# ZOOM-BASED SUPER-RESOLUTION IMAGE RECONSTRUCTION FROM IMAGES WITH DIFFERENT ORIENTATIONS

A Thesis by

Chandana K.K. Jayasooriya

Diplom in Electrical & Communications Engineering

Technical University of Berlin, Germany, 2004

Submitted to the Department of Electrical and Computer Engineering
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Master of Science

December 2006

ZOOM-BASED SUPER-RESOLUTION IMAGE RECONSTRUCTION
FROM IMAGES WITH DIFFERENT ORIENTATIONS

I have examined the final copy of this thesis for form and content, and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Electrical Engineering.

_____

Hyuck Kwon, Committee Chair

We have read this thesis and
recommend its acceptance:

_____

Steven Skinner, Committee Member

_____

M.B. Yildirim, Committee Member

# DEDICATION

To my beloved parents and wife

ACKNOWLEDGEMENTS

I would like to thank Dr. S. Jayaweera who guided me throughout this research work and gave me valuable advice and suggestions. I would like to thank Dr. H.M. Kwon, under whom I finished this work and who also helped me by commenting on my thesis writing.

My gratitude goes to committee members Dr. S. Skinner and Dr. M.B. Yildirim.

I would like to acknowledge the funding support from Kansas National Science Foundation (NSF) EPSCOR program under the First Award grant KUCR #NSF32241 and thank all my friends at Wichita State University who helped make my research work a success.

ABSTRACT

Construction of a mosaic image using a set of low resolution images taken at different zoom settings and at different angles was investigated in this thesis. The proposed reconstruction algorithm uses the zoom based super resolution technique, based on maximum likelihood estimate. A computationally less intensive point matching algorithm was introduced based on a known algorithm. The simulation results show that the implemented algorithm can find point correspondences of images successfully, even they are differently zoomed, that helps for image registration.

TABLE OF CONTENTS

LIST OF FIGURES

# Chapter 1

# Introduction

Researchers use different cues, like motion, blur, etc., in super-resolution (SR) image reconstruction techniques. A novel SR image reconstruction method was proposed by Joshi and Chaudhuri [1] based on low-resolution (LR) image observations zoomed at different levels. The condition for exact recovery of an analog signal from its discrete sample values is well known by Shannon's sampling theorem. The effect, called aliasing, arises (e.g., in case of audio signals) if the analog signal to be digitized contains frequencies higher than half the sampling frequency. Similarly, in the case of images, the finite resolution (number of pixels per unit length) of the image-capturing chip causes aliasing in captured images. If images of a scene are taken with different zoom settings, the amount of aliasing varies in each image. The image with the least zoom setting that contains the whole scene will be the most affected by aliasing because the entire scene must be represented by a limited number of pixels. The highest zoomed image, which contains only a small area of the scene, will be the least affected by aliasing because, even though the number of pixels available to represent the highest zoomed image remains the same as for the lowest zoomed image, the area that must be represented is much smaller. Therefore, the highest zoomed image will have the highest spatial resolution, whereas the lowest zoomed image will have the lowest spatial resolution. The objective of this proposed SR technique is to enhance the resolution of the image containing the entire scene to the resolution of the observed most-zoomed image. The SR image is modelled as a Markov random field (MRF), and a cost function is derived by

using a maximum *a posteriori* (MAP) estimation method. This cost function is optimized to recover the high-resolution field. In the region where multiple observations are available, this method uses a noise smoothing, and the same neighborhood property is utilized to super resolve the remaining regions [1].

In this thesis, the above-mentioned novel technique was used to obtain a mosaic image from observations taken at different zoom settings. Instead of modelling the mosaic image as a MRF, which leads the lexicographically ordered high-resolution image pixels to satisfy the Gibb's density function, in this thesis, it was assumed that no prior knowledge of the mosaic image was available. Thus a cost function was derived by using the maximum likelihood (ML) estimation method, whereby the high-resolution image was estimated. A point matching algorithm proposed by Kanazawa and Kanatani [2] along with the above-described super-resolution technique was implemented in this thesis, possibly to apply in a scenario as described below. The point matching algorithm was used to find corresponding points in observations in order to align them. The algorithm proposed by Kanazawa and Kanatani [2] uses "*confidence*" values for potential matches of correspondence points and updates them progressively by "*mean-field approximation*." Finally, by using the Random Sample Consensus (RANSAC) algorithm [3], the epipolar constraint [4] was imposed on the potential matches to find final corresponding points between images. In this thesis, this algorithm was slightly modified and applied to find point correspondences between images. Without applying the epipolar constraint, the homography matrix [4] was estimated using the RANSAC algorithm after assigning confidence values to potential point matches. The homography matrix contains information on how two images containing a common area are geometrically related. In other words, with the homography matrix, one image can be aligned or transformed to the other image, given that the the scene in the image is planar.

Wireless sensor networks are an emerging technology that involves many day-to-day applications. Indoor-outdoor environmental monitoring, security and tracking, health and wellness monitoring, and seismic and structural monitoring are a few examples. A

wireless sensor network can consist of a few sensor nodes to hundreds of them distributed randomly in an application area that gathers information, such as temperature, humidity, acceleration, light, and acoustics, depending on the application. Each sensor node may communicate with another sensor node and form a mesh network. The data captured by each sensor node must be routed to an application node (AN), which is assigned to that specific cluster of sensor nodes. An application node has more data processing capability than a sensor node. It receives raw data captured by sensor nodes and processes them to observe locally. Then this data is forwarded to a base station (BS) which is usually located a distance from the sensor nodes. A base station gathers and processes data from each application node in order to obtain a global view of the entire sensor network. A base station possesses vast storage-capacity and data-processing capabilities. Sensor networks are autonomous ad hoc networks. Once the sensor nodes are scattered in the application area, there is minimum human intervention. Unlike the base station, sensor nodes are battery powered, and usually the battery cannot be recharged or replaced economically depending on its deployment. Therefore, the concept of energy savings in sensor networks has become a fast-growing research area.

