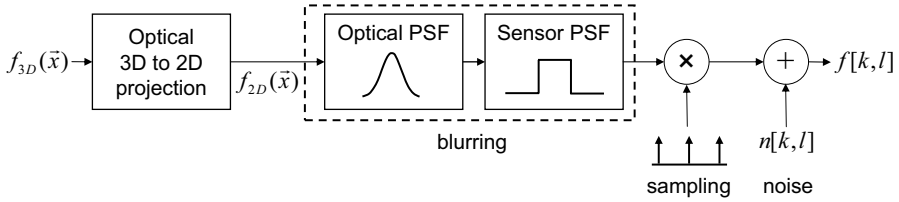


Figure 1.3: *Process of digital image formation: 3D to 2D projection, optical blur, sensor blur, sampling and additive noise.*

each sensor element. A convolution of both functions yields the total blurring function.

Sampling: The sensor characteristics of a camera system are determined by the pixel pitch as well as the fill-factor of the sensor elements. The sampling as depicted in Figure 1.3 relates to the pixel pitch. Likely, the recorded data by an infrared camera is under-sampled, which means that the sampling frequency is below two times the highest frequency of the continuous 2D scene $f_{2D}(\vec{x})$.

Noise: The noise in the recorded data is modeled by additive, independent and identically distributed Gaussian noise samples with standard deviation σ_n. Although state-of-the-art photon detectors (such as cameras) obey Poisson statistics [10] above the low light-level regime, where the readout noise dominates, we believe that independent additive Gaussian noise is a sufficiently accurate noise model. Other types of noise, such as fixed pattern noise and bad pixels, are not modeled explicitly in this thesis.

1.2.2 Optics

For optical systems that are circularly-symmetric, aberration-free and diffraction-limited, the PSF for incoherent illumination is given by the Airy disk. The minimum resolvable distance between two point sources is, according to the Rayleigh criterion, when the center of the Airy disk of the first point source coincides with the first minimum of the Airy disk of the second point source (see Figure 1.4). In an equation the minimum resolvable distance, Δl, can be defined as [1]:

$$\Delta l = 1.22 \frac{f\lambda}{D}, \tag{1.1}$$

with λ the wavelength of the incoming light, D the diameter of the lens aperture and f the focal length of the lens. The minimum resolvable distance can hence be decreased by decreasing the focal length and/or increasing the aperture

of the lens, which boils down to optics with a small F-number ($F = f/D$). However, the resolution of a camera system is not solely determined by its optics, but also by its sensor. Often the resolution is limited by the latter one.

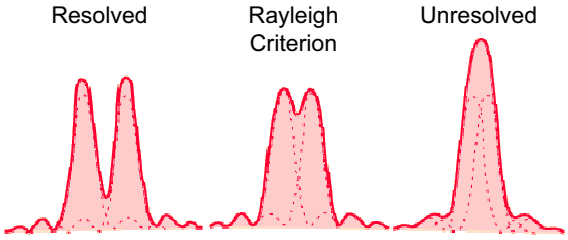


Figure 1.4: *The Rayleigh criterion: the center of the Airy disk of the first point source coincides with the first minimum of the Airy disk of the second point source.*

1.2.3 Sensor

Let us assume that the sensor is a Focal Plane Array (FPA), which is a 2D array of non-overlapping photosensitive elements. The spatial characteristics of a FPA are determined by the pixel pitch and the fill-factor of each sensor element. The fill-factor indicates the percentage of photosensitive area and the pixel pitch is the center-to-center distance of adjacent sensor elements.

Undersampling occurs if the sampling frequency (1/pixel pitch) is below the bandwidth of the image, often two times the highest frequency of the continuous 2D scene. The Nyquist-Shannon sampling theorem is not met and the imaged scene will be corrupted by aliasing. Some artifacts that occur by under-sampling in images are visualized in Figure 1.2 and are called *aliasing* artifacts [44]. Aliasing refers to the effect that causes different continuous signals to become indistinguishable after sampling.

Apart from the visible effects of under-sampling in images, the effects of aliasing can also be shown in the frequency domain (see Figure 1.5). Here, the camera's transfer function is modeled by the Modulation Transfer Function (MTF). We set the lens blur ($\sigma_{psf} = 0.3$) and the sensor blur (fill-factor = 81%) to realistic values that cause aliasing. Both factors are incorporated in the MTF. The scene spectrum is modeled with a quadratic decay, which is characteristic in natural images [54]. The non-aliased spectrum (before sampling) results after applying the MTF to the original spectrum.

However, if the scene is under-sampled, adjacent copies of the spectrum overlap. It looks like the spectral information above half the sampling frequency ($f_s/2$)

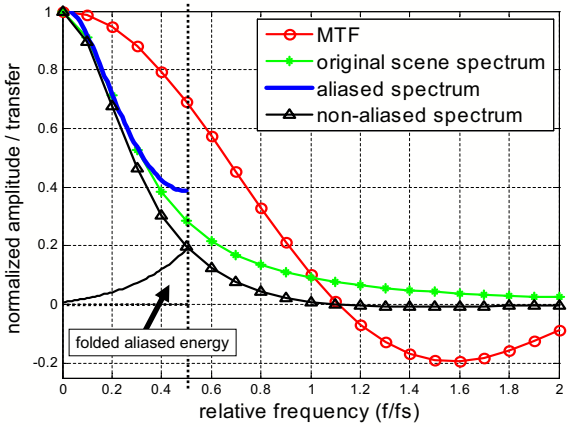


Figure 1.5: *The aliased spectrum (no marks) of a signal which is band-limited and under-sampled. The non-aliased spectrum (denoted with triangles) results after applying the MTF to the original spectrum. To obtain the aliased spectrum, the spectral energy above $= f_s/2$ needs to be folded (denoted with the arrow) and added to the non-aliased spectrum.*

is folded back and added to the part below $f_s/2$, because the central period of the periodic spectrum after sampling is limited to frequencies between $-f_s/2$ and $f_s/2$.

Increasing the number of sensor elements on a FPA by reducing the pixel pitch, while keeping the fill-factor constant, will reduce under-sampling, but has a negative effect on the SNR. Smaller sensor elements can capture less photons in an equal time interval and increasing the acquisition time, to compensate for this effect, will increase motion blur.

1.3 Multi-frame super-resolution reconstruction

Applying a multi-frame SR reconstruction method to an under-sampled image sequence increases the spatial sampling rate such that the aliased spectra are 'unfolded' and the spectrum, including the high frequencies, is recovered. SR reconstruction uses temporal information to improve the spatial resolution of the image sequence. Deblurring and an improvement in SNR can be obtained as well [55]. A spatial resolution improvement is only possible if uncorrelated subpixel motion between the imaged scene and the camera is present in the acquired image sequence. If a camera would be completely fixed and records a static scene,

no resolution improvement can be obtained for arbitrary image sequences since deblurring schemes cannot recover the information that was corrupted by aliasing. In this setup the camera acquires exactly the same information at each time instance, so there is no information gain of the scene over time. Therefore, uncorrelated subpixel motion between the scene and the camera is an important element to improve the spatial resolution of a recorded image sequence.

If a scene without moving objects is recorded with a moving camera (e.g. a camera on a moving platform), SR reconstruction can be applied to the whole scene. A typical algorithm/method for multi-frame SR reconstruction from a low-resolution (LR) image sequence involves three subtasks: *registration*, *fusion* and *deblurring* (see figure 1.6). Some SR reconstruction algorithms combine subtasks: e.g. Hardie's method [26] combines fusion and deblurring in a single step. First, the LR images are registered against a common reference with subpixel precision. During fusion an image at a higher resolution is constructed from the scattered input samples. Deblurring can be employed to (partially) correct for the optical and sensor blurring.

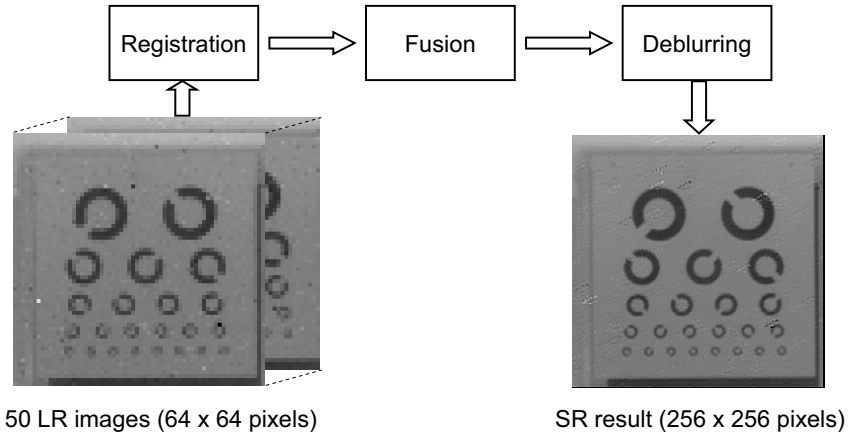


Figure 1.6: *A typical three-step solution for super-resolution reconstruction of a low-resolution image sequence: registration, fusion and deblurring. Here, Hardie's method is applied to 50 LR images with zoomfactor 4. Note the significant improvement in resolution and reduction of aliasing artifacts.*

1.3.1 Image registration

Image registration is the task of finding a geometric transformation between different views of the same scene. To be able to perform SR reconstruction, a subpixel precise registration is needed of the recorded image sequence. The effective resolution enhancement of a SR reconstruction algorithm is limited by 1) the camera characteristics and 2) the registration precision. The latter depends mainly on the characteristics of the captured data, such as the amount of gradient energy and the amount / type of noise [48]. If we know what the limitations are concerning image registration, we have an indication for an effective zoomfactor, which is the ratio of the sampling distance of the LR grid and the high-resolution (HR) grid. Note that the choice of an effective zoomfactor also depends on the amount of under-sampling and the number of available LR frames.

Registration of a recorded LR image sequence can be done in many different ways. However, two main approaches can be distinguished: feature-based methods and area-based methods. The main difference between both approaches is that the former one uses only a sparse set of feature points to fit the motion model, while the latter one uses all pixel information. For SR reconstruction often the area-based approach is used, because a better precision can be achieved.

In this thesis the most basic motion model is used: translation. Such a basic model suffices when a scene is captured from a (large) distance by a slowly moving camera. Small rotations perpendicular to the optical axis can also be modeled by translation. Rotations along the optical axis, however, cannot be described by a translation of the entire scene.

1.3.2 Super-resolution fusion

After image registration all LR samples are merged on a HR grid. This process is called super-resolution fusion [70]. Although many SR reconstruction methods [26], [43], [55], [72] combine fusion and deblurring, a decoupling of these subtasks reduces computational complexity [18], [21] and allows flexibility. However, this decoupling can only be done under the assumptions of rigid motion, common space-invariant blur and the same noise characteristics across all LR frames.

Almost 25 years ago, a first attempt of SR reconstruction was done by Tsai and Huang [59] by solving a set of equations that led to the Fourier coefficients of the sampled scene without aliasing. They assumed a set of uncorrupted shifted LR images. To allow other motion models than pure translation, most recent SR reconstruction methods are spatial-based. The first spatial-based SR reconstruction method was reported by Gross [23] in 1986.

Nowadays there is a tendency towards robust fusion techniques to cope with registration and intensity outliers [21], [72] or to deal with very few LR samples

[49]. In the latter work this is done by fusion of a spatiotonal adaptive neighborhood.

Other SR fusion techniques that are known from literature are interpolation-based. In [37] a Delaunay triangulation is used to perform an interpolation on a HR grid. However, a drawback of such interpolation based fusion techniques is the high computational complexity.

1.3.3 Image deblurring

The last step of a SR reconstruction method is image deblurring, also called deconvolution. The purpose of this step is to sharpen the HR fused image by undoing the blur caused by the lens and sensor. Let us assume that an image f is degraded with space invariant blur h and additive Gaussian noise n [36]:

$$g(\vec{x}) = (f * h)(\vec{x}) + n(\vec{x}), \tag{1.2}$$

with $*$ the convolution operator. Then deconvolution is the task of recovering image f from the degraded image g. One of the first solutions to this problem was given by Wiener [65], who minimized the Mean Squared Error (MSE) between the restored image $\hat{f}$ and the target image f.

However, (least-squares) deconvolution is an ill-posed problem which needs regularization to find a stable solution. A few regularization norms that penalize high-frequency oscillation in the restored image $\hat{f}$ are Tikhonov-Miller's quadratic norm [57], Rudin's Total Variation (TV) norm [52] and Farsiu's Bilateral Total Variation (BTV) norm [21].

An alternative to the deconvolution methods described above is to combine the deblurring step with the image fusion step. An example of a SR reconstruction method that combines those steps is Iterated Back Projection (IBP) [31], which applies least-squares deconvolution to multiple LR images to construct a HR image. Hardie et al. [25] proposed a Maximum A Posteriori (MAP) SR approach using Tikhonov-Miller regularization.

In this thesis we will not focus on deconvolution, but it will be used by several of the algorithms described. For example in Chapter 2 we show the effect of regularization on the perceived resolution of an image.

1.4 Moving objects in images

For many applications, such as surveillance, changes in the scene are of key interest. In this thesis we focus on moving objects, because they are responsible for changes in the observed scene. The way an object is represented on the image

plane is determined by the camera model discussed in section 1.2.1. In this thesis we assume objects to be rigid and that their shape and intensity is preserved. These assumptions permit us to process scenes in which the shape and intensity only change slowly over time.

The size of an object on the 2D image plane is determined by 1) the real-world object size, 2) the object's distance to the camera and 3) the focal length of the lens. To capture an object larger on the image plane, one can decrease the distance to the object and/or use a lens with a larger magnification (larger focal length).

If we talk about the size of an object in this thesis, we mean the number of pixels that the object covers in the image plane. By this definition the object size is also dependent on the sampling distance of the sensor. With a smaller sampling distance, the object will cover more pixels. Note that an object that is captured small on the image plane, may not be small in the real-world: a Boeing 747 is huge, but looks really small if it is observed from a large distance.

When moving objects are present in the scene or when a scene contains large depth-variations, the relative motion of the imaging sensor with respect to the scene becomes space-variant, i.e. it differs as a function of position in the image. Each moving object will have its own relative motion with respect to the camera. To be able to perform SR reconstruction on a moving object, the relative motion of the moving object with respect to the background has to be estimated by means of registration.

1.4.1 Large objects

If a moving object is large on the image plane (the total number of pixels depicting the object is large compared to the amount of object boundary pixels), SR reconstruction of a moving object can be applied in the same way as the rest of the scene after detection and segmentation of the moving object. This is described in Chapter 4 of this thesis. However, at the boundary of the moving object an error will be made because the pixels in that region contain both information of the scene's background and the moving object; such boundary pixels are called *mixed pixels*. For large moving objects these errors at the boundary are noticeable, but not of great importance for subsequent detection and recognition tasks.

1.4.2 Small objects

If small moving objects, i.e. objects of which the majority of pixels depicting the object are mixed pixels, are present in the scene, a standard approach for SR reconstruction will fail to reconstruct these objects. In this case, the mixed pixels will have to be processed differently to separate the mixed foreground (object)

and background information. A technique for solving this problem is presented in this thesis. The basic idea is that for each LR pixel in the recorded sequence the contributions of the background and foreground are estimated by describing the object boundary with a subpixel precise polygon.

1.4.3 Point objects

A point object is the smallest possible object and it occurs after blurring by the camera as a blurred point in the image plane. Depending on the blurring and sampling the point object's energy is likely to be spread among more than one pixel. Although it makes no sense to perform SR reconstruction on point objects, their detection can be improved by applying SR reconstruction on the background of the same scene. This is described in this thesis as well.

1.5 Research questions

In section 1.1 it was already stated that the main goal of this thesis is to improve the resolution, SNR and contrast of moving objects in under-sampled image sequences by means of SR reconstruction and that these improvements may help to increase the recognition rate of moving objects.

To reach this main goal, several research questions were formed before and during my PhD project. In this section all these research questions are described and most of them are answered in this thesis.

One of the first questions that arises when one is developing an algorithm to improve images is "How am I going to measure the performance of my algorithm?". Ideally we would like to have an quantitative and objective measure. Such a measure makes it easy to compare different algorithms. But what is a good quantitative and objective measure? And is it task specific or generic? Furthermore it is interesting to know if it is possible to predict the performance of SR reconstruction algorithms on real-world data by testing the performance on controllable simulated data. In **Chapter 2**, **Performance evaluation of SR reconstruction methods on real-world data**, we address these issues.

When we have found an answer to our first research question, we can focus again on our goal concerning moving objects. First we zoom in on the smallest possible objects: the point objects, also called point targets. We already stated that it makes no sense to perform SR reconstruction on point targets, so why addressing them in this thesis?

A major topic concerning point targets is their detection. Typically, point targets need to be detected in an early stage. However, if an image sequence is under-sampled and contains a lot of structure in the background, it is difficult to detect a point target. So an interesting research question is: "How can SR

reconstruction improve the detection of point targets?". This question is answered in **Chapter 3**, **SR reconstruction for moving point target detection**.

When moving objects are large, a 'standard' pixel-based SR reconstruction method can be applied to the pixels comprising the moving object. But "How do we apply simultaneously SR reconstruction to the object and the background?", "What is the minimum object size for this kind of approach?" and "How do we process the boundary region between object and background?" are typical questions that we would like to see answered. Also we would like to know how the performance of SR reconstruction on a moving object compares with the performance on the background. Answers to these questions can be found in **Chapter 4**, **SR reconstruction on large moving objects and background**.

Performing SR reconstruction on small moving objects is a hard and challenging problem, because the majority of pixels contained by the object are mixed (boundary) pixels. We have to find a way to separate the foreground and background information in each observed mixed pixel. "How can this be done?" is the main research question here. Furthermore, we have to find a way to register the moving object with high precision to be able to perform SR reconstruction in the first place. All in all there are a lot of challenges concerning SR reconstruction of small moving objects and they are tackled in **Chapter 5**, **SR reconstruction of small moving objects in simulated data**, and **Chapter 6**, **SR reconstruction of small moving objects in real-world data**.

Chapter 2

Performance evaluation of SR reconstruction methods on real-world data

Abstract

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 chapter 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 a 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]This chapter has been published in A.W.M. van Eekeren, K. Schutte, O.R. Oudegeest and L.J. van Vliet, Performance evaluation of super-resolution reconstruction methods on real-world data, *EURASIP Journal on Advances in Signal Processing*, 2007, Article ID 43953. [14]

2.1 Introduction

During the last decade numerous Super-Resolution (SR) reconstruction methods have been reported in the literature. Reviews can be found in [45, 20]. SR reconstruction is the process of combining a set of under-sampled (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 sub-tasks: registration and fusion. Occasionally an additional deblurring step is performed afterwards. First, the LR images are registered against a common reference with sub-pixel 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 work is reported in [4, 39]. Both study the problem of SR from an algebraic point of view. Robinson [51] 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 sub-task level of an SR reconstruction method.

This chapter 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) [7, 6]. 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 chapter 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 chapter is organized as follows. In Section 2.2, the registration of the real-world and simulated data is discussed. In Section 2.3, the different SR reconstruction methods are discussed. In Section 2.4, the TOD method is explained and the setup of the measurements is given. The results are presented in Section 2.5 and finally conclusions will be provided in Section 2.6.

2.2 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 [48]. A gradient-based shift estimator [40] finds the displacement $t_{\vec{x}}$ between two shifted signals as the least squares solution of (2.1)

$$MSE = \frac{1}{N} \sum_{R} \left(s_2(\vec{x}) - s_1(\vec{x}) - t_{\vec{x}} \frac{\partial s_1}{\partial \vec{x}} \right)^2 \tag{2.1}$$

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 (2.1) is biased, which is corrected in an iterative way. In the first iteration s_2 is shifted with the estimated sub-pixel 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 ($\sigma_{disp} \approx 0.01$ pixel for noise free data) unbiased registration, which approaches the Cramér-Rao bound [34].

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.

2.3 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.

2.3.1 Elad's Shift & Add method

After registration of all LR frames, Elad's [18] 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 & Add method is only a fusion method and does not incorporate deblurring.

2.3.2 Lertrattanapanich's triangulation-based method

In [37] 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 (2.2):

$$z(x,y) = c_1 + c_2x + c_3y + c_4x^2 + c_5y^2 + c_6x^3 + c_7x^2y + c_8xy^2 + c_9y^3. \quad (2.2)$$

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 direction. Lertrattanapanich's triangulation-based method performs fusion only.

2.3.3 Kaltenbacher's least-squares method without regularization

This method [33] is based on the idea of estimating the "underlying" unaliased frequency spectrum from multiple, aliased spectra. For sake of clarity, the 1-D case will be explained below. With the shift property, the Fourier transform F_i of a shifted frame i before sampling is

$$F_i(\omega) = F(\omega)e^{j\delta_i\omega}, \quad (2.3)$$

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 (2.3) converts to:

$$\tilde{F}_i(n) = \frac{1}{S}\sum_{m=-\infty}^{\infty} F_i\left(\frac{2\pi}{NS}n - m\omega_s\right). \quad (2.4)$$

Here, $\tilde{F}_i(n)$ is the discrete Fourier transform of LR input frame $i = 1, ..., P$. S is the sampling period and $\omega_s = 2\pi/S$ the sampling frequency, N is the amount of samples per LR frame and $n = 1, .., N$ 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 (2.4) 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:

$$\mathbf{G}_n = \mathbf{\Phi}_n\mathbf{F}_n, \quad (2.5)$$

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

$$\mathbf{G}_n(i) = \tilde{F}_i(n), \tag{2.6}$$

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

$$\mathbf{\Phi}_n(i,k) = e^{j2\pi\delta_i(\frac{n}{N}+(\lfloor K/2 \rfloor - k))}. \tag{2.7}$$

$\mathbf{F}_n$ 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 Φ_n

$$\mathbf{F}_n = (\mathbf{\Phi}_n^T \mathbf{\Phi}_n)^{-1} \mathbf{\Phi}_n^T \mathbf{G}_n \tag{2.8}$$

2.3.4 Hardie's method using a regularized inverse observation model

Hardie [26] employs a discrete observation model that relates the ideally sampled image $\mathbf{z}$ and the observed frames $\mathbf{y}$:

$$y_m = \sum_{r=1}^{H} w_{m,r} z_r + \eta_m \tag{2.9}$$

where $w_{m,r}$ represents the contribution of the r^{th} HR pixel in $\mathbf{z}$ to the m^{th} 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:

$$C_{\mathbf{z}} = \sum_{m=1}^{L} \left(y_m - \sum_{r=1}^{H} w_{m,r} z_r \right)^2 + \lambda \sum_{i=1}^{H} \left(\sum_{j=1}^{H} \alpha_{i,j} z_j \right)^2 \tag{2.10}$$

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

The cost function in (2.10) balances two types of errors. The left term is minimized when a candidate $\mathbf{z}$, projected through the observation model (2.9), 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 $\alpha_{i,j}$ (2.11) are selected to perform a Laplacian operation on $\mathbf{z}$ and ensure that the regularization term is minimized when $\mathbf{z}$ is smooth.

(2.11)

2.3.5 Farsiu's robust method

In comparison with Hardie's method, the reconstruction method proposed by Farsiu et al. [21] separates the fusion and deblurring processes of an SR reconstruction method: 1) the LR frames are fused with median Shift & Add (similar as described in Section 2.3.1, 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 (2.12).

$$C_{\mathbf{z}} = \|A(G\mathbf{z} - \mathbf{z_0})\|_1 + \lambda \sum_{l=0}^{P} \sum_{m=0}^{P} \alpha^{m+l} \|\mathbf{z} - S_h^l S_v^m \mathbf{z}\|_1 \tag{2.12}$$

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 is based on the bilateral Total Variation (TV) criterion [21]. 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.

2.3.6 Pham's structure-adaptive and robust method

Pham [49] recently proposed an SR reconstruction method using adaptive Normalized Convolution (NC). NC [35] is a technique for local signal modeling from projections onto a set of basis functions. Pham uses a first-order polynomial basis (2.13):

$$\hat{f}(\mathbf{s}, \mathbf{s}_0) = p_0(\mathbf{s}_0) + p_1(\mathbf{s}_0)x + p_2(\mathbf{s}_0)y, \tag{2.13}$$

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 [24], 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 $\mathbf{s}_0$:

$$\varepsilon(\mathbf{s}_0) = \int (f(\mathbf{s}) - \hat{f}(\mathbf{s}, \mathbf{s}_0))^2 c(\mathbf{s}) a(\mathbf{s} - \mathbf{s}_0) d\mathbf{s}, \tag{2.14}$$

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 2.1.

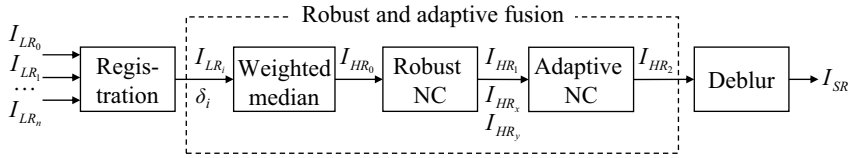


Figure 2.1: *Flow diagram of Pham's structure-adaptive and robust SR reconstruction method.*

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 (2.15) is performed, which results in a better estimate I_{HR_1} and two corresponding derivatives I_{HR_x} and I_{HR_y}.

$$c(\mathbf{s}, \mathbf{s}_0) = \exp\left(-\frac{|f(\mathbf{s}) - \hat{f}(\mathbf{s}, \mathbf{s}_0)|^2}{2\sigma_r^2}\right) \tag{2.15}$$

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).

2.4 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 [7]. The TOD method determines the smallest triangle size in an image of which the orientation can be discriminated. This evaluation method is preferred 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 [32]. 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 non-linear algorithms, which most SR reconstruction methods are.

2.4.1 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 2.2).

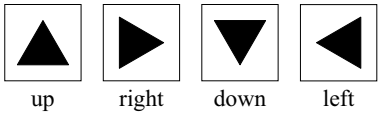


Figure 2.2: *The four different stimuli used in the TOD method.*

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:

$$\hat{\theta} = \min_{\theta,s}\{\frac{1}{N}\sum_{\vec{x}}(I_{HR}(\vec{x};\theta_f,s_f) - M(\vec{x};\theta,s))^2\}. \tag{2.16}$$

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 (2.16) 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 [7] it is shown that this increase can be described with a Weibull distribution:

$$p_c(x) = 0.25 + 0.75/1.5^{(\alpha/x)^\beta}, \tag{2.17}$$

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 2.3. From this fit the triangle size that corresponds with a 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, e.g. SNR, T_{75}'s are determined, a performance curve can be plotted. Such curves will be used in Section 2.5 to show the results.

