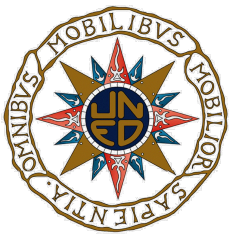
TESIS DOCTORAL

SUPER - RESOLUTION:

MULTI - FRAME REGISTRATION AND INTERPOLATION USING OPTIMAL PROJECTIONS ON FUNCTIONAL SPACES

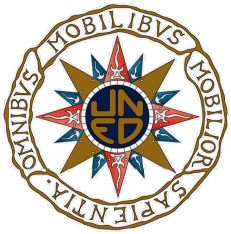
ALFONSO SÁNCHEZ-BEATO SUÁREZ
INGENIERO DE TELECOMUNICACIÓN



DEPARTAMENTO DE INFORMÁTICA Y AUTOMÁTICA
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA

UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA (UNED)

2 0 0 8



Departamento de Informática y Automática
Escuela Técnica Superior de Ingeniería Informática

SUPER - RESOLUTION:

MULTI - FRAME REGISTRATION AND INTERPOLATION USING OPTIMAL PROJECTIONS ON FUNCTIONAL SPACES

ALFONSO SÁNCHEZ-BEATO SUÁREZ
INGENIERO DE TELECOMUNICACIÓN

DIRECTOR DE TESIS
DR.D. GONZALO PAJARES MARTINSANZ

TUTOR
DR.D. JOSÉ SÁNCHEZ MORENO

Divide each difficulty into as many parts as is feasible
and necessary to resolve it.

Discourse on the Method, 1637

René Descartes

Agradecimientos

En primer lugar me gustaría agradecer a la Universidad Nacional de Educación a Distancia y a la Universidad Complutense de Madrid la excelente oportunidad que me han brindado a través del programa "Ingeniería de Sistemas y Automática" de realizar el doctorado. Debido a las complicaciones que supone el compaginar estudios y trabajo, me habría resultado difícil llevar a buen término la tesis de otra forma. Gracias a este programa he tenido además libertad para trabajar en temas en los que tengo un gran interés personal, lo que me ha permitido realizar un trabajo muy gratificante y realizador.

Quería agradecer especialmente a Gonzalo Pajares todo el apoyo que he recibido por su parte. Jamás ha escatimado ni tiempo ni esfuerzo en ayudarme, encontrándome cámaras fotográficas y artículos, reuniéndome con él o revisando mis trabajos. Le considero un amigo y espero que sigamos colaborando de una u otra forma tras la presentación de esta tesis.

Extiendo los agradecimientos al profesor Jesús Manuel de la Cruz y a la profesora Eva Besadas por su abierta disposición y apoyo en lo que respecta a las cámaras antes mencionadas junto con la ayuda proporcionada. Me gustaría dar también las gracias a Ángela Ribeiro y Juan Ranz por su ayuda con ciertos experimentos llevados a cabo con cámaras térmicas del CSIC.

En cuanto a los medios instrumentales para escribir esta tesis, he hecho uso de diversos programas Open Source, entre los que destaco TEXy LATEX, creados por Donald Knuth y Leslie Lamport respectivamente. Sin sus aportaciones, la tipografía y la presentación de esta tesis se habrían resentido considerablemente.

Finalmente, quiero agradecer a mis padres y a mi hermano su apoyo incondicional. En especial, quiero agradecer a mis padres todos los sacrificios que han hecho por mí

durante tanto tiempo. Siempre me han apoyado y aconsejado pensando en qué era mejor para mí. Ellos han sido siempre mi modelo a la hora de hacer frente a la vida. Esto es para vosotros, con todo mi cariño.

Abstract

The objective of super-resolution (SR) techniques is to produce a high resolution and high quality image using as starting point a series of low resolution and low quality images from the same scene. The images can come from many sources: a video, a sequence of photographs from the same scene, medical images, satellite images, etc. In general, SR is useful when it has been impossible to obtain suitable images due to high cost problems, physical limitations of the system or other causes.

Super-resolution is one of the imaging digital signal processing problems that is more open nowadays, besides the huge quantity of works that have appeared since the first paper that touched this problem, which was published in 1,984 ("Multi-frame image restoration and registration", by Tsai and Huang). Among the premises that are necessary for SR to be possible is the need for aliasing to be present in the images, which means that the images must be sampled under the Nyquist frequency, with movement among themselves. In this way, each low resolution image contains different information and it is possible to extract a bigger resolution image from the sequence. The problems or stages that we need to solve to obtain the final high resolution image are:

1. The register of the images. It presents special difficulties due to the fact that we are registering aliased images, so the common information among them is less than in other image processing problems.
2. Interpolation. We must pass from a series of irregular samples to a regular sampling grid that can be addressed algorithmically and easily visualized.
3. Restoration. Here we include noise reduction and deblurring the distortions due to the lenses and the sensor.

These problems can be solved individually or jointly: we do not intend to suggest with

the previous list that the SR problem is always separable in this way. The methods employed in SR have been widely varied: methods in the frequency domain, based on projections on convex sets, posing an algebraic system, based on polynomial local approximations, using wavelets, etc.

Besides of this, super-resolution techniques have not arrived yet to matureness. In general, they are quite restrictive in the allowed motion models, computationally very costly, and not always effective. Thus, there is an ample field of possible improvements that remains open.

In this dissertation we present novel algorithms in each of the previously exposed SR stages:

1. A simultaneous registration method for multiple images that establishes relationships among the motion transformations of all the images to achieve a more accurate registration. In the presented experiments we demonstrate better performance with regard to classical registration methods and also to other multi-frame registration methods.
2. Interpolation. We demonstrate that the implicitly assumed data model in most SR methods leads to the appearance of aliasing artifacts in the reconstructed images. We propose a novel method that projects orthogonally on the desired basis the low resolution images samples, following classical sampling theory. In this way, we are able to remove aliasing artifacts in the reconstructed images, achieving quantitatively and visually better quality regarding to other SR methods.
3. Restoration. One of the biggest problems in SR is the appearance of outliers due to problems in the sensor, misregistration, etc. To address these problems, people resort to robust SR methods (that is, with robust estimation mechanisms). We propose a median filter for irregular samples to remove outliers as a previous step before applying the SR method of choice. We have demonstrated that this method is able to obtain better results than other robust SR algorithms.

Resumen

El objetivo de las técnicas de súper-resolución (SR) es producir una imagen de alta resolución y alta calidad a partir de una serie de imágenes de baja resolución y escasa calidad de una misma escena. Las imágenes pueden provenir de muy variadas fuentes: un vídeo, una secuencia de fotos tomadas de una misma escena, imágenes médicas, de satélite, etc. En general, la SR resulta de utilidad cuando nos ha sido imposible tomar imágenes de una calidad adecuada debido a problemas de coste, limitaciones físicas del sistema u otras causas.

La SR es uno de los problemas de tratamiento digital de imágenes que se encuentra más abierto hoy en día a pesar de los numerosos trabajos aparecidos desde el primer artículo que trataba este problema, el cual fue publicado en 1.984 ("Multi-frame image restoration and registration", por Tsai y Huang). Entre las premisas que son necesarias para que la SR sea posible se encuentran que las imágenes presenten aliasing, es decir, que se encuentren muestreadas por debajo de la frecuencia de Nyquist, y que las imágenes presenten movimiento entre ellas. De esta forma, cada imagen de baja resolución contiene distinta información y es posible extraer una imagen de mayor resolución de la secuencia. Los problemas que es necesario resolver para alcanzar la imagen de alta resolución final son:

1. El registro de las imágenes. Éste presenta una especial dificultad debido a que registramos imágenes con aliasing, por la que la información que poseen en común es menor que en otros problemas de procesamiento de imágenes.
2. Interpolación. Debemos pasar de una serie de muestras irregulares a un muestreo regular en rejilla que pueda ser fácilmente tratado algorítmicamente y visualizado.

3. Restauración. Aquí se incluye la reducción de ruido y la corrección de desenfoque debido a las lentes y al sensor.

Estos problemas pueden resolverse de forma individual o bien conjuntamente, no pretendiendo sugerir con la lista anterior que el problema de la SR sea separable siempre de esta forma. Los métodos empleados en la SR han sido extremadamente variados: métodos en el dominio de la frecuencia, basados en proyección sobre conjuntos convexos, planteamiento de un sistema algebraico, aproximaciones polinómicas locales, basados en wavelets, etc.

A pesar de esto, las técnicas de súper-resolución aún no han llegado a la madurez. En general, son bastante restrictivas en los modelos de movimiento permitidos, computacionalmente muy costosas y no siempre eficaces. Existe, por tanto, un amplio campo de posibles mejoras que permanece abierto.

En esta tesis presentamos algoritmos novedosos en cada una de las etapas de la SR que hemos expuesto:

1. Un método de registro simultáneo de múltiples imágenes que establece relaciones entre las transformaciones de todas las imágenes para conseguir un registro más preciso. En los experimentos realizados demostramos un rendimiento superior con respecto a métodos clásicos de registro y otros métodos de registro para múltiples imágenes
2. Interpolación. Demostramos que el modelo implícitamente asumido en la mayoría de los métodos de SR lleva a la aparición de aliasing en las imágenes reconstruidas. Se propone un nuevo método que proyecta ortogonalmente sobre la base deseada las muestras de las imágenes de baja resolución, siguiendo la teoría clásica de muestreo. De esta forma, conseguimos que las imágenes reconstruidas no contengan artefactos debidos al aliasing, consiguiéndose así una mayor calidad tanto cuantitativamente como cualitativamente con respecto a otros métodos de SR.
3. Restauración. Uno de los mayores problemas en SR es la aparición de puntos no válidos (*outliers*) debidos a problemas del sensor, registro, etc. Para combatir estos problemas se recurre a métodos de SR robustos (es decir, con mecanismos de estimación robustos). Proponemos un filtro de la mediana sobre muestras irregulares como paso previo de eliminación de *outliers* antes de aplicar el algoritmo de

SR. Hemos demostrado que este método consigue mejores resultados que otros algoritmos de SR robustos.

Contents

List of Abbreviations **19**

List of Symbols **21**

List of Figures **23**

List of Tables **25**

1 Introduction **27**

1.1 Problem Description 27

1.2 Applications for Super-Resolution 33

1.3 Motivation and Objectives 34

1.4 Contributions of Thesis 34

1.5 Outline of Thesis 35

2 Literature Review **37**

2.1 Introduction 37

2.2 Early Approaches 38

2.3 Super-Resolution as an Estimation Problem 40

2.4 Super-Resolution in its Stages 42

2.4.1 Super-Resolution and Registration 42

2.4.2 Super-Resolution and Interpolation 43

2.4.3 Super-Resolution and Deblurring 44

2.5 Limits of Super-Resolution 45

2.5.1 Influence of Point Spread Function 46

2.5.2 Influence of Signal-to-Noise Ratio 49
2.6 Comparison of Methods 50

3 Multi-Frame Registration for Super-Resolution 55
3.1 Introduction 55
3.2 Registration Models 57
3.2.1 Translational Motion 59
3.2.2 Affine Motion 59
3.2.3 Projective Transformation 60
3.2.4 Hierarchical Motion Estimation 61
3.3 Method Description 61
3.4 Experiments 65
3.4.1 Methodology 65
3.4.2 Shift 66
3.4.3 Rotation 66
3.4.4 Comparison with other methods 68
3.5 Fast method 68
3.6 Influence of registration in Super-Resolution 70
3.7 Conclusions 72

4 Interpolation: Optimal Projections on Functional Spaces 73
4.1 Introduction 73
4.2 The Data Model 74
4.3 A Sampling Theory Interpretation 75
4.4 Proposed Algorithm 78
4.4.1 Prefiltering Using B-Spline Basis 78
4.4.2 Implementation for Cubic B-splines 82
4.5 Competing Approaches 83
4.5.1 Regularization Priors 83
4.5.2 The l_2 Norm Method 85
4.5.3 The l_1 Norm Method 85
4.5.4 The Shift and Add Method 86
4.5.5 The Zomet Method 87

4.5.6 The Delaunay Triangulation Method 88
4.6 Experiments 88
4.6.1 Experiments with Synthetic Data 88
4.6.2 Experiments with Real Data 91
4.7 Conclusion 93
4.A Interchange of Blurring and Warping Operators 94

5 Non-Stationary Noise Removal and Deblurring 107
5.1 Introduction 107
5.2 Median Filter for Irregular Samples 108
5.3 Multi-Frame Deblurring 110
5.4 Performance Analysis of Robust Noise Removal Methods 112
5.5 Performance Analysis of Deblurring Methods 115
5.6 Conclusions 118

6 Conclusions and Future Work 123
6.1 Conclusions 123
6.1.1 Multi-Frame Registration 124
6.1.2 Fusion of Data Using Projections on Functional Spaces 124
6.1.3 Super-Resolution Restoration 125
6.2 Future Work 125

Bibliography 129

Index 139

List of Abbreviations

BTV	Bilateral Total Variation.
CCD	Charged-Coupled Device.
CMOS	Complementary Metal Oxide Semiconductor.
CT	Computed Tomography.
FLIR	Forward Looking Infrared.
HR	High Resolution.
IBP	Iterative Back-Projection.
IIR	Infinite Impulse Response.
JPEG	Joint Photographic Experts Group.
LR	Low Resolution.
LSI	Linear-Shift Invariant.
LTI	Linear Time-Invariant.
MAP	Maximum a Posteriori.
ML	Maximum Likelihood.
MPEG	Moving Picture Experts Group.

MRI Magnetic Resonance Imaging.

MSE Mean Squared Error.

PET Positron Emission Tomography.

POCS Projection Onto Convex Sets.

PSF Point Spread Function.

PSNR Peak Signal-to-Noise Ratio.

RBF Radial Basis Functions.

SNR Signal-to-Noise Ratio.

SR Super-Resolution.

TV Total Variation.

UAV Unmanned Aerial Vehicles.

List of Symbols

$T_{i,j}$	Warping between images i and j.
Γ	Operator in Tikhonov regularization.
Ω	Regularization cost function.
β^n	B-spline of order n.
d_k	Downsampling operator for kth low resolution image.
g_k	Warping operator for kth low resolution image.
h_k	Blurring operator for kth low resolution image.
n_k	Additive noise for kth low resolution image.
w_k	Observed discrete kth low resolution image.
z	Ideal continuous ground truth image.
$\mathbf{A}_k$	Result of the multiplication $\mathbf{D}_k\mathbf{H}_k\mathbf{G}_k$.
$\mathbf{A}$	System matrix that relates $\mathbf{z}$ to $\mathbf{w}$.
$\mathbf{D}_k$	Downsampling operator for LR image k in matrix form.
$\mathbf{G}_k$	Warping operator for LR image k in matrix form.
$\mathbf{H}_k$	Blurring operator for LR image k in matrix form.
$\mathbf{n}_k$	Additive noise for LR image k in matrix form.
$\mathbf{n}$	Additive noise for all low resolution images stacked in one vector.
$\mathbf{p}_{i,j}$	Function parameters of warping between images i and j.
$\mathbf{u}_{i,j}$	Displacement vector between images i and j.
$\mathbf{w}_k$	Observed discrete kth low resolution image, vector form.

$\mathbf{w}$ Low resolution images stacked in one vector.

$\mathbf{z}_k$ Discretized high resolution image, vector form.

List of Figures

1.1 Measurement of a real-world scene 28
1.2 Tasks for achieving Super-Resolution 30
1.3 SR from a black and white video 31
1.4 SR from a color video 32

2.1 CCD sensor with microlenses. 47
2.2 Transfer functions and image spectra. 48
2.3 Anti-aliasing filter from a digital video camera. 49

3.1 Transformations and restrictions for multi-frame registration 62
3.2 Original images used in the experiments: Texture and Television 65
3.3 MSE ratios for shifts for Texture and Television images 67
3.4 MSE ratios for rotations for Texture and Television images 67
3.5 MSE: Comparison of methods for shifts and rotations 69
3.6 MSE ratios for fast methods 70
3.7 Influence of registration in super-resolution 71

4.1 Standard three-step sampling paradigm. 78
4.2 Integration areas for a quadratic B-spline 81
4.3 Radial function sequence and generated LR images 96
4.4 Super-resolved images for experiment one 97
4.5 Chart sequence and generated LR images 98
4.6 Super-resolved images for experiment two with SR factor 2 99
4.7 Super-resolved images for experiment two with SR factor 4 100

4.8 Mobile phone sequence and generated HR image 101
4.9 Zoom of area of the experiment three sequence 102
4.10 Clock sequence and generated HR image 103
4.11 Zoom of area of the experiment four sequence 104
4.12 Input images for experiment five and super-resolved image 105

5.1 Median filter for irregular samples 109
5.2 Simulated registration errors for resolution chart 113
5.3 Images produced by robust methods for resolution chart 114
5.4 Images produced by robust methods for clock sequence 115
5.5 Detail of images from robust methods for clock sequence 116
5.6 Images produced deblurring the chart sequence 117
5.7 Images produced deblurring the clock sequence 119
5.8 Detail of images produced deblurring the clock sequence 120

List of Tables

2.1 Comparison of SR methods 52

4.1 Parameter Values and Results for Experiment One 89
4.2 Parameter Values and Results for Experiment Two 91
4.3 Parameter Values for Experiment Three 92
4.4 Parameter Values for Experiment Four 93

5.1 Parameter Values and Results for Robust SR Experiment One 113
5.2 Parameter Values and Results for Robust SR Experiment Two 113
5.3 Parameter Values and Results for SR Deblurring Experiment One 118
5.4 Parameter Values and Results for SR Deblurring Experiment Two 118

Chapter 1

Introduction

1.1 Problem Description

In most imaging applications there is a tradeoff between the resolution and other parameters. For instance, to increase the resolution the pixel size is reduced in the sensor, but this reduces the amount of light available. It also generates shot noise that degrades the image quality severely. Another example is the processing and storage limitations that appear when taking a video with a mobile phone, which make the manufacturer limit the video resolution for these devices. Sensor technology is also important: while charged-coupled devices (CCD) and CMOS image sensors for current cameras offer sufficient resolution for most applications, forward looking infrared (FLIR) sensors have still many limitations. This happens also in medical applications: positron emission tomography (PET), computed tomography (CT), etc. All these devices offer a limited resolution.

Besides of increasing the resolution, the idea of reducing the noise, deblur or minimize other degradations of an image emerges naturally when we have a sequence of images that include the same region of interest in all of them. The aim of Super-Resolution (SR) is to solve simultaneously these two problems: it uses a sequence of images to increase resolution and to improve the quality of the image. We can then define Super-Resolution as a signal processing technique that creates images of bigger resolution and better quality (high resolution images, HR) from a series of images of small resolution and low quality (low resolution images, LR). This can be extended to a problem of simultaneous resolution enhancement in time as well as in space, which increases the

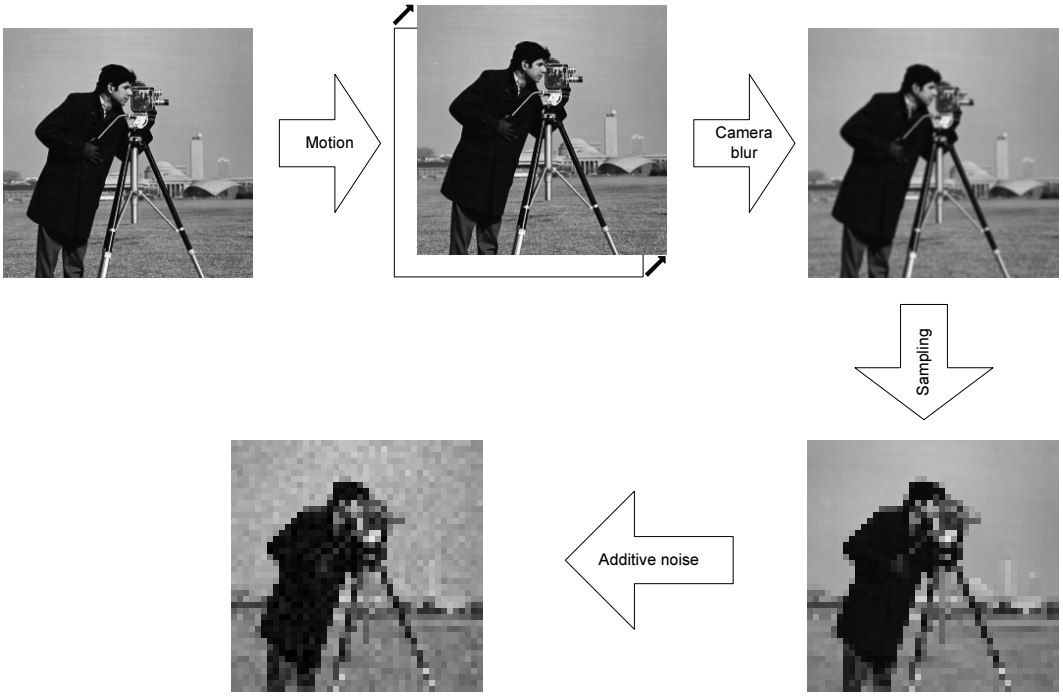


Figure 1.1: Transformations that a real-world scene suffers until it is converted to a digital image.

temporal (frames per second) as well as the spatial resolution of a video (Robertson and Stevenson, 2001; Shechtman et al., 2002).

Now that we have defined the concept, it is of the utmost importance to define the process of formation of the input images for SR algorithms. The sequence of steps that is followed when an image is acquired can be seen in Fig. 1.1. Initially, an ideal continuous image is the object of a transformation, which might be a shift, a rotation, an affine or projective transform, or a more complex motion. This transformation models the location of our imaging system (the LR image plane) as regards the real world scene plane that we will take as our reference for the reconstruction of the HR image. Then, the image is blurred by the camera, including this blur the Point Spread Function (PSF) of both the lens and the sensor (optical and acquisition blur). The image is finally sampled, that is, discretized, and additive noise due to the acquisition process is also incorporated to the samples. This process for a given LR image k can be modeled by the equation

$$w_k[m,n] = d_k\left(h_k(x,y) * (z(g_k(x,y)))\right) + n_k[m,n] \quad \text{for} \quad k = 1,\dots,N, \tag{1.1}$$

where the parentheses are used for functions defined in $\mathbb{R}^2$ and the square brackets

denote functions defined in $\mathbb{Z}^2$. In the equation we suppose that we have N LR images, $z(x, y)$ is the continuous HR image, $g_k(x, y)$ is the transformation from the reference plane to the imaging system plane for a given LR image k, $h_k(x, y)$ is the blurring for image k, $d_k : \mathbb{R}^2 \longrightarrow \mathbb{Z}^2$ is a sampling operator that converts from the continuous to the discrete domain, $n_k[m, n]$ represents the process additive noise, $*$ represents convolution, and $w_k[m, n]$ is the discrete image finally obtained. This equation is a slightly modified version of the one that can be found in Irani and Peleg (1993).

The discretization is the key to achieve super-resolution: if the image on the sensor plane is properly sampled at the Nyquist frequency, we will have the same information in each of the LR images and no resolution improvement is possible. At most, by using the available LR images, we could reduce the noise and improve the deblurring, but we could not increase the spatial resolution and show high frequency details. Therefore, the aliasing is fundamental: we need it in the LR images to make super-resolution possible. In fact, SR can be seen as the unfolding of the aliased high-frequency information present in a set of images from the same scene. This also means that if there is some kind of anti-aliasing filter in the imaging system it will not be possible to increase the resolution more than what is allowed by the low pass filter. Taking a look at (1.1), we can also point out that if d_k, h_k, and g_k are the same for all the available images it is again not possible to improve the resolution even when aliasing is present, as we would be sampling always at the same points of $z(x, y)$. The images would be the same but for the noise, converting it in a restoration problem again. As d_k and h_k usually are the same for a sequence of images, since we normally take all of them with the same imaging system, this implies that there must be a relative movement among the LR images to make SR possible. That is, in the terminology of computer vision, the images must be misregistered among themselves.

To solve the inverse problem of finding the function z present in (1.1) we need to accomplish the following tasks:

1. Image registration. We need to move all the samples of the LR images to a common reference frame. It is fundamental for achieving super-resolution, as we need to perform it with subpixel precision. In fact, it is the main bound for improving SR that current algorithms find (Robinson and Milanfar, 2006). This is because of the aliased nature, needed to make SR possible, of the LR images we

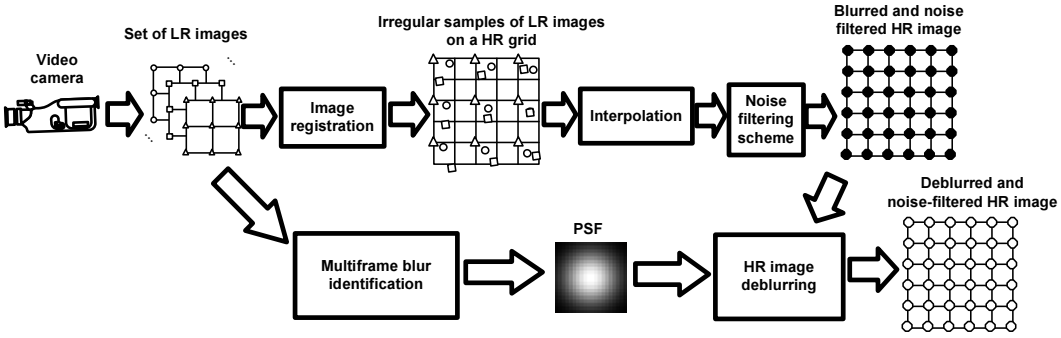


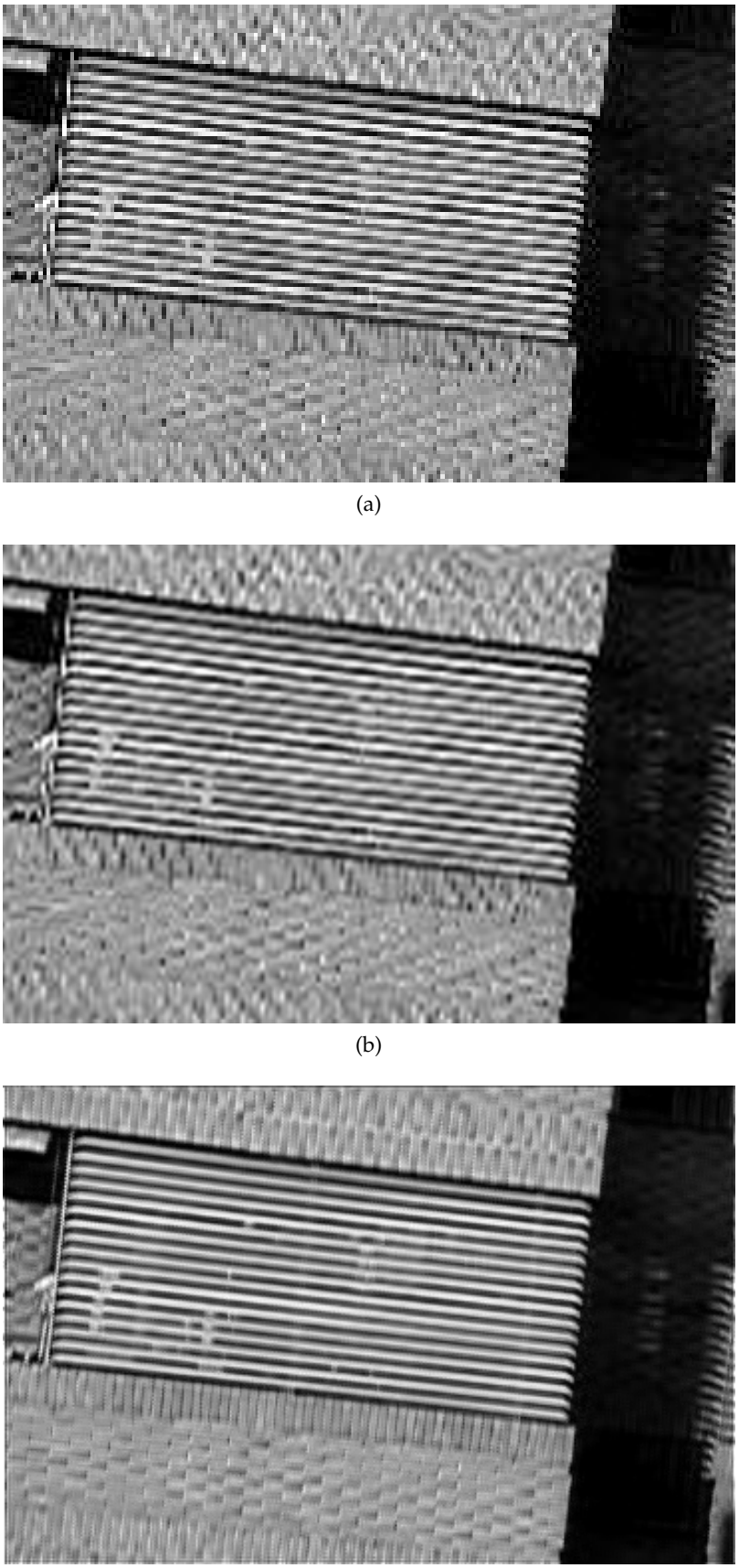
Figure 1.2: Tasks for achieving Super-Resolution

want to register.

2. Interpolation. Or fusion or resampling. Once we have all samples in a common reference, we will use that information to find values for the rectangular grid of the HR image.
3. Restoration. Including deblurring and noise dumping. The deblurring problem can be considered blind (unknown PSF) or not depending on the information we have of the imaging system. With very few exceptions (Borman and Stevenson, 2004; Sroubek et al., 2007), the PSF is normally assumed to be linear shift-invariant (LSI). The noise is usually considered white and Gaussian, although robust methods can be applied under different noise distributions (examples of robust methods are Zomet et al. (2001) and Farsiu et al. (2004b)).

These tasks do not need to be solved separately, many SR algorithms group some or all of them to have more chances of reaching a global optimization, but at the price of increased complexity. In Fig, 1.2 we can see an example of an SR framework that solves these problems separately. Firstly, the images are registered. Then, the LR samples are used to find a first approximation of the HR image through interpolation. This is combined with a noise filtering scheme. Finally, blind deblurring is performed: the PSF is estimated from the LR images and the HR image is deblurred with it as input.

