2017

Super Resolution of Wavelet-Encoded Images and Videos

Vildan Atalay
University of Central Florida

University of
Central
Florida

STARS
Showcase of Text, Archives, Research & Scholarship

SUPER RESOLUTION OF WAVELET-ENCODED IMAGES AND VIDEOS

by

VILDAN ATALAY
M.S. University of Florida, 2011

A dissertation submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy
in the Department of Computer Science
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term
2017

Major Professor: Hassan Foroosh

ABSTRACT

In this dissertation, we address the multiframe super resolution reconstruction problem for wavelet-encoded images and videos. The goal of multiframe super resolution is to obtain one or more high resolution images by fusing a sequence of degraded or aliased low resolution images of the same scene. Since the low resolution images may be unaligned, a registration step is required before super resolution reconstruction. Therefore, we first explore in-band (i.e. in the wavelet-domain) image registration; then, investigate super resolution.

Our motivation for analyzing the image registration and super resolution problems in the wavelet domain is the growing trend in wavelet-encoded imaging, and wavelet-encoding for image/video compression. Due to drawbacks of widely used discrete cosine transform in image and video compression, a considerable amount of literature is devoted to wavelet-based methods. However, since wavelets are shift-variant, existing methods cannot utilize wavelet subbands efficiently. In order to overcome this drawback, we establish and explore the direct relationship between the subbands under a translational shift, for image registration and super resolution. We then employ our devised in-band methodology, in a motion compensated video compression framework, to demonstrate the effective usage of wavelet subbands.

Super resolution can also be used as a post-processing step in video compression in order to decrease the size of the video files to be compressed, with downsampling added as a pre-processing step. Therefore, we present a video compression scheme that utilizes super resolution to reconstruct the high frequency information lost during downsampling. In addition, super resolution is a crucial post-processing step for satellite imagery, due to the fact that it is hard to update imaging devices after a satellite is launched. Thus, we also demonstrate the usage of our devised methods in enhancing resolution of pansharpened multispectral images.

to my parents Tulay and Ali, my sister Yelda, and my husband Burak

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Super resolution (SR) is the process of obtaining one or more high resolution (HR) images using one or more degraded or aliased low resolution (LR) images of the same scene. Degradation can be a consequence of motion, camera optics, atmosphere, insufficient sampling, etc. Hardware improvement is a straightforward solution to increase the resolution. An obvious way is to reduce the pixel sizes in sensors in order to increase the sampling rate. However, this reduces the image quality by introducing shot noise. It is because the amount of light on each photo-detector decreases as well. Therefore, this solution to obtain high resolution images has a limitation, which is nearly reached by the current sensor technology. A second way of solution is to increase the number of photo-detectors in a sensor (size of a sensor chip). However, this solution would increase the capacitance, causing lower charge transfer rate, increasing the time amount needed for image capturing, and the cost of hardware [3–5]. Therefore, software solutions, called SR methods, are needed as a post-processing step in order to acquire HR images from LR images, without increasing the cost, or introducing shot noise. Especially, when it is hard to update the imaging devices, e.g. in satellite imagery, or it is not feasible to mount heavy cameras, e.g. in unmanned aerial vehicles, SR is the appropriate solution to obtain HR images. For this reason, SR gained popularity in signal processing, computer vision, and machine learning communities, and has a wide range of applications in many areas including but not limited to video surveillance, remote sensing, medical imaging, microscopy imaging, target detection, video coding, and satellite imagery [6–10].

Super resolution methods can be divided into two groups depending on the number of LR images used; namely, single- and multiframe. Single image methods infer an HR image from a single LR frame. These methods have several drawbacks such as high computational complexity, ambiguities in matching HR and LR image patches, and trade-off between large/small image patches, and large/small databases. Therefore, **our focus in this dissertation is on multiframe super**

resolution methods, which need multiple aliased LR images of the same scene in order to reconstruct the lost high frequency information of the HR image. Fig. 1.1 shows an example for multiframe SR problem. Since observed LR images are required to be aliased, image registration, the process of aligning two or more images of the same scene, plays a crucial role in pre-processing for multiframe SR methods. Image registration methods used in SR include interpolation [11], differential [12], direct [13], phase domain [14–20] or wavelet domain approaches [21–29], which will be discussed thoroughly in Section 2.2. Moreover, image registration is also an important step for many areas of image and video processing, such as change detection, medical diagnosis, satellite imagery, video coding, and quality control in manufacturing [30, 31].
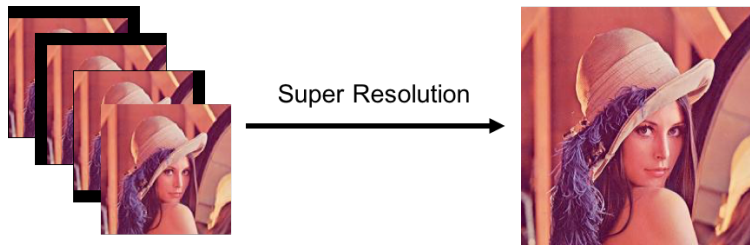
Figure 1.1: Multiframe SR example.

Recently, a considerable literature on super resolution and image registration has grown up around wavelets due to their properties of orthogonality, signal localization, and low computing requirements. Wavelets have also received scholarly and commercial attention in imaging (especially in medical imaging) [32, 33] with the aim of achieving better resolution, reduced distortions, higher SNR, and quick acquisition time; and image/video coding [34–39] for a better compression and to alleviate compression artifacts.

Motion compensated temporal filtering (MCTF) is a very advantageous method in video compression since less bitrate is required to code the prediction errors than the actual frames. Motion estimation (i.e., image registration) is performed before MCTF to obtain displacements between

frames. Later, reference frames are warped onto target frames in order to achieve prediction errors to be encoded. Motion compensated temporal filtering is also an essential step of numerous applications such as scalable video coding [40, 41], still image coding [38, 42], and denoising [43]. A second method to decrease the encoded bit rate in video compression, is super resolution. When downsampling is added as a pre-processing step of compression, super resolution can be employed as a post-processing step in order to reconstruct the required resolution. Video compression methods are required by many applications such as video surveillance, video conferencing, and mobile multimedia, which might have limited bandwidth [44].

Another application area of super resolution is satellite imagery. Typically, multispectral sensors provide low-spatial-high-spectral resolution for the multispectral (MS) volume, and high-spatial-low-spectral resolution for the panchromatic (PAN) images. This is often due to technological limitations inherent in satellite sensors. However, numerous remote sensing applications related to land-cover management, environmental monitoring, weather forecasting, and map updating require high-spatial-high-spectral resolution MS images. In order to obtain these high spatial resolution MS images, a large body of research is devoted to fuse information of MS and PAN bands, which is called pansharpening. Super resolution can also be used to enhance the spatial-resolution of multispectral images [45].

This dissertation is organized as follows. We will explain wavelet decomposition, image registration, motion compensated temporal filtering, super resolution, super resolution of compressed video, and pansharpening problems in the next chapter. We will provide a survey of literature for each problem in Chapter 3. Chapters 4, 5, 6, 7, and 8 are concerned with our methodologies derived for In-band (i.e., in the wavelet domain) Image Registration, Motion Compensated Temporal Filtering for Video Coding, Super Resolution of Wavelet-Encoded Images, Super Resolution of Downsampling Based Compressed Video, and Super Resolution of Pansharpened Multispectral Images, respectively. Finally, we conclude the dissertation and provide future work in Chapter 9.

CHAPTER 2: BACKGROUND

In this chapter, we first start with a brief explanation of wavelet decomposition, and later introduce the problem definitions and our contributions to subjects of image registration, motion compensated temporal filtering, super resolution, super resolution of compressed video, and super resolution of pansharpened multispectral images, respectively.

2.1   Multiresolution Wavelet Decomposition

In this section, we will first explore 1D multiresolution analysis, and later provide the 2D extension.

A signal can be decomposed into a linear combination of a basis for the space that the signal belongs to. Let $f(t) \in L^2(R)$ be a signal, where $L^2(R)$ is the vector space of square integrable signals. Then, $f(t)$ can be expressed as:

$$f(t) = \sum_{j,k} a_{j,k} \psi_{j,k}(t), \tag{2.1}$$

where a and ψ stand for expansion coefficients and expansion functions, respectively; and the expansion functions $\psi_{j,k}(t)$ are constructed by dilations (i.e., j) and translations (i.e., k) of a single function called the *mother wavelet* (i.e., $\psi(t)$), as in:

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k), \tag{2.2}$$

where $2^{j/2}$ is used for normalization. As an example, Haar mother wavelet is defined as follows:

$$\psi(t) = \begin{cases} 1, & 0 \le t < 1/2 \\ -1, & 1/2 \le t < 1 \\ 0, & \text{otherwise.} \end{cases} \tag{2.3}$$

In order to analyze the signals in a multiresolution framework, different scales of dilation can be used. The multiresolution analysis is based on nested approximation subspaces, $V_j \subset V_{j+1}$, where $j \in \mathbb{Z}$, with the following properties [46–48]:

1. $\bigcap_j V_j = \{0\}$ (separation) and $\overline{\bigcup_j V_j} = L^2(R)$ (density),

2. $f(t) \in V_j \iff f(2t) \in V_{j+1}$,

3. $f(t) \in V_j \iff f(t - k) \in V_j$,

4. $V_j = \text{span}\{\phi_{j,k}(t)\}$.

The intersection property indicates that the only constant function, 0, can be in all subspaces V_j, for $j \in \mathbb{Z}$; while, the union property implies, $\bigcup_j V_j$ is dense in $L^2(R)$. Scaling functions, $\phi_{j,k}(t)$, are also defined based on dilations and translations of a single function, $\phi(t)$:

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k). \tag{2.4}$$

As an example, Haar scaling function is defined by:

$$\phi(t) = \begin{cases} 1, & 0 \le t < 1 \\ 0, & \text{otherwise.} \end{cases} \tag{2.5}$$

The wavelet and approximation spaces can be defined as complements, which shows that signals can be decomposed into approximation and wavelet components:

$$V_{j+1} = V_j \oplus W_j, \tag{2.6}$$

where $\oplus$ shows orthogonal direct sum of the subspaces, for $W_j = \mathrm{span}\{\psi_{j,k}(t)\}$. Therefore, $L^2(R)$ can be written for an arbitrary scale, J, as:

$$L^2(R) = V_J \oplus W_J \oplus W_{J+1} \oplus \ldots \tag{2.7}$$

Note that, this implies orthogonality between functions in V_j and W_j:

$$\langle \phi_{j,k}(t), \psi_{j,l}(t) \rangle = \int \phi_{j,k}(t)\psi_{j,l}(t)dt = 0, \tag{2.8}$$

for $k \neq l$. Using Eq. (2.7), the signal $f(t)$ can also be written in the following form:

$$f(t) = \sum_{k \in \mathbb{Z}} a_k \phi_k(t) + \sum_{j \geq 0} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(t). \tag{2.9}$$

The scaling (i.e., a_k), and wavelet (i.e., $d_{j,k}$) coefficients are calculated using inner product and orthogonality:

$$\begin{aligned}
a_k &= \langle f(t), \phi_k(t) \rangle = \int f(t)\phi_k(t)dt, \\
d_{j,k} &= \langle f(t), \psi_{j,k}(t) \rangle = \int f(t)\psi_{j,k}(t)dt.
\end{aligned} \tag{2.10}$$
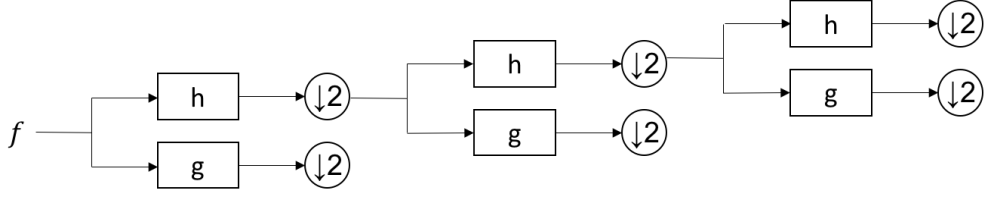
Figure 2.1: 3-level wavelet decomposition.

Using the second property of approximation subspaces, $\phi(t)$ and $\psi(t)$ can also be written as:

$$\begin{aligned}
\phi(t) &= \sqrt{2} \sum_n h(n)\phi(2t - n), \\
\psi(t) &= \sqrt{2} \sum_n g(n)\phi(2t - n),
\end{aligned} \tag{2.11}$$

where $h(n)$ and $g(n)$ are called scaling and wavelet function coefficients, or high and low pass filters, for $n \in \mathbb{Z}$.

Figure 2.1 shows a multiresolution wavelet decomposition for three levels, using high-pass and low-pass *analysis* filters, h and g. When this process is reversed and *synthesis* filters together with upsampling is used, the original signal can be reconstructed.

This analysis can be extended to 2D images in $L^2(R^2)$, by defining the approximation subspace in 2D (i.e., $V_j^{(2)}$) with the tensor product, $V_j^{(2)} = V_j \otimes V_j$. The orthonormal scaling and wavelet basis for $V_j^{(2)}$ and $W_j^{(2)}$ are defined by [47, 48]:

$$\begin{aligned}
\Phi_{j,k,l}(t, u) &= \phi_{j,k}(t)\phi_{j,l}(u), \\
\Psi^h_{j,k,l}(t, u) &= \psi_{j,k}(t)\phi_{j,l}(u), \\
\Psi^v_{j,k,l}(t, u) &= \phi_{j,k}(t)\psi_{j,l}(u), \\
\Psi^d_{j,k,l}(t, u) &= \psi_{j,k}(t)\psi_{j,l}(u).
\end{aligned} \tag{2.12}$$

7

Finally, the 2D signal $f(t, u)$ can be written in terms of the approximation and wavelet functions, for $k, l \in \mathbb{Z}$ and an arbitrary scale, J:

$$
\begin{aligned}
f(t, u) &= \sum_{k,l} a_{J,k,l} \Phi_{J,k,l}(t, u) + \sum_{j \geq J} \sum_{k,l} d^h_{j,k,l} \Psi^h_{j,k,l}(t, u) \\
&+ \sum_{j \geq J} \sum_{k,l} d^v_{j,k,l} \Psi^v_{j,k,l}(t, u) + \sum_{j \geq J} \sum_{k,l} d^d_{j,k,l} \Psi^d_{j,k,l}(t, u),
\end{aligned} \tag{2.13}
$$

where coefficients a and $d^{h,v,d}$ are calculated by inner product, as in 1D case:

$$
\begin{aligned}
a_{J,k,l} &= \int \int f(t, u) \phi_{J,k}(t) \phi_{J,l}(u) dt du, \\
d^h_{j,k,l} &= \int \int f(t, u) \psi_{j,k}(t) \phi_{j,l}(u) dt du, \\
d^v_{j,k,l} &= \int \int f(t, u) \phi_{j,k}(t) \psi_{j,l}(u) dt du, \\
d^d_{j,k,l} &= \int \int f(t, u) \psi_{j,k}(t) \psi_{j,l}(u) dt du.
\end{aligned} \tag{2.14}
$$

Again, as in 1D case, the scaling and wavelet coefficients are used with synthesis filters and upsampling, in order to reconstruct the original images. The formulas in Eqs. 2.9 and 2.13 demonstrate the wavelet decomposition of 1D signals and 2D images, which will be employed in image registration and super resolution methods in the following sections in this chapter.

We will explain the problem definitions and widely applied solutions for image registration, motion compensated temporal filtering, super resolution, super resolution of compressed video, and pansharpening of multispectral images, in the next sections.

2.2 Image Registration

Image registration is a classic problem dealing with geometrically aligning two or more images of the same scene, which are refered as reference and sensed images. Most of the image registration techniques consist of four main steps; feature detection, feature matching, transform model estimation, and image resampling and transformation [31]. While distinctive objects are detected in reference and sensed images in feature detection step; feature extraction deals with finding correspondences between these detected features, via similarity based methods (e.g. correlation or mutual information), frequency-based methods (i.e., Fourier or wavelet), optimization, or pyramid-based methods. After feature matching step, a transform model is estimated via global or local mapping models or interpolation and approximation functions; which is then employed for aligning the images in the final step.

Image registration methods are generally performed by exploiting either spatial or transform domain features. Spatial domain methods [49, 50] often rely on image features that capture local image distortions [12], which is the main drawback of these methods since they can be highly sensitive to distortions caused, for instance, by noise. Fourier-domain methods, on the other hand, are more resilient to spatially localized noise, and are computationally more efficient [51–54]. However, they can be sensitive to noise due to poor localization, and affected by aliasing due to low sensor resolution. In order to reduce computational complexity and avoid local extrema, image pyramids are also used in registration [55, 56], which are employed to find small amounts of motion in the lower levels and propagate the estimates to upper levels. Wavelets, due to their inherent multiresolution nature, provide similar advantages as in image pyramids [57], while avoiding the drawbacks of Fourier and spatial domain methods by exploiting their localization characteristics. We will provide a comprehensive literature review for these groups in Section 3.1.

Image registration techniques are widely used in many areas including, but not limited to, com-

puter vision, medical imaging, remote sensing, cartography, change detection, weather forecasting, target recognition, and monitoring global and land usage using satellite images [58–60]. A straighforward application of image registration is multiframe super resolution, due to required aliasing between LR images. Vast majority of multiframe SR methods assume that the image registration problem is solved a priori [61–64]. Others employ registration algorithms, generally with subpixel accuracy, for either global or local motion. While global motion can be defined by the motion of the object or the camera, local motion depends on the nature of the objects [5]. In this dissertation, we will focus on global motion model in registration for super resolution application.
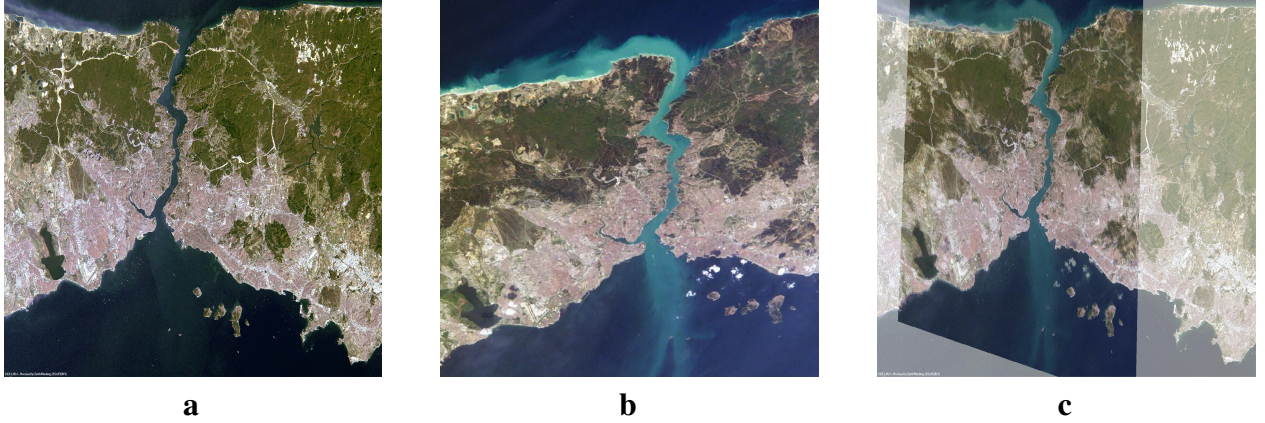
Figure 2.2: Image registration example **a** Reference image **b** Sensed image **c** Registrered and overlapped images.

A global motion between images can be modeled as translational, rigid, affine, or projective transformation [5]. We can describe a general affine model as follows:

$$
\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{2.15}
$$

where $(x, y, 1)$ and $(u, v, 1)$ are the homogeneous coordinates of the reference and sensed images,

respectively, and $a_{i,j}$, for $1 \leq i, j \leq 3$, stand for the motion parameters. The goal of image registration is to find the parameters of transformation matrix given coordinates of the two images. Figure 2.2 shows two satellite images (**a** and **b**) taken from Istanbul, Turkey, and the registration result (**c**) by warping **b** onto **a**, after parameters are estimated (Images courtesy of ESA Earth Observation Dataset and NASA ISS, respectively).

Several studies in SR literature employ Taylor series expansion in order to find the motion parameters [49, 65–67]. The general approach in these methods is to minimize the error between a sensed image and Taylor series expansion of a reference image. In one of the first examples of this group, Keren *et al.* [49] define a translational and rotational transform between reference image, $I(x, y)$, and sensed image, $J(x, y)$, as in:

$$J(x, y) = I(x \cos(\theta) - y \sin(\theta) + t_x, x \sin(\theta) + y \cos(\theta) + t_y), \tag{2.16}$$

where θ shows the rotation, and (t_x, t_y) stand for the translation in two dimensions between images. After performing Taylor series expansion, the error function to be minimized becomes [49]:

$$E(t_x, t_y, \theta) = \sum \left[\left(I(x, y) + \left(t_x - y\theta - \frac{x\theta^2}{2} \right) \frac{\partial f}{\partial x} + \left(t_y + x\theta - \frac{y\theta^2}{2} \right) \frac{\partial f}{\partial y} \right) - J(x, y) \right]^2 \tag{2.17}$$

for the overlapping area of the images. The error function is minimized by setting the derivatives with respect to unknowns to zero. This method is mainly valid for small translational and rotational motions; therefore, is later modified to use Gaussian pyramid to demonstrate large translations in higher resolutions as small translations in lower resolutions of the pyramid; and extended to be used in modeling more general transformations by dividing images into blocks with uniform motion [65].

When the motion is more complicated, as in real videos, parametrization of the motion model can

be difficult, leading to a group of methods called optical flow, which use dense correspondences between images to estimate motion of each pixel [68–71]. One of the first optical flow techniques used in literature is the Lucas-Kanade method [72] which defines the relationship between reference, $I(\mathbf{x})$, and sensed, $J(\mathbf{x})$, images by:

$$J(\mathbf{x}) = I(\mathbf{A}\mathbf{x} + \mathbf{t}), \tag{2.18}$$

where $\mathbf{x}$ stands for dimensions of the images (e.g. $\mathbf{x} = (x, y)$ in 2D), $\mathbf{A}$ shows the transform matrix for a general affine transform, and $\mathbf{t}$ is the translational displacement vector for all dimensions. A linear approximation of the reference image $I(\mathbf{x})$ is then defined as follows:

$$I\big((\mathbf{A} + \Delta\mathbf{A})\mathbf{x} + (\mathbf{t} + \Delta\mathbf{t})\big) \approx I(\mathbf{A}\mathbf{x} + \mathbf{t}) + (\Delta\mathbf{A}\mathbf{x} + \Delta\mathbf{t})\frac{\partial}{\partial x}I(\mathbf{x}), \tag{2.19}$$

where $\dfrac{\partial}{\partial x}$ stands for the gradient operator in n dimensions:

$$\frac{\partial}{\partial x} = \begin{bmatrix} \dfrac{\partial}{\partial x_1} & \dfrac{\partial}{\partial x_2} & \cdots & \dfrac{\partial}{\partial x_n} \end{bmatrix}^T, \tag{2.20}$$

where superscript T stands for transpose. In order to minimize the error function given in:

$$E = \sum_{\mathbf{x}} \big[I(\mathbf{A}\mathbf{x} + \mathbf{t}) - J(\mathbf{x})\big]^2, \tag{2.21}$$

derivatives with respect to unknown parameters in Eq. (2.19) are set to zero, and Newton-Raphson iteration is employed.

Optical flow methods are later used together with feature extraction methods in order to obtain better quality HR images [73]. However, optical flow based methods, in general, have high computational complexity and are prone to errors under noisy conditions and large motions [7].

Numerous multiframe super resolution studies [54, 74, 75] have also investigated the use of frequency based, specifically Fourier domain, registration methods, which exploit the transform domain representation of the images [31]. Majority of these methods depend on the use of shift property of Fourier transform, and the cross power spectrum defined in [76]. When the translational shift between the images to be registered is defined as follows:

$$J(x, y) = I(x - t_x, y - t_y), \tag{2.22}$$

the Fourier shift property states the relationship in the Fourier domain as:

$$F_J(u, v) = F_I(u, v)e^{-i(ut_x + vt_y)}, \tag{2.23}$$

where $F_{I,J}$ stand for the Fourier transformed images, and (u, v) are the coordinates in Fourier domain. The correlation between images is then found by normalized cross power spectrum as:

$$\frac{F_J(u, v)F_I(u, v)^*}{|F_J(u, v)F_I(u, v)^*|} = e^{-i(ut_x + vt_y)}, \tag{2.24}$$

where * shows conjugate gradient.

In an extensively employed Fourier-based method by Foroosh *et al.* [14], the above mentioned relationship is extended for subpixel accuracy. Moreover, Vandewalle *et al.* [54] propose another widely used method, where rotation and translation are estimated sequentially, by first transforming the image coordinates to polar coordinates, in which rotation can also be defined by translation. Methods in this group, as mentioned before, are prone to errors depending on aliasing and noise.

Several other registration algorithms have also been utilized in multiframe SR literature. While hierarchical methods employ coarse-to-fine refinement for registration [56, 77, 78]; Bayesian-based methods marginalize the unknowns that are either the high resolution image [79] or the registration

parameters [80]. Furthermore, learning based methods are also used as reported in [81], where temporal dependencies are learnt via convolutional neural network (CNN) based methods.

Since the performance of multiframe SR methods are largely dependent on the registration accuracy, a group of methods focus on estimating high resolution images and registration parameters simultaneously [82–84], where the relationship between LR images are incorporated into the SR observation model. Unfortunately, these methods suffer from complex models and extra parameters [7]. More recently, researchers have also investigated multiframe SR without explicit motion estimation, where rather than an accurate registration, a general motion model is enough to increase resolution [85–87]. The main drawback of these methods is the computational complexity [7].

Finally, wavelet-based multiframe SR methods have also used several types of image registration techniques, including hierarchical methods [71], Fourier-based methods [9], elastic image registration [24], and optical flow [29], while a vast majority of them assume that motion parameters are known beforehand [88, 89]. Wavelet based image registration techniques, in general, make use of features extracted from multiscale wavelet decomposition [23, 25–28] to find the correspondences between reference and sensed images. These methods generally find the local maxima in wavelet decomposition of the input images, and use these features to search for the best match. Stone *et al.* [22] demonstrate the use of low-pass subbands for feature extraction, proving that these subbands are not as sensitive to translational shifts under the condition that features are greater than twice the size of wavelet blocks. As an alternative approach to the problem, Wu *et al.* [21] propose a coarse-to-fine wavelet-based motion model based on hierarchical basis functions by Cai and Wang [90]. The motion model is defined as follows:

$$
\begin{aligned}
u(x,y) &= u_0(x,y) + \sum_{j=0}^{J} \left(u_j^h(x,y) + u_j^v(x,y) + u_j^d(x,y) \right), \\
v(x,y) &= v_0(x,y) + \sum_{j=0}^{J} \left(v_j^h(x,y) + v_j^v(x,y) + v_j^d(x,y) \right),
\end{aligned}
\tag{2.25}
$$

14

where u_0 and v_0 demonstrate the coarsest level motion vectors for approximation coefficients, $u^{h,v,d}$ are motion vectors for detail subbands expressed as:

$$
\begin{aligned}
u_0(x,y) &= \sum_{k,l \in \mathbb{Z}} a_{k,l} \Phi_{k,l}(x,y), \\
u_j^h(x,y) &= \sum_{k,l \in \mathbb{Z}} a_{j,k,l}^h \Psi_{j,k,l}^h(x,y), \\
u_j^v(x,y) &= \sum_{k,l \in \mathbb{Z}} a_{j,k,l}^v \Psi_{j,k,l}^v(x,y), \\
u_j^d(x,y) &= \sum_{k,l \in \mathbb{Z}} a_{j,k,l}^d \Psi_{j,k,l}^d(x,y),
\end{aligned}
\tag{2.26}
$$

$v^{h,v,d}$ are defined similarly. The motion vectors u and v are later found iteratively by minimizing a cost function with sum of squared differences.

Above mentioned image registration methods used for wavelet-based multiframe SR application do not utilize the relationship between wavelet subbands effectively. In this dissertation, our goal is to obtain a wavelet domain subpixel registration method that can achieve highly accurate results using wavelet coefficients directly, that will be used for wavelet-based multiframe SR, and wavelet-encoded imaging applications. **We propose a subpixel motion estimation method that can achieve direct wavelet domain registration from a sparse set of coefficients**. We make the following contributions towards this goal: (**i**) We devise a method of decoupling scale, rotation, and translation parameters in the wavelet domain, (**ii**) we derive explicit mathematical expressions that define in-band subpixel registration in terms of Haar wavelet coefficients, (**iii**) using the derived expressions, we propose an approach to achieve in-band subpixel registration, avoiding back and forth transformations. (**iv**) Our solution remains highly accurate even when a sparse set of coefficients are used, which is due to localization of signals in a sparse set of wavelet coefficients. Extensive experiments are used to validate our method both on simulated and real data under various scenarios.

In the next section, we will discuss a straightforward application of wavelet-based image registration in video coding.

2.3   Motion Compensated Temporal Filtering

Motion compensated temporal filtering (MCTF) has an important role in video coding, which is the process of reducing the size of a video file to be stored or transmitted. Compression in video coding methods can be achieved by either individually encoding each frame (i.e., intra-frame coding) using image compression techniques, or revealing the redundancy between frames (i.e., inter-frame coding) for temporal filtering, based on the fact that consecutive frames have similar data. Intra-frame coding techniques expose the redundancy in the same frame (i.e., I-frame), and code pixels or pixel blocks in a frame by prediction based on previously coded ones. Our focus in this dissertation is on inter-frame coding, since higher compression ratios can be achieved via revealing redundancies between frames.

A straightforward solution to reveal the redundancy between frames and reduce required bits for coding, is to encode the difference of two consecutive frames together with the reference frame, instead of coding both original frames. Since more similar neighboring frames increase the compression rate by decreasing the difference to be coded, coding motion compensated error, which has less energy than the difference, results in more efficient compression. Therefore, motion compensation is a key aspect in inter-frame coding. Figure 2.3 shows an example of two frames from widely used Foreman video sequence (**a** and **b**), the difference between these frames (**c**) and motion compensated error (**d**).
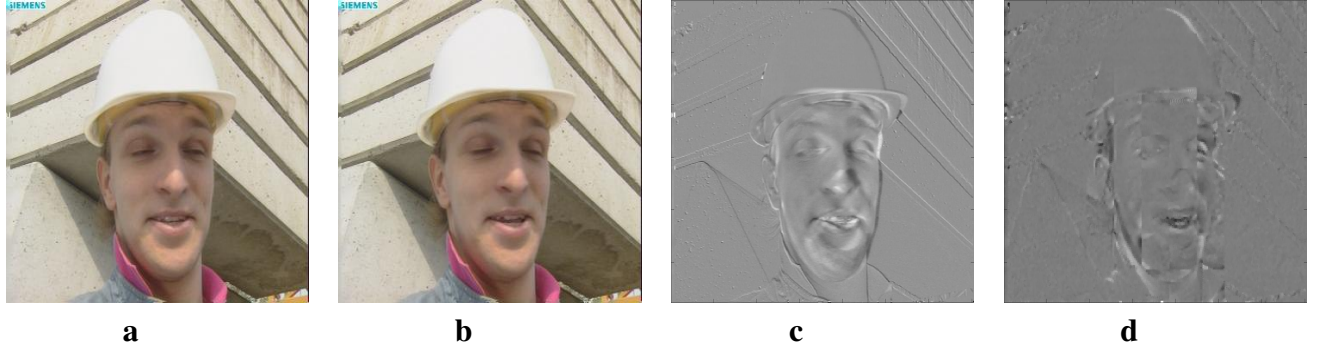
Figure 2.3: Motion compensation example **a** Reference image in Foreman video sequence **b** Target image **c** Difference **d** Motion compensated error.

A fundamental pre-step of motion compensation is motion estimation, which predicts where the pixels in the current frame originate in the reference frames. Prediction of inter-coded frames can be performed in either a uni-directional or a bi-directional framework. While uni-directional methods predict the target frames (i.e., P-frame) based on the previous frames only; bi-directional methods take advantage of both previous and future frames in order to code the current frame (i.e., B-frame). Although B-frames have the highest compression rates compared to others, there might be a delay caused by the need of future frames to be encoded before the current one.

Motion estimation can be carried out by block matching, pel-recursive matching, optical flow, or mesh matching techniques [91]. The general idea of these motion estimation methods is based on the optical flow formula:

$$|I(x, y, n) - I(x - t_x, y - t_y, n - 1)| = \epsilon, \tag{2.27}$$

where ϵ demonstrates the small change with respect to translational shifts only, since object motions are assumed to be pixel-valued translational, and occlusions and light variations are avoided. In general, the error between frames is minimized based on either mean square, or mean absolute difference with full search technique. Even though this exhaustive method has high computational

cost, and there exists accelerated methods in the literature [92, 93]; it is very regular, parallelizable and easy to implement in the hardware [94]; therefore, still is the most commonly used method.

For the motion estimation step, block matching method [95] divides the target frame into usually square blocks, and predicts where each block comes from in the reference frames. Although, in general, blocks of size 16×16 are employed, some methods use smaller blocks to track small objects in complex motions, and larger blocks to decrease image noise. Block matching method is generally accurate for pixel level translation; however, interpolation can be used in order to obtain half or quarter pixel accuracy. Although this method causes blocking artifacts due to quantization of Discrete Cosine Transform (DCT) coefficients, the widely used transform in coding, it still is the most accessible method for hardware due to its simplicity. Modifications to block matching method are proposed by the use of overlapping blocks [96], variable size blocks [97], and hierarchical block matching [98].

Other motion estimation methods include pel-recursive matching [99], which recursively calculates motion vectors for each pixel using gradient methods, and can also be modified to use hierarchical matching [100]; optical flow-based methods that approximate the derivatives for dense velocity estimate [91]; and mesh matching method [101] which divides the blocks into triangles, and estimate the affine transformation using the corners of triangles. Due to their high computational complexity, these methods are not often used in video coding.

After the motion estimation step is completed, motion compensation step is performed, where the blocks in the reference frames are warped onto the target frames, using the estimated motion vectors. Later, the error between the target and predicted target frames, together with the motion vectors and reference frames are transmitted through an encoder. At the receiver, the decoder rebuilts the target frames using provided data. This whole process is named as motion compensated temporal filtering.

MCTF methods perform temporal filtering in either an open-loop or a closed-loop fashion. While open-loop framework uses the original reference images in predictions, closed-loop architecture employs the reconstructed reference frames in the decoder side. We explore the open-loop architectures in this dissertation, since closed-loop methods are not as easy to implement for hardware, even though their coding efficiency is relatively higher [102]. Due to the drawbacks (e.g. blocking artifacts) of widely utilized DCT in video coding, the recent trend is to use Discrete Wavelet Transform (DWT) in MCTF methods. Open-loop DWT-based MCTF is performed either directly on input images, or on their transforms; thus, can be categorized into two groups depending on the order of temporal and spatial transforms as $t + 2D$ [103, 104] and $2D + t$ [38, 41], where $2D$ and t stands for *spatial* and *temporal* decompositions, respectively.
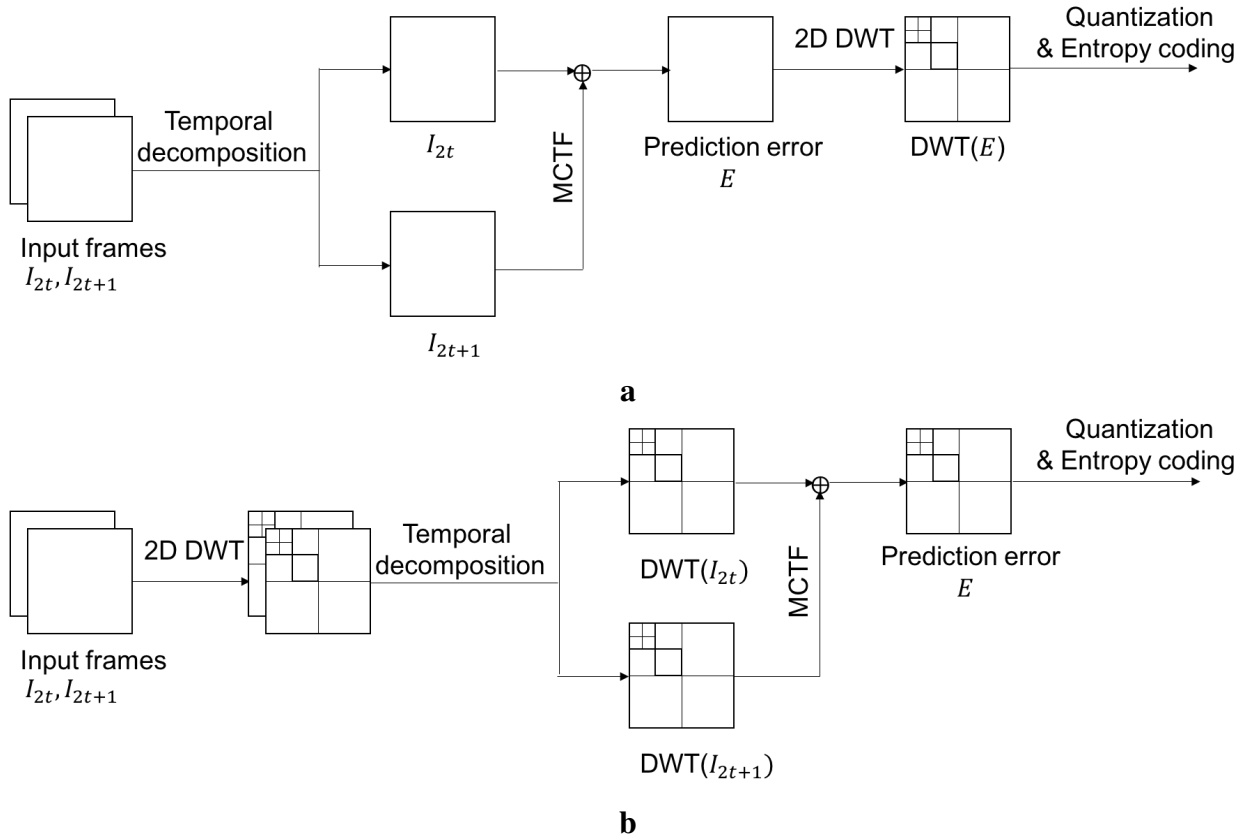
Figure 2.4: Wavelet based video coding **a** $t + 2D$ **b** $2D + t$.

Spatial-domain MCTF methods (i.e., $t + 2D$) first decompose the frame sequence temporally, employ motion compensation on the original frames, and encode the temporally filtered frames using wavelet-based coding schemes. In-band MCTF methods (i.e., $2D + t$), on the other hand, first decompose the frames in wavelet domain, and use motion compensation on the wavelet subbands, which obtain the error frames to be encoded directly in the wavelet domain. Figure 2.4 demonstrates the encoder side of the MCTF based video coding, for these two frameworks. Even though the spatial domain MCTF methods have higher performance compared to in-band ones, they have drifting and operational mismatch problems [39].

MCTF methods in the literature so far, perform motion estimation/motion compensation (ME/MC) either in the temporal domain before DWT, or in the wavelet domain with the help of redundancy (e.g. ODWT, DT-CW, etc.), due to the fact that complete (i.e., critically sampled) DWT is shift-variant and motion estimation directly on DWT subbands is a challenging task. However, redundancy in these methods leads to high computational complexity [105]. Inspired by the facts that shift variance keeps the perfect reconstruction and nonredundancy properties of wavelets, breaks the coupling between spatial subbands, and that wavelet codecs always operate on complete DWT subbands [38], we propose a novel in-band MCTF method, which avoids the need of shift invariance, and operates directly on the original DWT coefficients of the input sequences. Since Haar wavelets are widely utilized in MCTF methods due to the coding efficiency based on their short kernel filters [38], our method is built on Haar subbands. For accurate ME/MC, we define the exact relationships between the DWT subbands of input video sequences, which allows us to avoid upsampling, inverse DWT, and calculation of redundant DWT coefficients, while achieving arbitrary subpixel accuracy without interpolation, and high quality video even at very low-bitrates. Experimental results demonstrate the accuracy of the proposed method, confirming that our model for ME/MC effectively improves video coding quality.