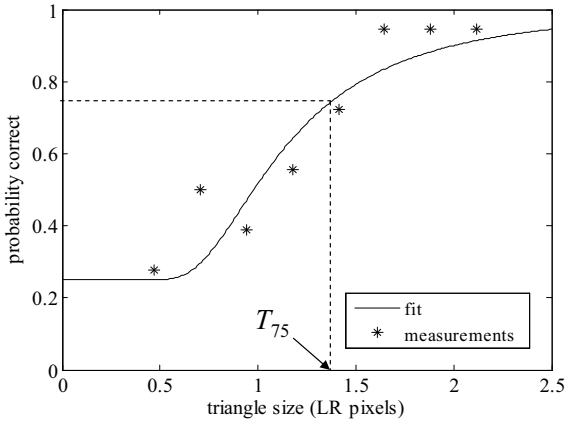


Figure 2.3: *Example of a possible Weibull distribution of probability correct observer response.*

2.4.2 Real-world data experiment

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

Setup

The setup of the experiment (including TOD) is depicted in Figure 2.4. 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 [61]) is present as depicted in the left part of Figure 2.4. 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 (SNR's) 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 (2.16) 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, ..., 280 pixels. In our evaluation each triangle is equidistantly shifted, blurred ($\sigma = 0.9 \times S$) and 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 [62]. 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.

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 SNR's. Here, the SNR (dB) is defined as:

$$SNR = 20 \log_{10} \left(\frac{|I_{TR} - I_{BG}|}{\sigma_{BG}} \right), \tag{2.18}$$

with I_{TR} 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 SNR's: 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 2.5, 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% (p.101 in [47]). Given a $\sigma_{tot} = 0.5$ the blurring of the lens is $\sigma_{psf} = 0.4$.

2.4.3 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 reconstruc-

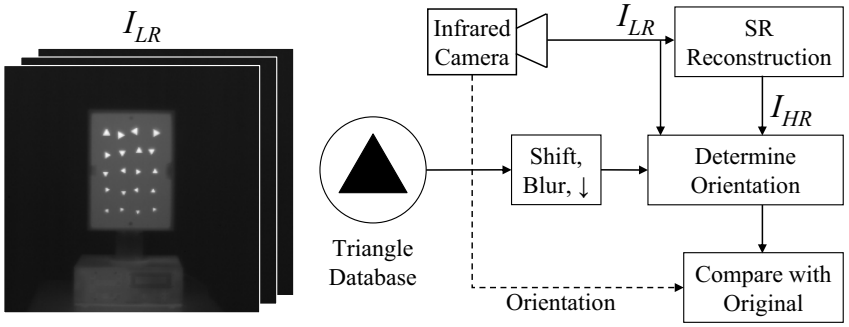


Figure 2.4: *Left: example of real-world data I_{LR}. Right: flow diagram of the real-world data experiment.*

tion methods as in the previous experiment an indication can be obtained of the predictability of the real-world performance of these algorithms.

Camera model

A data set is simulated with a camera model as depicted in Figure 2.5. Where I_{HYP_i} is a discrete representation of a scene sampled at the Nyquist rate with a $S\times$ 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 is modeled with a uniform filter U with width $S \cdot \sqrt{\text{ff}}$. The overall noise in the camera model is assumed to be Gaussian distributed.

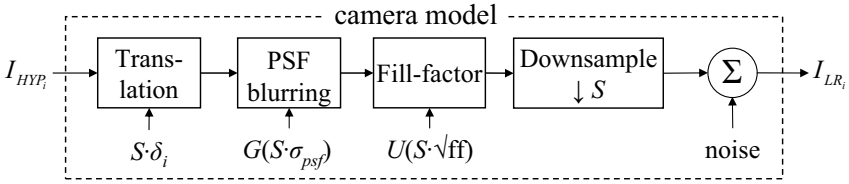


Figure 2.5: *Camera model used in the experiments.*

In this experiment two simulated data sets I_{LR} are generated: 1) $\sigma_{psf} = 0.3$, $\text{ff} = 0.8$, which results in a less blurred data set as derived in Section 2.4.2 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 sub-pixel shifts in the simulated data. Different amounts of Gaussian noise are added, resulting in a SNR varying from 12 dB to 42 dB.

Setup

The setup of the experiment on simulated data is depicted in Figure 2.6. 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 part of Figure 2.6, 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 (2.16). Note that for this experiment no gain and offset estimation is used in the triangle model M.

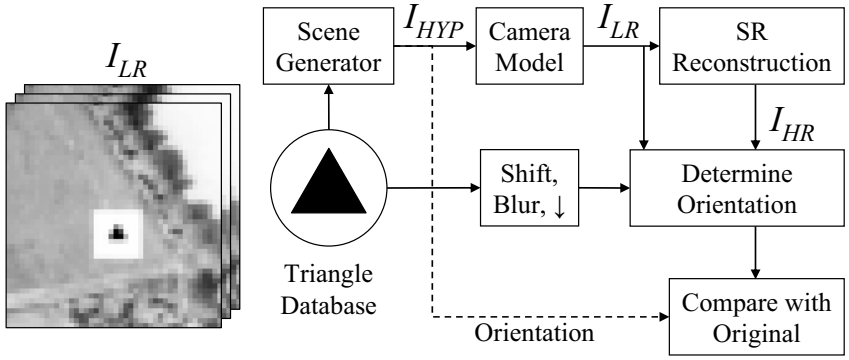


Figure 2.6: *Left: example of simulated data I_{LR}. Right: flow diagram of the simulated data experiment.*

2.4.4 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 (cam-

era's fill-factor, optical PSF, SNR). Here, camera model parameters ($\sigma_{psf} = 0.2$, ff $= 1$) are chosen that result in a more aliased data set than the previous simulated data sets. 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 2.6). The performance of the SR reconstruction methods is measured for the following conditions:

- Different number of frames
- Different SNRs
- 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 method 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.

2.4.5 TOD versus MSE

An alternative measure to TOD is the MSE:

$$MSE = \frac{1}{N} \sum_{\vec{x}} (I_{HR}(\vec{x}; \theta_f, s_f) - M(\vec{x}; \theta_f, s_f))^2. \tag{2.19}$$

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 2.7.

It is clear from Figure 2.7 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

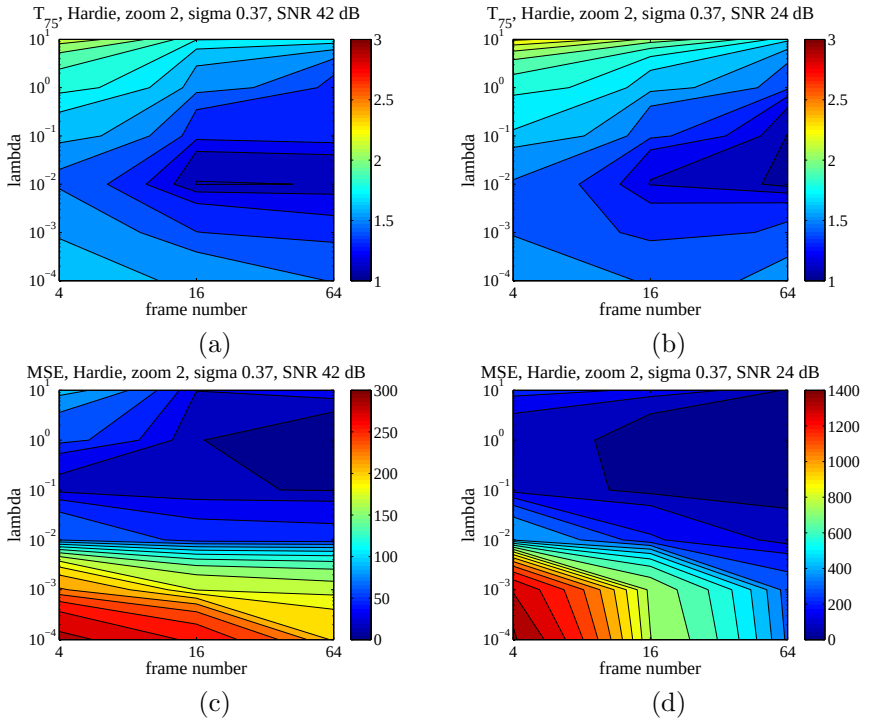


Figure 2.7: *(a) Contour plot T_{75}, SNR 42 dB, (b) Contour plot T_{75}, SNR 24 dB, (c) Contour plot MSE, SNR 42 dB, (d) Contour plot MSE, SNR 24 dB.*

than the 'optimal' λ resulting in the lowest MSE: 10^{-2} and 1, respectively. The corresponding SR results (not depicted in this chapter) 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.

2.5 Results

All results of the experiments can be found at the end of this chapter. 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.

2.5.1 Results real-world and simulated data experiment 1

The results of the "Real-world data experiment" and the "Simulated data experiment 1" can be seen in Figure 2.8. 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 ($\sigma_{psf} = 0.55$ and ff $= 0.8$, resulting in an "overall" $\sigma_{tot} \approx 0.6$) and a *performance* upper bound ($\sigma_{psf} = 0.3$ and ff $= 0.8$, resulting in $\sigma_{tot} \approx 0.4$) on the real-world performance. Note that in Figure 2.8 the *performance* upper bound is visually a lower bound and the *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 method perform slightly better.

2.5.2 Results simulated data experiment 2

In Figure 2.9 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 performance 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 non-evenly 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.

To illustrate the effect of an increasing zoom-factor, Figure 2.10 shows performance curves of all SR reconstruction methods for zoom-factor 1, 2 and 4. All methods processed the same 64 LR frames ($\sigma_{psf} = 0.2$ and ff $= 100\%$). From Figure 2.10 it is clear that the performance of zoom-factor 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-factor 2 & 4) is not obtained for low SNRs. Here, the temporal noise reduction is more relevant than the anti-aliasing. 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.

2.6 Conclusions

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

- 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.
- 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.
- 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.
- Comparing the performance of the unregularized Kaltenbacher's method with the regularized methods of Hardie, Farsiu and Pham (Figure 2.9), it can be concluded that in general regularization is not required for good performance when many input frames are available.
- The relative performance of the various methods change little as a function of SNR.
- The results presented in Figure 2.10 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.

2.7 Acknowledgment

The authors would like to thank T.Q. Pham for the implementation of several of the used SR reconstruction methods and thank P. Bijl for providing the infrared data.

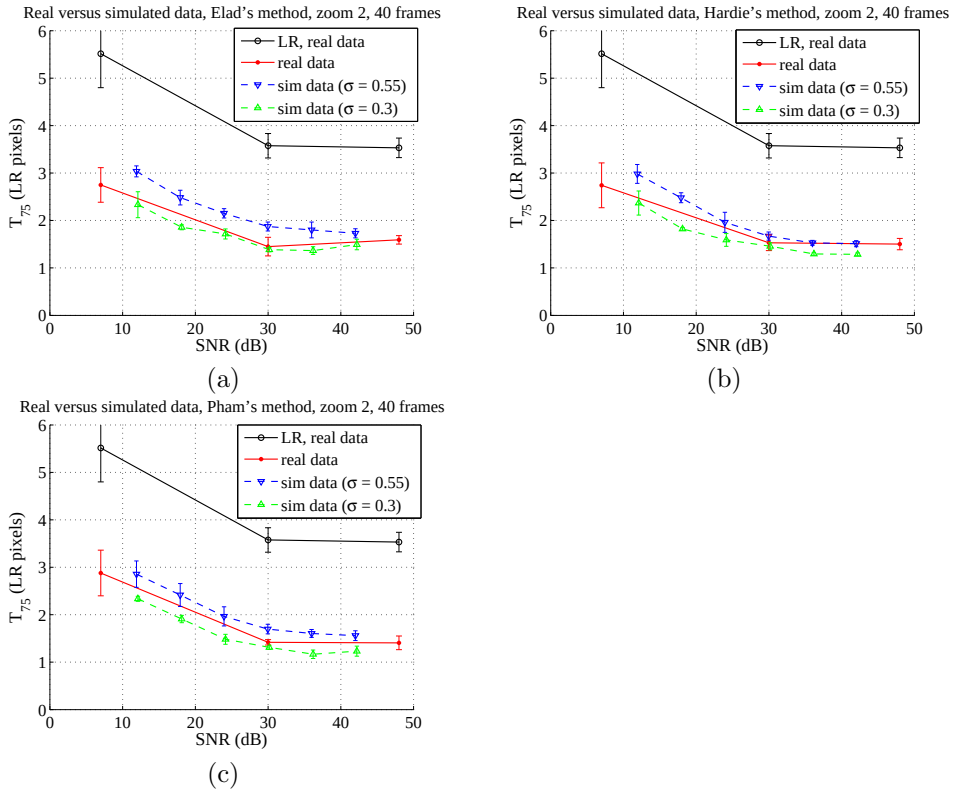


Figure 2.8: *Performance measurements on real-world and simulated data (40 frames). Blue line: sim. data created with $\sigma_{psf} = 0.55$ & ff=80%, green line: sim. data created with $\sigma_{psf} = 0.3$ & ff=80%. (a) Elad, (b) Hardie ($\sigma = 0.55$, $\lambda = 0.01$), (c) Pham ($\sigma = 1$, $\lambda = 10^{-3}$, $\beta = 10$). All data is processed with zoom-factor 2.*

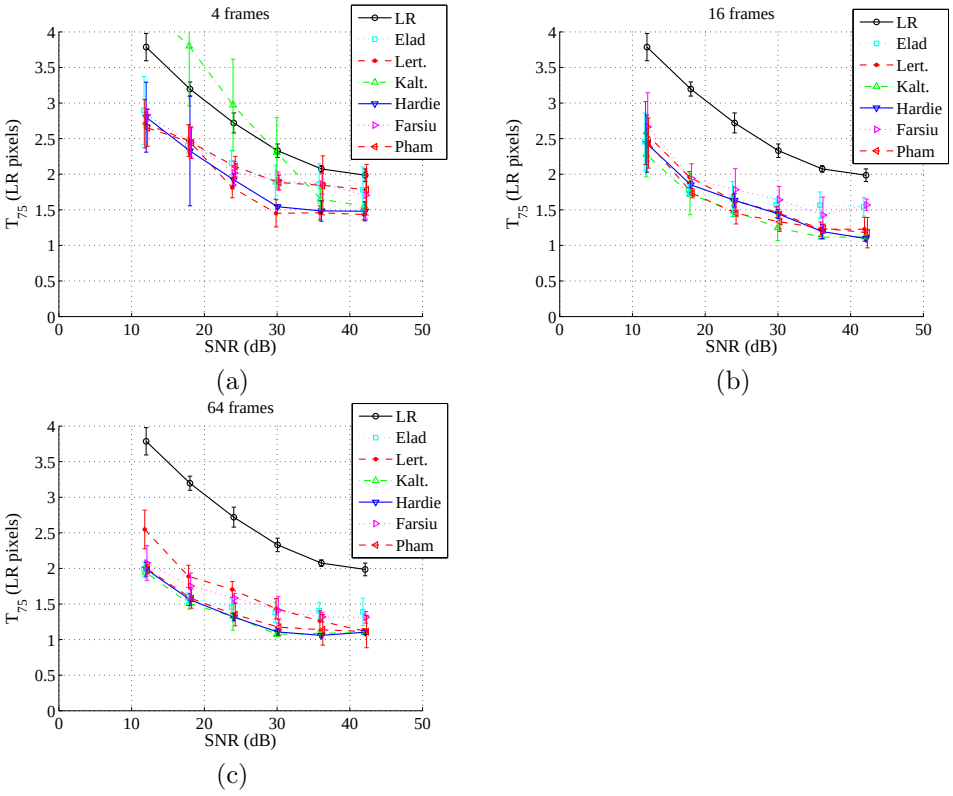


Figure 2.9: *Performance measurements on simulated LR data (σ_{psf} = 0.2, ff = 100%) processed with different SR reconstruction methods (zoom-factor 2) with optimized settings, (a) 4 frames, (b) 16 frames, (c) 64 frames.*

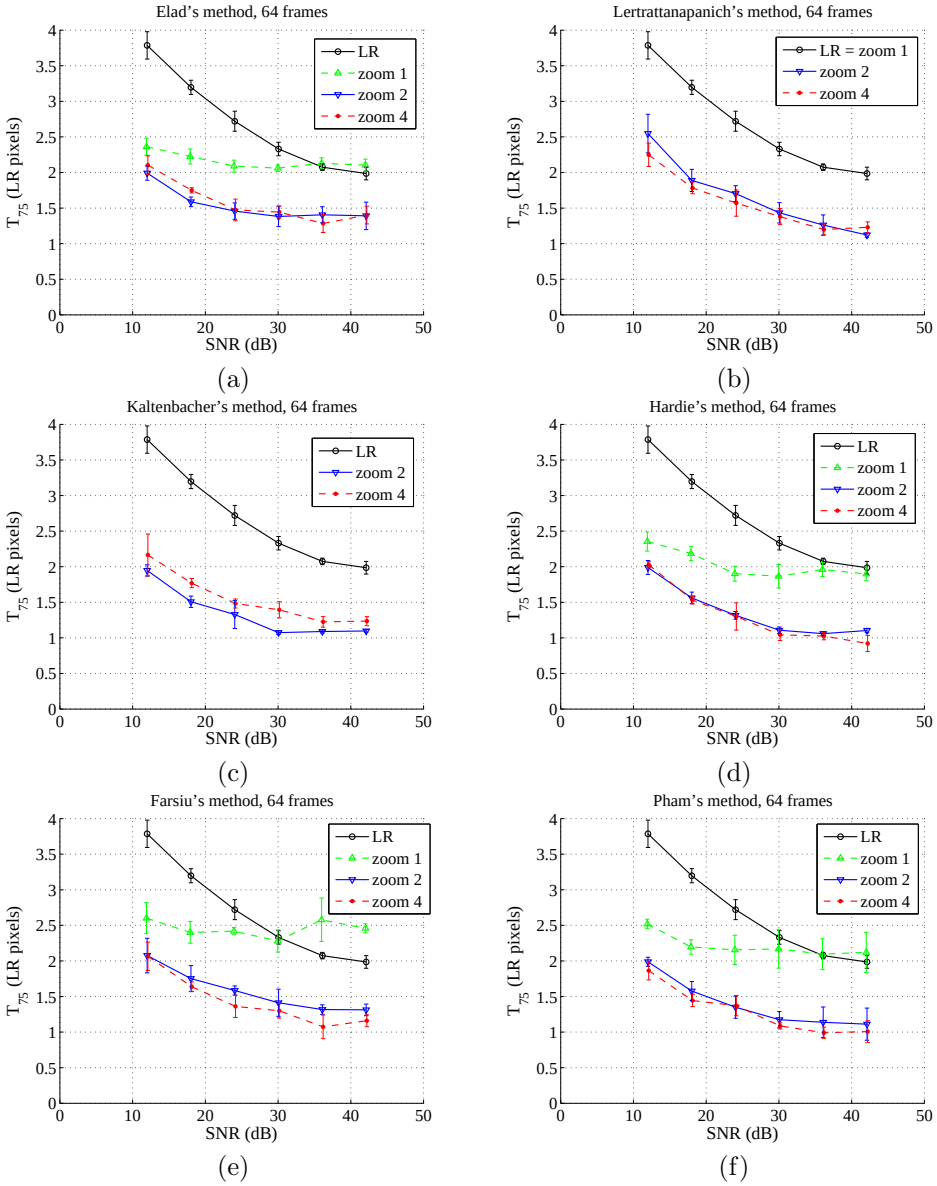


Figure 2.10: *Performance measurements on simulated LR data ($\sigma_{psf} = 0.2$, ff = 100%, 64 frames), processed with different methods with optimized settings for zoom-factor 1, 2 and 4. (a) Elad, (b) Lertrattanapanich, (c) Kaltenbacher (no zoom-factor 1 results could be obtained with our implementation), (d) Hardie, (e) Farsiu, (f) Pham.*

Chapter 3

Super-resolution reconstruction for moving point target detection

Abstract

When bright moving objects are viewed with an electro-optical system at long range, they will appear as small slightly blurred moving points in the recorded image sequence. Typically, such point targets need to be detected in an early stage. However, in some scenarios the background of a scene may contain much structure, which makes it difficult to detect a point target.

The novelty of this work is that super-resolution reconstruction is used for suppression of the background. With super-resolution reconstruction a high-resolution estimate of the background, without aliasing artifacts due to under-sampling, is obtained. After applying a camera model and subtraction, this will result in difference images containing only the point target and temporal noise.

In our experiments, based on realistic scenarios, the detection performance, after background suppression using super-resolution reconstruction, is compared to the detection performance of a common background suppression method. It is shown that using the proposed method, for an equal detection - false alarm ratio, the signal strength of a point target can be up to 4 times smaller. This implies that a point target can be detected at a longer range.

[1]Parts of this research are described in European patent application nr. 06077053.4: point target detection with super-resolution.

[2]This chapter has been published in J. Dijk, A.W.M. van Eekeren, K. Schutte, D.J.J. de Lange and L.J. van Vliet, Super-resolution reconstruction for moving point target detection, *Optical Engineering*, vol. 47, no. 8, 2008. [11]

3.1 Introduction

In surveillance applications moving targets need to be detected at a very early stage. Electro-optical surveillance systems observe 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. Furthermore, the background may also contain structure (clutter) of high contrast.

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.

One of the essential steps for background suppression is to determine the apparent motion between the frames, i.e. the registration step. The apparent motion of the background can, on a small scale, often be described by translational motion between two subsequent camera frames I_k and I_{k-1}.

A standard way of performing background suppression is by Shift, Interpolate and Subtract (SIS). One of the frames is corrected for the shift (dx, dy) using interpolation ($\tilde{I}_{k-1}$) and is subtracted from the other frame. In the experiments in this chapter we use bspline interpolation. After subtraction a difference image ΔD_k^{SIS} results:

$$\Delta D_k^{SIS}(x, y) = I_k(x, y) - \tilde{I}_{k-1}(x + dx, y + dy). \tag{3.1}$$

Note that for point targets with a small apparent motion with respect to the background, the point target's signal energy in the difference image ΔD_k^{SIS} is almost lost. Another problem of SIS is that due to under-sampling by the image sensor, aliasing artifacts in the recorded image sequence remain present in the difference image. Both will hamper point target detection.

In this chapter we propose to use super-resolution (SR) reconstruction to improve the detection of moving point targets. The SR reconstruction algorithm is used in the background suppression step. In previous work [49, 55] we developed Super-Resolution (SR) reconstruction techniques to improve the spatial resolution of under-sampled image sequences by exploiting the subpixel shift between the frames. Using SR for point target detection has the advantage that 1) the signal and aliasing contribution in the last frame can be predicted, which substantially reduces the aliasing related clutter in the difference image, 2) the temporal noise is reduced, which improves the Amplitude-to-Noise Ratio (ANR) of the point target in the difference image, and 3) the ANR in the background estimate is suppressed, which increases the ANR in the difference image. The latter is especially noticeable for point targets with a small apparent motion with respect to the background. Note that point targets which have *no* motion with respect to the background will be totally part of the background estimate. Therefore, they are not visible in the difference image.

After SIS or the SR background suppression step, standard detection algorithms such as thresholding or track-before-detect [67, 58] can be used. In this chapter results are shown for a 3-out-of-5 tracking algorithm [8] and for direct thresholding.

This chapter is organized as follows. In the next section the advantages of SR reconstruction for point target detection are discussed from a theoretical perspective. In section 3.3 the SR based point target detection method is presented. In section 3.4 the setup of these experiments is described. The experimental results are shown in section 3.5. Finally, conclusions are presented in section 3.6.

3.2 Theory

Super-Resolution (SR) reconstruction is a well-known technique to increase the spatial resolution of a sequence of aliased Low-Resolution (LR) images using temporal information. The zoom factor 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. Numerous SR reconstruction methods are described in the literature. Overviews are given by Park [45], Farsiu [21] and Van Eekeren [14]. Generally, SR reconstruction can be split up in three parts [48]: 1) registration, 2) fusion and 3) deblurring. The first part is necessary to align the content of all frames with subpixel accuracy. The two following steps will fuse the aligned data on a HR grid and deblur the result.

SR reconstruction can be used for point target detection to improve the background suppression step. With SR reconstruction it is possible to create a HR model of the background. This HR background model contains less or no aliasing and ideally does not contain the point target. With the HR background model $\vec{Z}$ (reordered in a vector) and a transfer matrix H_k, an estimate can be made of LR image $\vec{I}_k$ (reordered in a vector). The transfer matrix H_k describes 1) the model of the camera, 2) the estimated motion between $\vec{Z}$ and $\vec{I}_k$ and 3) the zoom factor. A difference image of frame k is then created by applying:

$$\Delta\vec{D}_k^{SR} = \vec{I}_k - H_k\vec{Z}. \tag{3.2}$$

$H_k\vec{Z}$ suffers from aliasing exactly the same way as $\vec{I}_k$, so the difference image is free from clutter due to aliasing. If the images were subtracted in the high-resolution space, this would not be the case, as the background image $\vec{Z}$ is aliasing-free and the HR version of $\vec{I}_k$ contains the interpolated aliasing.

In the next subsections the advantages of SR reconstruction for point target detection will be explained from a theoretical perspective.

3.2.1 Aliasing noise reduction

In a camera system the measured signal is limited by 1) the band-limitation of the optics and 2) the sampling of the sensor. Aliasing is an effect due to under-sampling. Both limitations are depicted in Figure 3.1.

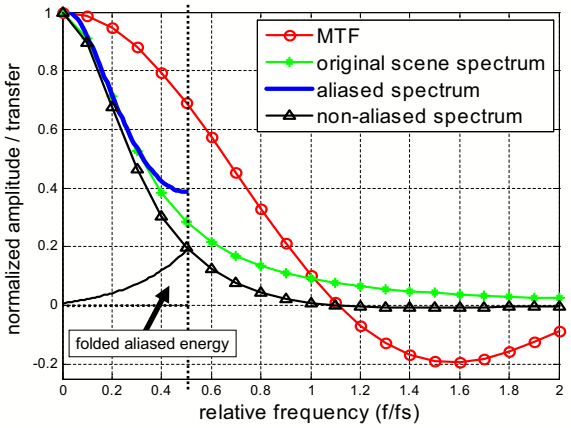


Figure 3.1: *The aliased spectrum (no marks) of a signal which is band-limited and under-sampled. The Modulation Transfer Function (MTF) is modeled by Gaussian lens blur ($\sigma_{psf} = 0.3$) and uniform sensor blur (area fill-factor = 81%). The non-aliased spectrum (denoted with triangles) results after applying the MTF to the original spectrum. To obtain the aliased spectrum, the spectral energy above the half-sampling frequency = $f_s/2$ needs to be folded (denoted with the arrow) and added to the non-aliased spectrum. Note that the aliased spectrum does not have any information above the half-sampling frequency. All spectra are normalized such that the DC value equals one.*

Here, the blurring of the lens is modeled with Gaussian blurring ($\sigma_{psf} = 0.3$) and the sensor is modeled as a 2-D array of non-overlapping square photosensitive elements with fill-factor = 81%. The scene spectrum is modeled with a quadratic decay, which is characteristic in natural images [54]. Note that this might slightly differ from the real scene spectrum of the images used in the simulations of which no spectrum was determined.