Figure 1.1 [5] shows an average current through different devices of a Mica2 sensor node by *Crossbow*® at different operational states. It shows that the current related to data transmission and reception is relatively higher than the current consumed by other operations at a sensor node. Transmission of data, even with the lowest power level and reception of data, needs 8.8 mA and 9.6 mA, respectively, which is higher than any current flow at any operational mode of the sensor node. This fact implies that the more data a sensor node has to transmit, the greater the power consumption.

Mainly due to the available limited power supply, sensor nodes undergo strict hardware limitations. Accordingly, a miniature digital camera mounted on a sensor node may have only 64x64 pixels (four kilo pixels) resolution compared to a camera on a mobile phone, which has more than one mega pixels), which is quite low. The lower the resolution of the

camera the lower the amount of data to be transmitted, which means less power consumption at the sensor node. Even though available source coding techniques can considerably reduce the amount of data to be transmitted from a source, these techniques cannot be implemented in a sensor node due available limited data processing capabilities of a sensor node.

| **Device** | **Current** | **Device** | **Current** |
|---|---|---|---|
| **CPU** | | **Radio** (900MHz) | |
| Active | 7.6 mA | Core | 60 µA |
| Idle | 3.3 mA | Bias | 1.38 mA |
| ADC Noise | 1.0 mA | Rx | 9.6 mA |
| Power down | 116 µA | Tx (-18 dBm) | 8.8 mA |
| Power Save | 124 µA | Tx (-13 dBm) | 9.8 mA |
| Standby | 237 µA | Tx (-10 dBm) | 10.8 mA |
| Ext Standby | 243 µA | Tx (-6 dBm) | 11.3 mA |
| | | Tx (-2 dBm) | 15.6 mA |
| **LED** (each) | 2.2 mA | Tx (0 dBm) | 17.0 mA |
| | | Tx (+3 dBm) | 20.2 mA |
| **Sensor Board** | 0.7 mA | Tx (+4 dBm) | 22.5 mA |
| | | Tx (+5 dBm) | 26.9 mA |

Figure 1.1: Typical currents through different devices of a Mica2 sensor node. Currents related to transmission and reception of data is much higher than other currents in sensor node.

The objective of this thesis was to reconstruct an image of a desired scene, combining several low-resolution images captured by several sensor nodes. Generally, sensor node positions can be scattered randomly in a given area. Thus, the orientation of cameras may also vary accordingly. Images taken by sensor nodes may contain overlapping regions but at different angles and with different zoom factors. The least-zoomed LR image that contains the largest area of a scene is selected as the base image. LR images with higher zoom factors that are taken at different angles than the base image but also containing an area common to the base image were selected as input images. These input images were warped to align with the base image and combined with it. The common area to both the base image and the warped-input images achieved a higher resolution than the corresponding areas of the least-zoomed image. The resulting image contained a larger view of the scene, compared to any of the individual LR images.

## 1.1 Literature Survey

In the modern world, digital imaging is involved in a variety of applications, such as medical diagnosis, surveillance, and digital photography, to name a few. Images with high-resolution (HR) is a necessity in many of these applications. For example, a high-resolution image will be very helpful for a doctor making a correct diagnosis or a HR arial photograph may be of high interest in a surveillance application. There are two direct methods to obtain high-resolution images - namely, by changing the dimensions of the image-capturing chip or a single pixel, that is, to increase the size of the image-capturing chip or to reduce the size of a pixel (i.e., increase the number of pixels per unit area) [6]. The first approach suffers from the increased delay in data transfer from the chip because of increased capacitance (an increase in chip size increases capacitance), and the second approach suffers from the decreased amount of available light because of the smaller dimension. Therefore, signal precessing techniques have been proposed and developed to obtain an HR image from a set of LR-image observations. This type of resolution enhancement of images is called super-resolution (SR) (or HR) image reconstruction in the literature. The major advantage of this approach is that HR images can be obtained by utilizing existing LR imaging systems.

There are two related problems to SR image reconstruction techniques. The first is the restoration of an image degraded by noise or blur. The size of the reconstructed image remains the same as the original degraded image in this case. The second problem involves increasing the size of an image using interpolation techniques. Due to downsampling, high-frequency components are lost or diminished in the process of acquiring an LR image. These lost high-frequency components cannot be recovered if a single image is used to construct a larger-sized image using interpolation. Multiple data sets of the same scene are utilized to make improvements in this case [6]. Apart from the loss of spatial resolution due to the limited physical size of the image-acquiring chip; optical distortion due to out of focus, diffraction etc.; motion blur due to limited shutter speed; and additive noise during transmission cause further degradation in the captured image.

Multiple LR images of the same scene must be available in order to obtain an HR image using SR techniques. The basic idea behind SR techniques is to exploit the new information available on these multiple LR images that have subpixel shifts among each other to construct an HR image. Multiple images of the same scene can be obtained by capturing images of a single scene using a single camera or by taking images of the same scene using multiple cameras located at different positions. Even a video sequence can be used to construct an HR image. For subpixel shifts to exist among LR images, some relative scene motion must have occurred while capturing the scene. When these relative scene motions are known or can be estimated, an HR image can be constructed, as shown in Figure 1.2 [6].