To end this section we provide two illustrative examples in figures 1.3 and 1.4. In the first of them, an image is super-resolved using a black and white low-resolution video taken with a common photographic camera. It is clearly seen that the heavily aliased areas where bricks appear in Fig. 1.3a have well-defined borders in the super-resolved image in Fig. 1.3c. In the second one, SR is applied to the three channels of a color video taken with a mobile phone. It is seen that resolution is improved in Fig. 1.4c regarding

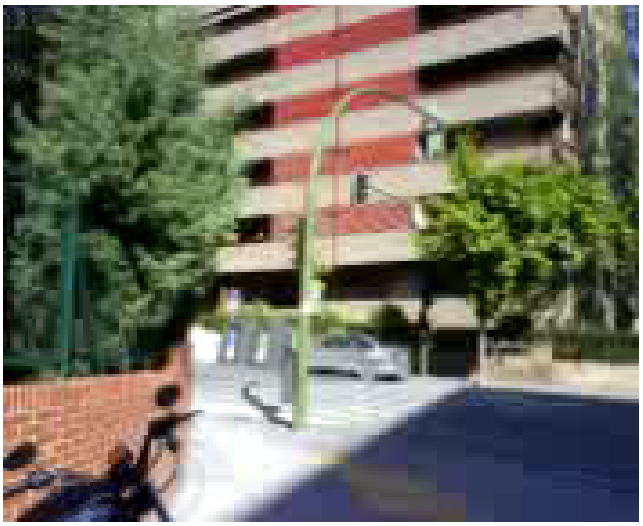


(a)

(b)

(c)

Figure 1.3: (a) A low resolution image from a black and white sequence, (b) its interpolation using cubic splines when multiplying by two its resolution, and (c) the super-resolved version using 20 images from the sequence with an SR factor of two.



(a)

(b)

(c)

Figure 1.4: (a) A low resolution image from a color sequence, (b) its interpolation usig cubic splines when multiplying by two its resolution, and (c) the super-resolved version using 10 images from the sequence with an SR factor of two.

to image Fig. 1.4a, especially in the borders of the building and that JPEG compression artifacts are reduced.

1.2 Applications for Super-Resolution

Applications for the techniques of super-resolution restoration from image sequences appear to be growing rapidly as the theory gains exposure. Continued research and the availability of fast computational machinery have made these methods increasingly attractive in applications requiring the highest restoration performance. If we consider that in many applications the increase in sensor resolution comes at a high cost and that increase could be carried out using SR techniques, we can easily see that in many cases is preferable to employ SR instead of increasing physically the sensor resolution. Of course in some cases this is difficult: complicated registration of the images or resolution already limited by the diffraction limit may prevent the use of SR techniques. But as SR theory starts to stabilize, applications start to soar.

SR restoration techniques have already been applied to problems in:

- Black and white photography (most papers apply SR to them, like in Irani and Peleg (1991) or Schultz and Stevenson (1996)).
- Satellite imaging (Tsai and Huang, 1984; Rochefort et al., 2006; Merino and Núñez, 2007).
- Astronomical imaging (Sheppard et al., 1996; Hunt, 2004).
- Video enhancement and restoration (Farsiu et al., 2006b).
- Video standards conversion (Patti et al., 1995).
- Confocal Microscopy (Wilson and Hewlett, 1991).
- Applied jointly with mosaicing (Capel and Zisserman, 1998).
- Aperture displacement cameras (Lenz and Lenz, 1990).
- Diffraction tomography (Schatzberg and Devaney, 1992).
- Restoration of MPEG-coded video streams (Segall et al., 2001).
- Magnetic Resonance Imaging (MRI) (Greenspan et al., 2002).
- Positron Emission Tomography (PET) (Kennedy et al., 2006).
- Applied to photographic camera color images (Farsiu et al., 2006a).
- Forward looking infrared cameras (Hardie et al., 1997; Pham et al., 2006).

The already wide number of applications continues to grow as the effectiveness of SR methods increases. In fact, there are commercial software packages already available to improve video resolution for the general public (MotionDSP, 2008).

1.3 Motivation and Objectives

The motivation behind SR is quite clear: there are many situations where the resolution that can give a sensor is limited because of physical or economical constraints. SR can improve the resolution in many cases where other techniques are not feasible. The ample range of applications that we exposed in the previous section exemplifies this. The specific objectives that we pursue in this dissertation are explained below.

In the super-resolution literature, the representation of the system (1.1) as a discrete linear system has been the mainstream in most recent publications. The trend is to solve the problem as a whole, performing jointly as most tasks as possible: registration, interpolation, deblurring and restoration. Although supposedly we could then achieve a global optimum that would be unattainable otherwise, this leads to a very complex problem with too many degrees of freedom.

One of the main ideas behind this dissertation is to demonstrate that we can achieve the same or even better performance in SR if we attack the different tasks separately but using at all moments all the available information. Therefore, we have made different contributions to solve separately the different parts of the SR problem and we have compared the results with methods that perform jointly most SR tasks. In all cases, we have achieved similar or better performance than these methods, as will be seen in the next chapters.

1.4 Contributions of Thesis

The contributions of this thesis span through the different stages of SR:

- *Registration.* We have introduced a new multi-frame registration method that forces the registration to be coherent among all frames. It differs from similar methods in a per-pixel weight that allows the method to work correctly under motion models different from just shifts among the images.

- *Interpolation*. We have demonstrated that the discrete model usually employed in SR implies that the function to be estimated is a bandwidth limited function, which is not always necessarily true. To avoid the aliasing artifacts that this assumption generates, we have created a new SR method that applies a prefiltering step to the data from the LR images. The proposed method employs B-splines as projection basis and Delaunay triangulation as a mean of passing from discrete samples to a continuous function.
- *Restoration*. One of the main problems in SR is the abundance of outlier samples that appear due to registration errors, different lighting conditions, etc. We introduce a novel mean filter for irregular samples that erases these outliers before the SR process begins, which makes it perform better than other robust SR methods (Zomet et al., 2001; Farsiu et al., 2004b) that treat outliers at the same time at which the HR image is being estimated.

1.5 Outline of Thesis

This dissertation is structured as follows:

1. Chapter 1 describes the SR problem and exposes the motivation and objectives of the thesis. The contributions that have been done are enumerated.
2. Chapter 2 provides a review of the most significative results in the field and relates them to the research made for the thesis.
3. Chapter 3 introduces a novel multi-frame registration method that allows multiple registration models.
4. Chapter 4 exposes a new SR method that erases aliasing artifacts from the reconstructed images.
5. Chapter 5 exposes methods for SR restoration, including deblurring and denoising. A novel mean filter for irregular samples is introduced.
6. Chapter 6 finalizes the dissertation, drawing some conclusions and also exploring directions for future research.

Chapter 2

Literature Review

2.1 Introduction

There have already been many proposals for achieving SR. We can cite as most significative those proposed in Tsai and Huang (1984), Kim et al. (1990), Ur and Gross (1992), Irani and Peleg (1993), Hardie et al. (1997), Elad and Feuer (1997), Shekarforoush and Chellappa (1999), Elad and Hel-Or (2001), Zomet et al. (2001), Lertrattanapanich and Bose (2002), Farsiu et al. (2004b), Borman and Stevenson (2004), Pham et al. (2006), and He et al. (2007). There are also some good tutorials that provide an excellent introduction to the field, like Borman and Stevenson (1998), Park et al. (2003), and Farsiu et al. (2004a).

In this chapter we will make a review of the most important approaches to SR in the literature and also of the results in the important topic of limits in SR. We have divided the remaining of the chapter in five sections: section 2.2 will expose the earliest methods invented for SR and will serve also as a historical introduction to the matter. Section 2.3 reviews SR methods that try to solve the problem as a whole defining SR as an ill-posed problem and solving it in the context of estimation theory. Section 2.4 looks at research that has split SR in different tasks and solved them separately. In section 2.5 we take a look at the very important issue of the limitations that real imaging systems pose for super-resolution. There has been not too much research in this topic, but we include it in this chapter due to its importance for practical systems. Finally, section 2.6 presents a table summarizing the differences among the most important SR methods.

2.2 Early Approaches

Super-Resolution was first proposed in a seminal paper by Tsai and Huang (1984), who suggested a frequency domain approach to super-resolution. This initial approach assumes a translational movement among the LR images and is based on the following principles:

- The shifting property of the Fourier transform.
- The aliasing relationship between the continuous Fourier transform of the ideal HR image and the discrete Fourier transform of the observed LR images.
- The assumption that the ideal HR image is band-limited.

Registration and interpolation is performed simultaneously, thanks to the shifting property of the Fourier transform. The original method provided no protection against noise, being addressed that drawback in Kim et al. (1990), Kim and Su (1993), and Bose et al. (1993).

The main contribution of the frequency domain approach was the theoretical establishment of a clear relationship between possible increments of spatial resolution and aliasing. However, the observation model is restricted to only global translations and LSI blur. It is also difficult to transfer any possible a priori knowledge that we had in the spatial domain to the frequency domain. Another important drawback is the implicit supposition that we are dealing with ground-truth band-limited images, limiting the data model. Chapter 4 presents a novel SR method that does not make this supposition.

Another early approach is Projection Onto Convex Sets (POCS). This is a set theoretic method that was first applied to SR by Stark and Oskoui (1989) and then extended to include observation noise in Tekalp et al. (1992). POCS is based on defining convex sets of restrictions on a vectorial space with as many dimensions as pixels has the HR image z. The successive projection of a point of the vectorial space on the convex sets ends up leading to the intersection of the above mentioned sets if that intersection exists. The projection on a set is understood as an operator that relates any point in the space to the nearest point that belongs to the set. In each iteration, to obtain the point z^{n+1} from z^n we would use the formula

$$z^{n+1} = P_m P_{m-1} \cdots P_1 z^n, \tag{2.1}$$

where each P_i would be a projection operator onto a restriction set. That restrictions sets are defined using the LR images. For each pixel of each LR image, the following restrictions set is defined:

$$C(m_1, m_2, k) = \left\{ y(n_1, n_2) : |r^{(y)}(m_1, m_2, k)| \leq \delta_0(m_1, m_2, k) \right\}, \tag{2.2}$$

where

$$r^{(y)}(m_1, m_2, k) = g_k(m_1, m_2) - \sum_{(n_1, n_2)} y(n_1, n_2) h(n_1, n_2; m_1, m_2, k) \tag{2.3}$$

is the residue of element y of the restrictions set with regard to the one that generated the LR image $g_k(m_1, m_2)$. The quantity δ_0 determines the size of the set and influences several properties of the algorithm. A low value could allow us to approach to a better HR image, but it also can make the algorithm numerically unstable. A big δ_0 can also protect us against noise. The set (2.2) includes all possible HR images that could have generated that pixel from LR image k.

The advantages of POCS are its simplicity and the easy way in which a priori information can be incorporated to the algorithm. The disadvantages are non-uniqueness of solution and slow convergence.

Irani and Peleg (1991) proposed an SR algorithm similar to iterative back-projection (IBP), a technique commonly used in computed tomography. This method was further improved to consider a more general motion model (Irani and Peleg, 1993). The IBP method is similar to some methods used for solving systems of linear equations. In each iteration we have an estimation $z^{(n)}$ of the HR image from which a set of low resolution images will be generated. Those images will be compared with the original LR images w_k to obtain $z^{(n+1)}$. To calculate the $w_k^{(n)}$, we will use

$$w_k^{(n)} = d^s \left((z^{(n)}(g_k(x, y)) * h \right), \tag{2.4}$$

where d^s is a subsampling operator by a factor s, h is the blurring operator, $g_k(x, y)$ is the warping of the coordinates from the kth LR image to the coordinates system of z, and $*$ represents a convolution. To obtain the next estimation of z, we have

$$z^{(n+1)} = z^{(n)} + \frac{1}{N} \sum_{k=1}^{N} \left((u^s \circ (w_k - w_k^{(n)}))(g_k^{-1}(x, y)) * p \right), \tag{2.5}$$

where N is the number of LR images, u^s is an oversampling operator by a factor s, and p is a back-projection kernel. This kernel should ideally be the inverse of h, but in that case it could not exist or the numerical calculations with it could be numerically unstable. Using a kernel that is not the inverse of h is valid whenever p fulfills $\|\delta - h * p\| < 1$; this increases the numerical stability of the algorithm but makes the convergence slower. In the presence of noise, the mean of the LR images that is made in (2.5) reduces the additive noise, which is a significant advantage of this algorithm. The main problem of this method is that it is not always convergent.

2.3 Super-Resolution as an Estimation Problem

Quite early (Cheeseman et al., 1996; Schultz and Stevenson, 1996), it was seen that super-resolution could be seen as an estimation problem that could be solved in a Maximum Likelihood (ML) or Maximum a Posteriori (MAP) framework. In these approaches, eq. (1.1) is discretized to build the linear system

$$\mathbf{w}_k = \mathbf{A}_k \mathbf{z} + \mathbf{n}_k \quad \text{for} \quad k = 1, \ldots, N, \tag{2.6}$$

where $\mathbf{w}_k$ is a vector with the lexicographically ordered pixels of the LR image k, $\mathbf{z}$ is a vector with the lexicographically ordered pixels of the HR image, $\mathbf{A}_k$ is a matrix where each row contains the contributions of different HR pixels to a certain pixel from the LR image, and $\mathbf{n}_k$ is additive noise. Vector $\mathbf{z}$ is supposed to contain the coefficients of a properly sampled HR image, being in that case the coefficients the values multiplying a *sinc* function as well as the values of the function in the sampling point.

The $\mathbf{w}_k$ and $\mathbf{n}_k$ vectors and the $\mathbf{A}_k$ matrices can be stacked to form the system of linear equations

$$\mathbf{w} = \mathbf{A}\mathbf{z} + \mathbf{n}, \tag{2.7}$$

where $\mathbf{w}$, $\mathbf{A}$, and $\mathbf{n}$ contain the stacked data from eq. (2.6) for the different LR images. The discretization that is made of the HR image has not received too much attention, and a criticism of that is one of the main contributions of this thesis, as we will see in chapter 4: we prove there that the restriction of $\mathbf{z}$ to being band-limited to half the sampling frequency can be too restrictive.

Elad and Feuer (1997) reinterpreted the linear system to approach the classic theory of restoration of a single image from linear blur and additive noise, proposing the equation

$$\mathbf{w}_k = \mathbf{D}_k\mathbf{H}_k\mathbf{G}_k\mathbf{z} + \mathbf{n}_k \quad \text{for} \quad k = 1, \ldots, N \tag{2.8}$$

as a decomposition of (2.6). The relationship with (1.1) can be seen here more directly: $\mathbf{G}_k$ represents the geometric warping, $\mathbf{H}_k$ the linear blurring and $\mathbf{D}_k$ the decimation operator that produces the k-th LR image. The vectors and matrices for all the LR images can be stacked again to produce (2.7). The linear system described by (2.8) is the canonical model that is currently used in super-resolution related research. In Elad and Feuer (1997) it is proved that the observation models of POCS, ML and MAP approaches can be seen as particular expressions of (2.8).

The ML solution to (2.7) is

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}} \|\mathbf{A}\mathbf{z} - \mathbf{w}\|_{l_p}^p, \tag{2.9}$$

where the l_p norm is used to measure distances. Usually $p = 2$ and then $\hat{\mathbf{z}}$ will be the least-squares solution to (2.7), which will be an optimal ML estimator of $\mathbf{z}$, provided that $\mathbf{n}$ is zero mean Gaussian white noise. Farsiu et al. (2004b) have used the l_1 norm instead of the l_2 norm implicitly used in least-squares estimation. In that case, the ML estimator is optimal when $\mathbf{n}$ is Laplacian white noise. The l_1 norm is more robust than the l_2 norm, therefore that norm improves the performance of SR under the presence of outliers.

To find the solution to (2.9) we must solve an ill-posed problem, as certainly that equation does not have a unique solution. An analysis of the ill-conditioning of (2.9) can be found in Nguyen et al. (2001a). We could not have enough LR frames and in that case the equation would be under-determined. Even when the problem is over-determined, the solution to (2.9) can have large perturbations for small amounts of noise in the measurements. Therefore, it is necessary to consider some form of regularization to find the most probable solution within the possible ones. This regularization must use some kind of a priori knowledge to compensate the missing information, so finally a MAP problem arises. To find now the HR image we will add a penalty factor in the

generalized minimization cost function, so we have

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}} \left[\|\mathbf{A}\mathbf{z} - \mathbf{w}\|_{l_p}^p + \lambda\Omega(\mathbf{z}) \right], \tag{2.10}$$

where λ, the regularization parameter, is a scalar for properly weighting the first term (similarity cost) against the second term (regularization cost) and Ω is the regularization cost function.

One of the most widely used regularization cost functions for SR is the Tikhonov cost function (Tikhonov, 1963)

$$\Omega(\mathbf{z}) = \|\Gamma\mathbf{z}\|_{l_2}^2, \tag{2.11}$$

where Γ is usually a high pass operator such as derivative, Laplacian, or even identity matrix. The idea behind this regularization term is to limit the total energy of the image. This will force spatial smoothness and remove the noise, but with the danger of also smoothing the sharp edges of the image. Other regularization terms have been proposed to correct this drawback, like bilateral Total Variation (bilateral TV) (Farsiu et al., 2004b).

2.4 Super-Resolution in its Stages

In this section we review the research that has split the SR problems in three main tasks: registration (have all samples in the same coordinates reference), interpolation (resampling in the HR grid) and restoration (removing noise and artifacts).

2.4.1 Super-Resolution and Registration

Registration is widely considered the most influential factor when considering sources of poor performance in super-resolution. Registration has to be extremely accurate for achieving real SR: if we multiply the resolution by a factor F, the registration precision must be in the order of $1/F$. A key factor is aliasing: we need it for SR to be possible, but its presence makes the images different among themselves, which makes registration much more difficult, as comparisons of images are affected.

Methods that apply SR in the Fourier domain perform registration and resolution improvement simultaneously, as described in section 2.2, but with the drawback of be-

ing that possible when there only shifts present among the images. In most cases, SR practitioners use optical flow (Lucas and Kanade, 1981), constrained to a certain type of movement, which could be a shift, solid rigid, affine, projective, etc. For robustness when the movement is not limited to just a few pixels, multi-resolution approaches like the one explained in Bergen et al. (1992) are used. In Pham et al. (2005a) it was proved, using the Cramer-Rao lower bound, that registration between two images using optical flow iteratively are optimal whenever we have perfect knowledge of the gradient. Unfortunately, when optical flow methods are employed, the aliasing present in the LR images degrades greatly gradient calculations.

To address these shortcomings, some authors have applied multi-frame registration methods to SR. The first example that we can find in the literature was exposed in Hardie et al. (1997). In Robinson et al. (2007) the authors explore multi-frame registration under aliasing using the method of variable projections. Another method that focuses on multi-frame registration under rotations and shifts is proposed in He et al. (2007). In general, all these methods are computationally very expensive and their performance depends on the size of the final HR image that is obtained, which is not desirable. Another approach can be found in Farsiu et al. (2005), where consistency among the registration parameters of the LR images is enforced. In chapter 3 of this thesis we will propose an improved alternative to this later method that is another contribution of our thesis.

Although the optical flow approach is the main registration technique used in SR due to its accuracy, another methods have also been used by some authors. Shekarforoush and Chellappa (1999) used the polyphase decomposition of the cross power spectrum to achieve registration. In Vandewalle et al. (2006), the authors proposed a frequency domain registration method that erased the aliased part of the spectrum of the LR images before proceeding to registration. Feature-based methods (bundle adjustment) have also been used recently (Pickup et al., 2007) to perform registration in a Bayesian framework where that task is jointly made with the estimation of the HR image.

2.4.2 Super-Resolution and Interpolation

As has been said before, super-resolution can be achieved solving directly the problem posed in (1.1) or separating it into a few tasks. Interpolation comes after registration and before image restoration, as can be seen in Fig. 1.2. In this stage, already knowing

the relative position among themselves of the LR frames, the data from the different frames is fusioned into the HR grid. This approach is only feasible when the warping and blurring operators can be interchanged, making the order in which the problem is solved (first interpolation, then deblurring) possible. In chapter 4 the conditions under we can do this are exposed.

The interpolation problem is essentially a resampling problem. All samples from the LR images are positioned in the same reference frame and to obtain the HR image we have to find the values in a grid within that common reference. Ur and Gross (1992) were the first to propose an interpolation approach to SR, but limited to shifted pictures. They based their method on the generalized multichannel sampling theorem of Papoulis (1977), in the refined form proposed later by Brown Jr (1981). Other authors (Shekarforoush and Chellappa, 1999) generalized Papoulis' theorem to non-periodic samples, but supposing the function to be estimated to be band-limited.

An approach using wavelets was proposed in Nguyen and Milanfar (2000). The problem of finding the pixels of the HR image was transformed to finding the wavelet coefficients that better approximated the LR samples. This was formulated in a least-squares framework that had as main drawback the computational complexity. An interpolation method that is closely related to the method we propose in chapter 4 is the one that can be found in Lertrattanapanich and Bose (2002). The method first transforms the LR samples to the continuous domain using Delaunay triangulation and then samples the built function to obtain the HR image. Another proposal (Pham et al., 2006) uses polynomial approximations to describe locally the HR image we want to find.

The method proposed in chapter 4 of the thesis can be classified as belonging to this type of algorithms. The main difference with the presented approaches lies in that it makes no supposition about the bandwidth of the imaging system and in that it is able to handle any type of stationary noise.

2.4.3 Super-Resolution and Deblurring

After the interpolation stage, we can proceed to deblur the image. There are two approaches here: we can suppose a known blur or suppose it unknown, having in that case what is called a blind super-resolution problem. In the first case we can use classical image restoration methods: a Wiener filter (Wiener, 1949), the Lucy-Richardson algorithm

(Richardson, 1972; Lucy, 1974), etc. We can also take advantage of the additional information present in the LR images and use it jointly with the interpolated image already calculated. This idea is followed in Farsiu et al. (2004b), where the error measurement between the deblurred and the interpolated images is weighted, pixel by pixel, by the number of LR pixels that contributed to one of the pixels of the interpolated image. In this way, we can combine a fast stage approach to SR with using information in the deblurring not usually available for classical restoration methods.

There are several works on blind SR. The first attempt proposed estimating the parameters of a Gaussian blurring kernel from the edges of the image (Chiang and Boult, 1997). Later, it was proposed to use generalized cross-validation and Gauss quadrature theory for identification of one parameter of a blurring kernel (Nguyen et al., 2001b). A more general framework has been proposed recently (Sroubek et al., 2007) that allows variant and non-symmetric blurring kernels. This approach uses the general theory of multichannel blind deconvolution and generalizes it to treat the SR problem. One of the main advantages of being able to handle variant kernels is that the algorithm is more resistant to registration errors, as they can be compensated with the different weights of the blurring kernel pixels in different locations.

In chapter 5 we present a comparison of approaches where deblurring and resolution improvement is performed jointly with approaches that perform interpolation and then apply a standard deblurring method. We will see there that not performing jointly deblurring with other SR tasks does not necessarily lead to worse performance than joint approaches.

2.5 Limits of Super-Resolution

A question that naturally arises when working in Super-Resolution is how to achieve maximum performance with minimum computational cost with a given imaging system and a given scene. That is, we want the best resolution and image quality without wasting resources. To reach this goal, we need to answer the questions

1. Which is the maximum real increment in resolution we can achieve?
2. How many LR images are needed to maximize the image resolution and quality? Is there a bound from which it makes no sense to add more LR images?

The first question is equivalent to asking which is the maximum SR factor (the number of times that the LR resolution is increased) that can be achieved without smoothing the image instead of increasing the resolution. In section 2.5.1 we review the current results from research in this problem. The second one can be answered in two scenarios: supposing LR images have no noise or supposing normal conditions where noise is unavoidable. In the first case, we would just need enough samples from the LR images so that the Nyquist criterion is fulfilled. In this case, we are interested on the amount of information that we need for maximizing the resolution. That is, if we know that an imaging system bandwidth is twice its sampling rate, we could increase its resolution by a factor $F = 2$. In that case, with $F^2 = 4$ LR images shifted among themselves we would have enough information for reconstructing the HR image. This is the case even when the sampling is irregular, if the image is band-limited (this was the main insight of the seminal paper by Tsai and Huang (1984)). In the second case, when noise is present, we need to know the relationship between the SNR and the number of LR images that we introduce in the algorithm, as we will be mainly interested on the improvement of quality we can achieve in the reconstructed image. We will look at this in section 2.5.2.

Not too many authors have studied these problems. The first paper to address them was Baker and Kanade (2002), which looked at maximum SR factors through the condition number of the system matrix of (2.7). Lin and Shum (2004) employed a similar algebraic approach, that developed explicit limits under translations and supposing a square PSF. In Pham et al. (2005b), limits for diffraction limited imaging systems are derived.

In this section we will try to answer the posed questions looking at the way the Point Spread Function and the Signal-to-Noise ratio limit Super-Resolution. We will use the previously cited works as our main sources of information. The general ideas extracted from the studied documentation about SR limits have been taken into account when selecting SR factors and number of input LR images in the experiments exposed in further chapters of the thesis.

2.5.1 Influence of Point Spread Function

Due to the diffraction limit of finite-size lenses, a point source in the scene corresponds to an intensity blob in the imaging plane, often referred to as the Airy disk or the optical

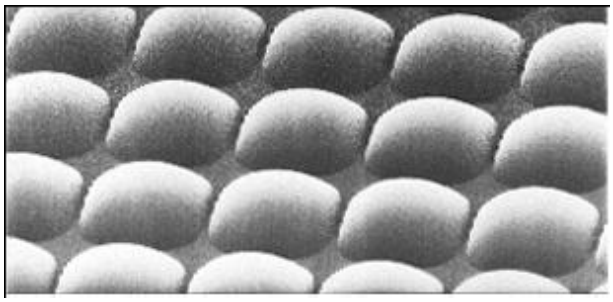


Figure 2.1: CCD sensor with microlenses.

Point Spread Function. Due to charge transport and sampling, the digitized image is further degraded. The combination of all degradations can be modeled by a system blur, which is a convolution of all PSFs along the imaging path. Two types of blur are present in all digital imaging systems: the optical blur which specifies the fine-resolving capability of the optics and the sensor integration blur caused by the accumulation of incoming light over a finite-size sensor element. Therefore, we define the system blur h as

$$h = h_{SEN} * h_{OPT}, \tag{2.12}$$

where h_{SEN} is the blurring due the integration on the sensor and h_{OPT} is the blurring due to the lenses.

The optical PSF for in-focus objects can be reduced to the diffraction limit by good lenses. The sensor PSF, however, cannot be reduced arbitrarily because the sensor has to be big enough to capture sufficient light for achieving a good SNR. In fact, the trend is to occupy all possible space to capture as much light as possible. In modern camera systems, the pixels tend to be quite small, so to decrease noise the integration area tends to occupy all available space, which means that the so-called fill-factor tends to be 100%. This decreases noise at the price of increasing the sensor blur. As an example, Fig. 2.1 shows the chip surface of a Super HAD CCD sensor from Sony. This is a state of the art CCD, employed in many professional cameras. The microlenses on it increase the fill factor to almost a 100% fill factor, but at the price of having square microlenses that create some defocus.

We will now study the expression of h_{OPT} and h_{SEN} in the frequency domain. In the case that the lens system is diffraction-limited and we have a circular aperture, the

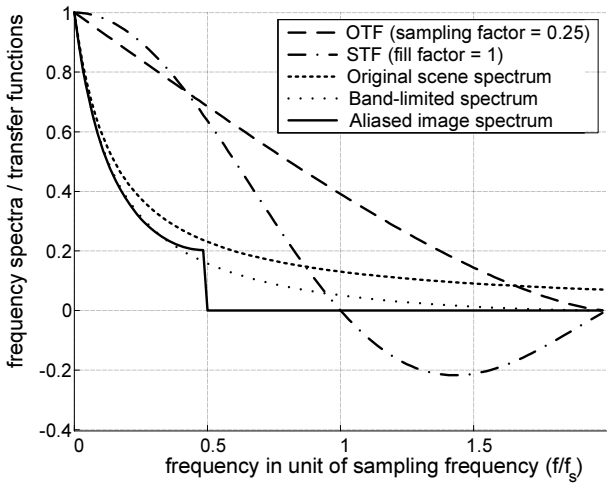


Figure 2.2: Transfer functions and image spectra.

Optical Transfer Function (OTF), which is the Fourier transform of h_{OPT} is

$$\mathrm{OTF}(f) = \frac{2}{\pi}\left(\arccos\left(\frac{f}{f_c}\right) - \frac{f}{f_c}\sqrt{1-\left(\frac{f}{f_c}\right)^2}\right) \quad \text{for} \quad f < f_c, \quad 0 \quad \text{otherwise}, \tag{2.13}$$

where $f = \sqrt{f_x^2 + f_y^2}$ is the radial frequency in a two-dimensional Fourier space $\{f_x, f_y\}$, and $f_c = 2 \cdot \mathrm{NA}/\lambda$ is the cutoff frequency of the lens (NA denotes the numerical aperture of the lens and λ is the wavelength of incoming light).

Supposing that the sensor response h_{SEN} can be approximated by a square box function, we will obtain its Fourier transform, the Sensor Transfer Function (STF)

$$\mathrm{STF}(f_x, f_y) = \mathrm{sinc}\left(\frac{\omega}{2}\frac{f_x}{f_s}\right) \cdot \mathrm{sinc}\left(\frac{\omega}{2}\frac{f_y}{f_s}\right), \tag{2.14}$$

where f_s is the sampling frequency and ω^2 is the fill factor of a square pixel whose width is ω-times the pixel pitch ($\omega \leq 1$).

The product of (2.13) and (2.14) will determine the analog image that is finally sampled by the sensor. In Fig. 2.2 we can see typical OTF and STF transfer functions plotted. Usually, the sampling frequency in modern devices is smaller than twice the cutoff frequency of the optics ($f_s < 2f_c$), which makes super-resolution feasible. This is the case displayed in Fig. 2.2, where the image finally sampled presents aliasing. The unfolding of the aliased part of the spectrum would present details from the original image that the system transfer function does not erase. In the case displayed in the figure, we could duplicate the resolution in the HR images taken with that imaging system.

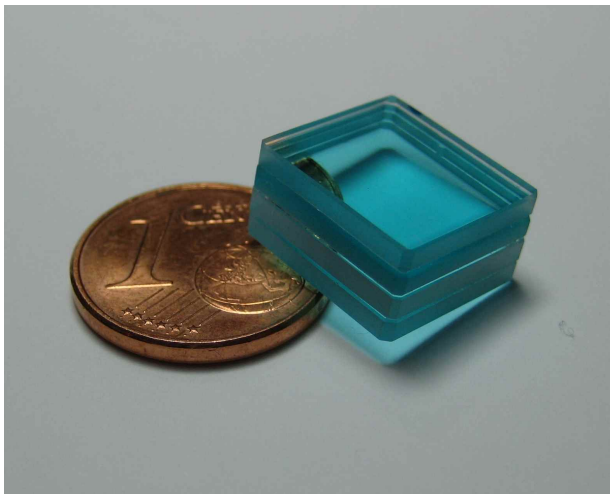


Figure 2.3: Anti-aliasing filter from a digital video camera.

In practical systems, the condition $f_s < 2f_c$ usually happens, which makes SR possible in theory. However, the presence of optical anti-alias filters in common photographic cameras makes SR difficult in some cases. These anti-aliasing filters try to reduce the Moiré patterns that appear when demosaicing undersampled images (Adams et al., 1998). The typical implementation of these filters in digital cameras is two layers of birefringent material such as lithium niobate, which spreads each optical point into a cluster of four points (Davies and Fennessy, 2001). Figure 2.3 shows an example of an optical anti-alias filter from a digital video camera.

