Master of Science Thesis in Electrical Engineering
Department of Electrical Engineering, Linköping University, 2020

Super-Resolution using Dynamic Cameras

Erik Dahlström

LINKÖPING
UNIVERSITY

Master of Science Thesis in Electrical Engineering

Super-Resolution using Dynamic Cameras

Erik Dahlström

LiTH-ISY-EX--20/5315--SE

Supervisor: **Gustav Häger**
ISY, Linköpings universitet
Magnus Olsson
Image System Motion Analasys

Examiner: **Lasse Alfredsson**
ISY, Linköpings universitet

Computer Vision
Department of Electrical Engineering
Linköping University
SE-581 83 Linköping, Sweden

Abstract

In digital image correlation, an optical full-field analysis method that can determine displacements of an object under load, high-resolution images are preferable. One way to improve the resolution is to improve the camera hardware. This can be expensive, hence another way to enhance the image is by various image processing techniques increase the resolution of the image. There are several ways of doing this and these techniques are called super-resolution. In this thesis the theory behind several different approaches to super-resolution is presented and discussed. The goal of this Thesis has been to investigate if super-resolution is possible in a scene with moving objects as well as movement of the camera. It became clear early on that image registration, a step in many super-resolution methods that will be explained in this thesis, was of utmost importance, and a major part of the work became comparing image registration methods. Data has been recorded and then two different super-resolution algorithms have been evaluated on a data set showing that super-resolution is possible.

Acknowledgments

First of all I would like to thank Image Systems for the opportunity to work on this thesis. Special thanks to Magnus Olsson and Tomas Chevalier for their input and support throughout this project. I would also like to thank all the employees at Image Systems for fun conversations during coffee breaks and making me feel welcome at their office. Secondly I would like to thank my supervisor Gustav Häger for good inputs and quick responses, and my examiner Lasse Alfredsson.

Robert Eklund at IKOS, Linköping University also deserves to be thanked. He helped me with inputs on the structure and grammar when writing this thesis.

And lastly I would like to thank my family for their encouraging words and especially Nicole Wattis, my girlfriend, for moral support and proofreading.

Linköping, June 2020
Erik Dahlström

Contents

Notation

Abbreviation

Abbreviation	Meaning
DIC	Digital Image Correlation
IBP	Iterative Back-Projection
CNN	Convolutional Neural Network
SRCNN	Super-Resolution Convolutional Neural Network
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity Index
HVU	Human Visual System

1

Introduction

In a vast number of digital image applications ranging from satellite imagery to medical imaging, images with high-resolution are often desired. High-resolution means that the pixel density in the image is high and the image can thus convey a more detailed description of the continuous scene it samples. There are several ways of increasing the resolution of an image. In a hardware perspective the direct method is to reduce the pixel size and thereby fit more pixel sensors onto the chip area. A reduced pixel size means that a reduced number of photons will hit the sensor element and that shot noise will occur resulting in degraded quality. Hence, there is a physical limit to reducing the size of the image sensor [1]. It would also be possible to make larger image sensors, but that would lead to larger and more expensive cameras, which might not be desirable. Other solutions would be increasing the size of the sensor chip at the cost of increased capacitance of the system, which will result in a slower transfer rate, or increasing the focal length of the camera lens, which would also result in larger and heavier cameras.

In super-resolution, signal processing techniques are used to overcome the limitations of imaging systems and sensor technology to produce high-resolution images with more details than the sampling grid of a given imaging system would provide.

The field of super-resolution was established in the 80s and initial work was done in the frequency domain, using several under-sampled low-resolution images to reduce aliasing in fused high-resolution images. This could be done if there were sub-pixel shifts between the low-resolution images. Today most work is done in the spatial domain because of the inability of the frequency domain approaches to handle complex motion models in the scene [2].

1.1 Motivation

Highly detailed images are desired in many applications, especially in *Digital Image Correlation* (DIC), an optical full-field analysis method that can determine displacements of an object under load. Image Systems, the company where this thesis was conducted, are specialized in DIC and work with some systems involving dynamic tracking mounts used when tracking different kinds of high-speed objects. There are several state-of-the-art algorithms that handle scenarios with either a stationary camera and a moving scene or a stationary scene and a moving camera that produce very good results. The case in which motion is used in both the scene and the camera is yet to be explored. If super-resolution is achievable under these conditions, it could result in better measurement results in future products. It would also be economical, in the sense that the quality of images produced from an image system would be increased without investing in new hardware.

1.2 TEMA

TEMA [3] is a state-of-the-art motion analysis software developed by Image Systems. With TEMA it is possible to track a point or a region very precisely in an image sequence and it is used for image registration in this work. Image registration and how TEMA was used is explained in detail in the theory chapter under section 2.3.2.1.

1.3 Problem formulation

When achieving super-resolution by using several low-resolution observations, the observations need to contain information unique to each other. A condition for super-resolution is thus that the shifts between the low-resolution observation are not integral. With TEMA, it is possible to track features between frames at sub-pixel precision.

The idea from the start was that this thesis was supposed to survey the field of super-resolution in order to see what has been done and if super-resolution might be of use for the DIC-applications that Image Systems are developing and also answer the following questions:

- Is super-resolution achievable in the case of motion in both the scene and the camera?
- Is the baseline algorithm that uses TEMA for motion estimation able to support super-resolution?
- How does one of the state-of-the-art algorithms compare to the baseline algorithm?

Regarding the baseline algorithm we apply the most simple implementation for the kind of motion present in the data set and then evaluate other methods against it. As the research progressed and the experiments where conducted, it became evident that for several super-resolution methods the most crucial step in the super-resolution system was the image registration. The focus of the thesis shifted thus to a comparison between two super-resolution systems for a challenging data set.

1.4 Limitations

This thesis will only use the data acquired with Image Systems when evaluating the algorithms. Super-resolution in the whole image will not be necessary. The baseline algorithm will implemented in Python. Possibilities to incorporate the algorithm in the TEMA software is not explored. Machine learning approaches often need large amounts of data for training and verification. The lack of machine learning expertise within Image Systems and the fact that there was no available data at the start of the thesis work excludes any experimenting with machine learning approaches.

1.5 Thesis outline

Chapter 2 introduces and explains the theoretical background used for super-resolution, conditions for super-resolution and various approaches. Chapter 3 discusses the methods used in this thesis. In chapter 4 the results will be presented and then discussed in chapter 5. The conclusions that can be drawn from this work as well as the direction of future work will be discussed in chapter 6.

2

Super-resolution overview

This chapter introduces the background theories of achieving super-resolution and different approaches on how to accomplish it.

2.1 Observation model

The basic principle of super-resolution is to use one low-resolution image or a sequence of low-resolution images of a scene, in order to create an image with higher spatial resolution that conveys finer detail or content with higher frequencies than the low-resolution images. This means that when recording a digital image, there is always a natural loss of spatial resolution and usually some kind of motion blur and noise due to limitations of the imaging system as illustrated in figure 2.1. This can occur because of shutter speed, noise inside the sensor or optical distortions. Some of this blur can be reduced by using techniques similar to the ones used in image restoration in order to recover some high frequency information. The imaging model, which is called the observation model in this thesis, describes how the low-resolution images relate to the desired high-resolution image. This inverse *ill-posed* problem is posed in its linear form as

$$g_k(k,l) = D_k(x,y)B_k(x,y)W_k(x,y)f(x,y) + n_k(k,l), \tag{2.1}$$

where g is the observed low-resolution images vectorised, D is the sub-sampling, done by the imaging system, B is the blurring matrix, due to atmospheric and optic limitations. W is the warping matrix and n is the additive noise inherent in all imaging systems, f is the desired high-resolution image and k ranges from 1 to the number of low-resolution observations. This problem is ill-posed since for a given low-resolution image g, several high-resolution images f satisfy the reconstruction constraint in the equation seen in equation 2.1.

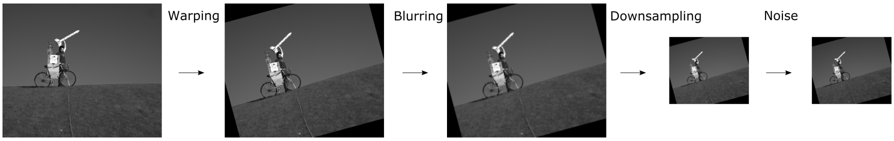


Figure 2.1: *The observation model used in most algorithms.*

2.2 Frequency domain methods

Some of the earliest work in the field of super-resolution was done in the frequency domain. In 1984, Huang and Tsai proposed an early multiple-image super-resolution algorithm when working with satellite images [4]. These kind of methods transform the low-resolution observations to the frequency domain where, in a combined image registration and reconstruction step, the high-resolution image is derived and then transformed back into the spatial domain, as shown in figure 2.2. This is done by using the shifting property of the Fourier transform, the aliasing relationship of the low-resolution input images and the high resolution output image, see figure 2.3, as well as the assumption that the original high-resolution image is bandlimited. Most work done with frequency domain methods can be divided into Fourier transform or Wavelet transform based methods, depending on the transforms used.

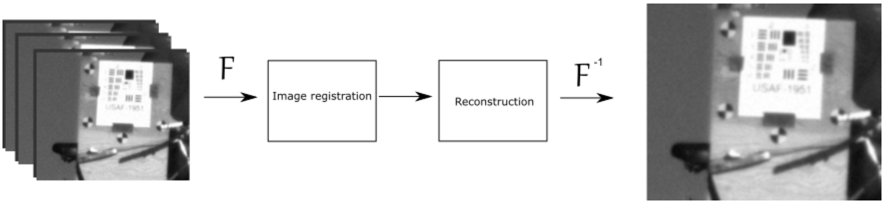


Figure 2.2: *The typical pipeline of a super-resolution algorithm in the frequency domain. Low-resolution images are fed to the system and transformed into the frequency domain. Then the image registration is performed and then reconstructed before being transformed back to the spatial domain as a high-resolution image.*

Since these methods rely on the shifting properties of the Fourier transform, the better motion estimation of the low-resolution images used, the better high-resolution image can be estimated. Consequently a good image registration is of utmost importance [5].

Frequency domain based methods work well in scenes with only global translations, rotations and scaling and are cheap to compute but not very flexible. Therefore, spatial domain based methods are often preferable before other methods [6].

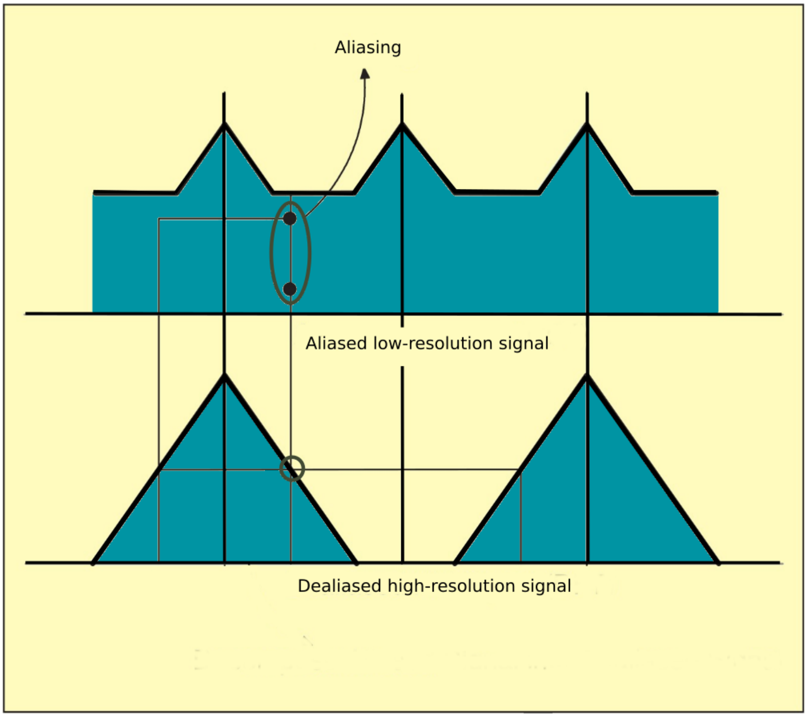

Aliasing
Aliased low-resolution signal
Dealiased high-resolution signal

Figure 2.3: *The aliasing relationship between the low-resolution image in the upper half of the figure and the high-resolution image at the bottom half, here represented as 1-d signals.*

2.2.1 Image registration and reconstruction in the frequency domain

This section will briefly explain how super-resolution works for algorithms using the Fourier transform. When using several low-resolution images, an image registration step is necessary. The motion-estimation of the scene used to register the images can be derived with the shifting properties of the Fourier transform [7]. A more thorough explanation of what image registration is will be given in section 2.3.2.1.