Applying SR reconstruction increases the sampling rate such that the aliased frequency spectra are unfolded and part of the high-frequency spectrum is recovered. This implies that a better, i.e. (almost) aliasing free, HR estimate of the background is obtained. Applying the camera model for frame k to this HR background image yields the same aliasing artifacts as in the recorded image I_k.

Subtraction of these two images is very effective in suppressing the background, because by sampling the HR image at exactly the same grid positions (including sub-pixel shift) as the corresponding LR image, exactly the same aliasing artifacts for frame k are created. After subtraction, the difference image will contain only temporal noise and the point target signal. Note that the main aliasing effect on the point targets is that their maximum energy per frame is not constant.

3.2.2 Temporal noise reduction

All cameras inadvertently add temporal noise to the scene information. Let us assume that there are N recorded frames available, containing additive Gaussian distributed noise with standard deviation σ_n. The resulting noise in a difference image after SIS (3.1) will be $\sqrt{2}\sigma_n$. Note that we assume that the images are corrected for non-uniformity correction. This can be done by the camera or based on the images. In this chapter, we do not evaluate the effects of non-uniformity reduction by SR on the detection of point targets.

Now, assume that SR reconstruction is used to calculate a difference image as in (3.2). Here, the estimated LR frame $H_k Z$ is based on N recorded frames, which reduces the noise standard deviation with factor $\sqrt{N}$. Therefore, the resulting noise in a difference image after SR reconstruction is:

$$\sigma_n^{\Delta SR} = \sqrt{\frac{N+1}{N}}\sigma_n. \tag{3.3}$$

If many frames are used, the noise in the resulting difference image will be only slightly higher than σ_n, the noise in a single LR image. The fraction of noise in the SR difference image compared to the SIS difference image is

$$\frac{\sigma_n^{\Delta SR}}{\sigma_n^{\Delta SIS}} = \frac{\sqrt{\frac{N+1}{N}}\sigma_n}{\sqrt{2}\sigma_n} \approx \frac{1}{\sqrt{2}}. \tag{3.4}$$

This means that for the same amount of false detections, the point target amplitude that can be detected with SR reconstruction in a temporal noise limited situation will be a factor $\sqrt{2}$ lower.

3.2.3 Point target amplitude preservation

Another advantage of background suppression using SR reconstruction is that the point target intensity in the difference image is preserved for large and small apparent motion of the point target with respect to the background. 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. The difference image is calculated with (3.2). In the non-ideal case, however, the point target can be visible in the projection of HR image Z. Note that the main aliasing effect on the point targets is that their maximum energy per frame is not constant.

To analyze the point target amplitude preservation, point targets with amplitude one are simulated. First, point targets are placed in a super scale image with a constant background, which is a factor 15 larger than LR image I_k. This super scale image is shifted and blurred with the MTF as described in Figure 3.1 and afterwards subsampled with factor 15. The simulation method is fully explained in section 3.4.3. To the resulting LR images a small amount of Gaussian distributed noise ($\sigma_n = 0.002$) is added.

The effect of point target amplitude preservation is measured as the maximum of the difference images ΔD_k^{SR}. If this maximum is around one, the point target amplitude is well preserved. The point target intensity in the difference image is simulated for two different SR reconstruction methods: Hardie (non-robust) and Zomet (robust). Both methods model the camera blur with Gaussian blur ($\sigma_{cam} = 0.41$) and use 48 frames for reconstruction. A more detailed explanation of both methods is given in section 3.3.2.

As a comparison the point target amplitude preservation of SIS for varying point target motion is simulated as well. Here, the difference images ΔD_k^{SIS} are calculated with (3.1).

The results are shown in Figure 3.2. As expected, the robust Zomet method (Z1 and Z2) performs best for point targets with a small apparent motion with respect to the background. There is no significant effect between the different zoom factors. Note that the robust Zomet method preserves the point target amplitude better than the non-robust Hardie method. This can be explained by the fact that the point target is treated as an outlier in Zomet's SR reconstruction. The difference between robust and non-robust SR is explained in more detail in section 3.3.2. If the point target velocity is large (>1.5 pixels), the point target profile of the previous recorded frame will hardly influence the point target profile in the current frame. For those cases the point target amplitude in the difference image is maximal. The high maximum value in the difference image after SIS for a high point target velocity can be explained by the bspline interpolation which is used for the shift. This interpolation can cause lobes which are below the background and add up to the point target in the non-interpolated frame.

Summarizing, using SR reconstruction for background suppression has the following advantages from a theoretical perspective: 1) aliasing artifacts are reduced, 2) temporal noise is reduced and 3) the point target is better preserved in the difference image for small apparent motion with respect to the background. Therefore, the largest gain of using SR reconstruction for background suppression is expected for recorded sequences with much structure in the background (causing significant aliasing artifacts) and a small apparent point target motion.

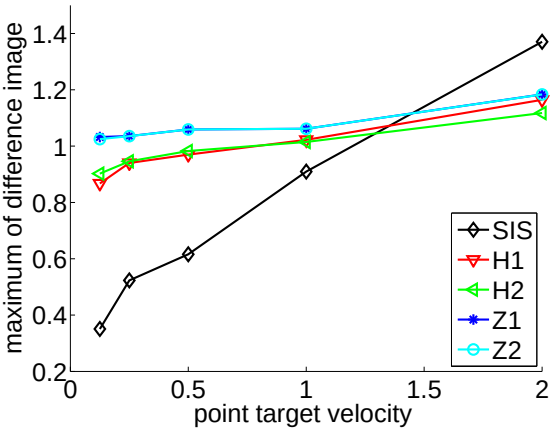


Figure 3.2: *Relative point target amplitude in a difference image as a function of the point target velocity. The relative point target amplitude for a difference image resulting from SIS (bspline) is indicated with diamonds. Zomet (robust SR) is indicated with Z1 and Z2 (zoom factor 1 and 2 respectively) and Hardie (non-robust SR) is indicated with H1 and H2. 48 frames are used for the SR reconstruction and Gaussian blur $\sigma = \sigma_{cam} = 0.41$ is used to describe the camera blur.*

3.3 Point target detection using super-resolution reconstruction

This section describes the point target detection method based on background suppression using SR reconstruction. Although this method is based on existing, well-known techniques, the combination and use of those techniques is innovative. First, the registration method is described, followed by the SR reconstruction method and finally the detection and tracking methods.

3.3.1 Registration

Registration aligns the content of all LR frames prior to SR reconstruction. This registration step is also needed for background subtraction with SIS. There are a variety of image registration techniques described in the literature [71]. We perform registration with a very precise iterative gradient-based shift estimator [49]. This gradient-based shift estimator [40] 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:

$$\text{it.1}: \min_{dx_{k1},dy_{k1}} \frac{1}{P}\sum_{x,y}\left(I_k - I_{k-1} - dx_{k1}\frac{\partial I_{k-1}}{\partial x} - dy_{k1}\frac{\partial I_{k-1}}{\partial y}\right)^2. \tag{3.5}$$

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 partial derivatives are calculated with a Gaussian gradient filter (see p.64 in [68]).

The solution of (3.5), (dx_{k1}, dy_{k1}), is biased, which is corrected in an iterative way:

$$\begin{aligned}\text{it.}n: \min_{dx_{kn},dy_{kn}} \quad & \frac{1}{P}\sum_{x,y}\Big(\tilde{I}_k(x + dx_{k(n-1)}, y + dy_{k(n-1)}) \ldots \\ & - I_{k-1} - dx_{kn}\frac{\partial I_{k-1}}{\partial x} - dy_{kn}\frac{\partial I_{k-1}}{\partial y}\Big)^2.\end{aligned} \tag{3.6}$$

In iteration n $(n > 1)$, I_k is translated by interpolation (indicated by 'tilde') with the estimated subpixel displacement $(dx_{k(n-1)}, dy_{k(n-1)})$ from the previous iteration. Now, the displacement (dx_{kn}, dy_{kn}) between the shifted I_k and I_{k-1} is estimated. This displacement is accumulated with the displacement obtained 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 [49]. The total estimated displacement with the iterative gradient-based shift estimator after M iterations is:

$$(dx_k, dy_k) = (dx_{k1}, dy_{k1}) + \ldots + (dx_{kM}, dy_{kM}). \tag{3.7}$$

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.

3.3.2 Robust super-resolution fusion and deblurring

The second and third step of super-resolution reconstruction are fusion and deblurring. Numerous SR reconstruction methods can be found in the literature; some methods work in the Fourier domain [59, 33], there exists robust methods [20], non-robust methods [26] and some methods [49] are adaptive. Van Eekeren [14] made a quantitative performance comparison between a selection of different SR reconstruction methods. One of the best performing methods is

the method proposed by Hardie et al. [26]. Like many other SR reconstruction methods it models the image formation process in the following way:

$$\vec{I}_k = D_k C_k F_k \vec{Z} + \vec{\theta}_k = H_k \vec{Z} + \vec{\theta}_k, \tag{3.8}$$

where $\vec{I}_k$ is the k^{th} LR frame, $\vec{Z}$ is the HR image scene and $\vec{\theta}_k$ denotes normally distributed additive noise. All reordered in a vector. F_k is the geometric warp matrix based on the results of the registration. C_k is the blurring matrix of the camera and D_k is the decimation matrix which resamples the image to low resolution. For simplification all matrices are combined in H_k. The blurring of the camera is modeled by Gaussian blurring. Note that it is allowed to represent basic operations such as warping and blurring in a matrix, because they are linear in the image intensities.

As already stated in section 3.2.3, a robust SR algorithm for background suppression is proposed because the point target is better preserved in the difference image. A robust algorithm is less sensitive to outliers in the background data, such as moving point targets. With enough frames available and sufficient apparent motion of the point target with respect to the background, a robust SR algorithm will treat the point target as an outlier. For this reason we use a robust method, proposed by Zomet et al. [72], in the experiments. This method is similar to Hardie's method but uses robustness in the minimization procedure.

This is best explained by comparing the minimization procedure used by Hardie with that used by Zomet. We start by giving a short derivation of the Hardie method, and then stress the differences with Zomet.

The total squared error of resampling HR image $\vec{Z}$ is given by:

$$L(\vec{Z}) = \frac{1}{2} \sum_{k=1}^{N} \sum_{i} (\vec{I}_k(i) - (H_k \vec{Z})(i))^2 \tag{3.9}$$

with N the total number of LR frames and i the LR pixels. Note that $\vec{Z}$ is based on all $\vec{I}_k$'s. Taking the derivative of L with respect to $\vec{Z}$ results in:

$$\nabla L(\vec{Z}) = \sum_{k=1}^{N} H_k^T (H_k \vec{Z} - \vec{I}_k) = \sum_{k=1}^{N} \vec{G}_k \tag{3.10}$$

with H_k^T the transposed of H_k. A gradient-based iterative minimization method updates the estimation in each iteration n by

$$\vec{Z}^{n+1} = \vec{Z}^n + \epsilon \nabla L(\vec{Z}) \tag{3.11}$$

with ϵ the step size in the direction of the gradient. The procedure above can be seen as a version of the Iterated Back Projection method [31]. In each

iteration the difference between the resampled HR image $H_k\vec{Z}$ and LR image $\vec{I}_k$ is projected back to the HR grid.

A replacement of the sum of back-projected images $\vec{G}_k$ in (3.10) with a scaled pixel-wise median introduces the robustness of Zomet's method:

$$\nabla L(\vec{Z}) \approx N \cdot \text{median}\left(\vec{G}_k\right)_{k=1}^{N}. \tag{3.12}$$

3.3.3 Detection and tracking of point targets

After background subtraction, the point targets need to be detected in the difference image. The difference images show the amplitude difference between a moving target and its local background, noise, and aliasing artifacts (SIS method). On such a difference image the detection of the objects is done. The simplest 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.

$$T_k = \begin{cases} 0, & \text{for } \Delta D_k(x,y) \leq \text{ threshold} \\ 1, & \text{for } \Delta D_k(x,y) > \text{ threshold} \end{cases} \tag{3.13}$$

This detection method works well for targets that have a sufficiently high Amplitude-to-Noise Ratio (ANR), so that the moving object can be detected in every frame. However, objects having an amplitude close to the local background may not be detected this way.

The specificity of a detection algorithm can be increased if it is performed on a series of subsequent difference images instead of a single difference image. Tracking can be used to associate the detections in the images. It is assumed that a target path is a continuous path over time, which means that the positions of correct detections are highly correlated over the frames. Uncorrelated detections are unlikely to be correct detections.

In this chapter a 3-out-of-5 tracking algorithm [8] is used to increase the specificity. This tracking algorithm performs first a special dilation on 5 subsequent frames after thresholding. This allows a limited displacement of the point target in the next frames. In order to keep the moving point target in track, a dilation with different kernel sizes is performed on frame 1, 2, 4 and 5. The center frame, 3, is not dilated. Kernel sizes are chosen such that point targets which have an apparent motion with respect to the background of maximum 2 pixels/frame can be tracked (see Figure 3.3). Afterwards a pixel-wise summation is performed. Pixels with a sum larger than or equal to 3 are marked as detection after tracking. This means that the targets are present in at least three of the five frames. Note that this tracking algorithm can be improved (using velocity and heading of the point target), which will result in a further reduction of the number of false alarms without losing sensitivity.

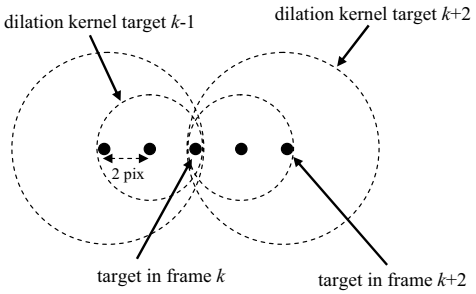


Figure 3.3: *The different kernel sizes used for the 'special' dilation in the 3-out-of-5 tracking algorithm. The smallest kernel size is 2 pixels, which support tracking of point targets with an apparent motion up to 2 pixels/frame.*

3.4 Experimental setup

The performance of the different algorithms is tested on images containing point targets. These images are constructed by inserting simulated point targets in a real image sequence. To simulate a realistic scenario, first a few real-world scenarios are analyzed.

3.4.1 Real-world scenario

An incoming missile at long distance is observed as a point in a recorded sequence. Such an incoming missile must be detected as early as possible. In this analysis two missiles, a Stinger [2] and an AA-10 [3], are chosen, because most of these missiles are radar silent. This means that they must be detected using an electro-optical sensor. First, let us analyze the observed velocity of missile. The apparent velocity (expressed in rad/s) of a missile with respect to the background from the observer's point of view can be described with:

$$v_t = \frac{v}{d} \sin \alpha. \tag{3.14}$$

Here, v is the velocity of a missile, d is the distance to the observer and α is the angle between the missile's path and the shortest path to the observer. The apparent motion of a missile in camera coordinates depends on the Instantaneous Field-Of-View (IFOV) and the frame rate of the camera.

Realistic specifications of an infrared camera for the task of missile detection are: a center wavelength of 4 μm, IFOV = 1.5 mrad, a frame rate of 15 Hz and a sensitivity of 0.025 K. The latter determines the amount of noise radiance. With

these specifications a missile, such as a Stinger, flying at Mach 2 at 2 km with $\alpha = 10°$ has an apparent motion of approximately 2.6 pixels/frame. For a scenario with a larger range (e.g. an AA-10 flying at Mach 4 at 60km with $\alpha = 10°$) the apparent motion is approximately 0.17 pixels/frame.

The observed missile intensity depends on its radiated energy at a certain wavelength. Propagation losses are ignored in our analysis. The observed missile and background are regarded as black bodies of which the energy per unit time per unit surface area per unit wavelength can be calculated with Planck's law [53]. The total observed radiance E_t on one sensor element is defined as

$$E_t = \beta E_m + (1 - \beta) E_{bg}, \tag{3.15}$$

with E_m the radiance of the missile, E_{bg} the radiance of the background and β the area fraction of the missile. The difference in radiance between an observed missile and its background is defined as: $\Delta E_m = \beta(E_m - E_{bg})$. For two different real-world scenarios ΔE_m is calculated and compared with the clutter radiance and the noise radiance.

High clutter scenario. In this scenario a Stinger is fired from the ground to an air-target at 3 km. The temperature of the background is chosen to be 290 K and the clutter is chosen to be $\Delta T_{cl} = 1$ K. The Stinger has a velocity of Mach 2 and has a diameter of 7 cm. Its velocity defines its aerodynamic temperature [30] to be 480 K. The diameter of the missile, the distance to the target, IFOV and transfer of the camera define the area fraction β to be $1.9 \cdot 10^{-4}$. This results in the ratio between the missile's radiance and the noise radiance: $\Delta E_m / E_n \approx 24$ with E_n the noise radiance. The ratio between clutter radiance and noise radiance is: $\Delta E_{cl} / E_n \approx 41$. This indicates a clutter-dominated scenario.

Low clutter scenario. In this scenario an AA-10 is fired from an aircraft at 9 km altitude to an air-target at a distance of 60 km. The air temperature is estimated (-10 K / +1 km altitude difference) to be 203 K and the clutter is estimated to be $\Delta T_{cl} = 0.1$ K. The AA-10 flies with Mach 4 and has a diameter of 23 cm. Its velocity defines its aerodynamic temperature [30] to be 736 K. The diameter of the missile, the distance to the target, IFOV and transfer of the camera define the area fraction β to be $5.1 \cdot 10^{-6}$. This results in the ratio between the missile's radiance and the noise radiance: $\Delta E_m / E_n \approx 9$. The ratio between clutter radiance and noise radiance is: $\Delta E_{cl} / E_n \approx 0.04$. This indicates a noise-dominated scenario.

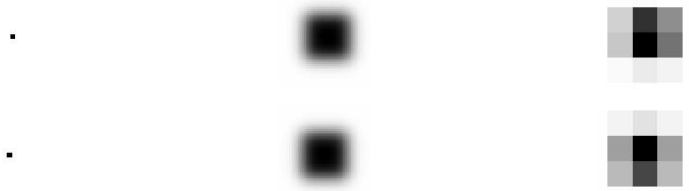
3.4.2 Simulated scenario

The point target images used for the experiments are constructed from a real image sequence which was recorded with an infrared camera (Radiance HS, 3 - 5 μm, 256 $\times$ 256, 15 fps). The recorded images, which have an intensity range

of [1000, 1172] grey values, contain noise with an estimated standard deviation of 1 grey value. Furthermore, they contain artifacts such as bad pixels and non uniformity. Before inserting the point targets, the recorded image sequence is corrected for the latter two types of artifacts [55]. This will improve the detection results and will make it easier to compare the results of our experiments. The camera movement of the recorded sequence is approximated by a frame-to-frame translation, which is on average over the frames $v_x = 3.10$ pixels/frame and $v_y = 0.64$ pixels/frame.

3.4.3 Simulated point targets

The point targets are simulated and added to the LR camera images. First the point targets are placed in a super scale image, which is a factor 15 larger than the LR camera image. Here the position of the point target is integer based. To obtain the LR image with the point target, a camera model is applied to the super scale image. In this camera model the MTF of the camera is modeled by a lens blur ($\sigma_{psf} = 0.3$ LR pixel) and a fill-factor (81% area). The camera MTF is plotted in Figure 3.1. The camera model also subsamples the super scale image with a factor of 15. This subsampling is done by taking each 15th pixel. The resulting LR image contains the point target with aliasing. The maximum point target energy depends on the LR sub-pixel position of the LR image. We define the amplitude of a point target as the average maximum intensity of the point target in all available LR images. The point target simulation is visualised in figure 3.4. Here can also be seen that due to aliasing the maximum energy per frame of the point targets is not constant.



(a) superscale point (b) point with camera model (c) downscaled point

Figure 3.4: *Two simulated point targets with different locations. At the left (a), the input point target in super scale is visualised. Note that the point is slightly larger than 1 pixel for visualisation purposes. In the center figures (b), the camera model is applied to the point target. It can be seen that the point target energy is divided over a large number of pixels. When the image is downscaled (c), the point target energy is still in more than one pixel.*

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 aliasing in a similar way as the background. Second, the target is a point target and has therefore a small footprint. The error that is made by adding instead of replacing in the super scale image is $\Delta E_{add} = \beta \Delta E_{cl}$, which is the clutter radiance that is not replaced. In our scenario's, the worst case of ΔE_{add} is $E_n 1.9 \cdot 10^{-4} \cdot 41 \approx E_n 10^{-2}$. This means that the maximum error that is made is 100 times smaller than the temporal noise.

For the experiments, the amplitude and apparent motion of the point target with respect to the background are varied according to the calculations of the different real-world scenarios. The point target amplitude is varied between 4 and 56 grey values and the apparent motion of the point target is varied between 0.125 LR pixels per frame (almost no movement with respect to the background) to 2 LR pixels per frame. For each velocity 8 different subpixel start locations of the point target are chosen. To simulate the different clutter scenarios, two different kind of sequences are constructed: one with the point target in a low clutter region and one with the point target in a high clutter region of the real image sequence. The upper part (256×128) of a few frames of a constructed point target sequence is shown in Figure 3.5. Here, the point target is placed in a low clutter region with an apparent motion w.r.t. the background of 2 pixels/frame.

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

Figure 3.5: *Three frames (256×128) of a constructed point target sequence. The position of the point target can be seen in Figure 3.6. 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.*

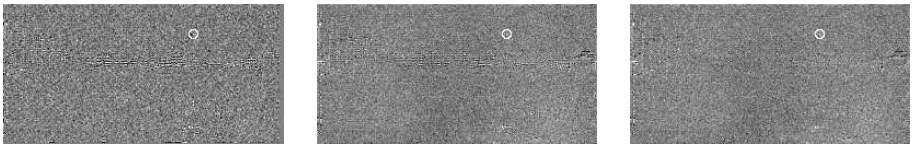
3.4.4 Processing details

The constructed LR images are registered using the techniques presented in section 3.3.1. The number of iterations used is 5. The σ of the Gaussian derivative filters is 1. Then the constructed LR images are processed by three different

background suppression methods: 1) SIS with bspline interpolation, 2) Zomet's robust SR reconstruction method with zoom factor 1 and 3) Zomet's robust SR reconstruction method with zoom factor 2. The camera model used in Zomet's method consists of a Gaussian blurring only. In our experiments this blurring has been set to $\tilde{\sigma}_{cam} = 0.41$, which is the best Gaussian fit to the real camera model ($\sigma_{psf} = 0.3$ and fill-factor = 81%). In the Zomet reconstruction 10 iterations are used. There is no regularisation used in the Zomet algorithm.

3.5 Results

Figure 3.6 shows the difference images for the three different background suppression methods. 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 SIS.



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

Figure 3.6: *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.*

To evaluate the performance of the different methods under different scenarios, sequences of 48 frames are used. The different scenarios are created using: 1) different clutter levels (indicated with CNR = Clutter Noise Ratio), 2) different point target amplitudes (indicated with ANR = Amplitude Noise Ratio) and 3) different point target velocities (indicated with PTV = Point Target Velocity). The latter two can be controlled, because they are simulated, but the clutter level cannot. Therefore, as low clutter region the sky-area in the image is selected and as high clutter region the building-area in the center. Clutter is defined as the maximum gradient magnitude present in the Region-of-Interest (ROI) used for a

specific scenario. In Figure 3.7 the ROIs (one for high and one for low clutter) are visualized for frame 24. Note that in our analysis the point target is always present in the ROI. Furthermore, an external mask is used to mask the bad pixels in the original camera scene.

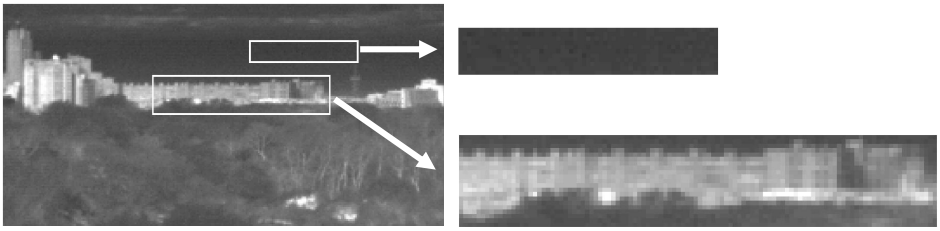


Figure 3.7: *Regions-of-Interest in frame 24 that are used for analyzing the point target detection in a low and high clutter scenario. All images are visualized between [1000, 1120]. Note that no simulated point target is present in these images.*

The detection results are presented in two different ways. First, the results are presented by means of Receiver Operating Characteristic (ROC) curves. These curves represent the relation between the true positive rate (sensitivity) and the false positive rate (1 - specificity) for different threshold values. Next, the performance for the different algorithms is compared for a representative operating point.

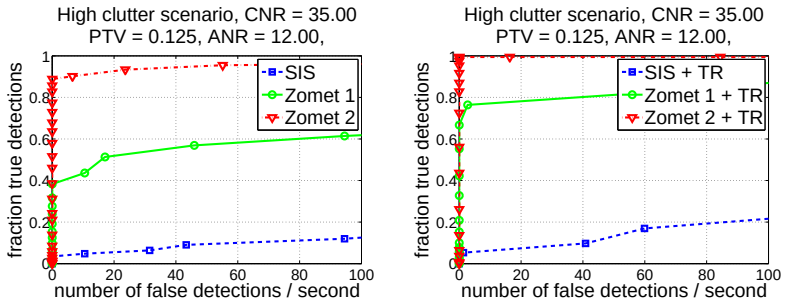
3.5.1 ROC curves

An ROC curve relates the sensitivity to the specificity of an algorithm. For a detection method such an ROC curve can be determined by varying the threshold and counting the number of true and false detections (knowing the ground truth). In our analysis, first the fraction of true detections is determined. A true detection occurs when in PT ROI, a 5×5 neighborhood around the point target, a detection is present. Here, a detection is defined as one or more connected pixels after thresholding or tracking. As every frame contains one point target, the number of true detections divided by the total number of frames equals the fraction of true detections. This is indicated on the vertical axis of the ROC curve.