The different image processing tasks that are usually performed in most imaging systems also present a problem if we cannot access the data in a raw format. Fortunately, the presence of this kind of formats which store the data unprocessed is now common in high-end systems. These problems for SR in real systems show that a deep knowledge of the characteristics of the imaging system we are dealing with is fundamental to apply SR to the data obtained with it.

2.5.2 Influence of Signal-to-Noise Ratio

Noise is present in all digital imaging systems due to a number of sources such as photon shot noise, readout noise, dark current noise, and quantization noise. While some noise sources can be effectively suppressed such as dark current noise by cooling and quantization noise by using more bits, others cannot. Photon shot noise, for example, is unavoidable due to the particle nature of light. Readout noise increases with a higher readout rate, which is desirable in high-speed cameras. The combined effect of these

noise sources is often modeled by Gaussian additive noise. If we denote the noise variance of an LR image as σ_I^2, the noise variance of the HR image obtained from a sequence of N LR images all with the same statistics is (Pham et al., 2005b)

$$\sigma_n^2 = \frac{F^2}{N}\sigma_I^2, \tag{2.15}$$

where F is the SR multiplying factor.

From (2.15) one could deduce that we could achieve any desired SNR just increasing the number of LR images, provided that we can take any number of LR images we wish. Unfortunately, the presence of noise in the LR images implies an error in registration if we do not have any a priori knowledge of the relative movement among the LR images. It was demonstrated in Pham et al. (2005a) that there is a fundamental lower bound in the registration error due to the inevitable noise present in the sensed images. Therefore, the noise in the reconstructed image is (Pham et al., 2005b)

$$\sigma_n^2 = \frac{F^2}{N}\sigma_I^2 + \overline{\|\nabla I\|^2}\,\sigma_{reg}^2, \tag{2.16}$$

where σ_{reg}^2 is the variance of the registration error and $\overline{\|\nabla I\|^2}$ is the average gradient energy. This equation shows that although we can almost suppress the first term at the right hand of the equation just by adding more LR images, the second term does not depend on N and imposes a limit on the SNR we can achieve. The equation can be used as a clue to how many LR images we should use as input to the SR algorithm we use, in function of the noise present in the data.

2.6 Comparison of Methods

As a summary, we compare in this section the main features of the most significative SR methods. We also compare them with the features of the method we propose in chapter 4, which appears as "Optimal projection" in table 2.1.

We can see the comparison in table 2.1. The approaches that appear in the table are frequency domain (Kim et al., 1990), POCS (Patti et al., 1997), IBP (Irani and Peleg, 1991), ML (in Elad and Hel-Or (2001), for a special case), MAP (Schultz and Stevenson, 1996), joint registration and interpolation under a MAP framework (He et al., 2007), Delaunay

triangulation (Lertrattanapanich and Bose, 2002), the interpolation method by Pham et al. (2006), and the optimal projections method (Sánchez-Beato and Pajares, 2008). The citations in the previous sentence are meant to be the most significant publications, although not the unique, that have followed the corresponding approach.

We will also briefly comment each row in the table:

- The motion model is limited to shifts and rotation for interpolating methods, as those are the only cases for which we can interpolate before deblurring the image. This fact is explained in detail in chapter 4.
- The noise model is usually assumed Gaussian, although some methods can handle Laplacian noise changing the norm in the function to be minimized. The Delaunay triangulation method offers no protection against noise. Our method is efficient under any stationary noise.
- A priori knowledge is easily incorporated for POCS and MAP approaches. There is usually a price to pay that is the need to tune a certain number of parameters that are dependent on the sequence that is being treated.
- Registration is performed jointly with other SR tasks only under a very high computational cost.
- Deblurring is performed jointly with SR in POCS, IBP, ML, and MAP approaches. The advantage here is that this allows more general motion models, but also forces these methods to be iterative.
- The computational cost usually increases when more tasks are performed simultaneously.
- Uniqueness of solution can be an issue for certain iterative approaches.
- The discretization is not usually handled properly, as the HR image is supposed to be band-limited in many cases, but in our method.
- If the optimization method is iterative we must be careful with its convergence and uniqueness of solution.
- If the method is parallelizable we can improve its speed just adding more processors and memory to our system.

Table 2.1: Comparison of SR methods

	Frequency domain	POCS	IBP	ML	MAP	Registration + MAP	Delaunay	Pham	Optimal projection
Motion model	Shifts	Any	Any	Any	Any	Shifts and rotations	Shifts and rotations	Shifts and rotations	Shifts and rotations
Noise model	Gaussian	Gaussian	Gaussian	Gaussian or Laplacian	Gaussian or Laplacian	Gaussian or Laplacian	None	Gaussian	Any stationary noise
Incorporates a priori knowledge	No	Yes	No	No	Yes	Yes	No	Yes	No
Registration	Yes	No	No	No	No	Yes	No	No	No
Deblurring	No	Yes	Yes	Yes	Yes	Yes	No	No	No
Computational cost	High	Medium	Low	High	High	Very high	Low	High	Low
Uniqueness of solution	Unique	Non-unique	Non-unique	Non-unique	Unique	Unique	Unique	Unique	Unique
Discretization	Not properly handled	Not properly handled	Not properly handled	Not properly handled	Not properly handled	Not properly handled	Not properly handled	Not properly handled	Properly handled, orthogonal projection
Optimization method	Iterative	Iterative	Iterative	Iterative	Iterative	Iterative	Non-iterative	Iterative	Non-iterative
Parallelizable	No	No	No	No	No	No	Yes	Yes	Yes

Looking at the table, we can emphasize the main advantages of the method proposed in this thesis: it can handle any stationary noise, it is non-iterative, and it does not suppose that the bandwidth of the imaging system is the same as that of the reconstructed image (which means that it is the only method that makes a proper discretization of the problem).

Chapter 3

Multi-Frame Registration for Super-Resolution

3.1 Introduction

A lot of research has been devoted to SR, especially to the interpolation and restoration steps. Specific registration techniques for SR have also been studied, although not with the intensity that the reconstruction stage has attracted. Although it seems that registration in SR is not different to registration applied to other problems, there are three interesting facts that must be taken into account:

1. LR images are of the same scene and overlap largely
2. Usually there are many LR images
3. Images must be aliased for super-resolution to be feasible

How could we take advantage of these facts? In this chapter we describe a new registration algorithm that makes a registration not just between a reference image and another image, repeating this for all LR images in the sequence, but among all images at the same time, taking advantage of all the information that is available. Govindu (2004) and previously many others (Unnikrishnan and Kelly, 2002; Davis, 1998; Shum and Szeliski, 1998; Kang et al., 2000) have used globally consistent registration methods applied to mosaics. Farsiu et al. (2005) have applied a similar algorithm to the one used here for super-resolution (we will analyse the differences in section 3.3). Vandewalle et al. (2006) use the fact that images must be aliased for SR to build a registration method for this problem. Finally, Capel and Zisserman (1998) have used bundle adjustment for

registration to a joint problem of mosaicing and super-resolution.

There are also methods that jointly super-resolve and register the LR images. This approach was examined by Hardie et al. (1997), although their implementation was limited to a translational motion model. Tipping and Bishop (2003) use more general motion models but at a very high computational cost, while He et al. (2007) allow translations and rotations in their approach. Usually, the simultaneous registration and super-resolution comes at a considerable computational cost, which limits this approach to small images. It is also not clear in which way obtaining at the same time the HR image aids the multi-frame registration: we have the same input data for both registration and super-resolution, so the additional data that the SR algorithms give to registration must reside in the a priori knowledge that is employed in MAP and other approaches. These priors are not designed to improve the registration of the LR images, but to regularize the problem, so maybe it is preferable to devise registration algorithms that take advantage of all the LR data for doing its task and add specific priors for registration instead of jointly registering and super-resolving. This is the approach we follow in this chapter.

Another interesting issue is that of the optimality of the registration, which has been treated by Robinson and Milanfar (2004) and by Pham et al. (2005a). Both papers use the Cramer-Rao lower bound for deriving theoretical limits for registration estimators. Pham shows that although most registration estimators are biased, the bias can be practically reduced to zero using an iterative gradient-based estimator, approaching the theoretical limit for an optimal estimator. This is remarkable, as most registration algorithms based on block matching use that approach. These results are obtained when two images are registered between themselves, so they are not directly applicable to the joint registration of multiple images, but they point out that probably the use of gradient-based algorithms for registration of multiple frames is preferable to feature-based algorithms like bundle adjustment, like those exposed by Hartley and Zisserman (2003). This is true when it is not a must to use feature-based algorithms, which happens when the images to register come from different imaging systems (usual in medical imaging where PET and CT images are merged, for instance), there are differences in luminosity, the overlapping is only partial (like in mosaics), etc. Nevertheless, SR is normally applied to video sequences obtained with a single imaging system within a

short time interval, which normally means constant luminosity and a high degree of overlapping in the resulting frames.

In this chapter we propose a method for joint registration of multiple frames based on the optical flow equation (Horn and Schunck, 1981) and on geometric restrictions among the different images to be registered. We compare this approach with the standard optical flow algorithm (shown to be optimal in Pham et al. (2005a) for pairs of images) and with other joint registration algorithms. As all gradient-based registration estimators, the method is valid only for small movements due to the approximation that is made by taking only the first two terms of the Taylor expansion of the optical flow equation. This means that we must apply a different method for initial registration of the frames (the multi-resolution approach of Bergen et al. (1992), for instance) and then apply our method as the final step to achieve subpixel accuracy.

The rest of the chapter is organized as follows. Section 3.2 presents the basic optical flow method and different motion models for the registration, including the ones most commonly used in SR. In section 3.3 we will expose our method and its differences against competing approaches. Section 3.4 will be devoted to the results obtained applying this method to controlled and uncontrolled experiments. Section 3.5 describes a fast variation of our method that achieves similar performance to the original but using a fraction of the computational power required by it. In section 3.6 we apply our method to super-resolution, comparing the results with those obtained using a standard registration algorithm. Finally, we present the conclusions in section 3.7.

3.2 Registration Models

We will use as starting point the classical approach of the conservation of the optical flow to register the images, as described by Lucas and Kanade (1981). If we have two images of the same scene (we will call them I_1 and I_2), the pixels in one of them must be the pixels of the other displaced a quantity that depends on the pixel position $\mathbf{x}$, which is expressed as

$$I_1(\mathbf{x}) = I_2(\mathbf{x} + \mathbf{u}_{1,2}(\mathbf{x})), \tag{3.1}$$

where $\mathbf{u}_{1,2}(\mathbf{x})$ is a vector that transforms the coordinates vector $\mathbf{x}$ from image I_1 to image I_2. Thus, we have

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} + \mathbf{u}_{1,2}(\mathbf{x}^{(1)}), \tag{3.2}$$

where the superscript specifies the image coordinate system for the given vector (in (3.1) all the coordinate vectors that appear have as reference image I_1, we use superscripts only when there are vectors defined in different coordinate systems in the same expression).

If we develop the right side of (3.1) and take the first expansion terms of the Taylor series, we have

$$I_1(\mathbf{x}) \approx I_2(\mathbf{x}) + \vec{\nabla} I(\mathbf{x}) \cdot \mathbf{u}_{1,2}(\mathbf{x}). \tag{3.3}$$

As we only take the first two terms, the gradient $\vec{\nabla} I(\mathbf{x})$ is supposed to be constant in a neighborhood of $\mathbf{x}$, so it is the same for the two images. For that reason, we do not use a subscript for the gradient, and in fact we use data from the two images to obtain a better approximation of the gradient value.

This is the general optical flow equation and is the basis for all area-based registration methods, like the one exposed in Lucas and Kanade (1981). In its more general form we can use it to calculate an individual displacement vector $\mathbf{u}_{1,2}$ for each pixel. However, there is not always enough information in the images to find the right $\mathbf{u}_{1,2}$ for all pixels. This is called the aperture problem in the literature (Pajares and de la Cruz, 2008). This problem can be solved in two ways. One option is to assume that the movement is local, and in that case the optical flow is calculated solving a regularized minimization problem by means of a set of partial differential equations (non-parametric or elastic registration, see Verdú-Monedero et al. (2008)). The other option is to suppose that the movement is global and that it can be precisely modeled by a transformation for which we must calculate certain parameters. In SR normally the second option is chosen, as it allows a more precise registration if the model really fits the data, so this is the choice we have taken for this research. The reason for the better accuracy of the parametric registration resides on having much more less degrees of freedom than in non-parametric registration. The transformation might be a translation, a rotation, a projective transformation, etc. All of them can be described with just a few parameters once we fix the type of transformation. The way we calculate the parameters for some of

the transformations is shown below. We also introduce the hierarchical model estimation (Bergen et al., 1992) that will be used to estimate the registration parameters before the final and more precise step.

3.2.1 Translational Motion

In this case,

$$\mathbf{x}^{(2)} = \begin{pmatrix} p_{1,2}^{(1)} \\ p_{1,2}^{(2)} \end{pmatrix} + \mathbf{x}^{(1)}. \tag{3.4}$$

Therefore, $\mathbf{u}_{1,2} = \mathbf{p} = \begin{pmatrix} p_{1,2}^{(1)} & p_{1,2}^{(2)} \end{pmatrix}^T$, being $\mathbf{p}$ the parameters vector, and to find these parameters we have to minimize

$$E(\mathbf{p}_{1,2}) = \sum_{\mathbf{x}} \left(\Delta I_{1,2}(\mathbf{x}) + \vec{\nabla} I(\mathbf{x}) \cdot \mathbf{p}_{1,2} \right)^2,$$

where $\Delta I_{1,2}(\mathbf{x})$ is $I_2(\mathbf{x}) - I_1(\mathbf{x})$. Here we have moved to the right all the terms in (3.1), squared the result, and then we have added that expression for all the positions of the pixels where the images coincide. The final equation can be solved by least squares as the parameters vector can be taken out of the summatory as it does not depend on $\mathbf{x}$. This transformation can model just shifts between the images, but is frequently used in SR due to its simplicity.

3.2.2 Affine Motion

With an affine transformation we can model rotations, translations, scaling in any axis, and reflections. The transformation is

$$\mathbf{x}^{(2)} = \begin{pmatrix} p_{1,2}^{(1)} \\ p_{1,2}^{(4)} \end{pmatrix} + \begin{pmatrix} 1 + p_{1,2}^{(2)} & p_{1,2}^{(3)} \\ p_{1,2}^{(5)} & 1 + p_{1,2}^{(6)} \end{pmatrix} \mathbf{x}^{(1)}. \tag{3.5}$$

In this expression we use $1 + p_{m,n}^{(2)}$ and $1 + p_{m,n}^{(6)}$ instead of just the parameter in the diagonal of the matrix to make the final equation easier to minimize. As the motion will be possibly almost a rotation, this also makes the parameters in the matrix have nearer values, which translates to a better numerical stability for optimization methods without the need of scaling the variables (Press et al., 1992). The displacement, obtained

subtracting $\mathbf{x}^{(1)}$ to the right part (3.5), is

$$\mathbf{u}_{1,2}(\mathbf{x}) = \begin{pmatrix} p_{1,2}^{(1)} \\ p_{1,2}^{(4)} \end{pmatrix} + \begin{pmatrix} p_{1,2}^{(2)} & p_{1,2}^{(3)} \\ p_{1,2}^{(5)} & p_{1,2}^{(6)} \end{pmatrix} \mathbf{x}.$$

If we define the matrix

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & x_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x_1 & x_2 \end{pmatrix},$$

we can then express $\mathbf{u}_{1,2}$ as

$$\mathbf{u}_{1,2}(\mathbf{x}) = \mathbf{X}(\mathbf{x})\mathbf{p}, \tag{3.6}$$

where $\mathbf{p}$ is the vector of parameters $\mathbf{p} = \begin{pmatrix} p_{1,2}^{(1)} & p_{1,2}^{(2)} & p_{1,2}^{(3)} & p_{1,2}^{(4)} & p_{1,2}^{(5)} & p_{1,2}^{(6)} \end{pmatrix}^T$. Finally, the expression that we have to minimize is

$$E(\mathbf{p}_{1,2}) = \sum_{\mathbf{x}} \left(\Delta I_{1,2}(\mathbf{x}) + \vec{\nabla} I(\mathbf{x}) \cdot \mathbf{X}(\mathbf{x})\mathbf{p}_{1,2} \right)^2,$$

where again we can extract $\mathbf{p}_{1,2}$ from the summatory and easily solve the equation by least squares.

3.2.3 Projective Transformation

This motion model has eight parameters and can model translations, rotations, scaling in any axis, reflections, panning, and tilting. This is the general transformation that a plane in the 3D space suffers when viewed from a pin-hole camera. For this motion

$$\mathbf{x}^{(2)} = \frac{\begin{pmatrix} p_{1,2}^{(1)} \\ p_{1,2}^{(4)} \end{pmatrix} + \begin{pmatrix} 1 + p_{1,2}^{(2)} & p_{1,2}^{(3)} \\ p_{1,2}^{(5)} & 1 + p_{1,2}^{(6)} \end{pmatrix} \mathbf{x}^{(1)}}{\begin{pmatrix} p_{1,2}^{(7)} & p_{1,2}^{(8)} \end{pmatrix} \mathbf{x}^{(1)} + 1}. \tag{3.7}$$

In this case there is not an easy expression for $\mathbf{u}_{1,2}$, it is just $\mathbf{u}_{1,2} = \mathbf{x}^{(2)} - \mathbf{x}^{(1)}$. The expression that we would want to minimize is

$$E(\mathbf{p}_{1,2}) = \sum_{\mathbf{x}^{(1)}} \left(\Delta I_{1,2}(\mathbf{x}^{(1)}) + \vec{\nabla} I(\mathbf{x}^{(1)}) \cdot (\mathbf{x}^{(2)} - \mathbf{x}^{(1)}) \right)^2, \tag{3.8}$$

where we cannot take out from the summatory the parameters vector. Thus, we cannot solve this equation by least squares and we must resort to an iterative numerical method. We have the added problem of (3.8) being not convex. This kind of nonlinear least squares problems can be solved using, for instance, a Levenberg-Marquardt iteration (Marquardt, 1963).

3.2.4 Hierarchical Motion Estimation

Equation (3.1) is valid only in a small neighborhood of a given point $\mathbf{x}$, as we are using only the first terms of a Taylor series. The images we are trying to register can be much further than the two or three pixels that the optical flow method allows us to register. Thus, we need an initial estimate of the parameters that approaches closely, up to a couple of pixels, the final ones. The usual approach is to build a pyramid of images, where we produce reduced versions of the original images that we register between themselves and then use the found registration parameters as initial estimates of the lower level parameters. Before proceeding to downsampling, the images are first low pass filtered. This is the method we follow in this chapter, as explained by Burt and Adelson (1983). Bergen et al. (1992) applied this method to different motion models in the context of registration.

3.3 Method Description

As mentioned before, we employ area-based registration in the proposed method. To apply this approach to a set of N images, we will register all images among themselves, that is, we will find the displacement vector for all possible combinations of images. We will define a transformation of coordinates $T_{k,m}$ to go from image k to image m. As a convention, k will always be less than m and we suppose that $T_{m,k} = T_{k,m}^{-1}$ is well-defined and unique. Therefore, we can register image m in the coordinates system of image k and vice versa. In Fig. 3.1a we can see the transformations that will be calculated by our algorithm in the case of four images, represented as a directed graph.

We also need to establish a link between the displacements to make them interdependent. To do so, and to fulfill the requirement of a globally consistent registration, we will require the transformations between the coordinate systems of the different LR

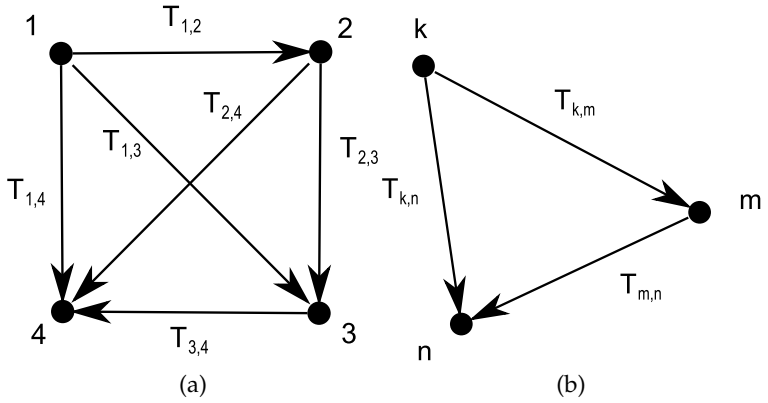


Figure 3.1: Transformations and restrictions for multi-frame registration. (a) Transformations calculated for 4 images and (b) restrictions for 3 images. The nodes represent the images and the edges represent the coordinate transformation between the images.

images to be consistent. The consistency is achieved by taking sets of 3 images and enforcing that the transformation of one of the coordinates of one image into other image must be the same independently of the path followed. If we take a triangle like the one in Fig. 3.1b, we will use as constraint $T_{k,n} = T_{m,n} \circ T_{k,m}$, where $\circ$ means composition of functions. The previously described restriction that $i < j$ for transformation $T_{i,j}$ implies that in each triangle the transformations in Fig. 3.1b follow the directions specified in the graph when the nodes fulfill the condition $k < m < n$. Using image k as origin, we will take all possible combinations of pairs of the images with index bigger than k to build all the needed constraints, and we will do this for values of k between 1 and $N-2$. Thus, we will force global consistency of registration, as there are common edges among the constraints. This implies that there are relationships among all images in the set and, therefore, there are global restrictions in the transformation functions.

Once we have these definitions we can build the cost function to be minimized so that registration can be achieved, which is

$$C = \sum_{m=1}^{N-1} \sum_{n=m+1}^{N} \sum_{\mathbf{x}} (\Delta I_{m,n}(\mathbf{x}) + \vec{\nabla} I(\mathbf{x}) \cdot (T_{m,n}(\mathbf{x}) - \mathbf{x}))^2 + \\ + \lambda \sum_{k=1}^{N-2} \sum_{m=k+1}^{N-1} \sum_{n=m+1}^{N} \sum_{\mathbf{x}} \| T_{m,n}(T_{k,m}(\mathbf{x}))) - T_{k,n}(\mathbf{x}) \|^2. \quad (3.9)$$

This equation is composed of the optical flow conservation equations derived from 3.1 (first term) and the constraints for consistency in the registration (second term). The displacement has been substituted by $\mathbf{u}_{m,n}(\mathbf{x}) = T_{m,n}(\mathbf{x}) - \mathbf{x}$. The constant λ weights the

importance we want to give to the constraint term and also has a conversion function between the second term, which is measured in pixels, to light intensity. The $\mathbf{x}$ are the coordinates of the pixels that are translated from one image to another. This equation is symmetrical as we pass the same number of times over all the edges in the graph for its second term. This can be easily seen: over the edge between nodes i and j ($i < j$) we pass $i - 1$ times until the index k of the first summatory equals i. For that value of the index we will pass $N - (i + 1)$ times over that edge, and for bigger values of k we will not pass again over it. So finally, we will pass $N - 2$ times over all edges in the graph.

This equation is very similar to the ones found in Farsiu et al. (2005) and could be considered within the framework proposed there, with the following differences:

1. We make the constraint term dependent on $\mathbf{x}$, so we take into account the magnitude of each pixel registration error when we minimize the function. This is not important when the images are just shifted. In that case the error between 2 frames is the same for all pixels, but this does not happen for more complex transformations (rotations, affine, etc.). This has also the following advantage: when the images only overlap partially (we consider our method a refinement step after more global registration methods have been applied, so we already have an idea of which parts of the images overlap) the constraint has more or less weight depending on the overlapping area. That is an improvement that can be useful even when the images are only shifted.
2. We do not estimate both $T_{m,n}$ and $T_{n,m}$. We consider $T_{m,n} = T_{n,m}^{-1}$ a hard constraint as we use data from both images to estimate the gradient in (3.1), so the equation stated when we want to know the transformation from n to m is equivalent to the one used to find the transformation from m to n. In this way we have less parameters to estimate.
3. We use less restrictions in the constraint term of (3.9), employing just the minimum number of sets of three images to consider all geometric constraints while achieving symmetry.
4. There is another reason for taking into account each pixel registration error: if we have, for example, images that are just shifted and we register them with a more general motion model (affine, for instance) without being weighted by the pixel errors, all the parameters are considered equally important, when in fact the

rotation parameters have far more influence in the registration error. This means that we look at the differences in value between the parameters, not to the influence they have in the overall error. Thus, the results obtained will be worse than if we had used just the shift motion model when we use the method described in Farsiu et al. (2005). Nevertheless, our method produces similar results in this case independently of using a shift model or an affine model.

We can particularize (3.9) for any particular registration model. Usually, the transformation will have some parameters that will be found minimizing the cost function. For instance, for the case of an affine transformation, we have, substituting the coordinates in (3.5) by the nomenclature used above,

$$T_{m,n} = \begin{pmatrix} p_{m,n}^{(1)} \\ p_{m,n}^{(4)} \end{pmatrix} + \begin{pmatrix} 1 + p_{m,n}^{(2)} & p_{m,n}^{(3)} \\ p_{m,n}^{(5)} & 1 + p_{m,n}^{(6)} \end{pmatrix} \mathbf{x}, \tag{3.10}$$

where the $p_{m,n}^{(i)}$ are the parameters that we need to calculate.

To solve the minimization problem of finding the parameters of a translation or an affine transformation among the input images for the cost function (3.9), we will use the quasi-Newton algorithm that the Matlab (The MathWorks, 2008) function *fminunc* implements. As the cost function is analytical for these motion models we also calculate its gradient, which greatly speeds up the search of the parameters.

To end this section we present, as a summary, the steps that are taken in the algorithm:

1. We select the sequence of images to be registered.
2. We choose the type of transformation that is applied between a pair of images, which determines the shape of the $T_{m,n}$ functions.
3. We minimize equation (3.9), obtaining as output the parameters of all the transformations.

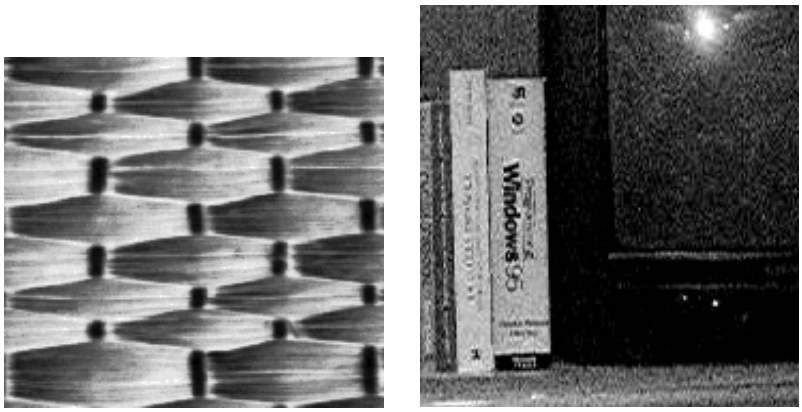


Figure 3.2: Original images used in the experiments: Texture and Television

3.4 Experiments

3.4.1 Methodology

The goal of the experiments carried out here is to verify the performance of the proposed registration approach as compared against other existing registration methods. We will test the method with the images displayed in Fig. 3.2 for the cases of shift and rotations, using different values for lambda and for different number of generated images. These images are first warped using random values for the registration parameters and then subsampled by a factor of two to obtain the different images to register. The subsampling is done to force a certain degree of aliasing in the generated images. The tests will be repeated for 30 different random realizations of the registration parameters for each point represented in the plots (that is, 30 realizations for each value of lambda and each number of images to register). We have tested the method only for shifts and rotations because of the difficulty of achieving a precise enough registration for SR under more complex motion models.

To measure the error we have to take into account that, as all frames are registered at the same time, there is no frame that can serve as reference. What we will do is to consider each frame as a reference and measure the registration error of the other frames with regard to that frame. The errors will be summed up and averaged to obtain the Mean Squared Error (MSE). If we call $T^I_{m,n}$ the ideal function that registers a position in image m to image n and M the total number of pixels of an image, we will have as MSE

$$\text{MSE} = \frac{1}{MN(N-1)} \sum_{m=1}^{N} \sum_{\substack{n=1 \\ m \neq n}}^{N} \sum_{\mathbf{x}} \| T_{m,n}(\mathbf{x}) - T^I_{m,n}(\mathbf{x}) \|^2.$$

Note that in this case we allow m to be bigger that n for $T_{m,n}$: in that case the transformation is calculated as $T_{n,m}^{-1}$.

3.4.2 Shift

In Fig. 3.3 we can see the results for the texture and the television image both depicted in Fig. 3.2. We have shifted the images in the x and y directions with random values following a uniform distribution between 0 and 1 pixel. The y-axis in Fig. 3.3 represents the ratio between the MSE for a certain number of images and a certain λ and the MSE for $\lambda = 0$. The x-axis is λ and different plots are displayed for different number of images. For two images we have the same case as $\lambda = 0$, so we have a line on $y = 1$ that if we surpass we are doing worse than the standard registration based on the conservation of the optical flow for two images. If we are below that line we are making a better registration, being the difference between one and the ratio we get the percentage of improvement.

We can see a significant improvement for both images, more than 60% for the texture image and more than 25% for the Television image. It can be seen that as the number of images increases the plots quickly converge, meaning that we do not need a lot of images to achieve optimal performance. Although for the Television image the results are worse than the reference for low λ, this is corrected when that parameter is increased. We also see that for high enough λ values the error remains almost constant and minimal, which means that the choice of λ is not critical: we can just stick to a high enough λ value (around 10, for instance) for the registration of any image sequence and obtain good results.

3.4.3 Rotation

We have tested the estimation of the rotation parameters rotating the images with random angles between -1° and 1°, following a uniform distribution. We can see the results in Fig. 3.4. We still get very good results for the texture image, with improvements of up to 25%. For the television image we obtain also a good performance for the algorithm, with improvements near to 25% for that image. In both cases the choice of λ is not critical for high enough λ, as happened in the tests for shifts.