If f_1 and f_2 are two consecutive images of a scene and f_2 differs from f_1 with a displacement of Δx and Δy, i.e.

$$f_2(x,y) = f_1(x-\Delta x, y-\Delta y), \tag{2.2}$$

then the Fourier transforms of f_1 and f_2 will be related as

$$F_2(\zeta,\eta) = e^{-j2\pi(\zeta\Delta x+\eta\Delta y)}F_1(\zeta,\eta). \tag{2.3}$$

The cross-power spectrum of the two images, f_1 and f_2, with Fourier transform F_1 and F_2, is defined as

$$S_{f_1f_2}(\zeta,\eta) = \frac{F_1(\zeta,\eta)F_2^*(\zeta,\eta)}{\left|F_1(\zeta,\eta)F_2^*(\zeta,\eta)\right|} = e^{j2\pi(\zeta\Delta x+\eta\Delta y)}, \tag{2.4}$$

where F_2^* is the complex conjugate of F_2. According to the shift theorem, the phase of the cross-power spectrum is equivalent to the phase shift between the images. By inverse Fourier transforming the cross-power spectrum, an impulse that is close to zero everywhere, except at the optimal values for Δx and Δy, is acquired.

If the shifted images are impulse sampled with sampling period T_1 and T_2 the following low-resolution images are acquired:

$$g_i[k,l] = f_i(kT_1+\Delta x, lT_2+\Delta y), \tag{2.5}$$

where $k = 0,1,2,\ldots,K-1$ and $l = 0,1,2,\ldots,L-1$.

The discrete Fourier transform (DFT) $G_i[\nu,\mu]$ of these low-resolution images, ν and μ is related to the continuous Fourier transform (CFT) $F_i[\nu,\mu]$ of the shifted images by their aliasing property as follows:

$$G_i[\nu,\mu] = \frac{1}{T_1T_2}\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}F_i\left(\frac{2\pi}{T_1}\left(\frac{\nu}{K}-n\right),\frac{2\pi}{T_2}\left(\frac{\mu}{L}-m\right)\right). \tag{2.6}$$

If F is bandlimited, and by combining the equations 2.6 and 2.3, the DFT coefficients of $G_i[\nu,\mu]$ with the samples of the unknown CFT of $f(x,y)$ can be expressed in matrix form as

$$\mathbf{G} = \Phi\mathbf{F}, \tag{2.7}$$

and is a set of linear equations that can be solved for F. An inverse DFT is then used to acquire the super-resolved image $f(x,y)$.

2.3 Spatial domain methods

The spatial domain based methods can be divided into two subcategories depending on how many images they use. Because of limitations of the motion model and the blur that the frequency domain based methods can not handle, most work has been done in the spatial domain [8].

2.3.1 Single-image methods

Multi-image super-resolution algorithms are highly dependent on a correct motion estimation to be effective. In situations where the motion model is complex, single-image super-resolution might be effective. According to Nasrollahi and Moeslund [9], single-image super-resolution can be divided into two subgroups, reconstruction based methods and learning based methods, also known as hallucination based. Both types of methods involve training the algorithm with different types of training images.

2.3.1.1 Reconstruction based super-resolution methods

Much work has been done with reconstruction based methods, for example [10][11]. The reconstruction based single-image super-resolution methods try to remove aliasing artefacts in the low-resolution input image. This can be done in several ways, such as in Image hallucination with primal sketch priors [10] where lost high frequency content is hallucinated into images where learned image primitives, such as corners and edges, are found. Parameters for the algorithms are estimated by applying smoothness priors and applying new constraints so that when the enhanced high-resolution image is down-sampled, it should render the low-resolution image.

2.3.1.2 Learning based methods

In recent years, machine learning has shown success in various image applications. Super-resolution is one of them [12][13][14][15] [16][17][18]. Most machine learning based methods consist of a *Convolutional Neural Network* (CNN), learning the mapping from low resolution to high resolution image patches, which is seen in figure 2.4. That information is then used as a priori for the reconstruction step. *Super-Resolution Convolutional Neural Network* (SRCNN) [19], published in 2016 was one of the first methods that surpassed conventional super-resolution methods. Other work such as Image super-resolution via sparse representation [20] have also shown very promising results. The drawback of CNNs and other example based machine learning methods is that large quantities of training and evaluation data is necessary. In Image super-resolution using deep convolutional networks [19], Dong, Loy, He and Tang trained their network with a data set consisting of 395,909 images. Notably, the network was also trained with another data set consisting of only 91 images and showed only slightly worse results.

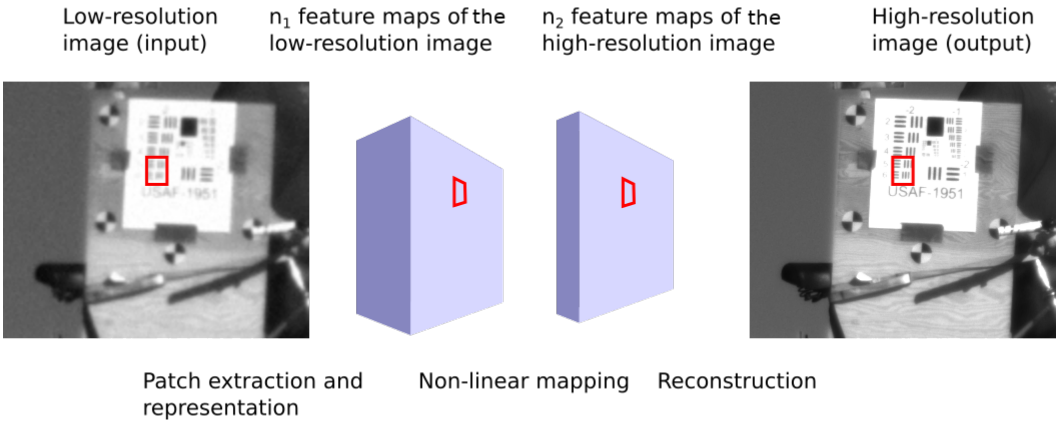

Low-resolution image (input)
n_1 feature maps of the low-resolution image
n_2 feature maps of the high-resolution image
High-resolution image (output)
Patch extraction and representation
Non-linear mapping
Reconstruction

Figure 2.4: *This is the architecture of the SRCNN. It consists of three layers. The first consists of $n_1 = 64$ feature maps, the second layer consists of $n_2 = 32$ feature maps and the last layer is the high-resolution output.*

The SRCNN consists of only three layers; a patch extraction layer, a non-linear mapping layer and a reconstruction layer. The input image is first re-sampled to the desired resolution by bilinear interpolation. The first layer extracts n_1-dimensional feature for each patch, the n_1 features are then non-linearly mapped to the n_2-features. The n_2-features represents the high-resolution patches. These patches are used for the final reconstruction of the high-resolution image.

2.3.2 Multi-image methods

Multi-image methods use several observations of the same scene. These observations need to be sub-sampled, which means that there is aliasing in the observations or that the observations are shifted relative to each other at sub-pixel precision. If they are not, the observations carry close to no additional information. To achieve these sub-pixel shifts there needs to be motion in the scene or movement of the camera taking these low-resolution images. The different steps involved in this kind of super-resolution method is described in the following steps [21].

2.3.2.1 Image registration

Image registration is a process that takes several data sets, in this case low-resolution observations, and estimates the motion present in the scene [22] as illustrated in figure 2.5. Once the motion is estimated, the low-resolution observations are transformed into the same coordinate system, and the sub-pixel information can be used to produce the desired high-resolution image. There are several ways of doing this, and the choice of image registration method depends on the kind of motion present in the scene.

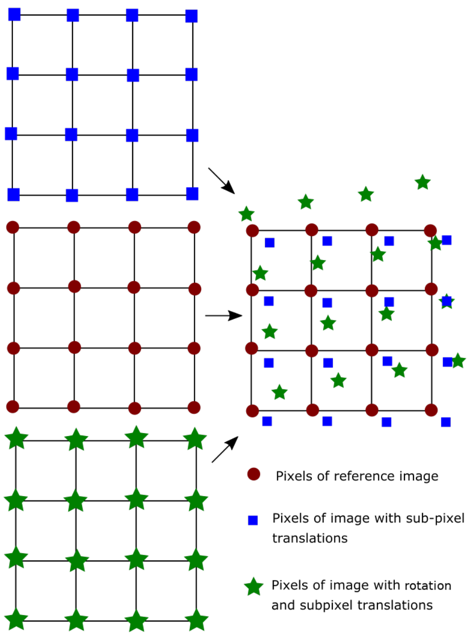

Pixels of reference image
Pixels of image with sub-pixel translations
Pixels of image with rotation and subpixel translations

Figure 2.5: *The image registration process.*

Optical flow [23] is the change in structured light in an image. It can be used for for estimating a motion field in a sequence of images. Consider a pixel, $I(x, y, t)$ in the reference image. It moves in the next image taken after Δt time, by Δx and Δy. Since it is still the same pixel, the intensity is presumed to be the same, i.e.

$$I(x_{next}, y_{next}, t_{next}) = I(x, y, t), \tag{2.8}$$

where $x_{next} = x + \Delta x$, $y_{next} = y + \Delta y$ and $t_{next} = t + \Delta t$. By a Taylor series approximation, expand the left side of equation and divide by the time difference Δt. We then get

$$f_x u + f_y v + f_t = 0, \tag{2.9}$$

where

$$f_x = \frac{\delta f}{\delta x}, f_y = \frac{\delta f}{\delta y}, u = \frac{\delta x}{\Delta t} \text{ and } v = \frac{\delta y}{\Delta t}. \tag{2.10}$$

The equation in 2.9 is called the optical flow equation and can not be solved directly because of the two unknowns. Several methods exist to solve this problem.

The Lucas-Kanade method [24] is a method that solves the optical flow equation by the use of a 3 × 3 pixel patch around a pixel. All pixels in that patch are assumed to have the same motion. If f_x, f_y and f_t is to be found for pixels in the patch, 9 equations are acquired to solve the problem with two unknowns, see equation 2.11. This works well for small motions in the image sequence.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum_i f_{x_i}^2 & \sum_i f_{x_i} f_{y_i} \\ \sum_i f_{x_i} f_{y_i} & \sum_i f_{y_i}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i f_{x_i} f_{t_i} \\ -\sum_i f_{y_i} f_{t_i} \end{bmatrix} \tag{2.11}$$

Cross-correlation based methods for image registration is a well researched area and they are used in several other image processing applications as [25][26]. Among other features, TEMA, described in in section 1.2, has a tracker that uses Zero Mean-normalization Cross-Correlation (ZNCC) to incrementally track a template in a sequence of images. This is done by first acquiring a displacement vector u. The reference position is x_0 and x is the deformed position in the template. Thus,

$$x = x_0 + u. \tag{2.12}$$

Let $\langle\Omega\rangle$ denote an undeformed subset and let $\langle\omega\rangle$ denote a deformed subset. These two subsets are linked by U, a shape-function. The shape-function is parameterized by a parameter vector p that approximates the displacement field around x_0. A first order Taylor series of U is used for the parametrization. Thus, we have

$$\langle\omega\rangle = U(\langle\Omega\rangle, p) + x_0 \tag{2.13}$$

and

$$u = U(x_0, p^*), \tag{2.14}$$

where p^*, the best p, is defined as

$$p^* = \max_p C(\langle\Omega\rangle, \langle\omega\rangle) = \max_p C(\langle\Omega\rangle, U(\langle\omega\rangle, p) + x_0), \tag{2.15}$$

where, C is the ZNCC correlation function. The *Broyden–Fletcher–Goldfarb–Shanno algorithm* (BFGS) [27][28][29][30] is used to find p^*. By letting X and Y represent the intensity values of two consecutive images and letting the template size be $M \times N$ pixels, C is defined as

$$C = \frac{\sum_{i=1}^{MN} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{MN} (x_i - \bar{x})^2 \sum_{i=1}^{MN} (y_i - \bar{y})^2}}. \tag{2.16}$$

TEMA calculates C for each possible template position in the search area in the next image, which is seen in figure 2.7. The ZNCC has values in the range $[-1, 1]$, where -1 is a bad match, 0 equals no correlation and 1 is a perfect match.