We devote the next section to another application area of image registration, namely, super resolu-

tion reconstruction.

2.4 Super Resolution Reconstruction

Even though our main focus in this dissertation is on multiframe SR methods, we provide a comprehensive review of both single and multiframe methods, in this section.

Super resolution methods can be categorized based on several criteria, including the domain of algorithms (i.e., frequency or spatial domain), number of low resolution images used to obtain a high resolution image (i.e., single or multiframe), or the employed methods (e.g. interpolation, regularization etc.) [5, 106, 107]. We will use a classification similar to [5] which divides the methods into two groups based on the domain used, namely, frequency and spatial domains. We will first discuss the early examples of frequency based methods, in the Fourier domain; and later, examine the spatial domain algorithms. Finally, we will investigate the more recent frequency based approach; i.e., wavelet-based techniques.

Frequency based methods, can be performed either in Fourier or wavelet domains. Methods in this group, first decompose the images into transform domain coefficients, make use of the aliasing between LR images in order to recover the high frequency information lost during the image acquisition process, and perform inverse transform to obtain HR images.

One of the first frequency-based SR methods proposed in the literature is by Tsai and Huang [3], where authors use the shift and aliasing properties of Fourier transform, on globally translated images taken by Landsat 4 satellite [5]. Let us demonstrate the continuous HR scene by $\mathbf{x}(x,y)$. Since the HR images are globally translated, we can define the kth HR image by $\mathbf{x}_k(x,y) = \mathbf{x}(x+t_{x_k}, y+t_{y_k})$, where (t_{x_k}, t_{y_k}) stand for the translations in x and y directions for the kth image, respectively, for $k = 1, \ldots, K$. The relationship between these HR images can be explained using

the shift property of Continuous Fourier Transform (CFT):

$$\mathscr{X}_k(u, v) = e^{j2\pi(ut_{x_k} + vt_{y_k})}\mathscr{X}(u, v), \qquad (2.28)$$

where $\mathscr{X}(u, v)$ and $\mathscr{X}_k(u, v)$ demonstrate the CFT of reference HR image, and k globally translated HR images, respectively. When the HR images are sampled with sampling period T_1, T_2 in two directions, sampled LR images can be defined by $\mathbf{y}_k[m, n] = \mathbf{x}_k(mT_1 + t_{x_k}, nT_2 + t_{y_k})$. Then, the relationship between these discrete LR and continous HR images can be defined using the aliasing property of Discrete Fourier Transform (DFT) and CFT of the scenes as in [5, 108]:

$$\mathscr{Y}_k[p, q] = \frac{1}{T_1 T_2}\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\mathscr{X}_k\left(\frac{2\pi}{T_1}\left(\frac{p}{M} - i\right), \frac{2\pi}{T_2}\left(\frac{q}{N} - j\right)\right), \qquad (2.29)$$

where $\mathscr{Y}_k[p, q]$ shows the DFT of LR images. By combining the equations in (2.28) and (2.29), we can obtain:

$$\mathscr{Y} = \mathbf{\Phi}\mathscr{X}, \qquad (2.30)$$

where $\mathbf{\Phi}$ demonstrates the matrix for the relationship. Equation (2.30) is later solved by a least squares algorithm. This method assumes that the LR images are impulse sampled, which does not take the blur into account. Later, it is modified to include the blurring effect and noise by Kim *et al.* [109], and regularization is added by Bose *et al.* [110] to decrease registration errors. Even though Fourier-based methods are intuitive, have low computational complexity, and parallelizable to reduce hardware complexity [106]; due to their global nature, they only allow linear space invariant blur (PSF), and are not appropriate for real-world applications [107]. Moreover, it is difficult to identify a global frequency-domain *a priori* knowledge to overcome ill-posedness.

Figure 2.5: Super resolution observation model.

Spatial-domain methods, on the other hand, tackle Fourier domain obstables by using interpolation, regularization, and machine learning methods, in order to take real world scenarios into account [64, 65, 111–114]. These methods can be performed using either single or multiple LR images. Let $\mathbf{x}(x, y)$ denote the desired HR image, and $\mathbf{y}_k$ be the kth observed LR image. The general multiframe super resolution observation model is given by:

$$\mathbf{y}_k = \mathbf{\Lambda}_k \mathbf{B}_k \mathbf{M}_k \mathbf{x} + \mathbf{n}_k, \quad k = 1, 2, ..., K, \tag{2.31}$$

where $\mathbf{M}_k$, $\mathbf{B}_k$, $\mathbf{\Lambda}_k$, and $\mathbf{n}_k$ denote matrices for motion, blurring effect, downsampling operator, and noise term for the kth LR image, respectively, and K is the number of observed LR images. $\mathbf{y}_k$ and $\mathbf{x}$ are represented as lexicographically ordered vectors. Figure 2.5 shows the observation model for a single frame. Given a sequence of observed LR images, $\mathbf{y}_k$, the goal of multiframe SR is to reconstruct an unknown HR image, $\mathbf{x}$.

In order to solve for $\mathbf{x}$ in Eq. (2.31), interpolation-based methods [64, 65, 115], first project all LR images onto HR grid, then fuse the information from all images using a general interpolation technique, such as the nearest neighbor, bilinear, and bicubic. These methods are generally performed in three steps: (1) registration, (2) interpolation, and (3) deblurring [106–108].

In one of the first works of this group, Iterative Back Projection (IBP) method is proposed by Irani

23

and Peleg [65], where an initial HR guess is iteratively refined, by back projecting the error between simulated and observed LR images, to the HR grid. The goal of IBP method is to minimize the following cost function using the notation from Eq. (2.31):

$$\arg\min \left\| \mathbf{\Lambda}_k \mathbf{B}_k \mathbf{M}_k \mathbf{x} - \mathbf{y}_k \right\|_2^2 . \tag{2.32}$$

The iterative scheme is then expressed by:

$$\mathbf{x}^{n+1} = \mathbf{x}^n + \frac{1}{K} \sum_{k=1}^{K} \left(\mathbf{y}_k - \mathbf{y}_k^n \right) h^{BP}, \tag{2.33}$$

where h^{BP} is the back-projection kernel and superscript n stands for the iteration number.

This method is later modified by Irani and Peleg [116] for a general motion model; and Zomet *et al.* [115] use median of differences instead of the average, for a faster algorithm. In order to avoid occilations between solutions, IBP can also be extended to include regularization. Even though IBP method is intuitive; there is no unique solution due to ill-posedness, and it might be hard to choose the back-projection kernel.

The projection onto convex sets (POCS) method, on the other hand, solves the restoration and interpolation problems together by taking prior knowledge into account [106]. The first POCS method was proposed by Stark and Oskoui [117], where the desired HR image is restricted to be a member of intersection of closed convex sets. These convex sets can be defined based on several constraints [106, 108]. For example, data consistency constraint can be modeled as

$$S_k = \{ \mathbf{x} \mid \delta_l \leq \left| \mathbf{\Lambda}_k \mathbf{B}_k \mathbf{M}_k \mathbf{x} - \mathbf{y}_k \right| \leq \delta_u \}, \tag{2.34}$$

where $\delta_{l,u}$ represents the lower and upper bound uncertainties [5]. Having a group of M convex

sets, the HR solution can be found iteratively, by projecting HR image onto these convex sets:

$$\mathbf{x}^{n+1} = \mathscr{P}_M \mathscr{P}_{M-1} \ldots \mathscr{P}_2 \mathscr{P}_1 \mathbf{x}^n, \tag{2.35}$$

where $\mathscr{P}$ demonstrates the projection operator.

The accuracy of POCS method is largely based on the motion estimation; therefore, in order to decrease this dependency, this method is later modified by Rajan and Chaudhuri [118] to disable the projections with estimation errors. Patti *et al.* [119] also extend this method to handle blur, aliasing, and space-varying PSF. Even though POCS methods can incorporate several types of priors easily; they have high computational cost, slow convergence, and the solution might differ based on the initial guess [106, 108].

In order to stabilize the generally ill-posed SR problem, regularization is incorporated into the solutions, where SR problem is cast into the Bayesian framework. As an example of regularization based methods, the maximum likelihood (ML) method was first articulated by Tom *et al.* [112], where the aim is to maximize the following cost function:

$$\hat{\mathbf{x}}^{ML} = \arg\max_{\mathbf{x}} p(\mathbf{Y}|\mathbf{x}). \tag{2.36}$$

Since the LR images, $\mathbf{Y}$, are independent from the HR image, $\mathbf{x}$, the conditional probability density function (pdf) can be defined by [5, 107]:

$$
\begin{aligned}
p(\mathbf{Y}|\mathbf{x}) &= \prod_{k=1}^{K} p(\mathbf{y}_k|\mathbf{x}) \\
&\propto \prod_{k=1}^{K} \exp\left(-\frac{1}{2\sigma_k^2}\left\|\mathbf{y}_k - \mathbf{\Lambda}_k \mathbf{B}_k \mathbf{M}_k \mathbf{x}\right\|^2\right) \\
&= \exp\left(-\sum_{k=1}^{K} \frac{1}{2\sigma_k^2}\left\|\mathbf{y}_k - \mathbf{\Lambda}_k \mathbf{B}_k \mathbf{M}_k \mathbf{x}\right\|^2\right).
\end{aligned}
\tag{2.37}
$$

This leads to expressing the cost function as in:

$$\hat{\mathbf{x}}^{ML} = \arg\min_{\mathbf{x}} \left\| \mathbf{y}_k - \mathbf{H}_k\mathbf{x} \right\|^2, \tag{2.38}$$

where $\mathbf{H} = \mathbf{\Lambda BM}$. The derivative of Eq. (2.37) with respect to $\mathbf{x}$ is then set to zero, in order to find a solution as follows:

$$\hat{\mathbf{x}}^{ML} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{Y}, \tag{2.39}$$

where the superscript T stands for transpose of a matrix. In the above formula for the maximum likelihood solution to HR image (i.e., $\hat{\mathbf{x}}^{ML}$), if $(\mathbf{H}'\mathbf{H})$ is singular, then the solution is not unique, which requires incorporating prior knowledge into the solution, leading to maximum a posteriori (MAP) framework. Methods in this group maximize the following cost function [107]:

$$\hat{\mathbf{x}}^{MAP} = \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{Y}). \tag{2.40}$$

By using the Bayes' rule, the above conditional pdf can be rewritten:

$$p(\mathbf{x}|\mathbf{Y}) \propto p(\mathbf{Y}|\mathbf{x})p(\mathbf{x}), \tag{2.41}$$

taking the logarithm,

$$
\begin{aligned}
log(p(\mathbf{x}|\mathbf{Y})) &= \log\big(p(\mathbf{Y}|\mathbf{x})\big) + \log\big(p(\mathbf{x})\big) \\
&= \log\left(\prod_{k=1}^{K} p(\mathbf{y}_k|\mathbf{x})\right) + \log(p(\mathbf{x})) \\
&= \sum_{k=1}^{K} \log(p(\mathbf{y}_k|\mathbf{x})) + \log(p(\mathbf{x})).
\end{aligned}
\tag{2.42}
$$

Here, $p(\mathbf{Y}|\mathbf{x})$ is defined as in Eq. (2.37). Then, $\hat{\mathbf{x}}^{MAP}$ can be reformulated as:

$$\hat{\mathbf{x}}^{MAP} = \arg\min_{\mathbf{x}} \left(\left\| \mathbf{y}_k - \mathbf{H}_k\mathbf{x} \right\|_2^2 + \lambda\Gamma(\mathbf{x}) \right), \tag{2.43}$$

where λ is the regularization parameter which is used to control the trade-off between residual and regularization terms, and Γ is a priori energy function. The most appropriate value for regularization parameter can be found by employing several algorithms, such as L-curve, U-curve, and generalized cross validation. Gaussian Markov Random Field (MRF) is one of the commonly used a priori functions:

$$\Gamma(\mathbf{x}) = \mathbf{x}'\mathbf{S}\mathbf{x}, \tag{2.44}$$

where $\mathbf{S}$ is a symmetric positive matrix. Another widely adapted a priori function is Total Variation (TV):

$$\Gamma(\mathbf{x}) = \left\| \nabla\mathbf{x} \right\|_1, \tag{2.45}$$

where ∇ stands for the gradient operator, and L_1 norm is employed.

Despite the fact that it is easy to incorporate prior knowledge into the SR framework using regularization based methods, they still suffer from high computational cost, and over-smoothing effects.

If the general image prior is not enough for regularization, example-based methods can be carried out [108, 113, 114], where an HR image is obtained from generally a single LR image by utilizing training sets of LR/HR images or patch pairs. In their seminal work, Freeman *et al.* [114] make use of the MRF model to infer the lost high frequency of an HR image using a single LR image.

Authors formulate the SR problem similar to Eq. (2.40),

$$x = \arg\max_{x} p(x|y), \tag{2.46}$$

where x and y stand for the high-frequency of HR, and mid-frequency of upsampled LR image [120], respectively; and express the likelihood and prior terms as:

$$p(y|x) = \prod_{k} p(y_k|x_k), \tag{2.47}$$

$$p(x) = \prod_{x_j \in \mathcal{N}(x_i)} p(x_i|x_j), \tag{2.48}$$

where the HR image patch, x_j, is in the neighborhood (i.e., $\mathcal{N}$) of another image patch, x_i. These likelihood and prior terms are learnt by training image sets, and desired high frequency is found by belief propagation.

Example-based methods can also be extended to use multiple low resolution images [107]. Even though learning-based methods provide promising results, their performance depends on several criteria. For example, while small patch sizes cause insignificant predictions; larger ones require immense datasets to find closer patches [108]. Since the input LR image and the training samples should be similar for a good prediction of an HR image, learning-based methods tend to limit their plot to use application specific prior knowledge. Furthermore, high computational cost is another shortcoming of the methods in this category.

Above mentioned methods can also be utilized in combinations; e.g. ML and POCS [121], MAP and POCS [8], ML and IBP [122] methods have been combined in the literature. As some general aspects of SR methods, color images are handled most usually by converting the images to either YIQ or YCbCr color spaces. Since human visual system is more sensitive to changes in luminance channel, the SR method is applied on this channel, and the other channels are upsampled using

interpolation [123]. As an alternative to this approach, all three channels can be reconstructed using the proposed SR algorithms, in YCbCr or RGB spaces [57, 78]. Finally, the SR methods are usually evaluated based on mean square error (MSE), peak-signal-to-noise ratio (PSNR), and structural similarity measure (SSIM) [124].

In order to overcome the drawbacks of aforementioned frequency and spatial domain methods, recent research in SR explores wavelet-based techniques [29, 63, 89, 125, 126]. The intuition behind these approaches is that the observed LR images can be used to model the low-pass subbands of the unknown HR images, in order to reconstruct the high frequency information lost during image acquisition.
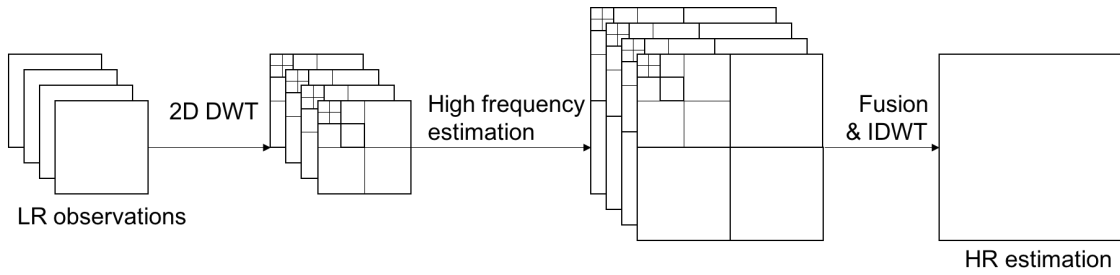
Figure 2.6: Wavelet-based SR framework.

Wavelet-based algorithms, similar to Fourier domain methods, first decompose the images into frequency domain subbands, in order to reveal the correlation of low and high frequency information of the images. The unknown high frequency subbands in finer scales are predicted, using the information provided by lower levels of observations. Finally, inverse wavelet transform is performed to obtain HR images. Figure 2.6 demonstrates this wavelet-based SR framework.

As one of the first wavelet-based SR methods, Nguyen and Milanfar [88] make use of the multires-

olution representation of wavelets defined as in Section 2.1:

$$
\begin{aligned}
f(t, u) &= \sum_{k,l} a_{J,k,l} \Phi_{J,k,l}(t, u) + \sum_{j \geq J} \sum_{k,l} d^h_{j,k,l} \Psi^h_{j,k,l}(t, u) \\
&+ \sum_{j \geq J} \sum_{k,l} d^v_{j,k,l} \Psi^v_{j,k,l}(t, u) + \sum_{j \geq J} \sum_{k,l} d^d_{j,k,l} \Psi^d_{j,k,l}(t, u).
\end{aligned}
\tag{2.49}
$$

By demonstrating the sample points provided by LR images in the HR grid and ignoring the detail subbands in Eq. (2.49), the observed data is written in the form:

$$
f(xs + \epsilon_t, ys + \epsilon_u) \approx a_{J,k,l} \phi_{J,k}(xs + \epsilon_t) \phi_{J,l}(ys + \epsilon_u),
\tag{2.50}
$$

where s is the resolution enhancement factor, $0 \leq \epsilon_t, \epsilon_u < s$, and $x = 0, \ldots, m-1, y = 0, \ldots, n-1$ with an LR image of size $m \times n$. Authors first find approximation coefficients by a regularized least squares approach using the above formula in a matrix form. Later, the difference between the original image and its estimate is used to find the remaining subbands sequentially.

Wavelet-based methods in the literature so far are based on interpolation [29, 63, 88, 89, 125, 127, 128], regularization [129, 130], or learning algorithms [126, 131, 132]. While interpolation based methods focus on interpolating the wavelet subbands (e.g. DWT or stationary wavelet transform) instead of the spatial domain images; regularization dependent methods take advantage of the multiscale nature of wavelet domain for image priors. On the other hand, learning based methods aim to learn the high frequency information of the observed LR images using the correspondence between wavelet transform of HR/LR image pairs.

One main problem with above mentioned multiview methods is that resampling from nonuniform samples (provided by multiple images) and the inversion of the point spread function (PSF) are highly nonlinear and ill-posed problems. Non-linearity and ill-posedness are typically overcome by linearization and regularization, often through an iterative optimization process, which essentially

trade off the very same information (i.e., high frequency) that we want to recover. We propose a different point of view for multiview SR that is very much like single-image methods which extrapolate the spectrum of one image selected as reference from among all views. However, for this, the proposed method relies on information provided by all other views, rather than prior constraints as in single-image methods which may not be an accurate source of information. In that sense, our method is half-way between single-image and multiview methods, taking advantage of the best of both worlds. We assume that displacements between the reference and other LR images are translational and known a priori or estimated. Our contributions are as follows: (**i**) We establish explicit closed-form expressions that define how the local high frequency information that we aim to recover for the reference HR image is related to the local low frequency information in the given sequence of LR views. (**ii**) We assume that the LR images correspond to the polyphase approximation coefficients of the first level wavelet transform of unknown HR images; therefore, the locality of derived expressions due to modeling using wavelets reduces the problem to an exact and linear set of equations that are well-posed and solved algebraically without requiring regularization or interpolation. Our approach is closed-form, and provides results that are superior to the state-of-the-art. We provide the derived formulae utilizing the Haar wavelet transform as an example due to their locality and low computational requirements; however, a general formulation for wavelets can be derived, as well. Our exceptional results are attributed to the accuracy, well-posedness, and linearity of the derived equations, and the inherent nature of wavelets, making them very effective in signal localization.

In the next two sections, we explain two application areas for super resolution reconstruction, using motion estimation; namely, compressed video and pansharpening.

2.5 Super Resolution of Compressed Video

As we have mentioned before in Section 2.3, the goal of video compression is to reduce the size of a video file. To this goal, the redundancy in the frames should be revealed, which is widely performed by decomposing the frames into a transform domain (e.g. DCT, DWT), and encoding the uncorrelated data, instead of the redundant original frames. However, quantization of the transform coefficients (e.g. DCT) in coding, causes compression artifacts such as blocking and ringing. Therefore, sampling these coefficients before compression, leads to more bits to encode each coefficient, which reduces the quantization error. This pre-processing step of down-sampling needs to be compensated at the decoder by a post-processing step of up-sampling. Nevertheless, since down-sampling discards the high frequency information of the frames, the straigthforward solution of general interpolation techniques are not sufficient for a high quality result at the decoder side. Therefore, super resolution methods come in handy for post-processing of compressed video. These methods have the advantage of both decreasing the complexity of encoder, and reducing the bit rate [133]. Figure 2.7 shows the flowchart for the use of super resolution with downsampling-based video compression.

One of the pioneering methods in literature which performs super resolution after decompression is proposed by Segall *et al.* [134]. In this work, registration, interpolation, restoration, and post-processing problems are jointly solved using the decompressed low resolution image sequence in a Bayesian framework. Authors incorporate the compression process into the original super resolution observation model by first defining the relationship between the original HR frame, $\mathbf{x}$, and observed LR frame, $\mathbf{y}$, by:

$$\mathbf{y}_i = \mathbf{\Lambda}\mathbf{H}\mathbf{x}_i, \tag{2.51}$$

32

for $i = 1, 2, 3...$, where i shows the number of frames. Here, $\mathbf{\Lambda}$ stands for the downsampling matrix, $\mathbf{H}$ shows the filtering of the ith HR image. Later, the relationship between frames within the HR sequence is defined by using the displacements between these images:

$$\mathbf{x}_i = \mathbf{C}(v_{i,j})\mathbf{x}_j + \mathbf{r}_{i,j}, \tag{2.52}$$

where $\mathbf{C}(v_{i,j})$ is the matrix that maps jth frame to ith frame, $(v_{i,j})$ is the displacements vector, and $\mathbf{r}$ stands for the registration error.
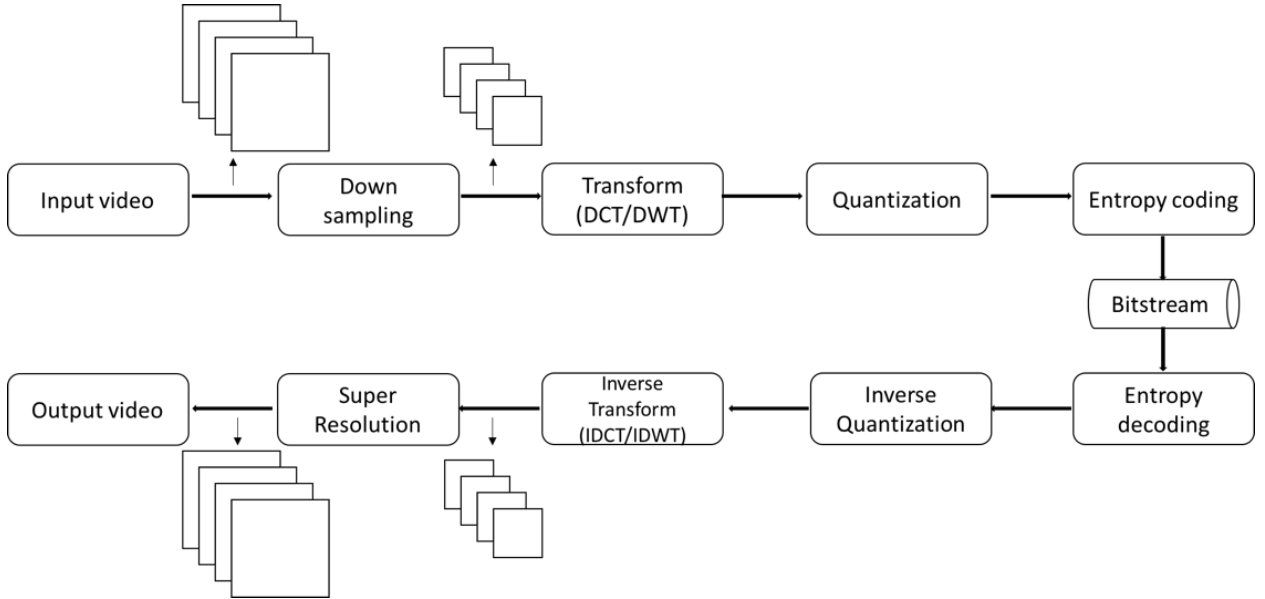
Figure 2.7: Super resolution of downsampling-based compressed video.

Using the motion compensation framework, the relationship between the compressed LR frame, $\mathbf{g}$, and the observed LR frame, $\mathbf{y}$, is defined as follows:

$$\mathbf{g}_i = \mathbf{T}^{-1}\mathbf{Q}[\mathbf{T}(\mathbf{y}_i - M(\mathbf{g}_j, v_{i,j}))] + M(\mathbf{g}_j, v_{i,j}), \tag{2.53}$$

33

where $\mathbf{T}$ and $\mathbf{T}^{-1}$ show the forward and inverse transforms, $\mathbf{Q}$ stands for the quantization, $M(\mathbf{g}_j, v_{i,j})$ shows motion compensation using jth compressed frame, and the motion vector that maps jth frame to ith frame.

Finally, by combining equations in (2.51), (2.52), and (2.53), the relationship between the HR frame, $\mathbf{x}$, and the compressed LR frames, $\mathbf{g}$, is defined as follows:

$$\mathbf{g}_i = \mathbf{\Lambda} \mathbf{H} \mathbf{C}(v_{i,j}) \mathbf{x}_j + \mathbf{e}_{i,j}, \tag{2.54}$$

where $\mathbf{e}$ stands for the error caused by compression, registration, and acquisition. This observation model is used in a Bayesian framework to solve registration and super resolution problems.

The relationship between the unknown HR and the compressed LR frames is employed and modified later by Belekos *et al.* [135] to include a new multichannel image prior model; and Zhang *et al.* [10] where individual frames are initially super resolved in a Bayesian framework, and are later used in a multiframe method.

Another group of super resolution of compressed video methods focus on mixed-resolution encoding which is demonstrated in Fig. 2.8. These methods encode some of the frames in full resolution, which are called the key frames, and the rest of the frames are downsampled, which are named as the non-key frames. The non-key frames are later upsampled in the decoder side, using a super resolution method. In their seminal work, Brandi *et al.* [133] propose a video coding scheme to encode I-frames (i.e., key frames) in full resolution, and downsample the P and B frames (i.e., non-key frames), so that in the decoder side, non-key frames are upsampled using information from the key frames. Later, this framework is employed by many researchers, adapting the number of key/non-key frames, and the super resolution method employed. More details on these studies are provided in Section 3.4.
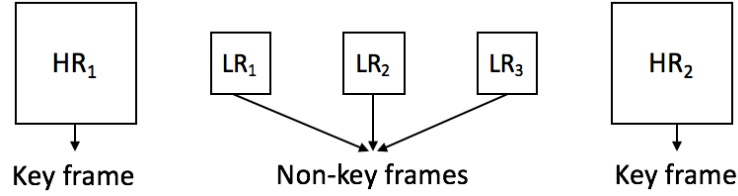
Figure 2.8: Mixed-resolution encoding for super resolution of compressed video.

Learning-based methods are also used in this area, which also will be explored in Section 3.4. However, since the performance of these methods are mostly based on the similarity between the LR inputs and the images in the database, these methods have the drawback of adding more noise to the reconstruction.

The goal of all downsampling based compressed video methods is to reconstruct the high frequency information, lost during downsampling and compression processes. The methods explained so far do not take advantage of the transform domain coefficients. As we have mentioned before, video coding methods tend towards wavelet-based encoding, which implies a wavelet-based super resolution is more reasonable than a spatial super resolution technique, in reconstruction of compressed video. Therefore, we propose a new framework for this problem, by combining the downsampling and transform steps shown in Fig. 2.7. Instead of downsampling the input HR frames, and decomposing the downsampled frames later into a transform domain, we first perform discrete wavelet transform on the input HR frames. Later, the low-pass subbands of these transformed frames, and the motion vectors estimated using the original HR images, are quantized and encoded. Since our devised technique for SR directly operates on the low-pass wavelet subbands to restore the high frequency information, it is well-suited for post-processing the downsampled compressed video. In the decoder side, we recover the high-pass subbands, and later perform inverse DWT. We employ our motion estimation method in Chapter 5 to estimate the motion vectors, and modify our super resolution method in Chapter 6 to an Iterative Back Projection (IBP) scheme, in order to

reconstruct the high-pass subbands.

We devote the next section to explain the pansharpening problem, and the use of super resolution in enhancing the resolution of pansharpened images.

2.6   Super Resolution of Pansharpened Images

In remote sensing imaging, the spatial resolution indicates the ground area captured in one pixel; while spectral resolution denotes the electromagnetic bandwidth of the signals [136]. Multispectral sensors capture a few spectral bands, and typically provide low-spatial-high-spectral resolution for the multispectral (MS) volume, and high-spatial-low-spectral resolution for the panchromatic (PAN) images; where PAN images have reflectance information for a large range of wavelengths. Low-spatial resolution of MS bands is often due to technological limitations inherent in satellite sensors. However, numerous remote sensing applications related to land-cover management, environmental monitoring, weather forecasting, and map updating require high-spatial-high-spectral resolution MS images. In order to obtain high spatial resolution MS images, a large body of research is devoted to fuse information of MS and PAN bands, which is called pansharpening. Figure 2.9 demonstrates an example of low-spatial-high-spectral resolution MS bands, high-spatial-low-spectral resolution Panchromatic image, and high-spatial-high-spectral MS bands after pansharpening, respectively.

Based on the classification by Amro *et al.* [136], pansharpening methods can be categorized into five groups as component substitution, relative spectral contribution, high-frequency injection, image statistics based, and multiresolution, where the goal is to fuse high frequency information of the PAN images into MS bands in order to increase resolution.

<center>a   b   c</center>
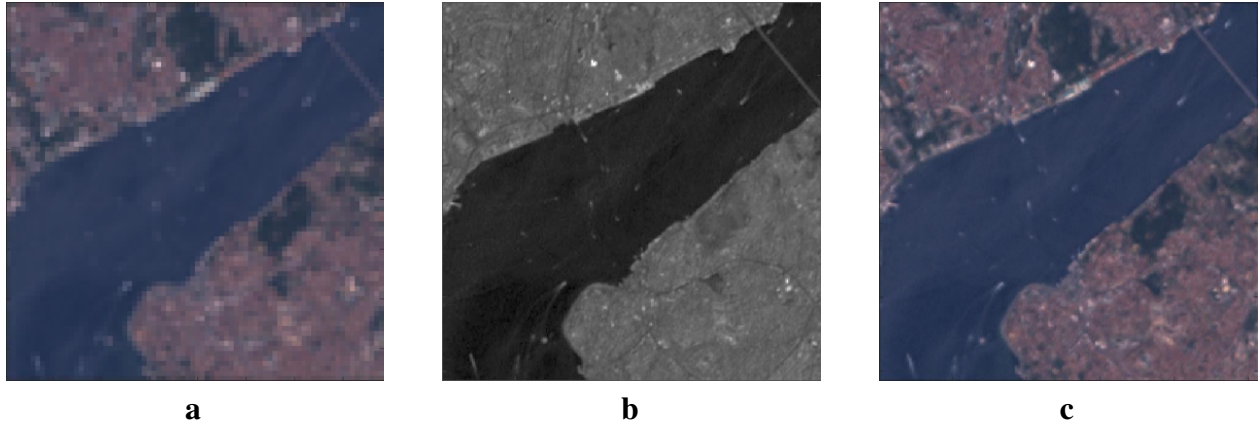
Figure 2.9: Pansharpening example **a** Low-spatial-high-spectral resolution MS bands **b** High-spatial-low-spectral resolution Panchromatic band **c** High-spatial-high-spectral resolution MS bands after pansharpening.

The first step of most pansharpening methods is to upsample the MS bands to the size of PAN image. Component substitution pansharpening methods transform these upsampled MS bands, perform histogram matching on MS bands and PAN images to substitute components of MS bands with the PAN images. Finally, backward transform is performed on the substituted components. Examples of this group include Intensity-Hue-Saturation (IHS) pansharpening [137], principal component substitution [138], and Gram-Schmidt spectral sharpening [139], which will be explained in Section 3.5.

Relative spectral contribution pansharpening methods, on the other hand, employ a linear combination of bands instead of using substitution after the histogram matching step. Brovey transform [140] is an example of relative spectral contribution pansharpening.

The third group of pansharpening methods perform high-frequency injection. Methods in this group, as in the previous two groups, first upsample the MS bands. Later, a low-pass filter is applied to the PAN image, and the difference between the original and filtered PAN images are found to be used as the high-frequency information. This high-frequency information is added to

the MS bands to obtain the pansharpened MS bands. High Pass Modulation method [141] is an example of this group.

Image statistics based models use the statistical relationship between each band of the MS and PAN images (e.g. Bayesian based techniques [142] and Price's method [143]), and use generally Bayesian approach to find the high-spatial-high-spectral MS bands.

The final group of multiresolution methods can also be categorized into two sets, as Laplacian Pyramid and wavelet-based. Laplacian pyramid based methods (e.g. [144]), as in the previous groups, first upsample the MS bands, find the Laplacian pyramid for the PAN image, and add the details from the pyramid to the MS bands after determining weights for the Laplacian pyramid coefficients. Wavelet-based methods (e.g. [145]), on the other hand, after upsampling the MS bands, perform forward transform on MS and PAN images, apply fusion on the coefficients in the transform domain, and finally perform inverse transform in order to achieve pansharpened images.

Based on the comparisons by Alparone *et al.* [146] and Bovolo *et al.* [147], a wavelet-based method called Additive Wavelet Luminance Proportional (AWLP) [148] is a state-of-the-art pansharpening technique, which implies the effective use of wavelets in pansharpening.

Pansharpening methods, however, do not consider the temporal information captured by the sensors. Multiframe SR methods as discussed before, on the other hand, fuse a sequence of degraded or aliased low resolution (LR) images of the same scene taken at different times, from different view points or by different sensors, to obtain a high resolution (HR) image. Fig. 2.10 shows an example of multispectral images taken at different times by the same sensor (i.e., Landsat 7 ETM+).

Figure 2.10: Landsat 7 ETM+ multispectral images taken at different dates (shown in green areas) around Sea of Marmara, Turkey (Images courtesy of USGS Glovis).

In order to take advantage of the temporal information in the multispectral images, Li *et al.* [45] propose a wavelet-based multiframe SR technique, where temporal information of low-spatial-resolution MS images are used to increase their spatial resolution, as opposed to the pansharpening methods, where PAN images are employed instead of temporal data. The authors propose using an SR method based on MAP approach with a universal Hidden Markov Tree model as a preprocessing step for multispectral image classification, where SR is applied band-by-band to multispectral images captured on different dates. However, although their method uses the temporal information, it does not take advantage of the high spatial resolution PAN image available with most multispectral sensors.

In this dissertation, we propose a wavelet-based multiframe SR method that takes advantage of both temporal and spatial information captured by multispectral sensors, in order to obtain a higher spatial resolution MS image which exceeds the spatial resolution of the available PAN image while keeping the high spectral resolution. To this end, we first apply a state-of-the-art pansharpening

method to a set of multispectral and corresponding PAN images taken at different times, to obtain high spatial resolution MS images. We then use our proposed wavelet-based approach for SR, band-by-band on the pansharpened MS images, in order to achieve a higher spatial resolution MS image. For the proposed SR method, we assume that pansharpened MS images (i.e., LR) correspond to the approximation coefficients of the first level discrete wavelet transform (DWT) of an unknown higher spatial resolution MS image (i.e., HR). We then solve the SR problem in a modified Iterated Back Projection (IBP) manner, as will be explained in Chapter 7. We demonstrate our results with comparisons, on Landsat 7 ETM+ datasets.

In the next chapter, we will present a literature review on the subjects explained so far.

CHAPTER 3: LITERATURE REVIEW

This chapter provides a survey on image registration, motion compensated temporal filtering, super resolution, super resolution of compressed video, and pansharpening for multispectral images, respectively.

3.1   Image Registration

We will explore subpixel image registration techniques in this section, dividing into spatial and frequency domain categories.

We will first investigate the spatial domain methods, in a chronological order. As an early spatial domain algorithm, Keren *et al.* [49] obtain subpixel accuracy by minimizing an error function which depends on the overlapping part of the reference and sensed images, using Taylor series expansion. In order to refrain from feature detection and tracking problems in image registration, Horn *et al.* [13] employ the so called "direct" methods where time derivative of brightness is used to find motion. Later, Davis *et al.* [149] analyze piece-wise linear interpolators, and present a reduced bias algorithm which takes advantage of the nearly linear relation between image displacement and bias that results from small displacements. The correlation coefficient function is maximized by Karybali *et al.* [150] in a closed-form solution for subpixel translation; whereas, a modified version of the correlation coefficient is maximized by Evangelidis and Psarakis [50] by both the forward additive approach and the inverse compositional method. More recently, feature extraction methods are introduced into spatial domain methods. While Ma *et al.* [59] select the control points in a preregistration process based on the scale invariant feature transform (SIFT), and subdivide the preliminary registered image into chips of pixels and match each chip with a cor-

responding one from the reference image utilizing normalized cross correlation; Gonçalvez *et al.* obtain subpixel accuracy integrating histogram-based image segmentation to registration problem of satellite images; and later extend their method to use SIFT and outlier removal [151]. Furthermore, Tzimiropoulos *et al.* [152] extend gradient correlation by modeling the dominant singular vectors of the 2-D gradient correlation matrix with a generic kernel, for pure translational subpixel shifts. For the next generation optical flow algorithms that can handle complex scenes, Baker et al. [153] provide a public dataset and evaluation methods. Finally, Ma *et al.* [154] propose a feature matching method that removes outlier features using an ML framework, and solve the registration problem with the expectation maximization technique.

We can summarize the Fourier domain subpixel registration methods as follows. The seminal work by Kuglin [76] for phase correlation method is extended for subpixel accuracy by several researchers. For example, Stone *et al.* [53] obtain subpixel accuracy by detecting and removing the Fourier components which become unreliable due to aliasing; Foroosh *et al.* [14] demonstrate that phase correlation peaks under subpixel translation are polyphase decomposition of a filtered unit impulse, which are centered at the correspondence; and Hoge *et al.* [155] use low complexity subspace identification. While Takita *et al.* [156] utilize a phase only correlation function to fit the analytical model of the correlation peak to 2D data; Balci and Foroosh [17–20] present solutions that work directly in the Fourier domain. More examples of methods that use phase correlation include, Vandewalle *et al.* [54] who perform registration based on low-frequency, aliasing-free parts of LR images; and Chen and Yap [51] who solve a set of nonlinear equations by developing a phase relationship between images. While Guizar *et al.* [52] present three new approaches which use nonlinear optimization with discrete Fourier transforms; Tzimiropoulos *et al.* [157] replace image functions with complex gray-level edge maps, and find motion parameters by normalized gradient correlation.

The earliest methods related to our work are based on image pyramids, with the aim of reducing

42

computational time and avoiding local extrema. In the coarse-to-fine scheme of pyramids, motion parameters are first estimated on the higher levels, and then refined in the lower levels with smaller search areas. In this scheme, Szeliski and Coughlan [158] use spline representations of motion field to deal with local deformations; and, Thévenaz *et al.* [56] minimize mean square intensity difference between reference and a set of sensed images using a modified version of Marquardt-Levenberg algorithm. While Chen *et al.* [60] maximize mutual information for registration; Kim *et al.* [159] apply Canny edge operator for a faster algorithm. Moreover, Qin *et al.* [160] estimate the displacements utilizing Taylor series expansion along with a coarse-to-fine Gaussian pyramid. Finally, local regions of interest in images are registered by estimating deformation parameters by Zhou *et al.* [161]; and automatic image registration of remotely sensed images is performed by using SIFT and mutual information by Gong *et al.* [162].

Lastly, we can outline the wavelet-based methods as follows. As mentioned in Section 2.2, most wavelet-domain methods use wavelet decomposition in order to extract features. For example, in an early wavelet domain method, Djamdji *et al.* [28] find the correspondence between the features extracted by finding the maxima of decomposition. Later, Turcajova and Kautsky [57] use separable fast discrete wavelet transform with normalized local cross correlation matching based on least square fit, and demonstrate the advantage of spline biorthogonal and Haar wavelets. In other examples of feature extraction methods, Le Moigne *et al.* [25] utilize maxima of Daubechies wavelet for correlation based registration; and Wu and Chung [163] utilize mutual information and sum of differences with wavelet pyramids. Later, a combination of feature-based and area-based registration, using wavelet-based features and relaxation based matching techniques is proposed by Hong and Zhang [26]. More recently, Wong and Clausi [23] employ an iterative wavelet-based feature matching method using a modified Geman-McClure M-estimation; and Alam *et al.* [164] utilize approximate coefficients of curvelets with a conditional entropy-based objective function.