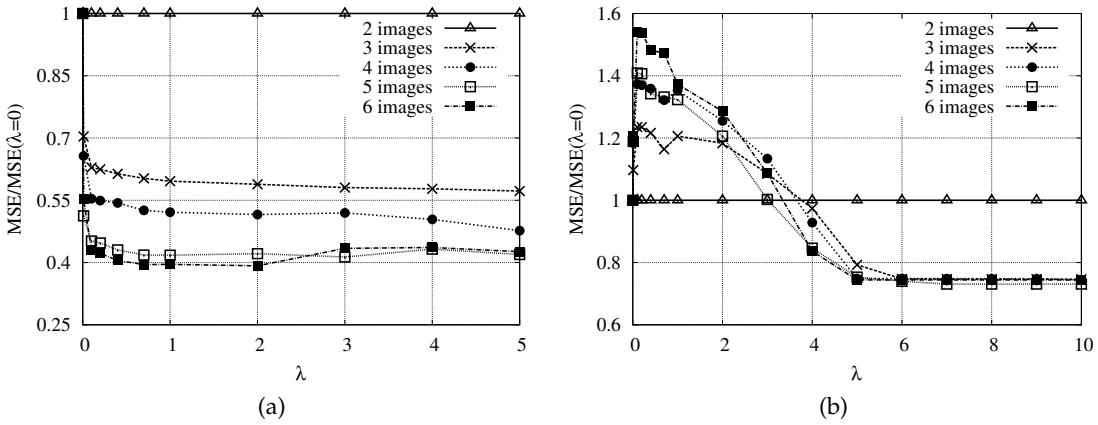


Figure 3.3: MSE ratios for shifts for (a) Texture and (b) Television images

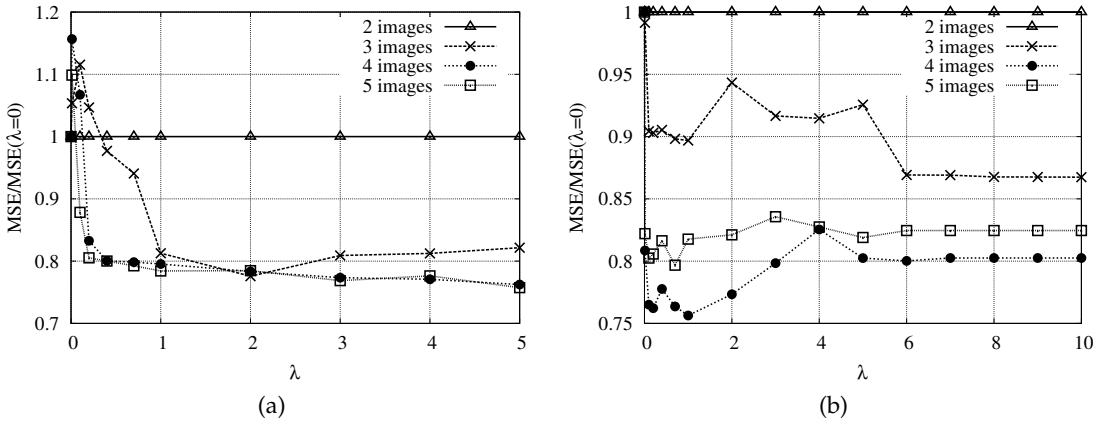


Figure 3.4: MSE ratios for rotations for (a) Texture and (b) Television images

3.4.4 Comparison with other methods

We already implicitly compared our approach with the classical optical flow method in the previous subsections. As our method is in some ways an extension to the one found in Farsiu et al. (2005), we will also compare both methods in this section. In both cases the implementation of the methods will try to find the parameters of an affine transformation, independently of the transformation we will apply to the input images. We have not made the comparison when the implementation just tries to find the parameters of a shift because in that case the results are almost the same for both methods. This happens because in that case all pixels make the same contribution to the total error and the weighting per pixel in (3.9), which is the main difference between the two methods, is not needed. Figure 3.5 displays the results of the MSE for 30 random realizations of shifts and rotations for the Texture image, for different values of λ. These results are for the case of four images to register. The λ for the Farsiu method is in fact the parameter that appears in Farsiu et al. (2005) divided by the number of pixels in the image. We do this to make it comparable to the λ in (3.9), where it multiplies the contribution to the error of all pixels instead of multiplying the parameters as in the method by Farsiu.

We can see that our method proves better for shift transformations. The reason for this is that our method does not consider all parameters as equally important, but we measure their impact on the registration of each pixel, as we explained in the previous section. Regarding rotations, we can see that our method has again better performance and that in this case the method by Farsiu cannot beat the standard optical flow algorithm (represented in the plots when $\lambda = 0$). This happens because there is not a linear relation between the rotation parameters and the registration error for this method when using a model more complex than just shifts.

3.5 Fast method

In this section we propose a fast method with similar accuracy to the one proposed in section 3.3. We will take advantage of the fast convergence when the number of images increases as we have seen in the previous section. The fast method will consist of the following steps:

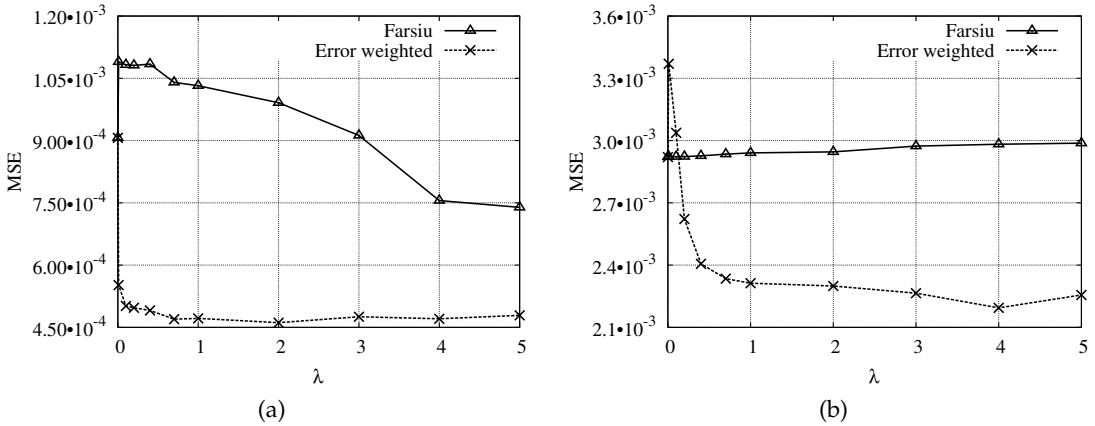


Figure 3.5: MSE: Comparison of methods for (a) shifts and (b) rotations

1. Apply the full algorithm for M images. This number does not need to be very high: 4 to 6 images is usually enough.
2. Add a new image to the set. The registration parameters between the previously registered images will be considered fixed, and we will only estimate the parameters of transformations between the new frame and the already registered ones. Therefore, we will have M new parameter vectors to estimate.
3. Increase M in one and return to the previous step while there are frames left ($M < N$).

The function that we must minimize when we add a new frame is

$$C = \sum_{m=1}^{M} \sum_{\mathbf{x}} (\Delta I_{m,M+1}(\mathbf{x}) + \vec{\nabla} I(\mathbf{x}) \cdot (T_{m,M+1}(\mathbf{x}) - \mathbf{x}))^2 + \\ + \lambda \sum_{k=1}^{M-1} \sum_{m=k+1}^{M} \sum_{\mathbf{x}} \| T_{m,M+1}(T_{k,m}(\mathbf{x}))) - T_{k,M+1}(\mathbf{x}) \|^2. \tag{3.11}$$

The equation uses the convention that we have already registered M images and we want to register the $M+1$ image. We will have M new parameter vectors to estimate and $\binom{M}{2}$ constraints in the second term of the equation.

In the end, the total number of registration vectors that we will calculate with this method is the same as in the full method, but having a smaller number of parameters to estimate when we apply the minimization algorithm. If, for instance, the minimization algorithm is $O(K^2)$ in time, being K the number of parameters to estimate, we will have a time complexity of $O(N^4)$ for the first method (N^2 vectors to estimate) and $O(N^3)$ for the second one (N vectors to estimate, approximately N times when N is

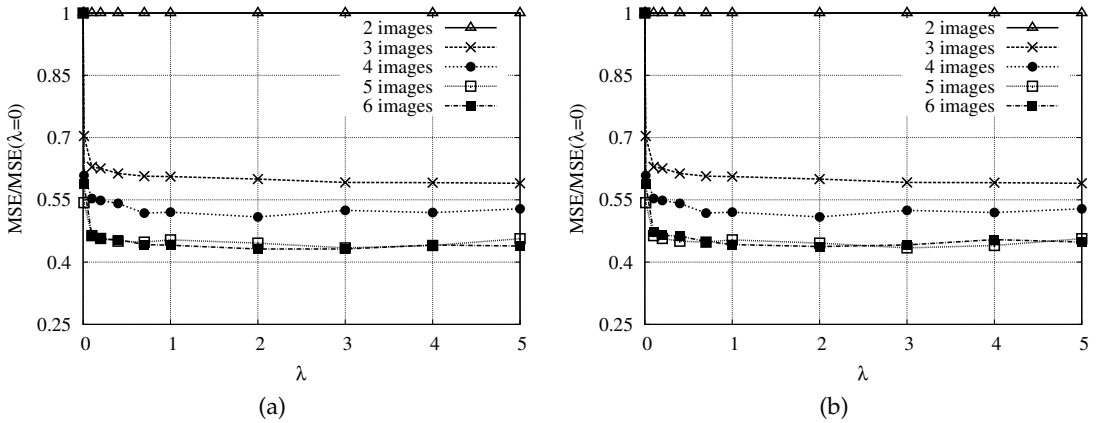


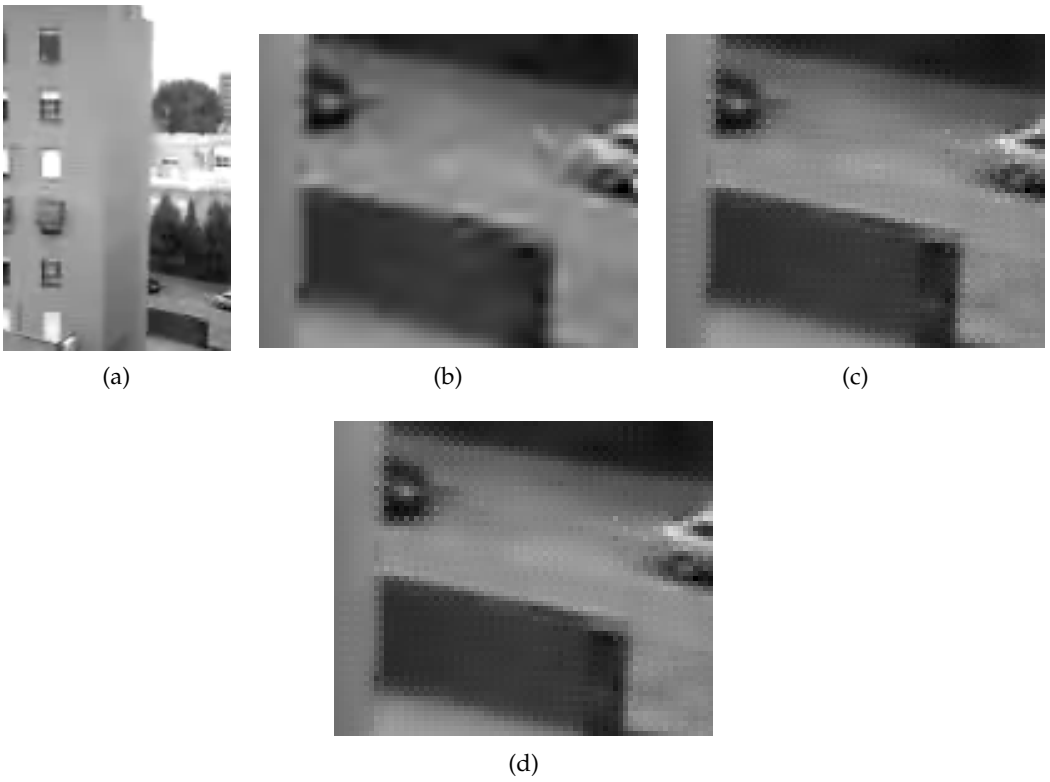
Figure 3.6: MSE ratios for (a) fast method and (b) fast method with M fixed. Initial $M = 4$.

big enough). The evaluation of function (3.11) and of its gradient is faster than the evaluation of (3.9), which also adds up. It can be seen in Fig. 3.6a that the effect in registration performance is low if we compare it with Fig. 3.3: the maximum difference is about 3% less improvement regarding to the case where $\lambda = 0$ when we jointly register 5 or 6 images.

We can make even faster the algorithm if we select a number of frames that we register among themselves and then we always use those frames to register the next frames, but without adding those frames to the reference set. In that case, the time complexity would be linear with N. The performance of this algorithm can be seen in Fig. 3.6b, which is very similar to the case when M is not fixed. As the complexity of this method is linear with the number of images, we conclude that this is the method of choice when we have many images to register.

3.6 Influence of registration in Super-Resolution

In this section we will make an experiment with real data, applying the registration method to super-resolution. We have used a sequence taken with the camera of a mobile phone. For registration, we have first applied the hierarchical pyramid approach by Bergen et al. (1992) and then our method as a final refinement step. After the registration, we have applied the SR method based on Delaunay triangulation exposed in Lertrattanapanich and Bose (2002), duplicating the original size of the low resolution frames. The reason for choosing this method is the high sensitivity to noise it has,



(a) (b) (c)

(d)

Figure 3.7: Influence of registration in super-resolution. Image (a) shows a frame of a sequence taken with a mobile phone camera, (b) is the cubic spline interpolation of a detail of the sequence, (c) is a super-resolved image obtained after registering with $\lambda = 0$, and (d) is the super-resolved image registered with $\lambda = 10$.

which makes it easier to show the effect of registration noise in the HR image (it is more difficult to see it in regularizing methods). In Fig. 3.7 we can see one of the low resolution frames taken with the mobile phone, the cubic spline interpolation of a detail that we will super-resolve, a super-resolution image obtained registering with $\lambda = 0$ and another SR image registered with $\lambda = 10$ (in both cases with 20 LR frames as input). The resolution is duplicated in the super-resolved images and in the interpolated fragment.

It can be observed that the noise is considerably higher in the image registered with $\lambda = 0$ (standard area based registration) than in the one registered with $\lambda = 10$, especially on the car to the right of the image and on the right border of the wall on the lower part of the image. To choose λ, different values were tried and the one for which the obtained image had a better visual quality was chosen. It was also seen that higher λ values did not improve the visual quality of the HR image, which strengthens the conclusion that it is easy to obtain near optimal results just taking a high value for λ from the beginning.

3.7 Conclusions

In this chapter we have presented a new method for the registration of a set of low resolution images that have large overlapping areas. This method outperforms similar approaches and the classical optical flow algorithm for shifts and rotations. An important characteristic of this method is that with 4 to 5 images we can obtain a very significant improvement. We have also presented a fast version of our method that is linear with the number of images and that hardly affects the quality of the registration. Finally, we applied our method to a real video sequence, obtaining visually better high resolution images when we apply super-resolution to the sequence.

After the experiments that we have made in this chapter, we conclude that the proposed registration method is adequate for super-resolution and we will apply it in the following chapters when registering sequences of images.

Chapter 4

Interpolation: Optimal Projections on Functional Spaces

4.1 Introduction

This chapter is based on Sánchez-Beato and Pajares (2008). In it we will concentrate on super-resolution as an interpolation problem. Similar approaches have been taken in Lertrattanapanich and Bose (2002), where a Delaunay triangulation method was adopted, or in Pham et al. (2006), where polynomial approximations are used to describe locally the HR image we want to find. The proposed method differs from these two on the backing we look for in sampling theory. Our objective is to show that the discrete approach commonly adopted for SR, like the one in Elad and Feuer (1997), implicitly assumes a data model that is not always valid and we provide a more general data model that works independently of the bandwidth of the imaging system.

We can cite as related with a sampling theory approach the first proposed methods for SR in Tsai and Huang (1984) and Kim et al. (1990), which were based on the shifting property of the Fourier transform, but they were valid only for translational motion models. There have also been methods based on the generalized sampling expansion theorem by Papoulis (1977) or its simplified form by Brown Jr (1981), like Ur and Gross (1992) or Shekarforoush and Chellappa (1999), but the former was valid only for translational motion and the latter assumes a band-limited HR image as most SR methods. Recently, a generalization of the Papoulis-Brown theorem to the multidimensional non bandlimited case has been done by Ahuja and Bose (2006). The method we

propose is valid for any motion model and allows a more general data model than the one implicitly assumed by the mentioned and other methods.

The main contribution of this chapter is a change of the commonly assumed data model for Super-Resolution using sampling theory. We propose a framework that employs the usual anti-aliasing prefilter to orthogonally project the input signal on the desired space that we define with a Riesz basis. In particular, we use Delaunay triangulation and a B-spline basis to define a practical SR algorithm. We demonstrate the effectiveness of this method with synthetic and real data experiments alike. Additionally, we extend the result that allows interchangeability of the warping and blurring operators for translational movement to rotation and shift for rotationally symmetrical Point Spread Function (PSF).

The rest of the chapter is organized as follows. Section 4.2 explains the weaknesses of the usually assumed data model for Super-Resolution. In section 4.3 we develop a sampling theory-based approach for the SR problem. Section 4.4 proposes a concrete algorithm for SR within the proposed framework. Competing approaches are exposed in section 4.5 and in section 4.6 the proposed method is tested with synthetic and real data experiments and the results are compared with them. Finally, the conclusions are drawn in section 4.7. Additionally, a proof of a theorem that is referenced in the chapter is exposed in the chapter appendix 4.A.

4.2 The Data Model

Usually (as in Elad and Feuer (1997), for instance), the image formation process for N low resolution measured images has been modeled with the equation

$$\mathbf{w}_k = \mathbf{D}_k\mathbf{H}_k\mathbf{G}_k\mathbf{z} + \mathbf{n}_k \qquad k = 1, \ldots, N, \tag{4.1}$$

where $\mathbf{w}_k$ is a vector with the low resolution pixels from the k^{th} low resolution image, $\mathbf{z}$ contains the pixels from the high resolution image and $\mathbf{n}_k$ comes from an additive noise process. $\mathbf{G}_k$ is the geometric motion operator between the HR and the various LR frames, whereas $\mathbf{H}_k$ models the blurring due to the PSF of the imaging system and $\mathbf{D}_k$ is the down-sampling operator.

It is important to recall that (4.1) is an approximation of the continuous model

$$w_k[m,n] = d_k\left(h_k(x,y) * (z(g_k(x,y)))\right) + n_k[m,n], \tag{4.2}$$

which we have slightly modified from the one in Irani and Peleg (1993). In this equation, z is the continuous image that we want to obtain, g_k is an operator that produces the geometric warping of the image, h_k is the PSF of the sensor, $*$ denotes a convolution, and d_k is the down-sampling operator. The discrete function w_k is composed by the pixels of the k^{th} measured low resolution image and n_k is the additive noise. To pass from (4.2) to (4.1) we discretize the HR image, in practice limiting its bandwidth. If $\mathbf{z}$ is the sampling of $z(x,y)$ with period $T_x = \frac{1}{2f_x}$ in the abscissas and $T_y = \frac{1}{2f_y}$ in the ordinates it implies that the least squares solution to (4.1) is optimal when the input sequence comes from an image that is bandlimited within frequencies f_x and f_y plus Gaussian white noise. If the continuous image $z(x,y)$ has information at higher frequencies, we do not follow the data model for which the least squares solution to (4.1) is optimal anymore. This happens also when we solve it in an ML (maximum likelihood) sense adopting a norm different from l_2: for instance, for the l_1 norm we would assume a data model composed of a bandlimited function plus Laplacian noise (Farsiu et al., 2004b). This mismatch between the input data and the model translates to the presence of aliasing artifacts due to the undersampling of the HR image, as we will see in later sections. The diffraction limit determines the maximum bandwidth of an optical system, so we could just increase the sampling rate of z until we include all possible information, but this is not always practical. We do not always know all the characteristics of the imaging system we are dealing with or the diffraction limit is too high and it is computationally too expensive for calculating such a big HR image. In those cases we must be careful with the optimality of ML or MAP (maximum a posteriori) solutions to (4.1) like those proposed in Farsiu et al. (2004b).

4.3 A Sampling Theory Interpretation

Our goal is to find a solution to (4.2) that is more flexible with the data model, valid for different noise types and imaging system bandwidths. The first thing we will do is to simplify the model converting it in an interpolation problem. To do so, we need to

interchange the warping/interpolation function g_k with the convolution with the PSF, so we can solve first the interpolation problem and then make the deconvolution. In Elad and Hel-Or (2001) it was established that $\mathbf{H}_k$ and $\mathbf{G}_k$ of equation (4.1) can change its application order for translational movement. We will extend that result for the continuous equation (4.2), establishing that g_k and the convolution with h_k can be interchanged when

1. The movement is translational. This fact was already well-known and can be found in Kim et al. (1990) and Ur and Gross (1992).
2. The movement is a rotation plus a shift and the blurring operator presents radial symmetry, that is, $h_k = h_k(x^2 + y^2)$. We prove this in section 4.A.

These results let us interchange blurring and interpolation operators with certain confidence, as most movements in a video sequence for temporally near frames can be represented as rotations and translations. Additionally, most times the PSF is supposed to have radial symmetry (in Elad and Hel-Or (2001) and Farsiu et al. (2004b), for instance, the authors assume a Gaussian PSF for synthetic and non-synthetic experiments alike). Although this is not always true for modern cameras, where the influence of the photon integration surface in the total PSF is usually more important than that of the optical PSF, that does not make a big difference for small rotations. Therefore, for the rest of the chapter, z will represent the blurred image before translations and rotations. We will also make the reasonable assumption of the equality of h_k and d_k for all frames, as we suppose them captured with the same imaging system.

Now we will draw our attention to the operator $\mathbf{G}_k$. This matrix resamples the original HR image into another rectangular grid where some of the pixels coincide with the k^{th} LR image that is finally formed. The way we perform this resampling or interpolation depends on the sampling model we have chosen. We must choose a basis to represent the signal. If it is, for instance, the tensor product of the *sinc* function, then each low resolution pixel could be expressed as (without losing generality we will assume $T_x = T_y = 1$)

$$w_k[m,n] = \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} z(i,j)\text{sinc}(x_{k,m,n} - i)\text{sinc}(y_{k,m,n} - j)$$

where $(x_{k,m,n}, y_{k,m,n})$ are the coordinates of the pixel $w_k[m,n]$ in the HR grid and N_x and N_y are the size of the HR image in the x and y directions respectively. This equation states that the measured pixels are a linear combination of the weighted basis functions at the low resolution pixel location and is valid for the registration models that allow the interchange of the blurring and warping operators. In fact, the different coefficients that multiply z for a given pixel $w_k[m,n]$ can be seen as the result of the multiplication of the matrices $\mathbf{D}_k$ and $\mathbf{G}_k$ once we have displaced to the right $\mathbf{H}_k$.

We can generalize the sampling model assuming that $z(x,y)$ belongs to the space of square-integrable functions $L_2(\mathbb{R}^2)$. The HR image we finally calculate, $\hat{z}$, is the projection of z on the space V defined by the Riesz basis $\{\varphi_{l,p} = \varphi(x-l, y-p)_{l,p\in\mathbb{Z}}\}$ (Unser, 2000), that is

$$\hat{z}(x,y) = \sum_{i\in\mathbb{Z}} \sum_{j\in\mathbb{Z}} c(i,j)\varphi_{i,j}.$$

The approximation of z in V with minimum error is obtained minimizing the distance

$$\hat{z}(x,y) = \arg\min_{t\in V} \|z-t\|^2. \tag{4.3}$$

The optimal solution to (4.3) is (see Unser (2000))

$$\hat{z} = \sum_{i\in\mathbb{Z}} \sum_{j\in\mathbb{Z}} \langle z, \tilde{\varphi}_{i,j}\rangle \varphi_{i,j},$$

where the $\tilde{\varphi}_{i,j}$'s are the dual basis functions of the $\varphi_{i,j}$'s. These functions are called the analysis and synthesis functions respectively, and they fulfill the biorthogonality condition

$$\langle \tilde{\varphi}_{i,j}, \varphi_{m,n}\rangle = \delta_{i-m,j-n}.$$

In the case of the *sinc* basis functions, the analysis and synthesis functions coincide and the biorthogonality condition translates to an orthogonality condition among the shifted *sinc* functions.

To find the coefficients $c(i,j)$ we can, instead of minimizing (4.3), follow the equivalent sampling process that can be seen in Fig. 4.1. We apply a prefiltering step, using the analysis functions, that orthogonally projects the input to the space V and next we sample the output of the filter, obtaining the coefficients $c(i,j)$. These coefficients will be used to obtain the reconstructed signal $\hat{z}$ that is the solution to (4.3), filtering with the

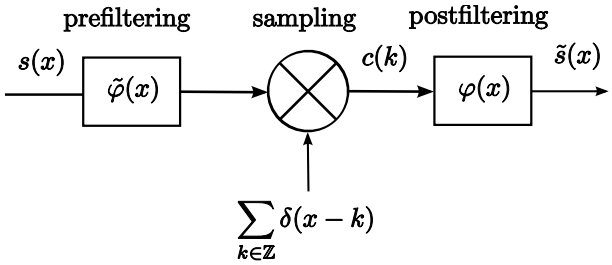


Figure 4.1: Standard three-step sampling paradigm.

synthesis functions. This is the approach we will follow in our method. The important thing to note here is that we introduce an anti-alias prefilter that no other SR method has used before and that produces an optimal solution to (4.3).

4.4 Proposed Algorithm

In this section we propose a method that fits in the previously described sampling paradigm. We first build a continuous function using Delaunay triangulation and then we project it on the space of polynomial splines of degree n, using a B-spline basis.

4.4.1 Prefiltering Using B-Spline Basis

We face the following dilemma: we need the continuous image z to project it on the chosen space of representable signals, but we only have the samples from the low resolution images. We also want to make sure that the projection we apply is orthogonal to V, which implies that we must apply a prefiltering step. To solve this problem we will try to find an approximation of the continuous (and noisy) function z using the available measured pixels. The ideal solution would be to use radial basis functions (RBF, see Powell (1990)), which are functions centered in the sampling points that satisfy certain smoothness constraints in the built function. Unfortunately, the use of RBFs would lead to a very ill-conditioned and computationally expensive problem due to the unbounded nature of these functions. Instead, we will build z using the Delaunay triangulation of the sampling points, which is a reasonable approximation when we have enough sampling points.

Delaunay triangulations maximize the minimum angle of all the angles of the triangles in the triangulation. We will use it to make a triangulation of the low resolution sampling points that we have and then build a piecewise linear function in two dimen-

sions. Each triangle will have an area $T_p \subset \mathbb{R}^2$. If we have P triangles, we will define the continuous function as

$$z = \sum_{p=1}^{P} z_p, \tag{4.4}$$

with

$$z_p = \begin{cases} k_p^1 x + k_p^2 y + k_p^3, & (x,y) \in T_p \\ 0, & \text{otherwise,} \end{cases}$$

where the constants k_p^1, k_p^2, and k_p^3 are defined so that z passes through all the low resolution pixel values we have. This can be considered a first order approximation of z in an irregular grid. Delaunay triangulation has already been used by Lertrattanapanich and Bose (2002) for super-resolution, but just sampling the built function z without the prefiltering step.

Now we must choose a basis to project z. The *sinc* function has the disadvantage of its infinite support and a low decay rate (with $1/|x|$). B-splines provide a good approximation of the *sinc* behavior (Unser, 1999), have compact support and an explicit expression (see Ahuja et al. (2005) for an analysis that concludes that B-spline basis are the most suitable basis for SR). For one-dimensional functions, that means projecting the function on the space S_T^n of continuous functions that are equal to a polynomial of degree n on each interval $[kT, (k+1)T), k \in \mathbb{Z}$. The period will approximately determine the maximum representable frequency in the HR image, as the S_T^n space rapidly converges to the band-limited space with maximum representable frequency $f = \frac{1}{2T}$ as n increases. Measured in the HR grid, we will have $T = 1$, and that is the convention we will follow in the next formulae. If we adopt the B-splines of order n as basis functions for z, we have

$$\hat{z} = \sum_{i \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} c(i,j) \beta^n(x-i) \beta^n(y-j).$$

In Unser et al. (1992) the prefiltering for B-splines of order n is carried out in three steps:

1. A convolution between the input signal and the continuous kernel $\beta^n(x)\beta^n(y)$ is performed
2. The resulting function is sampled to provide the discrete sequence $a(i,j)$
3. Finally, we do a discrete filtering using a direct B-spline filter of order $2n+1$

The first and second steps can be performed jointly. That is done calculating the

continuous convolution, sampled at the points (i, j), which is

$$a(i,j) = \iint \beta^n(u-i)\beta^n(v-j)z(u,v)dudv. \tag{4.5}$$

As the B-splines are piecewise polynomial and z is piecewise linear, we can solve (4.5) analytically integrating it in the intersection of the differentiable areas of both functions. We can easily split the B-spline functions in the areas where they are defined by polynomials. For instance, for the centered linear B-spline

$$\beta^1(x) = \begin{cases} -|x|+1, & 0 \leq |x| < 1 \\ 0, & 1 \leq |x|, \end{cases}$$

we can develop it as

$$\beta^1(x) = \begin{cases} x+1, & -1 < x < 0, \\ -x+1, & 0 \leq x < 1 \\ 0, & 1 \leq |x|. \end{cases}$$

That is, we have two integration areas different from zero defined by polynomials. The expression of the tensor product of two linear B-splines would produce four areas in the plane where the function will be defined as a polynomial different from zero. In general, for the tensor product of two B-splines of order n, we will define $(n+1)^2$ areas. We will name the polynomial in these areas $B_m^n(x,y), m = 1,..,(n+1)^2$, with m increasing as we move first increasing x and then increasing y. In Fig. 4.2 we can see an example for B-splines of order two. There are nine squares for which the function $\beta^2(x)\beta^2(y)$ has different polynomials. For each of these squares we calculate the intersection with the Delaunay triangles. Once we have that polygon (for instance, the dark area in Fig. 4.2), we will divide it in strips always defined by four lines, where two of them will be constant values of y. The intersection of those lines will define the final integration area. For each B-spline polynomial B_m^n and each coefficient $a(i,j)$ we will define an index set $P_{m,i,j}$ that will contain the indexes to the triangles that have a part of its area within the area where $B_m^n(x-i, x-j)$ is defined. For each triangle z_p with $p \in P_{m,i,j}$ we will have Q_r strips defined by the lines $x = s_{q,r}^1 y + s_{q,r}^2$, $x = s_{q,r}^3 y + s_{q,r}^4$, $y = s_{q,r}^5$, and $y = s_{q,r}^6$, with $q = 1,..,Q_r$ and being r the 4-tuple $r = (m,i,j,p)$. In the case of the polygon in the figure we would integrate two strips with different delimiting lines. In the general case what

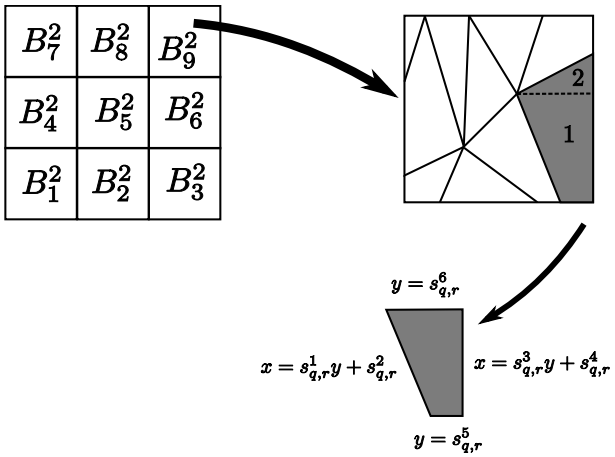


Figure 4.2: Integration areas for (4.6) for a quadratic B-spline. There are 9 integration areas with different polynomials. The figure shows how one triangle intersects with the integration area of B_9^2 and the intersection is divided in 2 strips. The equations for the limiting lines for one of them are also shown.

we have is

$$a(i,j) = \sum_{m=1}^{(n+1)^2} \sum_{p \in P_{m,i,j}} \sum_{q=1}^{Q_r} \int_{v=s_{q,r}^5}^{v=s_{q,r}^6} \int_{u=s_{q,r}^1 v + s_{q,r}^2}^{u=s_{q,r}^3 v + s_{q,r}^4} B_m^n(u-i, v-j) \cdot z_p(u,v) du dv$$

$$= \sum_{m=1}^{(n+1)^2} \sum_{p \in P_{m,i,j}} \sum_{q=1}^{Q_r} \int_{v=s_{q,r}^5}^{v=s_{q,r}^6} \int_{u=s_{q,r}^1 v + s_{q,r}^2}^{u=s_{q,r}^3 v + s_{q,r}^4} B_m^n(u-i, v-j) \cdot (k_p^1 u + k_p^2 v + k_p^3) du dv, \qquad (4.6)$$

which can be calculated analytically to obtain a polynomial in function of the parameters k_p^1, k_p^2, k_p^3 and $s_{q,r}^1, ..., s_{q,r}^6$. After obtaining these expressions for each B_m^n, we can easily find the contribution of a triangle for a certain coefficient just finding the integration strips and substituting the parameters in the expression. Equation (4.6) can be efficiently implemented first reserving memory for the coefficients and setting it to zero and then running through all the triangles in the triangulation, adding for each triangle its contribution to different coefficients. This implies that the method is linear in time with the number of triangles and linear in memory with the number of output pixels, which is a big improvement when comparing with MAP approaches that must solve a huge general system of linear equations.