The ZNCC needs to handle the case when the positions from $U(\langle\Omega\rangle, p) + x_0$ are non-integers. It is done by using Catmull-Rom bi-cubic spline interpolation [31]. Catmull-Rom interpolation is explained in section 2.3.2.5.

When the tracked object has changed position in the scene, C will get worse values because of the increasing difference between $\langle\Omega\rangle$ and $\langle\omega\rangle$ at p^*. Hence some incremental update is needed. This is done by taking a new reference for $\langle\Omega\rangle$ and making sure it produces the same values for x_0 and that u is adjacent. To do this a center point S_0 of $\langle\Omega\rangle$ is introduced as well as the displacement vector from the previous frame, if the displacement vector u_{k-1} of the current frame is u_k. The center point S_0 is an integer and is defined as

$$S_0 = x_0 + u_{k-1} + \delta, \tag{2.17}$$

where x_0 is the reference position and δ is an offset. Then by using the shape-function U to the deformed subset $\langle\Omega\rangle$ with S_0 as offset, $\langle\omega\rangle$ can be described as

$$\langle\omega\rangle = U(\langle\Omega\rangle, p) + S_0. \tag{2.18}$$

p^* can then be acquired by,

$$p^* = \max_p C(\langle\Omega\rangle, \langle\omega\rangle) = \max_p C(\langle\Omega\rangle, U(\langle\omega\rangle, p) + S_0). \tag{2.19}$$

Now the displacement vector between $\langle\Omega\rangle$ and $\langle\omega\rangle$ can be written as $v = U(S_0, p^*)$. By calculating the displacement vector from $x_0 + u_{k-1} = S_0 - \delta$ as $v' = U(S_0 - \delta), p^*)$, x can be obtained by,

$$x = S_0 - \delta + U(S_0 - \delta, p^*) \tag{2.20}$$

which results in,

$$u_k = x - x_0 = S_0 - \delta + U(S_0 - \delta, p^*) - x_0. \tag{2.21}$$

This is the u_k that will be used for the incremental tracking of the object in the template. All the relations between the displacements and vectors can be seen in figure 2.6.

Feature based methods use distinctive objects in the image, often a closed-boundary region, edges, intersecting lines, or points. These are manually or automatically detected by what is called a *detector*. There are several kinds of detectors: Scale-Invariant Feature Transform, (SIFT) proposed by Lowe [32] or Features from accelerated segment test, (FAST) proposed by Rosten and Drummond [33]. These points of interest are then described by what is called a *descriptor*. Also here there is a variety to chose from, one example is the Histogram of oriented gradients, (HoG) introduced by Dalal and Triggs in 2005 [34], based on the work done by McConnell in [35] in 1986. Once a certain number of points of interest, key-points, have been extracted from the images to be registered, the correspondence between the key-points in the each of the images is calculated. The result can then be improved with methods like RANdom SAmple Consesus, (RANSAC) [36]. RANSAC improves the result by sorting out outliers and by that solving the correspondence problem.

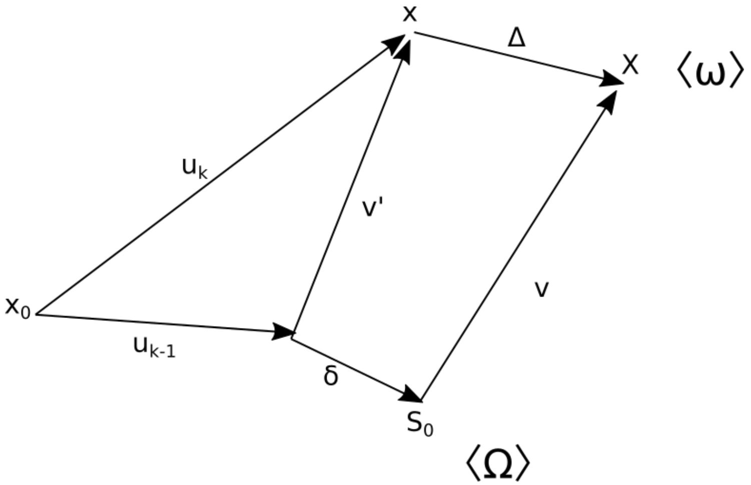


Figure 2.6: *The displacements and relations of the point clouds and vectors in one of the trackers featured in TEMA.*

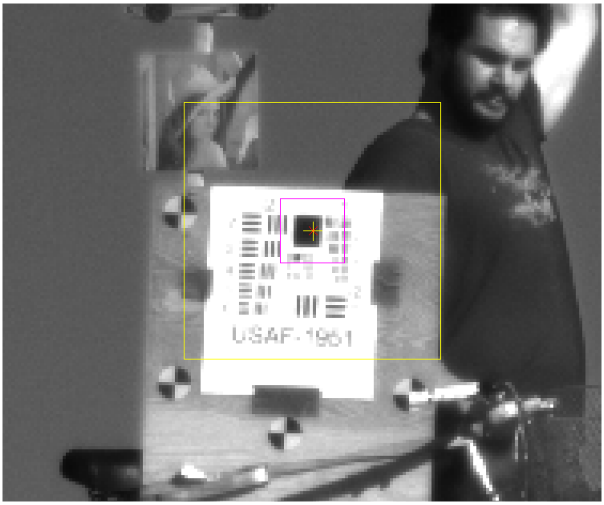


Figure 2.7: *The ZNCC tracker. The yellow square is the search area and the purple square is the template.*

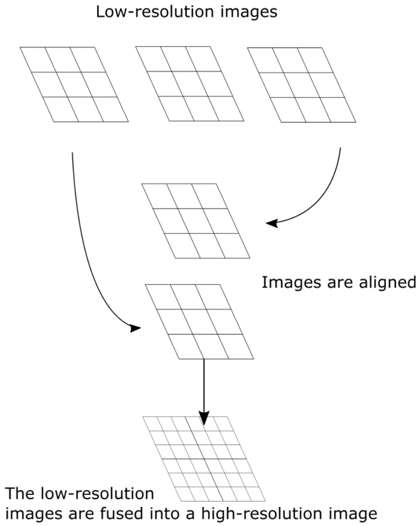

Low-resolution images
Images are aligned
The low-resolution images are fused into a high-resolution image

Figure 2.8: *The image reconstruction process.*

2.3.2.2 Reconstruction

With the registration complete and motion between the images estimated, the estimated motion is used to align the low-resolution images to construct a new high-resolution image. The images are aligned with transformation matrices, these are matrices that are derived depending on what kind of motion there is in the scene. The transformation of points in images can be done by a 3×3 matrix T. For a translation by Δx and Δy and rotation by θ the transformation matrix is given by

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} cos(\theta) & -sin(\theta) & \Delta x \\ sin(\theta) & cos(\theta) & \Delta y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \tag{2.22}$$

For scaling, s_x, s_y and shearing, k_x, k_y along the x-axis or y-axis the transformation matrix is given by

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & k_y & 0 \\ k_x & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \tag{2.23}$$

For perspective transformations, a homography is needed. It is given in equation 2.24. The elements in the matrix corresponds to image pairs in the reference image and the image to be registered.

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{2.24}$$

2.3.2.3 Image transformation and resampling

The images will be transformed into the coordinate system of the reference image and the pixel x, y are then mapped onto the reference image's coordinate system as x', y'.

$$f_{in}(x,y) := f_{out}(x',y') \tag{2.25}$$

This mapping can be done in two different ways. *Forward mapping* takes each pixel in the input image and copies the value onto the corresponding pixel in the output image according to the transform model as seen in equation 2.26. A risk with forward mapping is that some pixels in the output image does not get a value assigned to them and therefore interpolation in the output image might be necessary.

$$f_{in}(T(x,y)) := f_{out}(x',y') \tag{2.26}$$

In *inverse mapping*, each pixel in the output image is mapped back to the input images through the inverted transformation model as seen in equation 2.27. These pixels will be mapped to sub-pixel coordinates in the input sequence and hence interpolation in the input images is necessary.

$$f_{in}(x,y) := f_{out}(T^{-1}(x',y')) \tag{2.27}$$

2.3.2.4 Non-uniform interpolation

Non-uniform interpolation, based on the work by Yen [37], is a direct reconstruction method that creates a high-resolution grid and then by inverse mapping traces each pixel back to each low-resolution image used. The value for that pixel is then interpolated from the low-resolution images and then weighted depending on which image in the sequence of images it originates from. These values depend on which interpolation method is used. Examples of interpolation methods that can be used are *nearest neighbour*, *linear*, *bilinear* and *bicubic* interpolation, see figure 2.9.

2.3.2.5 Catmull-Rom interpolation

This interpolation method is a form of bicubic interpolation [39]. It is a recursive spline algorithm that uses extra data points for resampling as in figure 2.10. How splines work and are implemented can be read in [40].

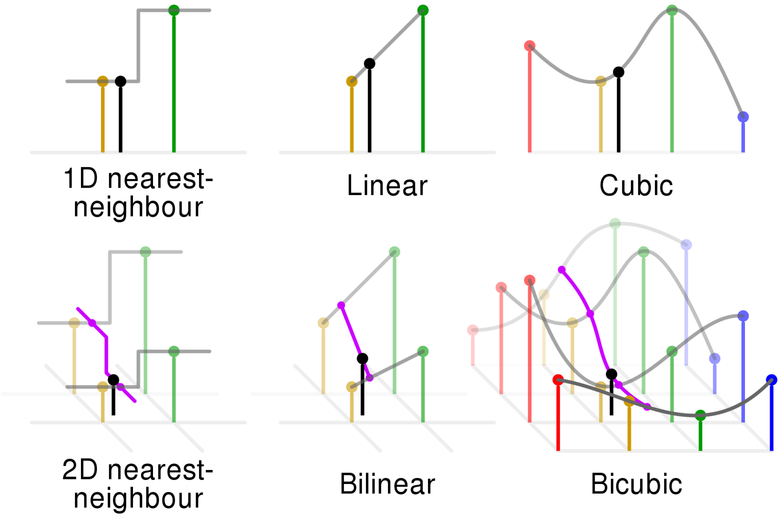


Figure 2.9: *Different types of interpolation methods in 1D and 2D [38].*

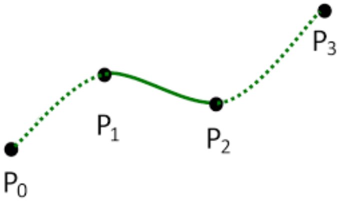


Figure 2.10: *Catmull-Rom splines use control points P_0 and P_3 to interpolate the value in the inner point between P_1 and P_2 [41].*

2.3.2.6 Iterative back-projection

Iterative Back-Projection (IBP) is an iterative reconstruction method first proposed by Irani and Peleg in [42]. The method they used resembles the method of solving the problem of reconstructing 2-D object from 1-D projections in *computer aided tomography*. IBP tries to minimize $||DBWf - g||_2^2$ from equation 2.1. It first makes a crude initial guess of the desired high-resolution image f^0. This crude estimate can be done in several ways, for example registering all low-resolution images on a high-resolution grid and then taking the mean of all low-resolution images in each pixel. It then tries to refine f^0 by reproducing the low-resolution observations, g_k^0, where $k = 1, 2, ..., K$ by using the observation model in equation 2.1. The error between the real low-resolution images and the reproduced ones is calculated by $\epsilon^{(i)} = \sqrt{\sum_{k=1}^{K}(g_k - g_k^i)_2^2}$, where i is the number of iterations. The error is then back-projected onto the estimated high-resolution image to improve the initial estimate. This is done iteratively according to

$$f^{i+1}(x, y) = f^i(x, y) + \frac{\lambda}{K}\sum_{k=1}^{K} W^{-1}(((g_k - g_k^i)\dot{D}) * \dot{B}), \tag{2.28}$$

where $\dot{D}$ is the up-sampling operator, $\dot{B}$ is deblurring matrix, f^{i+1} is the super-resolved image based on the high resolution estimate, f^i, after i iterations of the algorithm. This will continue for a set number of iterations, or until the error, ϵ^i is small enough. The *threshold*, for when ϵ^i is small enough, is a tunable parameter, as well as the step-size, λ, that controls how much of the error to back-project in each iteration.