3.2 Motion Compensated Temporal Filtering

In this section, we will explore the MCTF methods, based on the categorization provided in Section 2.3, which takes the order of decompositions (i.e., temporal and spatial) into account.

We can summarize the MCTF techniques which perform temporal decomposition before a spatial transform as follows. In an early method, Kim *et al.* [42] propose a 3-D extension of set partitioning in hierarchical trees (3D-SPIHT) for a low bit-rate embedded video coding scheme; moreover, both Secker and Taubman [165], and Pesquest-Popescu and Bottreau [166] use lifting formulation of three dimensional temporal wavelet decomposition for motion compensated video compression. More recently, Xiong *et al.* [39] extend spatiotemporal subband transform to in-scale motion compensation in order to exploit the temporal and cross-resolution correlations simultaneously, by predicting low-pass subbands from next lower resolution and high-pass subbands from neighboring frames in the same resolution layer. In order to achieve more accurate motion data, Esche *et al.* [167] propose an interpolation method for motion information per pixel using block based motion data. Furthermore, Chen and Liu [168] use an adaptive Lagrange multiplier selection model in rate-distortion optimization for motion estimation; and Rüfenacht *et al.* [169] anchor motion fields at reference frames instead of target frames to resolve folding ambiguities in the vicinity of motion discontinuities.

Even though methods in $t + 2D$ scheme have good performance, they have drifting and operational mismatch problems [39]. Therefore, performing spatial transform before temporal decomposition is introduced to overcome these drawbacks. However, since complete DWT is shift variant, in order to achieve in-band (i.e., directly in the wavelet domain) ME/MC, several methods are proposed to tackle this problem by redundancy. While Park and Kim [41] propose a low-band-shift method by constructing a wavelet tree by shifting low-band subband in each level for horizontal, vertical, and diagonal directions for one pixel and performing downsampling; Van der Auwera *et al.* [40]

use a bottom-up prediction algorithm for a bottom-up overcomplete discrete wavelet transform (ODWT). More recently, Andreopoulos *et al.* [38] define a complete to overcomplete discrete wavelet transform (CODWT), which avoids inverse DWT generally used to obtain ODWT. This method is later accelerated by Liu and Ngan [105] with the of use partial distortion search and anisotropic double cross search algorithms. Finally, Amiot *et al.* [43] demonstrate the usage of MCTF for denoising, using dual-tree complex wavelet (DT-CW) coefficients.

3.3 Super Resolution Reconstruction

We will review super resolution reconstruction methods in this section, by dividing them into spatial and frequency domain. We will first investigate the literature in Fourier domain, later go into detail of spatial domain methods; and finally explore wavelet based algorithms.

We can summarize the Fourier domain techniques as follows. In their seminal work, Tsai and Huang [3] exploit the relationship between Continuous Fourier Transform (CFT) of the unknown HR scene and Discrete Fourier Transform (DFT) of the shifted and sampled LR images. Later, Gilboa *et al.* [170] devise a method called forward-and-backward (FAB) adaptive diffusion that enhances, sharpens, and denoises image features simultaneously; and Vandewalle *et al.* [83] perform joint registration and reconstruction in the Fourier domain using multiple aliased images. More recently, in order to achieve SR using a single image, Rivenson *et al.* [171] utilize double random phase encoding in the imaging process; and Zhang and Cham [172] use Fourier transform within a learning based framework where DCT coefficients are inferred instead of pixel intensities. Since Fourier based methods cannot handle real world applications efficiently, they are not widely employed.

Spatial domain methods, on the other hand, can more easily be adapted for real world scenarios.

We can divide these methods into three groups in general, which use interpolation, regularization, and machine learning, that can also be employed in combinations. As an early example of interpolation based methods, Irani and Peleg [65] update the HR estimate by iteratively projecting the difference between the approximation and observed LR images back to HR grid. Later, Papoulis sampling theorem [173] is employed by Ur and Gross [174] in order to achieve a well-posed solution to merge LR images; and by Foroosh and Chellappa [175] to fuse nonuniform samples of multiple channels. Moreover, Zhou *et al.* [64] utilize multi-surface fitting.

Regularization, both in spatial and frequency domains, is widely used in SR methods, to overcome the ill-posed nature of the problem. We can summarize the spatial domain regularization based methods as follows. In an early example, Lorette *et al.* [176] investigate the contradiction between multichannel super resolution and regularization within the adaptive regularization framework. Later, Elad and Hel-Or [177] cast the SR problem into separate steps of deblurring and fusion within the ML scheme. While Tipping and Bishop [79] marginalize the unknown HR image in the Bayesian fashion; Pickup *et al.* [80] propose to marginalize the unknown motion parameters instead. In order to avoid the effects of motion estimation errors in SR, Shen *et al.* [84] simultaneously solve registration, segmentation, and SR problems in a MAP-based method, where a cyclic coordinate descent optimization is employed; and Babacan *et al.* [178] perform joint estimation of motion parameters and HR image in the Bayesian framework, where the unknowns are modeled in a stochastic fashion. As examples on usage of different image priors, Farsiu *et al.* [62] utilize a bilateral prior, which is minimized by L_1 norm; and Belekos *et al.* [135] incorporate a multichannel image prior to the MAP-based SR problem. Marquina and Osher [179] employ total variation constraint in a convolutional model, and solve the problem using Bregman iteration; whereas, Yuan *et al.* [75] constrain the SR process by using a spatially weighted TV model for different image regions. Finally, Huang *et al.* [180] propose to maximize the correlation between the local neighborhoods of HR and LR images.

A large body of research in SR focuses on learning based methods; which can employ feature pyramids, belief network, manifold learning, dictionary learning, compressive sensing, and neural networks [5]. Literature on multiframe learning-based SR include, Glasner *et al.* [181] who combine multiframe SR and example-based approaches based on the assumption that patches in natural images recur many times inside the image; Kappeler *et al.* [182] who propose a video super resolution method with a convolutional neural network (CNN) that is trained on both spatial and temporal dimensions of videos; and Huang *et al.* [81] who use a bidirectional recurrent convolutional network to model the temporal dependencies between video frames. On the other hand, most of the literature based on learning is devoted to single image SR methods. Yang *et al.* [108] use a compressed sensing framework, where correspondence of sparse representations between LR and HR patches are searched. Zhang *et al.* [183] employ clustering and supervised neighbor embedding. Johnson *et al.* [184] use training feed-forward networks using per-pixel loss together with perceptual loss functions. In order to recover the textual details in scenes, Huang *et al.* [185] employ a self-similarity based method where the internal patch search space is enlarged using geometric variations. Yang and Yang [186] learn priors seperately for several subspaces of the feature space. Timofte *et al.* [187] propose a fast method by using sparse learned dictionaries, global collaborative coding, and anchored neighborhood embedding; which is later combined with Yang's method [186], in [188]. Finally, the most recent research focuses on deep CNN framework. Kim *et al.* [189] propose a very deep CNN to achieve contextual information of large regions; Dong *et al.* [190] use deep CNNs to learn the mapping between LR and HR images; Dahl *et al.* [191] present a deep CNN architecture, in order to avoid the need of strong prior information required by conventional SR methods, to achieve HR images with high upscaling factors; and finally, Ledig *et al.* [192] present a method called a generative adversarial network (GAN), which aims at recontructing the details in higher magnification factors. The authors use a loss function that includes adversarial and content loss which helps to recover natural-looking images for $4\times$ downsampled images.

We will summarize wavelet based methods categorizing into interpolation, regularization, and learning based groups, as well. Interpolation based wavelet-domain SR approaches can be summarized as follows. In order to reduce noise in SR methods, Robinson *et al.* [63] apply a combined Fourier-wavelet deconvolution and denoising algorithm to multiframe SR. Authors first produce a sharp and noisy image by fast Fourier based image restoration, then reduce noise by space invariant nonlinear wavelet thresholding. The need to invert large matrices in their method results in solving the problem in the Fourier domain. On the other hand, to reduce degradation artifacts such as blurring and the ringing effect, Temizel and Vlachos [127] utilize zero padding in the wavelet domain followed by cycle spinning. Their method adopts a simplified edge profile and linear regression for edge degradations. Furthermore, to preserve edges, Demirel and Anbarjafari [125] use stationary and discrete wavelet transforms together in an interpolation-based framework. However, even though better than conventional interpolation techniques, the latter two methods still lack sharp edges. Tong and Leung [89] utilize Taylor series expansion to approximate the high frequency information. Their method is constrained to use LR images which have specific translations (namely 1 pixel in horizontal and 1 pixel in vertical directions). Moreover, Nguyen and Milanfar [88], contrary to the conventional interpolation based methods, use the regularity and structure in the interlaced sampling of LR images. Even though, for 2D images, they utilize reshaping property of the Kronecker product, which only doubles the complexity for 1D, their method is based on conjugate gradient which is still time consuming. For deblurring, Chan *et al.* [193] derive iterative algorithms, which decompose HR image obtained from an iteration into different frequency components and add them to the next iteration. Their method utilizes wavelet thresholding for denoising, where high-frequency components are penalized, making their method dependent on accurate noise estimation. Moreover, Ji and Fermuller [29] handle image registration and reconstruction together, by first estimating the homographies between multiple images, then reconstructing the HR image in a wavelet-based iterative back-projection scheme; and, Chavez-Roman and Ponomaryov [128] propose an interpolation-based method using sparse representation

of wavelet subbands.

Regularization based wavelet domain methods include the following works. Zhao *et al.* [129] solve a constrained optimization problem utilizing wavelet domain Hidden Markov Tree (HMT) model for the prior knowledge problem, since HMT characterizes the statistics of real world images accurately. In order to suppress the artifacts left after employing their method, cycle spinning is used which leads to blurring as in other interpolation based methods. Mallat and Yu [130] compute linear inverse estimators based on different priors by calculating sparse mixing weights for image blocks. Authors employ L_1 norm minimization using regularity of signals in these blocks.

Finally, we can summarize the learning-based methods in wavelet-domain as follows. Jiji *et al.* [131] handle the problem of representing the relationship between LR-HR frames with training their dataset with HR images by learning from wavelet coefficients at finer scales, followed by regularization in a least squares manner; and Gajjar and Joshi [132] follow Jiji's method and employ discrete wavelet transform for training, where a cost function based on MAP estimation is optimized with gradient descent method, employing an Inhomogeneous Gaussian Markov random field prior. Dong *et al.* [126], to solve the problem of varying contents in different images or image patches, learn various sets of bases from a precollected dataset of example image patches, and select one set of bases adaptively to characterize the local sparse wavelet domain. Patel and Joshi [194] design application-specific wavelet basis for hyperspectral image SR. They first learn high frequency information, later refine the initial estimate by a sparsity-based regularization method. Lastly, Mandal and Sao [195] propose edge preserving constraints in sparse coding framework with patch based dictionary learning. However, these methods are all based on optimization which requires high computational cost.

3.4   Super Resolution of Compressed Video

We devote this chapter to review the studies on super resolution of compressed video methods.

A large amount of literature on super resolution of compressed video is based on Bayesian framework. Examples in this category include the works by Gunturk *et al.* [8], where authors use quantization, additive noise, and image prior information in a stochastic framework; and by Segall *et al.* [134], where super resolution and post-processing problems are solved simultaneously in a Bayesian fashion. Belekos *et al.* [135] incorporate a new multichannel image prior to the model defined by Segall *et al.* [134].

To cope with the problems of complex motion and motion estimation errors, Zhang *et al.* [10] extend the Bayesian framework, by first initializing the HR frames in the Bayesian framework, and then estimating the final HR frames in a multiframe fashion using a cost function to reject outliers. Moreover, Barreto *et al.* [196] incorporate segmentation into super resolution of compressed video problem, in order to categorize blocks in frames based on the motion and texture. The segmentation information is later employed in decision of downsampling and super resolution processes.

Another set of super resolution of compressed video methods focuses on mixed-resolution encoding. These methods encode some of the frames at full resolution (i.e., key frames), and downsample the others (i.e., non-key frames) to be reconstructed at the decoder side. In one of the examples of this group by Kondo and Toma [197], decoded HR images are first downsampled to the size LR frames, then by using motion estimation between these LR frames, the unknown HR images are synthesized by replacing the pixel values from the HR key frames. Brandi *et al.* [133], on the other hand, find motion estimation parameters by using blocks from band-pass frames, instead of the low-pass versions. When the motion parameters are established, the high frequency information of the key frames are added to the non-key frames for super resolution.

Learning-based methods are also utilized for super resolution of compressed video problem. Shen *et al.* [44] construct a training set, and search the high frequency information of each LR frame patch in this training set, by taking the compression artifacts into account. Song *et al.* [198] adopt the mixed resolution scheme. Authors perform motion estimation in a hierarchical block matching fashion on the LR frames, and depending on the motion compensation error, utilize either temporal super resolution, or learning-based spatial super resolution. Finally, Kappeler *et al.* [199] train a CNN on both spatial and temporal dimensions of compressed videos, and avoid the need of frame type, motion estimation, and quantization information from the encoder.

Finally, wavelets have also been used in this area. Matsuo *et al.* [200] utilize wavelet-based super resolution which optimizes the super resolution parameters adaptively, in the pre-processing step of encoding. Estimated parameters for super resolution process is later encoded with the down-sampled frames, and employed in super resolution after decoding.

3.5   Pansharpening of Multispectral Images

In this section, we will summarize the pansharpening literature.

One of the pioneering methods in component substitution group is the Intensity-Hue-Saturation (IHS) pansharpening [201] which utilizes the IHS color space. Later, this method is enhanced by Tu *et al.* [137] to decrease the spectral distortion. Principal component analysis (PCA) is also used in this group by Kwarteng and Chavez [202], where PCA is used to substitute information from PAN image to MS bands, with the first principal component containing the spatial data, and other principal components corresponding to spectral information. Shah *et al.* [203] later combine this method with a contourlet-based technique, and employ an adaptive PCA method to find the optimal principal components to be used in substitution. Finally, Gram-Schmidt (GS) spectral

sharpening [139] can be categorized in the component substitution group, where Gram-Schmidt method is used to reduce the redundancy in MS bands.

Examples in relative spectral contribution group can be summarized as follows. The Brovey transform [140] is one of the classical methods in this group, which uses chromaticity transform. In order to decrease the possible spectral distortion in pansharpening methods, a smoothing filter-based intensity modulation is employed by Liu [204].

One of the works in high-frequency injection group is proposed by Chavez [205], where high-frequency information is extracted from the PAN image using a high-pass filter and injected to the MS bands. High-frequency modulation method [141] is another example work in this group, where the injection of high-frequency information is performed using a modulation coefficient [136].

Another group of pansharpening methods focus on the statistical properties of MS bands and PAN images. One of the first works in this group is proposed by Price [143], where a linear model is used to obtain the statistical relationship between low-spatial resolution MS, high-spatial resolution PAN, and high-spatial resolution MS bands. In addition, Fasbender *et al.* [206] propose a pansharpening method based on the Bayesian framework, where the statistical relationship between MS bands and PAN images are used in a weighted scheme, in order to match the user needs.

Finally, wavelet-based methods include, but not limited to, the works by Zhou *et al.* [207] where Landsat TM MS bands and SPOT PAN image are decomposed into wavelet subbands, and later the approximation subbands of the MS bands together with the detail subbands of the PAN image are used in inverse wavelet transform to merge the bands; and Kim *et al.* [145] where an improved additive-wavelet fusion method is proposed using the à trous algorithm which does not decompose the MS image to preserve the radiometric data. Moreover, Otazu *et al.* [148] propose a wavelet-based method that incorporates the "physical electromagnetic spectrum responses of sensors", in order to suppress the artifacts; and Garzelli and Nencini [144] model the relationship between

the wavelet coefficients of MS bands and PAN images at a coarser resolution, and later use this relationship to estimate the high spatial resolution MS bands. Alparone *et al.* [146] compare several pansharpening methods and conclude that the multiresolution based ones and the methods that employ adaptive models for the injection of highpass details outperform all the others.

In the following five chapters, we will provide our methodologies to solve each particular problem discussed so far.

CHAPTER 4: IN-BAND IMAGE REGISTRATION

In this chapter, we present our devised method for in-band subpixel registration [208] [1]. We first start by deriving an explicit relationship between Haar wavelet subbands of reference and sensed images under translational shifts, in Section 4.1. Later, we demonstrate decoupling of rotation, scale, and translation in the wavelet domain, in Section 4.2, and provide methodologies to recover subpixel shift parameters. We present the experimental results and comparisons with the state-of-the-art techniques in Section 4.3; and finally provide a summary and concluding remarks on our method in Section 4.4.

4.1 Subpixel Shifts in the Haar Domain

We first derive mathematical expressions that define in-band (i.e., direct wavelet-domain) shifts of an image, which will be used later for general registration under a similarity transformation (i.e., scale, rotation, and translation) [46].

4.1.1 Notation for Subpixel Registration Method

Table 4.1 summarizes the notations used throughout this chapter.

Superscripts of A, H, V, D show the level of wavelet decomposition. Subscripts x and y show horizontal and vertical directions, respectively; and new stands for the calculated shifted coefficients.

---

[1] The content in this chapter was in part reproduced from the following article: Vildan Atalay Aydin, Hassan Foroosh, In-band subpixel registration of wavelet-encoded images from sparse coefficient, Signal, Image and Video Processing, Year 2017. The copyright form for this article is included in Appendix A.

Table 4.1: Notation for Subpixel Registration Method.

$I(x, y)$	Reference image
$J(x, y, \sigma, \theta, t_x, t_y)$	Sensed image to be registered to I
σ, θ, t_x, t_y	Transformation parameters to be estimated: scale, rotation angle, and shifts along the two axes, respectively
A, H, V, D	Wavelet transform approximation, horizontal, vertical, and diagonal detail coefficients, respectively
ℓ	Number of hypothetically added levels
s	Perceived integer shift of wavelet coefficients after the hypothetically added levels (ℓ)

4.1.2 In-band Shifts

Here, we demonstrate the derived explicit mathematical expressions for an in-band translation of a given image.

Let $I(x, y)$ be a $2^N \times 2^N$ image, where N is a positive integer. The Haar transform of this image consists of N levels, where level l holds approximation coefficient $A_{i,j}^l$ and horizontal, vertical and diagonal detail coefficients $H_{i,j}^l$, $V_{i,j}^l$, and $D_{i,j}^l$, respectively, with $l = 0, ..., N - 1$, $i = 0, ..., 2^l - 1$ and $j = 0, ..., 2^l - 1$.

Let,

$$
\begin{aligned}
X_{i,j}^l &= H_{i,j}^{l-1} + V_{i,j}^{l-1} + D_{i,j}^{l-1}, \\
Y_{i,j}^l &= -H_{i,j}^{l-1} + V_{i,j}^{l-1} - D_{i,j}^{l-1}, \\
Z_{i,j}^l &= H_{i,j}^{l-1} - V_{i,j}^{l-1} - D_{i,j}^{l-1}, \\
W_{i,j}^l &= -H_{i,j}^{l-1} - V_{i,j}^{l-1} + D_{i,j}^{l-1}.
\end{aligned}
\tag{4.1}
$$

Also, let $E_{i,j}^l$ be the difference between $A_{0,0}^0$ and $A_{i,j}^l$, then, $A_{i,j}^l = A_{0,0}^0 + E_{i,j}^l$.

The following formula shows the relationship between $E_{i,j}^l$ and its parent level $l-1$;

$$E_{i,j}^l = \begin{cases} E_{i/2,j/2}^{l-1} + X_{i/2,j/2}^l, & \text{i is even, j is even} \\[2mm] E_{i/2,\lfloor j/2 \rfloor}^{l-1} + Y_{i/2,\lfloor j/2 \rfloor}^l, & \text{i is even, j is odd} \\[2mm] E_{\lfloor i/2 \rfloor,j/2}^{l-1} + Z_{\lfloor i/2 \rfloor,j/2}^l, & \text{i is odd, j is even} \\[2mm] E_{\lfloor i/2 \rfloor,\lfloor j/2 \rfloor}^{l-1} + W_{\lfloor i/2 \rfloor,\lfloor j/2 \rfloor}^l, & \text{i is odd, j is odd} \\[2mm] 0, & i = j = l = 0. \end{cases} \tag{4.2}$$

Equation (4.2) shows that $E_{i,j}^l$, for all l, can be calculated by using only the detail coefficients of Haar transform iteratively, since $E_{0,0}^0 = 0$. We utilize $E_{i,j}^l$ to calculate the detail coefficients of the shifted image which implies that the shifting process is in-band.

We can categorize a translational shift for a 2D image into two groups for horizontal and vertical shifts where a diagonal shift can be modeled as a horizontal shift followed by a vertical one. Unlike the common approach of modeling subpixel shifts by integer shifts of some upsampled version of the given image, our method models subpixel shifts directly in terms of the original level coefficients.

Observation 3.1. Let Haar transform of the image $I(x,y)$ have N levels, with $I(x,y)$ at the Nth level. Upsampling an image is equivalent to adding levels to the bottom of the Haar transform, and setting the detail coefficients to zero while keeping the approximation coefficients equal to the ones already in the Nth level, $E_{i,j}^{N+\ell_0} = E_{\lfloor i/2^{\ell_0} \rfloor,\lfloor j/2^{\ell_0} \rfloor}^N$, where $0 \leq \ell_0 \leq \ell$.

Observation 3.2. Shifting upsampled image by an amount of s is equivalent to shifting the original image by an amount of $s/2^\ell$, where ℓ is the upsampling factor.
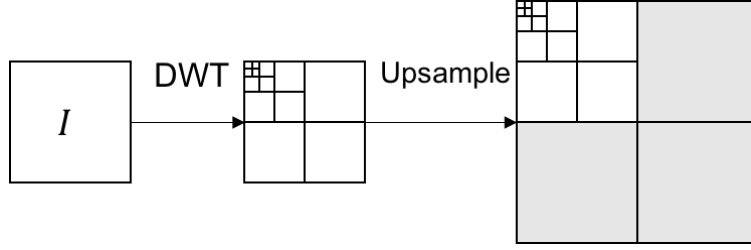
Figure 4.1: Upsampling illustration.

Fig. 4.1 demonstrates an example of the upsampling process described above for $\ell = 1$, which implies that only 1 level of zero detail coefficients are added. For upsampling, Haar Transform coefficients of the original reference image are utilized as approximation coefficients with more levels of detail coefficients which are set to be zero. Here, gray boxes demonstrate added zeros.

These observations allow us to shift a reference image for a subpixel amount without actually upsampling it, which saves memory, reduces computation, and avoids propagating interpolation errors.

Now, let $N' = N + \ell$ and $k = 1 + \ell, ..., N + \ell$. The horizontal detail coefficients of the shifted image in case of a horizontal translation are computed from the reference image coefficients by:

$$
\begin{aligned}
H_{i,j_{new}}^{N'-k} &= \sum_{m=2^{k-t-1}i}^{2^{k-t-1}(i+1)-1} \left(E_{\lfloor \frac{m\%2N}{2^{\ell}-1} \rfloor, \lfloor \frac{j_1\%2N}{2^{\ell}-1} \rfloor}^{N} + 2 \sum_{n=j_1+1}^{j_2-1} E_{\lfloor \frac{m\%2N}{2^{\ell}-1} \rfloor, \lfloor \frac{n\%2N}{2^{\ell}-1} \rfloor}^{N} \right. \\
&\quad \left. - 2 \sum_{n=j_2+1}^{j_3-1} E_{\lfloor \frac{m\%2N}{2^{\ell}-1} \rfloor, \lfloor \frac{n\%2N}{2^{\ell}-1} \rfloor}^{N} - E_{\lfloor \frac{m\%2N}{2^{\ell}-1} \rfloor, \lfloor \frac{j_3\%2N}{2^{\ell}-1} \rfloor}^{N} \right) \div (2 \times 4^{k-t-1}), \quad (4.3)
\end{aligned}
$$

where,

$$j_1 = 2^{k-t-1}j + \lfloor s_x/2^{t+1} \rfloor,$$

$$j_2 = 2^{k-t-2}(2j+1) + \lfloor s_x/2^{t+1} \rfloor,$$

$$j_3 = 2^{k-t-1}(j+1) + \lfloor s_x/2^{t+1} \rfloor.$$

Here, s_x is the horizontal shift amount at the $(N + \ell)$th level (where s and ℓ are calculated based on Observation 3.2), k is the reduction level, t is highest power of 2 by which the shift is divisible. For the subpixel shifts, $t = 0$, since the shift amount at the hypothetically added level is always an odd integer. t is essential to generalize the equation for even shifts. When $k = 1$, we set the coefficients utilizing j_2 in Eq. (4.3) to 0, since j_2 has a non-integer value. $H_{i,j_{new}}^{N'-k}$ for vertical shifts are obtained by interchanging the H's with V's, i's with j's and m's with n's in Eq. (4.3).

By examining Eq. (4.3), it can be seen that each level of horizontal detail coefficients of the shifted image can be calculated using the original levels of the reference image, since $E_{i,j}^N$ is calculated in Eq. (4.2) using only the detail coefficients in its parent levels.

Here, we only demonstrate the formulae for horizontal detail coefficients. Approximation, vertical and diagonal detail coefficients of the shifted image can be described in a similar manner.

4.2   Subpixel Registration

We first demonstrate that scale, rotation, and translation can be decoupled in the wavelet domain. This is similar to decoupling of rotation and translation in Fourier domain in magnitude and phase. We then describe the proposed method to solve the decoupled registration problem for the separated parameters.

Let us assume that sensed image is translated, rotated, and scaled with respect to a reference image, in that given order. Let also $p \in I$ and $q \in J$ be two points, where I and J are the reference and the sensed images, respectively. The point q can be defined in terms of the similarity transformation (scale, rotation, translation) and the point p in terms of homogeneous coordinates as follows:

$$\mathbf{q} \;=\; \mathbf{S\,R\,T\,p}, \tag{4.4}$$

$$
\begin{bmatrix} q_x \\ q_y \\ 1 \end{bmatrix}
=
\begin{bmatrix} \sigma & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} p_x \\ p_y \\ 1 \end{bmatrix},
$$

where we assume the same scale σ for both axes. Here, **S**, **R**, and **T** denote the scale, rotation, and translation matrices, and σ, θ, t_x and t_y denote the scale factor, the rotation angle in degrees, and the translations along the two axes, respectively. Although we assume the order of transformations as **S**, **R**, **T**, we first explain rotation recovery to demonstrate the decoupling in the wavelet domain. **Algorithm 1** shows the steps of the proposed in-band registration algorithm.

Algorithm 1 *In-band Registration for Similarity Transform*

- $\diamond$ *Input*: $I(x, y)$, $J(x, y, \sigma, \theta, t_x, t_y)$
- $\diamond$ *Objective*: Find similarity transform parameters
- $\diamond$ *Output*: $translation, rotation, scale$
- ▶ Generate wavelet coefficients of both images
- ▶ *Scale recovery* using curvature radius on coefficients
    - – Rescale sensed detail coefficients to the size of reference coefficients
- ▶ *Rotation recovery* using angle histograms of coefficients
    - – Rotate sensed detail coefficients for $-estimate$
- ▶ *Translation recovery* using in-band wavelet coefficient relationship (Section 4.1.2)

4.2.1 Rotation Recovery

Let $\mathbf{H}$ and $\mathbf{V}$ denote wavelet coefficients of the input images as in Section 4.1, where subscripts I and J stand for the images. Wavelet transform of Eq. (4.4), can be defined as follows:

$$\mathbf{H}_J = \sigma \cos(\theta)\mathbf{H}_I - \sigma \sin(\theta)\mathbf{V}_I,$$

$$\mathbf{V}_J = \sigma \sin(\theta)\mathbf{H}_I + \sigma \cos(\theta)\mathbf{V}_I. \tag{4.5}$$

Eq. (4.5) shows the relationship between the Haar wavelet coefficients of two images under similarity transformation, and indicate that the rotation and scale can be separated from translation, since translation parameters do not appear in these equations. In order to recover the rotation and scale independently, we also need to decouple σ and θ. One can see from Eq. (4.5) that dividing $\mathbf{V}_J$ by $\mathbf{H}_J$ eliminates the scale term, the result of which is an approximation to the slopes of local image gradients using Haar coefficients, since Haar coefficients can be viewed as an estimate of partial derivatives. To obtain an initial estimate of the rotation angle, we use wavelet thresholding [209] before finding the local slopes. This will both reduce noise and sparsify the coefficients. We then find an initial estimate of the rotation angle θ by maximizing the following cross-correlation:

$$\hat{\theta} = \arg\max_{\theta}(h_I \star h_J(\theta)), \tag{4.6}$$

where $\star$ denotes the cross-correlation, and h_I and h_J are the histogram of wavelet-coefficient slopes (HWS) for the thresholded coefficients, which we define as follows:

$$h_{\mathtt{img}} = \sum_{i=1}^{k} \arctan\left(\frac{\mathbf{V}_{\mathtt{img}}(i)}{\mathbf{H}_{\mathtt{img}}(i)}\right), \tag{4.7}$$

60

where k is the number of bins, and the subscript $\text{img} \in \{I, J\}$. We then refine the initial estimate $\hat{\theta}$, in the range $\hat{\theta} \pm 5°$ to get the best estimate $\hat{\theta}^*$:

$$\hat{\theta}^* = \arg\min_{\hat{\theta}} ||\mathbf{H}_J - \mathbf{R}\mathbf{H}_I||_2 + ||\mathbf{V}_J - \mathbf{R}\mathbf{V}_I||_2. \tag{4.8}$$

4.2.2 Scale Recovery

Since we already demonstrated that scale, rotation and translation can be decoupled in wavelet domain, we can perform scale estimation independently of rotation and translation. Let us assume that the two images have a scale ratio of σ. Then, the mean curvature radius calculated on thresholded wavelet coefficients would provide an accurate estimate of the scale factor:

$$\hat{\sigma} = \frac{1}{2} \left(\frac{\frac{1}{M_I} \sum_{i=1}^{M_I} \mathcal{R}(\mathbf{H}_I(i))}{\frac{1}{M_J} \sum_{i=1}^{M_J} \mathcal{R}(\mathbf{H}_J(i))} + \frac{\frac{1}{M_I} \sum_{i=1}^{M_I} \mathcal{R}(\mathbf{V}_I(i))}{\frac{1}{M_J} \sum_{i=1}^{M_J} \mathcal{R}(\mathbf{V}_J(i))} \right), \tag{4.9}$$

where $\mathcal{R}$ shows the radius of curvature.

4.2.3 Translation Recovery

Once the scale and rotation parameters are recovered and compensated for, the translations t_x and t_y along the two axes can be recovered independently by maximizing the following normalized

cross-correlation function:

$$
\begin{aligned}
\{\hat{t}_x, \hat{t}_y\} = \quad & \underset{t_x, t_y}{\arg\max} \\
& \left[\frac{\sum_{x,y} \left(\mathbf{H}_I(x + t_x, y + t_y) \mathbf{H}_J(x', y') \right)}{\sqrt{\sum_{x,y} \left(\mathbf{H}_I(x + t_x, y + t_y) \right)^2} \sqrt{\sum_{x,y} \left(\mathbf{H}_J(x', y') \right)^2}} \right. \\
& \left. + \frac{\sum_{x,y} \left(\mathbf{V}_I(x + t_x, y + t_y) \mathbf{V}_J(x', y') \right)}{\sqrt{\sum_{x,y} \left(\mathbf{V}_I(x + t_x, y + t_y) \right)^2} \sqrt{\sum_{x,y} \left(\mathbf{V}_J(x', y') \right)^2}} \right],
\end{aligned}
\tag{4.10}
$$

where $\mathbf{H}_I(x + t_x, y + t_y)$ and $\mathbf{V}_I(x + t_x, y + t_y)$ are the shifted versions of the reference detail coefficients (corresponding to $\mathbf{H}_{new}$ or $\mathbf{V}_{new}$ in the derivations of Section 4.1.2), calculated using Eq. (4.3) (or the equivalent for the vertical coefficients); and $\mathbf{H}_J(x', y')$ and $\mathbf{V}_J(x', y')$ are the sensed image detail coefficients after rotation and scale compensation.

Observation 3.2 implies that subpixel registration for wavelet-encoded images can be performed directly in the wavelet domain without requiring inverse transformation. Furthermore, if the encoded image is also compressed (e.g. only a sparse set of detail coefficients are available), one can still perform the registration. The latter could be for instance a case of compressed sensing imager based on Haar wavelet sampling basis. To maximize the cost function in Eq. (4.10), we use a branch and bound (BnB) algorithm, where split of rectangle areas in BnB are decided based on the two maximum cross correlations of four bounds.

Algorithm 2 demonstrates the main steps of the proposed method for translation recovery. Shifted horizontal/vertical detail coefficients for the updated bounds are calculated for a specified level (k) using Eq. (4.3) (similar equation for the vertical coefficients), followed by application of maximization of Eq. (4.10).

Algorithm 2 *Subpixel Shifts Estimation*

- ◇ *Input*: $I(x, y)$, $J(x', y', t_x, t_y)$ (scale and rotation corrected images)
- ◇ *Objective*: Find translational registration parameters
- ◇ *Output*: (t_x, t_y)
- ▶ Initialize bounds for subpixel shift estimates to $[-1, +1] \times [-1, +1]$.
- ▶ Do:
    - ○ Generate horizontal/vertical detail coefficients of shifted versions of the reference image $\mathbf{H}_I(x + t_{x_i}, y + t_{y_i})$ and $\mathbf{V}_I(x + t_{x_i}, y + t_{y_i})$, using Eq. (4.3) (similar equation for vertical), where t_{x_i} and t_{y_i} are the bounds at iteration i.
    - ○ Update bounds (reduce rectangles to half in size) based on the peak of cross correlation in Eq. (4.10) for detail coefficients of shifted reference image and sensed image.
  until maximum cross-correlation in Eq. (4.10) exceeds τ for an estimated bound, where τ is an accuracy measure (tolerance) for cross correlation.

When the algorithm converges within an ϵ distance to the true solution, it often starts osculating. So, as a modification to a general branch and bound method, we take the mid-point of osculations as the solution, which often happens to be the true solution.

The method requires the knowledge of $A_{0,0}^0$ for in-band shifts which may limit the approach to image sizes of $2^N \times 2^N$. However, the solution can be generalized to images with arbitrary sizes by simply applying the method to a subregion of size $2^N \times 2^N$ of the original images.

4.3   Experimental Results

To demonstrate the accuracy of our algorithm, we performed extensive experiments on both simulated and real data. In order to simulate reference and sensed images, a given high resolution image is shifted (using bicubic interpolation) and rotated, then both images are downsampled, which is a common technique employed in state-of-the-art literature [14], [54]. If different scale are assumed, then the sensed image is also scaled further. We performed thorough comparisons with state-of-the-art methods, which were given the same input images, and results were evaluated by

measuring alignment errors. Fig. 4.2 shows some of the standard test images together with the real data obtained from [1] and [2]. Captions for real data indicate the dataset and specific image names utilized as reference image.
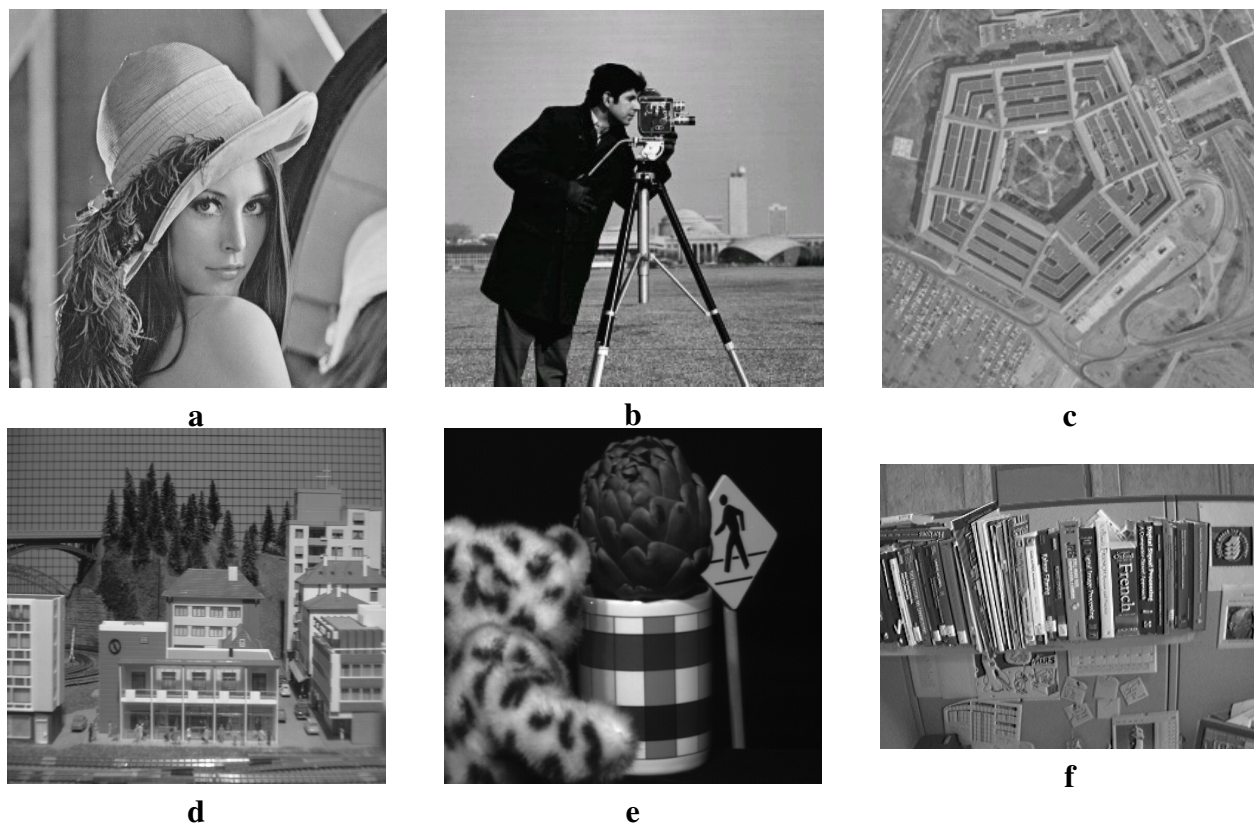
Figure 4.2: Example of simulated and real world images used for image registration experiments. **a** Lena **b** Cameraman **c** Pentagon **d** CIL - horizL0 [1] **e** Artichoke - 1 [1] **f** MDSP - Bookcase 1 [2].

4.3.1 Validation on Simulated Data

Here, we first performed experiments on translation, rotation and scale recovery separately. We then carried out tests for combination of transformations.

Table 4.2 summarizes some of the results for our translational method with simulated data, where the results are compared with the ground truth (GT) and other baseline methods by Keren *et al.* [49], Guizar *et al.* [52], Szeliski and Coughlan [158], in terms of estimated shifts, peak-signal-to-noise ratio (PSNR), and mean square error (MSE). The simulated images are generated by shifting the given reference image using bicubic interpolation, and downsampling the reference and shifted images. The results presented in Table 4.2 are simulated for a noise free case, and PSNR and MSE results for all methods are obtained by simulating the image generation process. Since the expressions derived in Section 4.1.2 are exact for any arbitrary shift that can be expressed as positive or negative integer powers of 2, in the noise-free case, exact or near-exact solutions can be achieved, which outperforms the state-of-the-art methods. For any other shift amount, we can get arbitrarily close within the closest integer power of 2, which when compared with the state-of-the-art, is still outstanding.

Table 4.3 shows the PSNR, MSE and computational time for our rotation method compared to the technique by Vandewalle *et al.* [54], averaged over 121 simulations. Although our technique can recover any rotation angle, since Vandewalle's method [54] recovers only angles in the range $[-30, 30]$, in order to be fair, we compared our results for every $0.5°$ in that range.

Table 4.2: Comparison of the proposed registration method with other baseline methods in estimated shifts, PSNR, and MSE.