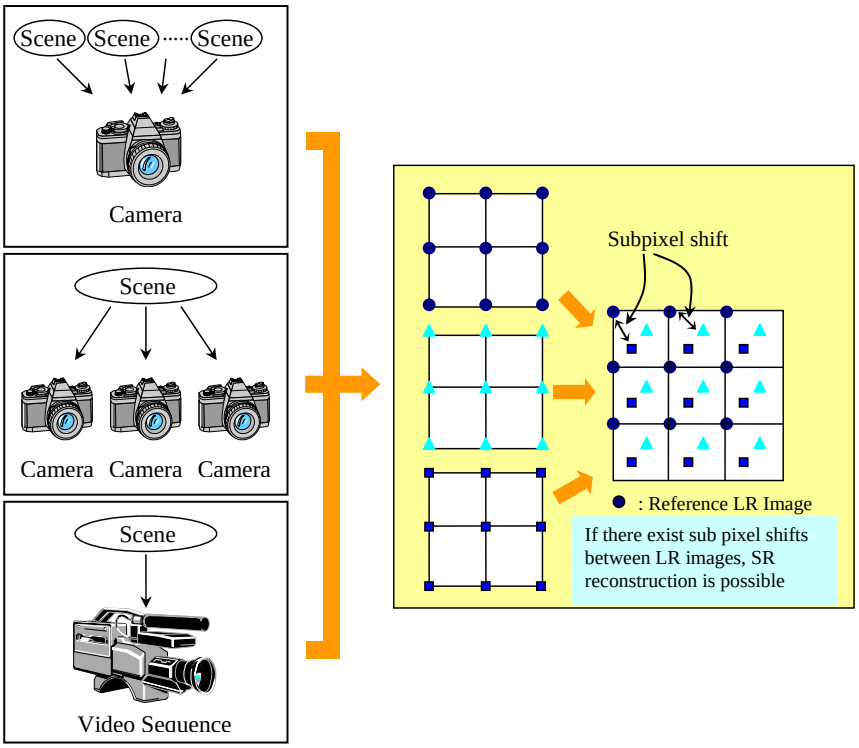


Figure 1.2: Basic premise for super-resolution. Multiple LR images can be captured by a single camera, multiple cameras or by a video sequence. A HR image can be constructed if subpixel shifts between LR images are available.

Image registration, the process of aligning two or more images of the same scene is the first step in most image-processing applications involving several images. Usually, one

image is called the base image, and it is the reference to which the other images (called input images) are compared. The goal of image registration is to align input images to the base image by applying a spatial transformation to the input images. Image registration requires that images contain overlapping regions so that spatial transformations can be calculated. Image registration can be divided into three steps. The first step is to extract feature points in images. Then, in step two, point correspondences in images that undergo a geometric transformation must be determined. Two sets of feature points corresponding to two images are always used for this point matching. Using these point correspondences, a transformation is computed in step three, and with this transformation an input image is aligned to the base image.

A widely used feature detector is the Harris operator [7]. But other approaches [8] [9] do not need to identify common features in image pairs for calculating transformation parameters. Generally, determining correspondences in two images that undergo a certain geometric transformation requires extensive computations. Cortelazzo and Lucchese [10] present a computationally efficient way of determining a set of corresponding features in the overlapping region of two images. In this approach, an affine approximation of the projective transformation is obtained, which reduces the search area for corresponding points. Then the projective transformation is determined from corresponding sets of features. The estimation of parameters of a projective transformation is a least-squares minimization problem based on a set of noisy correspondence points. Radke et al. [11] present an algorithm that reduces the computation complexity by reducing the least-squares problem to a two-dimensional nonquadratic minimization problem. Denton and Beveridge [12] presents a point matching algorithm to obtain estimates of projective transformations between image pairs. In this work, point matching is done using a local search method. Even though image registration can be divided into three steps, generally steps two and three are merged into one. Because point matching and estimating transformation parameters cannot be done totally separated, an estimation of transformation parameters is used to determine point matching

more robustly. Random Sample Consensus is a method, used for fitting a model (e.g., transformation parameters) to experimental data that contain a significant percentage of gross errors. Hence, most of the algorithms developed for image registration using point matching use RANSAC to estimate transformation parameters in a robust way.

## 1.2 Contributions

In this thesis, an algorithm was implemented to reconstruct high-resolution images by low-resolution observations taken with different zoom settings and different angles. Kumar et al. [9], [10], [12], and [13] present image-mosaicking algorithms based on images taken with the same zoom levels. An image-registration method was proposed to mosaic images taken with different zoom factors, such as two. The zoom factors were assumed to be known. Since the focal length of the camera, which is directly related to the zoom factor, could be transmitted to the base station, the above assumption was realistic. Then the maximum-likelihood estimate for the zoom-based super-resolution reconstruction is formulated and derived. Using small image segments, the ML estimate was calculated for different types of low-resolution images, which gave an idea of the criteria needed for gaining high-resolution images based on multiple low-resolution observations. The main contributions are the derivation of equation (3.2.8) in Section 3.2 for the low-resolution image formation model given in equation (3.2.1) and applying the RANSAC algorithm to calculate the homography matrix using putative correspondence points that satisfy equation (2.3.12) in Chapter 2.3. A fully automated image-mosaicking algorithm was implemented and tested for images taken at different angels and at different zoom levels.

## 1.3 Thesis Outline

The remainder of this thesis is organized as follows: Chapter 2.1 discusses the acquisition of low-resolution images from a high-resolution reference image, which is used for zoom-based super-resolution image reconstruction. An introduction to epipolar geometry

and the geometric transformation by which pixels are related in two images taken in different angles of the same planar scene is described in Chapter 2.2. Chapter 2.3 introduces and proposes the implemented point matching algorithm. In Chapter 3, Section 3.1 discusses the maximum *a posteriori* probability (MAP)-based super-resolution imaging algorithm under the assumption of Gaussian noise. In section 3.2, the maximum-likelihood estimate for the zoom-based super-resolution imaging, based on the model introduced in Chapter 2.1, is derived. Simulation results are presented in Chapter 4.