2.3.2.7 Restoration

Some multi-image super-resolution algorithms have a restoration step [1][43][44] [45]. This can either be part of the registration step or the reconstruction and aims to improve the image quality but not enhancing the resolution by removing noise or blur from the image system. It can be done by filtering with a Wiener restoration filter as seen in [43][44].

2.4 Summary and rationale behind a priori choice of baseline algorithm

The motion model present in the data set is far more complex than just global translations. Frequency domain methods were consequently disregarded after the initial research. Machine learning methods require large amounts of data for training and evaluation, hence a multi-image based method in the spatial domain might be a good choice. A non-uniform interpolation for reconstruction would make it possible to try different registration methods. This method could then be evaluated and compared to the proven IBP method.

3 Method

This chapter describes the two super-resolution frameworks used in this thesis as well as how the data set was recorded. The image registration methods used will be described. Then the reconstruction methods will be presented as well as the way the results are evaluated. The framework shown in figure 3.1 describes the pipeline of the system. A number of low-resolution images are fed to the image registration. There the motion is estimated and the images aligned. The aligned images are reconstructed with the chosen algorithm into a high-resolution image.

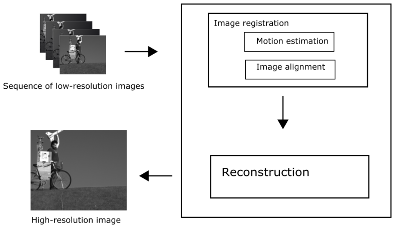


Figure 3.1: *The basic flow of the super-resolution process used in this thesis.*

3.1 Data

Most available data sets used in previous work only contains a static object or a static camera. With assistance of Image Systems new data sets were recorded. Simple motion like translations and rotations in the scene as well as small movements of the camera were desired. A resolution chart as in figure 3.3 was attached to a bicycle. The bicycle was then pulled as steadily as possible to avoid

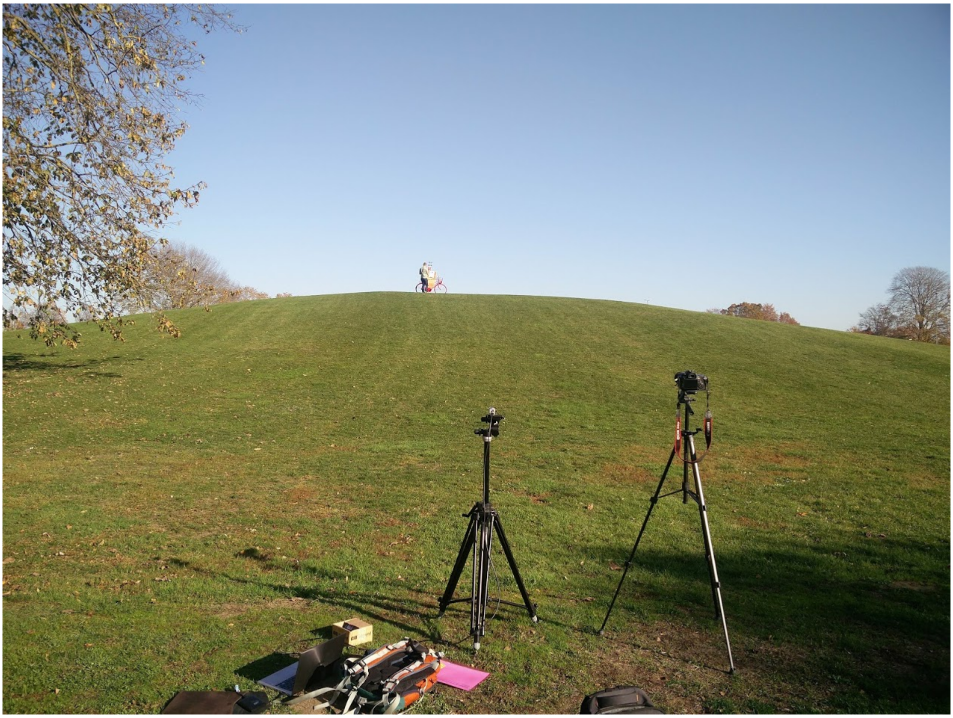


Figure 3.2: The setup for when the data used in this thesis was recorded.

zooming and warping effects in the images. The images were captured with a basler aca2440-75um camera and a Canon EOS 60D camera. The cameras were mounted on a stand, rotating as they were following the bicycle as it passed through the scene. The camera setup can be seen in figure 3.2. The location of the scene was set to a place with a uniform background in case some segmentation would be needed.

3.1.1 Synthetic data

The images captured with the cameras where of very high quality and it would be difficult to improve the resolution further. To be able to enhance the input images, synthetic images were created from the recorded data set. To simulate the down sampling effect of a camera system of lower quality the observation model in chapter 2 was used. The movement in the scene as well as the movement of the camera counts as the warping. The blurring was done by convolving the image with a 5×5 Gaussian kernel. The decimation was done by a factor of two and lastly a Gaussian noise with mean, $\mu = 0$, and standard deviation, $\sigma = 3$, was added to the decimated image. This process of simulating the down sampling effect of the camera was similar to the way Elad and Feuer did it in [46].

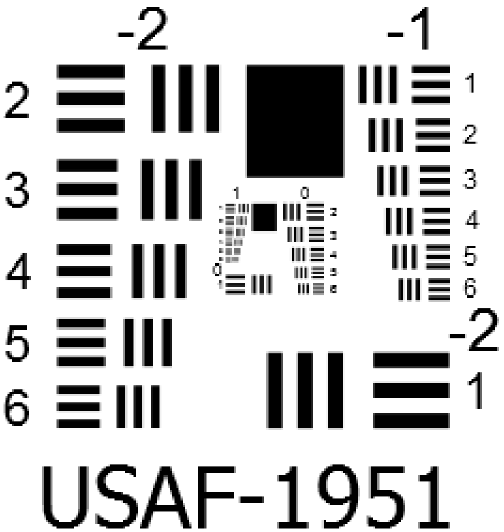


Figure 3.3: A spatial resolution test-chart often used when determining the spatial resolution an imaging system is capable of.

3.2 Image registration

The image registration for the baseline algorithm was done with the TEMA tracker developed by Image Systems. TEMA contains several tracking algorithms. There are several surveys done comparing different image registration methods [22][47]. The resolution chart is tracked in each frame using the *correlation plus* algorithm, a cross-correlation based tracker, that is described in section 2.3.2.1. It was chosen because it performed best when testing the different available trackers in TEMA.

3.3 Image reconstruction

Two reconstruction methods were implemented and evaluated. The methods were implemented using Python together with the Numpy and openCV libraries. Points of interest were passed from TEMA to the python framework. The points were used to calculate the translation and rotation of the resolution chart. The image reconstruction process is illustrated in figure 2.8.

3.3.1 Non-uniform interpolation

As described in section 2.3.2.4 the values for each pixel of the high-resolution grid are interpolated by bilinear interpolation with inverse mapping from a number of input low-resolution images. The input images used in most of the experiments were the images just before and directly after the image to be processed.

3.3.2 Iterative back-projection

The method was implemented as in [42]. First with the original image registration described in [48] and then with the image registration that uses TEMA described in section 3.2.

3.4 Evaluation

Image quality evaluation methods can be divided into subjective and objective methods [49]. Objective methods such as *Peak Signal-to-Noise Ratio* (PSNR) and *Structural Similarity Index* (SSIM) [50] compares numerical criteria with a ground truth, in this case in shape of a reference image. Even though PSNR and SSIM have become standard evaluation methods in image processing applications, their results can be somewhat misleading when discussing the quality improvements of an image [51]. The *Human Visual System* (HVS) take other metrics into account when assessing an image. For two images with almost the same PSNR value, the image quality can differ greatly if perceived by the HVS.

3.4.1 Peak signal-to-noise ratio

PSNR uses the MSE (mean square error) of the processed image, $h(x,y)$, and the high-resolution reference image, $f(x,y)$,

$$MSE = \frac{1}{mn}\sum_{x=0}^{m}\sum_{y=0}^{n}[f(x,y) - h(x,y)]^2. \tag{3.1}$$

If the processed image is similar to the reference image, the MSE will be close to zero. The PSNR is defined as

$$PSNR = 10 \times log_{10}\left(\frac{MAX_f^2}{MSE}\right) \tag{3.2}$$

and will go to infinity as MSE goes to zero. A high PSNR value indicates a high similarity between h and f. A low PSNR value implies that there are numerical differences between the images. The unit of PSNR is dB (decibel).

3.4.2 Structural similarity index

SSIM also uses a high resolution-image reference image, f, and a processed image, h, to be evaluated. But instead of calculating the difference of the two images, it compares similarities of the images by using the known quality perception of the HVS. Hence, SSIM is the product of three different measurements between the two images. The first measurement is the luminescence,

$$l(f,h) = \frac{(2\mu_f\mu_h + c_1)}{(\mu_f^2 + \mu_h^2 + c_1)}. \tag{3.3}$$

The second measurement is the contrast,

$$c(f,h) = \frac{(2\sigma_f \sigma_h + c_2)}{(\sigma_f^2 + \sigma_h^2 + c_2)}. \tag{3.4}$$

The third measurement is the structure,

$$s(f,h) = \frac{(2\sigma_{fh} + c_3)}{(\sigma_f^2 \sigma_h^2 + c_3)}. \tag{3.5}$$

Here, c_1, c_2 and c_3 are positive constants, used to avoid a zero denominator. With $c_3 = \frac{c_2}{2}$ the following equation is acquired,

$$SSIM(f,h) = l(f,h)c(f,h)s(f,h) = \frac{(2\mu_f \mu_h + c_1)(2\sigma_{fh} + c_2)}{(\mu_f^2 + \mu_h^2 + c_1)(\sigma_f^2 + \sigma_h^2 + c_2)}, \tag{3.6}$$

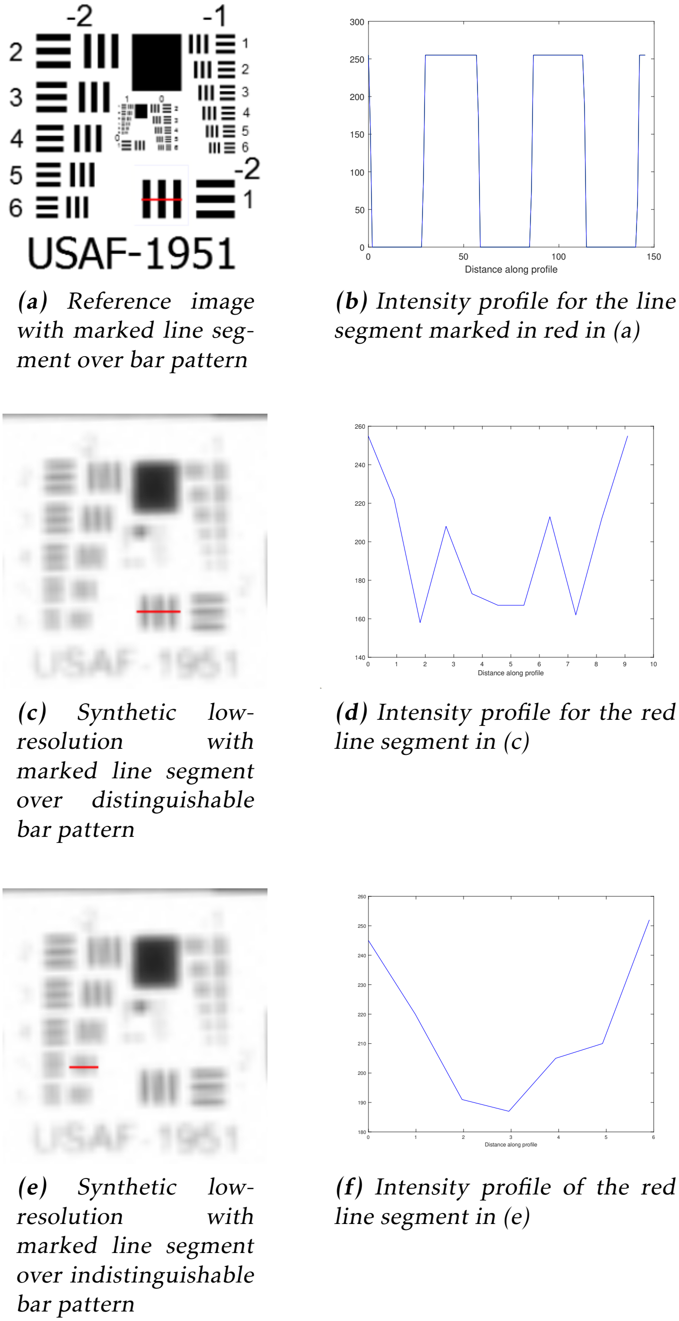
where μ_f is the average of f, μ_h is the average of h, σ_f^2 is the variance of f, σ_h^2 is the variance of h, and σ_{fh} is the covariance of f and h. The SSIM value ranges from -1 to 1, where a value of 1, means that the images are identical and a value of 0 indicates no structural similarity between the images.