Once we have the $a(i,j)$ coefficients we can perform the third step of the prefiltering. We have to calculate the $c(i,j)$ samples finding the B-spline coefficients of order $2n+1$ for the data $a(i,j)$. This can be done using a very fast recursive digital filter as explained

in Unser et al. (1991).

4.4.2 Implementation for Cubic B-splines

In the experiments we have used a cubic B-spline, which has a good tradeoff between computational complexity and close behavior to the sampling system constructed when we use as basis shifted *sinc* functions. Its representation is

$$\beta^3(x) = \begin{cases} 2/3 - |x|^2 + |x|^3/2, & 0 \leq |x| < 1 \\ (2 - |x|)^3/6, & 1 \leq |x| < 2 \\ 0, & 2 \leq |x|. \end{cases}$$

We have used (4.6) to find the expressions needed to calculate the $a(i,j)$ coefficients, obtaining quite complex polynomials that we will not reproduce here because of its lack of interest. To find the $c(i,j)$ coefficients we need the impulse response of the B-spline digital filter of order seven. This B-spline has Z-transform

$$S^7(z) = \frac{5040z^3}{1 + 120z + 1191z^2 + 2416z^3 + 1191z^4 + 120z^5 + z^6}, \tag{4.7}$$

which can be implemented as two recursive filters, one causal and another one anti-causal, as explained in Unser et al. (1991). This kind of techniques have also been exploited for Gaussian filtering (Young and van Vliet, 1995; Triggs and Sdika, 2006). The denominator of $S^7(z)$ has six real roots, three of them inside the unit circle and the other three outside. The former will be included in the causal filter and the latter in the anti-causal one. Separating the denominator of (4.7) in its causal and anti-causal parts (the α's are the coefficients that has the polynomial with roots in the unit circle and the γ's correspond to the polynomial with roots outside it), we have

$$S^7(z) = \frac{5040}{(1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \alpha_3 z^{-3})(\gamma_1 + \gamma_2 z + \gamma_3 z^2 + z^3)} = \frac{D(z)}{X(z)} \frac{Y(z)}{D(z)}.$$

We can now obtain the recursive equations that implement the IIR filter,

$$\begin{aligned} d(n) &= 5040x(n) - \alpha_1 d(n-1) - \alpha_2 d(n-2) - \alpha_3 d(n-3) \\ y(n) &= \tfrac{1}{\gamma_1} \left(d(n) - \gamma_2 y(n+1) - \gamma_3 y(n+2) - y(n+3) \right). \end{aligned}$$

We will apply initially the first equation moving forwards to find the $d(n)$ coefficients and then we will move backwards to obtain the final output of the filter using the second equation. For images we must apply this filter once in the x direction and once in the y direction. This is possible because the filter is separable.

4.5 Competing Approaches

In this section we will briefly describe the methods with which we have compared our proposal to. All of them lack the prefiltering step necessary to obtain an orthogonal projection to the reconstruction space. To solve (4.1), most of them pose a minimization problem with some regularization prior, so we will first describe these regularization terms, which may be used in different methods, and then we will describe the methods in itselves.

4.5.1 Regularization Priors

Recognizing that SR can be an ill-posed problem, which means that we can have many solutions to the problem, we must find the HR image that would be the most probable solution to it. As we do not have enough information in the already available data (the samples from the LR images), we have to suppose that there is some regularity in the ground truth that we can model as a priori knowledge. Therefore, we must search for some term that expresses the regularity of natural images and that will be the $\Omega(\mathbf{z})$ present in equation (2.10).

One of the most widely referenced regularization cost function in SR (Nguyen et al., 2001a) is the Tikhonov cost function (we reproduce here (2.11) for convenience)

$$\Omega_T(\mathbf{z}) = \|\Gamma\mathbf{z}\|_{l_2}^2, \tag{4.8}$$

where Γ is usually a high pass operator such as a derivative or Laplacian and we employ the squared l_2 norm. It can also be the identity matrix, limiting then the energy of the solution. If Γ is a high pass operator it forces spatial smoothness in the solution. As the noisy and edge pixels both contain high-frequency energy, they will be removed in the regularization process and the resulting denoised image will not contain sharp edges. Therefore, there will be a tradeoff between the denoising and the sharpness of

the reconstructed image.

In some of the experiments to follow, we have used a Tikhonov regularization with operator

$$\Gamma = \frac{1}{8}\begin{pmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \tag{4.9}$$

which is a matrix realization of the Laplacian kernel.

A very successful regularization method originally used for denoising and deblurring is the total variation (TV) method (Rudin et al., 1992). The TV criterion penalizes the total amount of change in the image as measured by the l_1 norm of the magnitude of the gradient and is defined as

$$\Omega_{TV}(\mathbf{z}) = \|\nabla \mathbf{z}\|_{l_1}, \tag{4.10}$$

where ∇ is the gradient operator. The most useful property of the TV criterion is that it tends to preserve edges in the reconstruction (Bovik and Gibson, 2000; Rudin et al., 1992), as it does not severely penalize steep local gradients.

Based on the spirit of the TV criterion and on a related technique known as the bilateral filter (Tomasi and Manduchi, 1998), Farsiu et al. (2004b) introduced a robust regularizer called bilateral Total Variation (bilateral TV or BTV). This regularizer is computationally cheap and preserves edges. The regularizing function looks like

$$\Omega_{BTV}(\mathbf{z}) = \sum_{l=-P}^{P}\sum_{m=1}^{P} \alpha^{|m|+|l|}\|\mathbf{z} - S_x^l S_y^m \mathbf{z}\|_{l_1} + \sum_{l=1}^{P} \alpha^{|l|}\|\mathbf{z} - S_x^l \mathbf{z}\|_{l_1}, \tag{4.11}$$

where matrices (operators) S_x^l and S_y^m shift $\mathbf{z}$ by l and m pixels in the horizontal and vertical directions respectively. This expression calculates derivatives of $\mathbf{z}$ at different scales and at different directions. The scalar weight α, $0 < \alpha < 1$, is applied to give a spatially decaying effect to the summation of the regularization terms: we penalize big gradients in nearby pixels. As the differences are calculated in different directions, the behavior is isotropic even although we are using the l_1 norm, thing that does not happen with (4.10). In that case, the l_1 norm of the gradient favors the x and y directions, as we calculate differences only in those directions. We have changed a little the expression that appears in Farsiu et al. (2004b), because there is a small asymmetry in the way they

pose the equation. To solve that issue we have separated the expression in two terms, being the first of them all directions but the abscissas and being the second that missing direction (in that term, $m = 0$).

We will use (4.8) with operator (4.9) and (4.11) as regularization priors in the experiments with different methods. Note that we do not need this kind of regularization in the proposed method, as if the SR problem we are dealing with is ill-posed, that issue is solved when the Delaunay triangulation is calculated.

4.5.2 The l_2 Norm Method

Using the l_2 norm without a regularization prior means solving (2.9) as a least squares problem. We would simply have to solve for $\mathbf{z}$ in

$$\mathbf{A}^T\mathbf{A}\mathbf{z} = \mathbf{A}^T\mathbf{w}. \tag{4.12}$$

The solution would be optimal in an ML sense when the samples have additive i.i.d. zero mean Gaussian distributed noise.

If we wish to introduce regularization terms, we will solve the MAP problem

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}} \left[\|\mathbf{A}\mathbf{z} - \mathbf{w}\|_{l_2}^2 + \lambda \|\Gamma\mathbf{z}\|_{l_2}^2 \right]. \tag{4.13}$$

This is one of the methods we have used for benchmarking our projective method, using as regularization term (4.8) with the Laplacian operator (4.9) as kernel. This was the method proposed by Elad and Feuer (1997). We have used it in the experiments as it can be considered the most "standard" SR method nowadays, having been widely implemented.

4.5.3 The l_1 Norm Method

The l_1 method differs from the l_2 method in that its optimality happens when the samples are contaminated by additive i.i.d. zero mean Laplacian distributed noised instead of Gaussian. In the experiments we have used an l_1 norm with a bilateral TV prior, as

suggested by Farsiu et al. (2004b). The image that we want to find is

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}} \left[\|\mathbf{A}\mathbf{z} - \mathbf{w}\|_{l_1} + \lambda \Omega_{BTV}(\mathbf{z}) \right]. \tag{4.14}$$

The parameters for the bilateral TV prior defined in (4.11) will be in all cases $P = 2$ and $\alpha = 0.8$. The reason for the inclusion of this methods are its good performance when there are outliers present, thing that happens frequently due to unavoidable registration errors and also because of the nice behavior of the bilateral TV prior when it is applied to natural images.

4.5.4 The Shift and Add Method

The shift and add method is a very fast non-iterative method proposed by Elad and Hel-Or (2001) that solves the SR problem under a special case where

1. All the decimation operators are assumed to be the same for all frames.
2. All the blur operations are assumed to be the same for all frames. Moreover, the blurring matrix is assumed to be block circulant, representing a linear space invariant blur.
3. All the warp operations correspond to pure translations. Thus, the warping matrices are all block circulant as well. Moreover, it is assumed that the warping is represented through the nearest neighbor displacement paradigm, which means that the displacement in the finer grid is rounded and the warping matrices apply only integer translations in the HR grid. This can lead to less quality in the image if the SR amplification factor is small.
4. The additive noise is white and i.i.d.

When all these conditions apply, it is very easy to calculate $\mathbf{z}$. Solving for $\mathbf{z}$ in (4.12), we have

$$\mathbf{z} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{w}. \tag{4.15}$$

Under the conditions specified above, $\mathbf{A}^T\mathbf{A}$ is a diagonal matrix, so we can calculate its inverse just with a division per HR pixel. Then we just need to multiply by matrix $\mathbf{A}^T$ to find the HR image.

The shift and add method obtains very good quality images with very low computational load, so it is a good method for comparison to see if the additional complexity

that other methods introduce is compensated by an increase in quality with regard to this method.

4.5.5 The Zomet Method

The Zomet method (Zomet et al., 2001) is another robust method that has very good performance when outliers are present. To expose it, we will recall equation (2.6), which is

$$\mathbf{w}_k = \mathbf{A}_k\mathbf{z} + \mathbf{n}_k \quad \text{for} \quad k = 1, \ldots, N.$$

In this equation, we are considering separately each LR image instead of stacking all vectors as in (2.7). The expression we have to minimize to find the HR image adopting the l_2 norm is

$$L(\mathbf{z}) = \sum_{k=1}^{N} \|\mathbf{w}_k - \mathbf{A}_k\mathbf{z}\|_{l_2}^2. \tag{4.16}$$

In most iterative methods, we find the gradient of (4.16) in each step to update the solution estimated in each iteration by the expression

$$\mathbf{z}^{n+1} = \mathbf{z}^n + \beta\nabla L(\mathbf{z}), \tag{4.17}$$

where β is a scale factor defining the step size in the direction of the gradient. The gradient of L is the sum of the gradients computed over the input images

$$\nabla L(\mathbf{z}) = \sum_{k=1}^{N} \mathbf{b}_k, \tag{4.18}$$

where

$$\mathbf{b}_k = 2\mathbf{A}_k^T(\mathbf{A}_k\mathbf{z} - \mathbf{w}_k). \tag{4.19}$$

In order to introduce robustness into the procedure, the sum of gradients coming from different images in (4.18) is replaced with a scaled median in each pixel of the HR image,

$$\nabla L(\mathbf{z})(i,j) = N \cdot \text{MEDIAN}\left\{\mathbf{b}_k(i,j)\right\}_{k=1}^{N}. \tag{4.20}$$

For a symmetric distribution a median can approximate the mean quite accurately given a sufficient set of samples. In case of distant outliers the median is much more robust than the mean. This is the reason why the Zomet method works.

If some prior assumptions can be made about the solution one can combine this robust scheme with a constrained optimization algorithm. In the experiments, we have done this adding a Tikhonov prior term to (4.16).

4.5.6 The Delaunay Triangulation Method

The method proposed by Lertrattanapanich and Bose (2002) uses a Delaunay triangulation as in our method to obtain the function that we have defined in (4.4). The difference with our method is that they simply sample the obtained continuous function in the sampling interval defined by the HR grid. This implies that possibly aliasing is happening in the HR image as in the previously exposed methods, with the added drawback of not offering any protection against noise. In any case, we have included this method in the set we compare to, due to its similarity with ours. This is because both use a Delaunay triangulation to pass from a set of discrete irregular samples to a continuous function.

4.6 Experiments

We will make five experiments to show the behavior of our approach, two of them with synthetic images, and three with real data taken with common low-end imaging systems.

4.6.1 Experiments with Synthetic Data

The first two experiments will show the improvement in resolution for high frequencies that brings our method when the HR image has less resolution than what could be provided by the bandwidth of the imaging system. This will be done with simulated video sequences. For the first of them we will generate a sequence of synthetic images sampling the function

$$f = \frac{1}{2}\left(1 + \cos\left(40 \arctan \frac{y}{x}\right)\right). \tag{4.21}$$

This function has also been used for super-resolution tests by Borman and Stevenson (2004). This function oscillates around the origin for any given radius and has the peculiarity that it is not bandlimited: when we approach to the origin, the frequencies rapidly increase. We have generated 20 low resolution images sampling this function

and then applied different super-resolution methods with an SR factor of four. The LR images have been randomly shifted with shifts following a uniform distribution between zero and one pixels. In Fig. 4.3 we can see two different LR frames, the cubic spline interpolation of the first of them and the sampling of the function at the super-resolution rate. It can be seen the high degree of aliasing present in the low resolution images and also, to a lesser extent, in the image sampled at higher rate. Figure 4.4 shows the results for different SR methods. The MDSP software package from the University of California has been used for the shift and add method (Elad and Hel-Or, 2001), the iterative MAP methods based on the l_2 and l_1 norms (Farsiu et al., 2004b), and for the Zomet method (Zomet et al., 2001). No deblurring has been made in any case, that is, $\mathbf{H}_k = I$. The parameters of the different methods have been tuned to maximize the PSNR of the images, defined as

$$\text{PSNR} = 10 \log_{10} \left(\frac{max(I)^2}{\frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|I(i,j) - K(i,j)\|^2} \right), \tag{4.22}$$

where we are comparing two images I and K of size $m \times n$. In Table 4.1 we can see the values for the different parameters for the iterative methods (being λ the weight of the regularization prior and β the step size for the gradient descent method) and the results of the experiment for all methods. Our method achieves the best PSNR among all tested algorithms. Besides that quantitative datum, it can be seen that the frequency resolution is greatly improved by our method comparing with shift and add, l_2 and l_1 with regularization and Zomet, as the aliased center area of the images is much smaller. The borders are also more clearly defined. The method by Lertrattanapanich and Bose (2002), also based on Delaunay triangulation, performs similarly, but this method has the disadvantage of its high sensitiveness to noise, as we will see in the next experiment. This is due to the direct sampling it makes of the function built with the triangulation.

Table 4.1: Parameter Values and Results for Experiment One

Experiment One	λ	β	Iterations	PSNR (dB)
Shift&Add	—	—	—	18.75
l_2 with Tikhonov	0.2	8	20	20.48
l_1 with bilateral TV	0.01	20	50	18.01
Zomet with Tikhonov	0.1	2	50	18.37
Delaunay	—	—	—	21.16
Projection	—	—	—	21.37

For the second experiment we will generate 20 low resolution images from Fig. 4.5a, downsampling the original by four. The LR images have been randomly shifted with shifts following a uniform distribution between zero and one pixels. These images are blurred with a Gaussian kernel with shape $\exp(-(x^2+y^2)/(2\sigma^2))$ with width parameter $\sigma^2 = 1$ and finally white Gaussian noise is added until achieving a PSNR of 17 dB (as defined in (4.22)). The presence of noise is important to simulate SR, as there are theoretical limits in registration (Pham et al., 2005a) that introduce it in real world SR. We have also generated an image with half the resolution of Fig. 4.5a, which can be seen in Fig. 4.5b. To generate this image we first apply a low pass filter to the original image and then we downsample it, following a standard sampling process that minimizes the approximation error. In the experiment, we will use the low resolution images to generate an HR image that multiplies the resolution by two, without reaching the resolution of the image that originated the LR sequence. This means that the usual SR methods will reconstruct an HR image following the incorrect assumption that the ground truth is band-limited to the frequencies determined by the reconstruction resolution.

In Fig. 4.6 we can see the results. The MDSP software package has been used again for all methods but the proposed one and the pure Delaunay triangulation approach. No deblurring has been made in any case, that is, $\mathbf{H}_k = I$. The parameters of the different methods have been tuned to maximize the PSNR of the images, defined as in (4.22). The reference image is Fig. 4.5b. In Table 4.2 we can see the values for the different parameters for the iterative methods (being λ the weight of the regularization prior and β the step size for the gradient descent method) and the results of the experiment for all methods. Our method achieves the best PSNR among all tested algorithms. Besides that quantitative datum, it can be seen that the frequency resolution is greatly improved by our method comparing with the rest. The borders of the lines that converge to the center of the image present aliasing artifacts in all methods but in ours. Most methods have good performance against noise but are not able to erase the high frequency artifacts. The method of Lertrattanapanich and Bose (Lertrattanapanich and Bose, 2002), also based on Delaunay triangulation, has a high sensitivity to noise, due to the direct sampling it makes of the function built with the triangulation.

To prove that the artifacts that appear in Fig. 4.6 are due to aliasing and not to noise, we will reconstruct an HR image with the same resolution as the original image, mul-

tiplying by four the resolution of the LR images. In this case the SR image has the same bandwidth as the ground truth and therefore the usual data model that assumes band-limited images up to the reconstruction resolution becomes true. We will compare our method with the iterative l_2 method with Tikhonov regularization. Again, the parameters of this method have been tuned to maximize the PSNR of the images. The parameters and the resulting PSNR values can be seen in Table 4.2. Now the PSNR is almost the same for both methods. In Fig. 4.7 the SR images are shown. The quality of both of them is similar, although some artifacts are still visible in the radii of the image for the l_2 method.

Table 4.2: Parameter Values and Results for Experiment Two

Experiment Two, x2	λ	β	Iterations	PSNR (dB)
Shift&Add	—	—	—	18.80
l_2 with Tikhonov	0.01	4	20	18.80
l_1 with bilateral TV	0.001	20	50	18.64
Zomet with Tikhonov	0.001	2	50	18.79
Delaunay	—	—	—	18.54
Projection	—	—	—	19.04
Experiment Two, x4	λ	β	Iterations	PSNR (dB)
l_2 with Tikhonov	0.1	8	20	20.26
Projection	—	—	—	20.35

4.6.2 Experiments with Real Data

The remaining experiments have been made with data taken with real imaging systems. For the third experiment we have taken a video with a mobile phone camera. We have used 20 low resolution frames for the experiment. For registering, we have first applied the hierarchical method exposed by Bergen et al. (1992), supposing a translational model. After obtaining a coarse approximation of the displacement, we have applied the multi-frame registration method of chapter 3 as a refinement step to obtain the final registration parameters. All the HR images obtained with the different methods have used the same motion vector and no deblurring has been made. Figure 4.8 shows one of the low resolution frames, its cubic spline interpolation and the result of applying our method doubling the resolution. We can see a significant improvement of quality: more details are present in the HR image and the artifacts provoked by the compression are completely removed. We have also compared with the same methods as in the previous

experiment and the results can be seen in Fig. 4.9, where we have zoomed a little area of the HR images. It is clearly seen that the higher resolution that our projection method has in frequency is translated to clearer borders in the upper window and in the door that appears in the bottom of the image. The iterative methods are not able to remove the aliasing artifacts that appear around the edges of the window and the door, whereas the Delaunay method from Lertrattanapanich shows noisier results. The weight of the regularization (λ), the gradient descent step and the bilateral TV parameters for the iterative methods have been chosen to maximize the visual quality of the final HR images. In Table 4.3 we can see the parameters finally used for the experiment.

Table 4.3: Parameter Values for Experiment Three

Experiment Three	λ	β	Iterations
Shift&Add	—	—	—
l_2 with Tikhonov	0.01	8	20
l_1 with bilateral TV	0.005	5	50
Zomet with Tikhonov	0.01	2	50
Delaunay	—	—	—
Projection	—	—	—

For the fourth experiment we have taken a black and white video with a low quality camera, a Nikon Coolpix 3100 with a resolution of 320x240 pixels. We have used 20 low resolution frames for the experiment. For registering, we have applied again the method proposed in Bergen et al. (1992) and after it the method exposed in chapter 3 as a final refinement step, supposing a translational model. All the HR images obtained with the different methods have used the same motion vector and no deblurring has been made. Fig. 4.10 shows one of the low resolution frames, its cubic spline interpolation and the result of applying our method doubling the resolution. We can see a significant improvement of quality: more details are present in the HR image and some of the words that appear in the image are now readable. We have also compared with the same methods as in the previous experiment and the results can be seen in Fig. 4.11, where we have zoomed a little area of the HR images. It is clearly seen that the higher resolution that our projection method has in frequency is translated to clearer borders in the clock and to less visible artifacts in the letters that appear in the image. The iterative methods are not able to remove these aliasing artifacts and the Delaunay method from Lertrattanapanich shows noisier results and some artifacts. The weight of the regular-

ization (λ), the gradient descent step (β) and the bilateral TV parameters for the iterative methods have been chosen to maximize the visual quality of the final HR images. In Table 4.4 we can see the parameters finally used for the experiment.

Table 4.4: Parameter Values for Experiment Four

Experiment Four	λ	β	Iterations
Shift&Add	—	—	—
l_2 with Tikhonov	0.2	2	20
l_1 with bilateral TV	0.01	5	50
Zomet with Tikhonov	0.1	2	50
Delaunay	—	—	—
Projection	—	—	—

In the fifth experiment we will show the applicability of our method to motion models with rotations and translations. Again, we have used the method by Bergen et al. (1992) and the method of chapter 3 as a final refinement step, but now allowing rotations and translations. The video sequence employed was captured with the same camera as in the previous experiment under low illumination conditions. We have used 12 LR frames, an SR factor of two, and the same registration method as in the previous experiment, but with parameters allowing rotations and shifts in the images. In Fig. 4.12 we can see two of the input LR images, the cubic spline interpolation of one of them and the SR image obtained by our method. It can be seen the improvement in the readability of the texts in the image and the drop of the image noise.

4.7 Conclusion

This chapter presents a new Super-Resolution method based on applying the anti-aliasing filter present in all sampling schemes. We raise the issue of the correctness of the data model that has been usually assumed in most SR approaches and propose a method that can handle more general data models, providing protection against aliasing in the HR images. We have shown in the experiments that this approach outperforms classical SR methods with synthetic and real data experiments. Another contribution of this chapter is the proof that we can first solve the interpolation problem and then make the deblurring not only when the motion is translational but also when there are rotations and shifts and the imaging system PSF is rotationally symmetric.

As a summary, we can enumerate the main advantages of the proposed method:

1. It prevents the presence of aliasing artifacts when the HR image is undersampled, thanks to the anti-aliasing filter. The filter also removes the high frequency noise.
2. It is non-iterative, unlike most proposed methods for SR. Therefore, it poses no convergence problems. This also means that no initialization for the algorithm is needed, which is an important issue for iterative methods that try to minimize a non-convex function.
3. It is scalable. The complexity of the triangulation is $O(n \log n)$ and the application of the prefilter is linear with n, so the final complexity is $O(n \log n)$, which makes it scalable.
4. The Delaunay triangulation provides a very strong protection against a possible ill-conditioning of the problem, as it gives the best possible conditioning amongst all triangulations. In exceptional cases, ill-conditioning may still occur though.
5. There is no parameter involved in the reconstruction, which is not a minor advantage. For MAP methods the gradient descent step and different parameters for the regularization prior are needed and it is not a trivial task to find the optimal values for a given set of images.
6. It is highly parallelizable. Once we have the triangulation, we can process each defined triangle independently of the others.

4.A Interchange of Blurring and Warping Operators

When the warping operator g is just a translation, it is easy to see that we can interchange it with the convolution. We want to show that $h(x,y) * z(x-a, y-b) = c(x-a, y-b)$, being $c(x,y) = h(x,y) * z(x,y)$. This follows immediately from the shift property of the convolution, as $c(x-a, y-b) = h(x-a, y-b) * z(x,y) = h(x,y) * z(x-a, y-b)$.

In this section we also demonstrate that the geometric warping operator g and the convolution with h of (4.2) can be applied in inverse order if g produces a rotation and translation and h presents radial symmetry.

Theorem. *Let z and h be square integrable functions in $\mathbb{R}^2$, with $h(x,y) = h_r(x^2+y^2)$ and define*

$$g(x,y) = \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Then

$$h(x,y) * z(g(x,y)) = c(g(x,y)), \tag{4.23}$$

being $c(x,y) = h(x,y) * z(x,y)$.

Proof. We take first the Fourier transform of the left part of (4.23), using the Fourier rotation and translation theorem (McGuire, 1998):

$$\begin{aligned}\mathscr{F}[z(g(x,y)) * h(x,y)] &= e^{jau}e^{jbv} \\ &\cdot Z(u\cos\theta - v\sin\theta, u\sin\theta + v\cos\theta)H(u,v).\end{aligned} \tag{4.24}$$

Now we take the transform of the right side of (4.23), obtaining

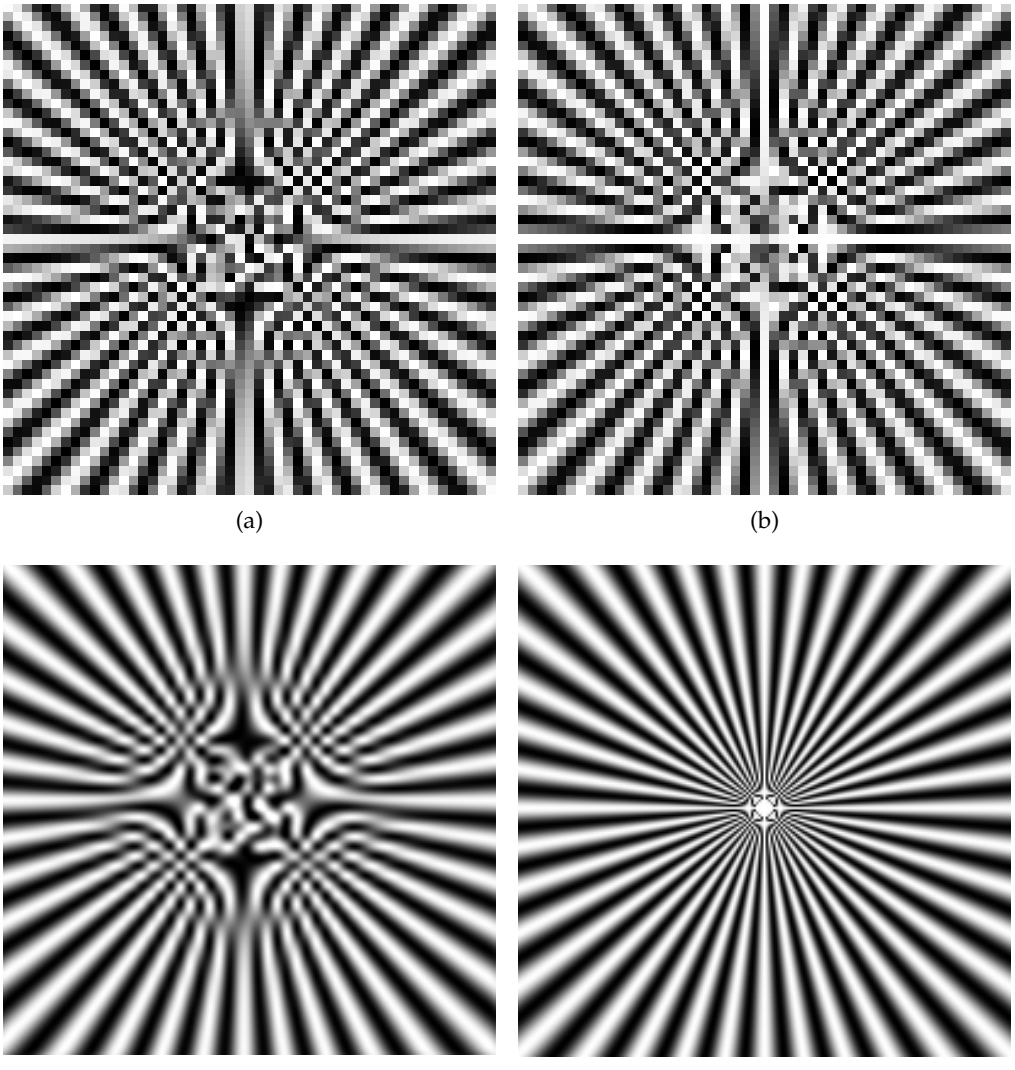
$$\begin{aligned}\mathscr{F}[c(g(x,y))] &= e^{jau}e^{jbv} \\ &\cdot Z(u\cos\theta - v\sin\theta, u\sin\theta + v\cos\theta) \\ &\cdot H(u\cos\theta - v\sin\theta, u\sin\theta + v\cos\theta).\end{aligned} \tag{4.25}$$

As $h(x,y)$ has radial symmetry, its Fourier transform also presents it, so $H(u,v) = H_r(u^2+v^2)$, and developing (4.25), we have

$$\begin{aligned}\mathscr{F}[c(g(x,y))] &= e^{jau}e^{jbv} \\ &\cdot Z(u\cos\theta - v\sin\theta, u\sin\theta + v\cos\theta)H_r(u^2+v^2),\end{aligned}$$

which is the same as (4.24) and proves (4.23).