# Chapter 2

# Zoom-Based Super-Resolution Image Reconstruction Algorithm

A novel technique was proposed by Joshi and Chaudhuri [1] for super-resolution reconstruction of a scene by using low-resolution (LR) observations at different zoom settings. The objective was to reconstruct an image of the entire scene at a resolution corresponding to the available most-zoomed LR observation. The LR image acquisition model and cost function, which was optimized to reconstruct the high-resolution image, are described in Sections 2.1 and 3.1, respectively. The cost function was derived using the maximum *a posteriori* estimation method, and the probability distribution of intensity values of the scene to be recovered (HR image) was assumed to be known (see Section 3.2). In this thesis, the above-mentioned probability distribution was assumed to be unknown. Accordingly, the resulting cost function to be optimized is given in equation (3.1.6).

## 2.1 Low-Resolution Image Acquisition Model

Assume that $\{\mathbf{Y}_m\}_{m=1}^{p}$ is a set of $p$ images, each of size $M_1 \times M_2$, of a desired scene that has been captured with different zoom settings. Suppose that these are ordered in an increasing order of zoom level so that $Y_1$ is the least-zoomed image and $\mathbf{Y}_p$ is the most-zoomed image. The most-zoomed image of the scene is assumed to have the highest resolution, whereas the least-zoomed image that captures the entire scene has the lowest spatial resolution. In this thesis, it was assumed that the zoom factors between successive

observations are known. A block diagram view of how the high-resolution and low-resolution observed images are related is given in Figure 2.1.

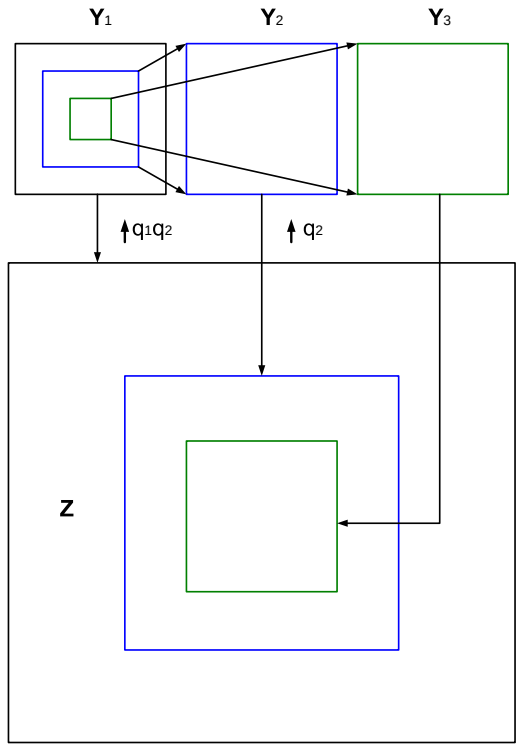


Figure 2.1: Illustration of the geometrical relationship between the LR observations $Y_1$, $Y_2$ and $Y_3$ and the HR image of the scene Z. $Y_1$ corresponds to the least zoomed and $Y_3$ to the most zoomed. $q_1$ and $q_2$ are decimation factors.

The goal was to obtain a super-resolution image of the entire scene, although there are multiple observations corresponding to only part of the scene. In other words, the goal was to up-sample the least-zoomed scene corresponding to the entire scene to the size of $(q_1 \cdot q_2 \cdot q_3 \cdots q_{p-1}) \cdot (M_1 \times M_2) = N_1 \times N_2$, where $q_m$ is the zoom factor between the two observed images $\mathbf{Y}_m$ and $\mathbf{Y}_{m+1}$, for $m = 1, \cdots, p$. For simplicity, in the following discussion, it is assumed that $q_m = q$, for all $m$. Note that here it also was assumed that the resolution at which the most-zoomed observed image is available is the resolution to which the entire scene needs to be super resolved. Then, given the most-zoomed image $\mathbf{Y}_p$, the remaining $p - 1$ observed images were modelled as decimated and noisy versions of the appropriate

regions in the high-resolution image $\mathbf{Z}$.

The observed images can then be modeled as [1]

$$\mathbf{y}_m \quad = \quad \mathbf{D}_m\mathbf{z} + \mathbf{n}_m \quad \text{for} \quad m = 1, 2, \cdots, p, \tag{2.1.1}$$

where $\mathbf{y}_m$ is the $M_1M_2 \times 1$ lexicographically ordered vector that contains the pixel values from the $m$-th low-resolution image $\mathbf{Y}_m$, $\mathbf{z}$ is the $N_1N_2 \times 1 (= q^{p-1}M_1M_2 \times 1)$ vector containing pixel values from the high-resolution image to be reconstructed, and $\mathbf{n}$ is an identically and independently distributed (i.i.d) Gaussian noise vector consisting of zero-mean and variance $\sigma^2$ components. The matrix $\mathbf{D}_m$ is the decimation matrix that depends on the given zoom factor $q$ (which is assumed to be known). Figure 2.2 shows the image acquisition model.

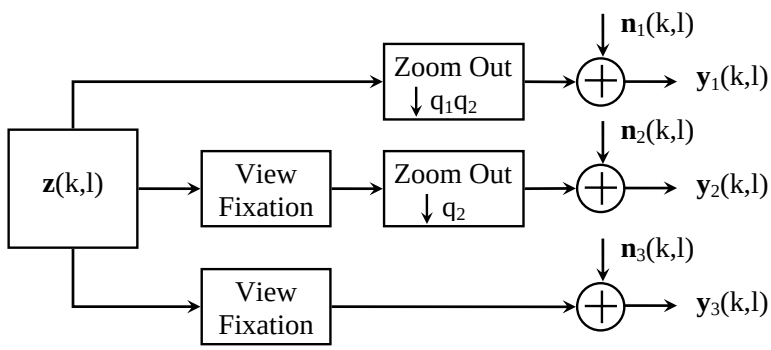


Figure 2.2: Illustration of the low-resolution image formation model for three different zoom levels. View fixation denotes the cropping of the HR image Z according to different zoom levels.