3.4.3 Visual inspection

Some details might appear after being processed by the super-resolution that can be difficult to assess by the objective evaluation methods. Before and after images are compared and if the text on the resolution chart is readable only after being processed, it can be said that super-resolution is achieved.

3.4.4 Bar pattern intensity profile comparison

Another way to evaluate the algorithms is to see if new high-frequency information is introduced into the image after the processing. Thus this method by the author compares the intensity profiles of a defined line segment that spans over the par patterns in the resolution chart in figure 3.3. The intensity profile of the reference image, figure 3.4b shows a clear square wave-pattern, where the peaks correspond to the high intensity of the white pixels along the defined line segment. If an image with an intensity profile like in figure 3.4f changes to a profile with two peaks that are visible as in figure 3.4d after being processed by a super-resolution algorithm, it can be said that super-resolution is acquired.



(a) *Reference image with marked line segment over bar pattern*

(b) *Intensity profile for the line segment marked in red in (a)*

(c) *Synthetic low-resolution with marked line segment over distinguishable bar pattern*

(d) *Intensity profile for the red line segment in (c)*

(e) *Synthetic low-resolution with marked line segment over indistinguishable bar pattern*

(f) *Intensity profile of the red line segment in (e)*

Figure 3.4: *Intensity profiles of the line segment spanning over the bar patterns in figure 3.3*

4

Results

In this chapter, results from this work will be shown. Firstly, the subjective results will be presented. This includes all the super-resolved images from both the IBP algorithm and non-uniform interpolation method as well as the intensity profiles for the different methods. Secondly, the results from the objective image quality evaluation metrics will be displayed. This includes the SSIM, PSNR and MSE values.

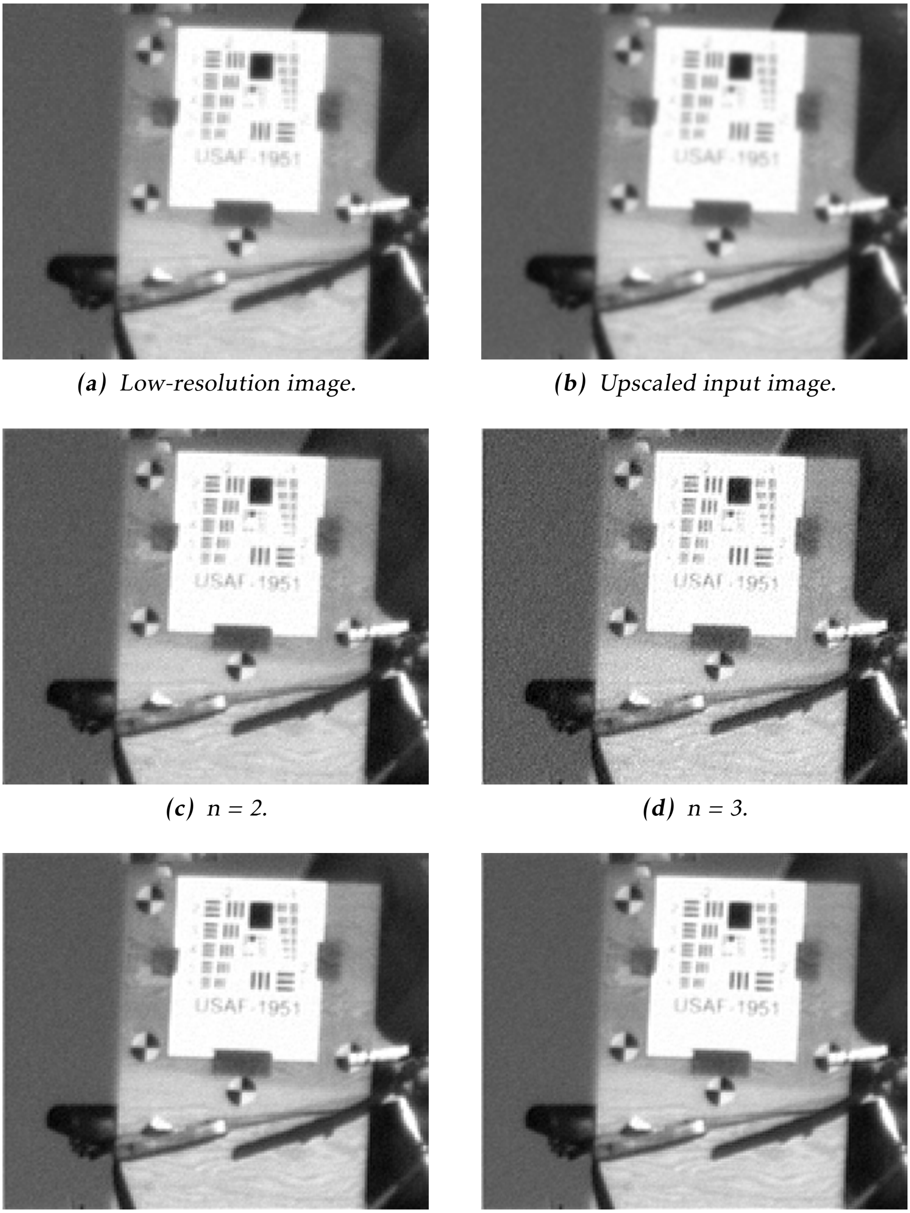
The images were cropped to show the board with bar patterns that contain several high frequency regions. All of the super-resolved images are of size (306×256) pixels. The input images are of (153×128) pixels in size but the input images presented here under the results are upscaled by a factor of 2 with nearest neighbour interpolation to make it easier to compare them to the super-resolved images.

4.1 Subjective results

Deciding if an image is of higher quality than another is highly subjective, simply looking at the images does not serve as a very reliable evaluation tool. The super-resolved images will still be presented here. It is after all a thesis about image enhancement.

4.1.1 Visual results of iterative back-projection algorithm

The IBP algorithm has some changeable parameters and can be configured in some different ways. Figure 4.1 shows the images super-resolved with IBP with a varying number of input low-resolution images. It is clear that all the super-resolved images in figures 4.1c-4.1h appear sharper than the resized input image in figure 4.1b.



(a) *Low-resolution image.*

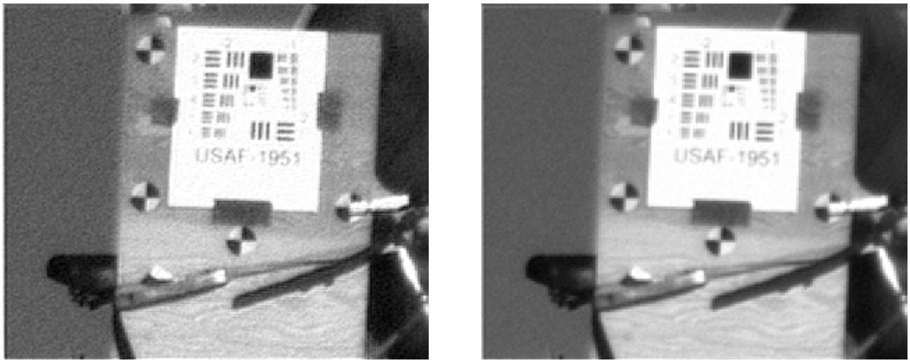
(b) *Upscaled input image.*

(c) *n = 2.*

(d) *n = 3.*

(e) *n = 4.*

(f) *n = 5.*



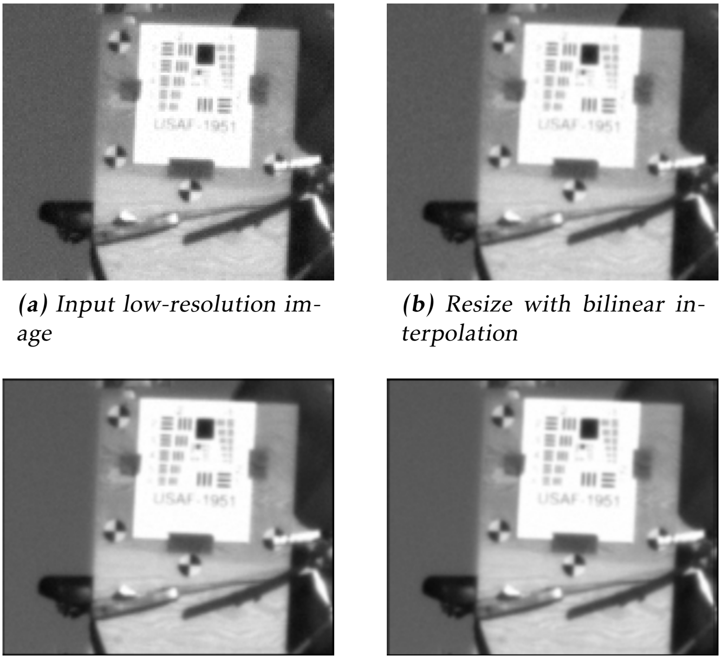
(g) *n = 6.* ***(h)*** *n = 7.*

Figure 4.1: *Visual results from the IBP algorithm. (a) shows the input image (153 × 128 pixels) to the system. The image is upscaled with nearest neighbour interpolation. (b) shows the image after bilinear resizing and (c) - (h) shows the result after IBP (306 × 256 pixels) with n number of sequential images.*

Figure 4.2: *A close up on the resized image to the left and with non-uniform interpolation super-resolved image to the right.*

4.1.2 Visual results from non-uniform interpolation

The figure 4.3 show the results of images that are super-resolved with non-uniform interpolation with a varying number of input images. Here it is not as evident that any new frequency content has emerged. Looking at the zoomed in images in figure 4.2 it could be argued that the super-resolved image looks like it conveys the content from the resolution-board in figure 3.3 more accurately and is more pleasing to the eye.