Img.	Exact shift	Keren [49]			Guizar [52]			Szeliski [158]			Proposed		
		Estimate	PSNR	MSE	Estimate	PSNR	MSE	Estimate	PSNR	MSE	Estimate	PSNR	MSE
a	0.5, 0.5	0.4878, 0.5427	56.91	0.11	0.56, 0.53	50.05	0.56	0.5017, 0.5009	80.91	0	0.5, 0.5	Inf	0
	0.25, -0.125	0.2456, -0.1212	72.23	0.003	0.29, -0.16	53.10	0.28	0.2518, -0.1243	80.67	0	0.25, -0.125	Inf	0
	-0.375, -0.4	-0.3826, -0.4146	63.99	0.02	-0.42, -0.42	52.67	0.31	-0.3732, -0.3990	80.36	0	-0.375, -0.4023	82.59	0
	-0.625, 0.75	-0.6958, -0.8268	47.81	0.95	-0.70, 0.81	48.00	0.90	-0.6231, 0.7508	80.17	0	-0.625, 0.75	Inf	0
b	0.33, -0.33	0.3347, -0.3008	54.19	0.24	0.27, -0.33	45.91	1.61	0.3275, -0.3316	72.49	0.003	0.3281, -0.3438	60.74	0.05
	0.167, 0.5	0.1641, 0.6154	42.08	3.91	0.11, 0.55	44.78	2.10	0.1633, 0.4977	69.04	0.007	0.1719, 0.5	67.76	0.01
	-0.875, -0.33	-0.8639, -0.2986	52.58	0.35	-0.92, -0.33	48.44	0.91	-0.8783, -0.3316	70.51	0.005	-0.875, -0.3438	60.40	0.06
	-0.125, 0.67	-0.1309, 0.8230	39.53	7.01	-0.08, 0.75	42.94	3.20	-0.1277, 0.6695	72.62	0.003	-0.125, 0.6719	77.60	0.001

Table 4.3: Comparison of average PSNR and MSE for rotation recovery for 121 simulations.

Image	Vandewalle [54]			Proposed		
	PSNR	MSE	Time (s)	PSNR	MSE	Time (s)
a	32.83	6.59	0.16	42.94	1.75	0.49
b	37.53	4.09	0.16	43.53	1.81	0.49

We also ran our scale recovery method for 50 images with scale amounts $1/4, 1/2, 1, 2, 4$. All experiments returned the exact scale in under 0.09 seconds. Since wavelet transform downsamples images by 2 in every level, we can only recover scales that are multiples of 2.

Results obtained for combination of transformations can be seen in Tables 4.4 and 4.5. While Table 4.4 shows comparisons to Vandewalle's method for rotation and translation, Table 4.5 presents our results obtained for several combinations of scale, rotation and translation. These tables also confirm that our method is accurate and outperforms or at least matches state-of-the-art.

Table 4.4: Comparison of PSNR, MSE, and time for rotation and translation recovery.

Image	Exact (x, y, θ)	Vandewalle [54]				Proposed			
		Estimate	PSNR	MSE	Time (s)	Estimate	PSNR	MSE	Time (s)
a	$(0.5, -0.25, 20)$	$(0.8, -0.5, 19.8)$	23.2	16.4	0.1	$(0.5, -0.25, 20)$	25.2	13.07	5.55
b	$(-0.375, -0.375, -10)$	$(-0.337, -0.62, -10.2)$	21.17	19.97	0.09	$(-0.406, -0.375, -10)$	22.05	19.7	23.4
c	$(-0.4375, 0.875, -30)$	$(-0.64, 0.526, -30)$	21.01	19.02	0.09	$(-0.39, 0.875, -30.3)$	20.2	20.86	66

Table 4.5: Our results for scale, rotation and translation registration.

Img	Exact (x, y, θ, σ)	Results	
		Estimate	Time (s)
a	$(0.5, 0.25, -50, 2)$	$(0.5, 0.25, -50.1, 2)$	25.7
a	$(0.5, 0.25, 50, 2)$	$(0.5, 0.25, 49.8, 2)$	4.94
b	$(-0.25, 0.25, 10, 1/4)$	$(-0.28, 0.28, 10.2, 1/4)$	105.3
c	$(-0.5, -0.375, 30, 1/2)$	$(-0.5, -0.375, 30.1, 1/2)$	93.6
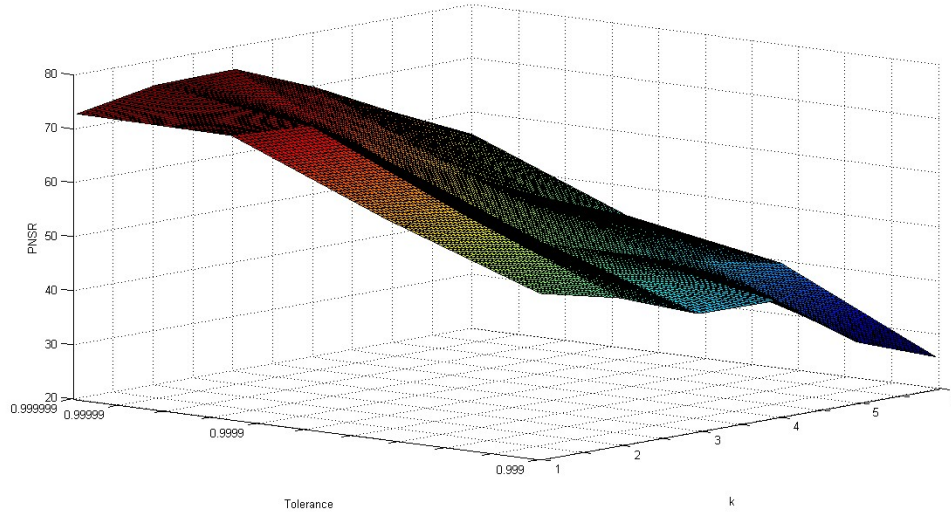
Figure 4.3: Comparison of k (x axis) and τ (y axis) with PSNR (z axis) for average of 50 images for GT shift of $(0.33, -0.33)$.

4.3.2 Optimal Parameters

In order to find the appropriate constants τ and k for translational shift, which are the measures of accuracy (tolerance for cross correlation function) and reduction level of Haar transform, respectively, and show the accuracy of the proposed method, we tested our algorithm with 50 simulated test images for shift amount $(0.33, -0.33)$. Results after removing the outliers (when a local maxima is reached) are shown in Fig. 4.3, where $PSNR = Inf$ is demonstrated as 100. As seen in the

67

figure, the constants τ and k can be adapted depending on the trade-off between time complexity and PSNR.

In case of the most general similarity transformation, k is decided based on the recovered scale $\hat{\sigma}$ by choosing $k = 1$ if $scale < 1$ or $k = \hat{\sigma} + 1$ otherwise.

4.3.3   Validation on Real Data

In order to ensure the accuracy of our method, real world images were also utilized as input. Results for real world examples (**d**, **e** and **f** in Fig. 4.2) including comparisons with the state-of-the-art methods by Evangelidis and Psarakis [50], and Vandewalle *et al.* [54] are summarized in Table 4.6. Since the GT for the used images is not known, the results are compared using PSNR and MSE as it is common practice in the literature. All methods are given the same input, where smaller image regions are used to adopt image sizes to work with our method as described in Section 4.2.3. As seen in Table 4.6, our method outperforms the baseline methods in real world examples in most cases as well.

Table 4.6: Comparison of our registration method with other methods for real world examples from [1] and [2] in PSNR and MSE.

Dataset	Reference img.	Sensed img.	Vandewalle [54]		Evangelidis [50]		Proposed	
			PSNR	MSE	PSNR	MSE	PSNR	MSE
Artichoke	1	2	26.88	11.59	31.5	6.78	31.8	6.72
Artichoke	27	28	26.86	11.17	42.08	1.93	31.06	6.92
CIL	HorizR0	HorizR1	24.02	13.2	12.4	50.4	24.73	12.66
CIL	VertR4	VertR5	20.66	21.57	22.2	18.46	20.75	22.5
MDSP Bookcase 1	2	3	26.58	11.90	12.52	60.31	25.10	14.11

4.3.4 The Effect of Noise and Sparseness

Our proposed approaches for scale and rotation estimation already suppress noise by hard wavelet thresholding. Therefore, here we discuss only noise in translation estimation. In Table 4.7, a comparison of the proposed method with Foroosh *et al.* [14], and Chen and Yap [51] under noisy conditions is presented. By adapting the tolerance value, τ, to be passed by the registration method, based on the level of noise and cross validation, very accurate shift values can be achieved, as can be seen from Table 4.7. It can be concluded that our method performs well in suppressing Gaussian noise, which also is superior to the state-of-the-art. In order to show the accuracy under noisy conditions, the proposed algorithm is tested for 50 images with 50 different shift amounts for each image, with Gaussian noise. Results, after removing outliers, are shown in Fig. 4.4 for average SNR with respect to τ and σ.

Since our method works entirely in-band (i.e., using only detail coefficients), the method is particularly applicable to wavelet-encoded imaging. Moreover, our approach can work with a sparse subset of coefficients, e.g. compressed sensing of wavelet-encoded images. Since our scale and rotation recovery methods already use sparse coefficients (i.e., hard-thresholded wavelet coefficients), we experimented on translational shifts under sparseness. We tested our method as the level of sparseness varied from 2% to 100% of detail coefficients, for several simulated images and different shifts. We then fitted a model to the average results to evaluate the trend which is shown in Fig. 4.5. It can be noticed that, in all cases, the worst registration PSNR when using only 2%-7% of detail coefficients was above 46dB. Beyond 50% sampled detail coefficients, the PSNR grows exponentially.
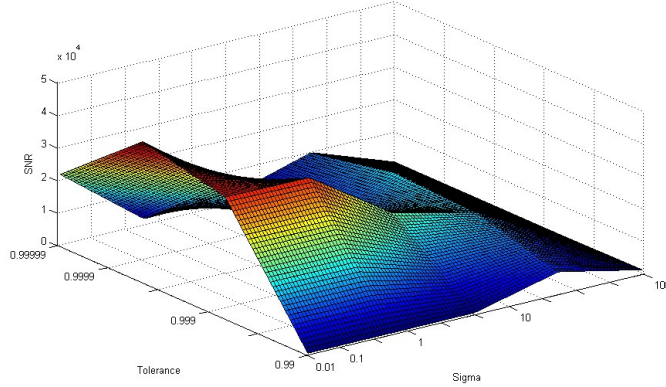
Figure 4.4: Average SNR (z axis) compared with changing τ (y axis) and σ (x axis), for horizontal axis.

Table 4.7: Comparison of registration results for noisy environments with "Pentagon" image for $(0.25, 0.75)$ shift.

SNR	Foroosh [14]	Chen [51]	Proposed
10 dB	0.38 0.65	0.29 0.68	0.25 0.75
20 dB	0.31 0.71	0.28 0.74	0.25 0.75
30 dB	0.30 0.73	0.27 0.74	0.25 0.75
40 dB	0.29 0.74	0.27 0.74	0.25 0.75

4.3.5   Computational Complexity and Convergence Rate

Time complexity of our method depends on in-band shifting, parameter selection, and the level of sparseness. In-band shifting method in Section 4.1.2, has a complexity of $O((L/2^{N-k+1})^2)$ for all $k = 1...N$, where L is size of the image (or a sparsified version). Parameter selection also affects the complexity since when τ is higher, the method attempts to match the images with higher accuracy, which would increase the run time. We provide running time of our method with comparisons in Tables 4.3, 4.4 and 4.5 on a machine with 2.7 GHz CPU and 8 GB RAM.

70
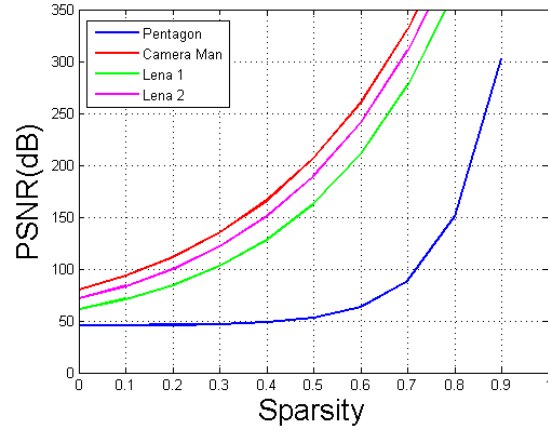
Figure 4.5: Average PSNR as a function of percentage of detail coefficients (level of sparsity) used to register for Pentagon, Cameraman, and two different shifts of Lena.

Fig. 4.6 demonstrates the convergence of our method to the global cross-correlation maximum for Lena image with GT $(0.5, 0.5)$ in blue circles, Pentagon image with GT $(0.25, 0.5)$ in green stars, and Cameraman image with GT $(0.33, -0.33)$ in red line. The convergence is visibly exponential and therefore we get a very rapid convergence to the solution.
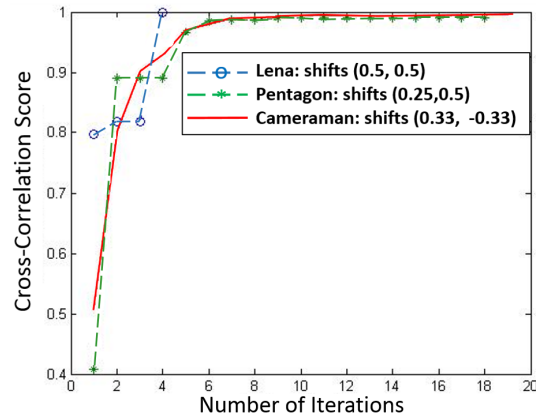
Figure 4.6: Examples illustrating the convergence to optimal cross-correlation as a function.

71

4.4 Conclusion

A subpixel registration technique for sparse Haar encoded images is demonstrated in this chapter. Only a sparse set of detail coefficients are sufficient to establish the cross-correlation between images for scale, rotation, and translation recovery. Our registration process is thus performed solely in-band, making the method capable of handling both in-band registration for wavelet-encoded imaging systems, and sparsely sensed data for a wavelet-based compressive sensing imager. Moreover, our method conveniently decouples scale, rotation and translation parameters, while exploiting Haar wavelet's important features, such as multiresolution representation and signal energy localization. Our method does not use image interpolation for estimating the registration parameters, since the exact set of in-band equations are derived for establishing the registration and fitting the parameters. Although the run time of our method is higher than compared methods, we achieve far better accuracy as a reasonable trade-off. Overall, our results show superior performance, and outperform the baseline methods in terms of accuracy and resilience to noise.

CHAPTER 5: MOTION COMPENSATED TEMPORAL FILTERING FOR LOW-BITRATE VIDEO CODING

We have shown, in the previous chapter, the effective use of wavelet subbands in image registration problem. In this chapter, we will reformulate the relationship between subbands, explained in Section 4.1.2, and investigate the effects of employing this relationship on motion estimation/motion compensation problems in a video coding framework (i.e., motion compensated temporal filtering (MCTF)) [210] [1]. The rest of this chapter is organized as follows. We first introduce our proposed solution to MCTF problem in Section 5.1. Later, we define the derived exact inter-subband relationships in Section 5.2, demonstrate the experimental results in Section 5.3, and finally provide a conclusion in Section 5.4.

5.1 Motion Compensated Temporal Filtering

In this section, we explain our proposed method for in-band MCTF, operating directly on DWT subbands.

The wavelet transform provides localization both in time and frequency; therefore, it is straightforward to use wavelets for MCTF. In order to perform ME/MC steps in MCTF, wavelet subbands of the transformed signal should be estimated. However, due to decimation and expansion operations of DWT, direct band-to-band estimation is generally not practical [41]. The proposed method overcomes this challenge by revealing the relationships between subbands of reference and target frames.

---

[1]The content in this chapter was in part reproduced from the following article: Vildan Atalay Aydin, Hassan Foroosh, Motion compensation using critically sampled dwt subbands for low-bitrate video coding, IEEE The International Conference on Image Processing, Year 2017. The copyright form for this article is included in Appendix B.
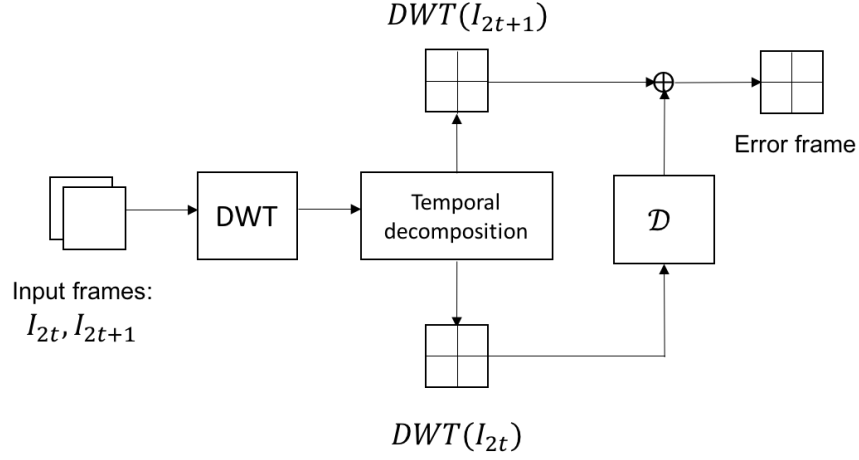
Figure 5.1: Proposed in-band MCTF model.

The proposed in-band MCTF method is demonstrated in Fig. 5.1, as in MCTF methods shown before in Fig. 2.4-b. Given a video sequence, first, DWT is performed on each frame for spatial decomposition, then a temporal decomposition is performed by splitting video frames into groups. ME/MC ($\mathcal{D}$ in Fig. 5.1) is performed by block matching, using reference frames ($DWT(I_{2t})$) to predict the target frames ($DWT(I_{2t+1})$). Employing the found motion vectors (MV), reference frames are mapped onto the target frames to generate error frames as in;

$$E_{2t} = \text{DWT}(I_{2t+1}) - \mathcal{D}(\text{DWT}(I_{2t})), \tag{5.1}$$

where E_{2t} stands for the error frame at time $2t$, which are then quantized and encoded/decoded by a wavelet codec, together with the MVs for the video coding application.

We employ Haar wavelet decomposition in spatial transform due to the benefits mentioned earlier. Since the method in Section 5.2 is accurate for any arbitrary subpixel translation defined as a multiple of 2^k, where k is the decomposition level, our method does not need interpolation for subpixel accuracy. A block matching method with unidirectional full search is used for ME/MC

74

steps which is a common method used for MCTF. Our cost function is based on squared error minimization using all subbands, as follows:

$$\{\hat{x}, \hat{y}\} = \arg\min_{x,y}\{(A - \hat{A})^2 + (H - \hat{H})^2 + (V - \hat{V})^2 + (D - \hat{D})^2\}, \tag{5.2}$$

where, $\{\hat{x}, \hat{y}\}$ are the estimated motion vectors, A, H, V, D denote the original target frame wavelet subbands, and $\hat{A}, \hat{H}, \hat{V}, \hat{D}$ are the predicted subbands for the same target image, using the method described in Section 5.2 and a reference frame.

5.2   Inter-subband Relationship

In-band (i.e., wavelet domain) shift method along with the related notation are provided in this section.

5.2.1   Notation for MCTF Method

Here, we provide the notations used throughout this chapter, in Table 5.1.

Table 5.1: Notation for MCTF Method.

I_t	Input video frame at time t
A, H, V, D	Haar wavelet transform approximation, horizontal, vertical, and diagonal subbands of input image, respectively
F, K, L	Matrices to be multiplied by approximation, horizontal, vertical, and diagonal DWT subbands, used for in-band shift of reference frame
ℓ	Number of hypothetically added levels in case of non-integer shifts
s	Integer shift amount after the hypothetically added levels (ℓ)

75

Bold letters in the following sections demonstrate matrices and vectors. Subscripts x and y indicate the horizontal and vertical directions, respectively. Finally, subscript k indicates the kth video frame, where $k = 1, 2, \ldots N$.

5.2.2   In-band Shifts

Our goal for the MCTF method described in Section 5.1 is to achieve ME/MC in the wavelet domain using DWT subbands, given a video frame sequence. For this purpose, wavelet subbands of the tranformed signal should be predicted using only DWT subbands of the reference frame. Therefore, we derive the relationship between the subbands of transformed and reference images, which can be described by in-band shift (i.e., in the wavelet domain) of the reference image subbands. Below, we derive the mathematical expressions which demonstrate these relationships.

Let $\mathbf{A}$, $\mathbf{H}$, $\mathbf{V}$, and $\mathbf{D}$ be the first level approximation, horizontal, vertical, and diagonal detail coefficients (i.e., subbands), respectively, of a $2D$ reference frame at time t, I_t, of size $2m \times 2n$, where m and n are positive integers. Since decimation operator in wavelet transform reduces the size of input frame by half in each direction for each subband, we require the frame sizes to be divisible by 2. Now, the $1st$ level subbands of translated frame in any direction (i.e., horizontal, vertical, diagonal) can be expressed in matrix form using the $1st$ level Haar transform subbands of reference frame as in the following equations:

$$
\begin{aligned}
\mathbf{A}_s &= \mathbf{F}_y\mathbf{A}\mathbf{F}_x + \mathbf{F}_y\mathbf{H}\mathbf{K}_x + \mathbf{K}_y\mathbf{V}\mathbf{F}_x + \mathbf{K}_y\mathbf{D}\mathbf{K}_x, \\
\mathbf{H}_s &= -\mathbf{F}_y\mathbf{A}\mathbf{K}_x + \mathbf{F}_y\mathbf{H}\mathbf{L}_x - \mathbf{K}_y\mathbf{V}\mathbf{K}_x + \mathbf{K}_y\mathbf{D}\mathbf{L}_x, \\
\mathbf{V}_s &= -\mathbf{K}_y\mathbf{A}\mathbf{F}_x - \mathbf{K}_y\mathbf{H}\mathbf{K}_x + \mathbf{L}_y\mathbf{V}\mathbf{F}_x + \mathbf{L}_y\mathbf{D}\mathbf{K}_x, \\
\mathbf{D}_s &= \mathbf{K}_y\mathbf{A}\mathbf{K}_x - \mathbf{K}_y\mathbf{H}\mathbf{L}_x - \mathbf{L}_y\mathbf{V}\mathbf{K}_x + \mathbf{L}_y\mathbf{D}\mathbf{L}_x.
\end{aligned}
\tag{5.3}
$$

As already mentioned in Section 5.2.1, $\mathbf{F}$, $\mathbf{K}$, and $\mathbf{L}$ stand for matrices to be multiplied by the lowpass and highpass subbands of the reference frame in order to perform in-band shift, where subscripts x and y indicate *horizontal* and *vertical* directions. $\mathbf{A}_s, \mathbf{H}_s, \mathbf{V}_s, \mathbf{D}_s$ are translated frame subbands (in any direction). The low/high-pass subbands of both reference and transformed frames are of size $m \times n$, $\mathbf{F}_x$, $\mathbf{K}_x$ and $\mathbf{L}_x$ are $n \times n$, whereas $\mathbf{F}_y$, $\mathbf{K}_y$ and $\mathbf{L}_y$ are $m \times m$.

Here, we first define the matrices for subpixel shift amounts. The algorithm to reach any shift amount using the subpixel relationship will be described later in this section.

For subpixel translation, contrary to the customary model of approximating a subpixel shift by up-sampling an image followed by an integer shift, our method models subpixel shift directly based on the original coefficients of the reference frame, without upsampling and the need for interpolation. To this end, we resort to the following observations, from Section 4.1:

(1) Upsampling an image I, is equivalent to adding levels to the bottom of its wavelet transform, and setting the detail coefficients to zero, while the approximation coefficients remain the same.

(2) Shifting the upsampled image by an amount of s is equivalent to shifting the original image by an amount of $s/2^\ell$, where ℓ is the number of added levels.

Transformed signals therefore can be found by using the original level subbands of the reference image with the help of a hypothetically added level (ℓ) and an integer shift value (s) at the added level.

Now, the aforementioned coefficient matrices, $\mathbf{F}_x$, $\mathbf{K}_x$, and $\mathbf{L}_x$ can be defined, in lower bidiagonal

Toeplitz matrix form as follows.

$$\mathbf{F}_x = \frac{1}{2^{\ell_x+1}} \begin{bmatrix} 2^{\ell_x+1} - |s_x| & & & & \\ |s_x| & 2^{\ell_x+1} - |s_x| & & & \\ & |s_x| & & & \\ & & \ddots & \ddots & \\ & & & |s_x| & 2^{\ell_x+1} - |s_x| \end{bmatrix},$$

$$\mathbf{K}_x = \frac{1}{2^{\ell_x+1}} \begin{bmatrix} -s_x & & & & \\ s_x & -s_x & & & \\ & s_x & & & \\ & & \ddots & \ddots & \\ & & & s_x & -s_x \end{bmatrix},$$

$$\mathbf{L}_x = \frac{1}{2^{\ell_x+1}} \begin{bmatrix} 2^{\ell_x+1} - 3|s_x| & & & & \\ -|s_x| & 2^{\ell_x+1} - 3|s_x| & & & \\ & -|s_x| & & & \\ & & \ddots & \ddots & \\ & & & -|s_x| & 2^{\ell_x+1} - 3|s_x| \end{bmatrix}, \tag{5.4}$$

where s_x and ℓ_x demonstrate the integer shift amounts at the hypothetically added level and the number of added levels for x direction, respectively.

$\mathbf{F}_y$, $\mathbf{K}_y$, and $\mathbf{L}_y$ matrices are defined in a similar manner by upper bidiagonal Toeplitz matrices,

78

using y direction values for s and ℓ.

As mentioned earlier, $\mathbf{F}_x$, $\mathbf{K}_x$ and $\mathbf{L}_x$ are $n \times n$, while $\mathbf{F}_y$, $\mathbf{K}_y$ and $\mathbf{L}_y$ are $m \times m$. Sizes of these matrices also indicate that in-band shift of subbands is performed using only the original level Haar coefficients (which are of size $m \times n$) without upsampling. When the shift amount is negative, diagonals of the coefficient matrices interchange. For the MCTF method proposed in Section 5.1, the matrices are adapted for boundary condition by adding one more column/row at the end, where subband sizes are also adjusted to be $(m + 1) \times (n + 1)$.

The relationship defined above for subpixel shifts, can be used to produce any shift amount based on the fact that wavelet subbands are periodically shift-invariant. Table 5.2 demonstrates the calculation of any shift using subpixels, where % stands for modulo, $\lfloor s \rfloor$ and $\lceil s \rceil$ are the greatest integer lower than, and smallest integer higher than the shift amount s. Using circular shift of subbands for the given amounts in each shift amount case, and setting the new shift amount to the new shift values in Table 5.2, we can calculate any fractional or integer amount of shifts using subpixels.

Table 5.2: Arbitrary shifts defined by circular shift and subpixel amount.

Shift amount	Circular shift	New shift amount
$s\%2 = 0$	$s/2$	0
$s\%2 = 1$	$\lfloor s/2 \rfloor$	1
$\lceil s \rceil \%2 = 0$	$\lceil s \rceil /2$	$s - \lceil s \rceil$
$\lfloor s \rfloor \%2 = 0$	$\lfloor s \rfloor /2$	$s - \lfloor s \rfloor$

If the shift amount (or the new shift amount in Table 5.2) is not divisible by 2, in order to reach an integer value at the $(N + \ell)$th level, the shift value at the original level is rounded to the closest decimal point which is divisible by 2^ℓ.

5.3 Experimental Results

In this section, we demonstrate the results obtained with our method compared to the methods which perform in-band MCTF for video coding. We report our results on CIF video sequence examples with resolutions 352×240 and 352×288. We set our block size to 8×8 or 16×16 depending on the resolution of the sequences (in order to have integer number of blocks in subbands) and the required accuracy. Even though our MCTF method is based on 1-level DWT, we perform 2 more spatial decomposition levels after ME/MC steps before encoding, since compared methods use 3 spatial decomposition levels in total. Motion vectors and error frames are encoded using context-adaptive variable-length coding (CAVLC) and global thresholding with Huffman coding methods, respectively.

Fig. 5.2 shows the comparison of our method with respect to two conventional in-band methods, which are direct wavelet subband matching (i.e., band-to-band) and wavelet-block low-band-shift (LBS) [41] for CIF video sequence "Football". The graph demonstrates rate-distortion curves for a predicted frame of the Football sequence, where the shown bitrates are for error frame only (same as in the compared methods), and the accuracy for our method is set to $1/4$ pixel. As seen in this figure, our method improves PSNR compared to conventional in-band methods by $0.1 - 1$ dB.

We demonstrate our results for several video sequences at different bitrates in Fig. 5.3, where bitrates include the luminance component only, for the error frame, and MVs. The graph on the left shows the results with $1/2$ pixel accuracy using 16×16 blocks, and the one on the right uses $1/4$ pixel accuracy with 8×8 blocks. We also show the residual images for a predicted frame of the Foreman sequence in Fig. 5.4, for 0.1 and 0.02 bpp, respectively. The examples show how our method reduces the residual signal energy even at very low bitrates by providing more accurate reconstruction (prediction).
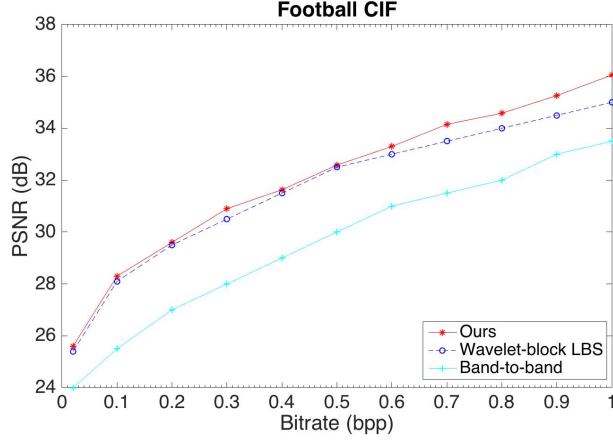
Figure 5.2: Rate-distortion comparison for Football sequence.
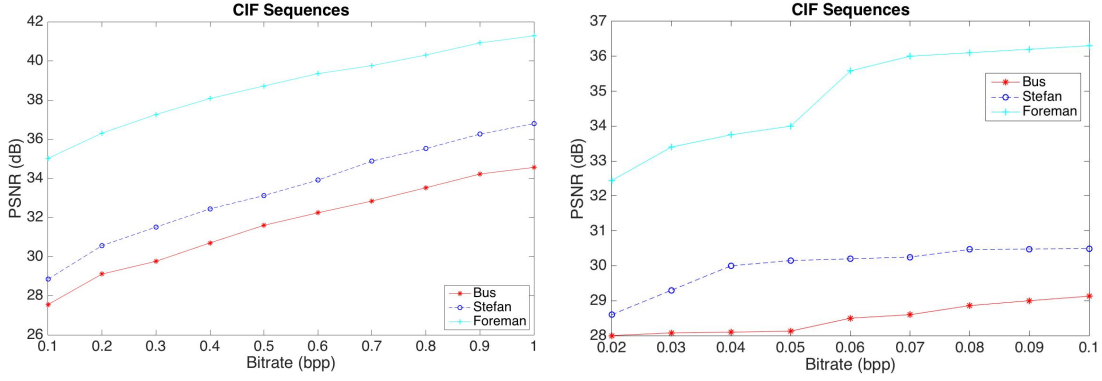
Figure 5.3: PSNR performance of proposed MCTF method, lower bitrates on the right.

Finally, while the compared wavelet-block LBS method has 10 frames memory requirement for 3-level transform, our method requires only 1 frame memory (since upsampling is avoided). Computational complexity of our method is based on the matrix multiplications defined in Section 5.2, and full search method. We also reduce computational complexity by avoiding IDWT.

<div align="center">

a   **b**

</div>
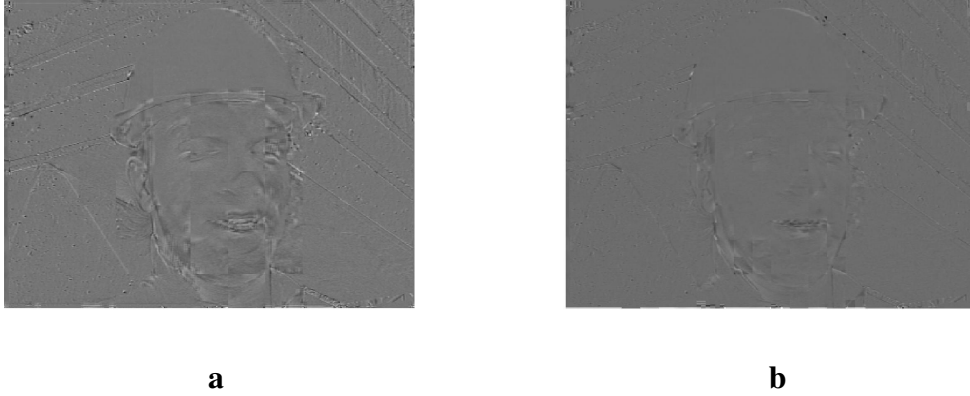
Figure 5.4: Residual images for predicted frames of Foreman for **a** 0.1 bpp **b** 0.02 bpp.

5.4   Conclusion

In this chapter, we propose a novel wavelet-based ME/MC method for MCTF focusing on low-bitrate video coding, where DWT is applied before temporal decomposition, and ME/MC steps are performed directly on DWT subbands. We avoid the need of shift-invariance property for non-redundant DWT (required by conventional methods), by deriving the exact relationships between DWT subbands of reference and transformed video frames. Our method avoids upsampling, inverse-DWT (IDWT), and calculation of redundant DWT while achieving high accuracy even at very low-bitrates. Experimental results show the accuracy of presented method, confirming that our model effectively improves video coding quality by reducing the residual energy in the error frames. The proposed ME/MC scheme can also be adapted for several image/video processing applications such as denoising, or scalable video coding.

CHAPTER 6: SUPER RESOLUTION OF WAVELET-ENCODED IMAGES

In this chapter, we will incorporate our devised methodology explained before in Section 5.2, in a Super Resolution framework [211] [1]. We will employ the relationship between low-pass subbands (i.e., LR images), in order to reconstruct the lost high-frequency information of a reference HR image. The remainder of this chapter is organized as follows. The notations used throughout the chapter along with a brief explanation on the use of derived closed-form linear relationships from Section 5.2 are provided in Section 6.1. Sections 6.2 and 6.3 present the proposed approach, and the stability analysis, respectively. Finally, in Section 6.4, we present the experimental results and comparisons with both single and multiframe state-of-the-art techniques; and in Section 6.5 we provide concluding remarks.

6.1 Subpixel Shifts of a Low Resolution Image

The relationship between low-pass subbands in the wavelet domain is explored in detail and the related notations are provided in this section.

6.1.1 Notation for SR Method

Here, we provide the notations used throughout this chapter in Table 6.1.

Bold uppercase letters in the following sections demonstrate matrices whereas bold lowercase ones indicate vectors. The subscripts h, v, d demonstrate horizontal, vertical, and diagonal translations,

---

[1]The content in this chapter was in part reproduced from the following article: Vildan Atalay Aydin, Hassan Foroosh, A linear well-posed solution to recover high-frequency information for super resolution image reconstruction, Multidimensional Systems and Signal Processing, Year 2017. The copyright form for this article is included in Appendix C.

and x, y stand for x and y directions, respectively; and the subscript k indicates the kth LR or HR image.

Table 6.1: Notation for SR Method.

$I(2m, 2n)$	Reference HR image
A, H, V, D	$1st$ level Haar wavelet transform approximation, horizontal, vertical, and diagonal detail coefficients of $I(2m, 2n)$, respectively
F, K	Matrices to be multiplied by approximation and detail coefficients (i.e., A, H, V, D) of the reference HR image, that are used to define in-band shift of the reference LR image (i.e., A)
ℓ	Number of hypothetically added levels in case of non-integer shifts
s	Integer shift amount after the hypothetically added levels (ℓ)

6.1.2   Subpixel Shifts

Our goal for the proposed SR method is to reconstruct the lost high frequency information of an unknown HR image, given a sequence of subpixel shifted LR images. For this purpose, we first derive the relationship that relates these LR images to the high frequency information of the unknown reference HR image. This relationship can be described by in-band shift (i.e., in the wavelet domain) of a reference LR image, as in Section 5.2.

In order to find the aforementioned relationship, we first assume that the reference HR image is known. The reference LR image is the approximation coefficients obtained by decomposing the HR image for 1-level Haar Transform. Then, we define shifted LR images based on the resultant Haar coefficients of the HR image. The shifting process is illustrated in Fig. 6.1, where shifted LR images (i.e., A_h, A_v, A_d) are described based on the first level approximation and detail coefficients of the reference HR image (i.e., A, H, V, D).

84
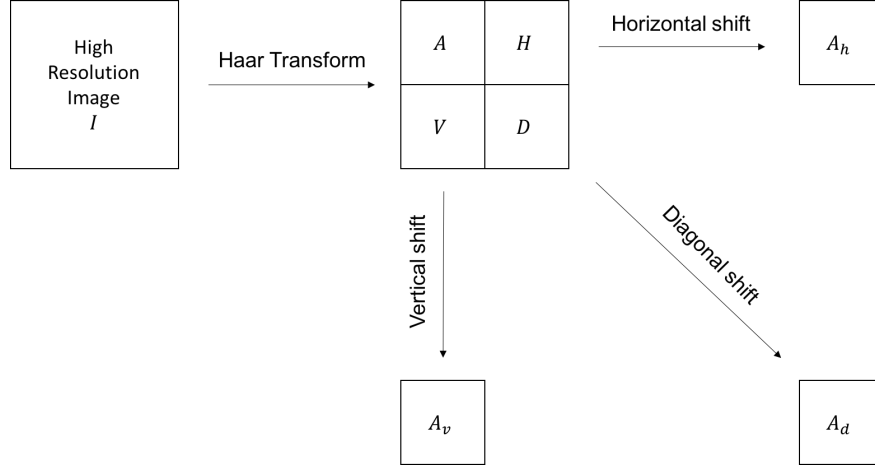
Figure 6.1: In-band shift of a reference LR image (i.e., A) in the Haar domain, using high frequency information of the related HR image (i.e., H, V, D).

Below, we derive the mathematical expressions which demonstrate this relationship, for each translation direction. The derived equations relate the high-frequency part (i.e., detail wavelet coefficients) of a reference HR image to the low-frequency information provided by the LR image sequence.

Here, we will address the derived relationship between LR images as a reminder from Section 5.2. Let $\mathbf{A}$, $\mathbf{H}$, $\mathbf{V}$, and $\mathbf{D}$ be the first level approximation (i.e., reference LR image), horizontal, vertical, and diagonal detail coefficients, respectively, of a $2D$ reference HR image, $I(2m, 2n)$, of size $2m \times 2n$, where m and n are positive integers. Since 1-level wavelet transform reduces the size of HR image by half in each direction for approximation and detail coefficients, we require the size of HR image to be divisible by 2. Now, a translated LR image in an arbitrary direction can be expressed in matrix form using the $1st$ level Haar transform of $I(2m, 2n)$ as in the following equation:

$$\mathbf{A}_s \quad = \quad \mathbf{F}_y\mathbf{A}\mathbf{F}_x + \mathbf{F}_y\mathbf{H}\mathbf{K}_x + \mathbf{K}_y\mathbf{V}\mathbf{F}_x + \mathbf{K}_y\mathbf{D}\mathbf{K}_x, \tag{6.1}$$

85

where $\mathbf{F}$ and $\mathbf{K}$ stand for matrices to be multiplied by the first level lowpass and highpass subbands of the reference HR image, subscripts x and y indicate *horizontal* and *vertical* directions; and $\mathbf{A}_s$ stands for a shifted image in any direction. The low/high-pass subbands together with $\mathbf{A}_s$ are of size $m \times n$, $\mathbf{F}_y$ and $\mathbf{K}_y$ are $m \times m$, whereas $\mathbf{F}_x$ and $\mathbf{K}_x$ are $n \times n$.

Examination of the translational shifts between two LR images reveals that horizontal translation reduces $\mathbf{K}_y$ to zero and $\mathbf{F}_y$ to the identity matrix. This could be comprehended by examining the coefficient matrices defined later in this section (namely, Eq. (6.3)), by making related vertical components zero (specifically, s_y and ℓ_y). This observation lets us define a horizontally shifted image $\mathbf{A}_h$ by using only approximation and horizontal detail coefficients. Likewise, vertical translation solely necessitates approximation and vertical detail coefficients, in which case $\mathbf{K}_x$ is reduced to zero and $\mathbf{F}_x$ is equal to the identity matrix.