The down-sampling process to acquire an LR image $\mathbf{y}$ from an HR image was achieved by multiplying the lexicographically ordered HR image $\mathbf{z}$ by the decimation matrix $\mathbf{D}$. Figure 2.4 illustrates the relationship of pixel values and grid between an HR image and an LR image. Here a decimation factor (q) of two is considered. The value of an LR image pixel is calculated by averaging four pixels of the HR image. Figure 2.5 shows the decimation matrix, that down samples a 4x4 pixel image to a 2x2 pixel image. Before multiplying by the decimation matrix, the 4x4 image is transformed to a 16x1 column vector transposing by row and stacking them from top to bottom. Applying this decimation matrix

to this lexicographically ordered column vector results in a 4x1 column vector, which has to be reordered by unstacking to a 2x2 matrix. By applying the transpose of this decimation matrix, a 2x2 pixel image can be upsampled to a 4x4 pixel image.

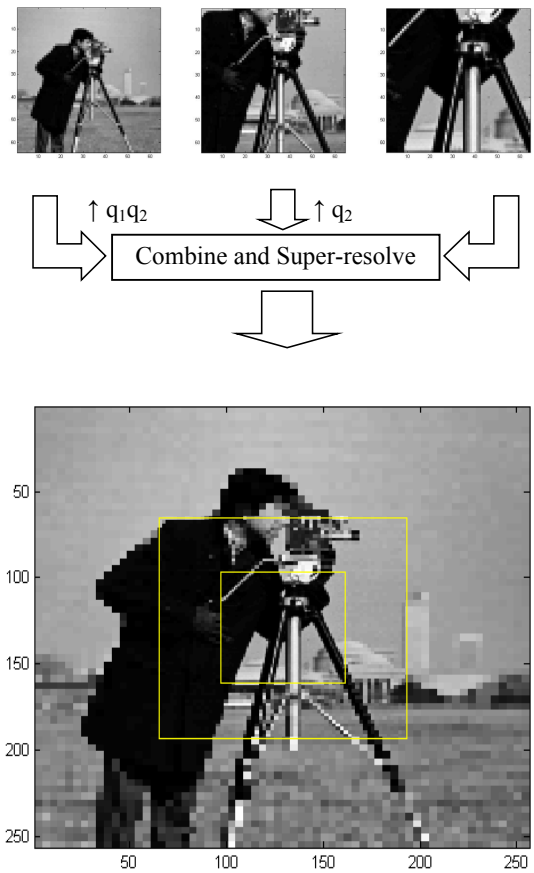


Figure 2.3: Upper row shows three LR observations. These images are combined to construct a SR image after upsampling them according to the zoom level that they were acquired. Lower row shows the constructed super resolved image. The squares represent the area covered by the second and third LR observations after upsampling.

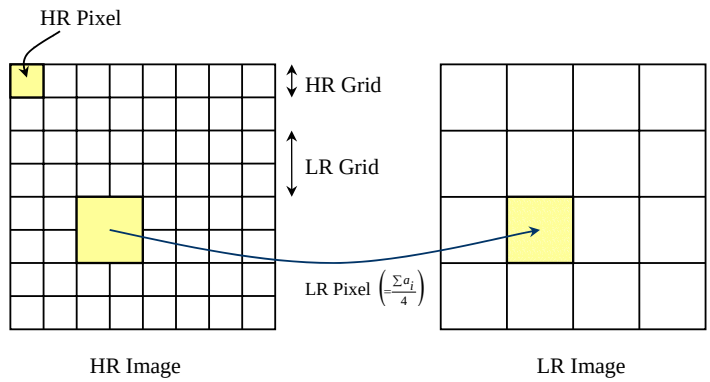


Figure 2.4: Pixel value of the LR image (left) is calculated by averaging four pixels of the HR image (right). In this case the decimation factor is four.

$$D = \frac{1}{q^2}\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix}$$

Figure 2.5: Decimation matrix used to downsample a 4x4 pixel image to a 2x2 pixel image. Here q=2. $D^T$ can be used to upsample a 2x2 pixel image to a 4x4 pixel image. D is applied to a lexicographically ordered column vector of the input image.

## 2.2 Geometry of Two Image Views

The geometrical relationship of pixel positions between two images of a single scene taken at different angles is developed in this section. The scene is assumed to be planar. Two images of the same object taken from different camera positions are considered (see Figure 2.6). Points $\mathbf{o}_1$ and $\mathbf{o}_2$ represent the center of each camera. The planar scene is represented by plane P, and point $\mathbf{p}$ on this plane is imaged by the cameras onto the image planes at $\mathbf{x}_1$ and $\mathbf{x}_2$. The line that connects camera centers $\mathbf{o}_1$ and $\mathbf{o}_2$ intersects the image planes at $\mathbf{e}_1$ and $\mathbf{e}_2$. These points are called epipoles, and the lines connect $\mathbf{e}_1$, $\mathbf{x}_1$ and $\mathbf{e}_2$, $\mathbf{x}_2$ epipolar lines [14]. The objective is to find the transformation matrix $\mathbf{H}$, called the homography matrix, which gives the coordinate relationship between points $\mathbf{x}_1$ and $\mathbf{x}_2$. In this section, uppercase bold letters represent matrices. Coordinate vectors are represented by either lowercase bold letters (coordinates of a point on an image plane with respect to $\mathbf{o}_1$ or $\mathbf{o}_2$) or by upperpcase letters (coordinates of a point on plane P with respect to $\mathbf{o}_1$ or $\mathbf{o}_2$). Also, column vectors are represented by uppercase letters.