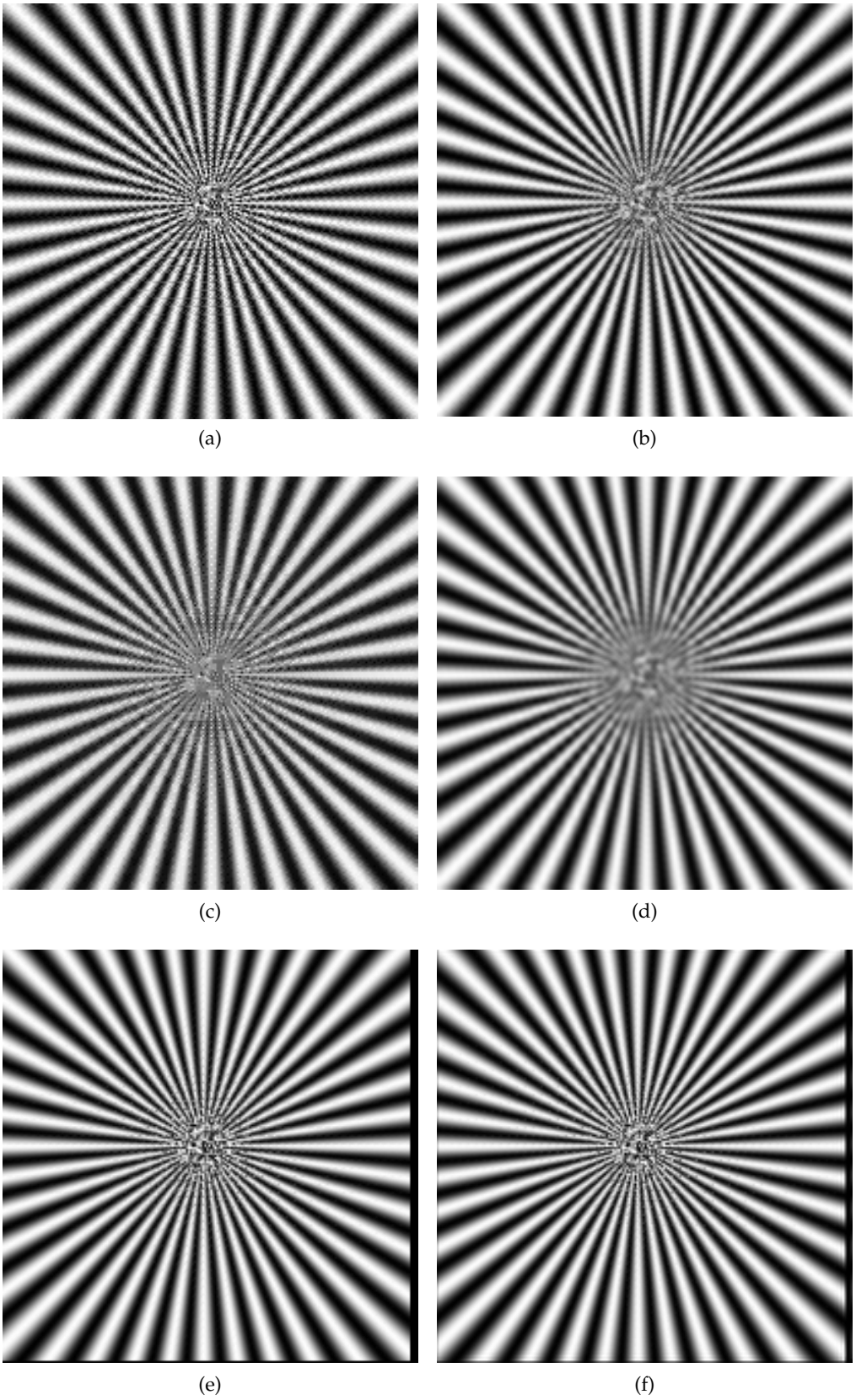
□



(a) (b)

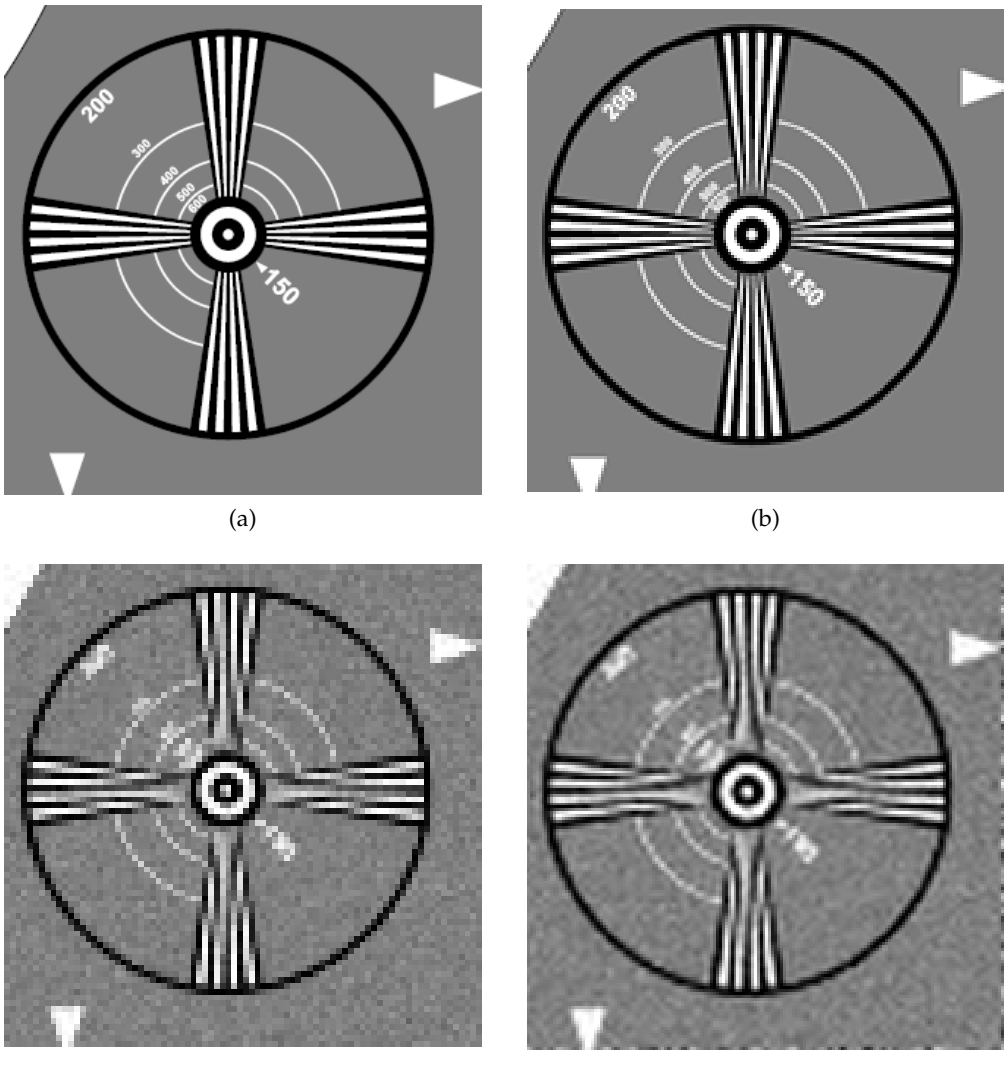
(c) (d)

Figure 4.3: Radial function sequence and generated LR images. Two low resolution frames (a) and (b), (c) a cubic spline interpolation and (d) the sampled function (4.21)



(a) (b) (c) (d) (e) (f)

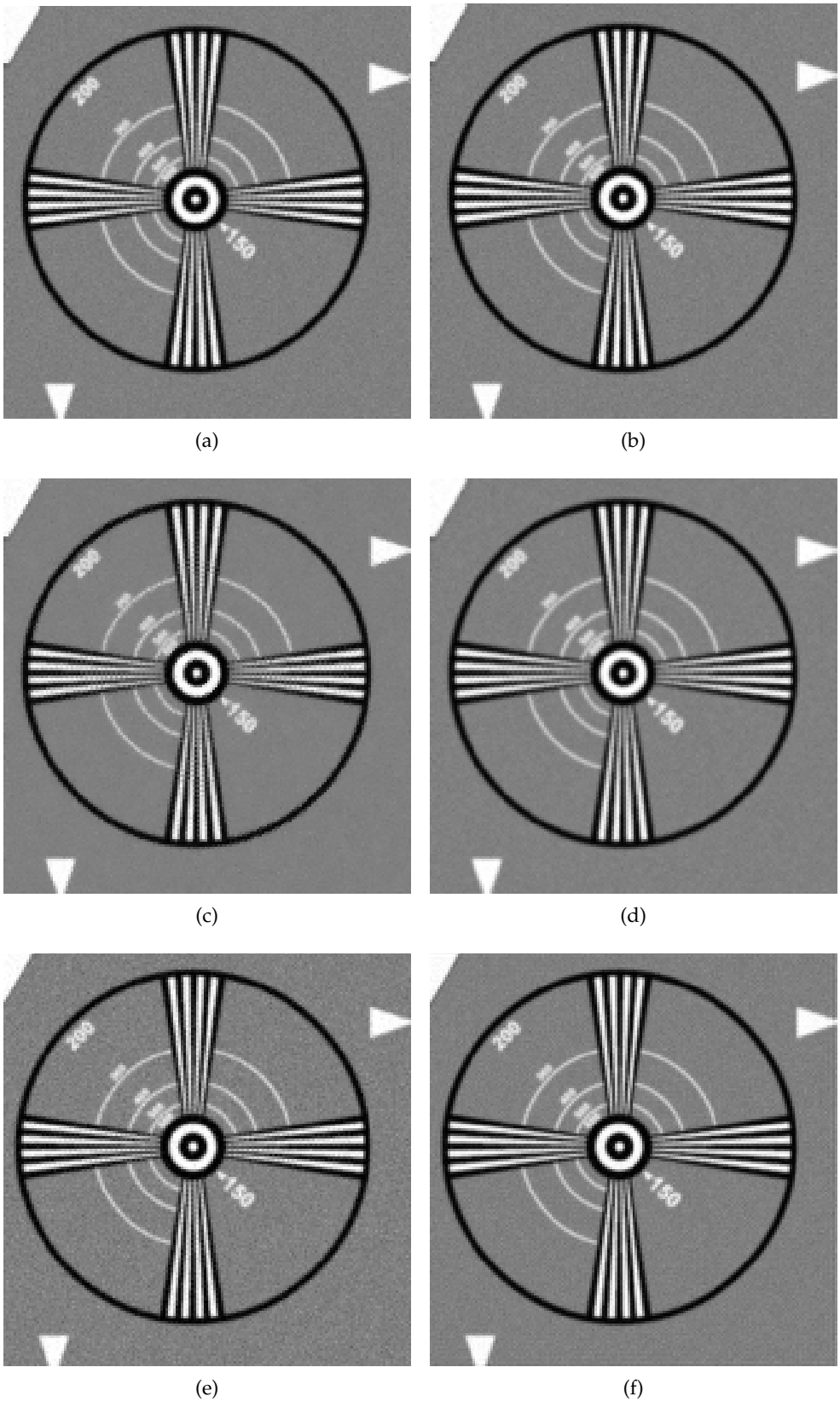
Figure 4.4: Super-resolved images for experiment one: (a) Shift&Add, PSNR=18.75 dB, (b) l_2 with Tikhonov regularization, PSNR=20.48 dB, (c) l_1 with bilateral TV, PSNR=18.01 dB, (d) Zomet method with Tikhonov regularization, PSNR=18.37 dB, (e) Delaunay based method from Lertrattanapanich and Bose (2002), PSNR=21.16 dB, and (f) our method, PSNR=21.37 dB.



(a) (b)

(c) (d)

Figure 4.5: Chart sequence and generated LR images. (a) The original image for experiment 1, (b) a half resolution version, (c) a noisy and blurry LR image generated from (a), and (d) its cubic spline interpolation.

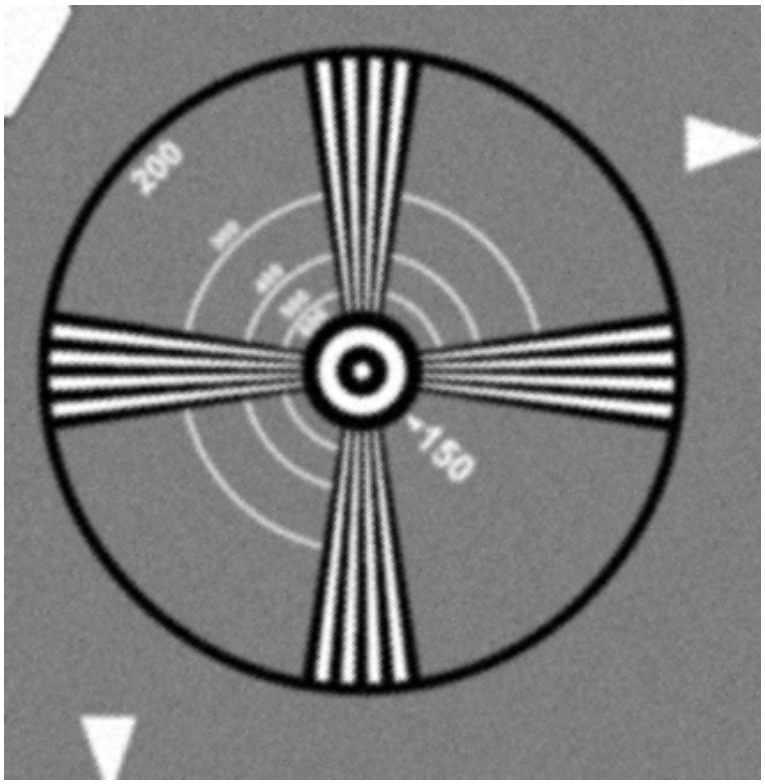


(a) (b)

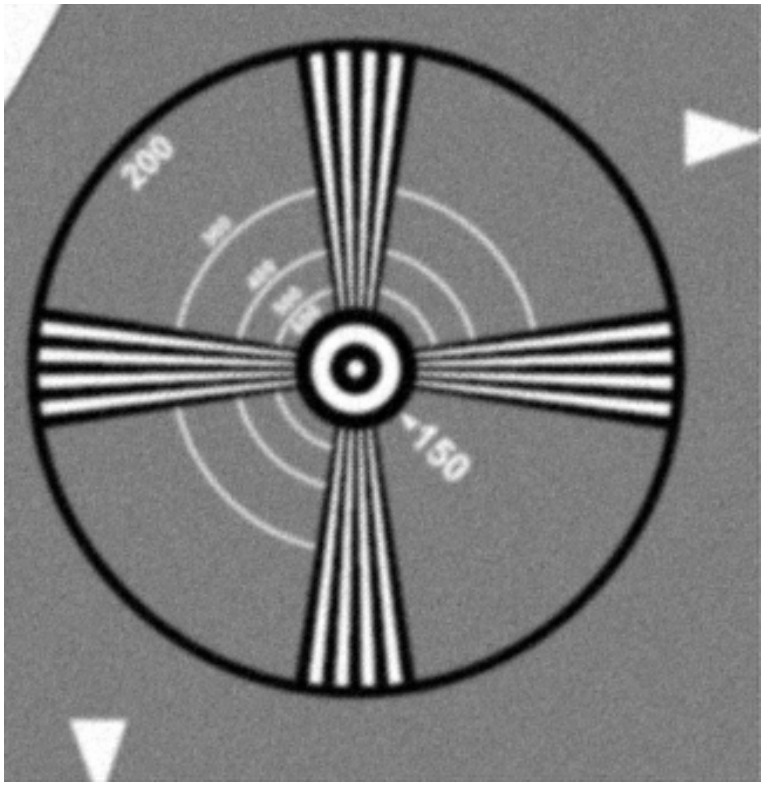
(c) (d)

(e) (f)

Figure 4.6: Super-resolved images for experiment two with SR factor 2: (a) Shift&Add, PSNR=18.80 dB, (b) l_2 with Tikhonov regularization, PSNR=18.80 dB, (c) l_1 with bilateral TV, PSNR=18.64 dB, (d) Zomet method with Tikhonov regularization, PSNR=18.79 dB, (e) Delaunay based method by Lertrattanapanich and Bose (2002), PSNR=18.54 dB, and (f) our method, PSNR=19.04 dB.

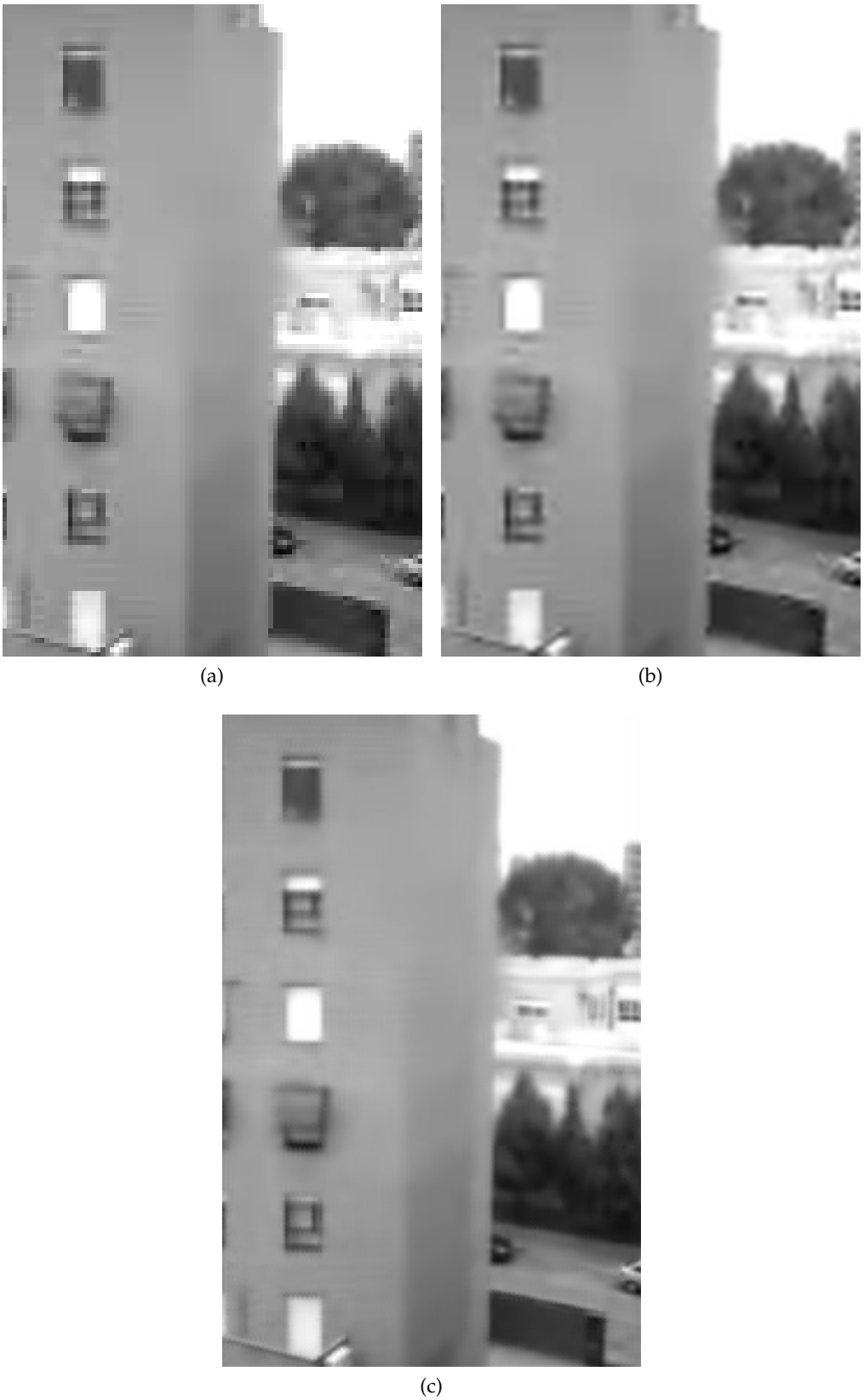


(a)



(b)

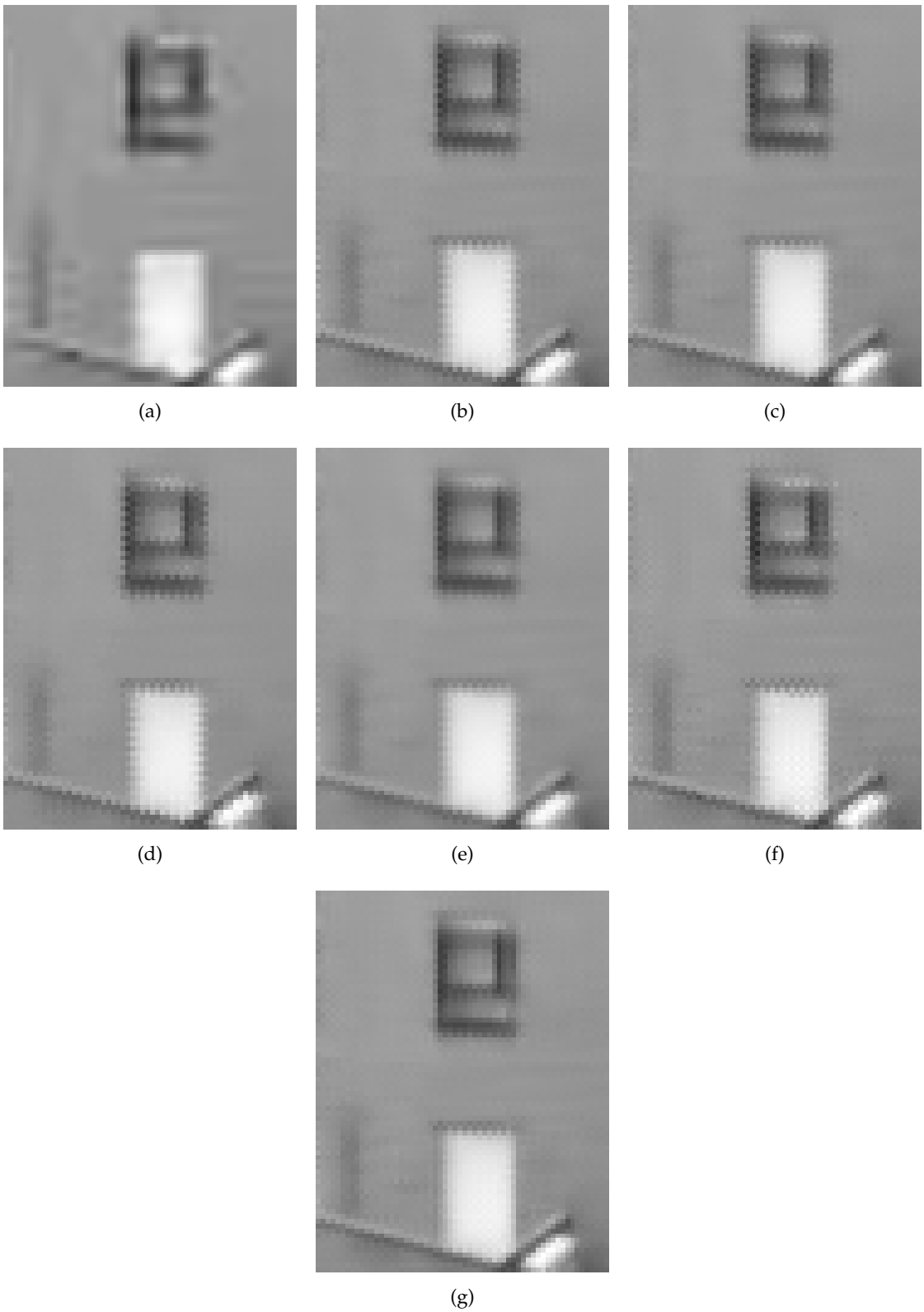
Figure 4.7: Super-resolved images for experiment two with SR factor 4: (a) l_2 with Tikhonov regularization, PSNR=20.26 dB, and (b) our method, PSNR=20.35 dB.



(a) (b)

(c)

Figure 4.8: Mobile phone sequence and generated HR image. (a) LR frame from a sequence taken with a mobile phone, (b) cubic spline interpolation of (a), and (c) SR frame obtained with our method.

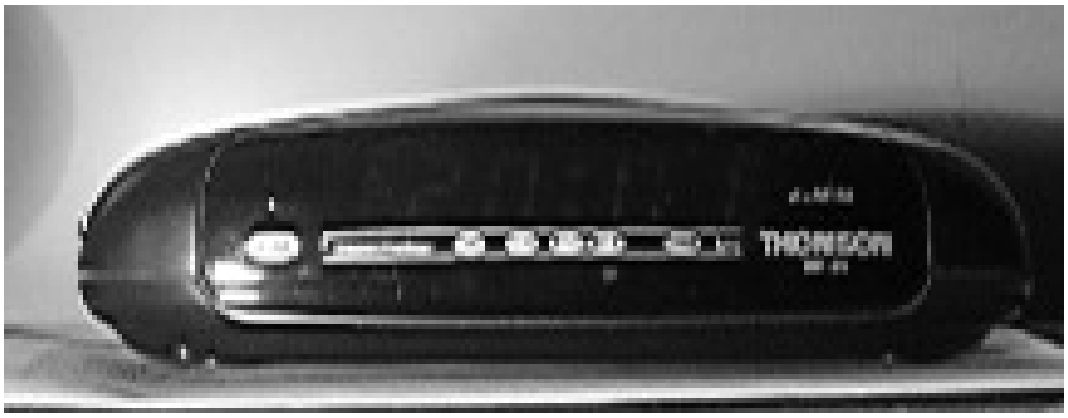


(a) (b) (c)

(d) (e) (f)

(g)

Figure 4.9: Zoom of area of the experiment three sequence with different SR methods. (a) Cubic spline interpolation of one of the LR frames, (b) Shift&Add, (c) l_2 with Tikhonov regularization, (d) l_1 with bilateral TV, (e) Zomet method with Tikhonov regularization, (f) Delaunay based method by Lertrattanapanich and Bose (2002) and (g) our method.



(a)

(b)

(c)

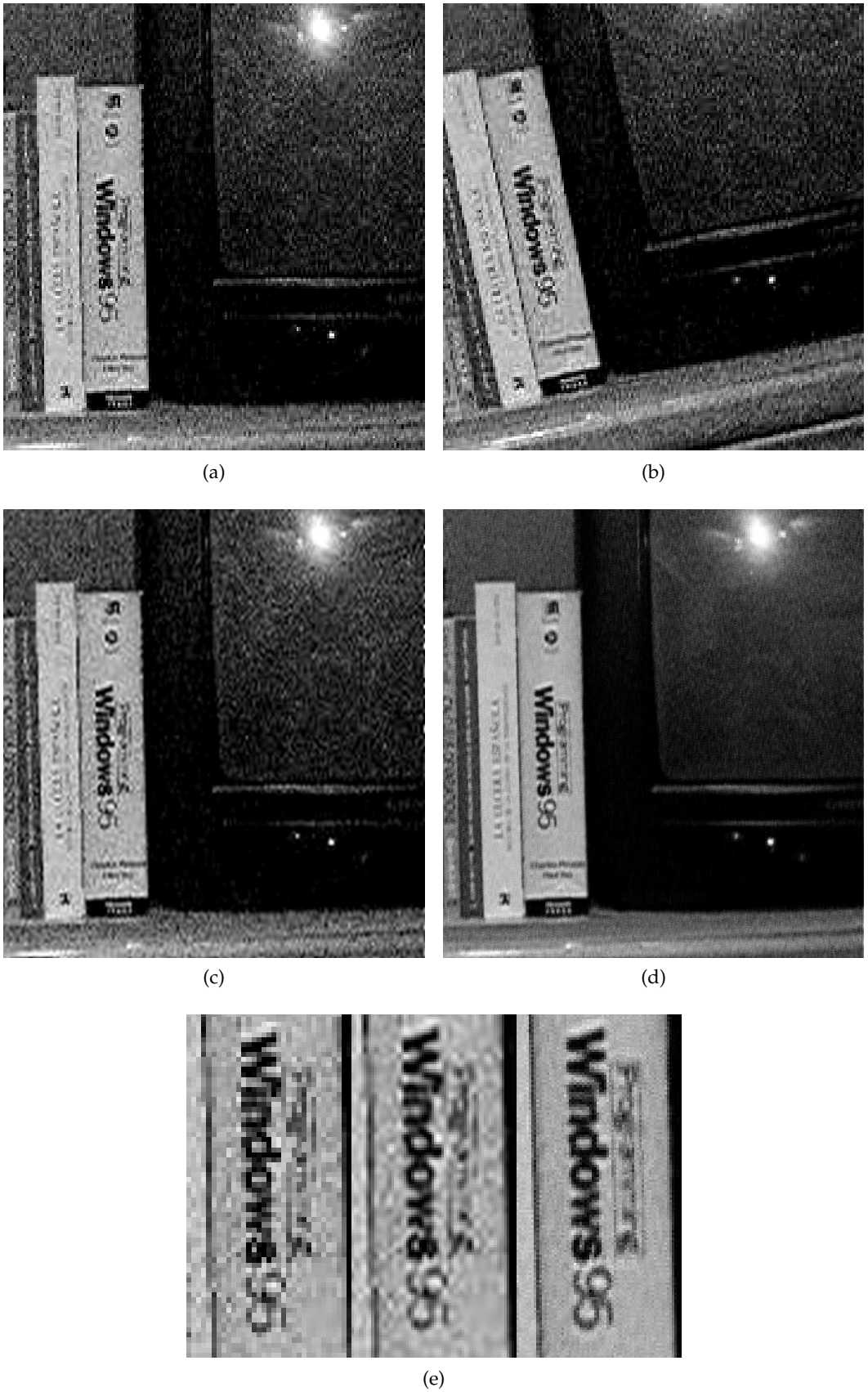
Figure 4.10: Clock sequence and generated HR image. (a) LR frame from the sequence used in experiment two, (b) cubic spline interpolation of (a), and (c) SR frame obtained with our method.