*(**a**) Input low-resolution image*

*(**b**) Resize with bilinear interpolation*

*(**c**) n = 3*

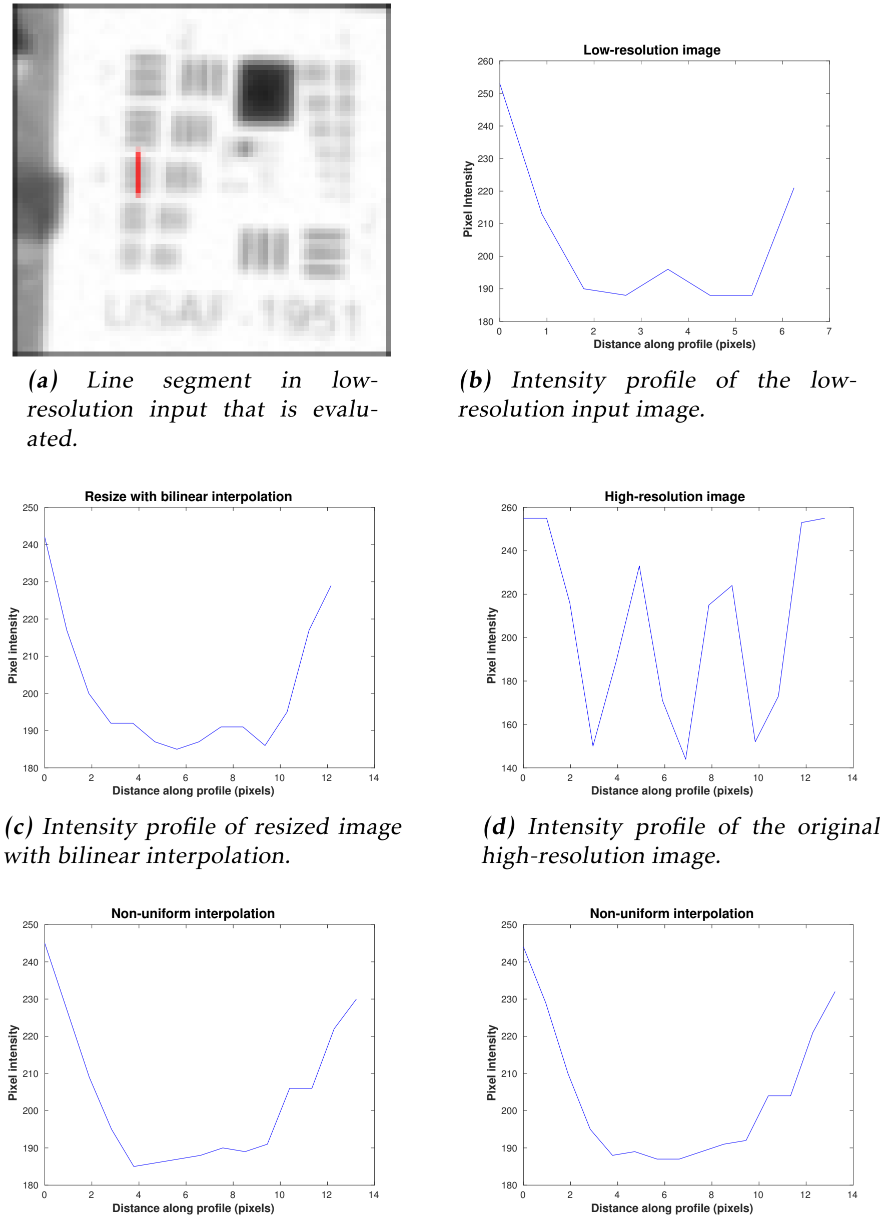
*(**d**) n = 5*

Figure 4.3: *(a) shows the upscaled low-resolution input image. (b) shows the same image resized with bilinear interpolation. (c) and (d) is the result after non-uniform interpolation three and five input images respectively.*

4.1.3 Intensity profiles

The intensity profiles for several different cases are shown. Figure 4.5 shows the intensity profiles for when some frequency content is recovered after the images have been processed by the algorithms. Figure 4.6 shows the intensity profiles for a bar pattern that is already recognizable, and shows a sharpening effect of the different methods. In figure 4.4 the intensity profile is shown for the non-uniform interpolation algorithm. In figure 4.7 the intensity profiles for IBP are presented. The input image has lost the bar pattern due to the observation model process.

The figure 4.7 shows the image profile of an area where the bar pattern is lost in the low-resolution image and the same intensity profiles from different images super-resolved with IBP.



(a) Line segment in low-resolution input that is evaluated.

(b) Intensity profile of the low-resolution input image.

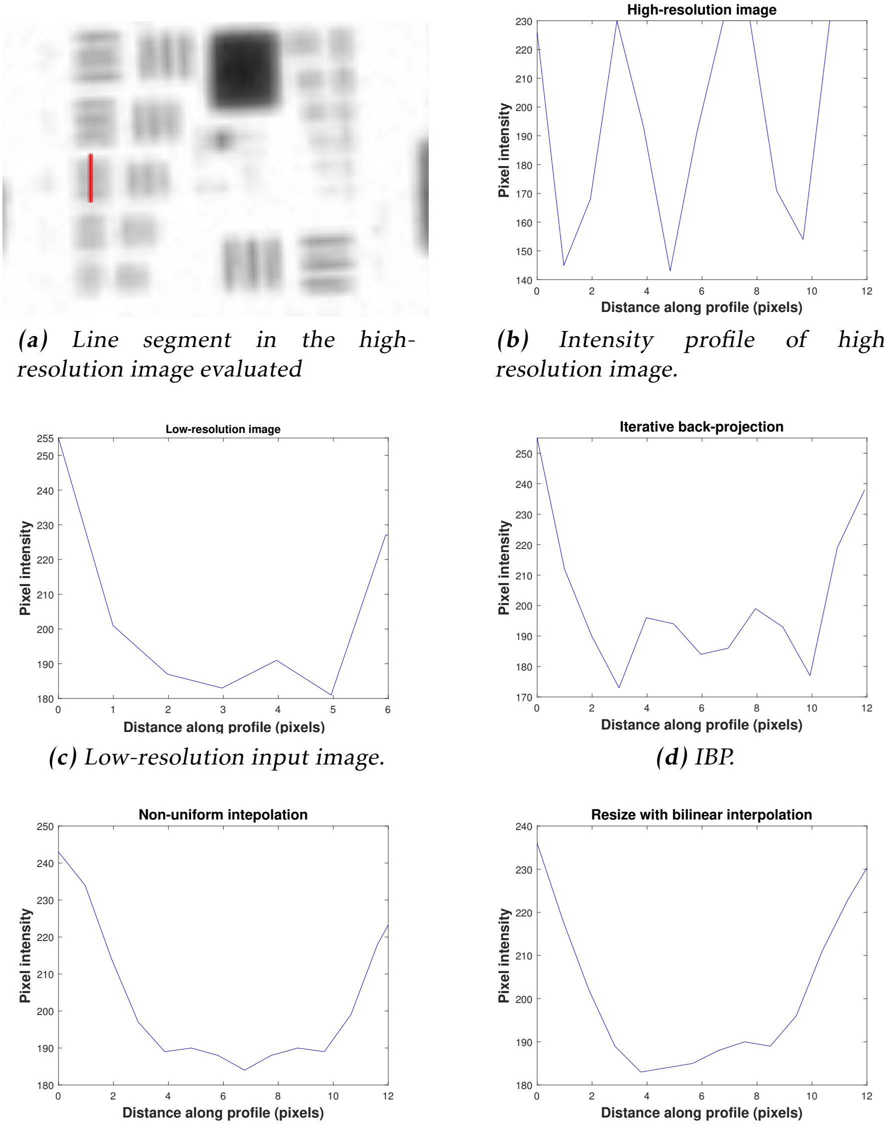
(c) Intensity profile of resized image with bilinear interpolation.

(d) Intensity profile of the original high-resolution image.

(e) Intensity profile of a super-resolved image using two contiguous frames.

(f) Intensity profile of a super-resolved image using four contiguous frames.

***Figure 4.4:** Intensity profiles of images super-resolved with non-uniform interpolation.*



(a) *Line segment in the high-resolution image evaluated*

(b) *Intensity profile of high-resolution image.*

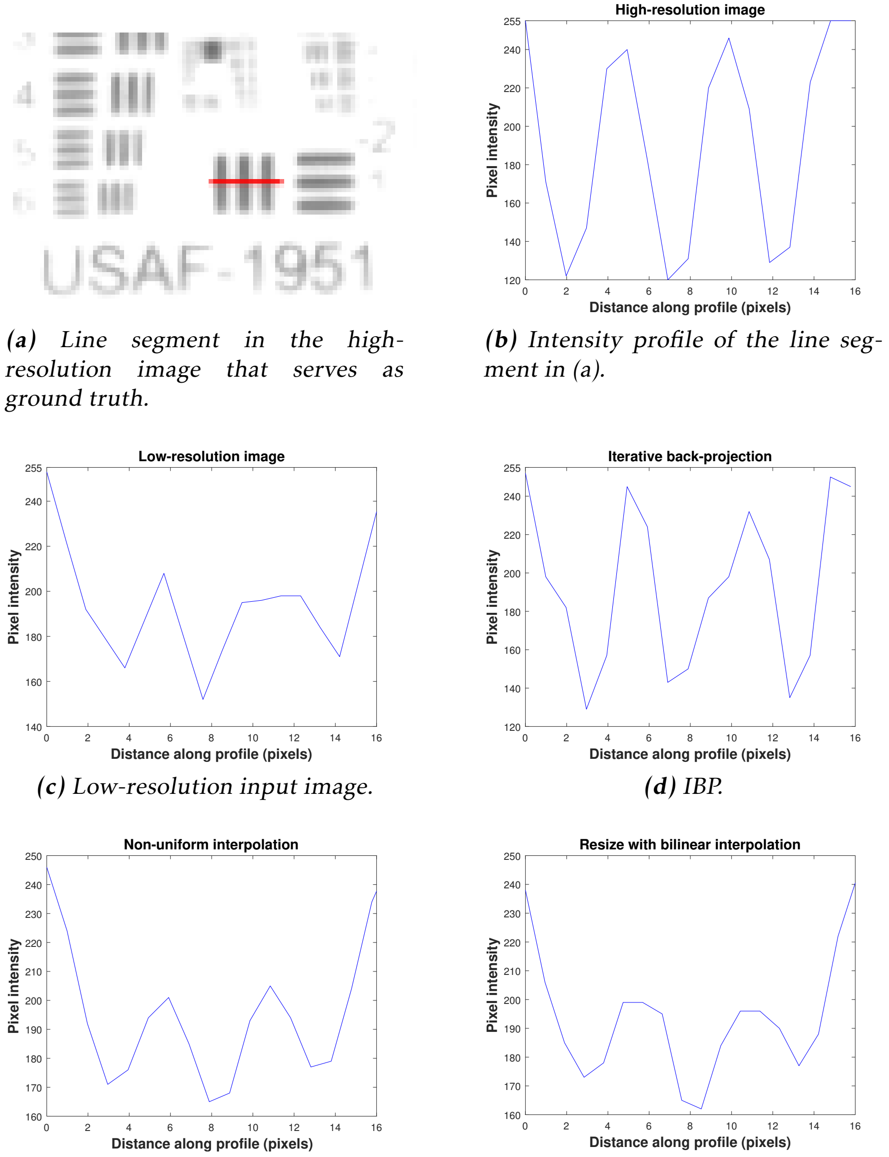
(c) *Low-resolution input image.*

(d) *IBP.*

(e) *Non-uniform interpolation*

(f) *Resize with bilinear interpolation*

Figure 4.5: *Pixel intensity profiles with (a) as input image. (b) shows the intensity profile of the high resolution image used as ground truth. (c) is the pixel intensity of the low-resolution input image. (d) and (e) are the intensity profiles of the super-resolved images while (f) is the intensity profile of a resized image, resized with bilinear interpolation.*