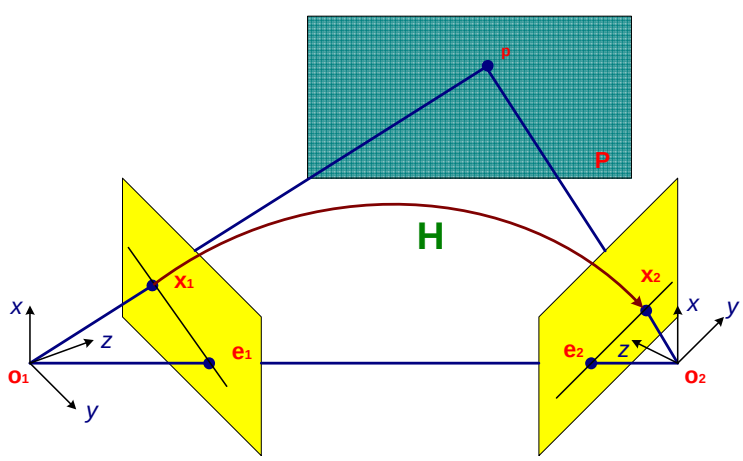


Figure 2.6: Relation between image points of two different views of a single planar scene. $\mathbf{o}_1$ and $\mathbf{o}_2$ are the camera center points. $\mathbf{x}_1$ and $\mathbf{x}_2$ are the image points of $\mathbf{p}$ on plane P. $\mathbf{x}_1$ and $\mathbf{x}_2$ are related by the homography matrix $\mathbf{H}$.

The relative positions between the camera centers $o_1$ and $o_2$ are related by a rotation matrix $\mathbf{R}$ and a translation vector $T$. The coordinate transformation between the

two camera positions can be written as

$$X_2 = \mathbf{R}X_1 + T \tag{2.2.1}$$

where $X_1$ and $X_2$ are the coordinates of the point $\mathbf{p}$ on the plane $P$ relative to camera positions $o_1$ and $o_2$, respectively [14]. Defining $N$ to be the unit normal vector of the plane $P$ with respect to the first camera position yields

$$\frac{1}{d}N^T X_1 = 1, \quad \forall X_1 \in P \tag{2.2.2}$$

where $d > 0$ is the distance between the optical center $o_1$ and the plane $P$. Substituting equation (2.2.2) in equation (2.2.1)

$$X_2 = (\mathbf{R} + \frac{1}{d}TN^T)X_1 \tag{2.2.3}$$

The matrix

$$\mathbf{H} = (\mathbf{R} + \frac{1}{d}TN^T) \tag{2.2.4}$$

is called the (planar) homography matrix. By denoting the coordinates of the image point of point $p$ with respect to coordinates at optical centers $o_1$ and $o_2$ by $\mathbf{x_1}$ and $\mathbf{x_2}$, respectively, then

$$\lambda_1 \mathbf{x_1} = X_1, \quad \lambda_2 \mathbf{x_2} = X_2 \tag{2.2.5}$$

where $\lambda_1$ and $\lambda_2$ are scaling factors. Substituting equations 2.2.4 and 2.2.5 in equation (2.2.3) a relationship between corresponding image points of the point $p$ is obtained as

$$\lambda_2 \mathbf{x_2} = \mathbf{H}\lambda_1 \mathbf{x_1} \quad \Leftrightarrow \quad \mathbf{x_2} \sim \mathbf{H}\mathbf{x_1} \tag{2.2.6}$$

where $\sim$ indicates equality up to a scalar factor. With the above relationship, it is possible to find coordinates of corresponding image points if the planar homography matrix $\mathbf{H}$ is known.

To calculate **H**, the four-point algorithm is used to make use of linear algebraic properties of four or more corresponding image points of two images that are lying on a planar surface of the object in 3D space.

## 2.3 Point Matching Algorithm

The point matching technique used in this thesis was based on the work of [2] which is described here. The goal of a point matching algorithm is, as it's name implies, to find corresponding points in images. To achieve this, first there must be some feature points extracted in each image. This can be done using the Harris corner detector [7]. These extracted image points are used as input to find point correspondences as described by the following algorithm.

The residuals calculated in equation (2.3.1) represent the local correlations between neighbors of point $p$ in one image and point $q$ in the other image as

$$J(p,q) = \sum_{(i,j)\in N} |T_p(i,j) - T_q(i,j)|^2 \tag{2.3.1}$$

where $T_p(i,j)$ and $T_q(i,j)$ are the intensity values of an $w$x$w$ neighborhood centered on $p$ and $q$ [2].

The basic procedure for point matching can be summarized as follows:

- Extract feature points in both images $(N, M)$.
- Compute the residuals $\{J(p_\alpha, q_\beta)\}$, $\alpha = 1, \cdots, N$ $\beta = 1, \cdots, M$.
- Search for $min\{J(p^*_\alpha, q^*_\beta)\}$, and establish a match between points $p^*_\alpha$ and $q^*_\beta$.
- Remove the column and row containing the value $\{J(p^*_\alpha, q^*_\beta)\}$ from the table.
- Repeat the above steps to create the resulting $(N-1)(M-1)$ table.
- Find $L$ number of matches after $L = min(N, M)$ repetitions.

Generally, this method does not give good results. Orientations of the two images are particularly different. In this procedure, a selected pair may not be correct, and a correct pair may be discarded.

Therefore, for all potential matches, confidence values are assigned via the Gibbs distribution, as defined in equation (2.3.2). According to this confidence value assignment, a pair $(p, q)$ that has high residual value will be given a low confidence value as

$$P = e^{-sJ(p,q)} \tag{2.3.2}$$

The attenuation constant $s$ can be determined as follows [2]: Since there can be $L = min(N, M)$ pairs among $N$x$M$ pairs, the average of the $L$ smallest residuals is set equal to the overall weighted average with respect to the confidence value distribution given in equation (2.3.2). This condition is given in equation (2.3.3). The average of $L$ smallest residuals and the weight factor is calculated according to equations (2.3.4) and (2.3.5), respectively. By substituting equation (2.3.3) with equations (2.3.4) and (2.3.5), equation (2.3.6) is obtained, which determines $s$ by substituting values greater than zero for $s$ and making the right-hand side of the equation zero.