The next part of the analysis is to determine the number of false detections. First, the true detections are removed if they are smaller than twice the size of the PT ROI. This is done to make sure that true detections are not counted as false detection as well, except when a true detection is large too. The latter situation can occur for small threshold values. After this removal, the number of false

detections in each frame is determined after labeling (4-connected neighbours) all detections. In the ROC curves that are presented here, the number of false detections per second is plotted on the horizontal axis. These numbers correspond with a frame rate of 15 frames/second and a frame size of 256×256 pixels. Because the evaluation is done on a smaller, defined ROI (see Figure 3.7), the measured number of false detections are scaled such that they correspond to a 256×256 pixel frame.

In Figure 3.8 the improvement in detection performance that can be obtained by using tracking is shown. The high clutter scenario is used with a low point target velocity (0.125 pixel/frame) and an ANR of 12. It can be seen that tracking improves the detection results for both background subtraction with Zomet and the SIS method.



(a) no tracking (b) 3-out-of-5 tracker

Figure 3.8: *ROC curves of the three different background suppression methods for a high clutter scenario (buildings), a point target velocity of 0.125 pixels/frame and an ANR of 12. Sub-figure (a) shows ROC curves obtained without using tracking and sub-figure (b) shows ROC curves after using a 3-out-of-5 tracker.*

Figure 3.9 shows ROC curves of a low clutter scenario. These curves show that the background subtraction methods using SR reconstruction outperform the SIS method. For fast moving point targets in a low clutter scenario (lower row of Figure 3.9), the Zomet method performs better due to its noise reduction capabilities. For slow moving point targets (upper row of Figure 3.9) the Zomet method performs also better because the point target is efficiently suppressed in the background estimation, resulting in more point target energy in the difference image. Therefore, the improvement using Zomet is much larger for point targets with a slow apparent velocity than for point targets with a high apparent velocity. As expected, the difference in performance between the two zoom factors of Zomet is not significant, because there is not much residual aliasing in the low clutter scenario.

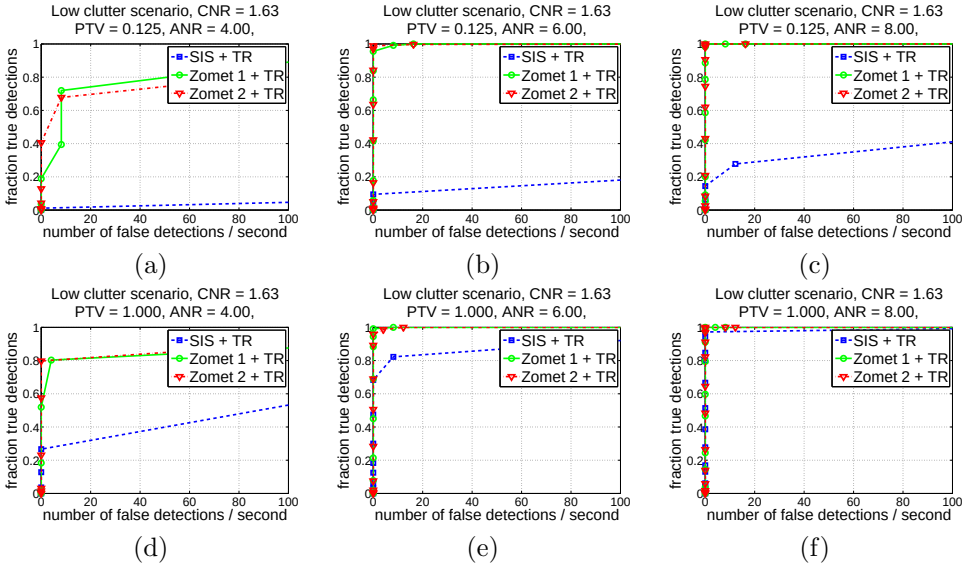


Figure 3.9: *ROC curves of the three different background suppression methods + tracking for a low clutter scenario (sky). SIS = Shift, Interpolate and Subtract, Zomet 1 & 2 = Zomet's robust SR method with zoom factor 1 and 2, respectively. TR = tracking. The upper row ((a),(b),(c)) shows results of data with low apparent point target velocity (0.125 pixels/frame) and the lower row ((d),(e),(f)) with high apparent point target velocity (1 pixels/frame). Each column shows a specific point target Amplitude Noise Ratio (ANR).*

Figure 3.10 shows ROC curves of the high clutter scenario. These results show the excellent performance of Zomet's SR method with zoom factor 2. This method reduces the aliasing artifacts, which are specifically present in the high clutter region, much better than both other methods. This can also be seen from the difference images depicted in Figure 3.6 in the building ROI. In the upper row of Figure 3.10 (small apparent point target motion with respect to the background) Zomet's SR method (both zoom factors) outperforms the SIS method. This can be explained by a higher point target amplitude in the difference image of Zomet's method. The difference image of Zomet with zoom factor 2 also contains less aliasing artifacts. Note that in Figure 3.10d the SIS method performs better than Zomet's method with zoom factor 1 for small ANR. This is explained by the fact that the background estimation generated with the Zomet 1 method will have aliasing errors. Because this background estimation is used for the detection in all frames, these aliasing errors will result in correlated false detections. In the

SIS method, there are also aliasing errors in the frames, but these errors will be uncorrelated over subsequent frames, and will therefore not lead to correlated detections. As correlated false detections are assumed to be correct detections by the tracking algorithm, the Zomet 1 method will lead to more false detections. This result shows that it is useful to apply SR reconstruction with a zoom factor larger than 1, as this reduces the aliasing noise.

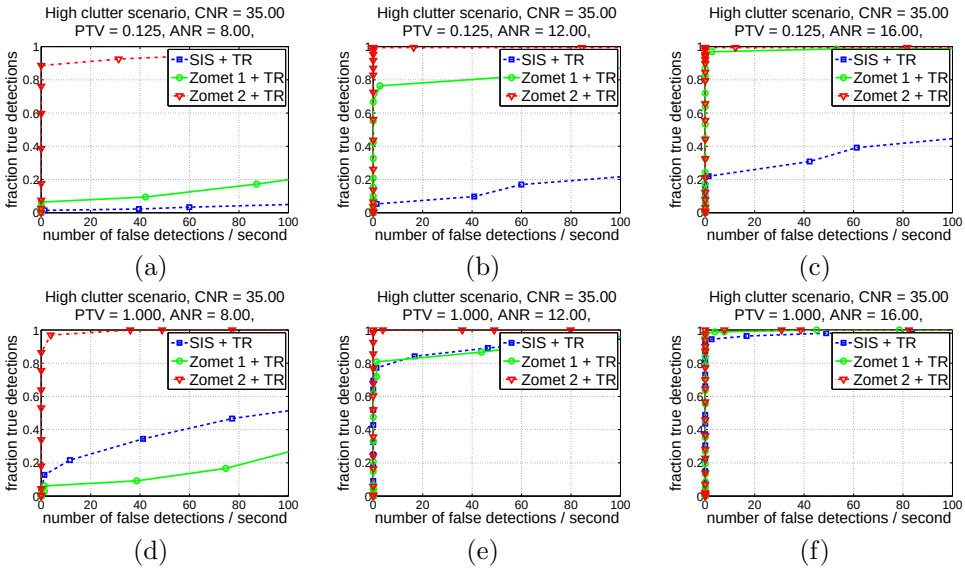


Figure 3.10: *ROC curves of the three different background suppression methods + tracking for a high clutter scenario (buildings). SIS = Shift, Interpolate and Subtract, Zomet 1 & 2 = Zomet's robust Super-Resolution method with zoom factor 1 and 2, respectively. TR = tracking. The upper row ((a),(b),(c)) shows results of data with low apparent point target velocity (0.125 pixels/frame) and the lower row ((d),(e),(f)) with high apparent point target velocity (1 pixels/frame). Each column shows a specific point target Amplitude Noise Ratio (ANR).*

3.5.2 Performance comparison

The area under the ROC curve is often used as a performance measure [9, 50]. In our case the interesting part of the ROC curve is when the number of false detections is small [60]. Therefore, the area under the ROC curve is determined up to a value of 20 false detections/second (for a 256×256 pixel frame and 15 frames/second).

In this chapter the performance of a specific scenario is determined by ANR_{80}, the ANR of the point target that corresponds with a 80% area under the ROC curve. This is the Amplitude-to-Noise Ratio for which 80% of the point targets are detected with 20 false detections/second. Note that a smaller ANR_{80} indicates a better performance. An operating point of 80% area under the ROC curve up to 20 false detections/second seems useful: the fraction of true detections is high enough to perform more advanced tracking while reducing the number of false positives even further.

For each scenario, ROC curves are measured for varying ANRs. Under each ROC curve the area under the ROC curve up to 20 false detections/second is calculated. By linear interpolation between those areas the ANR_{80} is determined. The resulting ANR_{80}s are shown in Table 3.1 including the precision. The precision of the ANR_{80} is determined from the standard deviation of each point (8 measurements) on the ROC curves. The inaccuracy of the interpolation is not taken into account in this precision. As can be seen in Table 3.1, the improvement of tracking is in the order of 1.5, except for point targets with the highest apparent target velocity (PTV) of 2 pixels per frame, where the effect of tracking is negligible. This might be explained by the limited association window used in the tracking algorithm.

The relative performance of the proposed detection method using Zomet's robust SR reconstruction for background suppression compared to the detection method based on SIS is presented in Table 3.2. Each number indicates the ratio of the ANR_{80} of SIS + TR (baseline) and the ANR_{80} of Zomet 1 + TR or Zomet 2 + TR. Larger numbers indicate a better performance.

For the smallest apparent target velocity the improvement is at least 2.3 for the Zomet method compared to SIS. For high clutter scenarios, Zomet 2 is significantly better than Zomet 1, while for low clutter scenarios the difference between Zomet 1 and Zomet 2 is not significant. The largest improvement compared to the baseline is obtained with Zomet 2 for the scenario of high clutter and a small apparent point target velocity. Here, the improvement is almost a factor 4.

The measured improvement in performance of a detection method using SR reconstruction to the expected theoretical improvement is also given in Table 3.2. The theoretical values are based on the analysis in section 3.2.2 (3.4) and 3.2.3 (Figure 3.2) and are given by: $TI = I_{noise} \cdot I_{ampl}$. Here, I_{noise} is the ANR improvement due to the temporal noise reduction, which is $\sqrt{2}$ and I_{ampl} is the ANR improvement due to point target suppression in the background by using robust SR reconstruction.

As we cannot quantitatively estimate the reduction of the aliasing noise, the values for the high clutter scenario are only indicative. As can be seen in this table, mostly the real performance is smaller than our theoretical expectations. For point targets with a high apparent motion with respect to the background, the improvement is primarily due to temporal noise reduction. Here, our measurements

are close to what is expected. For small apparent motion of the point targets, the difference between theory and measurements is somewhat larger. This may be explained by the fact that the simulated point targets are placed in real recorded data, which introduces an error in the position and therefore in the apparent motion of the point target. Relatively, this error is larger for smaller apparent motions.

3.6 Conclusions and discussion

In this chapter we present a new method for point target detection based on Super-Resolution (SR) reconstruction of the background. With a simulation based on a real-world sequence we show that the specificity and sensitivity of a point target detection method is improved. The improvement in specificity is based on two properties of the SR reconstruction algorithm: temporal noise reduction and anti-aliasing. Due to the temporal noise reduction and anti-aliasing the number of false alarms decreases, as there is less noise in the background estimation and therefore also less noise in the difference image on which the detection is based.

The sensitivity of point target detection is increased by the point target suppression capabilities of SR reconstruction in the background estimate. Therefore, the amplitude of the point target is preserved in the difference image. This effect is larger for point targets with lower apparent target velocity. Robust SR reconstruction is used, because this suppresses outliers and therefore has hardly any contribution of the point target in its background estimation, whereas for non-robust SR reconstruction methods a small portion of the point target energy will still be seen in the background estimation.

It can be seen that background suppression with SR reconstruction performs better than a standard Shift, Interpolate and Subtract (SIS) algorithm in almost all tested scenarios. As expected SR reconstruction with zoom factor 2 performs better than SR reconstruction with zoom factor 1 in high clutter scenarios. This effect is due to the fact that a better estimation of the background by using anti-aliasing, as is done with zoom factor 2, will decrease the number of false detections. In low clutter scenarios a higher zoom factor does not improve the performance.

The improvement using Super-Resolution reconstruction is only demonstrated for a limited dataset. However, these results provide indicators for the performance of these techniques using other imaging systems and for other scenes. The performance depends on the properties of the imaging system, such as the sharpness and the sampling frequency of the system. These properties result in different aliasing properties. Given the theory, the performance gain will be lower for systems with less aliasing. On the other hand, the performance gain will increase for systems with more aliasing. The results also depend on the amount of clutter

Table 3.1: *Point targets Amplitude Noise Ratios (ANR_{80}) which correspond with 80% area under the ROC curve up to 20 false detections/second for the Shift, Interpolate and Subtract (SIS) and Zomet method with zoom factor 1 and 2. For all methods the results with and without tracking (TR) are given. Note that a smaller ANR_{80} indicates a better performance.*

	SIS	SIS + TR	Zomet 1	Zomet 1 + TR	Zomet 2	Zomet 2 + TR
low clutter, PTV 0.125	23.3 ± 3.9	15.6 ± 3.7	8.0 ± 0.7	5.3 ± 1.3	9.4 ± 0.6	5.1 ± 0.8
low clutter, PTV 0.25	24.1 ± 2.0	15.4 ± 2.7	7.8 ± 0.7	5.1 ± 0.8	9.3 ± 0.9	5.3 ± 1.3
low clutter, PTV 0.5	15.5 ± 3.9	12.0 ± 2.9	7.7 ± 0.9	4.2 ± 1.1	8.7 ± 1.2	5.4 ± 1.2
low clutter, PTV 1	9.5 ± 0.8	6.2 ± 0.7	7.8 ± 1.4	4.4 ± 1.8	8.6 ± 1.8	4.0 ± 1.3
low clutter, PTV 2	7.4 ± 0.6	7.2 ± 1.2	6.0 ± 0.3	5.6 ± 0.7	6.7 ± 0.7	5.7 ± 0.9
high clutter, PTV 0.125	57.5 ± 12.9	29.6 ± 4.7	16.5 ± 1.0	12.8 ± 1.9	10.6 ± 0.7	7.7 ± 0.2
high clutter, PTV 0.25	44.8 ± 4.5	29.9 ± 7.3	16.2 ± 0.6	12.7 ± 1.2	10.4 ± 0.5	8.1 ± 0.7
high clutter, PTV 0.5	31.8 ± 3.6	21.7 ± 2.7	16.0 ± 0.6	11.9 ± 1.6	10.0 ± 0.6	7.6 ± 0.2
high clutter, PTV 1	20.4 ± 1.8	12.3 ± 1.5	15.7 ± 0.8	12.0 ± 1.1	9.5 ± 1.1	7.3 ± 1.0
high clutter, PTV 2	15.0 ± 0.9	11.8 ± 1.9	15.7 ± 0.5	14.8 ± 1.2	7.9 ± 1.1	5.8 ± 1.1

Table 3.2: *Relative performance of the proposed detection method using Zomet for background suppression and the detection method based on Shift, Interpolate and Subtract (SIS). For all methods only the results with tracking (TR) are given.*

	Zomet 1 + TR vs. SIS + TR	Zomet 2 + TR vs. SIS + TR	theoretical improvement
low clutter, PTV 0.125	2.9 ± 1.0	3.1 ± 0.9	4.1
low clutter, PTV 0.25	3.0 ± 0.7	2.9 ± 0.8	2.8
low clutter, PTV 0.5	2.9 ± 1.0	2.2 ± 0.8	2.4
low clutter, PTV 1	1.4 ± 0.6	1.6 ± 0.8	1.7
low clutter, PTV 2	1.3 ± 0.3	1.3 ± 0.3	1.2
high clutter, PTV 0.125	2.3 ± 0.5	3.8 ± 0.6	> 4.1
high clutter, PTV 0.25	2.4 ± 0.6	3.7 ± 0.5	> 2.8
high clutter, PTV 0.5	1.8 ± 0.3	2.9 ± 0.4	> 2.4
high clutter, PTV 1	1.0 ± 0.2	1.7 ± 0.3	> 1.7
high clutter, PTV 2	0.8 ± 0.1	2.0 ± 0.4	> 1.2

in the scene. In our simulations we tested two scenarios: a high clutter scenario which was clutter dominated and a low clutter scenario which was temporal noise dominated. This provides two measuring points at the extremes of the clutter-to-noise ratio. The performance gain of scenes with another clutter-to-noise ratio will therefore be in-between the low and high clutter improvement shown in this chapter.

Summarizing, we show that point target detection after background suppression with SR reconstruction is significantly better than detection results with the SIS method, especially in high clutter scenarios and for low apparent target motion w.r.t. the background. While maintaining an equal detection performance, the proposed method using SR reconstruction can detect point targets which have an up-to 4 times smaller Amplitude-to-Noise-Ratio in the scenarios studied. In practice this implies that a point target can be detected at longer range.

Chapter 4

Super-resolution reconstruction of large moving objects and background

Abstract

Unlike most Super-Resolution (SR) methods described in literature, which perform only SR reconstruction on the background of an image scene, we propose a framework that performs SR reconstruction simultaneously on the background *and* on moving objects. After registration of the background, moving objects are detected and to each moving object registration is applied. The fusion and deblurring tasks of the framework are performed by the well-known method of Hardie. To evaluate the performance of the framework the Triangle Orientation Discrimination (TOD) method is used. 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.

4.1 Introduction

Super-Resolution (SR) reconstruction is nowadays a well-known technique in image processing. Although the concept already exists for 20 years [59], not much

[1]The major part of this chapter has been published in A.W.M. van Eekeren, K. Schutte, J. Dijk, D.J.J. de Lange and L.J. van Vliet, Super-Resolution on moving objects and background, in *Proc. 13th International Conference on Image Processing*, vol.1, pp.2709–2712, IEEE, 2006. [13]

attention is given to a specific case: SR reconstruction on moving objects. This is remarkable, because moving objects (e.g. cars, flying objects) are often the interesting parts in an image sequence. The novelty of our proposed framework is that it performs SR reconstruction on moving objects as well.

In order to perform SR reconstruction, a sequence of Low-Resolution (LR) images is used which contains sub-pixel motion. After the LR images are registered accurately, a High-Resolution (HR) image sequence is obtained after fusion.

However, if moving objects are present in the LR image sequence, or when the LR image sequence consists of multiple depth layers, it is not trivial to use a 'standard' registration technique because multiple instances of apparent motion are present in the sequence. To overcome such problems we will make use of masks during registration [55]. In this chapter a framework will be presented which performs a correct registration and SR reconstruction on all regions with distinguishable apparent motion in an image sequence; so on the background *and* on moving objects.

In most papers, the performance evaluation of SR reconstruction methods leaves much to be desired, showing only some nice looking processed images. The exact SR performance remains unknown. With the large amount of SR methods available nowadays it is of great importance to have a proper performance measure for comparing these methods. To obtain a quantitative performance measure we propose to use the Triangle Orientation Discrimination (TOD) method [7].

The outline of this chapter is as follows. In section 4.2 a global setup is presented of the proposed framework. A description of the used evaluation method, a setup of the evaluation experiment and the obtained performance measurements of our framework can be found in section 4.3. Finally, in section 4.5 conclusions will be drawn.

4.2 Framework

We focus on a framework for SR reconstruction on the background *and* on moving objects within dynamic image sequences. A flow diagram of the proposed framework in steady state is depicted in Figure 4.1. In this flow diagram the blocks 'REGISTRATION background' and 'FUSION & DEBLUR background' are the elementary blocks for a SR reconstruction method of image sequences without moving objects. To perform SR reconstruction on moving objects as well, the blocks 'MOVING OBJECT DETECTION' and 'MERGING background and moving objects' are essential. As can be seen in Figure 4.1 after the 'MOVING OBJECT DETECTION' each moving object is processed in the same way (first 'REGISTRATION' followed by 'FUSION & DEBLUR') as an image sequence without moving objects. In our framework we use for the 'FUSION & DEBLUR' part a SR reconstruction method described by Hardie [26].

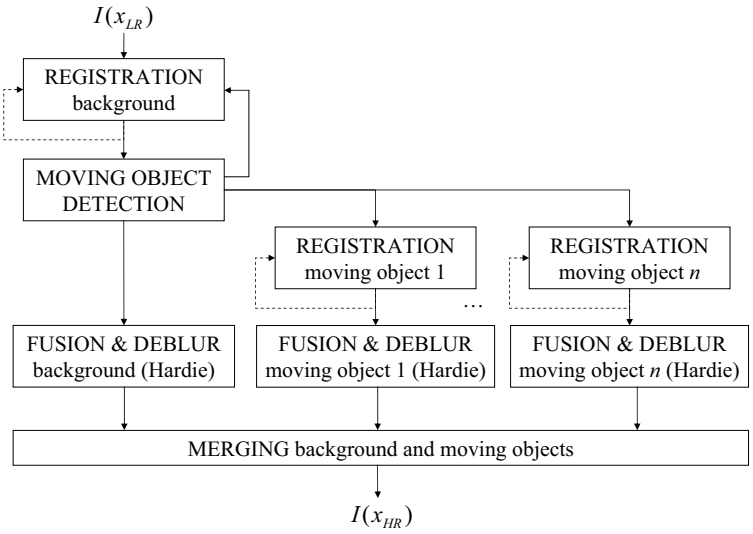


Figure 4.1: *Frame-based flow diagram of framework (steady state) for SR reconstruction on background and moving objects. Note that each moving object is processed separately in a 'standard' way.*

4.2.1 Registration

At *initialization* the first two frames of an image sequence are assumed to contain *no* moving objects. Hence we are able to form a proper model of the scene's background and the background motion vector field at initialization. When a moving object is entering the scene for the first time the registration of the background is slightly incorrect, because no mask from 'MOVING OBJECT DETECTION' is present yet. However, it is assumed that this registration is accurate enough to detect the moving object in the next block of the framework. Note that the dashed lines in Figure 4.1 denote the feedback of the previous calculated motion vector field.

In *steady state* the masks from 'MOVING OBJECT DETECTION' are used for a correct registration of the background in the next image frame (see feedback in Figure 4.1). The masks are tracked over time and a prediction is made of the moving objects locations in the next frame. Such regions will not be used to calculate the motion vector field of the background.

For correct registration of the moving objects it is essential that no object-border pixels (of which it is unknown whether they belong to the background or

moving object) are used. Therefore, the masks are eroded in advance.

To estimate the motion vector field, an iterative gradient-based shift estimator as proposed by Lucas and Kanade [40] is used. This estimator, which approaches the Cramer-Rao bound, results in a very precise unbiased registration [48].

4.2.2 Moving object detection

Moving object detection is an essential step for SR reconstruction of an image sequence with moving objects, because multiple regions with different apparent motion are present. Moving objects are detected by comparing a warped estimate of the background, $\hat{I}_{bg}(x_{LR})$, with the current frame $I(x_{LR})$ as depicted in Figure 4.2.

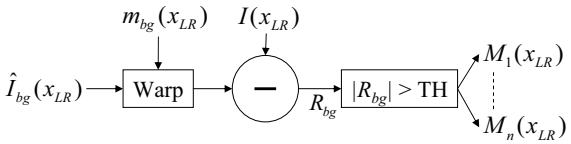


Figure 4.2: *Flow diagram of moving object detection.*

The estimate of the background, $\hat{I}_{bg}(x_{LR})$, is calculated by a simple 'Shift & Add' fusion (see [18] for more details) of a number (e.g. 10) of previous background images on a LR grid. A background image is defined here as an image which contains only background information: no moving objects are present or the moving objects are masked out.

The background estimate is warped to the current frame with the motion vector field $m_{bg}(x_{LR})$, which was calculated in the 'REGISTRATION background' step. Next, the warped background estimate is subtracted from the input image $I(x_{LR})$ and the residue image $R_{bg}(x_{LR})$ results.

All pixels in $|R_{bg}(x_{LR})|$ are compared with a threshold TH and are marked as moving object (1) or as background (0). Each group of pixels marked as moving object will obtain a corresponding mask $M_i(x_{LR})$, with i a unique label for this moving object. Mask $M_i(x_{LR})$ is used in 'REGISTRATION moving object i'. Furthermore, it is used to predict the region of the moving object in the next frame (feedback loop to 'REGISTRATION background') and it is used for MERGING moving object i with the background.

4.2.3 Fusion and deblurring

The 'FUSION & DEBLUR' blocks in the framework can be implemented by various methods reported in literature. Some methods combine fusion and deblurring, while others perform both parts sequentially. In this chapter we use the well-known combined SR reconstruction method of Hardie. It is used on the background as well as on the moving objects.

Hardie assumes a discrete observation model that relates the ideally sampled image $\mathbf{z}$ and the observed frames $\mathbf{y}$:

$$y_m = \sum_{r=1}^{N} w_{m,r} z_r + \eta_m \tag{4.1}$$

where $w_{m,r}$ represents the contribution of the r^{th} HR pixel in $\mathbf{z}$ to the m^{th} 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 represents additive noise.

If the registration parameters are estimated, the observation model can be completely specified. The HR image estimate $\hat{\mathbf{z}}$ is defined as the $\mathbf{z}$ that minimizes:

$$C_{\mathbf{z}} = \sum_{m=1}^{pL} M_m \left(y_m - \sum_{r=1}^{N} w_{m,r} z_r \right)^2 + \lambda \sum_{i=1}^{N} \left(\sum_{j=1}^{N} \alpha_{i,j} z_j \right)^2 \tag{4.2}$$

with p the number of LR frames, L the number of LR pixels per frame and $M(m)$ the mask value at pixel m. Note that in the Hardie's original cost function $M = 1$ for every LR pixel.

The cost function in (4.2) balances two types of errors. The first term is minimized when a candidate $\mathbf{z}$, projected through the observation model, matches the observed data. The second term is a regularization term, which is necessary because directly minimizing the first term is an ill-posed problem. The parameters $\alpha_{i,j}$ are selected as proposed by Hardie such that this term is minimized when $\mathbf{z}$ is smooth.

For SR reconstruction of moving object i all p masks $M_i(x_{LR})$ obtained from 'MOVING OBJECT DETECTION' are used, which result in HR image $I_{mob_i}(x_{HR})$. SR reconstruction of the background uses the same, but *inverted*, masks, and results in HR image $I_{bg}(x_{HR})$.