(a) (b) (c)

(d) (e) (f)

(g)

Figure 4.11: Zoom of area of the experiment four sequence with different SR methods. (a) Cubic spline interpolation of one of the LR frames, (b) Shift&Add, (c) l_2 with Tikhonov regularization, (d) l_1 with bilateral TV, (e) Zomet method with Tikhonov regularization, (f) Delaunay based method by Lertrattanapanich and Bose (2002) and (g) our method.



(a) (b) (c) (d) (e)

Figure 4.12: Input images for experiment five and super-resolved image. (a) and (b) LR frames from a video taken with a low-end digital camera, (c) cubic spline interpolation of (a), (d) SR frame obtained with our method, and (e) a detail of the 3 images.

Chapter 5

Non-Stationary Noise Removal and Deblurring

5.1 Introduction

In this chapter, we will treat specific restoration techniques for super-resolution. Restoration is understood as noise removal and deblurring. In SR, the objective is to achieve this task jointly with interpolation on the HR grid using all the available information, that is, using all the frames that we have.

Specific denoising procedures for SR are not numerous. The reason for this is that usually denoising is achieved jointly with data fusion and deblurring. As explained in section 2.3, the standard approach for SR is to consider it as an estimation problem where the original signal is embedded in noise. Solving equation (2.9) or (2.10) means solving an ML or MAP problem. Its solution will be optimal for Gaussian white noise when the norm is l_2 or optimal for biexponentially distributed white noise for norm l_1. When the SR problem is posed differently and the protection against noise is not very strong, specific denoising algorithms have been proposed, as the simple averaging performed after interpolation in Lertrattanapanich and Bose (2002).

In any case, the denoising in SR has usually assumed stationary noise, while in many real cases this is not enough. Non-stationary noise, like salt and pepper noise, or outliers due to, for instance, registration error, may appear. The techniques that deal with this kind of noise for image sequences are called robust super-resolution. In Zomet et al. (2001), a method that substituted the mean gradient among the LR images used in

iterative solutions of (2.9) by a median gradient calculated also using the LR images was proposed. Another approach to robust SR was proposed in Farsiu et al. (2004b), where instead of using the l_2 norm in (2.9), the more robust to outliers l_1 norm is employed. In this chapter we will propose an alternative approach based on the median filter and the results will be compared with those of the mentioned methods.

Deblurring is also usually performed jointly with denoising and data fusion. A specific method for deblurring separated from SR but using data from all LR frames is proposed in Farsiu et al. (2004b), where an iterative deblurring of the images is performed using as regularization a bilateral Total Variation prior. This algorithm weighted the contribution to the error to be minimized of each sample in the non-deblurred HR image using the number of LR samples that were used to calculate it. When no specific SR algorithm is used for deblurring, common methods as a Wiener filter or Lucy-Richardson deconvolution are used. In this chapter, we will compare the performance of these methods plus data fusion with integrated SR methods.

The rest of the chapter is organized as follows. In section 5.2 we will introduce a median filter for SR data, while in section 5.3 we will review different deblurring methods that can be applied after SR data fusion. In sections 5.4 and 5.5 experiments with simulated and real data are performed for robust SR methods and for deblurring methods for SR respectively. Finally, in section 5.6 we expose our conclusions.

5.2 Median Filter for Irregular Samples

A median filter applied to N samples first orders the samples in increasing order and then selects the sample in the middle as output, if N is odd. For even N we will select the two samples in the middle and its mean will be selected as output.

When applied to images, the pixels that surround a given pixel are selected and the median filter is applied to them. The selected area is usually rectangular and centered in the output sample coordinates. In some filters, the selected pixels are the center pixel and the pixels that are up, down, left and right of it, five in total. In any case the rectangular grid of the image limits the shape of the area of the selected pixels, which introduces an anisotropy that can produce artifacts in the filter output.

In super-resolution, it makes sense to apply a median filter not in the resampled

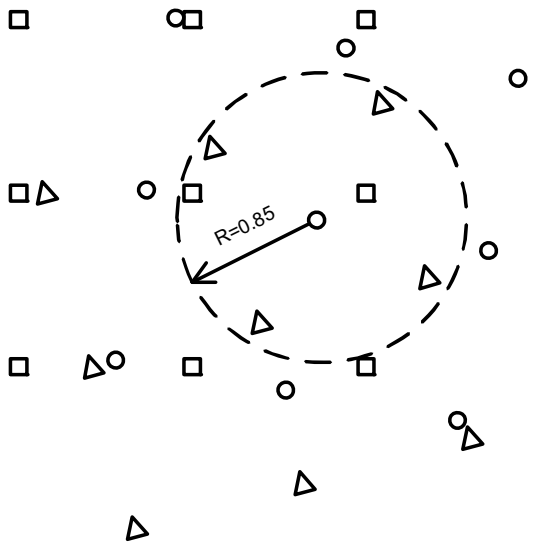


Figure 5.1: Median filter for irregular samples. Samples for three different images are represented. The point in the center of the circle is substituted by the median of all points in the circle, including itself. In the figure, the radius of the circle is equal to 0.85 LR pixels.

image, but directly in the LR samples, as the available information is much greater at that moment. The early erasing of non-stationary noise will also help to obtain a better resampled image, as most methods, including the projection method proposed in chapter 4, suppose stationary noise. In this way, we can handle both types of noise easily. We propose a median filter for irregular samples, as those coming from the registered LR images in SR, that will have as output locations the same irregular positions as the input samples. To obtain the selected input pixels for a given location, we will simply draw a circumference around that location and all the samples inside that circumference will be the input to the median filter. Note that this area is more regular than that of median filters applied to pixels in a grid, where the area has normally square shape. In Fig. 5.1 we can see an example where there are three registered LR images. The pixels from each image have a different shape to differentiate among them. The output of the median filter for the location at the center of the circumference in the figure will be the median of the seven LR pixels present in the drawn circle.

As can be easily seen, the radius of the circumference is an important design parameter. There is a trade-off between noise removal and smoothing: for a big radius, we will remove more outliers, but we also will make the output much smoother. In the other hand, a too small radius will make the final HR image sharper, but could also leave artifacts in it due to outliers.

5.3 Multi-Frame Deblurring

Deblurring is performed jointly with denoising and data fusion in many super-resolution approaches, but not in those that pose SR as an interpolation problem. Obviously, that includes the method that we propose in chapter 4. Also, all SR-as-interpolation methods cited in section 2.4.2 require a separate deblurring step. Among the most significative methods that require a final deblurring stage are Nguyen and Milanfar (2000), Lertrattanapanich and Bose (2002), and Pham et al. (2006). The model we will use for deblurring is

$$r(x,y) = h(x,y) * z(x,y) + n(x,y), \tag{5.1}$$

where $z(x,y)$ is the signal we want to estimate, $r(x,y)$ is the result of the SR interpolation step, $h(x,y)$ is the impulse response of a linear time-invariant (LTI) system, and $n(x,y)$ is some unknown additive noise, independent of $z(x,y)$.

When we suppose that we know the PSF, we can use a Wiener filter (see, for instance, Gonzalez and Woods (2007) or Pajares and de la Cruz (2008)) or the Lucy-Richardson algorithm (Richardson, 1972; Lucy, 1974) for deconvolution. In Farsiu et al. (2004b) a specific method for deblurring after interpolation with known PSF for SR is discussed. We will describe all these methods below. These will be also the methods that we will use for deblurring in the experiments performed in this chapter.

For the Wiener filter, we want to find an impulse response $g(x,y)$ such that

$$\hat{z}(x,y) = g(x,y) * r(x,y),$$

where $\hat{z}(x,y)$ is an estimate of $z(x,y)$ that minimizes the mean square error. The $g(x,y)$ that provides this estimate is the impulse response of the Wiener filter and its frequency response is calculated as

$$G(u,w) = \frac{1}{H(u,v)} \frac{|H(u,v)|^2}{|H(u,v)|^2 + \frac{N(u,v)}{Z(u,v)}} = \frac{1}{H(u,v)} \frac{|H(u,v)|^2}{|H(u,v)|^2 + \frac{1}{SNR(u,v)}}.$$

Here, $\frac{1}{H(u,v)}$ is the inverse of the original system, and $SNR(u,v) = \frac{Z(u,v)}{N(u,v)}$ is the signal-to-noise ratio. When there is zero noise (i.e. infinite signal-to-noise), the term inside the square brackets equals 1, which means that the Wiener filter is simply the inverse of the

system, as we might expect. However, as the noise at certain frequencies increases, the signal-to-noise ratio decreases, so the term inside the square brackets also drops. This means that the Wiener filter attenuates frequencies dependent on their signal-to-noise ratio.

The Lucy-Richardson algorithm, also known as Lucy-Richardson deconvolution, is an iterative procedure for recovering a latent image that has been blurred by a known point spread function.

We will discretize (5.1) to describe this algorithm. Pixels in the observed image can be represented in terms of the point spread function and the latent image as

$$r[i,j] = \sum_{m,n} h[i-m, j-n] \cdot z[m,n].$$

The statistics are performed under the assumption that the samples $z[i,j]$ are Poisson distributed, which is appropriate for photon noise in the data (Shepp and Vardi, 1982).

The basic idea is to calculate the most likely $z[i,j]$ given the observed $r[i,j]$ and known $h[i,j]$. This leads to an equation for $z[i,j]$ which can be solved iteratively according to

$$z[i,j]^{(t+1)} = z[i,j]^{(t)} \sum_{m,n} \frac{r[m,n]}{c[m,n]} \cdot h[m-i, n-j],$$

where

$$c[i,j] = \sum_{m,n} h[i-m, j-n] \cdot z[m,n]^{(t)}.$$

It has been shown empirically (Shepp and Vardi, 1982) that if this iteration converges, it converges to the maximum likelihood solution for $z[i,j]$.

In problems where the point spread function $h[i,j]$ is dependent on one or more unknown parameters, the Richardson-Lucy algorithm cannot be used. A later and more general class of algorithms, the expectation-maximization algorithms can however be applied to this type of problem.

The method described in Farsiu et al. (2004b) uses the additional data available when we have multiple frames for deblurring. The expression

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}} \left(\|\mathbf{A}(\mathbf{H}\mathbf{z} - \mathbf{r})\|_{l_p}^p + \lambda \Omega(\mathbf{z}) \right)$$

formulates the minimization criterion for obtaining $\hat{\mathbf{z}}$ from $\mathbf{r}$. In this case we have trans-

formed z and r to vectors formed by its samples, as in (2.10). Also as in chapter 1, $\mathbf{H}$ is a matrix that performs the LTI blurring, whereas l_p is the chosen norm and $\lambda\Omega(\mathbf{z})$ is a regularization term depending on $\mathbf{z}$, being Ω the regularization function and λ the regularization constant.

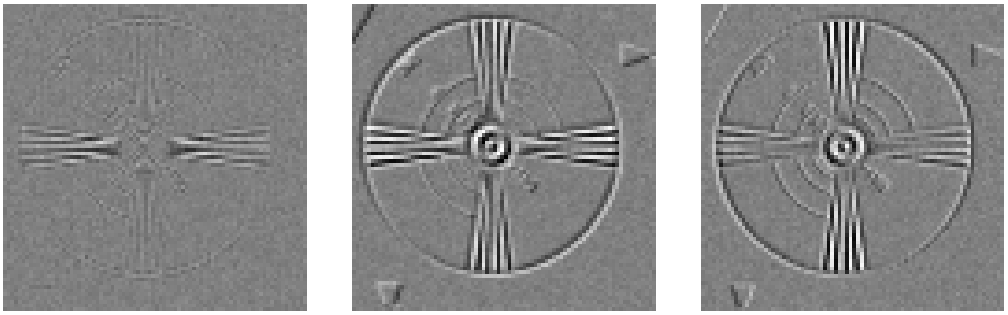
Matrix $\mathbf{A}$ is a diagonal matrix with diagonal values equal to the square root of the number of measurements that contributed to make each element of $\mathbf{r}$ (in the square case is the identity matrix). So, the undefined pixels of $\mathbf{r}$ have no effect on the HR estimate $\hat{\mathbf{z}}$. On the other hand, those pixels of $\mathbf{r}$ which have been produced from numerous measurements, have a stronger effect in the estimation of the HR frame $\hat{\mathbf{z}}$. The main contribution of this method is that we partially use the extra information that the fact of having multiple frames gives us. This is done in the weighting performed by the elements in the diagonal of matrix $\mathbf{A}$.

5.4 Performance Analysis of Robust Noise Removal Methods

In this section we will compare the proposed median filter for irregular samples with different robust super-resolution methods. We will apply our filter and then use the projection method proposed in chapter 4 for resampling in a rectangular grid. We will make experiments with synthetic and real data. The methods we will compare with are:

- The plain projective method proposed in the thesis.
- The robust SR method proposed in Farsiu et al. (2004b), which is based on using the l_1 norm in equation (2.9).
- The robust SR method proposed in Zomet et al. (2001), which is based on using the median of the gradient of the LR images to estimate the gradient used for calculating the next estimation when solving iteratively eq. (2.9).

The first experiment will use the same data set used in chapter 4. This data set was generated using a capture of a chart usually employed to measure the resolution of cameras, as in chapter 4, but generating just 10 low resolution images instead of 20. Firstly, the original image is blurred with a Gaussian kernel with variance $\sigma^2 = 1$ and then the LR images are generated using random shifts and downsampling the original by four. Finally, white Gaussian noise is added to the LR images until achieving a PSNR of 17 dB. To introduce outliers, we will add errors in the registering step for two of the



(a) (b) (c)

Figure 5.2: (a) An example difference image when there is no registration error, (b) difference image for error $(1,1)^T$, and (c) difference image for error $(-1,0.5)^T$.

images. The vector errors are $(1,1)^T$ and $(-1,0.5)^T$, measured in LR pixels. We can see the effect of these errors in the difference images with regard to the reference image of Fig. 5.2.

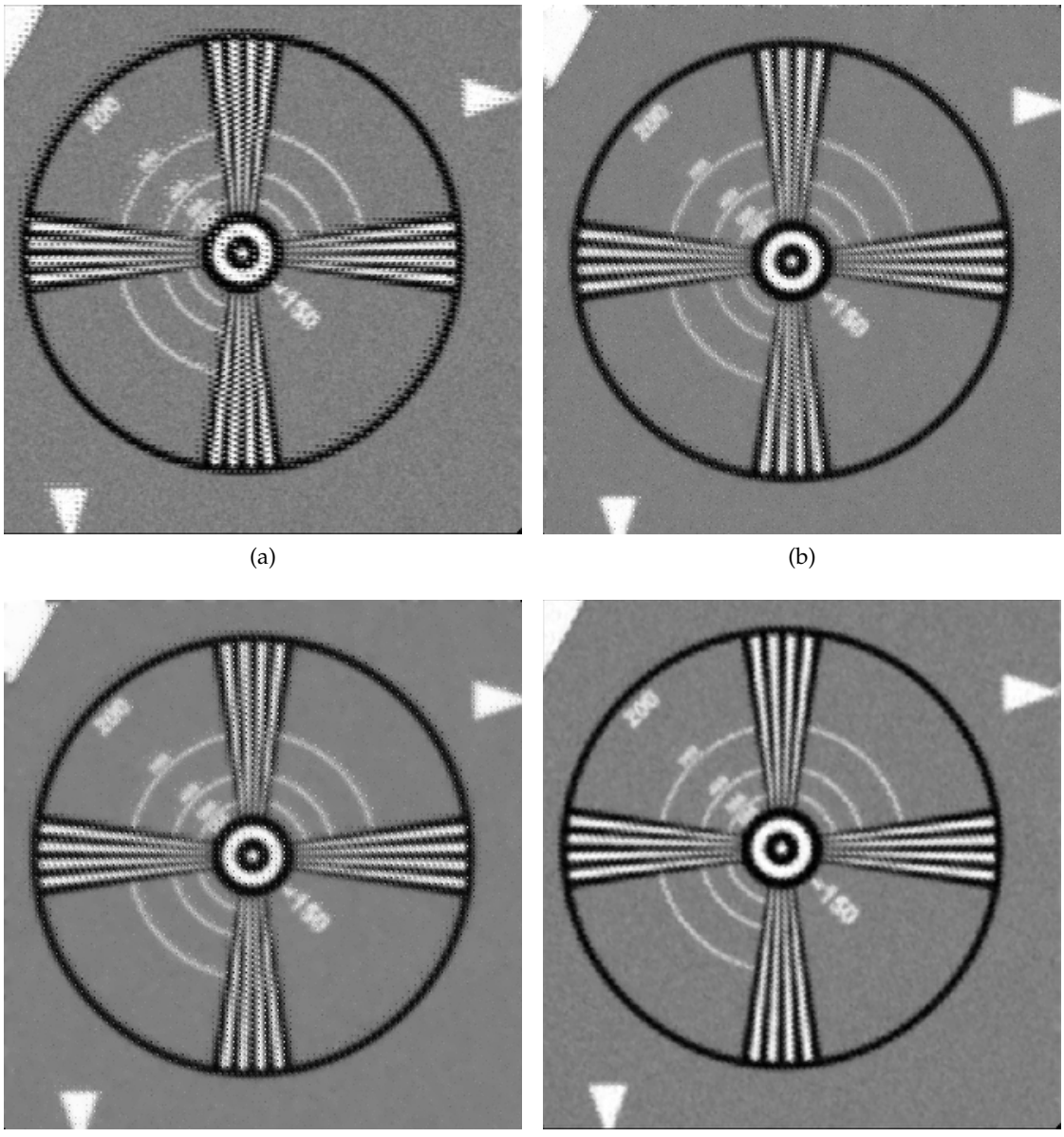
The results of applying the different methods to these data are shown in table 5.1 and in Fig. 5.3. The table shows the parameters needed for the different methods, which are tuned to maximize the PSNR when compared to the original image, and the final PSNR. The proposed method achieves the greatest PSNR. Besides this quantitative datum, we can also compare the visual appearance of the output of the different methods in Fig. 5.3. It is clearly seen that the only method that completely erases the influence of the outliers is the median filter for irregular samples plus projection method. The other robust methods obtain good results when compared to the projection method, but many outliers still remain.

Table 5.1: Parameter Values and Results for Robust SR Experiment One

Experiment One	λ	β	Iterations	R	PSNR (dB)
Projection	—	—	—	—	16.90
l_1 with bilateral TV	0.005	20	50	—	17.30
Zomet with bilateral TV	0.0005	20	50	—	17.33
Robust Projection	—	—	—	0.35	18.68

Table 5.2: Parameter Values and Results for Robust SR Experiment Two

Experiment One	λ	β	Iterations	R
Projection	—	—	—	—
l_1 with bilateral TV	0.01	5	50	—
Zomet with bilateral TV	0.0001	1	50	—
Robust Projection	—	—	—	0.25



(a) (b)

(c) (d)

Figure 5.3: Images produced by robust methods for resolution chart. (a) Our original projective method, (b) l_1 norm method with bilateral TV regularization, (c) the median gradient method of Zomet with bilateral TV regularization, and (d) the projective method with a previous median filter applied.

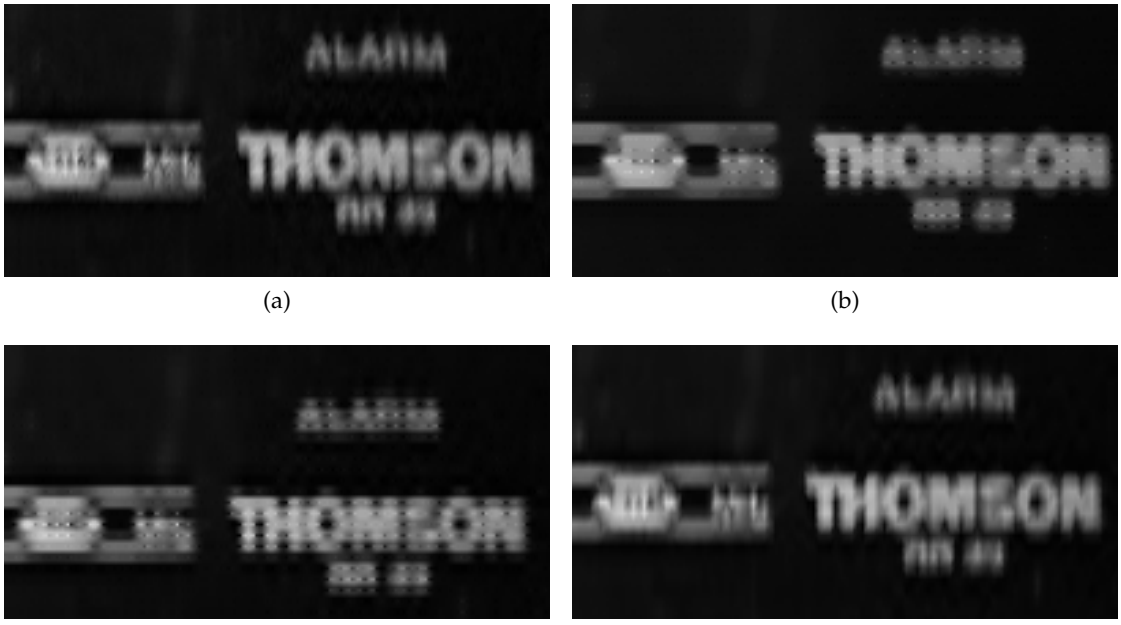
Figure 5.4: Images produced by robust methods for clock sequence. (a) Our original projective method, (b) l_1 norm method with bilateral TV regularization, (c) the median gradient method of Zomet with bilateral TV regularization, and (d) the projective method with a previous median filter applied.

The second experiment has been made using real data, in particular the clock sequence also used in chapter 4. We have used 20 LR images as well, but instead of applying an SR factor of two, we have multiplied by four the resolution to avoid the influence of aliasing in the HR image for methods without prefilter. We have tuned the different parameters to obtain the best visual results. The used parameters can be seen in table 5.2.

The results are shown in figures 5.4 and 5.5. The l_1 method and the Zomet method exhibit many artifacts in the edges of the images. The image obtained with the projective method is a bit noisy, as can be seen in the white figures at the left part of the image located in the line in the middle of the clock. The robust method that combines a median filter and our projective method erases all artifacts and is also less noisy than the image obtained just projecting. The price to pay is a bit more blurry image: this can be appreciated in the letters at the right of the image when we compare with Fig. 5.5a.

5.5 Performance Analysis of Deblurring Methods

In this section we perform two experiments comparing the l_2 norm method that jointly interpolates and deblurs the HR image proposed in Farsiu et al. (2004b) with the deblurring methods exposed in section 5.3.



(a) (b)

(c) (d)

Figure 5.5: Detail of images from robust methods for clock sequence. (a) Our original projective method, (b) l_1 norm method with bilateral TV regularization, (c) the median gradient method of Zomet with bilateral TV regularization, and (d) the projective method with a previous median filter applied.

The first experiment uses the chart sequence already employed in previous sections. In this case we have generated 20 low resolution images. Firstly, the original image is blurred with a Gaussian kernel with width parameter $\sigma^2 = 1$ and then the LR images are generated using random shifts and downsampling the original by four. Finally, white Gaussian noise is added to the LR images until achieving a PSNR of 17 dB. This corresponds to a variance $\sigma^2 = 0.01$, which we need to know for the Wiener filter. We have tuned the parameters for the different methods to maximize the PSNR in the final image. These parameters and the obtained PSNR can be seen in table 5.3. The best PSNR is achieved when applying the projection method and then the Lucy-Richardson algorithm. This demonstrates that solving separately the tasks needed to obtain SR does not necessarily lead to poorer results than a joint approach where all tasks are solved together, as the PSNR for the l_2 method that jointly resamples and deblurs is smaller than projection plus Lucy-Richardson deconvolution. Projection plus l_1 norm deconvolution with bilateral TV regularization also outperforms the l_2 norm method. Figure 5.6 shows the images obtained by the different methods. Visually all images are similar, but the one resulting from the Wiener filter, which is a bit blurry.

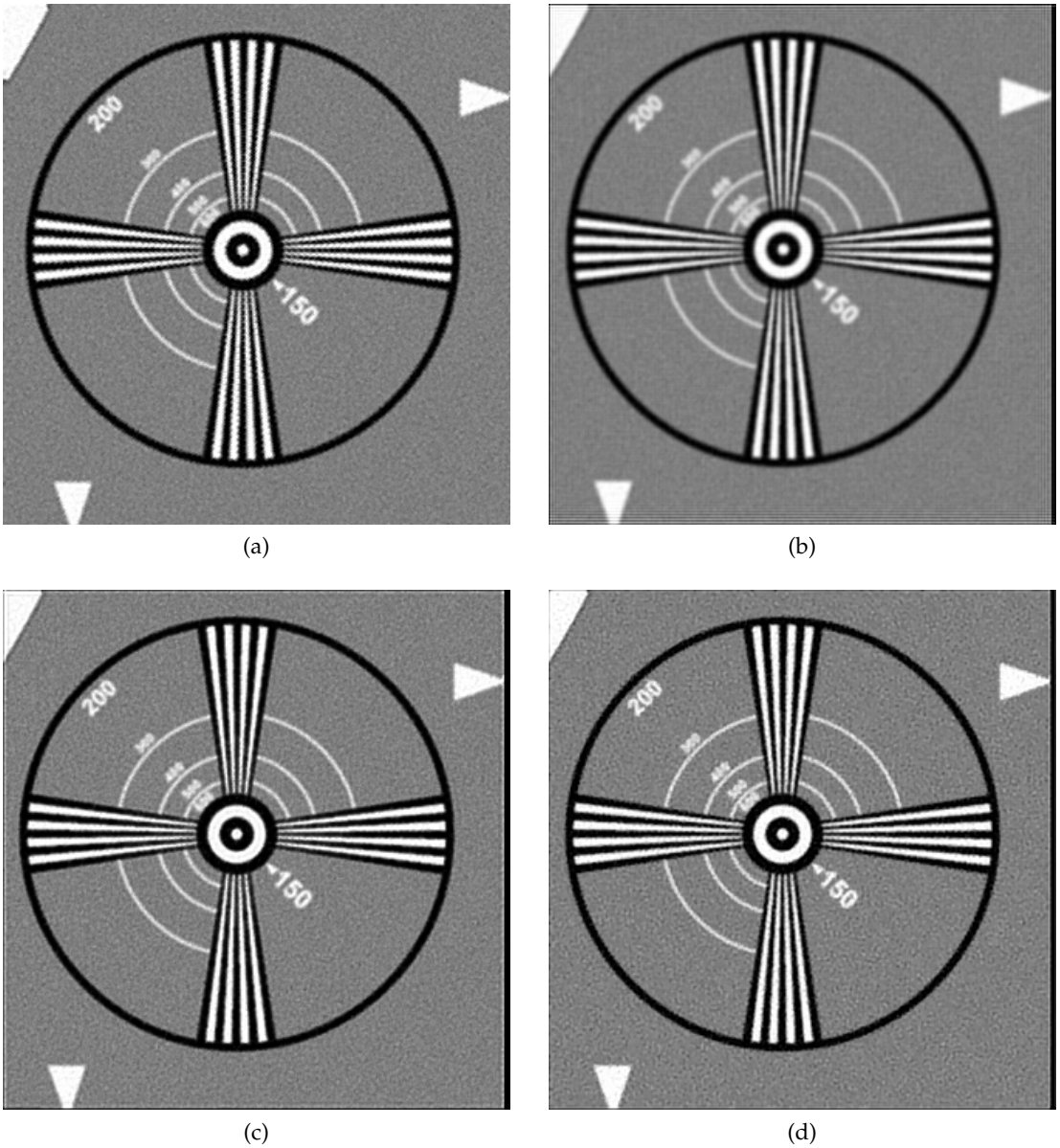


Figure 5.6: Images produced deblurring the chart sequence. (a) Deblurring and resampling performed jointly using the l_2 norm, (b) projection method plus Wiener deconvolution, (c) projection method plus Lucy-Richardson algorithm, and (d) projection method plus l_1 norm deconvolution with bilateral TV regularization.

Table 5.3: Parameter Values and Results for SR Deblurring Methods Experiment One

Experiment One	λ	β	Iterations	PSNR (dB)
l_2 with bilateral TV	0.01	4	20	21.63
Projection + Wiener	—	—	—	21.50
Projection + Lucy-Richardson	—	—	5	22.05
Projection + l_1 deconvolution with BTV	0.01	4	50	22.02

The second experiment uses the same set of real data as the second experiment carried out with the robust methods. We have used 20 LR images as well, but instead of applying an SR factor of two, we have multiplied by four the resolution to avoid the influence of aliasing in the HR image for methods without prefilter. We have tuned the different parameters to obtain the best visual results. The used parameters can be seen in table 5.4. We have also supposed a known PSF with Gaussian shape and width parameter $\sigma^2 = 1$. We have sampled this PSF in a 5x5 pixels grid to perform the deblurring. In the case of the Wiener filter, a noise of variance $\sigma^2 = 0.01$ gave us the best visual results.