$$\frac{1}{Z}\sum_{\lambda=1}^{NM} J_\lambda e^{-sJ_\lambda} = \overline{J} \tag{2.3.3}$$

$$\overline{J} = \frac{1}{L}\sum_{\lambda=1}^{L} J_\lambda \tag{2.3.4}$$

$$Z = \sum_{\lambda=1}^{NM} e^{-sJ_\lambda} \tag{2.3.5}$$

$$\Phi(s) = \sum_{\lambda=1}^{NM} \left(J_\lambda - \overline{J}\right) e^{-sJ_\lambda} \tag{2.3.6}$$

Then each pair of confidence values is assigned ($P_\lambda^{(0)}$ confidence value for the $\lambda$th pair). Pairs which satisfy equation (2.3.7) are selected as tentative candidates for correct

matches and the flow vector $\overrightarrow{r}_\mu$ is calculated. This vector connects the two points of the $\mu$th match, which starts in the first image and ends in the second. Then the confidence weighted mean $\overrightarrow{r}_m$ and the confidence weighted covariance matrix $V$ of the optical flow are given in equations (2.3.8) and (2.3.10). The confidence of spatial consistency assumes that the scene does not have an extraordinary three dimensional shape. The confidence of spatial consistency of the $N$x$M$ potential matches are assigned via the Gaussian distribution given in equation (2.3.11).

$$P_\lambda^{(0)} > e^{-\frac{k^2}{2}} \tag{2.3.7}$$

$$\overrightarrow{r}_m = \frac{1}{Z}\sum_{\mu=1}^{n_0} P_\mu^{(0)} \overrightarrow{r}_\mu \tag{2.3.8}$$

$$Z = \sum_{\mu=1}^{n_0} P_\mu^{(0)} \tag{2.3.9}$$

$$V = \frac{1}{Z}\sum_{\mu=1}^{n_0} P_\mu^{(0)} \left(\overrightarrow{r}_\mu - \overrightarrow{r}_m\right)\left(\overrightarrow{r}_\mu - \overrightarrow{r}_m\right)^T \tag{2.3.10}$$

$$P_\lambda^{(1)} = \exp^{-\left(\left(\overrightarrow{r}_\lambda - \overrightarrow{r}_m\right), V^{-1}\left(\overrightarrow{r}_\lambda - \overrightarrow{r}_m\right)\right)} \tag{2.3.11}$$

Assuming that the scene is planar the geometric transformation of image pixels can be approximated by a homography matrix as shown in Chapter 2.2. Now, tentative candidate pairs are chosen, which satisfy the condition given in equation (2.3.12).

$$P_\lambda^{(0)} P_\lambda^{(1)} > e^{-\frac{2k^2}{2}} \tag{2.3.12}$$

These tentative candidates are used as the input data set for the RANSAC algorithm, which selects four pairs randomly and calculates the homography matrix **H**. Using this homography matrix, inliers that have a shorter distance than the symmetric transfer error [14] given in equation (2.3.13) are selected. At last the RANSAC algorithm finds four

correspondence points with the maximum number of inliers and the corresponding homography matrix. The symmetric transfer error, which is used as the cost function for choosing the inliers, is illustrated in Figure 2.7.

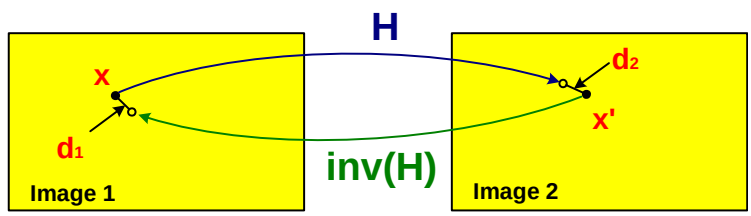


Figure 2.7: Estimated homography matrix, which does not map the correspondence points $\mathbf{x}$ and $\mathbf{x}'$ perfectly. The geometric error in image 2 and image 1 occurred during forward and backward transformation are given by $d_1$ and $d_2$ respectively.

Assume $\mathbf{x}$ and $\mathbf{x}'$ are two correspondence points in image 1 and image 2, respectively. The estimated homography matrix $\mathbf{H}$ neither maps $\mathbf{x}$ to $\mathbf{x}'$ (true corresponding point to $\mathbf{x}$) in forward transformation nor $\mathbf{x}'$ to $\mathbf{x}$ in backward transformation perfectly. The geometric error in image 2 corresponding to the forward transformation ($\mathbf{H}$) and the geometric error in image 1 corresponding to the backward transformation ($\mathbf{H}^{-1}$) are denoted by $d_2$ and $d_1$, respectively. Equation (2.3.13) gives the sum of geometric errors. The first term represents the transfer error in image 1, and the second term represents the transfer error in image 2.

$$\sum_i d_1 \left( \mathbf{x}_i, H^{-1}\mathbf{x}_i^{'} \right)^2 + d_2 \left( \mathbf{x}_i^{'}, H\mathbf{x}_i \right)^2 \tag{2.3.13}$$

# Chapter 3

# Detector Structure for Gaussian Noise Model

## 3.1 Traditional MAP-Based Super-Resolution Algorithm

This section briefly outlines the maximum *a posteriori* probability super-resolution imaging algorithm under the assumption of Gaussian noise. It is generalized to obtain the maximum-likelihood SR scheme for Gaussian noise when the prior knowledge of an image distribution is not available.