Although Hardie was originally designed as a reconstruction method for static sequences, we have applied it to dynamic sequences. The number of frames p used is 40 and the actual frame is used as reference frame. Note that the Hardie method is not very efficient for dynamic processing, because for each new frame the cost function has to be minimized again.

4.2.4 Merging

The final step of the framework merges the reconstructed moving objects with the reconstructed background. Because our algorithm is not capable of doing an accurate segmentation on boundaries of background and moving objects, no SR reconstruction is performed here. This boundary region is filled with bi-linear interpolated LR pixels $\tilde{I}(x_{HR})$ as depicted in Figure 4.3.

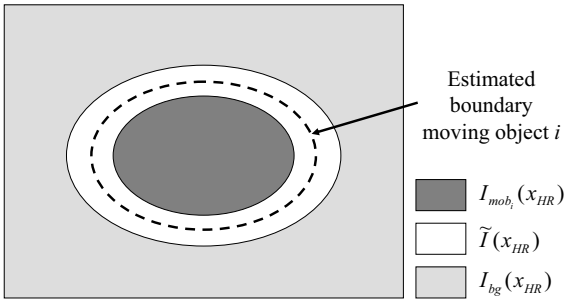


Figure 4.3: *Visualization of a resulting HR image $I(x_{HR})$ after merging background and moving object i.*

The estimated boundary of moving object i at the current frame is defined by a Nearest Neigbor (NN) interpolation of mask $M_i(x_{LR})$ and is denoted $M_i(x_{HR})$. A dilation of $M_i(x_{HR})$ gives the outer border of the boundary region and an erosion of $M_i(x_{HR})$ gives the inner border.

4.3 Evaluation experiment

To measure the performance of SR reconstruction, we use the Triangle Orientation Discrimination (TOD) method (Bijl and Valeton [7]). This evaluation method is preferred over other methods (e.g. comparing of MTF (Modulation Transfer Function)) because 1) the measurement is done in the spatial domain and therefore well localized, and 2) it employs a specific vision task.

4.3.1 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 4.4).

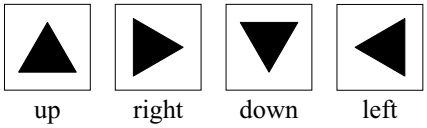


Figure 4.4: *The four different stimuli used in the TOD method.*

The observer task is a four-alternative forced-choice, so the observer has to indicate which of the four orientations is perceived, even if he is not sure. The probability of a correct observer response increases with the triangle size. For a number of observations at different contrast a 75% correct triangle size is determined and the corresponding TOD curve plotted. This curve describes the performance of the imaging system under measurement.

4.3.2 Setup

A flow diagram of our evaluation method setup using the TOD method is depicted in Figure 4.5.

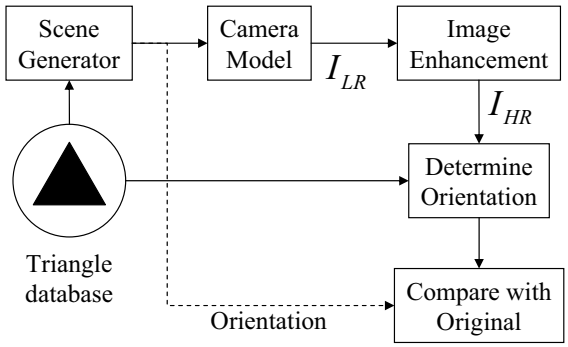


Figure 4.5: *Flow diagram of the evaluation experimental setup using the TOD method.*

The camera model incorporates the PSF of the lens, the fill-factor of the pixels on the focal plane array sensor, and several noise sources including the readout- and photon noise. In our experiment a 'regular' camera with $\sigma_{psf} = 0.3$ and a fill-factor of 0.9 is simulated. The overall noise is assumed to be Gaussian

distributed.

Our evaluation method employs an automatic observer that tries to discriminate the orientation of the triangles before and after processing. This is done by correlating such a triangle with a database containing 'ideal' triangles (see Figure 4.4) of different size and sub-pixel position. The orientation of the triangle that maximizes the similarity measure is taken as the result.

4.3.3 Experiment

We evaluated two different SR reconstruction techniques with respect to a raw input (LR) sequence (40 frames) containing one triangle: 1) a simple 'Shift & Add'(S&A) fusion [18] with zoom-factor 1, which only has a noise reducing effect and 2) a Hardie reconstruction with zoom-factor 2. The S&A fusion is performed only on a LR sequence without moving objects (background), while Hardie is performed on a sequence with moving objects (8x8 and 16x16 LR pixels) as well. In the latter case the SR performance is measured as well on the background as on the moving object. To obtain one point in the TOD curve (see Figure 4.6) we processed for each triangle orientation 6 different sizes with 8 different subpixel positions and noise realizations. This means that each point in the TOD curve results from 192 different observations.

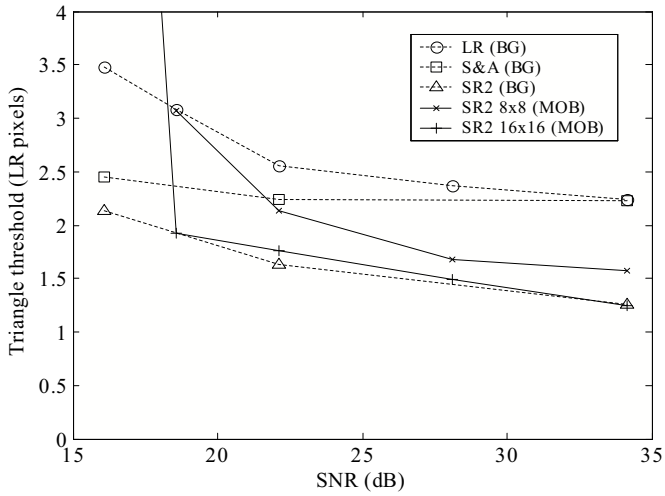


Figure 4.6: *TOD curves of the evaluated SR reconstruction techniques. Dotted lines represent SR on background and solid lines SR on moving objects.*

4.4 Results

The resulting TOD curves from the experiment are depicted in Figure 4.6. SNR is defined here as $\text{SNR} = 20\log_{10}(\Delta I/\sigma_n)$ with ΔI the image intensity range and σ_n the standard deviation of the additive Gaussian noise. Note that the triangle threshold (LR pixels) is inversely proportional to the SR performance (smaller threshold is higher performance). From Figure 4.6 it can be seen that the absolute SR performance improvement of Hardie SR2 w.r.t. LR on the background is approximately constant for the measured SNRs. Inspecting the performance of SR2, LR and S&A shows that 1) for low SNRs the improvement is mainly due to noise reduction (S&A), whereas 2) for high SNRs the improvement is completely due to an increase in resolution. Furthermore it can be seen that the SR performance on moving objects decreases with decreasing object size and that for middle and high SNRs the SR performance on large objects is comparable with the SR performance on the background.

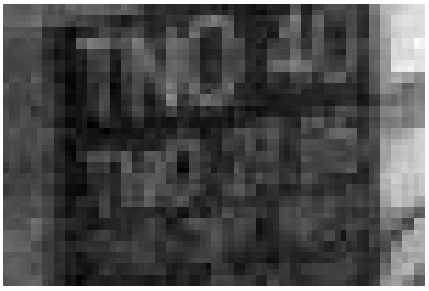
To show the performance of our framework on real data, we processed an image sequence containing a moving van (containing no triangles) captured with an infrared (IR) camera (see Figure 4.7). Carefully studying this result, more detail is e.g. visible at the trees in the background and on the center of the van. Note that at the boundary of the van and the background LR pixels are merged.

4.5 Conclusions

We can conclude that our framework has a comparable SR performance for movings objects and background under the conditions that 1) objects are 'large', e.g. 16x16 pixels, *and* 2) the SNR is not too small. For small moving objects the amount of information inside the object is too small to do a sufficient registration. The decrease in SR performance for moving objects for decreasing SNRs can be explained by a decreasing performance of the *registration* and the *moving object detection* part of our framework. A processed image sequence captured with an infrared camera, shows that the proposed framework is also performing well on real data.

(a) one of 80 LR frames (400×400)

(b) result after SR with zoomfactor 2



(c) Detail of LR frame in (a)

(d) Detail of SR result in (b)

Figure 4.7: *Two times SR reconstruction of an Infrared image sequence containing a moving object (minibus). Note that (c) and (d) are contrast enhanced for visualization.*

Chapter 5

Super-resolution reconstruction of small moving objects in simulated data

Abstract

Moving objects are often the most interesting parts in image sequences. When images from a camera are undersampled and the moving object is depicted small on the image plane, processing of the image sequence afterwards may help to improve the visibility / recognition of the object. This chapter addresses this subject and presents an approach which performs Super-Resolution (SR) specifically on small moving objects using a polygon-based object description. This approach is not bound to a finite pixel grid and has the advantage that less unknowns have to be estimated in the optimization process in comparison to a standard pixel-based SR algorithm. This chapter describes the setup of the proposed polygon-based SR algorithm and shows its superior performance in comparison to a pixel-based SR algorithm.

5.1 Introduction

Although the concept of Super-Resolution (SR) already exists for more than 20 years [59], not much attention is given to a specific case: SR reconstruction on

[1]This chapter has been published in A.W.M. van Eekeren, K. Schutte, O.R. Oudegeest and L.J. van Vliet, Super-Resolution on moving objects using a polygon-based object description, in *Proc. 13th Annual Conference of the Advanced School for Computing and Imaging*, pp.317–321, ASCI, 2007. [15]

moving objects. In [13] we presented an algorithm which simultaneously performs SR on the foreground and background of a scene. Both are processed separately with a standard pixel-based SR algorithm. This algorithm works fine if a moving object (foreground) is relatively large (the number of boundary pixels is small in comparison to the total number of object pixels) and contains enough inner structure to permit precise registration. However, if an object is small (the number of boundary pixels comprises more than fifty percent of the total number of object pixels) and the object itself contains hardly any internal structure, a standard pixel-based SR approach will fail. On one hand there is not enough gradient information to perform a proper registration of the object and on the other hand a standard pixel-based SR approach makes an error across the object boundary (assuming a cluttered background). This boundary region consists of so-called 'mixed pixels', which contain partly information from the object and partly from the background.

To tackle the aforementioned problems we propose to perform SR on small moving objects using a polygon-based object description. Assuming rigid objects with homogeneous intensity that move (constant speed and acceleration are assumed) along a known trajectory through the real world, a proper registration is done by fitting a model trajectory through the object's center of mass in each frame. The boundary of a moving object is modeled with a polygon description. With this model a relation is obtained between the position of the edges of the polygon and the estimated intensities of the mixed pixels. By minimizing the model error between the measured intensities and the estimated intensities in a least squared error sense, which is similar to conventional pixel-based SR algorithms, a sub-pixel accurate polygon description is obtained. The performance of the proposed polygon-based SR algorithm is compared to a conventional pixel-based SR algorithm using a quantitative measure obtained from Triangle Orientation Discrimination (TOD).

The chapter is organized as follows. In section 5.2 the setup of the proposed polygon-based SR algorithm is presented. Section 5.3 describes the TOD-based performance measurements of the polygon-based SR algorithm compared to a conventional pixel-based SR algorithm and results are presented. In the last section conclusions are drawn.

5.2 Algorithm

Our algorithm can be split up into three parts: 1) obtaining a background model and moving object detection, 2) model-based trajectory fitting for object registration and 3) obtaining a sub-pixel accurate polygon description by solving an inverse problem. The latter is the main part of the algorithm and will be explained first.

5.2.1 Polygon description

Given that a proper detection (object mask sequence) and registration (object shift vector $\vec{t}$) of a moving object is obtained, a polygon description (ordered set of vertices) of the boundary of a moving object can be estimated. A flow diagram of the polygon description part is depicted in Figure 5.1.

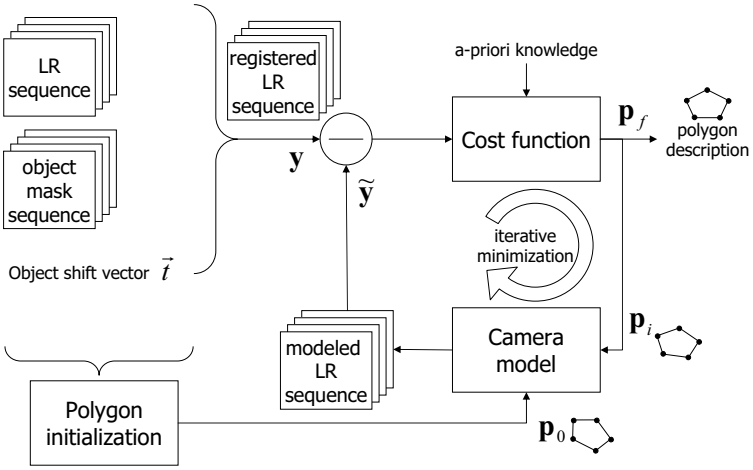


Figure 5.1: *Flow diagram of the polygon description part of the algorithm.*

Polygon initialization

The optimization scheme depicted in Figure 5.1 must be initialized with an initial polygon description of the object. This is done by first applying a pixel-based SR algorithm to the object mask sequence and threshold that result. This so-called SR object mask is used to determine the vertices of the initial polygon description $\mathbf{p}_0$. In this chapter the centre of gravity and the size of the SR object mask are used to construct $\mathbf{p}_0$ as an equilateral triangle.

For larger and more complex shapes, the boundary pixels of the SR object mask can be selected as vertices of $\mathbf{p}_0$. To simplify this polygon description (and limit the number of unknowns) a Douglas-Peucker algorithm can be used [12].

Camera model

The Camera model maps a polygon description to an estimated (modeled) Low-Resolution (LR) image sequence. The estimated intensity of each LR pixel $\tilde{y}_m$ is calculated given a polygon description $\mathbf{p}$ and assuming a uniform rectangular pixel support (5.1).

$$\tilde{y}_m(\mathbf{p}) = I_{bg}A_{bg} + I_{fg}A_{fg} \tag{5.1}$$

Here I_{bg} is the local background intensity, which is assumed to be known from a pixel-based SR method. The foreground (object) intensity I_{fg} is assumed to be constant and known. A_{bg} is the partial area of the support of pixel m that overlaps the background given the polygon description $\mathbf{p}$. Hence, the partial area covering the foreground is defined as: $A_{fg} = 1 - A_{bg}$. The width (w) of the pixel support is also assumed constant.

Cost function

The cost function that is minimized for $\mathbf{p}$ is defined as:

$$C = \sum_{m}^{N} \left(y_m - \tilde{y}_m(\mathbf{p})\right)^2 + \mu... \tag{5.2}$$

where the first part (summation) minimizes the model error for all pixels m and the last part, which is left empty, can be used to regularize the process by incorporating a-priori knowledge (e.g. a model of the expected object and/or a penalty for self-intersecting polygons). y_m are the measured intensities of the registered LR pixels and $\tilde{y}_m$ are the corresponding estimated intensities (5.1). The actual minimization of the cost function is done with the Levenberg-Marquardt method with a mixed quadratic and cubic line search procedure.

To show qualitatively the performance of the polygon description part of the algorithm, we fitted a polygon with 5 vertices to a simulated LR image sequence (10 frames) containing a sub-pixel shifted binary object. The LR image sequence is noise free and is simulated with the same camera model as described in the previous section. The resulting polygon (red) is superimposed on the first LR frame (see Figure 5.2).

5.2.2 Background modeling and moving object detection

Our detection of moving objects is based on the assumption that a moving object deviates from a static background. Ideally, a static background is estimated using a robust pixel-based SR algorithm [21, 49]. Such robust SR algorithms provide

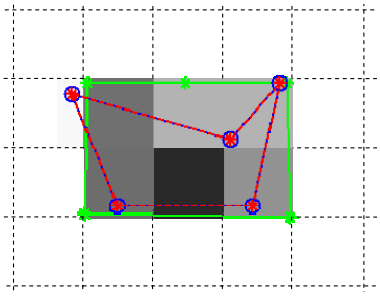


Figure 5.2: *Polygon with 5 vertices fitted to undersampled data (10 frames) of a binary object. The blue circles represent the ground-truth vertices used to simulate the LR data. The green stars are the vertices used as initialization (by hand) and the red stars are the final vertices found with the proposed polygon-based SR algorithm. No a-priori knowledge is incorporated in the cost function used in this example.*

a proper background estimate even in the presence of small moving objects. In the experiments as described in section 5.3.2 a median of all image frames is used as robust background estimate; in this case it is allowed as in the experiment no motion in the background scene is present. The SR background model, after shifting and down-sampling with a camera model, can be compared to each frame of the image sequence. After thresholding these difference images and removal of spurious detections with morphological operations, an object mask for each frame results.

5.2.3 Registration

With the object masks, obtained from the detection part, a global position of the object in each frame is known. For performing SR however, a very precise sub-pixel registration is needed. When sufficient internal structure is present, gradient-based registration can be performed. In the setting of *small* moving objects this is usually not the case and therefore another approach is needed.

Assuming that the contrast between object and background is sufficient, in each frame the Center Of Gravity (COG) of the object is calculated. When the motion model of the moving object is known, such a model is fit in a robust way to the calculated COGs. We assume a linear motion model in the real world, which seems realistic given the setting of small moving objects: the objects are far away from the observer (viewer) and will have a small displacement in the image plane due the high frame rate of today's image sensors. When enough (≥ 20) frames

are available, a robust fit to the COGs results in a sufficient precise and accurate sub-pixel registration of the moving object.

5.3 Experiments

To test the performance of our polygon-based SR algorithm for moving objects, it is compared in a quantitative way (TOD) with the performance of a well-performing pixel-based SR algorithm developed by Hardie [26]. The performance is tested on simulated data containing a small moving object on a cluttered background.

5.3.1 Triangle orientation discrimination

The TOD method is an evaluation method designed for system performance of a broad range of imaging systems [7], but it can also be used to test the performance of image enhancement techniques [6, 14]. It is based on the observer task to discriminate four different oriented equilateral triangles (see Figure 5.3).

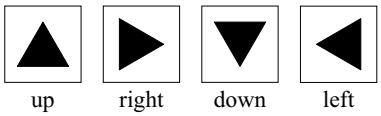


Figure 5.3: *The four different stimuli used in the TOD method.*

In the experiments two automatic observers are used: one for the pixel-based SR result I_{SR} and one for the polygon-based SR result $\mathbf{p}_f$. The first one makes its choice $\hat{\theta}$ based on the minimum MSE between the triangle in I_{SR} and a pixel-based triangle model M:

$$\hat{\theta} = \min_{\theta,s}\{\frac{1}{N}\sum_{\vec{x}}(I_{SR}(\vec{x};\theta_k,s_k) - M(\vec{x};\theta,s))^2\}. \tag{5.3}$$

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 k denotes one member of these sets. Although (5.3) 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 second observer is used for the polygon-based SR results ($\mathbf{p}_f$) and minimizes a similar error as in (5.3):

$$\hat{\theta} = \min_{\theta,s}\{\int_x (\mathbf{p}_f(\theta_k, s_k) - \mathbf{p}_{tr}(\theta, s))^2 dx\}, \tag{5.4}$$

where $\mathbf{p}_{tr}$ indicates a polygon model of an equilateral triangle with orientation θ and size s. The probability of a correct observer response increases with the triangle size. In [7] it is shown that this relationship can be described with a Weibull distribution:

$$p_c(x) = 0.25 + 0.75/1.5^{(\alpha/x)^\beta}, \tag{5.5}$$

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 5.4. From this fit the triangle size that corresponds with a 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.

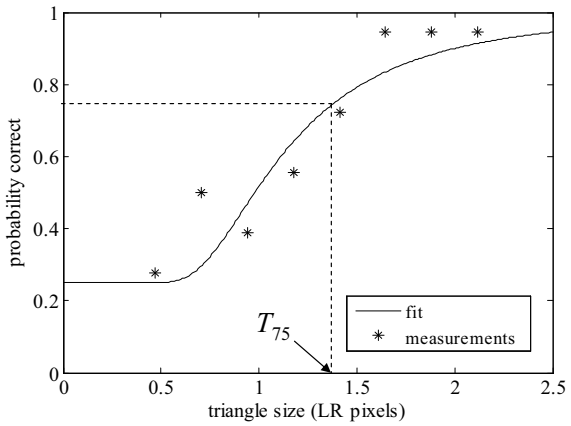


Figure 5.4: *Example of a Weibull fit to measured probability correct observer responses.*

5.3.2 Comparison polygon versus pixel-based approach

The setup of the experiment is depicted in Figure 5.5.

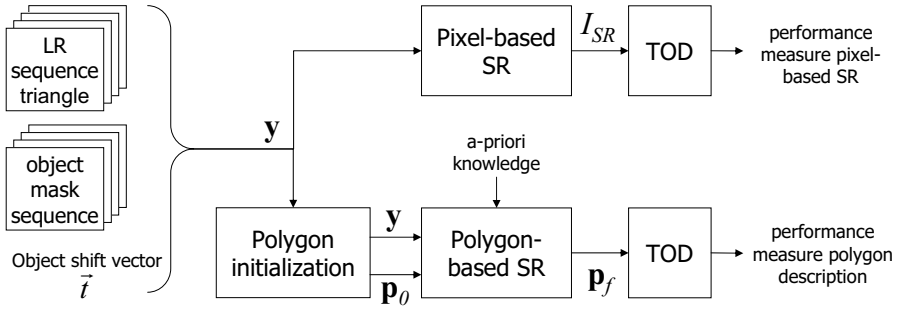
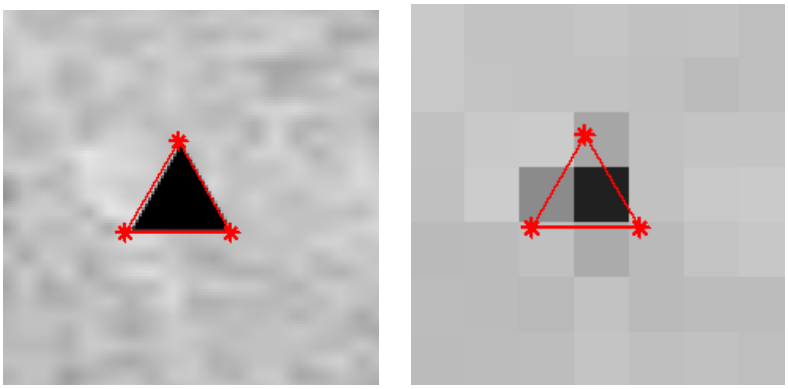


Figure 5.5: *Flow diagram of the comparison between a conventional pixel-based SR algorithm and the proposed polygon-based SR algorithm on moving objects. The performance measure is obtained with the TOD method.*

Here, the 'LR sequence triangle' is a simulated, under-sampled, image sequence containing a moving triangle with homogeneous intensity. The LR data is obtained from Hyper-Resolution (HYR) data, which has a 17× denser sampling grid, through a simulated camera model. This camera model simulates the camera's sensor (100% fill-factor), but no lens blurring is taking in to account. Note that both the pixel-based SR algorithm as well as the polygon-based SR algorithm are implemented in such a way that *no* model errors are introduced. To be able to apply the TOD method, different LR sequences with various triangle sizes and SNR's are simulated. An example of a HYR image and its corresponding LR image is depicted in Figure 5.6. On both images the sub-pixel accurate polygon description as found by our SR algorithm is superimposed in red.

As is depicted in the upper part of Figure 5.5, the detected moving triangle data together with the measured object positions ($\mathbf{y}$) is processed with a pixel-based SR algorithm (Hardie). Afterwards the orientation of the triangle is determined with the TOD method (5.3). In the lower part of Figure 5.5 the data is processed with the proposed polygon-based SR algorithm, which allows us to use a-priori knowledge about the object. The polygon-based method is tested in two different modes: one using *no* a-priori knowledge and one using a-priori knowledge. The first mode implies that the regularization term (μ...) in (5.2) is empty.

In the second mode a strong restriction is set on the interior angles of the polygon description. Because it is known that an equilateral triangle must be found, the interior angles must be approximately 60 degrees (assuming three vertices). The functional that is minimized for this mode is:



(a) (b)

Figure 5.6: *(a) Part of original HYR image containing a moving triangle of 32 pixels. (b) Corresponding LR image (triangle size is here reduced to 31/17 = 1.88 pixels). On both images the sub-pixel accurate polygon description as found by our SR algorithm is superimposed in red.*

$$C = \sum_{m}^{N} \left(y_m - \tilde{y}_m(\mathbf{p})\right)^2 + \cdots \\ \mu(1 - e^{-\sum_j^P (\alpha_j - \alpha_0)^2 / 2\sigma_\alpha^2}). \tag{5.6}$$

Here, α_j is the inner angle of polygon $\mathbf{p}_i$ at vertex j and α_0 is the preferred angle. σ_α expresses the steepness of the joint probability density function. The value of σ_α should be seen in relation to the preferred angle α_0. In our experiments $\alpha_0 = \pi/3$ (equilateral triangle constraint) and $\sigma_\alpha = \pi/9$ are used. Note that a flat prior is used for the angle difference.

Furthermore, for both modes the foreground intensity and the width of the pixel support (w = 1 LR pixel) are assumed to be known. Also the initialization step is the same: according to the centre of mass and the size of the SR object mask four different oriented equilateral triangles are used as initial polygon description $\mathbf{p}_0$. This is done to prevent that the optimization process gets stuck in a local minimum. Note that already some a-priori knowledge is assumed by initializing with three vertices. The polygon description ($\mathbf{p}_f$) that results in the smallest value of functional C is used for the TOD measurement (5.4).

5.3.3 Results

The results obtained from the comparison experiment are depicted in Figure 5.7. Note that a smaller T_{75} indicates a better performance. The Signal-to-Noise Ratio (SNR) is defined here as $20\log_{10}(|\bar{I}_{bg} - I_{fg}|/\sigma_n)$, with $\bar{I}_{bg}$ the mean intensity of the local background, I_{fg} the foreground intensity and σ_n the standard deviation of the noise.