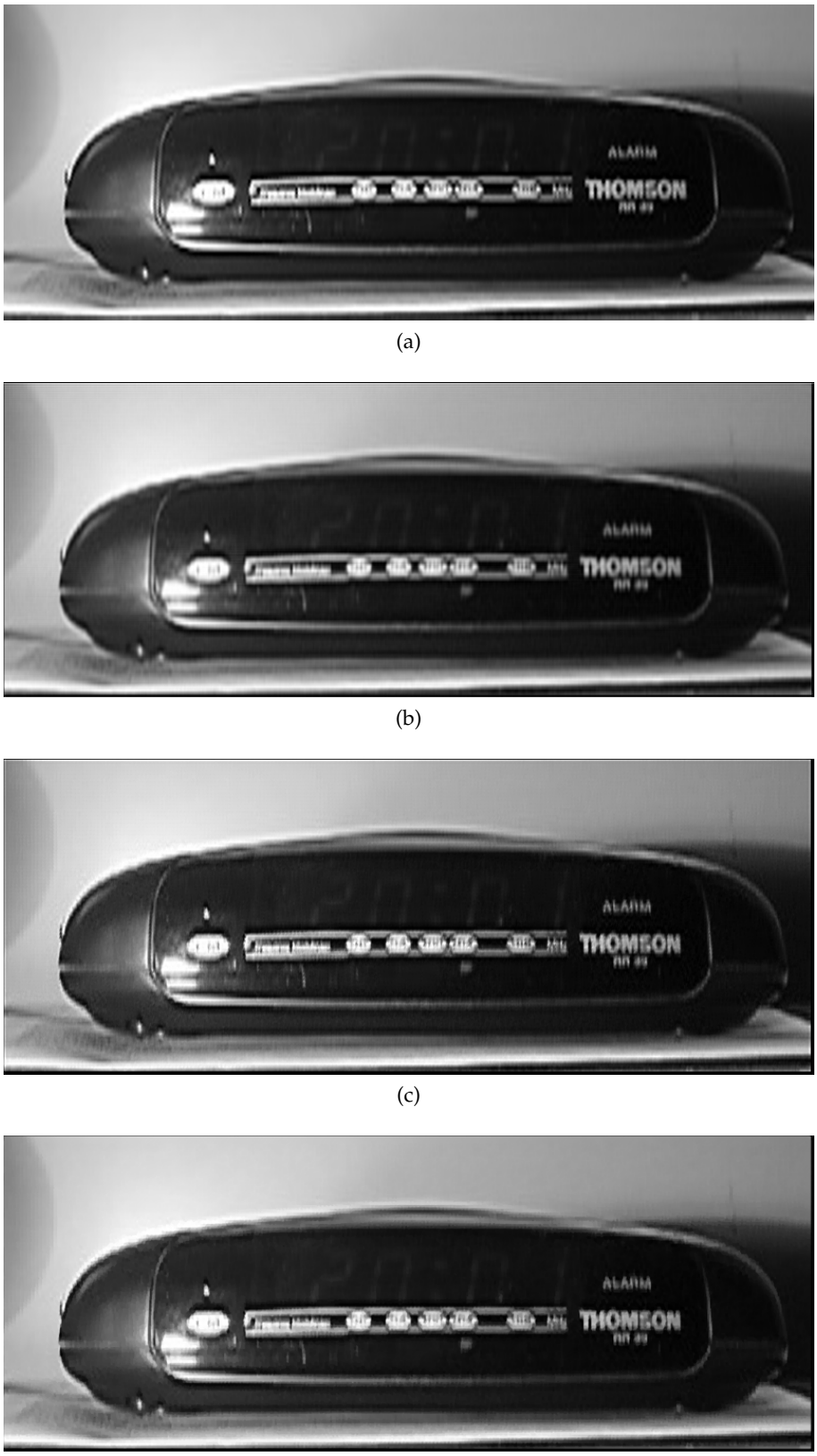
In figures 5.7 and 5.8 we can see the results. The best contrast is obtained again by the Lucy-Richardson algorithm, compared to the image obtained using the l_2 norm. The Wiener filtered image is a bit blurry, which can be seen in the letters that appear in the image, and the image with l_1 deconvolution shows some artifacts due to excessive regularization that appear as homogeneous areas in parts of the image.

Table 5.4: Parameter Values and Results for SR Deblurring Methods Experiment Two

Experiment Two	λ	β	Iterations
l_2 with bilateral TV	0.2	2	20
Projection + Wiener	—	—	—
Projection + Lucy-Richardson	—	—	3
Projection + l_1 deconvolution with BTV	0.05	2	50

5.6 Conclusions

In this chapter we have studied specific techniques for deblurring and denoising in SR. A novel robust denoising method for SR is proposed. The term "robust" refers to its capability of handling non-stationary noise as the one that appears when outliers are present or for salt and pepper noise. This method is based on a median filter that is applied to an irregular sampling that usually appears in SR. Its performance is checked



(a)

(b)

(c)

(d)

Figure 5.7: Images produced deblurring the clock sequence. (a) Deblurring and resampling performed jointly using the l_2 norm, (b) projection method plus Wiener deconvolution, (c) projection method plus Lucy-Richardson algorithm, and (d) projection method plus l_1 norm deconvolution with bilateral TV regularization.

(a)

(b)

(c)

(d)

Figure 5.8: Detail of images produced deblurring the clock sequence. (a) Deblurring and resampling performed jointly using the l_2 norm, (b) projection method plus Wiener deconvolution, (c) projection method plus Lucy-Richardson algorithm, and (d) projection method plus l_1 norm deconvolution with bilateral TV regularization.

against other robust methods for SR, outperforming them in the experiments that have been carried out.

We have also studied deblurring methods for SR. We have compared in the experiments the joint resampling and deblurring that is performed in most SR methods with a separated approach that uses our projective method plus classical deconvolution methods or a deconvolution-after-resampling method that takes advantage of multi-frame data. We have shown that the separated approach is able to outperform the joint approach even when there is no danger of aliasing artifacts in the HR image.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

We will summarize in this chapter the contributions of our work to the SR problem and the conclusions that can be drawn from them. We have followed throughout this dissertation a stage-based approach where all super-resolution tasks are solved separately but using all the available information at a given time. We have shown that this approach does not necessarily lead to poorer results than an SR as a whole approach and that it has important advantages. These advantages are:

- The solution to each stage can be tailored more easily to the posed problem. For instance, it is difficult to apply a prefiltering scheme to avoid aliasing in the reconstructed image if we have to make this jointly with deblurring.
- Performance can be much better than in joint approaches, where the degrees of freedom are many more and the complexity of the problem grows exponentially with them.
- Joint SR methods can be optimal for a given HR size to be reconstructed, but we can have the paradoxical situation in which the registration parameters for a given HR size are different from those for another reconstruction size. We have avoided this in the proposed registration method.

In the next sections we will break down the contributions to different SR stages made in this research.

6.1.1 Multi-Frame Registration

As we have already stated, one of the main problems for practical super-resolution is the precision in the registration. In chapter 3 we have proposed a new multi-frame registration method that outperforms plain optical flow registration and also competing approaches. The main novelty of this method resides on the per-pixel weighting that is made when calculating the registration penalty function. This makes it work with any motion model. In the performed experiments, we have shown a remarkable increase for registration precision under translations and under rigid motion (rotations plus translations). It is also demonstrated that other multi-frame registration methods cannot handle properly motion models different from simple shifts, whereas ours is able to do so. We have also shown the positive effects this registration method has in the visual aspect of the HR image, when applied to an SR problem. An important characteristic of the proposed method is that with 4 to 5 images in the sequence we can obtain a very significant improvement in the registration accuracy.

6.1.2 Fusion of Data Using Projections on Functional Spaces

The main contribution of chapter 4 is the discovery of flaws in the usually linear model assumed in most popular SR methods. We show that this model implicitly assumes a band-limited reconstructed image. This assumption is not always true, as we prove with different experiments. This leads to aliasing artifacts in the HR image. To avoid this problem, we have proposed a method that employs standard sampling theory, with a prefiltering step that erases aliasing in the reconstructed image. This is equivalent to orthogonally project the image to the desired basis. To be able to do this projection we first transform the sampling points to a continuous function through a Delaunay triangulation. Although we could use any base for the projection, we chose B-splines because of their compact support and their closed expression as piecewise polynomials.

The proposed method has also the important advantage of being non-iterative, which makes it scale well. It also removes all noise in the high frequency part of the spectrum, being effective against different types of noise, provided that the noise is stationary. Another remarkable feature of the method is that is highly parallelizable once we have the Delaunay triangulation.

6.1.3 Super-Resolution Restoration

As we have already exposed, the projection method of chapter 4 is effective against stationary noise, but not for non-stationary noise like salt and pepper noise or the one provoked by the presence of outliers. In chapter 5 a non-linear filter for irregular samples is presented. This filter applies the median operator, which has been widely used in imaging for reducing non-linear noise, to the sampled within a circle drawn around a certain sample. The result of the filter substitutes the value of the sample at the center of the circle. We have proved that this simple filter that is applied to the samples before starting the SR restoration is able to outperform all so-called robust SR methods against which we have tested it.

Additionally, in chapter 5 we also apply standard deblurring methods to the images obtained by the interpolation method of chapter 4. We prove that with this combination we outperform standard SR methods that perform jointly data fusion and deblurring. This result supports our supposition that states that a stage-based SR can outperform the joint approach when proper methods that take advantage of all available information at each stage are applied.

6.2 Future Work

The ways in which SR can be enhanced are multiple. Although there has already been a lot of research in the field, SR is still an open problem. The applications to which it has not been widely applied are many, because of the practical problems one finds when the theory is applied.

Some interesting ideas that deserve to be considered for future research in SR are:

- The proposed multi-frame registration method is efficient, but not optimal. The presence of unavoidable aliasing in the images to be registered increases the error in the estimation of the movement parameters. A method that allows multi-frame registering and removes the aliasing using all frames would be needed to achieve optimal results.
- The projection method has proved to be effective with black and white images, but an extension to color images would make it more useful. This should not be done just applying the algorithm to the different color planes, but using the corre-

lation among the different color channels as a priori knowledge. If the images are mosaicked, the demosaicking should be done simultaneously with the projection.

- SR is of most interest for cases where the resolution of the imaging system is low, the noise is high, or the quality is bad in general. An interesting area of application for SR would be for cheap handsets like cheap photographic cameras or mobile phones. The videos taken with these systems show many artifacts and degradations of the image. The images are compressed with MJPEG, H.263 or other standards, which introduces an additional challenge. SR of compressed color video is, therefore, an interesting are for further research.
- The projection method could also be improved incorporating a priori knowledge in the solution of the system. This knowledge is especially important when we have just a few frames available. We could use bilateral TV or edge preserving priors for this.
- Another interesting area of research in SR is the increase not only of spatial resolution, but also of temporal resolution. This means increasing or changing the frames per second (fps) of a given video sequence. This has, for instance, applications to video rate conversions. We could think of an extension of the projection method to address this problem. To do so, we would have to calculate intermediate motion vectors between two HR images to find the adequate position for the sampling points in the new frame to be created.
- The proposed median filter for irregular samples is able to correct the presence of outliers, but sometimes removes some of the detail in the images. An interesting extension would be to make it adaptive to the shape of the filtered image. The irregular nature of the samples makes this a challenging extension. Weighted median filters, like those described in Yin et al. (1996), could be used for this task.
- Due to the imprecision inherent to registration (Pham et al., 2005a), it is very difficult to register images to which a global motion model cannot be applied. Thus, the application of SR to general video sequences that require a general optical flow motion model for registration is currently impractical. This is an almost completely unexplored area for investigation, although some efforts have already been made (Baker and Kanade, 1999; Fransens et al., 2004; Protter et al., 2008).

- Another practical possible application of SR could be to images captured by Unmanned Aerial Vehicles (UAV). Due to the nature of these vehicles they are equipped with low resolution video systems to avoid overloading them and compromise their maneuverability.
- An idea that has not received much attention is to apply SR to sequences with motion blur, a degradation that can easily appear in video sequences. This problem is very challenging as normally it happens not globally in the frames but for objects that move inside the scene. Therefore, these objects must be tracked so we can remove then the motion blur. This has a link with general optical flow SR, as that would be the motion model that has to be employed. An early approach to this problem can be found in Bascle et al. (1996).

Bibliography

Adams, J., Parulski, K., and Spaulding, K. Color processing in digital cameras. *IEEE MICRO*, pp. 20–30, 1998.

Ahuja, N., Lertrattanapanich, S., and Bose, N. K. Properties determining choice of mother wavelet. *IEE Proceedings–Vision, Image and Signal Processing*, vol. 152, no. 5, pp. 659–664, 2005.

Ahuja, N. A. and Bose, N. K. Multidimensional generalized sampling theorem for wavelet based image superresolution. In *IEEE International Conference on Image Processing (ICIP)*, pp. 1589–1592. 2006.

Baker, S. and Kanade, T. Super-resolution optical flow. Tech. Rep. CMU-RI-TR-99-36, Robotics Institute, Carnegie Mellon University, October 1999.

———. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1167–1183, 2002.

Bascle, B., Blake, A., and Zisserman, A. Motion deblurring and super-resolution from an image sequence. In *Proceedings of the 4th European Conference on Computer Vision*, vol. 2, pp. 573–582. Springer-Verlag, London, UK, 1996.

Bergen, J. R., Anandan, P., Hanna, K. J., and Hingorani, R. Hierarchical model-based motion estimation. In *European Conference on Computer Vision*, vol. 588, pp. 237–252. 1992.

Borman, S. and Stevenson, R. L. Super-resolution from image sequences–a review. In *Proceedings of the Midwest Symposium on Circuits and Systems*, pp. 374–378. 1998.

———. Linear models for multiframe super-resolution restoration under nonaffine registration and spatially varying psf. In *Proceedings of SPIE: Computational Imaging II*, vol. 5299, pp. 234–245. San Jose, CA, 2004.

Bose, N. K., Kim, H. C., and Valenzuela, H. M. Recursive total least squares algorithm for image reconstruction from noisy, undersampled frames. *Multidimensional Systems and Signal Processing*, vol. 4, no. 3, pp. 253–268, 1993.

Bovik, A. C. and Gibson, J. D. *Handbook of Image and Video Processing*. Academic Press, Inc., Orlando, FL, USA, 2000.

Brown Jr, J. Multi-channel sampling of low-pass signals. *IEEE Transactions on Circuits and Systems*, vol. 28, no. 2, pp. 101–106, 1981.

Burt, P. and Adelson, E. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, 1983.

Capel, D. and Zisserman, A. Automated mosaicing with super-resolution zoom. In *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 885–891. 1998.

Cheeseman, P., Kanefsky, B., Kraft, R., Stutz, J., and Hanson, R. Super-resolved surface reconstruction from multiple images. In G. R. Heidbreder (ed.), *Maximum Entropy and Bayesian Methods*, pp. 293–308. Kluwer Academic Publishers, Dordrecht, the Netherlands, 1996.

Chiang, M.-C. and Boult, T. E. Local blur estimation and super-resolution. In *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 821–826. 1997.

Davies, A. and Fennessy, P. *Digital imaging for photographers*. Focal Press, fourth edn., 2001.

Davis, J. Mosaics of scenes with moving objects. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 354–360. 1998.

Elad, M. and Feuer, A. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1646–1658, 1997.

Elad, M. and Hel-Or, Y. A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur. *IEEE Transactions on Image Processing*, vol. 10, no. 8, pp. 1187–1193, 2001.

Farsiu, S., Elad, M., and Milanfar, P. Constrained, globally optimal, multi-frame motion estimation. In *IEEE/SP 13th Workshop on Statistical Signal Processing*, pp. 1396–1401. July 2005.

———. Multiframe demosaicing and super-resolution of color images. *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 141–159, 2006a.

———. Video-to-video dynamic super-resolution for grayscale and color sequences. *EURASIP Journal on Applied Signal Processing*, vol. 2006, no. 1, pp. 232–232, 2006b.

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

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

Fransens, R., Strecha, C., and Van Gool, L. A probabilistic approach to optical flow based super-resolution. In *Proceedings of Conference on Computer Vision and Pattern Recognition Workshop*, vol. 12. 2004.

Gonzalez, R. C. and Woods, R. E. *Digital Image Processing*. Prentice Hall, 2007.

Govindu, V. M. Lie-algebraic averaging for globally consistent motion estimation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1. 2004.

Greenspan, H., Oz, G., Kiryati, N., and Peled, S. MRI inter-slice reconstruction using super-resolution. *Magnetic Resonance Imaging*, vol. 20, no. 5, pp. 437–446, 2002.

Hardie, R. C., Barnard, K. J., and Armstrong, E. E. Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1621–1633, 1997.

Hartley, R. and Zisserman, A. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.

He, Y., Yap, K. H., Chen, L., and Chau, L. P. A nonlinear least square technique for simultaneous image registration and super-resolution. *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2830–2841, 2007.

Horn, B. K. P. and Schunck, B. G. Determining optical flow. *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

Hunt, B. R. Quantifying the super-resolution capabilities of the CLEAN image processing algorithm. In *Proceedings of SPIE*, vol. 300. SPIE, 2004.

Irani, M. and Peleg, S. Improving resolution by image registration. *CVGIP: Graph. Models Image Process.*, vol. 53, no. 3, pp. 231–239, 1991.

———. Motion analysis for image enhancement: Resolution, occlusion, and transparency. *Journal of Visual Communication and Image Representation*, vol. 4, no. 4, pp. 324–335, 1993.

Kang, E., Cohen, I., and Medioni, G. A graph-based global registration for 2d mosaics. In *Proceedings of the 15th International Conference on Pattern Recognition*. 2000.

Kennedy, J. A., Israel, O., Frenkel, A., Bar-Shalom, R., and Azhari, H. Super-resolution in PET imaging. *IEEE Transactions on Medical Imaging*, vol. 25, no. 2, pp. 137–147, 2006.

Kim, S. P., Bose, N. K., and Valenzuela, H. M. Recursive reconstruction of high resolution image from noisy undersampled multiframes. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 6, pp. 1013–1027, 1990.

Kim, S. P. and Su, W. Y. Recursive high-resolution reconstruction of blurred multiframe images. *IEEE Transactions on Image Processing*, vol. 2, no. 4, pp. 534–539, 1993.

Lenz, R. and Lenz, U. Calibration of a color CCD camera with 3000x2300 picture elements. In *Close-range photogrammetry meets machine vision*, vol. 1395, pp. 104–111. 1990.

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

Lin, Z. and Shum, H. Y. Fundamental limits of reconstruction-based superresolution algorithms under local translation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 83–97, 2004.

Lucas, B. D. and Kanade, T. An iterative image registration technique with an application to stereo vision. *International Joint Conference on Artificial Intelligence*, vol. 81, pp. 674–679, 1981.

Lucy, L. B. An iterative technique for the rectification of observed distributions. *The Astronomical Journal*, vol. 79, no. 6, pp. 745–754, 1974.

Marquardt, D. W. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

McGuire, M. An image registration technique for recovering rotation, scale and translation parameters. Tech. Rep. TR 98–018, NEC Research Institute, 1998.

Merino, M. T. and Núñez, J. Super-resolution of remotely sensed images with variable-pixel linear reconstruction. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 5, pp. 1446–1457, 2007.

MotionDSP. Ikena reveal. http://www.motiondsp.com/, 2008.

Nguyen, N. and Milanfar, P. A wavelet-based interpolation-restoration method for superresolution (wavelet superresolution). *Circuits, Systems, and Signal Processing*, vol. 19, no. 4, pp. 321–338, 2000.

Nguyen, N., Milanfar, P., and Golub, G. A computationally efficient superresolution image reconstruction algorithm. *IEEE Transactions on Image Processing*, vol. 10, no. 4, pp. 573–583, 2001a.

———. Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement. *IEEE Transactions on Image Processing*, vol. 10, no. 9, pp. 1299–1308, 2001b.

Pajares, G. and de la Cruz, J. M. *Visión por Computador: Imágenes digitales y aplicaciones*. Ra–Ma, second edn., 2008. In Spanish.

Papoulis, A. Generalized sampling expansion. *IEEE Transactions on Circuits and Systems*, vol. 24, no. 11, pp. 652–654, 1977.

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

Patti, A. J., Sezan, M. I., and Murat Tekalp, A. Superresolution video reconstruction with arbitrary sampling lattices and non-zero aperture time. *IEEE Transactions on Image Processing*, vol. 6, no. 8, pp. 1064–1076, 1997.

Patti, A. J., Sezan, M. I., and Tekalp, A. M. High resolution standards conversion of low resolution video. In *International Conference on Acoustics, Speech, and Signal Processing*, vol. 4. 1995.

Pham, T. Q., Bezuijen, M., van Vliet, L. J., Schutte, K., and Hendriks, C. L. L. Performance of optimal registration estimators. In Z. ur Rahman, R. A. Schowengerdt, and S. E. Reichenbach (eds.), *Proceedings of the SPIE*, vol. 5817, pp. 133–144. Orlando, FL, USA, 2005a.

Pham, T. Q., van Vliet, L. J., and Schutte, K. Influence of signal-to-noise ratio and point spread function on limits of superresolution. In *Proceedings of the SPIE*, vol. 5672, pp. 169–180. San Jose, CA, USA, 2005b.

———. Robust fusion of irregularly sampled data using adaptive normalized convolution. *EURASIP Journal on Applied Signal Processing*, vol. 2006, no. 1, pp. 236–236, 2006.

Pickup, L. C., Capel, D. P., Roberts, S. J., and Zisserman, A. Overcoming registration uncertainty in image super-resolution: maximize or marginalize? *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 2, pp. 20–20, 2007.

Powell, M. J. D. The theory of radial basis function approximation in 1990. In W. Light (ed.), *Advances in numerical analysis II: Wavelets, subdivision algorithms, and radial basis functions, Proc. 4th Summer Sch., Lancaster/UK*, pp. 105–210. Oxford University Press, 1990.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. *Numerical Recipes in C*. Cambridge University Press, 1992.

Protter, M., Elad, M., Takeda, H., and Milanfar, P. Generalizing the non-local-means to super-resolution reconstruction. *IEEE Transactions on Image Processing*, 2008. Accepted for publication.

Richardson, W. H. Bayesian-based iterative method of image restoration. *Journal of the Optical Society of America*, vol. 62, no. 1, pp. 55–59, 1972.

Robertson, M. A. and Stevenson, R. L. Temporal resolution enhancement in compressed video sequences. *EURASIP Journal on Applied Signal Processing*, vol. 2001, no. 4, pp. 230–238, 2001.

Robinson, D., Farsiu, S., and Milanfar, P. Optimal registration of aliased images using variable projection with applications to super-resolution. *The Computer Journal*, 2007.

Robinson, D. and Milanfar, P. Fundamental performance limits in image registration. *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1185–1199, 2004.

———. Statistical performance analysis of super-resolution. *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1413–1428, 2006.

Rochefort, G., Champagnat, F., Le Besnerais, G., and Giovannelli, J. F. An improved observation model for super-resolution under affine motion. *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3325–3337, 2006.

Rudin, L. I., Osher, S., and Fatemi, E. Nonlinear total variation based noise removal algorithms. *Physica D*, vol. 60, pp. 259–268, 1992.

Sánchez-Beato, A. and Pajares, G. Noniterative interpolation-based super-resolution minimizing aliasing in the reconstructed image. *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1817–1826, October 2008.

Schatzberg, A. and Devaney, A. J. Super-resolution in diffraction tomography. *Inverse Problems*, vol. 8, pp. 149–164, 1992.

Schultz, R. and Stevenson, R. Extraction of high-resolution frames from video sequences. *IEEE Transactions on Image Processing*, vol. 5, no. 6, pp. 996–1011, 1996.

Segall, C. A., Molina, R., Katsaggelos, A. K., and Mateos, J. Bayesian high-resolution reconstruction of low-resolution compressed video. In *Proceedings International Conference on Image Processing*, vol. 2. 2001.

Shechtman, E., Caspi, Y., and Irani, M. Increasing space-time resolution in video. In *Proceedings of the 7th European Conference on Computer Vision-Part I*, pp. 753–768. Springer, May 2002.

Shekarforoush, H. and Chellappa, R. Data-driven multichannel superresolution with application to video sequences. *Journal of the Optical Society of America A*, vol. 16, no. 3, pp. 481–492, 1999.

Shepp, L. A. and Vardi, Y. Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, vol. 1, no. 2, pp. 113–122, 1982.

Sheppard, D., Hunt, B. R., and Marcellin, M. W. Super-resolution of imagery acquired through turbulent atmosphere. In *Conference Record of the Thirtieth Asilomar Conference on Signals, Systems and Computers*, vol. 1. 1996.

Shum, H. Y. and Szeliski, R. Construction and refinement of panoramic mosaics with global and local alignment. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 953–958. 1998.

Sroubek, F., Cristobal, G., and Flusser, J. A unified approach to superresolution and multichannel blind deconvolution. *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2322–2332, 2007.

Stark, H. and Oskoui, P. High-resolution image recovery from image-plane arrays, using convex projections. *Journal of the Optical Society of America A*, vol. 6, pp. 1715–1726, 1989.

Tekalp, A. M., Ozkan, M. K., and Sezan, M. I. High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, pp. 169–172. March 1992.

The MathWorks. Matlab. http://www.mathworks.com, 2008.

Tikhonov, A. N. Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.*, vol. 4, no. 4, pp. 1035–1038, 1963.

Tipping, M. E. and Bishop, C. M. *Advances in Neural Information Processing Systems*, vol. 15, chap. Bayesian image super-resolution, pp. 1279–1286. MIT Press, Cambridge, MA, 2003.

Tomasi, C. and Manduchi, R. Bilateral filtering for gray and color images. In *Proceedings of the Sixth International Conference on Computer Vision*. 1998.

Triggs, B. and Sdika, M. Boundary conditions for Young-van Vliet recursive filtering. *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2365–2367, June 2006.

Tsai, R. Y. and Huang, T. S. Multiframe image restoration and registration. *Advances in Computer Vision and Image Processing*, vol. 1, no. 2, pp. 317–339, 1984.

Unnikrishnan, R. and Kelly, A. A constrained optimization approach to globally consistent mapping. In *IEEE/RSJ International Conference on Intelligent Robots and System*, vol. 1. 2002.

Unser, M. Splines: a perfect fit for signal and image processing. *IEEE Signal Processing Magazine*, vol. 16, no. 6, pp. 22–38, 1999.

———. Sampling–50 years after Shannon. *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, 2000.

Unser, M., Aldroubi, A., and Eden, M. Fast b-spline transforms for continuous image representation and interpolation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 3, pp. 277–285, 1991.

———. Polynomial spline signal approximations: filter design and asymptotic equivalence with Shannon's sampling theorem. *IEEE Transactions on Information Theory*, vol. 38, no. 1, pp. 95–103, 1992.

Ur, H. and Gross, D. Improved resolution from subpixel shifted pictures. *CVGIP: Graphical Models and Image Processing*, vol. 54, no. 2, pp. 181–186, 1992.

Vandewalle, P., Susstrunk, S., and Vetterli, M. A frequency domain approach to registration of aliased images with application to super-resolution. *EURASIP Journal on*

Applied Signal Processing, Special Issue on Super-Resolution Imaging, vol. 2006, pp. 1–14, 2006.

Verdú-Monedero, R., Larrey-Ruiz, J., and Morales-Sánchez, J. Frequency implementation of the Euler-Lagrange equations for variational image registration. *IEEE Signal Processing Letters*, vol. 15, pp. 321–324, 2008.

Wiener, N. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*. MIT Press, Cambridge, Mass., 1949.

Wilson, T. and Hewlett, S. J. Superresolution in confocal scanning microscopy. *Optical Letters*, vol. 16, no. 14, pp. 1062–1064, 1991.

Yin, L., Yang, R., Gabbouj, M., and Neuvo, Y. Weighted median filters: a tutorial. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 43, no. 3, pp. 157–192, March 1996.

Young, I. T. and van Vliet, L. J. Recursive implementation of the Gaussian filter. *Signal Processing*, vol. 44, no. 2, pp. 139–151, 1995.

Zomet, A., Rav-Acha, A., and Peleg, S. Robust super-resolution. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1. 2001.

Index

a priori knowledge, 38, 41, 56
aliasing, 29, 38, 90
analysis function, 77
anti-aliasing prefilter, 74
astronomical imaging, 33

B-splines, 78, 82
bilateral filter, 84
biorthogonality condition, 77
blur, 28
BTV, *see* regularization, Bilateral Total Variation (BTV)
bundle adjustment, 43, 55

CCD, *see* Charged-Coupled Devices
Charged-Coupled Devices (CCD), 27
CMOS, *see* Complementary Metal Oxide Semiconductor
Complementary Metal Oxide Semiconductor (CMOS), 27
compression artifacts, 33, 91
Computed Tomography (CT), 27, 39
Cramer-Rao lower bound, 43, 56
CT, *see* Computed Tomography

Delaunay triangulation, 44, 73, 78, 88
diffraction limit, 33

FLIR, *see* Forward Looking Infrared
Forward Looking Infrared (FLIR), 27, 33
Fourier transform, 38, 73

Gaussian noise, 30, 41, 75, 107
Gaussian PSF, 76, 90
generalized sampling theorem, 44, 73

hierarchical motion estimation, 59, 61
high resolution images, 27, 40
HR images, *see* high resolution images

IBP, *see* Iterative Back-Projection
IIR filter, *see* Infinite Impulse Response filter
ill-posed problem, 41, 83
image
  band-limited, 38, 75
  blind deblurring, 30, 44
  deblurring, 30, 44, 107
  denoising, 30, 107
  feature-based registration, 43, 56
  interpolation, 30, 35, 42
  mosaics, 55
  multi-frame deblurring, 110
  multi-frame registration, 34, 43, 56
  registration, 29, 34, 42

restoration, 30, 35, 42, 107
Infinite Impulse Response filter, 82
Iterative Back-Projection (IBP), 39

Joint Photographic Experts Group (JPEG), 33
JPEG, *see* Joint Photographic Experts Group

l_1 norm, 41, 85
l_2 norm, 41, 85
Laplacian noise, 41
Levenberg-Marquardt iteration, 61
Linear Shift-Invariant (LSI), 30, 38
Linear Time-Invariant (LTI), 110
low resolution images, 27, 40
LR images, *see* low resolution images
LSI, *see* Linear Shift-Invariant
LTI, *see* Linear Time-Invariant
Lucy-Richardson algorithm, 44, 108

Magnetic Resonance Imaging (MRI), 33
MAP, *see* Maximum a Posteriori
Maximum a Posteriori (MAP), 40, 85
Maximum Likelihood (ML), 40, 85
MDSP software package, 89
Mean Squared Error (MSE), 65
median filter, 108
    for irregular samples, 108, 112
ML, *see* Maximum Likelihood
mobile phone video, 91
motion model, 28, 51, 57
    affine, 59
    projective, 60
    registration parameters, 58
    translational, 59
Moving Picture Experts Group (MPEG), 33
MPEG, *see* Moving Picture Experts Group
MRI, *see* Magnetic Resonance Imaging
MSE, *see* Mean Squared Error

non-stationary noise, 107, 109
Nyquist frequency, 29

optical flow, 43, 57, 126
Optical Transfer Function (OTF), 48
orthogonality condition, 77
OTF, *see* Optical Transfer Function
outlier, 35, 41, 107

Peak Signal-to-Noise Ratio (PSNR), 89
PET, *see* Positron Emission Tomography
photon noise, 111
POCS, *see* Projection Onto Convex Sets
Point Spread Function (PSF), 28, 46, 47, 74, 110
Poisson noise, 111
Positron Emission Tomography (PET), 27, 33
Projection Onto Convex Sets (POCS), 38
PSF, *see* Point Spread Function
PSNR, *see* Peak Signal-to-Noise Ratio

Radial Basis Functions (RBF), 78
RBF, *see* Radial Basis Functions
regularization, 41, 83
    Bilateral Total Variation (BTV), 42, 84, 108
    Tikhonov, 42, 83

Riesz basis, 74

salt and pepper noise, 107
sampling theory, 73
satellite imaging, 33
Sensor Transfer Function (STF), 48
shift and add method, 86
Signal-to-Noise ratio (SNR), 46, 50
sinc function, 40, 79
SNR, *see* Signal-to-Noise ratio
SR, *see* Super-Resolution
stationary noise, 107, 109
STF, *see* Sensor Transfer Function
Super-Resolution, 27
    applications, 33
    as an estimation problem, 40
    blind, 44
    continuous model, 28, 75
    data model, 74
    discrete model, 74
    early approaches, 38
    image formation, 28
    in its stages, 42
    in the frequency domain, 38
    limits, 45
    robust algorithms, 30, 107, 112
synthesis function, 77

UAV, *see* Unmanned Aerial Vehicles
Unmanned Aerial Vehicles (UAV), 127

wavelets, 44
white noise, 30, 41
Wiener filter, 44, 108

Z-transform, 82
Zomet method, 87