(a) *Line segment in the high-resolution image that serves as ground truth.*

(b) *Intensity profile of the line segment in (a).*

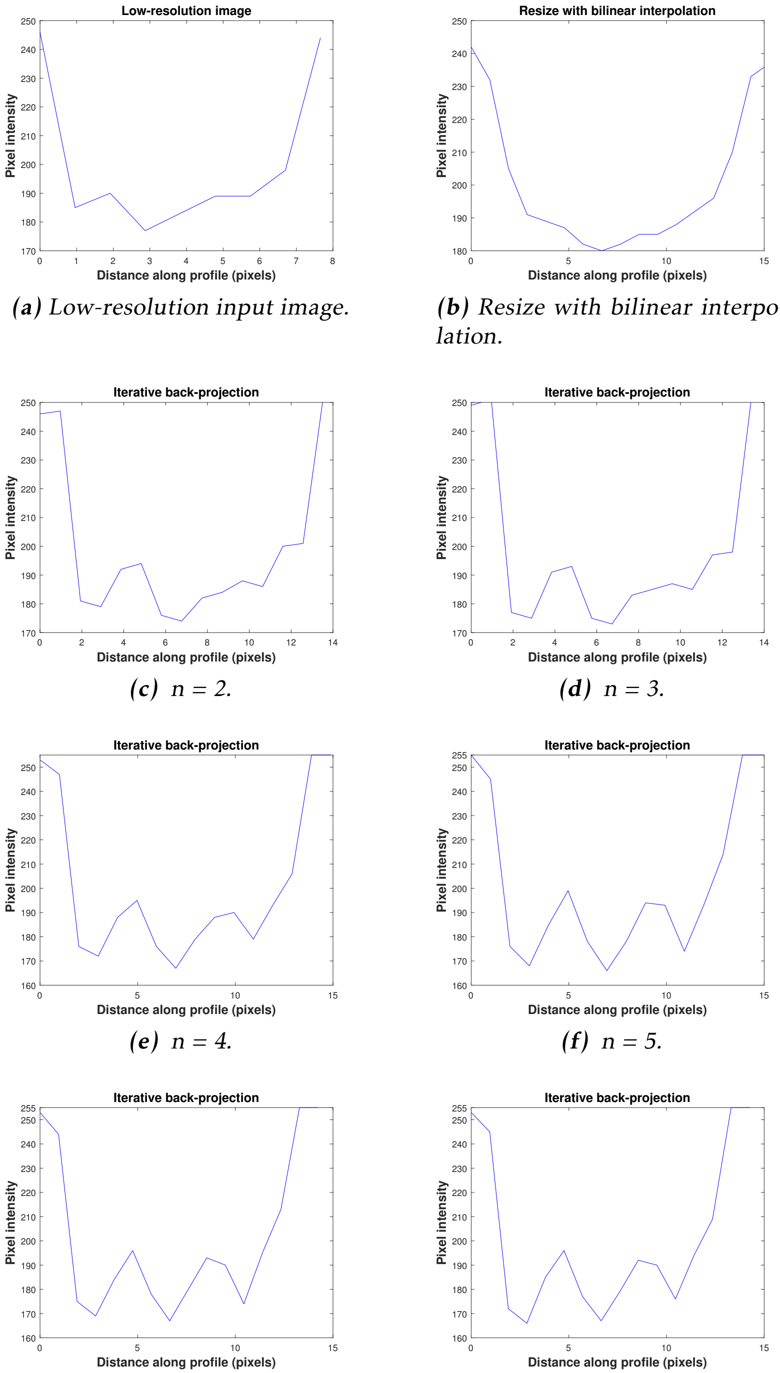
(c) *Low-resolution input image.*

(d) *IBP.*

(e) *Non-uniform interpolation.*

(f) *Resize with bilinear interpolation.*

Figure 4.6: *Pixel intensity profiles with (a) as input image. (b) shows the intensity profile of the high resolution image used as ground truth. (c) is the pixel intensity of the low-resolution input image. (d), (e) and (f) shows how the different methods preserve sharpness.*



*(**a**) Low-resolution input image.*

*(**b**) Resize with bilinear interpolation.*

(**c**) $n = 2$.

(**d**) $n = 3$.

(**e**) $n = 4$.

(**f**) $n = 5$.

(**g**) $n = 6$.

(**h**) $n = 7$.

Figure 4.7: *These plots show the intensity profiles of the profile marked in red in figure 4.6a, after being super resolved with IBP. The n indicates how many successive low-resolution images were used.*

4.2 Objective image quality results

All super-resolved images as well as images upscaled with bicubic interpolation were evaluated against the original high-resolution frame of the same scene.

4.2.1 Objective image quality results of the Iterative back-projection algorithm

The MSE, SSIM and PSNR are explained in section 3.4 and the results are listed in tables 4.1, 4.2 and 4.3 with the best value for each set of parameters in bold text. All of these results are from the same input image but with different lambda and threshold values. In table 4.4 the SSIM and PSNR values of an image that is resized with bilinear interpolation are presented.

n	MSE	SSIM	PSNR (dB)	iterations for convergence
2	45588,24	0,78	32,19	14
3	45499,28	0,64	30,55	19
4	45431,82	0,82	32,79	10
5	45433,28	0,83	32,86	9
6	45251,2	0,69	30,99	14
7	44740,88	0,83	32,95	8

Table 4.1: *Results from images super-resolved with IBP.* $\lambda = 0,01$, $threshold = 0,0001$ *and* n *tells how many sequential low-resolution images that are used.*

n	MSE	SSIM	PSNR (dB)	iterations for convergence
2	44861,16	0,85	33,65	9
3	45172,60	0,85	33,40	9
4	45497,16	0,84	33,16	9
5	45553,49	0,85	33,25	8
6	45408,86	0,84	33,07	8
7	44713,88	0,83	32,95	8

Table 4.2: *Results from images super-resolved with IBP.* $\lambda = 0,01$, $threshold = 0,001$ *and* n *tells how many sequential low-resolution images that are used.*

n	MSE	SSIM	PSNR (dB)	iterations for convergence
2	44591,03	**0,86**	**33,75**	24
3	44818,19	**0,86**	33,72	24
4	45157,17	**0,86**	33,59	24
5	45316,49	**0,86**	33,56	24
6	44996,89	**0,86**	33,63	24
7	44503,91	**0,86**	33,58	24

Table 4.3: *Results from images super-resolved with IBP. $\lambda = 0,001$, $threshold = 0,0001$ and n tells how many sequential low-resolution images that are used.*

Interpolation method	MSE	SSIM	PSNR (dB)
bilinear	43639,23	**0,86**	33,63

Table 4.4: *Results from images Resized with bilinear interpolation for reference.*

4.2.2 Objective image quality results of the non-uniform interpolation method

The MSE, SSIM and PSNR of the non-uniform interpolation are listed in table 4.5 with the best values written in bold text. For comparison the results of the same low-resolution image but resized with bilinear interpolation are presented in table 4.6.

k	n	MSE	SSIM	PSNR (dB)
2	3	42492,78	0,89	34,10
2	5	42047,07	**0,9**	34,11
3	3	43522,6	0,89	**34,12**
3	5	42436,81	0,89	33,94

Table 4.5: *Results from images super-resolved with non-uniform interpolation method. n is the number of low-resolution images used and k is which image in the sequence that is super-resolved.*

k	MSE	SSIM	PSNR (dB)
2	43639,23	0,86	33,52
3	43639,23	0,88	33,89

Table 4.6: *The PSNR and SSIM of the resized input image with bilinear interpolation compared to the high-resolution version of the scene.*

5 Discussion

In this chapter the results are discussed together with the methods used to reach those results. Methods that were not used, but could have been beneficial will also be discussed.

5.1 Data

The data set that was recorded for this thesis tried to imitate a scene were super-resolution would be useful for DIC applications together with TEMA. The footage was recorded with the premise that the images could be further enhanced with super-resolution. After the literature study and some early experimentation, it became clear that further efforts to simplify the actual motion in the scene would have been beneficial. The image registration is of utmost importance and the best results were achieved in the sequences where the movement of both the camera and the object in the scene between the images was small. Hence, a sub-pixel movement of the object that was tracked, and moving rigidly with as few changes in perspective as possible would have been ideal simplifications.

5.2 Image registration

Within the scope of this thesis, only a small object in the scene needed to be tracked. If all the motion in the scene would have been accounted for, another more advanced motion estimation algorithm would have been necessary.

5.3 Results

The results in the previous chapter indicate that accurate image registration is of great importance for a successful super-resolution method. To achieve any super-resolution at all, the movement between each frame needs to be small. Even though precautions were taken to limit the motion model to translations and rotations, the minor perspective transformation that still exists in the scene worsens the result a lot.

5.3.1 Non-uniform interpolation

The super-resolved images in figure 4.3c and 4.3d show no radical visual improvements compared to images re-sized with bilinear interpolation in figure 4.3b. The quantitative results tell another story. Both the results in table 4.5 and in the intensity profile in figure 4.5 indicate improvements compared to the re-sized image. In figure 4.5e an emergence of high resolution content can be seen, even if it is less impressive than the peaks in figure 4.5d. The number of low-resolution images used in the algorithm does not seem to improve the results. This is the same tendency as when there is too much movement between the low-resolution images. If the movement makes the image registration more imprecise, the result might improve the more images are used if there is less movement than in the data set used in this thesis.

5.3.2 Iterative back-projection

For the IBP method, both the visual and quantitative results indicate a clear improvement. The number of low-resolution images used seems to be a major factor for the results. When studying figure 4.7 the peaks indicating the recovered frequencies of the bar pattern emerge after four or more low-resolution images are used as input. If the image registration was better the results could probably be improved further. The problem that occurs is that when too many low-resolution images were used, the algorithm would not converge. This happens because the movement in the scene makes the input images later in the sequence differ too much from the image to be super-resolved, becoming images with too much different information instead of complementary information.

5.4 Methods

Super-resolution is a wide field, and it is beyond the scope of a master thesis to explore everything. The last few years with the success of machine learning in several image processing applications most of the progress within super-resolution has been made with machine learning approaches. Image Systems did at the time of the thesis work have more expertise in traditional image processing. With that in mind, a decision was made quite early to focus on the more traditional methods. The limitations of the frequency domain methods were made clear in the literature study, hence spatial domain methods were to be explored. The opportunity to use TEMA, a state-of-the-art tracker, further narrowed it down to multi-image methods.

5.4.1 Non-uniform interpolation

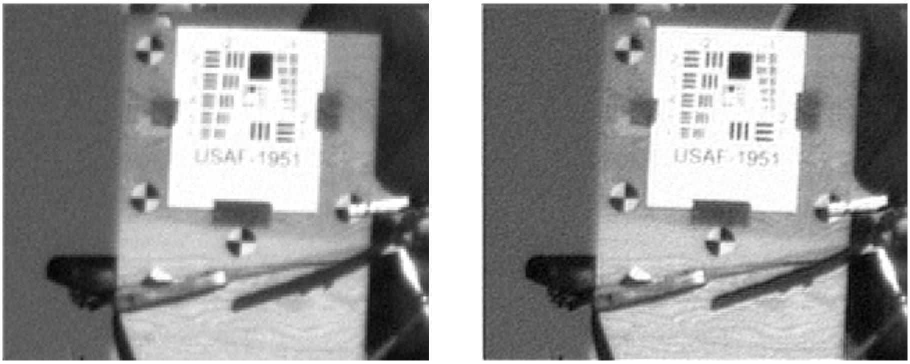
Non-uniform interpolation is maybe the most straightforward method there is. Because of the dependence on a good image registration, it ought to serve as the perfect baseline algorithm. Once implemented the result was disappointing. The disappointing results might be because of the perspective changes and the movements between each frame.

5.4.2 Iterative back-projection

The IBP was published in 1991 [42] and several similar methods have been developed before and after [52][53][54]. Numerous other works have been based on it as well [55][56]. Because of that it makes sense to evaluate the method on this data set. The algorithm has several tunable parameters. Some of the parameter settings had difficulties to converge. When using too many input images or when the search step length was too large or small the algorithm would not converge. The sometimes large movement between the frames probably made the images too different. It also had some trouble with noise artefacts as seen in figure 5.1 with some parameter settings. Though the image was sharper, and high frequency content was extracted, the images looked noisy. This could maybe be addressed with further parameter tuning. Some parameters were changed in the results gathered in table 4.1, 4.2 and 4.3 but the problem would not go away entirely. It was less evident in table 4.3 with a small step length and the lowest threshold for the error-function that resulted in a large number of iterations of the algorithm, 24 iterations to be precise.

5.4.3 Machine learning

A machine learning approach for example using a CNN could probably work well. Since many of the learning based methods use feature mapping instead of image registration, the challenges with complex image registration could be ignored. Another approach would be to use CNNs for the image registration itself. It has been done before in [57], [58], [59] and [60] to name a few. The drawback would be that deep learning methods usually depends on large quantities of training data to perform well, even though some methods have produced promising results with smaller data sets as in SRCNN [19]. The CNN is trained on low-resolution and high-resolution image pairs, where areas in the images are mapped to each other. If the low-resolution images are synthetically produced the annotation process is very simple and it would be possible to create large training data sets.