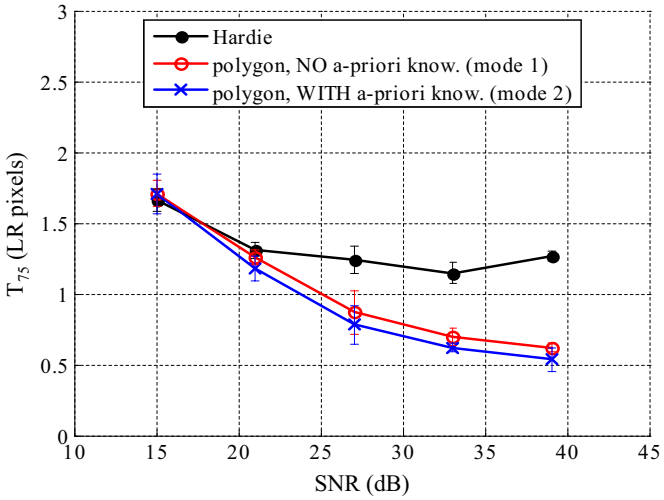


Figure 5.7: *TOD performance comparison between a polygon and a pixel-based approach on simulated data (effective uniform rectangular blur of 1 LR pixel, 20 frames). Black line: Hardie, zoom 4, $w = 1$, optimized λ. Red line: polygon description with 3 vertices using no a-priori knowledge (mode 1), Blue line: polygon description with 3 vertices using a-priori knowledge (mode 2).*

5.4 Conclusions

From the obtained results the following conclusions can be drawn:

- For high and medium SNR's our SR algorithm using a polygon-based object description performs significantly better than a conventional well-performing pixel-based SR algorithm according to the TOD measure considering a small moving object on a cluttered background.

- Although the main improvement in performance is obtained *without* using a-priori knowledge, a little more can be gained using some a-priori knowledge.
- For low SNR's the performance is comparable to the pixel-based SR algorithm. The tendency of the performance curves in Figure 5.7 shows us that the pixel-based method seems to be performing better when SNR < 15 dB.

Chapter 6

Super-resolution reconstruction of small moving objects in real-world data

Abstract

This chapter presents an approach which performs multi-frame super-resolution (SR) reconstruction on small moving objects, i.e. objects that consist solely of boundary pixels. The method improves the visibility, detection as well as automatic recognition of small moving objects. It is capable for this task because it models explicitly the space-time variant partial area effect of 'mixed pixels', which consist of both the moving object and the local background of the scene. The presented approach simultaneously estimates a subpixel precise polygon boundary as well as a high-resolution intensity description of a small moving object. Experiments on simulated and real-world data show excellent performance of the proposed multi-frame SR reconstruction method.

6.1 Introduction

In many applications the most interesting events are related to changes occurring in the scene: e.g. moving persons or moving objects. In this chapter we focus on multi-frame Super-Resolution (SR) reconstruction of small moving objects,

[1]This chapter has been submitted as A.W.M. van Eekeren, K. Schutte and L.J. van Vliet, Multi-Frame Super-Resolution Reconstruction of Small Moving Objects, to *IEEE Trans. Image Processing*, IEEE, Nov. 2008. [16]

i.e. objects that are comprised solely of boundary pixels, in under-sampled image sequences. These so-called 'mixed pixels' depict both the foreground (moving object) and the local background of a scene. Especially for *small* moving objects, resolution improvement is useful.

Multi-frame SR reconstruction[2] improves the spatial resolution of a set of sub-pixel displaced Low-Resolution (LR) images by exchanging temporal information for spatial information. Although the concept of SR reconstruction already exists for more than 20 years [59], only little attention is given to SR reconstruction on moving objects. In [5, 13, 17, 19, 27, 64] this subject has been addressed.

Although [5] and [19] apply different SR reconstruction methods, iterated-back-projection [31] and projection onto convex sets [46] respectively, both use a validity map in their reconstruction process. This makes these methods robust to motion outliers. Both methods perform well on large moving objects (the number of mixed pixels is small in comparison to the total number of object pixels) with a simple motion model, such as translation. Hardie et al. [27] use optical flow to segment a moving object and subsequently apply SR reconstruction to it. In the experiment they present, the background is static and SR reconstruction is done solely on a masked large moving object.

In previous work [13] we presented an algorithm that performs, after segmentation, simultaneously SR reconstruction on a large moving object and background using Hardie's SR reconstruction [27]. However, in [13] no SR reconstruction is applied to the boundary (mixed pixels) of the moving object because of a cluttered background. In [17] we presented the first results of SR reconstruction of small moving objects with simulated data. However, at that time no experiments were done on real-world data.

In [64] SR reconstruction is performed on moving vehicles of approximately 10 by 20 pixels. For object registration a trajectory model is used in combination with consistency of local background and vehicle. However, in the SR reconstruction approach no attention is given to mixed pixels.

An interesting subset of moving objects are faces. Efforts done in that area using SR reconstruction include [29] and [66], in which the modeling of complex motion is a key element. However, the faces in the used LR input images are far larger than the small objects in this chapter.

When a moving object is small (it consists solely of mixed pixels) and the background is cluttered, even the state-of-the-art pixel-based SR reconstruction methods mentioned above will fail. Any pixel-based SR reconstruction method makes an error at the object boundary, because it is unable to separate the space-time variant background and foreground information within a mixed pixel.

[2] In the remainder of this chapter 'SR reconstruction' refers to 'multi-frame SR reconstruction'.

To tackle the aforementioned problem we propose to perform SR reconstruction on small moving objects using a simultaneous boundary and intensity estimation of a moving object. Assuming rigid objects that move with constant speed through the real world, a proper registration is done by fitting a trajectory through the object's location in each frame. The boundary of a moving object is modeled with a subpixel precise polygon and the object's intensities are modeled on a High-Resolution (HR) pixel grid.

After applying SR reconstruction to the background, the local background intensities are known on a HR grid. When the intensities of the moving object and the position of the edges of the boundary are known as well, the intensities of the mixed pixels can be calculated. By minimizing the model error between the measured intensities and the estimated intensities similar to state-of-the-art pixel-based SR algorithms, a subpixel precise boundary and intensity description of the moving object are obtained.

Especially for small moving objects our approach improves the recognition significantly. However, the use of our SR reconstruction method is not limited to small moving objects. It can also be used to improve the resolution of boundary regions of larger moving objects. This might give an observer some useful extra information about the object.

The chapter is organized as follows. First, we model the relationship between the real-world and the data observed by an electro-optical system. In Section 6.3 the proposed SR reconstruction method for small moving objects is presented. Section 6.4 describes experiments on simulated and real-world data, and shows the performance of the proposed method. In the final section conclusions are presented.

6.2 Real-world data description

This section describes the model of the real-world at a 2D High-Resolution (HR) grid and how this is observed by an optical camera system.

6.2.1 2D high-resolution scene

We model a camera's field-of-view at frame k as a 2D HR image, consisting of R pixels, sampled at or above the Nyquist rate without significant degradation due to motion, blur or noise. Let us express this image in lexicographical notation as the vector $\mathbf{z}_k = [z_{k,1}, ..., z_{k,R}]^T$. $\mathbf{z}_k$ is constructed from a translated HR background intensity description $\mathbf{b} = [b_1, ..., b_V]^T$, consisting of V pixels, and a translated HR foreground intensity description $\mathbf{f} = [f_1, ..., f_Q]^T$, consisting of Q pixels. This is depicted in the left part of Figure 6.1. Note that the foreground $\mathbf{f}$ has a different apparent motion with respect to the camera than the background $\mathbf{b}$.

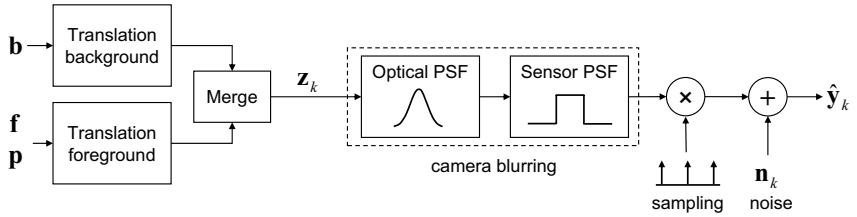


Figure 6.1: *Flow diagram of the construction of a 2D HR scene* $\mathbf{z}_k$ *at frame* k *and the degradation to a LR frame* $\hat{\mathbf{y}}_k$ *via a camera model.*

The foreground (small moving object) is not solely described by its intensity description $\mathbf{f}$, but also by a subpixel precise polygon boundary $\mathbf{p} = [v_{1x}, v_{1y}, ..., v_{Px}, v_{Py}]^T$ with P the number of vertices. The following assumptions are made about a moving object: 1) the aspect angle of the object stays the same and 2) the object is moving at constant speed. These are realistic assumptions if a moving object is far away and given the high frame rate of today's image sensors.

At frame k the HR background and the HR foreground are *translated* and *merged* to the 2D HR image $\mathbf{z}_k$ in which the r^{th} pixel is defined by:

$$\begin{aligned} z_{k,r} &= c_{k,r}(\mathbf{p})\tilde{f}_{k,r} + (1 - c_{k,r}(\mathbf{p}))\tilde{b}_{k,r} \\ &= c_{k,r}(\mathbf{p}) \sum_{q=1}^{Q} t_{k,r,q} f_q + (1 - c_{k,r}(\mathbf{p})) \sum_{v=1}^{V} s_{k,r,v} b_v, \end{aligned} \tag{6.1}$$

for $k = 1, 2, ..., K$ and $r = 1, 2, ..., R$.
Here, K is the number of frames. The summation of weights $t_{k,r,q}$ represent the translation of foreground pixel f_q to $\tilde{f}_{k,r}$ by bilinear interpolation and in a similar way the summation of $s_{k,r,v}$ translates background pixel b_v to $\tilde{b}_{k,r}$. The weight $c_{k,r}$ represents the foreground contribution at pixel r in frame k depending on the polygon boundary $\mathbf{p}$. The foreground contribution varies between 0 and 1, so the corresponding background contribution is then by definition equal to $(1 - c_{k,r})$.

A visualization of merging the translated background, $\tilde{\mathbf{b}}_k = [\tilde{b}_{k,1}, ..., \tilde{b}_{k,R}]^T$, and the translated foreground, $\tilde{\mathbf{f}}_k = [\tilde{f}_{k,1}, ..., \tilde{f}_{k,R}]^T$, is depicted in Figure 6.2. The polygon boundary $\mathbf{p}$ defines the foreground contributions $\mathbf{c}_k$ and the background contributions $(1 - \mathbf{c}_k)$ in HR frame k.

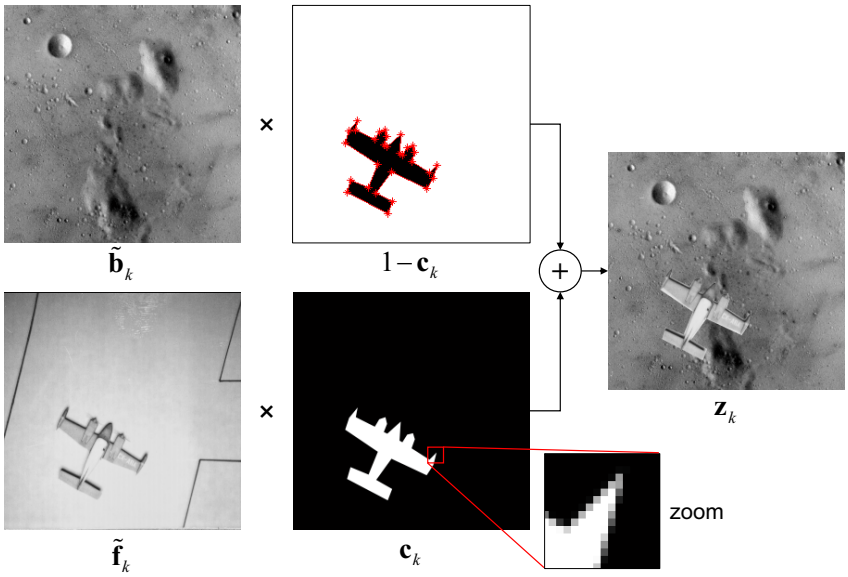


Figure 6.2: *Flow diagram of merging foreground and background to obtain HR image* $\mathbf{z}_k$*. The polygon boundary* $\mathbf{p}$ *is plotted on top of the background contributions* $(1-\mathbf{c}_k)$ *for visualization. Note that in* $\mathbf{c}_k$ *and* $(1-\mathbf{c}_k)$ *black* $(=0)$ *indicates* no *contribution, white* $(=1)$ *indicates* full *contribution and grey indicates a partial contribution.*

6.2.2 Camera model

Using the 2D HR image $\mathbf{z}_k$, the LR camera frame $\hat{\mathbf{y}}_k$ is constructed by applying the physical properties of an optical camera system.

Blurring: The optical Point-Spread-Function (PSF), together with the sensor PSF, will cause a blurring at the image plane. In this chapter the optical blur is modeled by a Gaussian function with standard deviation σ_{psf}. The sensor blur is modeled by a uniform rectangular function representing the fill-factor of each sensor element. A convolution of both functions represents the total blurring function.

Sampling: The sampling as depicted in Figure 6.1 relates to the sensor pitch.

Noise: The temporal noise in the recorded data is modeled by additive, independent and identically distributed Gaussian noise samples $\mathbf{n}_k$ with standard deviation σ_n. For the recorded data used, independent additive Gaussian noise is a sufficiently accurate model. Other types of noise, like fixed pattern noise and

bad pixels, are not modeled explicitly.

All in all, the observed m^{th} LR pixel from frame k is modeled as follows:

$$\hat{y}_{k,m} = \sum_{r=1}^{R} w_{k,m,r} z_{k,r} + \eta_{\sigma_n} = \tilde{y}_{k,m} + \eta_{\sigma_n}, \tag{6.2}$$

for $k = 1, 2, ..., K$ and $m = 1, 2, ..., M$.
Here, M is the number of LR pixels in $\hat{\mathbf{y}}_k$. The weight $w_{k,m,r}$ represents the contribution of HR pixel $z_{k,r}$ to estimated LR pixel $\tilde{y}_{k,m}$. Each contribution is determined by the blurring and sampling of the camera. η_{σ_n} represents an additive, independent and identically distributed Gaussian noise sample with standard deviation σ_n.

6.3 SR method description

This section presents the method to perform SR reconstruction on small moving objects based on the inversion of the forward model of section 6.2.

Our method can be split into three parts: 1) constructing a HR background and detecting the moving object, 2) fitting a trajectory model to the detected instances of the moving object through the image sequence to obtain subpixel precise object registration, 3) obtaining a HR object description, containing a subpixel precise boundary and a HR intensity description, by solving an inverse problem. We explain the latter part first, because this is the most innovating part of our method.

6.3.1 High-resolution object reconstruction

To find an optimal HR object description (consisting of a polygon boundary $\mathbf{p}$ *and* an intensity description $\mathbf{f}$), we propose to minimize the following cost function:

$$\begin{aligned} C_{\mathbf{p},\mathbf{f}} \quad = \quad & \frac{1}{KM\sigma_n^2} \sum_{k=1}^{K} \sum_{m=1}^{M} \left(y_{k,m} - \tilde{y}_{k,m}(\mathbf{p},\mathbf{f},\mathbf{b})\right)^2 + \\ & \frac{\lambda_f}{Q} \sum_{h,v=\{0,1\}}^{h+v=1} \|\mathbf{f} - S_x^h S_y^v \mathbf{f}\|_H + \\ & \lambda_p \left(\frac{\|\mathbf{p}\|}{P}\right)^2 \sum_{p=1}^{P} \Gamma_p(\mathbf{p}), \end{aligned} \tag{6.3}$$

where the first summation term represents the normalized data misfit contributions for all pixels k, m. Normalization is performed with the total number of LR pixels and the noise variance. Here, $y_{k,m}$ are the measured intensities of the observed LR pixels and $\tilde{y}_{k,m}$ are the corresponding estimated intensities obtained using the forward model of section 6.2. Although the estimated intensities $\tilde{y}_{k,m}$ are also dependent on the background $\mathbf{b}$, only $\mathbf{p}$ and $\mathbf{f}$ are varied to minimize (6.3). The HR background $\mathbf{b}$ is estimated in advance as described in section 6.3.2.

Minimization of (6.3) is an ill-posed problem, therefore we apply regularization to the foreground intensities and to the polygon boundary. The second term of the cost function $C_{\mathbf{p},\mathbf{f}}$ regularizes the amount of intensity variation within the object according to a criterion similar to the Total Variation (TV) criterion [38]. Here, S_x^h is the shift operator that shifts $\mathbf{f}$ by h pixels in horizontal direction and S_y^v shifts $\mathbf{f}$ by v pixels in vertical direction.

The actual minimization of the cost function is done in an iterative way with the Levenberg-Marquardt algorithm [42]. This optimization algorithm assumes that the cost function has a first derivative that exists everywhere. However, the L1-norm used in the TV criterion does not satisfy this assumption. Therefore we introduce the hyperbolic norm ($\|.\|_H$):

$$\|\mathbf{x}\|_H = \sum_i \left(\sqrt{x_i^2 + \alpha^2} - \alpha \right) \tag{6.4}$$

This norm has the same properties as the L1-norm for large values ($x_i \gg \alpha$) *and* it has a first (and second) derivative that exists everywhere. For the experiments performed $\alpha = 1$ is used.

The third term regularizes the variation of the polygon boundary $\mathbf{p}$. Regularization is needed to penalize unwanted protrusions, such as spikes, which cover a very small area compared to the total object area. This constraint is embodied by the measure Γ_p, which is small when the polygon boundary $\mathbf{p}$ is smooth:

$$\Gamma_p = 1/A_p \quad \text{with} \quad A_p = 0.5 a_p b_p \sin(\gamma_p/2). \tag{6.5}$$

Γ_p is the inverse of A_p, which is the area spanned by the edges (a_p and b_p) at vertex $\mathbf{v}_p$ and half the angle between those edges $\gamma_p/2$ as indicated by the right part of (6.5).

From example (a) in Figure 6.3 it is clear why the area is calculated with half the angle $\gamma_p/2$: if we would take the full angle γ_p, A_p would be zero, which would result in $\Gamma_p = \infty$. Example (b) shows that the measure Γ_p will be very large for small angles. Note that this measure is large as well for $\gamma_p \approx 2\pi$ (inward pointing spike).

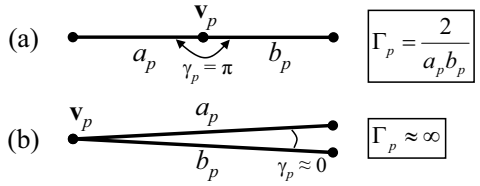


Figure 6.3: *Two examples (a) and (b) of the calculation of Γ_p at vertex $\mathbf{v}_p$ of polygon $\mathbf{p}$. Γ_p is minimal in (a) when $\gamma_p = \pi$ rad and in (b) Γ_p is almost infinity.*

Note that in (6.3) normalization is performed on Γ_p with the square of the mean edge length $(\|\mathbf{p}\|/P)^2$, with P the number of vertices and $\|\mathbf{p}\|$ the total edge length of $\mathbf{p}$. This normalization prevents extensive growth of edges in order to minimize Γ_p.

As mentioned above, the actual minimization of the cost function is performed in an iterative way with the Levenberg-Marquardt (LM) algorithm [42]. To allow this, we put the cost function in (6.3) in the LM framework, which expects a format like $\min_\beta \sum_i^N (x_i - \tilde{x}_i(\beta))^2$ where x_i is the measurement and $\tilde{x}_i(\beta)$ is the estimate depending on parameter β.

In a straightforward case a vector with all residual values, e.g. $[\overbrace{..., (x_i - \tilde{x}_i), ...}^{N}]$, forms the input of the LM algorithm. In our case it is slightly more complex to construct such a vector, which looks like:

$$\left[\overbrace{..., \frac{1}{\sqrt{KM}\sigma_n}(y_{k,m} - \tilde{y}_{k,m}), ...}^{KM}, \overbrace{..., \sqrt{\frac{\lambda_f}{Q}(\sqrt{(f_i - f_j)^2 + \alpha^2} - \alpha)}, ...}^{2Q}, \overbrace{..., \frac{\|\mathbf{p}\|}{P}\sqrt{\lambda_p \Gamma_p}, ...}^{P}\right], \tag{6.6}$$

with the letters on top indicating the number of elements used in each part of the cost function, which makes the total size of this vector $[1 \times (KM + 2Q + P)]$.

The cost function in (6.3) is iteratively minimized to find simultaneously an optimal $\mathbf{p}$ and $\mathbf{f}$. A flow diagram of this iterative minimization procedure in steady state is depicted in Figure 6.4. Here the *Cost function* is defined in (6.3) and the *Camera model* is defined in (6.1) and (6.2). Note that the measured data used for the minimization procedure is a small Region Of Interest (ROI) around the moving object in each frame.

The optimization scheme depicted in Figure 6.4 is initialized with an object

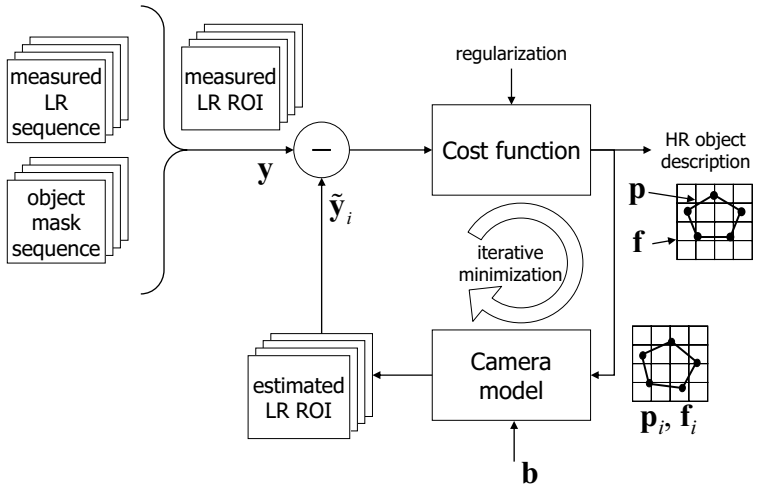


Figure 6.4: *Flow diagram of estimating a high-resolution description of a moving object (*$\mathbf{p}$ *and* $\mathbf{f}$*).* $\mathbf{y}$ *indicates the measured intensities in a region of interest containing the moving object in all frames and* $\tilde{\mathbf{y}}_i$ *are the corresponding estimated intensities at iteration* i*. Note that the initial HR object description (*$\mathbf{p}_0$ *and* $\mathbf{f}_0$*) is derived from the measured LR sequence and the object mask sequence.*

boundary $\mathbf{p}_0$ and an object intensity description $\mathbf{f}_0$. These can be obtained in several ways; we have chosen to use a simple and robust initialization method.

The initial object boundary is obtained by first calculating the median (frame-wise) width and the median (frame-wise) height of the mask in the object mask sequence. Afterwards we construct an ellipse object boundary with the previous calculated width and height. At initialization the vertices are evenly distributed over the ellipse. The number of vertices is fixed during minimization.

For initializing the object intensity distribution $\mathbf{f}_0$, a homogeneous intensity is assumed. This intensity is initialized with the median intensity over all masked pixels in the measured LR sequence.

Furthermore, the optimization procedure is performed in two steps. The first step consists of the initialization described above and 5 iterations of the LM algorithm. After this step it is assumed that the found object boundary and intensity description are approaching the global minimum. However, to improve the estimation of the object intensities near the object boundary, a second initialization step is proposed. In this step all intensities of HR foreground pixels ($\mathbf{f}_5$) which are close to *and* located completely within the object boundary are propagated

outwards. Afterwards, 15 more iterations of the LM algorithm are performed to let $\mathbf{p}$ and $\mathbf{f}$ converge.

6.3.2 Background SR reconstruction and moving object detection

The detection of moving objects is based on the assumption that a moving object deviates from a static background. In previous work [11] we have shown that for a LR image sequence containing a moving point target, a robust pixel-based SR reconstruction method is effective in estimating a HR background and detecting the moving point target. The same approach is applied to the case of small moving objects. However, the relative motion compared to the background must be sufficient given the number of frames. Assuming K LR frames containing a moving object of width W (LR pixels), the apparent lateral translation must be more than $2(W+1)/K$ LR pixels/frame for a proper background reconstruction.

In the literature several robust SR reconstruction methods are described [21, 49, 72]. We use the method developed by Zomet et al. [72], which is robust to intensity outliers, such as small moving objects. This method uses the same discrete camera model as given in (6.2). Its robustness is introduced by a robust back-projection, that is based on applying a frame-wise median operation instead of a mean operation. The latter one is often applied by non-robust SR reconstruction methods that use Iterated Back Projection [31].

A LR representation of the background, obtained by shifting, blurring and down-sampling of the HR background estimate $\mathbf{b}$, can be compared to the corresponding LR frame of the recorded image sequence:

$$\delta_{k,m} = \left(y_{k,m} - \sum_{r=1}^{R} w_{k,m,r} \tilde{b}_{k,r} \right). \tag{6.7}$$

Here, the weights $w_{k,m,r}$ represent the blur and down-sample operation, $\tilde{b}_{k,r}$ is the r^{th} pixel of the shifted HR background $\mathbf{b}$ in frame k and $y_{k,m}$ is the measured intensity of the m^{th} pixel in frame k. All difference pixels $\delta_{k,m}$ form a residual image sequence in which the moving object can be detected.

First thresholding is performed on the residual image sequence, followed by tracking. Thresholding is done with the *chord method* from Zack et al. [69], which is illustrated by Figure 6.5. With this histogram based method an object mask sequence $\mathbf{m}_T = \delta_{k,m} > T_\delta$ results for $k = 1, 2, ..., K$ and $m = 1, 2, ..., M$ with K the number of observed LR frames and M the number of pixels in each LR frame.

After thresholding, multiple detections may occur in each frame of $\mathbf{m}_T$. We apply tracking to find the most similar detection in each frame to a reference detection. This reference detection is defined by the median width (W_R), the

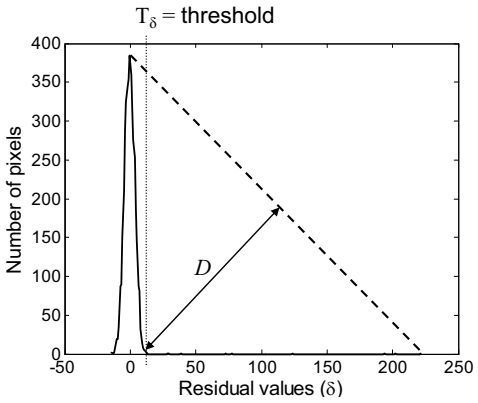


Figure 6.5: *The chord method is based on finding the value of δ that gives the maximum distance D. This value T_δ is used as threshold value.*

median height (H_R) and median residual energy (E_R) of the largest detection in each frame (median is taken frame-wise). Next, we search in each frame k the detection with the smallest normalized Euclidian distance Δ_k (regarding its width $W_{k,i}$, height $H_{k,i}$ and residual energy $E_{k,i}$) to the reference detection:

$$\Delta_k(\hat{i}) = \min_i \left(\sqrt{(\frac{W_{k,i} - W_R}{W_R})^2 + (\frac{H_{k,i} - H_R}{H_R})^2 + (\frac{E_{k,i} - E_R}{E_R})^2} \right), \qquad (6.8)$$

with $\hat{i}$ the index of the detection in frame k with the smallest normalized Euclidian distance to the reference detection. After this tracking step an object mask sequence $\mathbf{m}_{TT}$ results with in each frame at most one detection.