As a result, the equation shown above in Eq. (6.1), can be expressed for each translation direction as in:

$$
\begin{aligned}
\mathbf{A}_h &= \mathbf{A}\mathbf{F}_x + \mathbf{H}\mathbf{K}_x, \\
\mathbf{A}_v &= \mathbf{F}_y\mathbf{A} + \mathbf{K}_y\mathbf{V}, \\
\mathbf{A}_d &= \mathbf{F}_y\mathbf{A}\mathbf{F}_x + \mathbf{F}_y\mathbf{H}\mathbf{K}_x + \mathbf{K}_y\mathbf{V}\mathbf{F}_x + \mathbf{K}_y\mathbf{D}\mathbf{K}_x.
\end{aligned}
\tag{6.2}
$$

Now, we can define the matrices, $\mathbf{F}_x$, $\mathbf{F}_y$, $\mathbf{K}_x$, and $\mathbf{K}_y$ in bidiagonal Toeplitz matrix form as follows.

$$
\mathbf{F}_x = \frac{1}{2^{\ell_x+1}}
\begin{bmatrix}
2^{\ell_x+1} - |s_x| & & & & & \\
|s_x| & 2^{\ell_x+1} - |s_x| & & & & \\
& |s_x| & & & & \\
& & \ddots & \ddots & & \\
& & & & |s_x| & 2^{\ell_x+1} - |s_x|
\end{bmatrix},
$$

$$
\mathbf{F}_y = \frac{1}{2^{\ell_y+1}}
\begin{bmatrix}
2^{\ell_y+1} - |s_y| & |s_y| & & & \\
& 2^{\ell_y+1} - |s_y| & |s_y| & & \\
& & \ddots & \ddots & \\
& & & & |s_y| \\
& & & & 2^{\ell_y+1} - |s_y|
\end{bmatrix},
$$

$$
\mathbf{K}_x = \frac{1}{2^{\ell_x+1}}
\begin{bmatrix}
-s_x & & & & \\
s_x & -s_x & & & \\
& s_x & & & \\
& & \ddots & \ddots & \\
& & & s_x & -s_x
\end{bmatrix},
$$

87

$$\mathbf{K}_y = \frac{1}{2^{\ell_y+1}} \begin{bmatrix} -s_y & s_y & & & \\ & -s_y & s_y & & \\ & & \ddots & \ddots & \\ & & & -s_y & s_y \\ & & & & -s_y \end{bmatrix}, \tag{6.3}$$

where $s_{x,y}$ and $\ell_{x,y}$ demonstrate the integer shift amounts at the hypothetically added level and the number of added levels for x and y directions, respectively.

As mentioned earlier, $\mathbf{F}_x$ and $\mathbf{K}_x$ are $n \times n$, while $\mathbf{F}_y$ and $\mathbf{K}_y$ are $m \times m$. These matrices show that a 2-pixel neighborhood in the approximation and detail coefficients of a reference HR image is utilized to shift a reference LR image in-band. When the shift amount is negative, diagonals of the matrices interchange. We leave these matrices as square for them to be nonsingular in the SR process, otherwise these matrices could be adapted for periodic boundary condition by making them rectangular as in our devised MCTF method.

6.2   Super Resolution Reconstruction

We will explain our proposed model in this section. As in the underlying idea of wavelet-based SR algorithms, we assume that the given LR image sequence is the lowpass subbands (i.e., approximation coefficients of 1-level Wavelet Transform) of unknown HR images. The goal is to reconstruct the unknown highpass subbands (i.e., detail coefficients of 1-level Wavelet Transform) of one of these HR images which is chosen as the reference one. The SR method described below is the inverse process of the method described in Section 6.1, where HR images are unknown, and high frequency information for one of these underlying HR images is estimated by solving a related linear system.
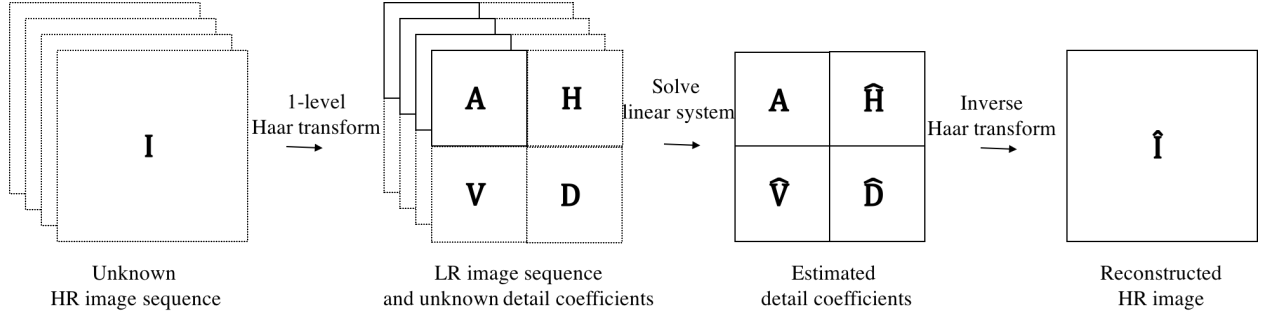
Figure 6.2: Proposed method for Super Resolution Image Reconstruction.

The relationship between two subpixel shifted LR images depends on the highpass subbands of the underlying reference HR image, as demonstrated in the previous section. This fact is used to construct a linear system of equations based on known LR images (i.e., $\mathbf{A}, \mathbf{A}_h, \mathbf{A}_v, \mathbf{A}_d$ in Section 6.1) and unknown highpass subbands of the reference HR image (i.e., $\mathbf{H}, \mathbf{V}, \mathbf{D}$ in Section 6.1) using related formulae from Eq. (6.2) depending on the translation direction. Since there are three unknowns (i.e., horizontal, vertical, and diagonal detail coefficients of the unknown HR image), three shifted LR images together with the reference LR image are required to solve the linear system. Once this system is solved for the unknowns, inverse Haar transform utilizing the reference LR image and the estimated highpass subbands of the underlying unknown reference HR image gives the reconstructed HR image.

Fig. 6.2 shows a pictorial explanation of the proposed method, where solid boxes indicate known or estimated coefficients and dotted boxes show unknown ones. Images with the hat symbol (i.e., ˆ) stands for estimated coefficients. As the figure demonstrates, assuming the LR sequence is first level approximation coefficients of the wavelet transform, we estimate the unknown high frequency information of the reference HR image in order to reconstruct the estimated HR image.

In the scope of this chapter, we assume that the registration between images are known a priori or

has been estimated. Even though the equations derived in Section 6.1 are for subpixel shifts, we apply the proposed SR method to the intersection area of any given shift, which may include an integer part, as well.

The proposed algorithm can also be explained step by step as in **Algorithm** - Super Resolution Image Reconstruction.

Algorithm *Super Resolution Image Reconstruction*
- *Input*: Observed LR images, registration information

- *Objective*: Estimate high frequency information of the unknown reference HR

- *Output*: Reconstructed HR image

▶ Imaging Process

    ⋄ Shift a given High Resolution image randomly in order to obtain three more HR images
    ⋄ Blur with Gaussian kernel, downsample, and add Gaussian noise with zero variance to acquire observed shifted LR images
    ⋄ If shift amount is not subpixel, remove integer parts to find intersection of images which can be defined as subpixel translation

▶ Super Resolution Image Reconstruction Process

    ⋄ Solve a linear system of equations comprised of the formulae in Eq. (6.2) based on the direction of the translation, for highpass subbands $\mathbf{H}$, $\mathbf{V}$, and $\mathbf{D}$ of the reference HR image, using observed LR images and known displacements
    ⋄ Perform inverse Haar Transform on the reference LR image and estimated highpass subbands (i.e., detail coefficients) to reconstruct the HR image

6.3   Stability Analysis

In this section, we will investigate the stability of our method.

As mentioned in Section 6.2, our method constructs a linear system of equations based on given LR images and related shifts. Since the LR images (i.e., $\mathbf{A}$, $\mathbf{A}_h$, $\mathbf{A}_v$, and $\mathbf{A}_d$) and the displacements

between them are known, $\mathbf{H}, \mathbf{V}$, and $\mathbf{D}$ are the only unknowns of the constructed system. This linear system may appear in four forms which include:

- 1 horizontally, 1 vertically, 1 diagonally shifted image

- 1 horizontally, 2 diagonally shifted images

- 1 vertically, 2 diagonally shifted images

- 3 diagonally shifted images

along with the reference LR image, $\mathbf{A}$, where second and third cases demonstrate the same properties. Thus, we will consider the first, second, and last cases in our analysis.

Case 1 *(1 horizontal, 1 vertical, 1 diagonal)* **:** This case constructs a linear system of equations exactly as shown in Eq. (6.2). This linear system is solved first for $\mathbf{H}$ using the equation for $\mathbf{A}_h$, then for $\mathbf{V}$ using the equation for $\mathbf{A}_v$, and finally for $\mathbf{D}$ using the equation for $\mathbf{A}_d$ and substituting the information found for $\mathbf{H}$ and $\mathbf{V}$. Since the coefficient matrices are invertible, this system is stable.

Case 2 and 3 *(1 horizontal/vertical, 2 diagonal)* **:** Here, we will explore Case 2 with 1 horizontally and 2 diagonally shifted images. Case 3 will demonstrate similar features as mentioned above.

This case includes one $\mathbf{A}_h$ and two $\mathbf{A}_d$ from Eq. (6.2) for one horizontal and two diagonally shifted images, where the linear system takes the form:

$$
\begin{aligned}
\mathbf{A}_h &= \mathbf{A}\mathbf{F}_x + \mathbf{H}\mathbf{K}_x, \\
\mathbf{A}_{d_1} &= \mathbf{F}_{y_1}\mathbf{A}\mathbf{F}_{x_1} + \mathbf{F}_{y_1}\mathbf{H}\mathbf{K}_{x_1} + \mathbf{K}_{y_1}\mathbf{V}\mathbf{F}_{x_1} + \mathbf{K}_{y_1}\mathbf{D}\mathbf{K}_{x_1}, \\
\mathbf{A}_{d_2} &= \mathbf{F}_{y_2}\mathbf{A}\mathbf{F}_{x_2} + \mathbf{F}_{y_2}\mathbf{H}\mathbf{K}_{x_2} + \mathbf{K}_{y_2}\mathbf{V}\mathbf{F}_{x_2} + \mathbf{K}_{y_2}\mathbf{D}\mathbf{K}_{x_2}.
\end{aligned}
\tag{6.4}
$$

Again, as in Case 1, the first equation is stable, therefore $\mathbf{H}$ can be found easily. Solving equations for $\mathbf{A}_{d_1}$ and $\mathbf{A}_{d_2}$ for $\mathbf{V}$ results in:

$$\mathbf{V}(\mathbf{F}_{x_1}\mathbf{K}_{x_1}^{-1} - \mathbf{F}_{x_2}\mathbf{K}_{x_2}^{-1}) = \mathbf{K}_{y_1}^{-1}\mathbf{S}\mathbf{K}_{x_1}^{-1} - \mathbf{K}_{y_2}^{-1}\mathbf{T}\mathbf{K}_{x_2}^{-1}, \tag{6.5}$$

where

$$\mathbf{S} = \mathbf{A}_{d_1} - \mathbf{F}_{y_1}\mathbf{A}\mathbf{F}_{x_1} - \mathbf{F}_{y_1}\mathbf{H}\mathbf{K}_{x_1},$$

$$\mathbf{T} = \mathbf{A}_{d_2} - \mathbf{F}_{y_2}\mathbf{A}\mathbf{F}_{x_2} - \mathbf{F}_{y_2}\mathbf{H}\mathbf{K}_{x_2}.$$

In order to tackle the instability problem caused by inverting multiplication and summation of matrices in Eq. (6.5), we right multiply this equation with $\mathbf{K}_{x_1}$. Since $\mathbf{K}_{x_1}$ and $\mathbf{K}_{x_2}$ differ only by the shift value for the two diagonally shifted images, $\mathbf{K}_{x_2}^{-1}\mathbf{K}_{x_1}$ results in the identity matrix multiplied by a scalar which depends only on the shifts. Thus, the equation for $\mathbf{V}$ becomes:

$$\mathbf{V}(\mathbf{F}_{x_1} - \mathbf{F}_{x_2}\alpha) = \mathbf{K}_{y_1}^{-1}\mathbf{S} - \mathbf{K}_{y_2}^{-1}\mathbf{T}\alpha, \tag{6.6}$$

where

$$\alpha = \mathbf{K}_{x_2}^{-1}\mathbf{K}_{x_1},$$

where α is defined as a constant. Truncated Singular Value Decomposition (TSVD) is used with the resulting equation in (6.6) to find $\mathbf{V}$. Rank of TSVD method is decided based on minimizing the following cost function:

$$\arg\min_r ||\mathbf{X}_r^{-1}\mathbf{X}_r - \mathbf{U}||_F, \tag{6.7}$$

where $\mathbf{U}$ shows the identity (i.e., unit) matrix, subscript F is *Frobenius norm*, and $\mathbf{X}_r$ stands for *rank-r* approximation of a matrix $\mathbf{X}$.

In order to successfully truncate $\mathbf{X}$ at r, we follow a theorem by [212] (Theorem 3.2), which implies that there must be a well-determined gap between the two consecutive singular values at r (i.e., σ_r) and $r+1$ (i.e., σ_{r+1}).

As one can see in Eq. (6.6), the stability of our method is partially dependent on the closeness of shift amounts.

Case 4 *(3 diagonal)* **:** The final case includes three diagonally shifted images together with the reference image. Therefore, the linear system is constructed as:

$$
\begin{aligned}
\mathbf{A}_{d_1} &= \mathbf{F}_{y_1}\mathbf{AF}_{x_1} + \mathbf{F}_{y_1}\mathbf{HK}_{x_1} + \mathbf{K}_{y_1}\mathbf{VF}_{x_1} + \mathbf{K}_{y_1}\mathbf{DK}_{x_1}, \\
\mathbf{A}_{d_2} &= \mathbf{F}_{y_2}\mathbf{AF}_{x_2} + \mathbf{F}_{y_2}\mathbf{HK}_{x_2} + \mathbf{K}_{y_2}\mathbf{VF}_{x_2} + \mathbf{K}_{y_2}\mathbf{DK}_{x_2}, \\
\mathbf{A}_{d_3} &= \mathbf{F}_{y_3}\mathbf{AF}_{x_3} + \mathbf{F}_{y_3}\mathbf{HK}_{x_3} + \mathbf{K}_{y_3}\mathbf{VF}_{x_3} + \mathbf{K}_{y_3}\mathbf{DK}_{x_3}.
\end{aligned}
\tag{6.8}
$$

Solving the system above in Eq. (6.8) for $\mathbf{H}$, we find a generalized Sylvester equation as in:

$$
\mathbf{P}_1\mathbf{HQ}_1 - \mathbf{P}_2\mathbf{HQ}_2 = \mathbf{R},
\tag{6.9}
$$

where

$$\mathbf{P}_1 = \mathbf{K}_{y_1}{}^{-1}(\mathbf{F}_{y_1} - \mathbf{K}_{y_1}\mathbf{K}_{y_2}{}^{-1}\mathbf{F}_{y_2}),$$

$$\mathbf{P}_2 = \mathbf{K}_{y_2}{}^{-1}(\mathbf{F}_{y_2} - \mathbf{K}_{y_2}\mathbf{K}_{y_3}{}^{-1}\mathbf{F}_{y_3}),$$

$$\mathbf{Q}_1 = \mathbf{K}_{x_1}(\mathbf{F}_{x_1} - \mathbf{F}_{x_2}\mathbf{K}_{x_2}{}^{-1}\mathbf{K}_{x_1})^{-1},$$

$$\mathbf{Q}_2 = \mathbf{K}_{x_2}(\mathbf{F}_{x_2} - \mathbf{F}_{x_3}\mathbf{K}_{x_3}{}^{-1}\mathbf{K}_{x_2})^{-1},$$

$$\mathbf{R} = [(\mathbf{K}_{y_1}{}^{-1}\mathbf{A}_{d_1} - \mathbf{K}_{y_1}{}^{-1}\mathbf{F}_{y_1}\mathbf{A}\mathbf{F}_{x_1})$$

$$-(\mathbf{K}_{y_2}{}^{-1}\mathbf{A}_{d_2} - \mathbf{K}_{y_2}{}^{-1}\mathbf{F}_{y_2}\mathbf{A}\mathbf{F}_{x_2})\mathbf{K}_{x_2}{}^{-1}\mathbf{K}_{x_1}]$$

$$\times \ (\mathbf{F}_{x_1} - \mathbf{F}_{x_2}\mathbf{K}_{x_2}{}^{-1}\mathbf{K}_{x_1})^{-1}$$

$$- \ [(\mathbf{K}_{y_2}{}^{-1}\mathbf{A}_{d_2} - \mathbf{K}_{y_2}{}^{-1}\mathbf{F}_{y_2}\mathbf{A}\mathbf{F}_{x_2})$$

$$-(\mathbf{K}_{y_3}{}^{-1}\mathbf{A}_{d_3} - \mathbf{K}_{y_3}{}^{-1}\mathbf{F}_{y_3}\mathbf{A}\mathbf{F}_{x_3})\mathbf{K}_{x_3}{}^{-1}\mathbf{K}_{x_2}]$$

$$\times \ (\mathbf{F}_{x_2} - \mathbf{F}_{x_3}\mathbf{K}_{x_3}{}^{-1}\mathbf{K}_{x_2})^{-1}.$$

By examining $\mathbf{P}_i$ for $i = 1, 2$, in the generalized Sylvester equation, $\mathbf{K}_{y_i}\mathbf{K}_{y_{i+1}}{}^{-1}$ could be changed by multiplication by a scalar (as in $\mathbf{K}_{x_2}{}^{-1}\mathbf{K}_{x_1}$ in Eq. (6.6)), which leaves $\mathbf{F}_{y_i} - \mathbf{K}_{y_i}\mathbf{K}_{y_{i+1}}{}^{-1}\mathbf{F}_{y_{i+1}}$ as an upper bidiagonal matrix, since $\mathbf{F}_{y_i}$ is also upper bidiagonal. Moreover, since $\mathbf{K}_y$ is an upper bidiagonal matrix, inverse of $\mathbf{K}_y$ is an upper triangular matrix [213]. Therefore, by multiplication of two upper triangular matrices, we obtain upper triangular matrices for $\mathbf{P}_i$. By following similar analysis, we observe that $\mathbf{Q}_i$ are lower triangular matrices.

Here, we refer to a theorem by [214] for a generalized Sylvester equation to have a unique solution. Interested reader can find the proof for this theorem in the referred paper; we include the theorem here to make this dissertation self-contained.

Theorem: The matrix equation in (6.9) has a unique solution if and only if

1. $\mathbf{P}_1 - \lambda\mathbf{P}_2$ and $\mathbf{Q}_2 - \lambda\mathbf{Q}_1$ are regular matrix pencils, and

2. $\rho(\mathbf{P}_1, \mathbf{P}_2) \cap \rho(\mathbf{Q}_1, \mathbf{Q}_2) = \emptyset$,

where λ shows the generalized eigenvalues of the matrix pencils, ρ defines the spectra of the generalized eigenvalues, and $(.,.)$ demonstrates a matrix pencil.

The matrix pencils constructed as $(\mathbf{P}_1, \mathbf{P}_2)$, and $(\mathbf{Q}_1, \mathbf{Q}_2)$, using the given $\mathbf{P}_1, \mathbf{P}_2, \mathbf{Q}_1$ and $\mathbf{Q}_2$ in Eq. (6.9), are not guaranteed either to be regular, or to have empty intersection of generalized eigenvalue spectra. For instance, when any of the two LR images have negative horizontal shift amount, the related $\mathbf{Q}_i$ has zero diagonals, and a zero element on the diagonal makes the matrix pencil singular when a matrix pencil is upper/lower triangular [215]. Since we know that $\mathbf{P}_i$, and $\mathbf{Q}_i$, for $i = 1, 2$, are upper and lower triangular matrices, respectively, forming upper/lower triangular matrix pencils, two images with negative horizontal shifts satisfy requirements for singular matrix pencils.

Based on these facts, solution methods utilized for generalized Sylvester equation cannot be employed here. Therefore, in order to find a solution to the system in Eq. (6.8), we first vectorize the equations using Kronecker tensor product, before solving for the unknowns:

$$\mathbf{a}_{d_i} = (\mathbf{F}_{y_i} \otimes \mathbf{F}_{x_i}^T)\mathbf{a} + (\mathbf{F}_{y_i} \otimes \mathbf{K}_{x_i}^T)\mathbf{h} + (\mathbf{K}_{y_i} \otimes \mathbf{F}_{x_i}^T)\mathbf{v} + (\mathbf{K}_{y_i} \otimes \mathbf{K}_{x_i}^T)\mathbf{d}, \qquad (6.10)$$

for $i = 1, 2, 3$. Here, lowercase bold letters indicate column-vise vectorized versions of $\mathbf{A}_{d_i}, \mathbf{A}, \mathbf{H}, \mathbf{V}$, and $\mathbf{D}$, and these vectors have size $mn \times 1$. The Kronecker tensor products in parenthesis result in matrices of size $mn \times mn$, where m and n are the size of LR images.

By solving Eq. (6.10) for $\mathbf{h}$, we find the following equation which appears similar to the equation

for $\mathbf{V}$ in Eq. (6.5):

$$[\mathbf{W}_1^{-1}\mathbf{Y}_1 - \mathbf{W}_2^{-1}\mathbf{Y}_2]\mathbf{h} = \mathbf{W}_1^{-1}\mathbf{Z}_1 - \mathbf{W}_2^{-1}\mathbf{Z}_2, \tag{6.11}$$

where

$$\begin{aligned}
\mathbf{W}_i &= (\mathbf{KF})_i - (\mathbf{KK})_i(\mathbf{KK})_{i+1}^{-1}(\mathbf{KF})_{i+1}, \\
\mathbf{Y}_i &= (\mathbf{FK})_i - (\mathbf{KK})_i(\mathbf{KK})_{i+1}^{-1}(\mathbf{FK})_{i+1}, \\
\mathbf{Z}_i &= \mathbf{a}_{d_i} - (\mathbf{KK})_i(\mathbf{KK})_{i+1}^{-1}\mathbf{a}_{d_{i+1}} \\
&\quad -[(\mathbf{FF})_i - (\mathbf{KK})_i(\mathbf{KK})_{i+1}^{-1}(\mathbf{FF})_{i+1}]\mathbf{a},
\end{aligned}$$

for $i = 1, 2$, and

$$\begin{aligned}
\mathbf{FF} &= \mathbf{F}_y \otimes \mathbf{F}_x^T, \\
\mathbf{FK} &= \mathbf{F}_y \otimes \mathbf{K}_x^T, \\
\mathbf{KF} &= \mathbf{K}_y \otimes \mathbf{F}_x^T, \\
\mathbf{KK} &= \mathbf{K}_y \otimes \mathbf{K}_x^T.
\end{aligned}$$

Here, in order to solve for $\mathbf{h}$, we follow a similar approach to the one used to reach Eq. (6.6) from Eq. (6.5), where Eq. (6.11) is left multiplied by $\mathbf{W}_1$ in order to reduce the instability. Again, TSVD is utilized to solve the equation with the same cost function used in Eq. (6.7).

As in Cases 2 and 3, the stability of our solution depends partly on the closeness of shift values which affects the matrix inversions.

6.4   Experimental Results

In this section, we first present the implementation details, followed by results for the proposed method along with comparisons to the recent state-of-the-art and conventional techniques. Comparisons are made based on qualitative and quantitative evaluations on both commonly adapted test examples and real world images to demonstrate the influence of compression artifacts and sensor noise on the proposed method. LR image sequences are synthetically generated. Computational time efficiency of the proposed method against other methods are also presented. Moreover, HR and LR reference images for all test cases and zoomed parts in detailed areas for each image are provided.

6.4.1   Implementation Details

LR image sequences are synthesized by the method explained in Section 6.2 (Imaging Process in **Algorithm**). LR images are divided into overlapping blocks of size 32×32, in order to reduce memory usage and decrease computational time.

To simulate the motion estimation error for the proposed method, HR reference image is shifted randomly for a shift amount which is not necessarily divisible by 2^ℓ and shifts are rounded to the closest decimal divisible by 2^ℓ for the calculations, as described before in Section 6.1.

For the cases when the shift amounts are not subpixel (which might be integer or include an integer part), we find the intersection area of the images which can be described as subpixel shift. We apply the same method to the intersected area, where boundaries are lost for the maximum integer amount among all shifts.

In order to reduce the boundary problem caused by square coefficient matrices $\mathbf{F}_{x,y}$ and $\mathbf{K}_{x,y}$ which

97

does not include the information in the boundaries, the last rows and columns of calculated $\mathbf{H}$, $\mathbf{V}$, and $\mathbf{D}$ are extrapolated.

Color images are handled by the conventional approach, in *YCbCr* color space, where only the illuminance channel of images are dealt with the proposed method, since human visual system is more sensitive to changes in illuminance channel. The chrominance channels are upsampled using bicubic interpolation.

6.4.2 Qualitative Comparison

We compared our method with both multiframe and single image SR techniques including interpolation based ones which are Bicubic interpolation and Robust Super Resolution by Zomet *et al.* [115], regularization-based methods by Babacan *et al.* [178] and Farsiu *et al.* [62], a learning-based method by Yang *et al.* [108], a wavelet-based method by Mallat and Yu [130], and finally a wavelet-domain learning-based method by Dong *et al.* [126]. Compared methods were given the same input images and knowledge of registration (if required).

Figs. 6.3 to 6.8 show results obtained with our method, the state-of-the-art, and conventional ones. As can be seen from these figures, in zoomed areas particularly, the proposed method generates sharper edges with less artifacts compared to other methods.

While bicubic interpolation ((**c**) parts in all related figures) tend to introduce blur to the images, Robust SR technique by Zomet *et al.* [115] leaves jaggy artifacts on the edges which are easily seen in (**d**) parts of Figs. 6.3, 6.4, and 6.5 of Lena, Car tag, and Resolution chart images. Babacan's method [178] ((**e**) parts in figures) alleviates the jaggy artifacts in most cases, yet the final results remain overly smoothed; whereas, Farsiu's method [62] does not reconstruct details that can be observed especially in part (**f**) of Figs. 6.5 and 6.6 of Resolution chart and Mandrill images. While

Yang's method [108] tend to obtain blurry images as in part (**g**) of Figs. 6.4 and 6.8; Mallats results [130] include many noticeable visual artifacts that can be observed especially in part (**h**) of Figs. 6.3, 6.7, and 6.8. Dong's method [126], even though better at removing artifacts and achieving natural looking results compared to the other methods, also is prone to leave blurry images as can be recognized without difficulty in (**i**) parts of Resolution chart, Mandrill and Car images in Figs. 6.5, 6.6, and 6.8. In addition, compared methods fail to recover fine details which is mostly recognized in car tags and numbers in Figs. 6.4, 6.5, 6.7, and 6.8.

a HR image   **b** LR reference image   **c** Bicubic   **d** Zomet [115]   **e** Babacan [178]

f Farsiu [62]   **g** Yang [108]   **h** Mallat [130]   **i** Dong [126]   **j** Ours
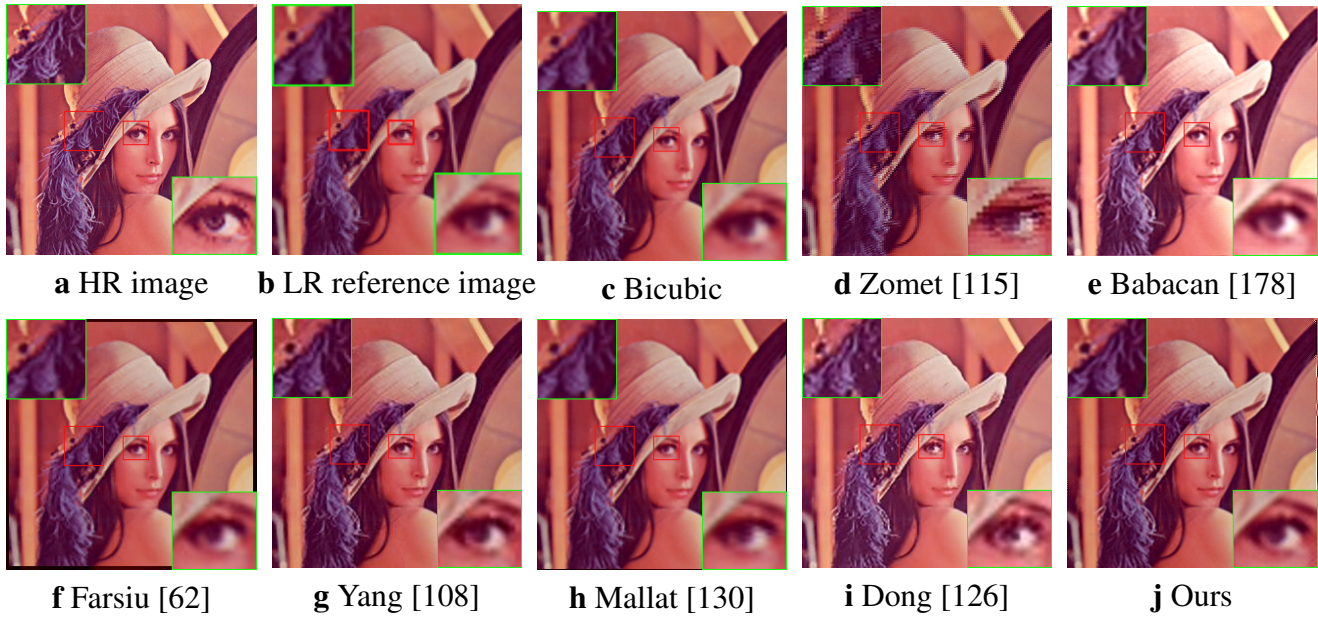
Figure 6.3: Lena image comparison (×2), with zoomed parts in green.

On the other hand, the proposed method, seen in (**j**) parts of Figs. 6.3, 6.4, and 6.5 to 6.8, is able to recover sharp edges without visual artifacts, or blurring the images which leads to generating the closest results to the ground truth. Particularly, Resolution chart, Mandrill and Car images in Figs. 6.5, 6.6, and 6.8 demonstrate the high quality achieved with the proposed method. Recovered texture details with our method can be observed in all test cases upon a closer look, specifically in the feather texture of Lena's hat in Fig. 6.3 and hair texture in cheeks of Mandrill image in Fig.

6.6. Overall, the proposed method removes artifacts and blur while preserving sharp edges without sacrificing a natural look.

a HR image   **b** LR reference image   **c** Bicubic   **d** Zomet [115]   **e** Babacan [178]

f Farsiu [62]   **g** Yang [108]   **h** Mallat [130]   **i** Dong [126]   **j** Ours
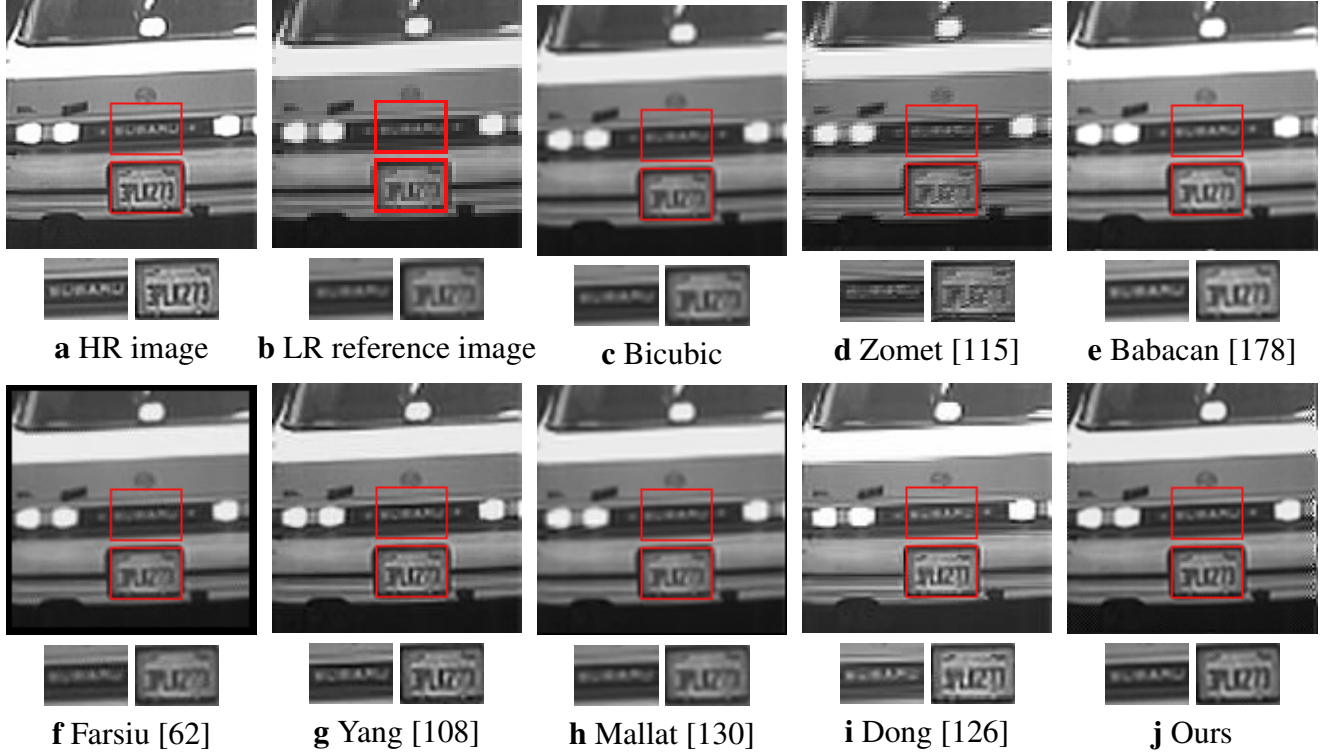
Figure 6.4: Car tag image (MDSP dataset [2]) comparison ($\times 2$), with zoomed parts below reconstructed ones.