(a) *Super-resolved with IBP.*

(b) *Super-resolved with IBP with some ringing noise but sharp.*

Figure 5.1: *Both images are super-resolved with IBP. In the image to the left, two input images are used. The image to the right, is sharp but somewhat noisy. Six low-resolution input images were used.*

6

Conclusion

In this chapter the questions asked in chapter 1.3 will be answered with the conclusions drawn from the results in chapter 4 and the following discussion during chapter 5.

6.1 Has super-resolution been achieved?

The purpose of this thesis was to examine if super-resolution was achievable when there is motion in the scene as well as motion of the camera. To answer that some questions were proposed in the introductory chapter.

6.1.1 Is super-resolution achievable in the case of motion in both the scene and the camera?

Yes it is. Even though the non-uniform interpolation and IBP algorithms are possibly the simplest implementations there are, the results from both methods show the emergence of high resolution information and it can because of that claim that super-resolution has been achieved.

6.1.2 Is the baseline algorithm that uses TEMA for motion estimation able to support super-resolution?

Yes. Figure 4.5d and 4.4e shows that new frequency content has emerged and therefore super-resolution is achieved.

6.1.3 How does one of the state-of-the-art algorithms compare to the baseline algorithm?

The IBP algorithm managed to outperform the non-uniform interpolation method when comparing the intensity profiles. It was evident that new frequency content had emerged. When looking at the PSNR and SSIM the non-uniform interpolation method performed clearly best, but both algorithms performed better than bilinear interpolation. The lower scores in PSNR and SSIM might be because of ringing and grainy artefacts in the images enhanced with IBP. With some post-processing of the the images, or some further parameter tuning, it is not unthinkable that the IBP method would improve the PSNR and SSIM scores and outperform non-uniform interpolation method in the objective evaluation methods as well.

6.2 Importance of image registration

When implementing these two algorithms and during the experimentation on the data set it has been evident that these multi-image based methods are completely dependent on an accurate image registration.

6.2.1 Data set

In the experiments, it could be noted that the smaller the movements in between each frame, the better the results. It would be a good idea to start out smaller and evaluate algorithms on longer sequences with very small motion and then work the way up to successively more challenging cases. With smaller movement between the frames and a data set with purely translations and rotations of the tracked object in the scene, the results would probably be more impressive.

6.3 Direction of future work

First of all, further parameter tuning and experimentation would be interesting as well as evaluation on data sets with even more complex motion present. In this thesis super-resolution for dynamic scenes and cameras have been explored for only two methods. There are a vast amount of other methods to examine. Because of the moderate success of these two multi image methods as well as the popularity and success of the learning based methods, it would be very interesting to investigate how they perform on these kind of data sets. Also to investigate how much training that would be needed for different kinds of motion models in the data sets.

Bibliography

[1] 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, May 2003. Cited on pages 1 and 18.

[2] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. Advances and challenges in super-resolution. *International Journal of Imaging Systems and Technology*, 14(2):47–57, 2004. Cited on page 1.

[3] Image Systems Motion Analysis. Tema, 2019. Cited on page 2.

[4] T. Huang and R. Tsai. Multi-frame image restoration and registration. *Advances in Computer Vision and Image Processing*, pages 317–339, 1984. Cited on page 6.

[5] P. Vandewalle, S. Süsstrunk, and M. Vetterli. A frequency domain approach to registration of aliased images with application to super-resolution. *EURASIP Journal on Advances in Signal Processing*, 2006(1):071459, Dec 2006. Cited on page 6.

[6] S. Borman and R. L. Stevenson. Super-resolution from image sequences-a review. In *Super-Resolution from Image Sequences-A Review*, 1998. Cited on page 6.

[7] B. S. Reddy and B. N. Chatterji. An fft-based technique for translation, rotation, and scale-invariant image registration. *Trans. Img. Proc.*, 5(8):1266–1271, August 1996. Cited on page 8.

[8] S. Chaudhuri and J. Manjunath. *Motion-Free Super-Resolution*. Springer Publishing Company, Incorporated, 1st edition, 2010. Cited on page 9.

[9] K. Nasrollahi and T. Moeslund. Super-resolution: A comprehensive survey. *Machine Vision and Applications*, 25:1423–1468, 08 2014. Cited on page 9.

[10] Sun, J., Zheng, N.N., Tao, H. , and Shum, H. Y. . Image hallucination with primal sketch priors. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 2, pages II–729, June 2003. Cited on page 9.

[11] S. Roth and M. J. Black. Fields of experts. *International Journal of Computer Vision*, 82(2):205, Jan 2009. Cited on page 9.

[12] J. Kim, J. K. Lee, and K. M. Lee. Accurate image super-resolution using very deep convolutional networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1646–1654, June 2016. Cited on page 9.

[13] W.S. Lai, J.B. Huang, N. Ahuja, and M.H. Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. Cited on page 9.

[14] Y. Tai, J. Yang, and X. Liu. Image super-resolution via deep recursive residual network. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. Cited on page 9.

[15] K. Zhang, W. Zuo, S. Gu, and L. Zhang. Learning deep cnn denoiser prior for image restoration. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. Cited on page 9.

[16] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang. Deep networks for image super-resolution with sparse prior. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015. Cited on page 9.

[17] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. Cited on page 9.

[18] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. Cited on page 9.

[19] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *CoRR*, abs/1501.00092, 2015. Cited on pages 9 and 39.

[20] J. Yang, J. Wright, T. S. Huang, and M. Yi. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11):2861–2873, 11 2010. Cited on page 9.

[21] M. C. Chiang and T. E. Boult. Efficient super-resolution via image warping. *Image and Vision Computing*, 18(10):761 – 771, jul 2000. Cited on page 10.

[22] B. Zitová and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977 – 1000, oct 2003. Cited on pages 10 and 21.

[23] J. J. Gibson. *The Perception Of The Visual World*. Boston: Houghton Mifflin, 1950. Cited on page 11.

[24] L. D. Bruce and T. Kanade. An iterative image registration technique with an application to stereo vision. In *In IJCAI81*, pages 674–679, 1981. Cited on page 11.

[25] B. Pan, K. Qian, H. Xie, and A. Asundi. Two-dimensional digital image correlation for in-plane displacement and strain measurement: a review. *Measurement Science and Technology*, 20(6):062001, apr 2009. Cited on page 12.

[26] D. I. Barnea and H. F. Silverman. A class of algorithms for fast digital image registration. *IEEE Transactions on Computers*, C-21(2):179–186, Feb 1972. Cited on page 12.

[27] C. G. Broyden. The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations. *IMA Journal of Applied Mathematics*, 6(1):76–90, 03 1970. Cited on page 12.

[28] R. Fletcher. A new approach to variable metric algorithms. *The Computer Journal*, 13(3):317–322, 01 1970. Cited on page 12.

[29] D. Goldfarb. A family of variable-metric methods derived by variational means. *Mathematics of computation*, 24(109):23–26, 1970. Cited on page 12.

[30] D. F. Shanno. Conditioning of quasi-newton methods for function minimization. *Mathematics of computation*, 24(111):647–656, 1970. Cited on page 12.

[31] E. Catmull and R. Rom. A class of local interpolating splines. *Computer Aided Geometric Design - CAGD*, 74, 12 1974. Cited on page 12.

[32] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, Nov 2004. Cited on page 13.

[33] E. Rosten and T. Drummond. Machine learning for high-speed corner detection. In *Proceedings of the 9th European Conference on Computer Vision - Volume Part I*, ECCV'06, pages 430–443, Berlin, Heidelberg, 2006. Springer-Verlag. Cited on page 13.

[34] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, 2, 06 2005. Cited on page 13.

[35] R.K. McConnell. Method of and apparatus for pattern recognition, 1 1986. Cited on page 13.

[36] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981. Cited on page 13.

[37] J. Yen. On nonuniform sampling of bandwidth-limited signals. *IRE Transactions on Circuit Theory*, 3(4):251–257, 1956. Cited on page 16.

[38] Wikipedia, the free encyclopedia. Comparison of 1d and 2d interpolation, 2016. [Online; accessed February 14, 2019]. Cited on page 17.

[39] R. Keys. Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(6):1153–1160, December 1981. Cited on page 16.

[40] I. Ragnemalm. *Polygons feel no pain: A course book in Computer Graphics with OpenGL*. Ragnemalm Utveckling och Underhållning, 2017. Cited on page 16.

[41] Wikipedia, the free encyclopedia. Catmull-rom spline interpolation with 4 points, 2013. [Online; accessed November 9, 2019]. Cited on page 17.

[42] M. Irani and S. Peleg. Improving resolution by image registration. *CVGIP: Graphical Models and Image Processing*, 53(3):231 – 239, may 1991. Cited on pages 18, 22, and 39.

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

[44] M. S. Alam, J. G. Bognar, R. C. Hardie, and B. J. Yasuda. Infrared image registration and high-resolution reconstruction using multiple translationally shifted aliased video frames. *IEEE Transactions on Instrumentation and Measurement*, 49(5):915–923, Oct 2000. Cited on page 18.

[45] B. Narayanan, R. C. Hardie, K. E. Barner, and M. Shao. A computationally efficient super-resolution algorithm for video processing using partition filters. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(5):621–634, May 2007. Cited on page 18.

[46] M. Elad and A. Feuer. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE transactions on image processing*, 6(12):1646–1658, 1997. Cited on page 20.

[47] L. Gottesfeld Brown. A survey of image registration techniques. *ACM Comput. Surv.*, 24(4):325–376, December 1992. Cited on page 21.

[48] D. Keren, S. Peleg, and R. Brada. Image sequence enhancement using subpixel displacements. In *Proceedings CVPR '88: The Computer Society Conference on Computer Vision and Pattern Recognition*, pages 742–746, June 1988. Cited on page 22.

[49] A. Hore and D. Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th International Conference on Pattern Recognition*, pages 2366–2369, Aug 2010. Cited on page 22.

[50] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: From error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13:600 – 612, 05 2004. Cited on page 22.

[51] K. Nelson, A. Bhatti, and S. Nahavandi. Performance evaluation of multi-frame super-resolution algorithms. 12 2012. Cited on page 22.

[52] B. C. Tom and A. K. Katsaggelos. Resolution enhancement of video sequences using motion compensation. In *Proceedings of 3rd IEEE International Conference on Image Processing*, volume 1, pages 713–716 vol.1, Sep. 1996. Cited on page 39.

[53] B. Bascle, A. Blake, and A. Zisserman. Motion deblurring and super-resolution from an image sequence. In B. Buxton and R. Cipolla, editors, *Computer Vision — ECCV '96*, pages 571–582, Berlin, Heidelberg, 1996. Springer Berlin Heidelberg. Cited on page 39.

[54] S. Peleg, D. Keren, and L. Schweitzer. Improving image resolution using subpixel motion. *Pattern Recognition Letters*, 5:223–226, 03 1987. Cited on page 39.

[55] B. Cohen and I. Dinstein. Polyphase back-projection filtering for image resolution enhancement. *IEE Proceedings - Vision, Image and Signal Processing*, 147(4):318–322, Aug 2000. Cited on page 39.

[56] A. Zomet and S. Peleg. Efficient super-resolution and applications to mosaics. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, volume 1, pages 579–583 vol.1, Sep. 2000. Cited on page 39.

[57] J. Caballero, C. Ledig, A. P. Aitken, A. Acosta, J. Totz, Z. Wang, and W. Shi. Real-time video super-resolution with spatio-temporal networks and motion compensation. *CoRR*, abs/1611.05250, 2016. Cited on page 39.

[58] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. van der Smagt, D. Cremers, and T. Brox. Flownet: Learning optical flow with convolutional networks. *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec 2015. Cited on page 39.

[59] N. Mayer, E. Ilg, P. Hausser, P. Fischer, D. Cremers, A. Dosovitskiy, and T. Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2016. Cited on page 39.

[60] J. J. Yu, A. W. Harley, and K. G. Derpanis. Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness. *CoRR*, abs/1608.05842, 2016. Cited on page 39.