6.3.3 Moving object registration

The object mask sequence $\mathbf{m}_{TT}$, obtained after thresholding and tracking, gives a pixel-accurate indication of the position of the object in each frame. For performing SR reconstruction, a more precise (subpixel) registration is needed. When moving objects contain sufficient internal pixels with sufficient structure or have sufficient contrast with their local background, gradient-based registration [48] can be performed. In the setting of *small* moving objects this is usually not the case and another approach is needed.

When a motion model for a moving object is known, such a model can be fitted to the object positions in time. We assume a constant motion model in the

real world, which seems realistic given the nature of small moving objects: the objects are far away from the observer and will have a small acceleration due to the high frame rate of today's image sensors.

First, an approximately pixel-precise position of the object in each frame is determined by calculating the weighted Center Of Mass (COM) of the masked pixels. The weighted COM of the masked pixels in frame k is defined by

$$\mathbf{a}_k = \frac{1}{\sum_{n=1}^{M} m_n \cdot y_{k,n}} \left[\sum_{n=1}^{M} i_n \cdot m_{k,n} \cdot y_{k,n}, \sum_{n=1}^{M} j_n \cdot m_{k,n} \cdot y_{k,n} \right]^T \tag{6.9}$$

with M the number of LR pixels in frame k, (i_n, j_n) de (x,y)-coordinate of pixel n, $m_{k,n}$ the corresponding mask value (0 or 1) and $y_{k,n}$ is the measured intensity.

To fit a trajectory, all object positions in time must be known to a reference point in the background scene. This is done by adding the previously obtained background translation $\mathbf{s}_k$ to the calculated object position for each frame: $\tilde{\mathbf{a}}_k = \mathbf{a}_k + \mathbf{s}_k$.

To obtain all object positions with subpixel precision, a robust fit to the measured object positions $\tilde{\mathbf{a}}_k$ is performed. Assuming constant motion, all object positions can be described by a reference object position $\mathbf{a}_R$ and a translation $\mathbf{v}$. Both the reference object position and the translation of the object are estimated by minimizing the following cost function:

$$C_{\mathbf{a}_R,\mathbf{v}} = \sum_{k=1}^{K} \left(1 - \exp\left(-\frac{d_k^2(\mathbf{a}_R, \mathbf{v})}{2\sigma_t^2} \right) \right), \tag{6.10}$$

where d_k denotes the Euclidean distance in LR pixels between the measured object position and the estimated object position at frame k:

$$d_k = \left\| \tilde{\mathbf{a}}_k - (\mathbf{a}_R + (k-1)\mathbf{v}) \right\|. \tag{6.11}$$

The cost function in (6.10) is known as the Gaussian norm [63]. This norm is robust to outliers (e.g. false detections in our case). The smoothing parameter σ_t is set to 0.5 LR pixel. Minimizing the cost function in (6.10) with the Levenberg-Marquardt algorithm results in a subpixel precise and accurate registration of the moving object. If e.g. 50 frames ($K = 50$) are used, the registration precision is improved with factor ≈ 7.

6.4 Experiments

The proposed SR reconstruction method for small moving objects is tested on simulated data as well as on real-world captured data. The experiments on simulated data show the performance of our method under varying, but controlled conditions. Real-world data is used to test our method under realistic conditions and to study the impact of changes in object intensities caused by reflection, lens abberations and small changes in aspect ratio of the object along the trajectory.

6.4.1 Test 1 on simulated data

We constructed a simulated under-sampled image sequence containing a small moving car using the camera model depicted in Figure 6.1. Gaussian optical blurring ($\sigma_{psf} = 0.3$ LR pixel) and rectangular uniform sensor blurring (100% fill-factor) are used to model the camera blur and Gaussian distributed noise is added. The car describes a linear trajectory with respect to the background and is modeled with two intensities, which both are above the median background intensity. The low object intensity is exactly in between the median background intensity and the high object intensity. The boundary of the car is modeled by a polygon with 7 vertices.

In Figure 6.6(b) the simulated car is depicted at a HR grid. The car in this image serves as a ground-truth reference for obtained SR reconstruction results. In the LR image sequence the car covers approximately 6 pixels (all mixed pixels) as can be seen in the upper row of Figure 6.6. In the LR domain the Signal-to-Noise Ratio (SNR) of the car with the background is 29 dB and the Signal-to-Clutter Ratio (SCR) is 14 dB. The SNR is defined as:

$$\mathrm{SNR} = 20\log_{10}\left(\frac{\frac{1}{K}\sum_{k=1}^{K}\bar{I}_{fg}(k) - \frac{1}{K}\sum_{k=1}^{K}\bar{I}_{bg}(k)}{\sigma_n}\right), \tag{6.12}$$

with K the number of frames, $\bar{I}_{fg}(k)$ the mean foreground intensity in frame k and $\bar{I}_{bg}(k)$ the mean local background intensity in frame k. $\bar{I}_{fg}(k)$ is calculated by taking the mean intensity of LR pixels that contain at least 50% foreground and $\bar{I}_{bg}(k)$ is defined by the mean intensity of all 100% background pixels in a small neighborhood around the object. The SNR gives an indication on the contrast of the object with its local background compared to the noise level. The SCR is defined as:

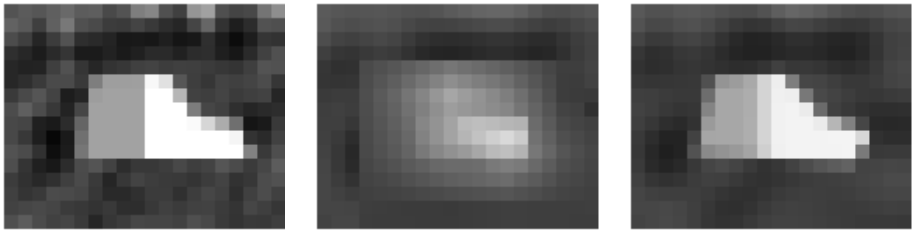
$$\mathrm{SCR} = 20\log_{10}\left(\frac{\frac{1}{K}\sum_{k=1}^{K}\bar{I}_{fg}(k) - \frac{1}{K}\sum_{k=1}^{K}\bar{I}_{bg}(k)}{\frac{1}{K}\sum_{k=1}^{K}\sigma_{bg}(k)}\right), \tag{6.13}$$

with $\sigma_{bg}(k)$ the standard deviation of the local background in frame k. The SCR is a measure of the amount of contrast of the object with the mean local background compared to the variation in the local background.

The result shown in Figure 6.6(c) is obtained by applying the pixel-based SR reconstruction approach described in [13] with zoomfactor 4, using 85 frames for reconstruction of the background and 50 frames for reconstruction of the foreground. The same camera model is used as in the construction of the data.

Using the same data, camera model and zoomfactor, the SR reconstruction result after applying our proposed method is depicted in Figure 6.6(d). The parameters used during reconstruction are in step 1: $\lambda_f = 10^{-4}$, $\lambda_p = 10^{-6}$ and in step 2: $\lambda_f = 10^{-3}$, $\lambda_p = 10^{-3}$. The object boundary is approximated with 8 vertices, which is one more than used for constructing the data, so the boundary is slightly over-fitted. Comparing the results in Figure 6.6(c) and (d) shows that the result of our new method bears a much better resemblance to the ground truth reference in Figure 6.6(b).

(a) Zoom in on moving car in three different LR frames (5×4 pixels).



(b) Ground truth. (c) 4× pixel-based SR. (d) 4× proposed SR.

Figure 6.6: *Four times SR reconstruction of a simulated under-sampled image sequence containing a small moving car. (a) shows three LR frames. (b) shows the HR ground truth reference. (c) shows a state-of-the-art pixel-based SR result and (d) shows the SR result of our new method.*

6.4.2 Test 2 on simulated data

This experiment is done on simulated image sequences similar to the one used in the previous experiment. To investigate the performance of our method under different conditions, we varied 1) the clutter (variance) of the local background and 2) the noise level. The clutter of the background is varied by multiplying the background with a certain factor after subtracting the median intensity. Afterwards the median intensity is added again to return to the original intensity domain. The intensities and the size of the car are not changed. The car still covers approximately 6 LR pixels (area) and the minimum object intensity (at the back of the car) is exactly in between the median local background intensity and the maximum object intensity.

Both the HR background and the HR foreground are reconstructed with zoom-factor 4 using 85 frames and 50 frames respectively. The camera model used during reconstruction is the same as used during constructing the data. For reconstruction of the moving object the same settings are used as in the previous experiment. The object boundary is again approximated with 8 vertices.

The quality of the different SR results is expressed by a Normalized Mean Squared Error (NMSE) with a ground truth reference $\mathbf{z}_{gt} = \mathbf{c}_{gt}\tilde{\mathbf{f}}_{gt}$ of the object. Note that this measure considers only the foreground intensities, the background intensities are set to zero.

$$\text{NMSE} = \frac{1/N \sum_{n=1}^{N} (\mathbf{z}_{gt}(n) - \mathbf{z}_{est}(n))^2}{\max(\mathbf{z}_{gt})^2}, \tag{6.14}$$

with N the number of HR pixels, $\mathbf{z}_{est}$ the estimated foreground intensities of the SR result and $\mathbf{z}_{gt}$ its ground truth reference. The normalization is done with the squared maximum intensity in $\mathbf{z}_{gt}$.

In Figure 6.7 the NMSE is depicted for varying SNR and SCR. We divided the results in three different regions: good (NMSE < 0.01), medium ($0.01 <$ NMSE < 0.03) and bad (NMSE > 0.03). For each region a SR result is shown to give a visual impression of the performance. It is clear that the SR result in the 'good region', with a realistic SNR and SCR, bears good resemblance to the GT reference. Note that the visible background in those SR results is not used to calculate the NMSE. Figure 6.7 shows that the performance decreases for a decreasing SNR. Furthermore, the boundary between the 'good' and 'medium' region indicates a decrease in performance under high clutter conditions (SCR < 5 dB).

6.4.3 Test on real-world data

The data for this experiment is captured with an infrared camera (the 1T from Amber Radiance). The sensor is composed of indium antimonide (InSb) detectors

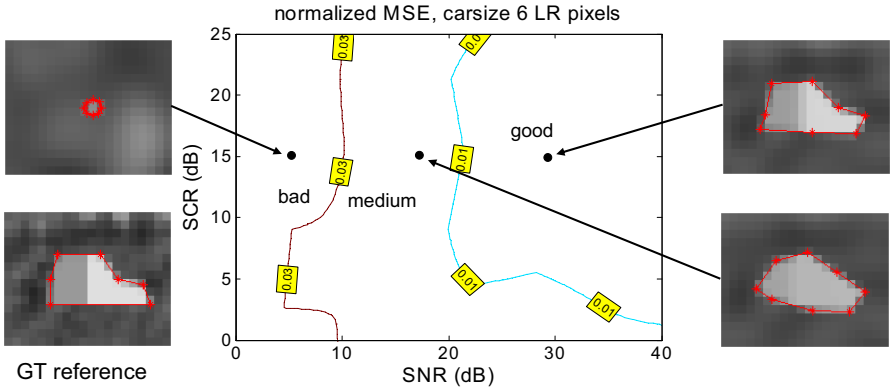


Figure 6.7: *Quantitative performance (normalized MSE) of the proposed SR reconstruction method on a simulated image sequence containing a moving car (6 pixels) for varying SNR and SCR. A smaller NMSE indicates a better performance, which is also visualized by three different SR results.*

(256×256) with a response in the $3-5\mu$m wavelength band. Furthermore, we use optics with a focal length of 50mm and a viewing angle of 11.2° (also from Amber Radiance). We captured a vehicle (Jeep Wrangler) at 15 frames/second, driving with a continuous velocity ($\approx$ 1 pixel/frame apparent velocity) approximately perpendicular to the optical axis of the camera. See Figure 6.8 for a top view of this setup. While capturing the data, the platform of the camera was gently shaken to provide subpixel motion of the camera. Panning was used to keep the moving vehicle within the field of view of the camera.

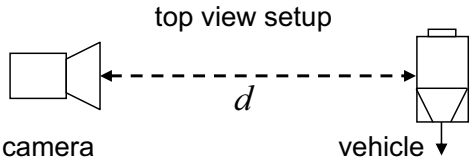
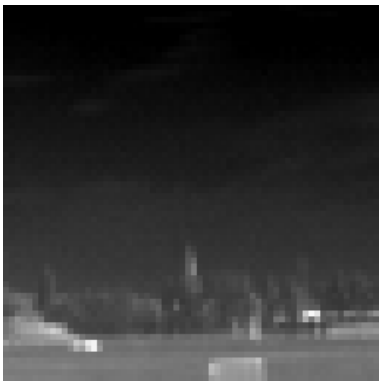


Figure 6.8: *Top view of the setup to capture real-world data.*

We selected the distance such that the vehicle appeared small ($\approx$ 10 LR pixels in area) in the image plane. In the left column of Figure 6.9 a part of a LR frame (64×64 pixels) and a zoom-in on the vehicle are shown. The vehicle is driving

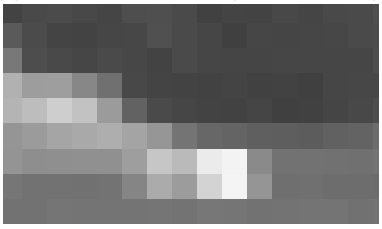
from left to right at a distance d of approximately 1150 meters. The SNR of the vehicle with the background is 30 dB and the SCR is 13 dB. In the previous experiment we have shown that for these values our method is capable to obtain good reconstruction. In the right column of Figure 6.9 the result after applying our SR reconstruction method shows that this is indeed the case.

The HR background is reconstructed from 85 frames with zoomfactor 4. The camera blur is modeled by Gaussian optical blurring ($\sigma_{psf} = 0.3$), followed by uniform rectangular sensor blurring (100% fill-factor). The HR foreground is reconstructed from 50 frames with zoomfactor 4 and the camera blur is modeled in the same way as for the background. The object boundary is approximated with 12 vertices and during the reconstruction the following settings are used: $\lambda_f = 10^{-4}$, $\lambda_p = 10^{-6}$ in both step 1 and 2.

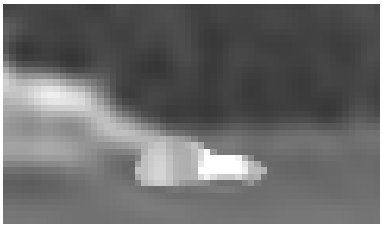


(a) LR reference frame (64×64 pixels).

(b) Zoomfactor 4 SR result with our proposed method.



(c) Zoom in on moving object in (a).

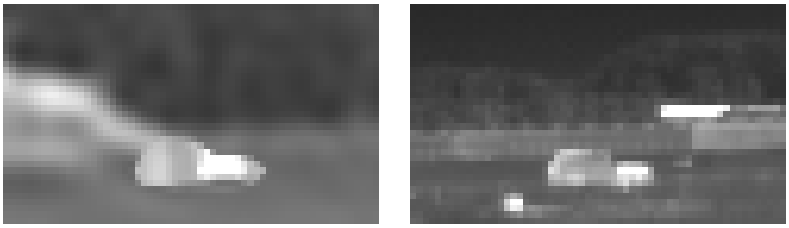


(d) Zoom in on moving object in (b).

Figure 6.9: *Four times SR reconstruction of a vehicle captured by an infrared camera (50 frames) at large distance. (a) and (c) show the LR captured data, (b) and (d) show the SR reconstruction result after applying our proposed method.*

Note that much more detail is visible in the SR result than in the LR image.

The shape of the vehicle is much more pronounced and the hot engine of the vehicle is well visible. For comparison we depict in Figure 6.10 the SR result next to a captured image of the vehicle at a 4× smaller distance. For visualization purposes, the intensity mapping is not the same for both images. So a greylevel in (a) may not be compared with the same greylevel in (b). This intensity mismatch is explained by the fact that both sequences were captured at a different time, which causes a change in reflection by the sun and heating of the vehicle. The shape of the vehicle is reconstructed very well and the hot engine is located at a similar place.



(a) 4× SR result. (b) Object 4× closer to camera.

Figure 6.10: *4× SR result of a jeep compared with the same jeep captured at a 4× smaller distance.*

6.5 Conclusions

The proposed multi-frame SR reconstruction method improves the visual recognition of small moving objects under realistic Signal-to-Noise Ratios and Signal-to-Clutter Ratios. We showed that our method performs well in reconstructing a small moving object where state-of-the-art pixel-based SR reconstruction methods fail. Our method not only performs well on simulated data, but also on a real-world image sequence captured with an infrared camera.

The main novelty of the proposed SR reconstruction method is the use of a combined boundary and intensity description of a small moving object. This enables us to estimate simultaneously the object boundary with subpixel precision *and* the foreground intensities from the mixed pixels, which are partly influenced by the background and partly by the foreground.

Also we have introduced a *hyperbolic* error norm on the foreground intensity differences in the cost functional of our SR reconstruction method. This robust error norm permits use of the popular Levenberg-Marquardt minimization procedure.

Chapter 7

Conclusions and discussion

In this thesis we presented different multi-frame Super-Resolution (SR) reconstruction methods to improve the detection, resolution, Signal-to-Noise Ratio (SNR) and contrast of moving objects in under-sampled image sequences. Applying these methods improves the detection and recognition rate of moving objects, especially for *small* moving objects.

In this chapter an overview is given of the answers to the research questions defined in the introduction of this thesis, some future research directions will be given as well.

7.1 Performance evaluation

In **Chapter 2** it is shown that Triangle Orientation Discrimination (TOD) [7] provides a good quantitative measure for the performance of SR reconstruction methods. The performance of different methods can be compared for a specific condition of the Low-Resolution (LR) input data. Considering the imaging conditions (camera's fill-factor, optical Point Spread Function (PSF), SNR) the TOD method enables an objective choice to determine which SR reconstruction method to use.

Furthermore, it is shown that the performance of a SR reconstruction method on real-world data can be predicted well by measuring the performance on simulated data, if a proper estimate of the parameters of the real-world camera system is available. This makes it possible to optimize the complete chain of a vision system in an early stage. The parameters of the camera and the algorithm must be chosen such that the performance of the vision task is optimized.

It is also shown that regularization is not required for good performance when a large number of recorded LR frames are available, i.e. stop criteria in iterative methods will act as regularization. For low SNRs and many LR frames the performance improvement is mainly due to temporal noise reduction. Results also show that using a larger zoomfactor does not guarantee 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 maximum achievable resolution gain.

7.2 Point targets

The detection of point targets in an under-sampled image sequence can be improved by applying robust SR reconstruction for background suppression. From **Chapter 3** it can be concluded that this has the following advantages: 1) less aliasing artifacts in high clutter regions, 2) better point target amplitude preservation for point targets with a small apparent motion and 3) less noise in the difference image after background suppression.

It is shown that the use of SR reconstruction improves the specificity and sensitivity of a point target detection method. The improvement in specificity is based on two properties of the SR reconstruction algorithm: temporal noise reduction and anti-aliasing. Due to temporal noise reduction and anti-aliasing the number of false alarms decreases, as there is less noise in the background estimation and therefore also less noise in the difference image on which the detection is based.

The sensitivity of point target detection is increased by the point target suppression capabilities of SR reconstruction in the background estimate. Therefore, the amplitude of the point target is preserved in the difference image. This improvement is larger for point targets with a low apparent target velocity. Robust SR reconstruction is used, because this suppresses outliers and therefore yields hardly any contribution of the point target in its background estimation, whereas for non-robust SR reconstruction methods a small portion of the point target energy will still be present in the background estimation.

It can be seen that background suppression with SR reconstruction performs better than a standard Shift, Interpolate and Subtract algorithm in almost all tested scenarios. As expected, SR reconstruction with zoomfactor 2 performs better than SR reconstruction with zoomfactor 1 in high clutter scenarios. This effect is due to the fact that a better estimation of the background is obtained by the anti-aliasing capabilities of the zoomfactor 2 approach. Hence, it will decrease the number of false detections. In low clutter scenarios a higher zoomfactor does not improve the performance, i.e. the local LR data was not hampered by significant aliasing.

7.3 Large objects

In **Chapter 4** a method is presented to perform SR reconstruction on the background of a scene as well as on large moving objects in the scene. First registration is performed on the background after which the moving objects are detected. Now, each moving object is registered separately and SR reconstruction is applied to each masked sequence after registration. Simultaneously, SR reconstruction is applied to the background and finally the super-resolved background and super-resolved objects are merged into a high-resolution image frame. It is shown that this method works on scenes with large moving objects.

From results we can conclude that our framework has a comparable SR performance for movings objects and background under the conditions that 1) objects are 'large' ($\geq 16 \times 16$ LR pixels) *and* 2) the SNR is high enough (> 20 dB). For smaller moving objects the amount of information inside the object is too small to perform gradient-based registration with sufficient subpixel precision. Furtermore, the SR performance on moving objects for low SNRs is bounded by the performance of the registration and the moving object detection part of our framework. A processed image sequence, captured with an infrared camera, shows that the proposed framework also performs well on real-world data.

7.4 Small objects

In **Chapter 5** and **Chapter 6** a method is presented to perform SR reconstruction specifically on small moving objects in an image sequence. An object is small if the majority of its constituting pixels are so-called mixed pixels. Mixed pixels contain information from the background and from the foreground (object). This method describes a small moving object with a subpixel precise polygonal boundary and a high-resolution (HR) intensity grid.

The proposed SR reconstruction method improves the recognition of small moving objects under realistic Signal-to-Noise Ratios and Clutter-to-Noise Ratios. First we showed that our method performs significantly better in reconstructing a small moving object than the state-of-the-art in pixel-based SR reconstruction methods. Our method not only performs well on simulated data, but also on sequences captured with an infrared camera.

The main advantage of the proposed SR reconstruction method is that it can estimate the object boundary with subpixel precision and therefore is able to separate the foreground and background information within the mixed pixels. Especially for small moving objects our approach improves the recognition significantly.

Also we have introduced a *hyperbolic* error norm on the foreground intensity differences in the cost functional of our SR reconstruction method. This robust

error norm permits use of the popular Levenberg-Marquardt minimization procedure.

7.5 Future work

Even after four years of research it is inevitable to be left with some unsolved issues. In our research this is no different. This section will address several issues that remain for future research.

Moving object detection

Although this thesis did not focus on detection and tracking, it is an important aspect and the first step in applying SR reconstruction to moving objects. If objects are becoming smaller and/or the SNR is decreasing, it is harder to detect them. Especially good tracking plays an important role here. More research is needed to improve this part of our proposed methods.

Moving object registration

A good registration is a key element of successful SR reconstruction. A sufficient precise registration ($< 1/10$ of a LR pixel using a 4 times denser HR grid) is needed to fuse the LR samples on a HR grid. In this thesis we apply gradient-based registration on the scene's background and on moving objects if sufficient gradient information is available. From the results in **Chapter 4** we can conclude that on moving objects with a size of 16×16 pixels containing a sharp edge *and* for a middle to high SNR it is just possible to obtain a sufficient precise gradient-based registration.

These results are in line with the observations of Pham et al. [48], which indicate that the precision of gradient-based registration is proportional to the noise variance divided by the signal energy within a region of interest. However, if moving objects are smaller than 16×16 pixels or their signal energy level is low, another way of registration is needed. In **Chapter 6** we have proposed a registration method for small moving objects, which fits a model-based trajectory through the object's location in each frame.

The object location is determined by the center of gravity of the masked object pixels after detection and tracking. It is not said that this is the best way to find the object's location, but it was good enough under the tested (realistic) conditions.

A completely different approach to perform registration of moving objects would be to incorporate the registration in the iterative optimization procedure discussed in **Chapter 6**. This means that the proposed cost function needs to be

minimized for the optimal registration parameters as well. It is hard to give an indication of the precision that can be obtained by this approach. This is a topic for future research.

Reconstruction

For the development of our SR reconstruction method for small moving objects we put some constraints on the object; it must be rigid, the viewing angle may not change and the intensity distribution must stay the same. For several scenarios these constraints are valid, but it is not very difficult to think of a scenario for which these constraints are invalid.

A different motion model must be incorporated in our reconstruction step to deal with changes in viewing angle and/or variations in scale. We believe that the parameters of such a motion model must be estimated simultaneously with the boundary and intensity description of a moving object, because estimating those parameters in advance seems very difficult in the setting of *small* moving objects. A drawback of a more complex motion model is that it requires more parameters, which enlarges the search space even further. Another issue in this context is that the moving object must be described with a 2.5D model to allow a change in viewing angle.

The object boundary description with a polygon leaves also some room for improvement. In our implementation the number of vertices is fixed during the optimization procedure. If the number of vertices could be varied, it would be possible to e.g. delete vertices that connect edges which have approximately the same orientation. It might also be interesting to experiment with other types of boundary descriptions such as splines.

Except for reconstruction of small moving objects, our SR reconstruction method is also capable of reconstructing the boundary region of large moving objects or image regions with a large change in depth of field. We believe that our proposed method can improve the visual quality and recognition in these regions. However, some effort is needed to apply our method to these tasks.

Bibliography

[1] http://en.wikipedia.org/wiki/angular_resolution.

[2] http://en.wikipedia.org/wiki/fim-92_stinger.

[3] http://en.wikipedia.org/wiki/vympel_r-27.

[4] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(9):1167–1183, 2002.

[5] M. Ben-Ezra, A. Zomet, and S. K. Nayar. Video super-resolution using controlled subpixel detector shifts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(6):977–987, 2005.

[6] 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*, volume 6207. SPIE Defense and Security, 2006.

[7] P. Bijl and J. M. Valeton. Triangle orientation discrimination: the alternative to minimum resolvable temperature difference and minimum resolvable contrast. *Optical Engineering*, 37(7):1976–1983, 1998.

[8] S. S. Blackman. *Multiple-Target Tracking with Radar Application*, page 10. Artech House, 1986.

[9] A. P. Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159, 1997.