6.4.3 Quantitative Comparison

To further investigate the effectiveness of our method, we also conduct a comparison based on objective measurements PSNR, RMSE and SSIM [124], which are the evaluation metrics for SR methods used by state-of-the-art, summarized in Table 6.2. Comparisons are based on the illuminance channel of images reconstructed with all methods. While Zomet's and Babacan's methods perform the worst based on most measurements for all images, Dong's method has much better results compared to the other methods since in order to alleviate the correspondence ambiguity, their

method uses different patches in a single image to learn various sets of bases. However, as can be seen also from Table 6.2, in most cases, quantitative comparisons confirm visual ones which shows that our method outperforms the state-of-the-art. Even though the quantitative measurements for some images are better for Yang's method, with a closer look in the zoomed squares of images, it can be seen that details are recovered better in our method.

a HR image   **b** LR reference image   **c** Bicubic   **d** Zomet [115]   **e** Babacan [178]

f Farsiu [62]   **g** Yang [108]   **h** Mallat [130]   **i** Dong [126]   **j** Ours
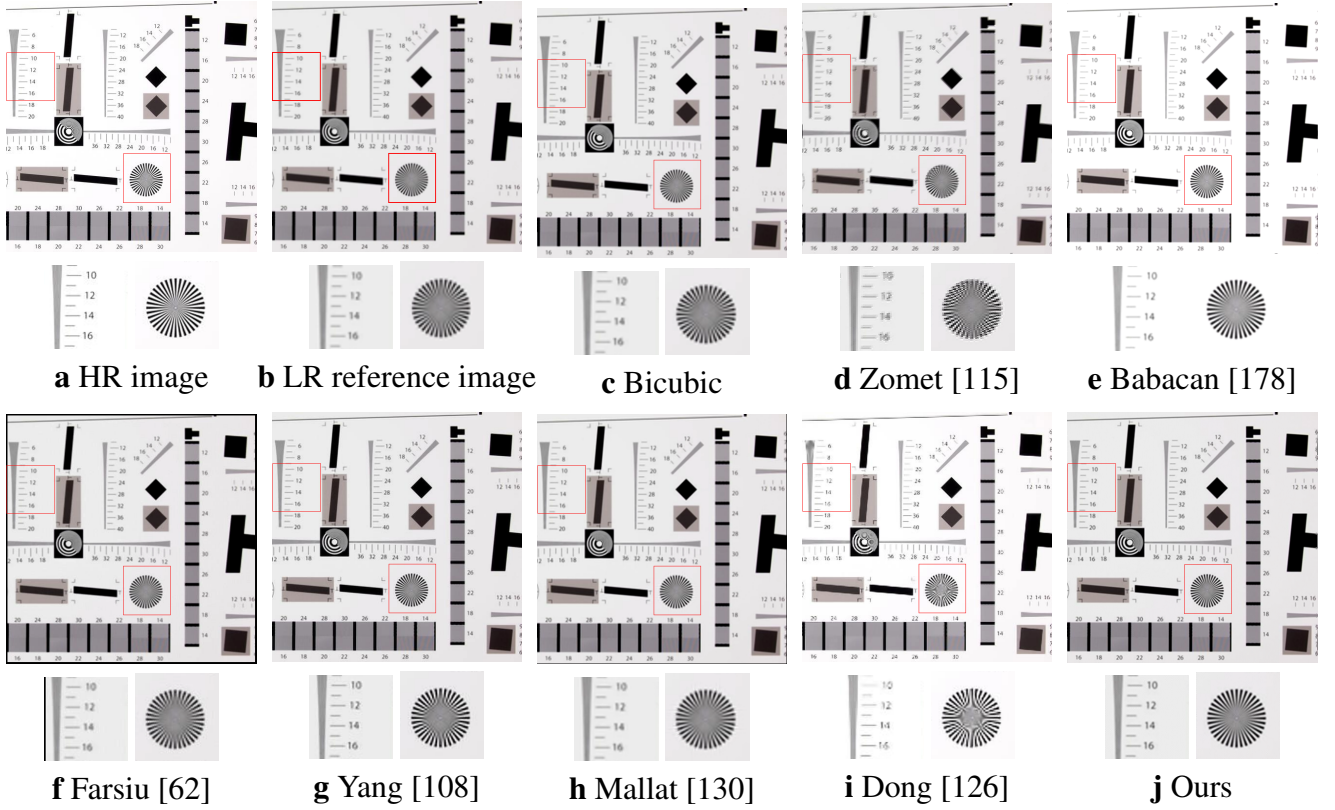
Figure 6.5: Resolution chart image comparison ($\times 2$), with zoomed parts below reconstructed ones.

a HR image   b LR reference image   c Bicubic   d Zomet [115]   e Babacan [178]

f Farsiu [62]   g Yang [108]   h Mallat [130]   i Dong [126]   j Ours
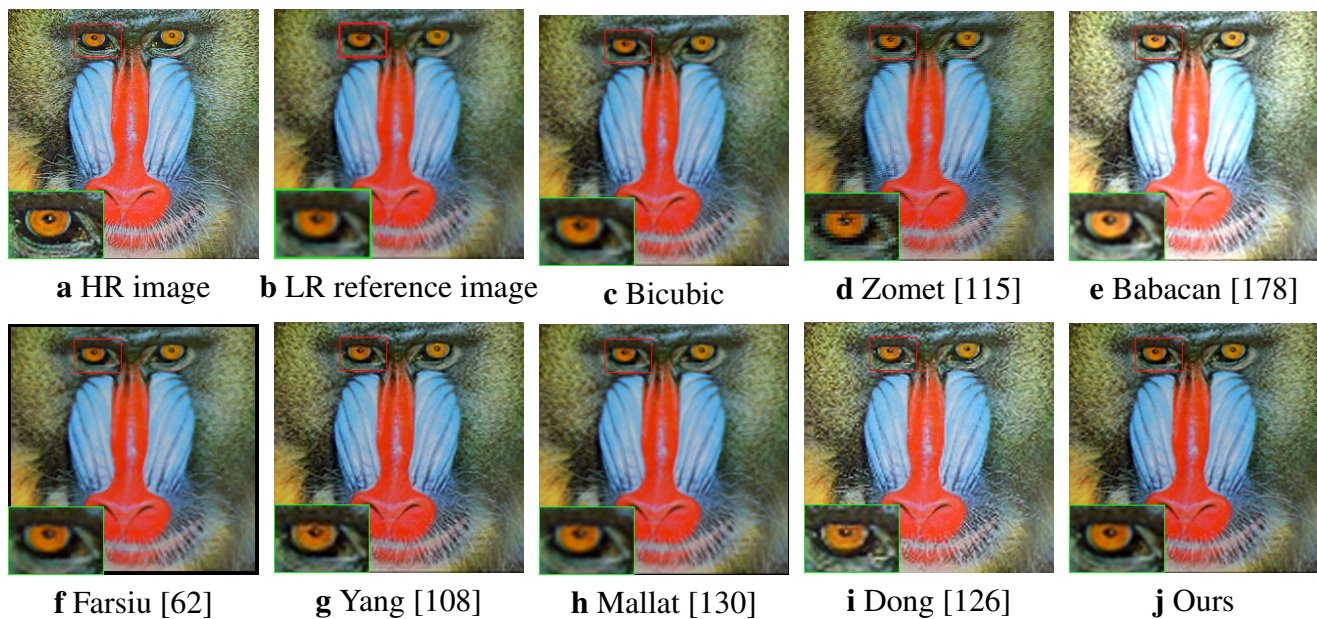
Figure 6.6: Mandrill image comparison (×2), with zoomed part in green.

a HR image   b LR reference image   c Bicubic   d Zomet [115]   e Babacan [178]

f Farsiu [62]   g Yang [108]   h Mallat [130]   i Dong [126]   j Ours
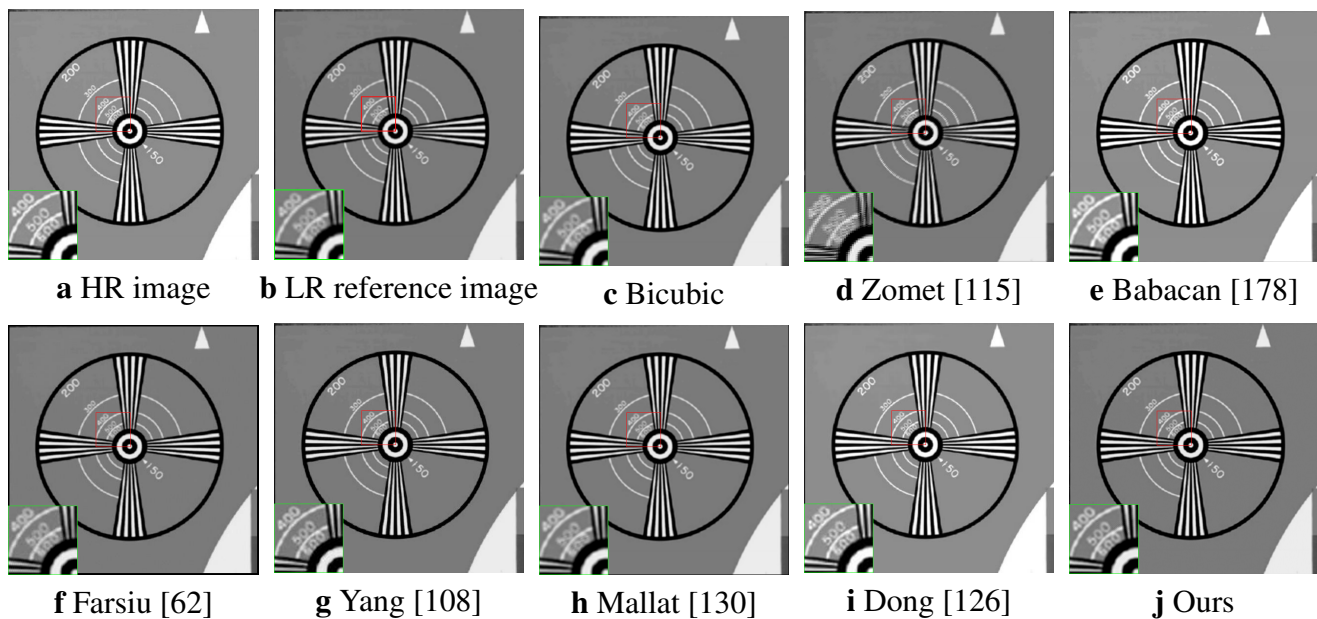
Figure 6.7: Circles (MDSP dataset [2]) image comparison (×2), with zoomed part in green.

| **a** HR image | **b** LR reference image | **c** Bicubic | **d** Zomet [115] | **e** Babacan [178] |

| **f** Farsiu [62] | **g** Yang [108] | **h** Mallat [130] | **i** Dong [126] | **j** Ours |
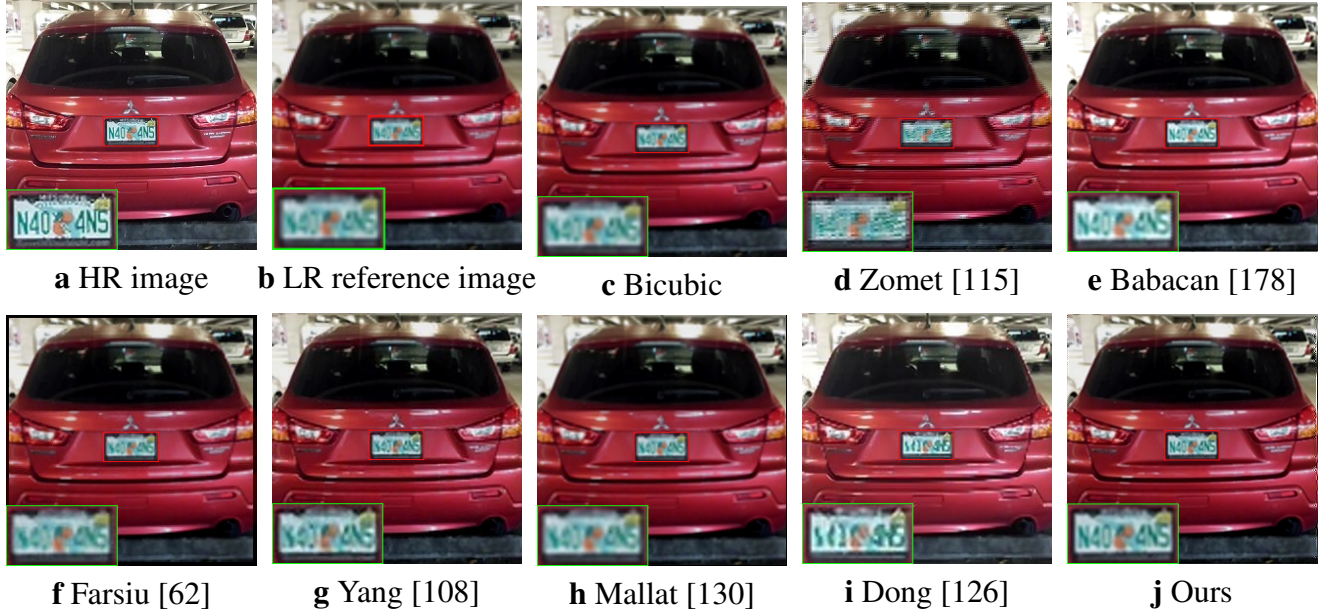
Figure 6.8: Car image comparison ($\times 2$), with zoomed part in green. Image is captured by a cell-phone camera.

6.4.4   Computational Efficiency

The computational complexity of the proposed method depends on matrix multiplications ($\mathcal{O}(n^3)$) along with the TSVD method ($\mathcal{O}(nr^2)$ where r is the approximation rank). Since all blocks have the same size and use the same shift information, matrix inversions are handled only once, and the proposed super resolution method is applied to all blocks in parallel. Our method can also be applied as a sparse method in order to reduce time complexity, considering the fact that coefficient matrices are either bidiagonal or at most triangular matrices.

Time complexity of the proposed method and state-of-the-art is compared in Table 6.3, where average time taken for different size LR images is shown in seconds, where all methods are run in the same computational framework. Block size for our method in all compared cases are set to 32×32. As can be seen from the table, the proposed method outperforms regularization based

methods of Babacan *et al.* [178] (where outliers of [178] are removed for a fair comparison) and Mallat *et al.* [130], and learning based methods of Dong *et al.* [126] and Yang *et al.* especially when the image sizes are relatively large.

Table 6.2: Comparison of proposed SR method with other methods in PSNR, RMSE, SSIM.

		Lena	Car tag	Res. chart	Mandrill	Circles	Car
Bicubic	PSNR	27.33	24.88	24.76	21.8	32.36	24.53
	RMSE	9.34	13.35	13.35	16.64	5.56	13.65
	SSIM	0.86	0.84	0.97	0.61	0.99	0.84
Zomet [115]	PSNR	24.54	21.83	24.14	20.69	24.13	23.03
	RMSE	13.67	20.63	19.29	18.91	16.02	18.83
	SSIM	0.78	0.72	0.95	0.55	0.97	0.73
Babacan [178]	PSNR	17.66	21.00	18.50	13.79	21.89	22.42
	RMSE	33.39	22.72	30.29	52.12	20.49	19.30
	SSIM	0.81	0.84	0.94	0.63	0.97	0.84
Farsiu [62]	PSNR	25.61	23.72	23.78	21.44	27.68	23.78
	RMSE	11.4	15.32	15.01	17.33	9.50	14.94
	SSIM	0.82	0.82	0.93	0.57	0.98	0.81
Yang [108]	PSNR	28.95	28.24	27.39	22.40	33.20	26.06
	RMSE	7.74	10.11	10.87	15.55	5.00	12.10
	SSIM	0.90	0.91	0.98	0.69	1.00	0.87
Mallat [130]	PSNR	25.64	23.04	23.79	21.26	27.25	23.32
	RMSE	11.34	16.37	15.53	17.73	10.06	15.59
	SSIM	0.83	0.80	0.95	0.57	0.98	0.8
Dong [126]	PSNR	28.45	24.94	24.89	21.48	34.63	24.23
	RMSE	8.22	12.40	12.47	17.27	4.07	13.46
	SSIM	0.89	0.89	0.96	0.67	1.00	0.86
Ours	PSNR	28.99	28.07	28.07	23.39	37.29	26.63
	RMSE	7.73	9.08	9.28	13.41	3.17	10.49
	SSIM	0.89	0.93	0.99	0.76	1.00	0.89

A comparison of block sizes with time in seconds and PSNR for the proposed method is shown in Fig. 6.9. The results are calculated for 100 different images for 100 random shift amounts, and the average time and PSNR are shown in the graphs (after removing the outliers). As one can see from the graph, as block sizes increase, PSNR improves; however, time complexity increases at the same time. Therefore, the block sizes can be decided based on the application depending on the importance of time or accuracy. Although the graph demonstrates the results for square sized blocks, the block sizes are decided based on the image sizes, which can as well be rectangular.

Table 6.3: Comparison of proposed SR method with other methods in time (s).

LR Image size	Bicubic	Zomet [115]	Babacan [178]	Farsiu [62]	Yang [108]	Mallat [130]	Dong [126]	Proposed
32×32	0.3	9.7	19	1.2	2.6	14	12	7
64×64	0.3	9.6	36	1.2	12	64	33	12
128×128	0.3	9.8	217	1.2	59	287	140	35
256×256	0.3	9.8	976	1.2	157	1298	440	122

a

b
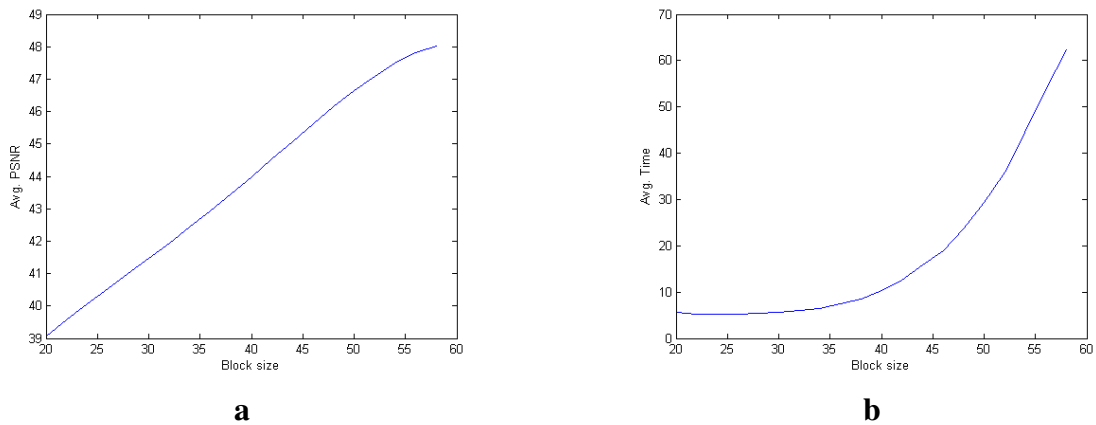
Figure 6.9: **a** Average PSNR based on block size **b** Average runtime (*s*) based on block size.

6.5 Conclusion

As a final remark, a direct wavelet-based super resolution technique is proposed by first deriving exact in-band relationships between two subpixel shifted images, then utilizing these relationships in a linear system to reconstruct high frequency information of a low resolution reference image. Our results outperform the conventional as well as advanced recently published methods. We attribute this to the accuracy, well-posedness and the linearity of the equations derived in Section 6.1 and the inherent local nature of wavelets, making them very effective in signal localization. In summary, we present herein a method for super-resolution by effectively estimating the high frequency information in the Haar domain, which in a sense is a hybrid approach between single image and multiview methods, taking advantage of the best of both worlds.

CHAPTER 7: SUPER RESOLUTION OF DOWNSAMPLING-BASED COMPRESSED VIDEO

The previous two chapters demonstrate the use of the relationship between wavelet-subbands of two translationally shifted images, for motion estimation in video coding and super resolution problems. In this chapter, we will utilize the motion estimation method described in Chapter 5, and modify our super resolution method in Chapter 6 into an iterative back projection framework, in order to employ these methods in super resolution of downsampling-based compressed video problem. We describe our proposed solution to this problem in Section 7.1, explain the modification to our super resolution solution in Section 7.2, present experimental results in Section 7.3, and conclude this chapter in Section 7.4.

7.1 Super Resolution of Compressed Video

We explain our proposed method for super resolution of compressed video, in this section.

We propose a wavelet-based super resolution method, for which the flowchart is shown in Fig. 7.1. Given a video sequence, as we have experimented before in Chapter 5, we first perform DWT on each frame of the sequence. Using the wavelet subbands of the original HR frames, we calculate the motion estimation block by block, between a reference frame, and target frames in the group of pictures. Motion estimation process is performed as described before in Section 5.1.

After the motion estimation step, high frequency information of all the frames in the sequence are discarded, in order to obtain the downsampled frames to be encoded. These LR frames, together with the motion estimation information for each block within them, are later encoded into a bitstream.
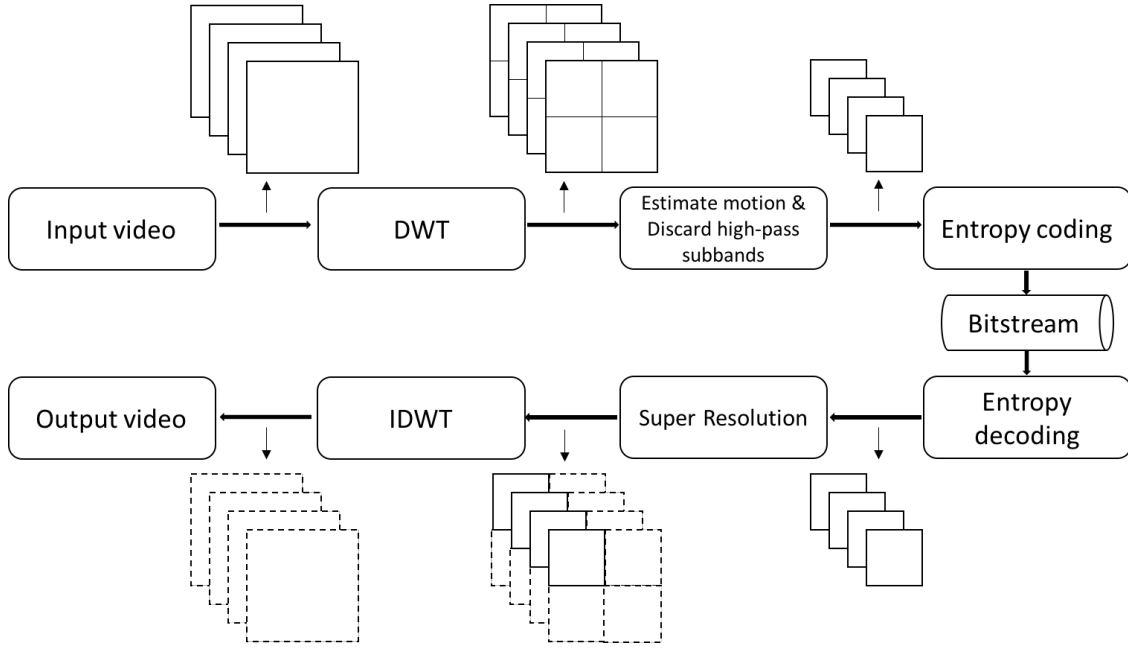
Figure 7.1: Proposed super resolution of compressed video model.

In the decoder side, after the LR frames and motion vectors are decoded, a wavelet-based multi-frame super resolution reconstruction method, which will be described in the next section, is performed on these frames. In Fig. 7.1, estimated high frequency information of decoded LR frames are shown in dashed lines.

After the super resolution process is performed on each frame in the group, inverse discrete wavelet transform (IDWT) is performed to estimate the required HR frames, which are also demonstrated in dashed lines in Fig. 7.1.

7.2   Iterative Back-Projection Based Super Resolution

In this section, we describe the modification of our super resolution method in Chapter 6.

7.2.1 Notation for Super Resolution of Compressed Video Method

Here we provide the notation used throughout this chapter.

Table 7.1: Notation for Super Resolution of Compressed Video Method.

I	Reference HR frame
A, H, V, D	Haar wavelet transform approximation, horizontal, vertical, and diagonal subbands of HR frame, respectively

As before, uppercase bold letters in the following sections demonstrate matrices and lowercase bold ones show vectors. Subscripts x and y indicate the horizontal and vertical translation directions, respectively.

7.2.2 Modified Observation Model

In this subsection, we describe the modification to the original super resolution observation model.

Let $I(\sigma m \times \sigma n)$ denote the desired HR frame, and A_k be the kth observed LR image, with a downsampling ratio of σ. The SR observation model, as defined before in Section 2.4, is given by:

$$\mathbf{a}_k = \boldsymbol{\Lambda}_k \mathbf{B}_k \mathbf{M}_k \mathbf{i} + \mathbf{n}_k, \quad k = 1, 2, ..., K, \tag{7.1}$$

where $\mathbf{M}_k$, $\mathbf{B}_k$, $\boldsymbol{\Lambda}_k$, and $\mathbf{n}_k$ denote motion, blurring effect, downsampling operator, and noise term for the kth LR image, respectively, and K is the number of images. In the above formula, LR and HR frames are rearranged in lexicographic order; therefore, $\mathbf{i}$ is of size $\sigma^2 mn \times 1$, $\mathbf{a}_k$ and $\mathbf{n}_k$ are $mn \times 1$, $\mathbf{B}_k$ and $\mathbf{M}_k$ are $\sigma^2 mn \times \sigma^2 mn$, and $\boldsymbol{\Lambda}_k$ is $mn \times \sigma^2 mn$.

Assuming the same downsampling ratio for all LR images, we can modify the above observation

model as follows:

$$\mathbf{a}_k = \mathbf{\Lambda B}_k \mathbf{M}_k \mathbf{i} + \mathbf{n}_k, \quad k = 1, 2, ..., K. \tag{7.2}$$

As in [63], we change the order of motion and blur matrices, which is assumed to be equivalent to the above model in (7.2):

$$\mathbf{a}_k = \mathbf{\Lambda M}_k \mathbf{B}_k \mathbf{i} + \mathbf{n}_k. \tag{7.3}$$

Given a sequence of decompressed LR images, $\mathbf{a}_k$, our goal is to reconstruct the unknown HR frames, $\mathbf{i}$.

7.2.3 Super Resolution Reconstruction

As before, we assume that the given LR image sequence (i.e., decompressed LR frames) corresponds to the lowpass subbands of unknown HR images. The goal is to reconstruct the unknown highpass subbands of these HR frames. Fig. 7.2 shows the SR process for one of the LR frames, where gray areas in dashed lines indicate unknown images and coefficients, input LR sequence in white solid lines stand for the decoded LR frames, and dashed lines present the estimated subbands and images. This process is performed for all LR frames in the decoder side.

SR process consists of two parts as image registration and image reconstruction. For the image registration step, we employ our devised methodology in Chapter 5, by dividing the images into blocks, and estimating motion for each block using uni-directional full search. We calculate the motion vectors from a reference LR frame to three more target LR frames, and use the information of these motion vectors to create the F, K, and L matrices, derived in Section 5.2.
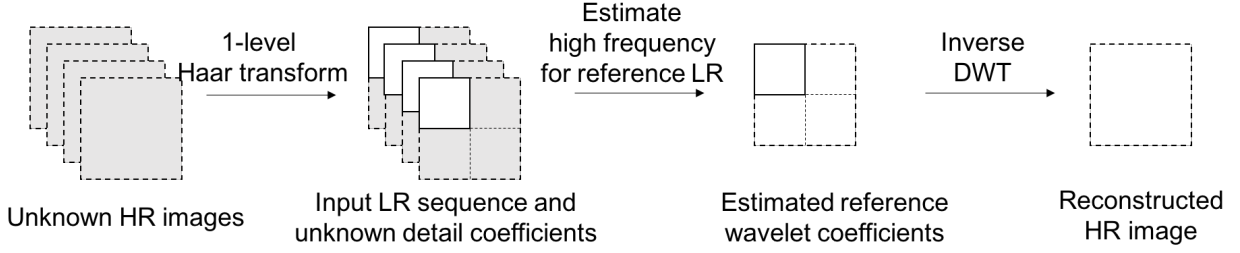
Figure 7.2: Super resolution of decoded LR frames.

For the image reconstruction part, in order to decrease the motion estimation errors caused by moving objects in the scenes, as a modification to the general IBP algorithm described first by Irani and Peleg [65], we propose updating the high-frequency information for the reference LR frame only, and we use the proposed method in Section 5.2 to calculate the high-pass coefficients of the target images. Therefore, the high-pass subbands of the unknown HR image is estimated iteratively by the following formula:

$$
\begin{bmatrix} \hat{\mathbf{H}} \\ \hat{\mathbf{V}} \\ \hat{\mathbf{D}} \end{bmatrix}^{(n+1)} = \begin{bmatrix} \hat{\mathbf{H}} \\ \hat{\mathbf{V}} \\ \hat{\mathbf{D}} \end{bmatrix}^{(n)} + \lambda \sum_{k} (\mathbf{A}_k - \hat{\mathbf{A}}_k^{(n)}) h^{BP}, \tag{7.4}
$$

where n stands for the iteration number, λ is a step size, $\hat{\mathbf{A}}_k^{(n)}$ is the estimated LR frame at iteration n, and h^{BP} is a back projection kernel as defined in [65].

We estimate $\hat{\mathbf{A}}_k$ using the reference frame, the high-pass subbands $\mathbf{H}$, $\mathbf{V}$, and $\mathbf{D}$ for the reference frame at the current iteration, and the motion vectors. Back projection is also performed using our method in Section 5.2 by first upsampling the difference between observed and estimated LR frames, translating the subbands of difference images in-band, back to the reference image grid, and deblurring, respectively. Before updating the reference image subbands, we employ bilateral

filtering on the error images as in [216], in order to avoid the ringing effects caused by IBP.

The proposed algorithm can also be explained step by step as in the following algorithm.

Algorithm *Super Resolution of Compressed Video*
- *Input*: Decoded LR frames and motion vectors for each block

- *Objective*: Reconstruct high frequency information for decoded LR frames

- *Output*: HR frames

◇ Choose a reference LR frame

◇ Initialize horizontal, vertical, and diagonal detail coefficients

◇ Construct coefficient matrices (i.e., F, K, L) derived in Section 5.2 for all target LR frames, using motion vectors for each block

◇ Until a predefined maximum number of iterations, do:

  – Use wavelet subbands of reference LR frame with constructed matrices to estimate target LR frames
  – Upsample the difference between observed and target LR frames, and perform DWT on the error frames
  – Translate high-pass subbands of the error frames back to reference frame coordinates
  – Filter subbands of the error frames with bilateral filtering
  – Update reference subbands

◇ Perform inverse DWT on reference LR frame and estimated high-pass subbands

7.3   Experimental Results

This section is devoted to demonstrate the results and comparisons for the super resolution of compressed video method. We use CIF video sequences of resolution 352×288. We set the block size for motion estimation method to 32×32; therefore, the results are based on 320×256 regions of the sequences. Since H.264 is the state-of-the-art encoder used by most super resolution

112

of compressed video methods [44, 133, 135, 197, 199], we encode the LR frames using H.264 intra coding, with several values of quantization parameter. Moreover, motion vectors are encoded using context-adaptive variable-length coding (CAVLC). We compare our results to bilinear interpolation, bicubic interpolation, and iterative back projection (IBP) method [65]. The comparisons are based on the PSNR value, which is the measurement method used for evaluation, in super resolution of compressed video studies.
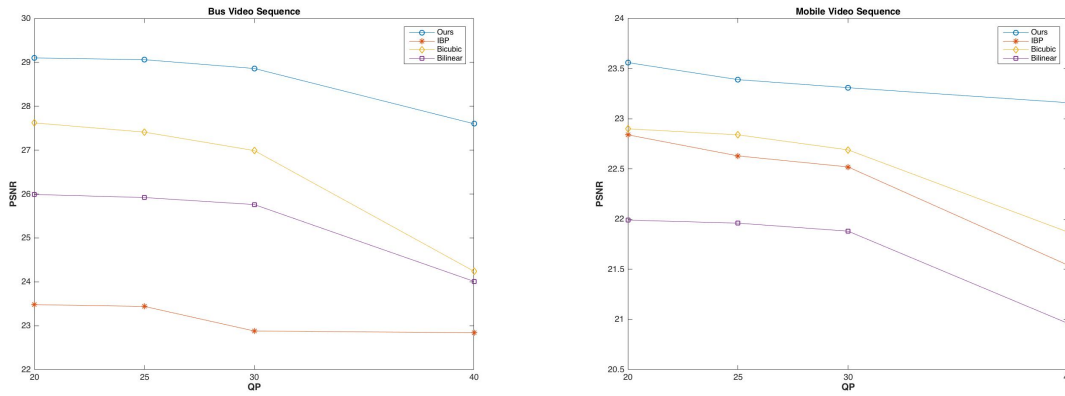
Figure 7.3: PSNR results for different quantization parameters (QP) of H.264 encoder.

Figure 7.3 demonstrates the results obtained by our method, bilinear, bicubic interpolations and the IBP algorithm. We can see from the graphs that, even when the quantization parameter is high (i.e., $QP = 40$), the PSNR results of our method outperform the compared methods.

Figures 7.4 and 7.6 show a subjective comparison of the methods, for a zoomed part of Bus and Mobile video sequences, respectively, with $QP = 20$; while, Figures 7.5 and 7.7 demonstrate the results for $QP = 40$. The reference image is selected as the first frame in the sequence, and reconstruction is performed using four frames. The details of reconstruction for the proposed method can especially be observed around the white pillar of the car in Bus sequence, and the sunshade of the store (on the right) in Mobile sequence. We can observe the edges of the sunshade in Fig. 7.7, even with $QP = 40$.

113

Figure 7.4: Comparison of reconstructions with QP = 20 for Bus video sequence **a** HR reference image **b** LR reference image **c** Bilinear interpolation **d** Bicubic interpolation **e** IBP **f** Ours.
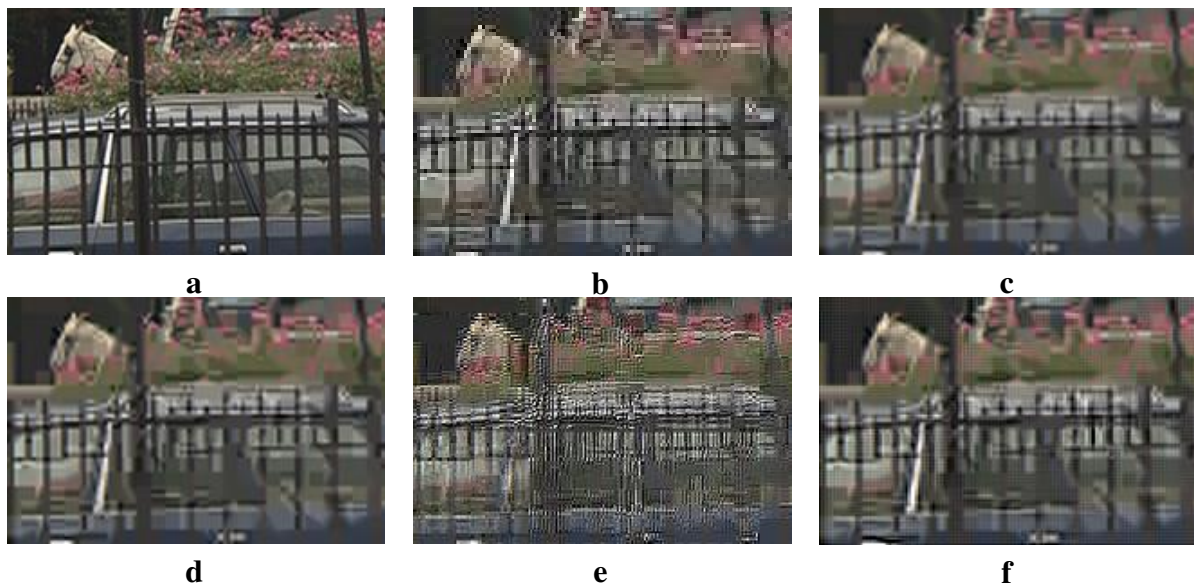
Figure 7.5: Comparison of reconstructions with QP = 40 for Bus video sequence **a** HR reference image **b** LR reference image **c** Bilinear interpolation **d** Bicubic interpolation **e** IBP **f** Ours.
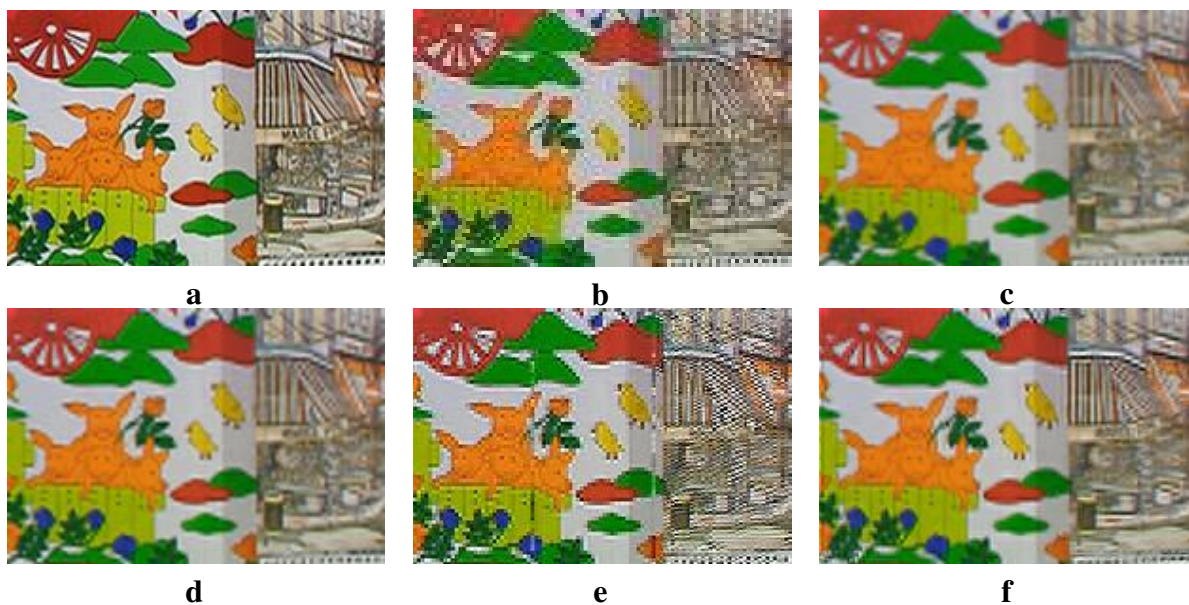
Figure 7.6: Comparison of reconstructions with QP = 20 for Mobile video sequence **a** HR reference image **b** LR reference image **c** Bilinear interpolation **d** Bicubic interpolation **e** IBP **f** Ours.
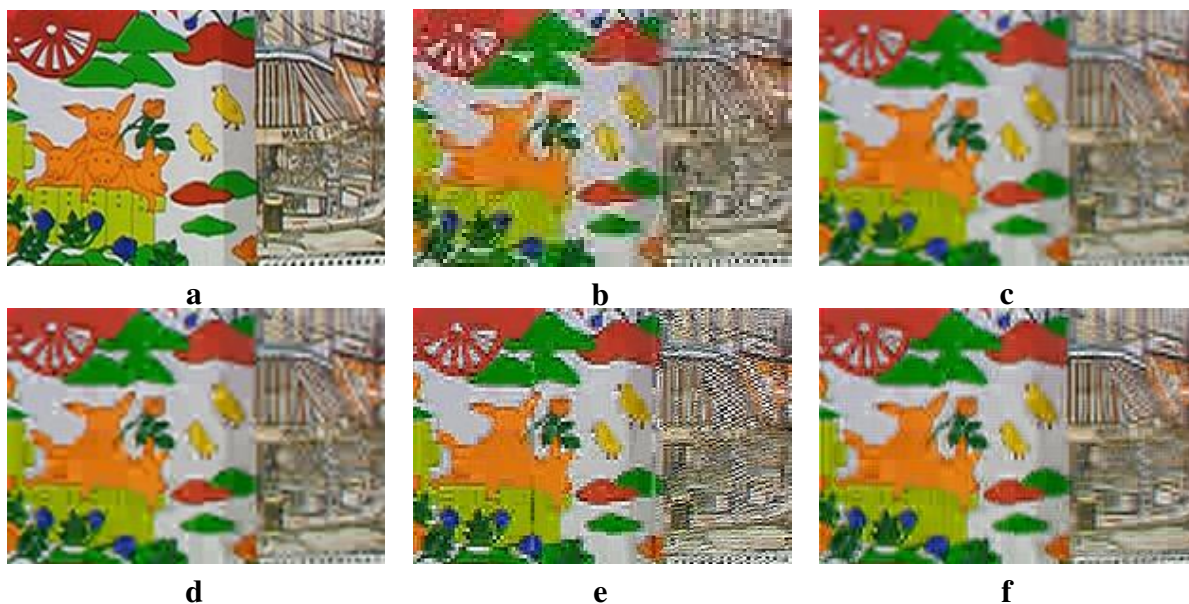
Figure 7.7: Comparison of reconstructions with QP = 40 for Mobile video sequence **a** HR reference image **b** LR reference image **c** Bilinear interpolation **d** Bicubic interpolation **e** IBP **f** Ours.

7.4   Conclusion

In this chapter, we present a framework for video compression, which encodes low resolution images, instead of full resolution images, in order to decrease the required bitrate to encode the video files. These low resolution images are encoded together with the motion information, and are later upsampled by a wavelet-based super resolution method in the decoder side. Since the goal of super resolution of compressed video problem is to reconstruct the high-frequency information lost during downsampling and compression, our super resolution methodology, which utilize the wavelet low-pass subbands to recover the high-pass subbands and work directly in the wavelet-domain, is well-suited for this problem. Experimental results also demonstrate the accuracy of the proposed framework, which can also be enhanced further, using a bidirectional search with variable block sizes for motion estimation, to decrease the error caused by moving objects.

CHAPTER 8: SUPER RESOLUTION OF PANSHARPENED MULTISPECTRAL IMAGES

Chapters 4 and 7 demonstrate that in-band image registration and super resolution methods result in more accurate motion estimation and higher quality reconstruction of images. This chapter is devoted to demonstrate the effective usage of these methods in a satellite imagery area, for enhancing the resolution of pansharpened multispectral images. This chapter is organized as follows. In Section 8.1, we introduce our solution to increase the resolution of multispectral images together with the pansharpening method used; and in Sections 8.2 and 8.3, we demonstrate the experimental results and conclude this chapter, respectively.

8.1   Super Resolution of Pansharpened Multispectral Images

In this section, we will first present the notation used throughout the chapter, and later explain the employed pansharpening and super resolution methodologies.

8.1.1   Notation for Super Resolution of Pansharpened Multispectral Images Method

Table 8.1: Notation for Super resolution of Pansharpened Multispectral Images Method.

I	Reference HR frame
A, H, V, D	Haar wavelet transform approximation, horizontal, vertical, and diagonal subbands of HR frame, respectively
F, K	Matrices to be multiplied by approximation and detail coefficients (i.e., A, H, V, D) of the reference HR image

Again, uppercase bold letters in the following sections demonstrate matrices, lowercase bold ones

show vectors, and subscripts x and y indicate the directions.

8.1.2 Pansharpening

We chose Additive Wavelet Luminance Proportional (AWLP) [148] method for the pansharpening step of our algorithm, based on the comparison results obtained by Alparone *et al.* [146] and Bovolo *et al.* [147].

The AWLP method is an extended version of the additive wavelet luminance technique (AWL) which is designed for three-band (RGB) multispectral images and works in the IHS domain. This method injects high frequency information of the PAN image to MS images proportional to their original values in order to preserve the radiometric signature of MS images. The AWLP method generalizes the AWL method to include arbitrary number of bands, as in the formula below:

$$A_{P_i} = A_i + \frac{A_i}{\sum_{i=1}^{L} A_i} \sum_{j=1}^{n} DWT(P), \tag{8.1}$$

where A and A_P are low resolution and pansharpened MS bands, and P shows the PAN image, respectively. L is the number of bands, n is the number of DWT decomposition levels, and $DWT(P)$ is the DWT decomposition of the PAN image. Fig. 8.1 demonstrates pansharpening for four sets of multispectral bands and PAN images.

8.1.3 Super Resolution of Pansharpened Images

Spatial resolution of PAN images differ based on the multispectral sensors. For example, while Landsat 7 ETM+ sensors provide 15 m spatial resolution PAN images, Quickbird sensors increase this amount upto 61 cm. In order to increase the available spatial resolution of PAN images while

keeping the spectral resolution provided by MS bands, we propose using both temporal and spatial information accessible via most multispectral sensors. In order to achieve our goal, we use pansharpened MS images to perform multiframe SR in order to exceed the spatial resolution of available PAN images. Fig. 8.1 shows a pictorial explanation of the proposed scheme.
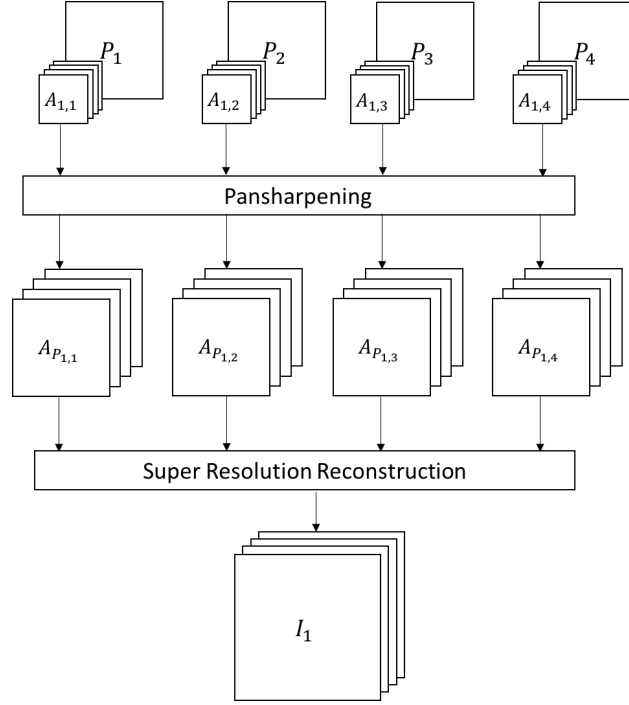
Figure 8.1: Flowchart of the proposed method for super resolution of pansharpened multispectral images.

We can demonstrate the SR observation model, using the pansharpened multispectral images as follows:

$$\mathbf{a}_{P_{i,k}} = \mathbf{\Lambda}\mathbf{M}_{i,k}\mathbf{B}_{i,k}\mathbf{i}_i + \mathbf{n}_{i,k}, \tag{8.2}$$

where $\mathbf{a}_P$ shows the pansharpened multispectral image and $\mathbf{i}$ demonstrates the desired higher resolution multispectral image, in lexicographical order. $\mathbf{\Lambda}$, $\mathbf{M}$, $\mathbf{B}$, and $\mathbf{n}$ indicate downsampling (same

scale for both directions), motion, blur, and noise, as before. Subscript i and k indicate the number of bands and number of images in the LR set, respectively, for $i = 1, ..., L$, $k = 1, ..., K$.

Here, we assume that the pansharpened MS bands are the lowpass subbands of unknown higher resolution MS bands. The goal is to reconstruct the unknown highpass subbands of the reference MS bands. We can write the iterative procedure as follows:

$$\mathbf{I}_i^{n+1} = \mathbf{I}_i^n + \lambda \sum_k (\mathbf{A}_{P_{i,k}} - \hat{\mathbf{A}}_{P_{i,k}}) h^{BP}, \tag{8.3}$$

$$
\begin{aligned}
\mathbf{I}_i^{n+1} &= \mathbf{I}_i^n \\
&+ \lambda \sum_k \left\{ \left(\mathbf{A}_{i,k} + \frac{\mathbf{A}_{i,k}}{\sum_{i=1}^{L} \mathbf{A}_{i,k}} \sum_{j=1}^{n} DWT(\mathbf{P})_k \right) - \mathrm{tr}(\mathbf{X}_i) \right\} h^{BP},
\end{aligned} \tag{8.4}
$$

where tr indicates the trace of matrix $\mathbf{X}$, and,

$$
\mathbf{X} = \begin{bmatrix} \mathbf{F}_{y_k} & & & \\ & \mathbf{F}_{y_k} & & \\ & & \mathbf{K}_{y_k} & \\ & & & \mathbf{K}_{y_k} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{i,1} + \frac{\mathbf{A}_{i,1}}{\sum_{i=1}^{L} \mathbf{A}_{i,1}} \sum_{j=1}^{n} DWT(\mathbf{P})_k & & & \\ & \hat{\mathbf{H}}^n & & \\ & & \hat{\mathbf{V}}^n & \\ & & & \hat{\mathbf{D}}^n \end{bmatrix} \begin{bmatrix} \mathbf{F}_{x_k} & & & \\ & \mathbf{K}_{x_k} & & \\ & & \mathbf{F}_{x_k} & \\ & & & \mathbf{K}_{x_k} \end{bmatrix}
$$

Equation 8.4 shows the relationship between the low-spatial resolution MS bands ($\mathbf{A}$), PAN images ($\mathbf{P}$), and the unknown higher-spatial resolution MS bands ($\mathbf{I}$) that we want to recover. We perform the proposed scheme on each band (i.e. i) of multitemporal pansharpened MS images.

As we have explained before in Section 7.2, SR process consists of two parts as image registration and image reconstruction. Since satellite images are assumed to have translational shifts between them, we generalize our in-band subpixel shift method to be used for this task. To this goal, we incorporate our motion estimation method in Chapter 5 to the subpixel method in Chapter 4. We first perform full search for pixel accuracy, by using our in-band shift method in Section 5.2. In

order to decrease the computational complexity, we divide the search into two steps. First, full search for every five pixels is performed. The result obtained in the first step is later refined to pixel accuracy with full search for every pixel, in a $[-5, 5]$ range. Finally, we perform the proposed subpixel registration method in Section 4.2.3.

Since we assume that bands of an MS image captured by the same sensor are already aligned, we perform registration between MS images taken at different dates, on one band only, and apply the same registration parameters to all bands.

The proposed algorithm can also be explained step by step as in **Algorithm** - Super Resolution Reconstruction of Pansharpened MS images.

Algorithm *SR of Pansharpened MS images*
- *Input*: Multitemporal MS and PAN bands

- *Objective*: Obtain high resolution MS images, exceeding the PAN band spatial resolution

- *Output*: High spatial/high spectral resolution MS images

▶ Pansharpening

  ◇ Perform pansharpening using AWLP, on a set of multispectral bands and PAN images, taken at different dates

▶ Super Resolution

  ◇ Image registration

    ∗ Register the first band of all target images to the first band of reference pansharpened image

  ◇ Image reconstruction (do for all bands)

    ∗ Initialize high-frequency information for the HR image
    ∗ Construct coefficient matrices defined in Section 5.2 for all LR images, using registration parameters found for translation
    ∗ Until a predefined number of maximum iterations, do:
      · Use wavelet subbands with constructed matrices to estimate LR images
      · Update HR image using Eq. (8.4)

121

8.2   Experimental Results

We test our proposed method using Landsat 7 ETM+ images taken at different dates, which have seven MS bands together with a PAN band. The spectral resolution of MS bands range from 0.45μm to 2.35μm, while PAN bands span $0.52 - 0.9\mu$m spectrum; and the spatial resolution of MS bands are 30 m, whereas PAN bands are 15 m. We select four multispectral image groups together with their PAN bands from a region in Istanbul, Turkey, captured on July 2, September 4, in 2000, and May 18, August 6, in 2001, shown in Fig. 8.2 (Images are courtesy of USGS Glovis). We conduct two sets of tests which are categorized as simulated and real experiments. All tests are carried after MS and PAN bands are fused in pansharpening.

Figure 8.2: Real pansharpened MS images (RGB bands).

8.2.1   Simulated Dataset

Since there is no ground truth for the proposed method, first test is a simulated experiment where one of the pansharpened MS images is chosen as reference, all its bands are shifted in horizontal, vertical, and diagonal directions for one pixel, convolved with a Gaussian filter, and downsampled, which is a conventional method used for simulated SR experiments [217]. The pansharpened MS

image bands chosen as reference is then used as the ground truth in comparisons.

For image registration, one of the bands of the reference pansharped MS image is chosen as the reference band, and image registration is performed for the same band of all datasets. Since we define our in-band shift method using subpixels and circular shifts, image regions are adjusted in order to cover the same area, after the registration step. We then initialize our HR estimate using the inverse transform of known reference LR and upsampled wavelet subbands of this LR image. The iterative method described in Section 8.1 is then applied in order to estimate the reconstructed HR image. We compare our results both qualitatively and quantitatively with conventional interpolation techniques and the Iterated Back Projection (IBP) method [65], since the proposed method is a modified IBP model. All compared methods are given the same pansharpened MS images as input. Quantitative comparisons are based on Peak-Signal-to-Noise-Ratio (PSNR) and Structural Similarity Index (SSIM).

Fig. 8.3 (a) shows the reference image used in simulated tests. All figures for simulated and real experiments, show a composite of R, G, and B bands.

Fig. 8.3 shows reference HR and reference LR images together with the compared methods including Bilinear interpolation, Bicubic interpolation and IBP method [65]. In order to comprehend the results, Figs. 8.4 and 8.5 provide zoomed areas of all images in Fig. 8.3. As can be seen from these figures, the proposed method preserves spectral information of the MS bands while increasing the spatial resolution. Quantitative comparisons of the figures confirm that the proposed method reconstructs edges better than the compared ones, while preventing ringing artifacts as in the compared IBP method.

a HR reference   **b** LR reference   **c** Linear

d Bicubic   **e** IBP [65]   **f** Ours
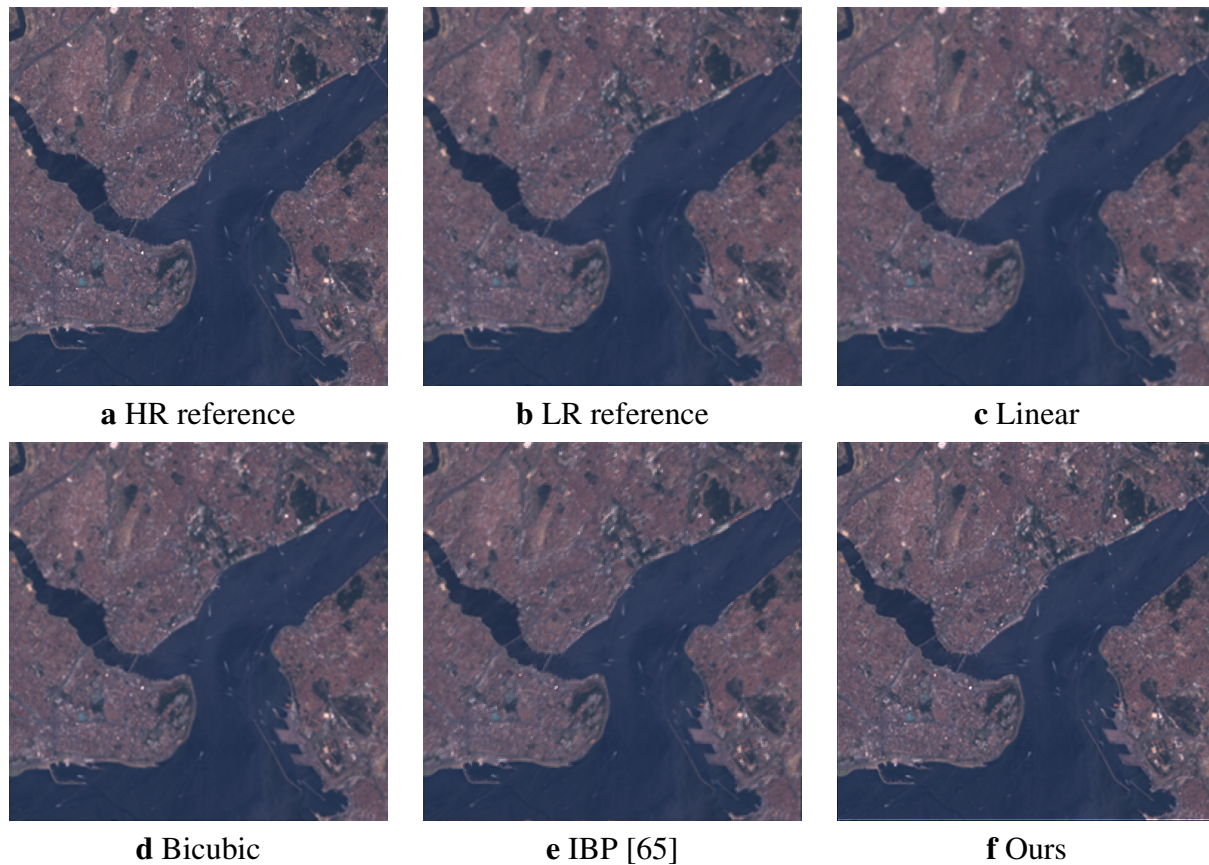
Figure 8.3: Simulated dataset results comparison.

Table 8.2 demonstrates the qualitative results based on PSNR and SSIM values for each band, for a resolution enhancement factor of two. Since pansharpening methods use MS bands numbered $1, 2, 3, 4, 5,$ and 7 for Landsat 7 ETM+, we compare the results for these bands. Qualitative comparisons also validate the quantitative ones. In general, the proposed method preserves the spectral information better while increasing the spatial resolution.

a HR reference   b LR reference   c Linear

d Bicubic   e IBP [65]   f Ours
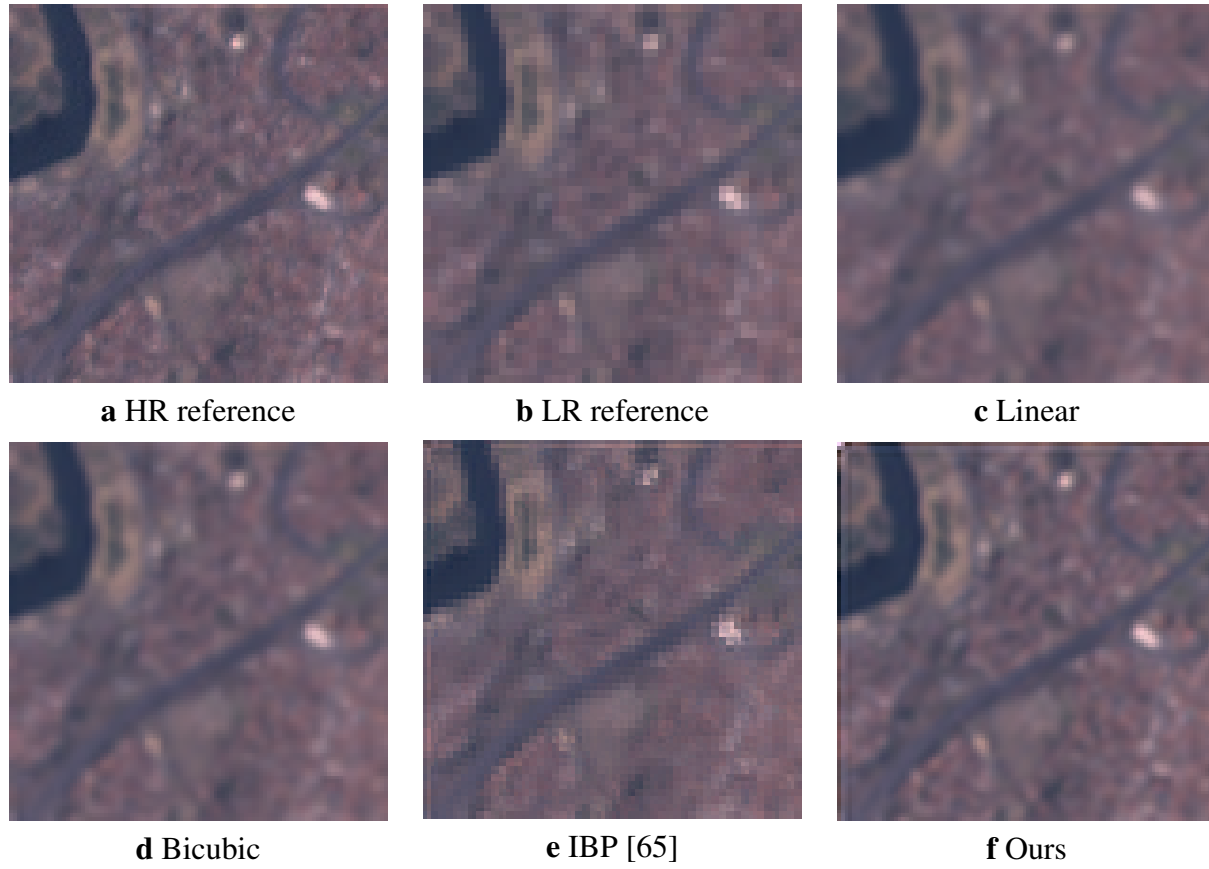
Figure 8.4: Zoomed areas of simulated dataset results in Fig. 8.3.

Table 8.2: Comparison of proposed method with other methods in PSNR, MSE, SSIM.

Band	Linear		Bicubic		IBP		Proposed	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
1	32.47	0.87	34.12	0.90	27.62	0.72	37.08	0.95
2	34.61	0.92	36.71	0.95	29.51	0.82	40.21	0.97
3	36.15	0.93	38.22	0.95	31.54	0.84	41.70	0.98
4	33.20	0.94	35.04	0.96	27.04	0.85	37.88	0.98
5	31.79	0.81	33.23	0.85	27.03	0.64	35.88	0.92
7	32.49	0.85	34.23	0.89	27.41	0.69	37.21	0.94

| **a** HR reference | **b** LR reference | **c** Linear |
| **d** Bicubic | **e** IBP [65] | **f** Ours |
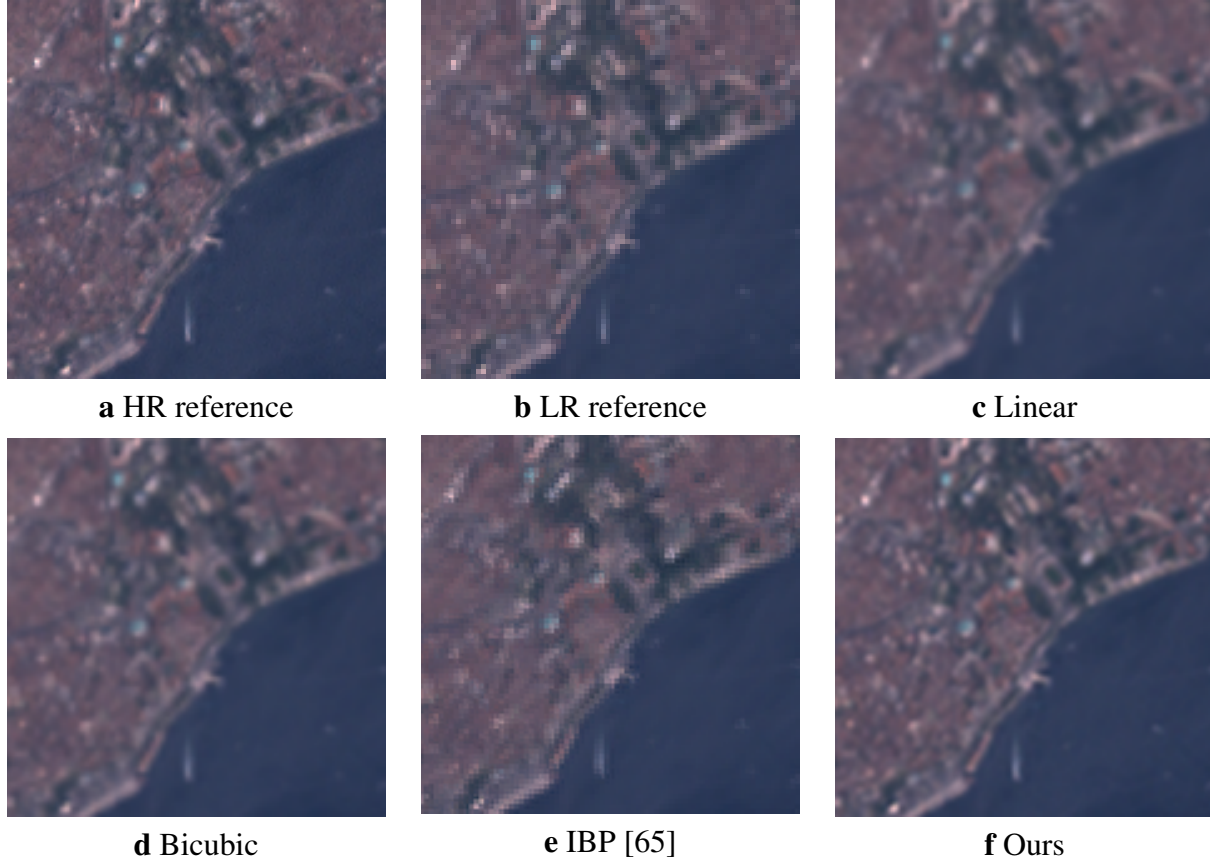
Figure 8.5: Zoomed areas of simulated dataset results in Fig. 8.3.

8.2.2 *Real Dataset*

For the second test, we use pansharpened MS bands as our LR image set, and estimate an HR image without a ground truth; where we compare the results qualitatively. Fig. 8.2 shows the real dataset used for the experiments.

Fig. 8.6 demonstrates the reconstruction results obtained by Nearest Neighbor, Bilinear and Bicubic interpolation, IBP method, and the proposed model. When ground area changes over time (e.g. see in Fig. 8.2, caused by ships on the Sea of Marmara), the accuracy of SR decreases due to the fact that multitemporal images are fused with the proposed method. The zoomed subregions in

126

Figs. 8.7 and 8.8 of the results in 8.6, that have mostly constructions and forests that have the least change during the used time period, demonstrate the reconstruction details better.

a LR reference   **b** Nearest neighbor   **c** Linear

d Bicubic   **e** IBP [65]   **f** Ours
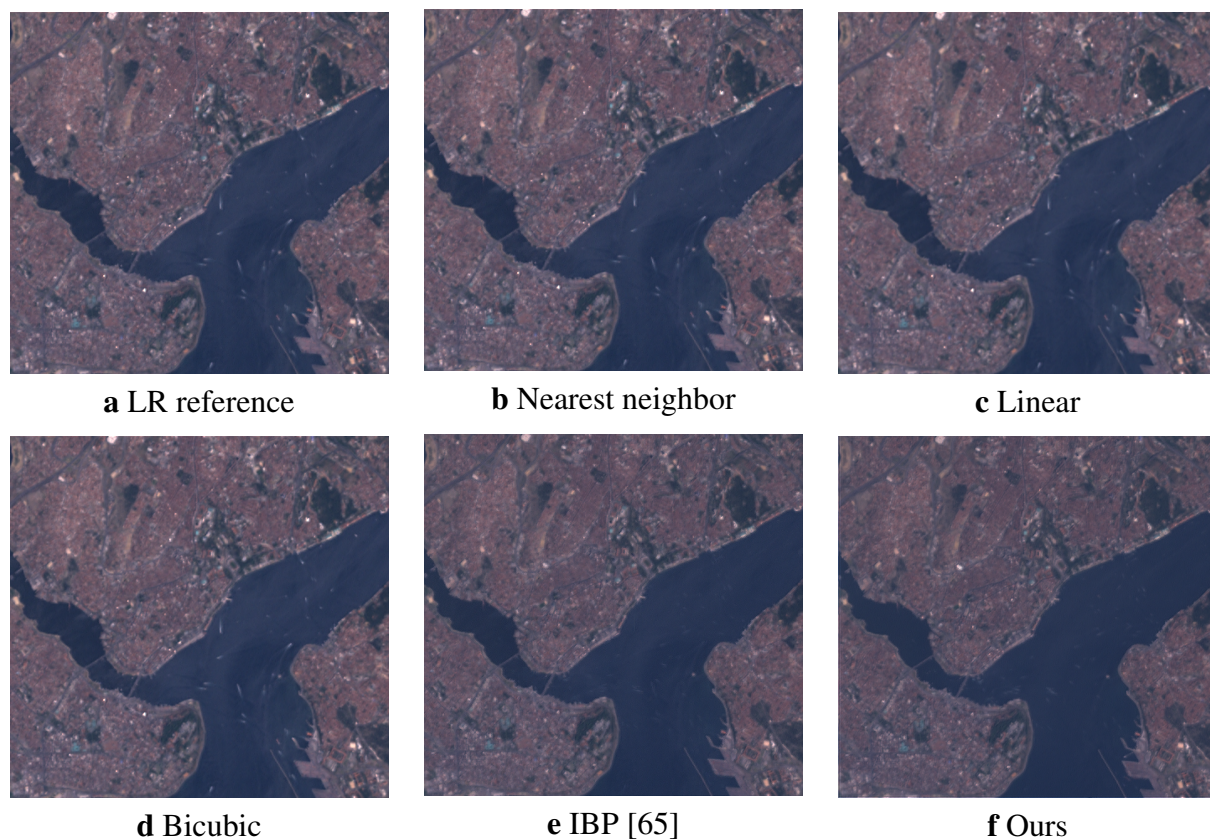
Figure 8.6: Real dataset results comparison.

Experiments with the real dataset also approve the ones with the simulated results. One can see that the edges of roads and constructions in Fig. 8.6 (f) are well reconstructed. As seen in the zoomed areas in Figs. 8.7 and 8.8, proposed method smoothes jaggy artifacts, generates sharper edges, while keeping the spectral information at the same time.

a LR reference   b Nearest neighbor   c Linear

d Bicubic   e IBP [65]   f Ours
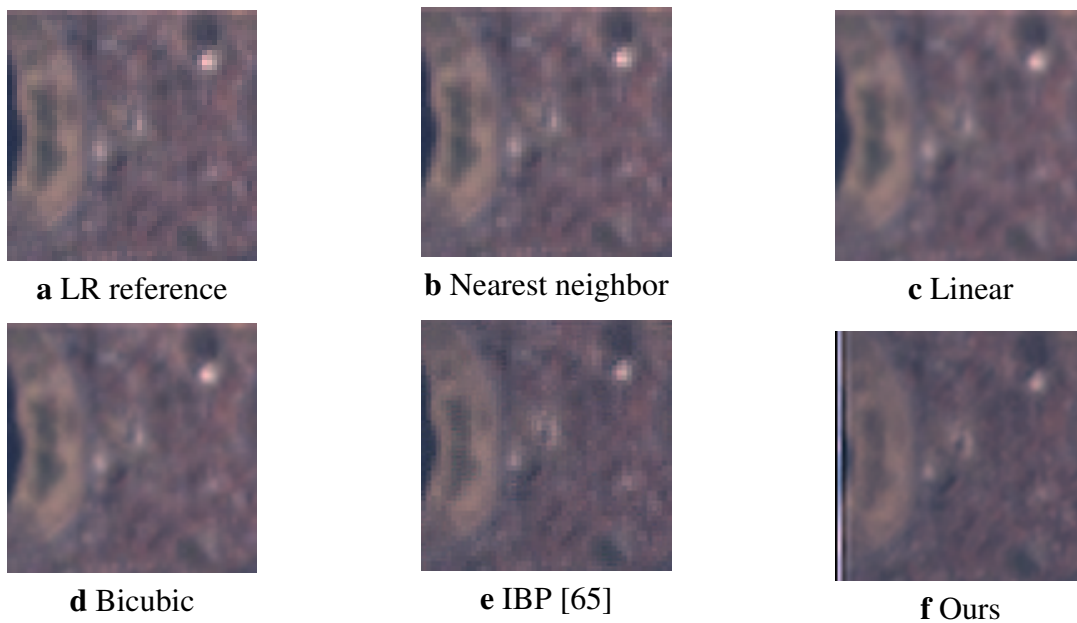
Figure 8.7: Zoomed areas of real dataset results in Fig. 8.6.

a LR reference   b Nearest neighbor   c Linear

d Bicubic   e IBP [65]   f Ours
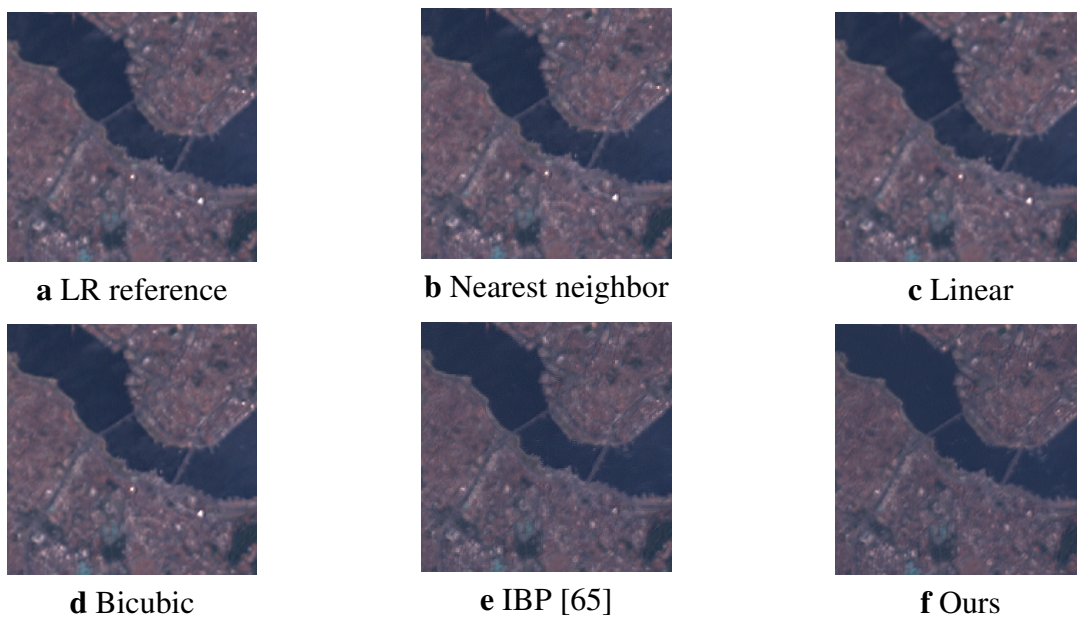
Figure 8.8: Zoomed areas of real dataset results in Fig. 8.6.

8.3   Conclusion

Numerous multispectral satellite imaging applications such as change detection, weather forecasting, and land-cover classification require high-spatial/spectral resolution MS images; yet, MS sensors provide low-spatial-resolution MS and high-spatial-resolution PAN images separately. Pansharpening or SR methods are used in order to obtain high-spatial resolution MS bands without losing spectral data. However, these methods either use only the spatial information of PAN images, or only multitemporal data in MS bands. In this chapter, we propose employing pansharpening and SR methods together, to exceed the spatial resolution available in the PAN bands, by using both spatial and temporal data captured by multispectral sensors. We perform a state-of-the-art pansharpening method before the proposed image reconstruction technique. SR is performed in a modified IBP manner by revealing the interband relationship of multitemporal image subbands. Experimental results demonstrate that the proposed scheme indeed exceeds the spatial resolution of PAN bands, while keeping the spectral information of MS images.

CHAPTER 9: CONCLUSION

We devote this chapter to summarize and conclude the works explained in the previous chapters.

In this dissertation, we mainly focus on the multiframe super resolution problem. The resolution of an image can be expressed as the details within the image, and several applications such as medical imaging, satellite imagery, and video coding require images with high resolution. Hardware solutions which can be used to increase the resolution of images by adapting the sensor specifications, has already reached its limitations. Super resolution techniques are therefore required as the software solutions to this problem.

The goal of multiframe SR is to fuse information from several LR images. In order for multiframe SR problem to use data from different LR frames, these images should be degraded or aliased with respect to each other. Therefore, a pre-processing step of image registration (i.e., motion estimation) is required to obtain the change between the LR images.

Recent trend in multiframe SR and image registration techniques tends toward wavelet-based methodologies. The main reason behind the use of wavelets is their localization property, both in time and frequency. Due to the benefits such as low computing requirements and orthogonality, wavelets have also attracted attention in imaging (e.g. medical imaging), and image/video coding technologies. Therefore, we concentrate on motion estimation and super resolution methods in the wavelet-domain, which can be employed in wavelet-based imaging and coding frameworks.

We first present, in Chapter 4, a wavelet-based image registration technique. In order to use our method in a wavelet-domain technology with wavelet-encoded images, we use wavelet subbands of the images to be registered, and estimate motion directly in the wavelet-domain (i.e., in band), without back and forth transformations. We propose a registration technique, where a sparse set

of high-pass subbands are used to obtain the registration parameters for scale, rotation, and translation. To achieve our goal, we first decouple scale, rotation, and translation parameters in the wavelet domain (as an analogy to the Fourier domain). Based on the assumption that Haar wavelet subbands can be used to approximate the partial derivatives, we devise methods to recover scale and rotation sequentially, using the relationship between wavelet detail coefficients of two images. After scale and rotation are compensated, translation recovery is performed by using the direct relationship between the subbands of two subpixel shifted images. Extensive experimental results demonstrate the accuracy of the proposed method compared to the state-of-the-art techniques. We utilize standard test images, together with real world examples in our experiments, which are compared under several conditions (e.g. noisy environments, similarity transform, Euclidean transform etc.). Even though the computational complexity of the proposed method is higher than the compared state-of-the-art methods, we observe that the high accuracy of our method can be an acceptable trade-off.

In Chapter 5, we modify our image registration technique, to be used in a practical area, which is motion compensated temporal filtering. MCTF is a method that is widely used for video coding schemes, in order to decrease the bits required to encode video files. The goal of MCTF is to encode the difference between two consecutive frames, instead of the original frames. In order to decrease the difference further, MCTF employs a motion estimation step, and compensates the motion between the frames to find the difference. The most common approach in this area is to find the motion in the spatial domain, and later encode the wavelet transform coefficients. However, this process has drawbacks such as blocking artifacts. A solution to this problem is proposed by finding the motion directly in the transform domain, and later encoding these coefficients. However, since wavelet-transform is shift-invariant, the trivial solution is to use redundancy via overcomplete or dual-tree wavelets. We propose avoiding the need of shift-invariance by using the direct relationship between the two frames. Since the assumption behind MCTF methods is that two consecutive

131

frames in a video differ only by a small translational amount, our devised methodology for image registration can easily be adopted for this problem. Therefore, we propose an MCTF method that follows the concept of performing wavelet transform before motion estimation, and estimate motion using wavelet-domain coefficients directly. We then encode the motion compensated error in the wavelet-domain. Our proposed methodology prevents upsampling, inverse wavelet transform, and redundancy, which leads to reduced memory usage. We test our method using standard CIF video sequences, and demonstrate the accuracy of the proposed method.

After accomplishing the required pre-processing step of image registration for the multiframe SR problem, we explore the proposed wavelet-based multiframe SR problem in Chapter 6. The intrinsic idea behind wavelet-domain multiframe SR methods is to consider the observed LR images as the approximation coefficients (i.e., low-pass subbands) of the wavelet transform for unknown HR images. Since the wavelet transform decomposes the image into approximation and detail coefficients, the goal is then to recover the unknown high-frequency (i.e., high-pass) subbands. In general, this problem is solved in literature by estimating the high-frequency information for all LR images, which is later fused before performing inverse wavelet transform using a reference LR image and estimated subbands.

We propose a different point of view to the wavelet-based multiframe SR problem, by expressing the relationship between the LR frames, in the wavelet domain. To this goal, we first derive equations to express the relationship between low-pass subbands of two subpixel shifted images. This relationship, which depends on the high-pass subbands of the reference LR frame, is then employed to construct a linear system of equations. By solving the linear system, high-pass subbands are estimated. Since we extrapolate the high-frequency information of one of the LR frames using several others, our method can be considered as a hybrid technique between single and multiframe techniques, taking advantage of best of the both groups. The results on standard test images, obtained by our proposed method outperforms the state-of-the-art based on accuracy and

computational complexity (especially compared to regularization-based techniques).

Multiframe super resolution methods are required by many applications. In this dissertation, we present two straightforward applications of our proposed method, which are super resolution of compressed video and super resolution of pansharpened multispectral images. Due to the fact that video coding and pansharpening methods are performed efficiently in the wavelet domain, methods that use wavelet subbands directly are necessary for these areas. Therefore, we first demonstrate the use of our in-band SR method in a compressed video application. As mentioned before, the goal of video coding is to reduce the bits to store or transmit video files. Instead of encoding the motion compensated error as in MCTF methods, a growing amount of literature focuses on encoding LR frames, and employing SR techniques after decoding. Following this concept, we propose a new design of compressed video framework, which first performs wavelet transform on the original frames. We then use our proposed motion estimation method, in order to find the change between blocks in consecutive frames. The motion vectors and LR frames are then encoded into a bitstream. After the decompression step, LR frames and motion vectors are used in an SR technique. In order to decrease the errors caused by motion estimation, we reformulate our SR methodology into an iterative back projection framework. Instead of updating the entire HR estimate in every iteration, we assume that the LR frames are the real low-pass subbands of the wavelet transform of unknown HR frames; and update high-pass subbands only. In order to reduce the ringing artifacts caused by the original IBP algorithm, we perform bilateral filtering on the back projected errors. Again, as in the MCTF method, we use standard CIF video sequences to show the accuracy of the proposed framework.

Finally, we also present the effective usage of our wavelet-based multiframe SR method, in a satellite imaging area. Satellite sensors provide images with different spectral and spatial resolutions. We focus on the multispectral sensors, which provide a few spectral bands with low-spatial-high-spectral resolution (i.e., multispectral bands), and one panchromatic image (i.e., PAN) with

high-spatial-low-spectral resolution. Many applications including change detection and weather forecasting need high resolution MS bands, which can be achieved by injecting the high-frequency information of PAN bands to each MS band. These methods are employed to increase the resolution of MS bands to the spatial resolution of available PAN images. Our goal, on the other hand, is to exceed this resolution, in order to reach the resolution that could be captured by different sensors, in a more costly manner. Therefore, we propose to use information from both spatial and temporal domains. Given LR MS band sequences taken at different times, from the same ground area, we first pansharpen these bands using the spatial information of PAN images. Later, we assume that these pansharpened images are low-pass subbands of unknown higher resolution MS bands; and, utilize the pansharpened images in our wavelet-based SR methodology. We employ a state-of-the-art pansharpening method, that is performed in the wavelet domain. Later, we generalize our subpixel translation recovery in Chapter 4 to find translational shift between images. As in the proposed super resolution of compressed video methodololgy, we use our SR method in iterative back projection fashion to find the high-resolution images for MS bands, sequentially. We demonstrate our results on Landsat ETM+ images taken at different times, and demonstrate the effective recovery of high-pass subbands, while still keeping the spectral information.

As a conclusion, we have highlighted in this dissertation, the effective and efficient usage of the inter-subband relationship of wavelets, in practical areas. Our findings provide evidence that the growing trend in wavelet-based imaging technologies would benefit from the methodologies described in the previous chapters, due to the fact that these techniques are designed solely using wavelet-subbands. In the future, we would like to employ our devised MCTF and SR of compressed video methods in a straightforward product of scalable video coding, where the resolution of the decoded video depends on the user needs. By using a base layer of bitstream, together with enhancement layers, scalable video coding adapts the resolution on the decoder side based on the needs of the receiver end. Furthermore, another possible area of future research would be to inves-

tigate the generalization of the relationship defined for Haar wavelet transform, to other wavelet transforms, which could also be generalized to include affine transform of subbands, instead of only translational shifts. The generalization to other wavelets and affine transformation between images, can be applied in a dense matching scheme, where instead of matching the pixels of images, dense image patches are aligned to match different scenes. Dense matching is a growing interest in the recent years, which can be used for applications such as face recognition and image classification. Moreover, considering the fact that learning-based methods have emerged as a powerful tool in SR techniques, we would like to employ our wavelet-based methods in learning-based frameworks as future research.

APPENDIX A: COPYRIGHT PERMISSION

SIGNAL, IMAGE AND VIDEO PROCESSING

This Agreement between Mrs. Vildan Atalay Aydin ("You") and Springer ("Springer") consists of your license details and the terms and conditions provided by Springer and Copyright Clearance Center.

License Number	4202730414716
License date	Oct 05, 2017
Licensed Content Publisher	Springer
Licensed Content Publication	Signal, Image and Video Processing
Licensed Content Title	In-band subpixel registration of wavelet-encoded images from sparse coefficients
Licensed Content Author	Vildan Atalay Aydin, Hassan Foroosh
Licensed Content Date	Jan 1, 2017
Type of Use	Thesis/Dissertation
Portion	Full text

APPENDIX B: COPYRIGHT PERMISSION

IEEE COPYRIGHT AND CONSENT FORM

To ensure uniformity of treatment among all contributors, other forms may not be substituted for this form, nor may any wording of the form be changed. This form is intended for original material submitted to the IEEE and must accompany any such material in order to be published by the IEEE.

Motion Compensation using Critically Sampled DWT Subbands for Low-Bitrate Video Coding
Vildan Atalay Aydin, Hassan Foroosh
2017 IEEE International Conference on Image Processing (ICIP)

COPYRIGHT TRANSFER

Name
Vildan Atalay

Date
08/02/2017

APPENDIX C: COPYRIGHT PERMISSION

MULTIDIMENSIONAL SYSTEMS AND SIGNAL PROCESSING

140

This Agreement between Mrs. Vildan Atalay Aydin ("You") and Springer ("Springer") consists of your license details and the terms and conditions provided by Springer and Copyright Clearance Center.

License Number	4202731122968
License date	Oct 05, 2017
Licensed Content Publisher	Springer
Licensed Content Publication	Multidimensional Systems and Signal Processing
Licensed Content Title	A linear well-posed solution to recover high-frequency information for super resolution image reconstruction
Licensed Content Author	Vildan Atalay Aydin, Hassan Foroosh
Licensed Content Date	Jan 1, 2017
Type of Use	Thesis/Dissertation
Portion	Full text

LIST OF REFERENCES

[1] Cmu/vasc image database. `http://vasc.ri.cmu.edu/idb/images/motion/`.

[2] Mdsp super-resolution and demosaicing datasets. `https://users.soe.ucsc.edu/~milanfar/software/sr-datasets.html`.