[10] E. L. Dereniak and G. D. Boreman. *Infrared Detectors and Systems*. New York: John Wiley & Sons, 1996.

[11] J. Dijk, A. W. M. van Eekeren, K. Schutte, D. J. J. de Lange, and L. J. van Vliet. Super-resolution reconstruction for moving point target detection. *Optical Engineering*, 47(8), 2008.

[12] D. Douglas and T. Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *The Canadian Cartographer*, 2(10):112–122, 1973.

[13] A. W. M. van Eekeren, K. Schutte, J. Dijk, D. J. J. de Lange, and L. J. van Vliet. Super-resolution on moving objects and background. In *Proc. 13th International Conference on Image Processing*, volume 1, pages 2709–2712. IEEE, 2006.

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

[15] A. W. M. van Eekeren, K. Schutte, O. R. Oudegeest, and L. J. van Vliet. Super-resolution on moving objects using a polygon-based object description. In *Proc. 13th Annual Conference of the Advanced School for Computing and Imaging*, pages 317–321. ASCI, 2007.

[16] A. W. M. van Eekeren, K. Schutte, and L. J. van Vliet. Super-resolution reconstruction of intensity and boundary information of small moving objects. *IEEE Trans. Image Processing.* submitted Nov. 2008.

[17] A. W. M. van Eekeren, K. Schutte, and L. J. van Vliet. Super-resolution reconstruction on small moving objects. In *Proc. 15th International Conference on Image Processing*, volume 1, pages 1248–1251. IEEE, 2008.

[18] M. Elad and Y. Hel-Or. A fast super-resolution reconstruction algorithm for pure translational motion and common space invariant blur. *IEEE Trans. Image Processing*, 10(8):1187–1193, 2001.

[19] P. E. Eren, M. I. Sezan, and A. M. Tekalp. Robust, object-based high resolution image reconstruction from low-resolution video. *IEEE Trans. Image Processing*, 6(10):1446–1451, 1997.

[20] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Advances and challenges in super-resolution. *Int. Journal of Imaging Systems and Technology*, 14(2):47–57, 2004.

[21] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. *IEEE Trans. Image Processing*, 13(10):1327–1344, 2004.

[22] M. A. Goodale and A. D. Milner. Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1):20–25, 1992.

[23] D. Gross. *Super-resolution from sub-pixel shifted pictures.* Master's thesis, Tel Aviv University, Israel, 1986.

[24] R. M. Haralick and L. Watson. A facet model for image data. *Computer Graphics and Image Processsing*, 15(2):113–129, 1981.

[25] R. C. Hardie, K. J. Barnard, and E. E. Armstrong. Joint map registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Trans. on Image Processing*, 6(12):1621–1633, 1997.

[26] 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*, 37(1):247–260, 1998.

[27] R. C. Hardie, T. R. Tuinstra, J. Bognart, K. J. Barnard, and E. E. Armstrong. High resolution image reconstruction from digital video with global and non-global scene motion. In *Proc. 4th International Conference on Image Processing*, volume 1, pages 153–156. IEEE, 1997.

[28] Ibn al Haytham. *Book of Optics.* 1011-1021.

[29] R. J. M. den Hollander, D. J. J. de Lange, and K. Schutte. Super-resolution of faces using the epipolar constraint. In *Proc. British Machine Vision Conference (BMVC'07)*, 2007.

[30] R. D. Hudson. *Infrared System Engineering*, pages 100–103. John Wiley & Sons, 1969.

[31] M. Irani and S. Peleg. Improving resolution by image registration. *Graphical Models and Image Processing*, 53:231–239, 1991.

[32] J. Johnson. Analysis of image forming systems. In *Proc. Image Intensifier Symposium*, pages 249–273, 1958.

[33] E. Kaltenbacher and R. C. Hardie. High resolution infrared image reconstruction using multiple, low resolution, alisased frames. In *Proc. of IEEE National Aerospace and Electronics Conference*, volume 2, pages 702–709, 1996.

[34] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice-Hall, 1993.

[35] H. Knutsson and C. F. Westin. Normalized and differential convolution. In *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 515–523, New York, NY, USA, 1993.

[36] R. L. Lagendijk and J. Biemond. *Iterative identification and restoration of images*. Kluwer, 1990.

[37] S. Lertrattanapanich and N. K. Bose. High resolution image formation from low resolution frames using delaunay triangulation. *IEEE Trans. Image Processing*, 11(12):1427–1441, 2002.

[38] Y. Li and F. Santosa. A computational algorithm for minimizing total variation in image restoration. *IEEE Trans. Image Processing*, 5(6):987–995, 1996.

[39] Z. Lin and H. Y. Shum. Fundamental limits on reconstruction-based superresolution algorithms under local translation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(1):83–97, 2004.

[40] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Imaging Understanding Workshop*, pages 121–130, 1981.

[41] H. A. Mallot. *Computational vision: information processing in perception and visual behavior*. MIT Press, 2000.

[42] J. J. Moré. *The Levenberg-Marquardt Algorithm: Implementation and Theory*, volume 630, pages 105–116. Springer Verlag, 1978.

[43] M. K. Ng, H. Shen, E. Y. Lam, and L. Zhang. A total variation regularization based super-resolution reconstruction algorithm for digital video. *EURASIP Journal on Advances in Signal Processing*, pages 1–16, 2007. Article ID 74585.

[44] A. V. Oppenheim, A. S. Wilsky, and I. T. Young. *Signals and Systems*, pages 527–531. Prentice-Hall, 1983.

[45] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, 20(3):21–36, 2003.

[46] A. J. Patti, M. I. Sezan, and A. M. Tekalp. Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time. *IEEE Trans. Image Processing*, 6(8):1064–1076, 1997.

[47] T. Q. Pham. *Spatiotonal Adaptivity in Super-Resolution of Under-sampled Image Sequences*. Phd thesis, Quantitative Imaging Group, TU Delft, 2006. ISBN: 90-75691-14-9.

[48] 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*, volume 5817, pages 133–144. SPIE, 2005.

[49] T. Q. Pham, L. J. van Vliet, and K. Schutte. Robust fusion of irregularly sampled data using adaptive normalized convolution. *EURASIP Journal on Advances in Signal Processing*, pages 1–12, 2006. Article ID 83268.

[50] S. A. Rizvi, N. M. Nasrabadi, and S. Z. Der. A clutter rejection technique for flir imagery using region-based principal component analysis. In *Proc. of IEEE International Conference on Image Processing*, volume 4, pages 415–419, 1999.

[51] D. Robinson and P. Milanfar. Statistical performance analysis of super-resolution. *IEEE Trans. Image Processing*, 15(6):1413–1428, 2006.

[52] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[53] G. B. Rybicki and A. P. Lightman. *Radiative Processes in Astrophysics*. New York: John Wiley & Sons, 1979.

[54] A. van der Schaaf. *Natural Image Statistics and Visual Processing*. Phd thesis, 1998. ISBN: 90-367-0868-0.

[55] K. Schutte, D. J. J. de Lange, and S. P. van den Broek. Signal conditioning algorithms for enhanced tactical sensor imagery. In *SPIE proceedings: Infrared Imaging Systems: Design, Analysis, Modeling and Testing XIV*, volume 5076, pages 92–100, 2003.

[56] C. E. Shannon. Communication in the presence of noise. In *Proc. Institute of Radio Engineers*, volume 37, pages 10–21, 1949.

[57] A. N. Tikhonov and V. Y. Arsenin. *Solutions of ill-posed problems*. Washington: Wiley, 1977.

[58] S. M. Tonissen and Y. Bar-Shalom. Maximum likelihood track-before-detect with fluctuating target amplitude. *IEEE transactions on aerospace and electronic systems*, 34(3):796–809, 1998.

[59] R. Y. Tsai and T. S. Huang. Multiframe image restoration and registration. In *Advances in Computer Vision and Image Proscessing*, volume 1, pages 317–339. JAI Press, 1984.

[60] A. P. Tzannes and D. H. Brooks. Detection of point targets in image sequences by hypothesis testing: a temporal test first approach. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 6, pages 3377–3380, 1999.

[61] J. M. Valeton, P. Bijl, E. Agterhuis, and S. Kriekaard. T-cat, a new thermal camera acuity tester. In *Infrared Imaging Systems: Design, Analysis, Modeling and Testing XI*, volume 4030. SPIE Defense and Security, 2000.

[62] L. J. van Vliet and P. W. Verbeek. Better geometric measurements based on photometric information. In *IEEE Instrumentation and Measurement Technology Conf. IMTC94*, pages 1357–1360, 1994.

[63] J. van de Weijer and R. van den Boomgaard. Least squares and robust estimation of local image structure. *International Journal of Computer Vision (IJCV)*, 64(2-3):143–155, 2005.

[64] F. W. Wheeler and A. J. Hoogs. Moving vehicle registration and super-resolution. In *Proc. of IEEE Applied Imagery Pattern Recognition Workshop (AIPR07)*, 2007.

[65] N. Wiener. *Extrapolation, interpolation, and smoothing of stationary time series*. New York: Wiley, 1949.

[66] J. Wu, M. Trivedi, and B. Rao. High frequency component compensation based super-resolution algorithm for face video enhancement. In *Proc. IEEE 17th International Conference on Pattern Recognition (ICPR'04)*, volume 3, pages 598–601, 2004.

[67] Y. Xiong, J. X. Peng, M. Y. Ding, and D. H. Xue. An extended track-before-detect algorithm for infrared target detection. *IEEE transactions on aerospace and electronic systems*, 33(3):1087–1092, 1997.

[68] I. T. Young, J. J. Gerbrands, and L. J. van Vliet. *Fundamentals of image processing*. Delft University of Technology, 1995.

[69] G. W. Zack, W. E. Rogers, and S. A. Latt. Automatic measurement of sister chromatid exchange frequency. *Journal of Histochemistry and Cytochemistry*, 25(7):741–753, 1977.

[70] W. Zhao, H. Sawhney, M. Hansen, and S. Samarasekera. Super-fusion: a super-resolution method based on fusion. In *Proc. IEEE 16th International Conference on Pattern Recognition (ICPR'02)*, volume 2, pages 269–272, 2002.

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

[72] A. Zomet, A. Rav-Acha, and S. Peleg. Robust super-resolution. In *IEEE Computer Society Conference on Computer Vision and Patter Recognition*, volume 1, pages 645–650, 2001.

Summary

Super-Resolution of Moving Objects in Under-Sampled Image Sequences

Multi-frame super-resolution (SR) reconstruction methods are capable of improving the resolution, signal-to-noise ratio and contrast of moving objects in under-sampled image sequences. These improvements will help to increase the detection and recognition rate of moving objects of various sizes. We distinguish point targets, objects that consist solely of "mixed" boundary pixels (small objects), and objects with sufficient internal structure for registration (large objects).

To objectively compare various SR reconstruction methods and to optimize their parameter settings, we used a quantitative performance criterion based on a method known as Triangle Orientation Discrimination. It is shown that the performance of a SR reconstruction method on real-world data can be predicted well by measuring the performance on simulated data. This makes it possible to optimize the complete chain of a vision system in advance. Furthermore it is shown that regularization is not required for good performance when many recorded frames are available for reconstruction.

Moving point targets against a cluttered background cannot be detected straightforwardly in under-sampled frames. We propose a method based on SR reconstruction for improving the detection of moving point targets in image sequences. A point target is the smallest possible object and it appears after blurring by the camera as a "blurred" point (the point spread function) in the image plane. Although it makes no sense to perform SR reconstruction on moving point targets themselves, their detection can be improved by applying SR reconstruction to the background of the scene. From the high-resolution background image we estimate the current low-resolution frame without point target. The difference image obtained this way has the following advantages: 1) less aliasing artifacts in high clutter regions, 2) better point target amplitude preservation for point

targets with a small apparent motion and 3) less temporal noise. These advantages result in fewer false detections for the same sensitivity. All in all, a better detection performance of point targets is obtained.

When an image sequence contains multiple motion fields, due to e.g. moving objects, it is not possible to apply one of the "standard" methods for SR reconstruction of the entire image. We developed a method to perform SR reconstruction on the background of a scene as well as on large moving objects in the scene. We define a moving object to be *large* if the total number of pixels contained by the object is large compared to the number of boundary pixels of the object. First, registration of the background enables us to detect the moving objects. Afterwards, each moving object is tracked, the sequence registered for each object separately, and SR reconstruction is applied. Simultaneously, SR reconstruction is applied to the background and finally the super-resolved background and super-resolved objects are merged into one high-resolution image frame. It is shown that this method performs well on scenes with large moving objects. The measured performance for large moving objects and for the background of a scene is similar for medium and high signal-to-noise ratios.

The biggest challenge was to develop a method to perform SR reconstruction specifically on small moving objects in an under-sampled image sequence. A small moving object is defined as an object which consists solely of "mixed" boundary pixels. Mixed pixels contain partly information from the varying background and partly from the foreground (object). We propose to perform SR reconstruction on small moving objects using a simultaneous boundary and intensity description of a moving object. Assuming rigid objects that move (constant speed is assumed) through the real world, a proper registration is accomplished by fitting a trajectory through the object's location in each frame. The boundary of a moving object is modeled with a subpixel precise polygon and the object's intensities are represented on a high-resolution pixel grid. After applying SR reconstruction to the background, the local background intensities are known on a high-resolution grid. When the intensities of the moving object and the position of the edges of the polygon boundary are known as well, the intensities of the mixed pixels can be estimated. By iteratively minimizing the model error between the measured and the estimated intensities, a subpixel precise boundary and intensity descriptions of the moving object are obtained. Results show that the proposed method works well on both simulated *and* real-world data and it is shown that, for reconstructing small moving objects, our method outperforms state-of-the-art pixel-based SR reconstruction methods. Furthermore, a *hyperbolic* error norm on the foreground intensity differences is introduced in the cost functional of our SR reconstruction method. This robust error norm permits use of an L1-based regularization term by the popular Levenberg-Marquardt minimization procedure.

Samenvatting

Super-Resolutie van Bewegende Objecten in Onderbemonsterde Beeld Sequenties

Meerdere-beeld super-resolutie (SR) reconstructie technieken zijn in staat om de resolutie, de signaal-ruis verhouding en het contrast van bewegende objecten in onderbemonsterde beeld sequenties te verbeteren. Deze verbeteringen dragen bij aan een betere detectie en herkenning van bewegende objecten van verschillende groottes. We onderscheiden punt doelen, objecten die slechts uit "gemengde" rand pixels bestaan (kleine objecten), en objecten met voldoende interne structuur voor registratie (grote objecten).

Om verschillende SR reconstructie methodes objectief te vergelijken en om hun instellingen te optimaliseren, gebruiken we een kwantitatief prestatie criterium dat gebaseerd is op driehoek oriëntatie discriminatie. Er wordt aangetoond dat de prestatie van een SR reconstructie methode op echte data voorspeld kan worden door het meten van de prestatie op gesimuleerde data. Dit maakt het mogelijk om de gehele keten van een visueel systeem vooraf te optimaliseren. Verder wordt ook aangetoond dat regularisatie niet noodzakelijk is voor een goede prestatie als veel opgenomen beelden beschikbaar zijn voor reconstructie.

Bewegende punt doelen tegen een afwisselende achtergrond kunnen niet zo maar gedetecteerd worden in onderbemonsterde beelden. Wij stellen een methode voor die gebaseerd is op SR reconstructie om de detectie van bewegende punt doelen in beelden te verbeteren. Een punt doel is het kleinst mogelijke zichtbare object en na waarneming met een camera verschijnt het als een uitgesmeerde punt in het beeldvlak. Hoewel het niet zinvol is om SR reconstructie te doen op bewegende punt doelen, kan hun detectie verbeterd worden door SR reconstructie toe te passen op de achtergrond van de opgenomen scene. Met het verkregen hoge resolutie achtergrond beeld schatten we het huidige lage resolutie beeld zonder punt doel. Nu kan een verschil beeld worden bepaald met de volgende voordelen:

1) minder bemonstering artefacten in gebieden met veel structuur, 2) beter behoud van punt doel amplitude voor punt doelen met een kleine beweging en 3) minder temporele ruis. Deze verbeteringen resulteren in minder foute detecties bij een gelijke gevoeligheid. Kortom, een betere detectie van punt doelen wordt bereikt.

Als een beeld sequentie door bijvoorbeeld bewegende objecten meerdere bewegingsvelden bevat, is het niet mogelijk om "standaard" SR reconstructie technieken toe te passen op het hele beeld. Wij hebben een methode ontwikkeld die zowel op de achtergrond van de scene als op grote bewegende objecten in de scene SR reconstructie toepast. Wij noemen een bewegend object groot als het totaal aantal pixels van dat object groot is ten opzichte van het aantal rand pixels van hetzelfde object. Allereerst doen we registratie van de achtergrond, zodat we de bewegende objecten kunnen detecteren. Vervolgens wordt ieder object apart gevolgd, geregistreerd en wordt er SR reconstructie op toegepast. Tegelijkertijd wordt er SR reconstructie gedaan op de achtergrond en tenslotte wordt de hoge resolutie achtergrond samengevoegd met de hoge resolutie bewegende objecten. Er wordt aangetoond dat deze methode goed werkt op scènes met grote bewegende objecten. De gemeten prestatie op grote bewegende objecten is vergelijkbaar met de prestatie op de achtergrond voor gemiddelde en hoge signaal-ruis verhoudingen.

De grootste uitdaging was om een methode te ontwikkelen om SR reconstructie toe te passen op kleine bewegende objecten. Een klein bewegend object is hier gedefiniëerd als een object dat slechts uit gemengde rand pixels bestaat. Gemengde pixels bevatten gedeeltelijk informatie van de wisselende achtergrond en gedeeltelijk informatie van het object. Wij stellen een SR reconstructie methode voor die gebruik maakt van een simultane rand en intensiteit beschrijving van het bewegend object. Uitgaande van niet-vervormbare objecten met een constante snelheid in de echte wereld, wordt er een goede registratie bereikt door een bewegingspad te schatten op basis van de object posities in elk beeld. De rand van het object wordt beschreven met een subpixel nauwkeurige polygoon en de object intensiteiten worden beschreven op een hoog resolutie grid. Na het toepassen van SR reconstructie op de achtergrond zijn de lokale achtergrond intensiteiten bekend. Als de intensiteiten van het bewegend object en de positie van de polygoon randen ook bekend zijn, kunnen de intensiteiten van de gemengde pixels worden geschat. Door iteratief de model fout tussen de gemeten en geschatte intensiteiten te minimaliseren, wordt er een subpixel nauwkeurige rand en intensiteit beschrijving van het bewegend object verkregen. Resultaten laten zien dat de voorgestelde methode werkt op zowel gesimuleerde als op echte data en dat onze methode significant betere prestaties levert op kleine bewegende objecten dan zeer geavanceerde pixel gebaseerde SR reconstructie methodes. Verder introduceren we een hyperbolische fouten norm voor de voorgrond intensiteit verschillen in de kosten functie van onze SR reconstructie methode. Deze robuuste fouten norm maakt het mogelijk om een L1-gebaseerde regularisatie term te gebruiken in de populaire Levenberg-Marquardt minimalisatie procedure.

Curriculum Vitae

Adam W.M. van Eekeren was born in Roosendaal en Nispen, the Netherlands on October 25^{th}, 1977. He received his grammar school diploma in 1996 at the Gymnasium Juvenaat Heilig Hart in Bergen op Zoom. The same year he started Electrotechnical engineering at the Eindhoven University of Technology (TU/e), where he obtained cum laude his propaedeutics. In 2002 he did an external graduation project at Philips Medical Systems (PMS) which he concluded with his thesis entitled 'Noise Reduction in Digital Subtraction Angiography, using Morphological Image Processing', under the supervision of Dr.ir. Peter Rongen (PMS), Ir. Herman Stegehuis (PMS) and Prof.dr.ir. J.W.M. Bergmans (TU/e).

After a good long holiday he continued in 2003 his research activities at Philips Research Eindhoven with a one year project entitled 'Cell detection using the level set method', under supervision of Drs. Jurgen Rusch and Dr. B. Bakker. In 2004 he proceeded his career as a PhD-student on the topic of 'Super-resolution of moving objects in under-sampled image sequences'. He worked on this subject 4 days a week at the Electro Optics group at TNO Defense, Security and Safety in The Hague and one day a week at the Quantitative Imaging Group at the Delft University of Technology. His supervisors were Dr. Klamer Schutte (TNO) and Prof.dr.ir. Lucas J. van Vliet (TUD).

While finishing his manuscript, Adam continued in September 2008 his career in the same group at TNO Defense, Security and Safety in The Hague. He is working at various image processing topics such as change detection, 3D reconstruction and image enhancement. An up-to-date curriculum vitae can be found at http://www.linkedin.com/in/adamvaneekeren.

List of publications

Journal publications

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

[2] J. Dijk, **A.W.M. van Eekeren**, K. Schutte, D.J.J. de Lange, and L.J. van Vliet. Super-resolution reconstruction for moving point target detection. *Optical Engineering*, 47(8), 2008.

[3] **A.W.M. van Eekeren**, K. Schutte, and L.J. van Vliet. Multi-frame super-resolution reconstruction of small moving objects. *IEEE Trans. Image Processing*, submitted Nov. 2008.

Conference proceedings

[1] **A.W.M. van Eekeren**, K. Schutte, J. Dijk, D.J.J. de Lange, and L.J. van Vliet, Super-resolution on moving objects and background, in: B.P.F. Lelieveldt, B. Haverkort, C.T.A.M. de Laat, J.W.J. Heijnsdijk (eds.), *Proc. ASCI 2006, 12th Annual Conf. of the Advanced School for Computing and Imaging* (Heijen, NL, June 14-16, 2006), ASCI, Delft, 2006, pp. 53-57.

[2] **A.W.M. van Eekeren**, K. Schutte, J. Dijk, D.J.J. de Lange, and L.J. van Vliet, Super-resolution on moving objects and background, in: (eds.), *ICIP2006, Proceedings 13th International Conference on Image Processing*, (Atlanta, GA, Oct.8-11), Vol. 1, IEEE Signal Processing Society Press, Los Alamitos, 2006, pp. 2709-2712.

[3] **A.W.M. van Eekeren**, K. Schutte, O.R. Oudegeest, and L.J. van Vliet, Super-resolution on moving objects using a polygon-based object description, in: F.W. Jansen, G.E.O. Pierre, C.J. Veenman, J.W.J. Heijnsdijk (eds.), *Proc. ASCI 2007, 13th Annual Conf. of the Advanced School for Computing and Imaging* (Heijen, NL, June 13-15, 2007), ASCI, Delft, 2007, pp. 317-321.

[4] J. Dijk, **A.W.M. van Eekeren**, K. Schutte, and D.J.J. de Lange, Point target detection using super-resolution reconstruction, in: Automatic Target Recognition XVII. (eds. F.A. Sadjadi), *Proceedings of the SPIE*, Volume 6566, 2007. doi:10.1117/12.725074

[5] **A.W.M. van Eekeren**, K. Schutte, and L.J. van Vliet, Super-resolution on small moving objects, in: (eds.), *ICIP2008, Proceedings 15th International Conference on Image Processing*, (San Diego, California, Oct. 12-15), Vol. 1, IEEE Signal Processing Society Press, Los Alamitos, 2008, pp. 1248-1251.

[6] J. Dijk, **A.W.M. van Eekeren**, K. Schutte, D.J.J. de Lange, and L.J. van Vliet, Performance study on point target detection using super-resolution reconstruction, in *Proceedings of the SPIE*, Volume 7335, 2009.

Premier depots

[1] **A.W.M. van Eekeren**, K. Schutte, L.J. van Vliet, Algorithm of performing super-resolution reconstruction on small moving objects, nr. 08150423.5, 2008.

[2] **A.W.M. van Eekeren**, K. Schutte, Apparatus and method for producing improved quality signals, nr. 08161222.8, 2008.

Acknowledgements

Who would have thought that once I would become a doctor? After I graduated I was probably the last one who would have thought that. In the following year however, when I worked at Philips Research, I had an insight that doing a PhD was maybe not a bad idea at all. So first of all let me thank the colleagues at Philips Research and especially Jurgen for getting that insight.

The supervision of my PhD was in the very good hands of Lucas (TUD) and Klamer (TNO). It is incredible how often you were on the same line of thought (for which I am grateful). Although I must admit that sometimes I had some difficulty keeping up with you.

Lucas, thanks a lot for all the discussions, sharing your knowledge and wonderful ideas. Your door was always open (even when you were not there) when I wanted to ask you something. I really enjoyed working with you and thanks for all the things you taught me.

Klamer, I want to thank you for all the fruitful discussions we had. I admire your passion and your gift to put your finger always on the spot of the problem. You encouraged me to find new challenging goals. Thanks for all that and I see forward to working with you in the next years.

Of course I didn't do my research alone. Judith, thanks for being my roomy for several years. I really enjoyed working with you, your helpfulness and mental support when I was merging my code. Dirk-Jan, thank you for helping me with all my questions about the DASCA code. Also I want to thank Piet Bijl for his assistance while using the TOD evaluation method.

A software method without testing it on real-world data is not useful. Hans, Peter, Harald, Frank, many thanks for your assistance during all of my experiments. Also I want to thank Tuan, my PhD-predecessor, from who I really learnt a lot. The tools you developed were incredibly useful during my research. Olivier, my master student, thanks for the nice piece of work you did.

During my PhD I worked four days a week at TNO. I want to thank all my past and present colleagues for their help, knowledge, lunch breaks/walks and yearly sailing event. Because it is hard to say goodbye, I decided to stay a little longer! ;-)

Although I spent only one day a week at the Quantitative Imaging group in Delft, I really enjoyed being there. Thank you all for the C++ during the coffee breaks, the interesting talks and discussions. Matthan, it was great having you as a roommate. Mandy, thanks for all the paperwork. Ronald, thank you for solving all of the computer problems and your loud, but positive presence. Wim, I cannot imagine QI without you.

In the first year of my PhD I cancelled a holiday while being very busy. A big mistake I never made again! I want to thank all of my climbing friends for the great trips to the rocks and snow, so I could mentally reload. Especially I want to remember Wouter. Although you are not among us anymore, you are a friend forever!

Last but not least I want to thank my family and dearest friends for their interest in my research and their support. Krista, jij bent mijn leukste tweelingzus! ;-) Bedankt voor jouw enthousiasme, mental support en gastvrijheid in Davos. Pa en ma, ik kan me geen betere ouders voorstellen! Bedankt voor alles dat jullie voor mij hebben gedaan en me gevormd hebben tot wie ik nu ben.

Adam van Eekeren
April 2009