[3] RY Tsai and Thomas S Huang. Multiframe image restoration and registration. *Advances in computer vision and Image Processing*, 1(2):317–339, 1984.

[4] Xuelong Li, Yanting Hu, Xinbo Gao, Dacheng Tao, and Beijia Ning. A multi-frame image super-resolution method. *Signal Processing*, 90(2):405–414, 2010.

[5] Kamal Nasrollahi and Thomas B Moeslund. Super-resolution: a comprehensive survey. *Machine Vision & Applications*, 25(6):1423–1468, 2014.

[6] Emmanuel J Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.

[7] Linwei Yue, Huanfeng Shen, Jie Li, Qiangqiang Yuan, Hongyan Zhang, and Liangpei Zhang. Image super-resolution: The techniques, applications, and future. *Signal Processing*, 128:389–408, 2016.

[8] Bahadir K Gunturk, Yucel Altunbasak, and Russell M Mersereau. Super-resolution reconstruction of compressed video using transform-domain statistics. *IEEE Transactions on Image Processing*, 13(1):33–43, 2004.

[9] CN Ravi Kumar, VK Ananthashayana, et al. Super resolution reconstruction of compressed low resolution images using wavelet lifting schemes. In *Computer and Electrical Engineering, 2009. ICCEE'09. Second International Conference on*, volume 2, pages 629–633. IEEE, 2009.

[10] Xiaohong Zhang, Min Tang, and Ruofeng Tong. Robust super resolution of compressed video. *The Visual Computer*, 28(12):1167–1180, 2012.

[11] Mohamed M Fouad, Richard M Dansereau, and Anthony D Whitehead. Geometric image registration under arbitrarily-shaped locally variant illuminations. *Signal, Image and Video Processing*, 6(4):521–532, 2012.

[12] Paul Viola and William M Wells. Alignment by maximization of mutual information. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 16–23. IEEE, 1995.

[13] Berthold KP Horn and EJ Weldon. Direct methods for recovering motion. *International Journal of Computer Vision*, 2(1):51–76, 1988.

[14] Hassan Foroosh, Josiane B Zerubia, and Marc Berthod. Extension of phase correlation to subpixel registration. *Image Processing, IEEE Transactions on*, 11(3):188–200, 2002.

[15] Hassan Shekarforoush, Marc Berthod, and Josiane Zerubia. Subpixel image registration by estimating the polyphase decomposition of cross power spectrum. In *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on*, pages 532–537. IEEE, 1996.

[16] Hassan Foroosh and Murat Balci. Sub-pixel registration and estimation of local shifts directly in the fourier domain. In *Image Processing, 2004. ICIP'04. 2004 International Conference on*, volume 3, pages 1915–1918. IEEE, 2004.

[17] Murat Balci and Hassan Foroosh. Estimating sub-pixel shifts directly from the phase difference. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 1, pages I–1057. IEEE, 2005.

[18] Murat Balci and Hassan Foroosh. Inferring motion from the rank constraint of the phase matrix. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, volume 2, pages ii–925. IEEE, 2005.

[19] Murat Balci and Hassan Foroosh. Subpixel estimation of shifts directly in the fourier domain. *IEEE Transactions on Image Processing*, 15(7):1965–1972, 2006.

[20] Murat Balci and Hassan Foroosh. Subpixel registration directly from the phase difference. *EURASIP Journal on Applied Signal Processing*, 2006:231–231, 2006.

[21] Yu-Te Wu, Takeo Kanade, Ching-Chung Li, and Jeffrey Cohn. Image registration using wavelet-based motion model. *International Journal of Computer Vision*, 38(2):129–152, 2000.

[22] Harold S. Stone, Jacqueline Le Moigne, and Morgan McGuire. The translation sensitivity of wavelet-based registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(10):1074–1081, 1999.

[23] Alexander Wong and David A Clausi. Aisir: Automated inter-sensor/inter-band satellite image registration using robust complex wavelet feature representations. *Pattern recognition letters*, 31(10):1160–1167, 2010.

[24] Feng Li, Xiuping Jia, Donald Fraser, and Andrew Lambert. Super resolution for remote sensing images based on a universal hidden markov tree model. *IEEE Transactions on Geoscience and Remote Sensing*, 48(3):1270–1278, 2010.

[25] Jacqueline Le Moigne, William J Campbell, and Robert F Cromp. An automated parallel image registration technique based on the correlation of wavelet features. *IEEE Transactions on Geoscience and Remote Sensing*, 40(8):1849–1864, 2002.

[26] Gang Hong and Yun Zhang. Wavelet-based image registration technique for high-resolution remote sensing images. *Computers & Geosciences*, 34(12):1708–1720, 2008.

[27] Leila MG Fonseca and Max HM Costa. Automatic registration of satellite images. In *Computer Graphics and Image Processing, 1997. Proceedings., X Brazilian Symposium on*, pages 219–226. IEEE, 1997.

[28] JP Djamdji A Bijaoui and R Maniere. Geometrical registration of images: the multiresolution approach. *Photogrammetric Eng. And Remote Sensing J*, 59(5):645–653, 1993.

[29] Hui Ji and Cornelia Fermuller. Robust wavelet-based super-resolution reconstruction: theory and algorithm. *IEEE PAMI*, 31(4):649–660, 2009.

[30] Qi Tian and Michael N Huhns. Algorithms for subpixel registration. *Computer Vision, Graphics, and Image Processing*, 35(2):220–233, 1986.

[31] Barbara Zitova and Jan Flusser. Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000, 2003.

[32] Lawrence P Panych and Ferenc A Jolesz. A dynamically adaptive imaging algorithm for wavelet-encoded mri. *Magnetic resonance in medicine*, 32(6):738–748, 1994.

[33] Matthieu Guerquin-Kern, M Haberlin, Klaas Paul Pruessmann, and Michael Unser. A fast wavelet-based reconstruction method for magnetic resonance imaging. *IEEE transactions on medical imaging*, 30(9):1649–1660, 2011.

[34] Kamrul Hasan Talukder and Koichi Harada. Haar wavelet based approach for image compression and quality assessment of compressed image. *arXiv preprint arXiv:1010.4084*, 2010.

[35] Ruifeng Xu, Sumanta N Pattanaik, and Charles E Hughes. High-dynamic-range still-image encoding in jpeg 2000. *IEEE Computer Graphics and Applications*, 25(6):57–64, 2005.

[36] Shiqian Ma, Wotao Yin, Yin Zhang, and Amit Chakraborty. An efficient algorithm for compressed mr imaging using total variation and wavelets. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[37] Thomas Naselaris, Kendrick N Kay, Shinji Nishimoto, and Jack L Gallant. Encoding and decoding in fmri. *Neuroimage*, 56(2):400–410, 2011.

[38] Yiannis Andreopoulos, Adrian Munteanu, Geert Van der Auwera, Jan PH Cornelis, and Peter Schelkens. Complete-to-overcomplete discrete wavelet transforms: theory and applications. *IEEE Transactions on Signal Processing*, 53(4):1398–1412, 2005.

[39] Ruiqin Xiong, Jizheng Xu, and Feng Wu. In-scale motion compensation for spatially scalable video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(2):145–158, 2008.

[40] G Van der Auwera, A Munteanu, P Schelkens, and J Cornelis. Bottom-up motion compensated prediction in wavelet domain for spatially scalable video coding. *Electronics Letters*, 38(21):1251–1253, 2002.

[41] Hyun-Wook Park and Hyung-Sun Kim. Motion estimation using low-band-shift method for wavelet-based moving-picture coding. *IEEE TIP*, 9(4):577–587, 2000.

[42] Beong-Jo Kim, Zixiang Xiong, and William A Pearlman. Low bit-rate scalable video coding with 3-d set partitioning in hierarchical trees (3-d spiht). *IEEE Transactions on Circuits and Systems for Video Technology*, 10(8):1374–1387, 2000.

[43] Carole Amiot, Catherine Girard, Jérémie Pescatore, Jocelyn Chanussot, and Michel Desvignes. Fluorosocopic sequence denoising using a motion compensated multi-scale temporal filtering. In *ICIP*, pages 691–695. IEEE, 2015.

[44] Minmin Shen, Ping Xue, and Ci Wang. Down-sampling based video coding using super-resolution technique. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(6):755–765, 2011.

[45] Feng Li, Xiuping Jia, and Donald Fraser. Superresolution reconstruction of multispectral data for improved image classification. *IEEE Geoscience and Remote Sensing Letters*, 6(4):689–693, 2009.

[46] Mais Alnasser and Hassan Foroosh. Phase-shifting for nonseparable 2-d haar wavelets. *IEEE Transactions on Image Processing*, 17(7):1061–1068, 2008.

[47] Stephane G Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7):674–693, 1989.

[48] Nhat Xuan Nguyen. *Numerical algorithms for image superresolution*. Stanford University, 2000.

[49] Danny Keren, Shmnel Peleg, and Ran Brada. Image sequence enhancement using sub-pixel displacements. In *Computer Vision and Pattern Recognition, 1988. Proceedings CVPR'88., Computer Society Conference on*, pages 742–746. IEEE, 1988.

[50] Georgios D Evangelidis and Emmanouil Z Psarakis. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1858–1865, 2008.

[51] Li Chen and Kim-Hui Yap. An effective technique for subpixel image registration under noisy conditions. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 38(4):881–887, 2008.

[52] Manuel Guizar-Sicairos, Samuel T Thurman, and James R Fienup. Efficient subpixel image registration algorithms. *Optics letters*, 33(2):156–158, 2008.

[53] Harold S Stone, Michael T Orchard, Ee-Chien Chang, and Stephen A Martucci. A fast direct fourier-based algorithm for subpixel registration of images. *IEEE Transactions on geoscience and remote sensing*, 39(10):2235–2243, 2001.

[54] Patrick Vandewalle, Sabine Süsstrunk, and Martin Vetterli. A frequency domain approach to registration of aliased images with application to super-resolution. *EURASIP Journal on applied signal processing*, 2006:233–233, 2006.

[55] Arthur Goshtasby, Stuart H Gage, and Jon F Bartholic. A two-stage cross correlation approach to template matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (3):374–378, 1984.

[56] Philippe Thevenaz, Urs E Ruttimann, and Michael Unser. A pyramid approach to subpixel registration based on intensity. *IEEE transactions on image processing*, 7(1):27–41, 1998.

[57] Radka Turcajova and Jaroslav Kautsky. A hierarchical multiresolution technique for image registration. In *Proceedings of SPIE Mathematical Imaging: Wavelet Applications in Signal and Image Processing, Colorado*, 1996.

[58] Ulas Bagci and Li Bai. Multiresolution elastic medical image registration in standard intensity scale. In *Computer Graphics and Image Processing, 2007. SIBGRAPI 2007. XX Brazilian Symposium on*, pages 305–312. IEEE, 2007.

[59] Jianglin Ma, Jonathan Cheung-Wai Chan, and Frank Canters. Fully automatic subpixel image registration of multiangle chris/proba data. *IEEE transactions on geoscience and remote sensing*, 48(7):2829–2839, 2010.

[60] H-M Chen, Manoj K Arora, and Pramod K Varshney. Mutual information-based image registration for remote sensing data. *International Journal of Remote Sensing*, 24(18):3701–3706, 2003.

[61] Michael Elad and Arie Feuer. Super-resolution reconstruction of image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(9):817–834, 1999.

[62] Sina Farsiu, M Dirk Robinson, Michael Elad, and Peyman Milanfar. Fast and robust multi-frame super resolution. *IEEE transactions on image processing*, 13(10):1327–1344, 2004.

[63] M Dirk Robinson, Cynthia A Toth, Joseph Y Lo, and Sina Farsiu. Efficient fourier-wavelet super-resolution. *IEEE Transactions on Image Processing*, 19(10):2669–2681, 2010.

[64] Fei Zhou, Wenming Yang, and Qingmin Liao. Interpolation-based image super-resolution using multisurface fitting. *IEEE Transactions on Image Processing*, 21(7):3312–3318, 2012.

[65] Michal Irani and Shmuel Peleg. Improving resolution by image registration. *CVGIP*, 53(3):231–239, 1991.

[66] Nhat Nguyen, Peyman Milanfar, and Gene Golub. A computationally efficient superresolution image reconstruction algorithm. *IEEE transactions on image processing*, 10(4):573–583, 2001.

[67] Di Zhang, Huifang Li, and Minghui Du. Fast map-based multiframe super-resolution image reconstruction. *Image and Vision Computing*, 23(7):671–679, 2005.

[68] Simon Baker and Takeo Kanade. *Super-resolution optical flow*. Carnegie Mellon University, The Robotics Institute, 1999.

[69] Thomas Brox, Andrés Bruhn, Nils Papenberg, and Joachim Weickert. High accuracy optical flow estimation based on a theory for warping. *Computer Vision-ECCV 2004*, pages 25–36, 2004.

[70] Michael K Ng, Huanfeng Shen, Edmund Y Lam, and Liangpei Zhang. A total variation regularization based super-resolution reconstruction algorithm for digital video. *EURASIP Journal on Advances in Signal Processing*, 2007.

[71] WenYi Zhao and Harpreet Sawhney. Is super-resolution with optical flow feasible? *Computer VisionECCV 2002*, pages 599–613, 2002.

[72] Bruce D Lucas, Takeo Kanade, et al. An iterative image registration technique with an application to stereo vision. 1981.

[73] Loïc Baboulaz and Pier Luigi Dragotti. Exact feature extraction using finite rate of innovation principles with an application to image super-resolution. *IEEE Transactions on Image Processing*, 18(2):281–298, 2009.

[74] Liangpei Zhang, Qiangqiang Yuan, Huanfeng Shen, and Pingxiang Li. Multiframe image super-resolution adapted with local spatial information. *JOSA A*, 28(3):381–390, 2011.

[75] Qiangqiang Yuan, Liangpei Zhang, and Huanfeng Shen. Multiframe super-resolution employing a spatially weighted total variation model. *IEEE Transactions on circuits and systems for video technology*, 22(3):379–392, 2012.

[76] CD Kuglin. The phase correlation image alignment method. In *Proc. International Conference on Cyber-netics Society*, pages 163–165, 1975.

[77] James R Bergen, Patrick Anandan, Keith J Hanna, and Rajesh Hingorani. Hierarchical model-based motion estimation. In *European conference on computer vision*, pages 237–252. Springer, 1992.

[78] Yan-Ran Li, Dao-Qing Dai, and Lixin Shen. Multiframe super-resolution reconstruction using sparse directional regularization. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(7):945–956, 2010.

[79] Michael E Tipping and Christopher M Bishop. Bayesian image super-resolution. In *Advances in neural information processing systems*, pages 1303–1310, 2003.

[80] Lyndsey C Pickup, David P Capel, Stephen J Roberts, and Andrew Zisserman. Bayesian methods for image super-resolution. *The Computer Journal*, 52(1):101–113, 2007.

[81] Yan Huang, Wei Wang, and Liang Wang. Bidirectional recurrent convolutional networks for multi-frame super-resolution. In *Advances in Neural Information Processing Systems*, pages 235–243, 2015.

[82] Yu He, Kim-Hui Yap, Li Chen, and Lap-Pui Chau. A nonlinear least square technique for simultaneous image registration and super-resolution. *IEEE Transactions on Image Processing*, 16(11):2830–2841, 2007.

[83] Patrick Vandewalle, Luciano Sbaiz, Joos Vandewalle, and Martin Vetterli. Super-resolution from unregistered and totally aliased signals using subspace methods. *IEEE Transactions on Signal Processing*, 55(7):3687–3703, 2007.

[84] Huanfeng Shen, Liangpei Zhang, Bo Huang, and Pingxiang Li. A map approach for joint motion estimation, segmentation, and super resolution. *IEEE Transactions on Image processing*, 16(2):479–490, 2007.

[85] Hiroyuki Takeda, Sina Farsiu, and Peyman Milanfar. Kernel regression for image processing and reconstruction. *IEEE Transactions on image processing*, 16(2):349–366, 2007.

[86] Matan Protter, Michael Elad, Hiroyuki Takeda, and Peyman Milanfar. Generalizing the nonlocal-means to super-resolution reconstruction. *IEEE Transactions on image processing*, 18(1):36–51, 2009.

[87] Hiroyuki Takeda, Peyman Milanfar, Matan Protter, and Michael Elad. Super-resolution without explicit subpixel motion estimation. *IEEE Transactions on Image Processing*, 18(9):1958–1975, 2009.

[88] Nhat Nguyen and Peyman Milanfar. A wavelet-based interpolation-restoration method for superresolution (wavelet superresolution). *Circuits, Systems, and Signal Processing*, 19(4):321–338, 2000.

[89] CS Tong and KT Leung. Super-resolution reconstruction based on linear interpolation of wavelet coefficients. *Multidimensional Systems and Signal Processing*, 18(2-3):153–171, 2007.

[90] Wei Cai and Jianzhong Wang. Adaptive multiresolution collocation methods for initial-boundary value problems of nonlinear pdes. *SIAM Journal on Numerical Analysis*, 33(3):937–970, 1996.

[91] John W Woods. *Multidimensional signal, image, and video processing and coding*. Academic press, 2011.

[92] Bede Liu and Andre Zaccarin. New fast algorithms for the estimation of block motion vectors. *IEEE Transactions on Circuits and Systems for Video technology*, 3(2):148–157, 1993.

[93] Yih-Chuan Lin and Shen-Chuan Tai. Fast full-search block-matching algorithm for motion-compensated video compression. *IEEE Transactions on communications*, 45(5):527–531, 1997.

[94] Yao Wang, Jörn Ostermann, and Ya-Qin Zhang. *Video processing and communications*, volume 5. Prentice Hall Upper Saddle River, 2002.

[95] Jaswant Jain and Anil Jain. Displacement measurement and its application in interframe image coding. *IEEE Transactions on communications*, 29(12):1799–1808, 1981.

[96] Michael T Orchard and Gary J Sullivan. Overlapped block motion compensation: An estimation-theoretic approach. *IEEE Transactions on Image Processing*, 3(5):693–699, 1994.

[97] MH Chan, YB Yu, and AG Constantinides. Variable size block matching motion compensation with applications to video coding. *IEE Proceedings I (Communications, Speech and Vision)*, 137(4):205–212, 1990.

[98] Robert Thoma and Matthias Bierling. Motion compensating interpolation considering covered and uncovered background. *Signal Processing: Image Communication*, 1(2):191–212, 1989.

[99] AN Netravali and JD Robbins. Motion-compensated coding: Some new results. *Bell Labs Technical Journal*, 59(9):1735–1745, 1980.

[100] Matthias Bierling and Robert Thoma. Motion compensating field interpolation using a hierarchically structured displacement estimator. *Signal Processing*, 11(4):387–404, 1986.

[101] Kiyoharu Aizawa and Thomas S Huang. Model-based image coding advanced video coding techniques for very low bit-rate applications. *Proceedings of the IEEE*, 83(2):259–271, 1995.

[102] Jeffrey S McVeigh and Siu Wai Wu. Partial closed-loop versus open-loop motion estimation for hdtv compression. *International Journal of Imaging Systems and Technology*, 5(4):268–275, 1994.

[103] Peisong Chen and John W Woods. Bidirectional mc-ezbc with lifting implementation. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(10):1183–1194, 2004.

[104] Yongjun Wu, Konstantin Hanke, Thomas Rusert, and John W Woods. Enhanced mc-ezbc scalable video coder. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(10):1432–1436, 2008.

[105] Yu Liu and King Ngi Ngan. Fast multiresolution motion estimation algorithms for wavelet-based scalable video coding. *Signal Processing: Image Communication*, 22(5):448–465, 2007.

[106] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. *Signal Processing Magazine, IEEE*, 20(3):21–36, 2003.

[107] Jing Tian and Kai-Kuang Ma. A survey on super-resolution imaging. *Signal, Image and Video Processing*, 5(3):329–342, 2011.

[108] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010.

[109] SP Kim, Nirmal K Bose, and HM Valenzuela. Recursive reconstruction of high resolution image from noisy undersampled multiframes. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(6):1013–1027, 1990.

[110] NK Bose, HC Kim, and HM Valenzuela. Recursive implementation of total least squares algorithm for image reconstruction from noisy, undersampled multiframes. In *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, volume 5, pages 269–272. IEEE, 1993.

[111] Marianna Pensky. Solution of linear ill-posed problems using flexible dictionaries. *arXiv preprint arXiv:1408.3386*, 2014.

[112] Brian C Tom, Aggelos K Katsaggelos, and Nikolas P Galatsanos. Reconstruction of a high resolution image from registration and restoration of low resolution images. In *Image*

Processing, 1994. Proceedings. ICIP-94., IEEE International Conference, volume 3, pages 553–557. IEEE, 1994.

[113] Michael Elad and Dmitry Datsenko. Example-based regularization deployed to super-resolution reconstruction of a single image. *The Computer Journal*, 52(1):15–30, 2009.

[114] William T Freeman, Thouis R Jones, and Egon C Pasztor. Example-based super-resolution. *Computer Graphics and Applications, IEEE*, 22(2):56–65, 2002.

[115] Assaf Zomet, Alex Rav-Acha, and Shmuel Peleg. Robust super-resolution. In *CVPR*, volume 1, pages I–645. IEEE, 2001.

[116] Michal Irani and Shmuel Peleg. Motion analysis for image enhancement: Resolution, occlusion, and transparency. *Journal of Visual Communication and Image Representation*, 4(4):324–335, 1993.

[117] Henry Stark and Peyma Oskoui. High-resolution image recovery from image-plane arrays, using convex projections. *JOSA A*, 6(11):1715–1726, 1989.

[118] Deepu Rajan and Subhasis Chaudhuri. Generalized interpolation and its application in super-resolution imaging. *Image and Vision Computing*, 19(13):957–969, 2001.

[119] Andrew J Patti, M Ibrahim Sezan, and A Murat Tekalp. Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time. *IEEE transactions on image processing*, 6(8):1064–1076, 1997.

[120] Zhouchen Lin, Junfeng He, Xiaoou Tang, and Chi-Keung Tang. Limits of learning-based superresolution algorithms. *International journal of computer vision*, 80(3):406–420, 2008.

[121] Michael Elad and Arie Feuer. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE transactions on image processing*, 6(12):1646–1658, 1997.

[122] Jian Sun, Nan-Ning Zheng, Hai Tao, and Heung-Yeung Shum. Image hallucination with primal sketch priors. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–729. IEEE, 2003.

[123] Xinbo Gao, Kaibing Zhang, Dacheng Tao, and Xuelong Li. Image super-resolution with sparse neighbor embedding. *IEEE Transactions on Image Processing*, 21(7):3194–3205, 2012.

[124] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *TIP*, 13(4):600–612, 2004.

[125] Hasan Demirel and Gholamreza Anbarjafari. Image resolution enhancement by using discrete and stationary wavelet decomposition. *IEEE TIP*, 20(5):1458–1460, 2011.

[126] Weisheng Dong, D Zhang, Guangming Shi, and Xiaolin Wu. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *Image Processing, IEEE Transactions on*, 20(7):1838–1857, 2011.

[127] A Temizel and T Vlachos. Wavelet domain image resolution enhancement using cycle-spinning. *Electronics Letters*, 41(3):119–121, 2005.

[128] Herminio Chavez-Roman and Volodymyr Ponomaryov. Super resolution image generation using wavelet domain interpolation with edge extraction via a sparse representation. *IEEE Geoscience and Remote Sensing Letters*, 11(10):1777–1781, 2014.

[129] Shubin Zhao, Hua Han, and Silong Peng. Wavelet-domain hmt-based image super-resolution. In *ICIP*, volume 2, pages II–953. IEEE, 2003.

[130] Stéphane Mallat and Guoshen Yu. Super-resolution with sparse mixing estimators. *Image Processing, IEEE Transactions on*, 19(11):2889–2900, 2010.

[131] CV Jiji, Manjunath V Joshi, and Subhasis Chaudhuri. Single-frame image super-resolution using learned wavelet coefficients. *International journal of Imaging systems and Technology*, 14(3):105–112, 2004.

[132] Prakash P Gajjar and Manjunath V Joshi. New learning based super-resolution: use of dwt and igmrf prior. *Image Processing, IEEE Transactions on*, 19(5):1201–1213, 2010.

[133] Fernanda Brandi, Ricardo de Queiroz, and Debargha Mukherjee. Super resolution of video using key frames. In *Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on*, pages 1608–1611. IEEE, 2008.

[134] C Andrew Segall, Aggelos K Katsaggelos, Rafael Molina, and Javier Mateos. Bayesian resolution enhancement of compressed video. *IEEE Transactions on image processing*, 13(7):898–911, 2004.

[135] Stefanos P Belekos, Nikolas P Galatsanos, and Aggelos K Katsaggelos. Maximum a posteriori video super-resolution using a new multichannel image prior. *Image Processing, IEEE Transactions on*, 19(6):1451–1464, 2010.

[136] Israa Amro, Javier Mateos, Miguel Vega, Rafael Molina, and Aggelos K Katsaggelos. A survey of classical methods and new trends in pansharpening of multispectral images. *EURASIP Journal on Advances in Signal Processing*, 2011(1):79, 2011.

[137] Te-Ming Tu, Ping Sheng Huang, Chung-Ling Hung, and Chien-Ping Chang. A fast intensity-hue-saturation fusion technique with spectral adjustment for ikonos imagery. *IEEE Geoscience and Remote sensing letters*, 1(4):309–312, 2004.

[138] Pats Chavez, Stuart C Sides, Jeffrey A Anderson, et al. Comparison of three different methods to merge multiresolution and multispectral data- landsat tm and spot panchromatic. *Photogrammetric Engineering and remote sensing*, 57(3):295–303, 1991.

[139] Craig A Laben and Bernard V Brower. Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening, January 4 2000. US Patent 6,011,875.

[140] Alan R Gillespie, Anne B Kahle, and Richard E Walker. Color enhancement of highly correlated images. ii. channel ratio and chromaticity transformation techniques. *Remote Sensing of Environment*, 22(3):343–365, 1987.

[141] Robert A Schowengerdt. *Remote sensing: models and methods for image processing*. Academic press, 2006.

[142] Rafael Molina, Miguel Vega, Javier Mateos, and Aggelos K Katsaggelos. Variational posterior distribution approximation in bayesian super resolution reconstruction of multispectral images. *Applied and Computational Harmonic Analysis*, 24(2):251–267, 2008.

[143] John C Price. Combining multispectral data of differing spatial resolution. *IEEE Transactions on Geoscience and Remote sensing*, 37(3):1199–1203, 1999.

[144] Andrea Garzelli and Filippo Nencini. Interband structure modeling for pan-sharpening of very high-resolution multispectral images. *Information Fusion*, 6(3):213–224, 2005.

[145] Yonghyun Kim, Changno Lee, Dongyeob Han, Yongil Kim, and Younsoo Kim. Improved additive-wavelet image fusion. *IEEE Geoscience and Remote Sensing Letters*, 8(2):263–267, 2011.

[146] Luciano Alparone, Lucien Wald, Jocelyn Chanussot, Claire Thomas, Paolo Gamba, and Lori Mann Bruce. Comparison of pansharpening algorithms: Outcome of the 2006 grs-s data-fusion contest. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10):3012–3021, 2007.

[147] Francesca Bovolo, Lorenzo Bruzzone, Luca Capobianco, Andrea Garzelli, Silvia Marchesi, and Filippo Nencini. Analysis of the effects of pansharpening in change detection on vhr images. *IEEE Geoscience and Remote Sensing Letters*, 7(1):53–57, 2010.

[148] Xavier Otazu, María González-Audícana, Octavi Fors, and Jorge Núñez. Introduction of sensor spectral response into image fusion methods. application to wavelet-based methods. *IEEE Transactions on Geoscience and Remote Sensing*, 43(10):2376–2385, 2005.

[149] C Quentin Davis and Dennis M Freeman. Statistics of subpixel registration algorithms based on spatiotemporal gradients or block matching. *Optical Engineering*, 37(4):1290–1298, 1998.

[150] Irene G Karybali, Emmanouil Z Psarakis, Kostas Berberidis, and Georgios D Evangelidis. An efficient spatial domain technique for subpixel image registration. *Signal Processing: Image Communication*, 23(9):711–724, 2008.

[151] Hernani Goncalves, Luís Corte-Real, and Jose A Goncalves. Automatic image registration through image segmentation and sift. *IEEE Transactions on Geoscience and Remote Sensing*, 49(7):2589–2600, 2011.

[152] Georgios Tzimiropoulos, Vasileios Argyriou, and Tania Stathaki. Subpixel registration with gradient correlation. *IEEE Transactions on Image Processing*, 20(6):1761–1767, 2011.

[153] Simon Baker, Daniel Scharstein, JP Lewis, Stefan Roth, Michael J Black, and Richard Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31, 2011.

[154] Jiayi Ma, Huabing Zhou, Ji Zhao, Yuan Gao, Junjun Jiang, and Jinwen Tian. Robust feature matching for remote sensing image registration via locally linear transforming. *IEEE Transactions on Geoscience and Remote Sensing*, 53(12):6469–6481, 2015.

[155] William Scott Hoge. A subspace identification extension to the phase correlation method [mri application]. *Medical Imaging, IEEE Transactions on*, 22(2):277–280, 2003.

[156] Kenji Takita, Takafumi Aoki, Yoshifumi Sasaki, Tatsuo Higuchi, and Koji Kobayashi. High-accuracy subpixel image registration based on phase-only correlation. *IEICE transactions on fundamentals of electronics, communications and computer sciences*, 86(8):1925–1934, 2003.

[157] Georgios Tzimiropoulos, Vasileios Argyriou, Stefanos Zafeiriou, and Tania Stathaki. Robust fft-based scale-invariant image registration with image gradients. *IEEE transactions on pattern analysis and machine intelligence*, 32(10):1899–1906, 2010.

[158] Richard Szeliski and James Coughlan. Hierarchical spline-based image registration. In *International Journal of Computer Vision*. Citeseer, 1994.

[159] Kee-Baek Kim, Jong-Su Kim, Sangkeun Lee, and Jong-Soo Choi. Fast image registration using pyramid edge images. *International Journal of Intelligent Engineering and Systems*, 2:1–8, 2009.

[160] Feng-qing Qin, Xiao-hai He, Wei-long Chen, Xiao-min Yang, and Wei Wu. Video superresolution reconstruction based on subpixel registration and iterative back projection. *Journal of Electronic Imaging*, 18(1):013007–013007, 2009.

[161] Fei Zhou, Wenming Yang, and Qingmin Liao. A coarse-to-fine subpixel registration method to recover local perspective deformation in the application of image super-resolution. *IEEE transactions on image processing*, 21(1):53–66, 2012.

[162] Maoguo Gong, Shengmeng Zhao, Licheng Jiao, Dayong Tian, and Shuang Wang. A novel coarse-to-fine scheme for automatic image registration based on sift and mutual information. *IEEE Transactions on Geoscience and Remote Sensing*, 52(7):4328–4338, 2014.

160

[163] Jue Wu and Albert CS Chung. Multimodal brain image registration based on wavelet transform using sad and mi. In *Medical Imaging and Augmented Reality*, pages 270–277. Springer, 2004.

[164] Md Mushfiqul Alam, Tamanna Howlader, and S M Mahbubur Rahman. Entropy-based image registration method using the curvelet transform. *Signal, Image and Video Processing*, pages 1–15, 2014.

[165] Andrew Secker and David Taubman. Motion-compensated highly scalable video compression using an adaptive 3d wavelet transform based on lifting. In *ICIP*, volume 2, pages 1029–1032. IEEE, 2001.

[166] Béatrice Pesquet-Popescu and Vincent Bottreau. Three-dimensional lifting schemes for motion compensated video compression. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, volume 3, pages 1793–1796. IEEE, 2001.

[167] Marko Esche, Michael Tok, and Thomas Sikora. Adpative dense vector field interpolation for temporal filtering. In *ICIP*, pages 1918–1922. IEEE, 2013.

[168] Ying Chen and Guizhong Liu. Adaptive lagrange multiplier selection model in rate distortion optimization for 3d wavelet-based scalable video coding. In *ICIP*, pages 3190–3194. IEEE, 2014.

[169] Dominic Rüfenacht, Reji Mathew, and David Taubman. Hierarchical anchoring of motion fields for fully scalable video coding. In *ICIP*, pages 3180–3184. IEEE, 2014.

[170] Guy Gilboa, Nir Sochen, and Yehoshua Y Zeevi. Forward-and-backward diffusion processes for adaptive image enhancement and denoising. *IEEE transactions on image processing*, 11(7):689–703, 2002.

[171] Yair Rivenson, Adrian Stern, and Bahram Javidi. Single exposure super-resolution compressive imaging by double phase encoding. *Optics express*, 18(14):15094–15103, 2010.

[172] Wei Zhang and Wai-Kuen Cham. Hallucinating face in the dct domain. *IEEE Transactions on Image Processing*, 20(10):2769–2779, 2011.

[173] Athanasios Papoulis. Generalized sampling expansion. *IEEE transactions on circuits and systems*, 24(11):652–654, 1977.

[174] Hanoch Ur and Daniel Gross. Improved resolution from subpixel shifted pictures. *CVGIP: Graphical Models and Image Processing*, 54(2):181–186, 1992.

[175] H Shekarforoush and R Chellappa. Data-driven multichannel superresolution with application to video sequences. *JOSA A*, 16(3):481–492, 1999.

[176] A Lorette, H Shekarforoush, and J Zerubia. Super-resolution with adaptive regularization. In *Image Processing, 1997. Proceedings., International Conference on*, volume 1, pages 169–172. IEEE, 1997.

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

[178] S Derin Babacan, Rafael Molina, and Aggelos K Katsaggelos. Variational bayesian super resolution. *Image Processing, IEEE Transactions on*, 20(4):984–999, 2011.

[179] Antonio Marquina and Stanley J Osher. Image super-resolution by tv-regularization and bregman iteration. *Journal of Scientific Computing*, 37(3):367–382, 2008.

[180] Hua Huang, Huiting He, Xin Fan, and Junping Zhang. Super-resolution of human face image using canonical correlation analysis. *Pattern Recognition*, 43(7):2532–2543, 2010.

[181] Daniel Glasner, Shai Bagon, and Michal Irani. Super-resolution from a single image. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 349–356. IEEE, 2009.

[182] Armin Kappeler, Seunghwan Yoo, Qiqin Dai, and Aggelos K Katsaggelos. Video super-resolution with convolutional neural networks. *IEEE Transactions on Computational Imaging*, 2(2):109–122, 2016.

[183] Kaibing Zhang, Xinbo Gao, Xuelong Li, and Dacheng Tao. Partially supervised neighbor embedding for example-based image super-resolution. *Selected Topics in Signal Processing, IEEE Journal of*, 5(2):230–239, 2011.

[184] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, pages 694–711. Springer, 2016.

[185] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5197–5206, 2015.

[186] Chih-Yuan Yang and Ming-Hsuan Yang. Fast direct super-resolution by simple functions. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 561–568, 2013.

[187] Radu Timofte, Vincent De Smet, and Luc Van Gool. Anchored neighborhood regression for fast example-based super-resolution. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1920–1927, 2013.

[188] Radu Timofte, Vincent De Smet, and Luc Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Asian Conference on Computer Vision*, pages 111–126. Springer, 2014.

[189] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1646–1654, 2016.

[190] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2016.

[191] Ryan Dahl, Mohammad Norouzi, and Jonathon Shlens. Pixel recursive super resolution. *arXiv preprint arXiv:1702.00783*, 2017.

[192] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint arXiv:1609.04802*, 2016.

[193] Raymond H Chan, Tony F Chan, Lixin Shen, and Zuowei Shen. Wavelet algorithms for high-resolution image reconstruction. *SIAM Journal on Scientific Computing*, 24(4):1408–1432, 2003.

[194] Rakesh C Patel and Manjunath V Joshi. Super-resolution of hyperspectral images: Use of optimum wavelet filter coefficients and sparsity regularization. *IEEE Transactions on Geoscience and Remote Sensing*, 53(4):1728–1736, 2015.

[195] Srimanta Mandal and Anil Kumar Sao. Edge preserving single image super resolution in sparse environment. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 967–971. IEEE, 2013.

[196] Dacil Barreto, LD Alvarez, Rafael Molina, Aggelos K Katsaggelos, and GM Callico. Region-based super-resolution for compression. *Multidimensional Systems and Signal Processing*, 18(2):59–81, 2007.

[197] Satoshi Kondo and Tadamasa Toma. Video coding with super-resolution post-processing. In *Image Processing, 2006 IEEE International Conference on*, pages 3141–3144. IEEE, 2006.

[198] Byung Cheol Song, Shin-Cheol Jeong, and Yanglim Choi. Video super-resolution algorithm using bi-directional overlapped block motion compensation and on-the-fly dictionary training. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(3):274–285, 2011.

[199] Armin Kappeler, Seunghwan Yoo, Qiqin Dai, and Aggelos K Katsaggelos. Super-resolution of compressed videos using convolutional neural networks. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 1150–1154. IEEE, 2016.

[200] Yasutaka Matsuo, Toshie Misu, Shinichi Sakaida, and Yoshiaki Shishikui. Video coding with wavelet image size reduction and wavelet super resolution reconstruction. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 1157–1160. IEEE, 2011.

[201] W Joseph Carper, Thomas M Lillesand, and Ralph W Kiefer. The use of intensity-hue-saturation transformations for merging spot panchromatic and multispectral image data. *Photogrammetric Engineering and remote sensing*, 56(4):459–467, 1990.

[202] PS Kwarteng and AY Chavez. Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis. *Photogramm. Eng. Remote Sens*, 55:339–348, 1989.

[203] Vijay P Shah, Nicolas H Younan, and Roger L King. An efficient pan-sharpening method via a combined adaptive pca approach and contourlets. *IEEE transactions on geoscience and remote sensing*, 46(5):1323–1335, 2008.

[204] JG Liu. Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details. *International Journal of Remote Sensing*, 21(18):3461–3472, 2000.

[205] Pat S Chavez Jr. Digital merging of landsat tm and digitized nhap data for 1: 24 000-scale image mapping. *Photogrammetric Engineering and Remote Sensing*, 52(10):1637–1646, 1986.

[206] Dominique Fasbender, Julien Radoux, and Patrick Bogaert. Bayesian data fusion for adaptable image pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 46(6):1847–1857, 2008.

[207] J Zhou, DL Civco, and JA Silander. A wavelet transform method to merge landsat tm and spot panchromatic data. *International Journal of Remote Sensing*, 19(4):743–757, 1998.

[208] Vildan Atalay Aydin and Hassan Foroosh. In-band sub-pixel registration of wavelet-encoded images from sparse coefficients. *Signal, Image and Video Processing*, pages 1–9, 2017.

[209] David L Donoho and Jain M Johnstone. Ideal spatial adaptation by wavelet shrinkage. *biometrika*, 81(3):425–455, 1994.

[210] Vildan A Aydin and Hassan Foroosh. Motion compensation using critically sampled dwt subbands for low-bitrate video coding. In *Proc. IEEE International Conference on Image Processing (ICIP)*, 2017.

[211] Vildan Atalay Aydin and Hassan Foroosh. A linear well-posed solution to recover high-frequency information for super resolution image reconstruction. *Multidimensional Systems and Signal Processing*, pages 1–22, 2017.

[212] Per Christian Hansen. The truncated svd as a method for regularization. *BIT Numerical Mathematics*, 27(4):534–553, 1987.

[213] Raf Vandebril, Marc Van Barel, and Nicola Mastronardi. Matrix computations and semiseparable matrices: Linear systems (volume 1), 2007.

[214] King-wah Eric Chu. The solution of the matrix equations axb- cxd= e and (ya- dz, yc-bz)=(e, f). *Linear Algebra and its Applications*, 93:93–105, 1987.

[215] Zhaojun Bai, James Demmel, Jack Dongarra, Axel Ruhe, and Henk van der Vorst. *Templates for the solution of algebraic eigenvalue problems: a practical guide*, volume 11. Siam, 2000.

[216] Shengyang Dai, Mei Han, Ying Wu, and Yihong Gong. Bilateral back-projection for single image super resolution. In *Multimedia and Expo, 2007 IEEE International Conference on*, pages 1039–1042. IEEE, 2007.

[217] Hongyan Zhang, Liangpei Zhang, and Huanfeng Shen. A super-resolution reconstruction algorithm for hyperspectral images. *Signal Processing*, 92(9):2082–2096, 2012.