Given the LR images with different zoom factors, the maximum *a posteriori* probability estimate of the high-resolution image $Z$ is given by

$$\begin{aligned} \hat{\mathbf{z}} &= \arg\max_{\mathbf{z}} P\left(\mathbf{z}|\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p\right) \\ &= \arg\max_{\mathbf{z}} P\left(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p|\mathbf{z}\right) P\left(\mathbf{z}\right) \end{aligned} \tag{3.1.1}$$

where equatoin (3.1.1) follows from the Bayes rule, and $P(\mathbf{z})$ is the prior distribution of the high-resolution image $\mathbf{z}$. According to Kang and Chaudhuri [15], the scene to be recovered was modeled as a Markov random field [16, 17], leading to a prior given by the so-called Gibbs density function. Note that, since noise vectors $\mathbf{n}_m$s are independent, conditioned on $\mathbf{z}$, the LR images $\mathbf{y}_m$s are independent. Using this and the fact that log(.) is a monotonic function of its argument, the MAP estimator in equation (3.1.1) for the super-resolved image

can be written as

$$\hat{\mathbf{z}} = \arg\max_{\mathbf{z}} \left[ \sum_{m=1}^{p} \log\left(p\left(\mathbf{y}_m|\mathbf{z}\right)\right) + \log\left(p\left(\mathbf{z}\right)\right) \right] \tag{3.1.2}$$

In the case of Gaussian noise and Gibbs prior density, it can be shown that the above MAP estimator can be written as

$$\hat{\mathbf{z}}_{MAP} = \arg\min_{\mathbf{z}} \left[ \sum_{m=1}^{p} \frac{\|\mathbf{y}_m - D_m\mathbf{z}\|^2}{2\sigma^2} + V(\mathbf{z}) \right] \tag{3.1.3}$$

where $V(\mathbf{z})$ is a smoothness-related term that comes from the assumed prior distribution of the HR image [1, 18].

$$V(\mathbf{z}) \quad = \quad \sum_{i,j} \left[\mu e_{zs} + \gamma e_{zp}\right] \tag{3.1.4}$$

where $\mu e_{zs}$ and $\gamma e_{zp}$, respectively, are the smoothness term and the penalty term necessary to prevent occurrence of spurious discontinuities. Note that here $\mu$ represents the penalty term for departure from the smoothness. Each turn-on of a line-process variable is penalized by a quantity $\gamma$ so as to prevent spurious discontinuities [1, 18].

Sometimes it is advantageous to combine prior knowledge into the estimator via a regularization parameter $\gamma$, which assigns relative weights to the prior and posterior as

$$\hat{\mathbf{z}}' = \arg\min_{\mathbf{z}} \left[ \sum_{m=1}^{p} \frac{\|\mathbf{y}_m - D_m\mathbf{z}\|^2}{2\sigma^2} + \gamma V(\mathbf{z}) \right] \tag{3.1.5}$$

Strictly speaking, $\hat{\mathbf{z}}'$ is not the true MAP estimator. When the assumed prior model is a poor approximation, equation (3.1.5) may provide better performance. Taking this approach even further, the so-called maximum likelihood super-resolution image may be obtained by setting $\gamma = 0$ as

$$\hat{\mathbf{z}}_{ML} \quad = \quad \arg\min_{\mathbf{z}} \left[ \sum_{m=1}^{p} \frac{\|\mathbf{y}_m - D_m\mathbf{z}\|^2}{2\sigma^2} \right] \tag{3.1.6}$$

The maximum likelihood method is suitable when one does not have access to a prior distribution for the image to be reconstructed. Note that the use of a prior knowledge in the zoom-based super-resolution reconstruction leads to smoothing of the image.

## 3.2 Derivation of Maximum Likelihood Estimate

Maximum likelihood estimation is used for parameter estimation when any prior information about the parameter to be estimated is not known [19]. The super-resolved image vector $\mathbf{z}$ is the parameter vector to be estimated, and it's probability distribution $p(\mathbf{z})$ is assumed to be unknown in this thesis. In this section, the maximum likelihood estimate (MLE) for the super-resolved image is calculated using the image observation model described in Section 2.1.

As discussed in Section 2.1, the low-resolution image observation can be formulated as

$$\mathbf{y}_m = \mathbf{D}_m\mathbf{z} + \mathbf{n}_m \quad \text{for} \quad m = 1, 2, \cdots, p \tag{3.2.1}$$

where $\mathbf{y}_m$ is the $M^2 \times 1$ lexicographically ordered vector that contains the pixel values from the $m$-th low-resolution image $\mathbf{Y}_m$, $\mathbf{D}_m$ is the decimation matrix, $\mathbf{z}$ is the high-resolution image vector, $\mathbf{n}_m$ is the additive Gaussian noise vector, and $p$ is the number of LR observations.

The MLE of the super-resolved image vector $\mathbf{z}$ is calculated as

$$\hat{\mathbf{z}}_{ML} = \arg\max_{\mathbf{z}}\{log\ p\left(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p|\mathbf{z}\right)\} \tag{3.2.2}$$

where $p\left(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p|\mathbf{z}\right)$ is the joint probability distribution function of the LR observation vectors $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p$, conditioned on $\mathbf{z}$, and it is called the *likelihood function* [19].

The joint probability distribution function (pdf) of $n$ jointly Gaussian random variables $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ is given by Leon-Garcia [20] as

$$p(x_1, x_2, \cdots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T\Sigma^{-1}(\mathbf{x} - \mathbf{m})\right) \tag{3.2.3}$$

where $\mathbf{x}$ and $\mathbf{m}$ are the column vectors containing the values $x_1, x_2, \cdots, x_n$ and the mean values of the $n$-random variables $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$, respectively. The covariance matrix is denoted by $\Sigma$, and the determinant of the matrix is denoted by $|\cdot|$. If the random variables are independent, then the covariance matrix becomes a diagonal matrix and each diagonal

element is the variance of the corresponding random variable. Furthermore, if the random variables are identical with variance $\sigma^2$, then $|\Sigma|$ becomes $\sigma^{2n}$.

As shown in equation (3.2.2), to calculate $\hat{\mathbf{z}}_{ML}$, the conditional pdf $p\left(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p|\mathbf{z}\right)$ is required. Assuming the additive noise vectors in equation (3.2.1) are independent and identically distributed Gaussian random variables, according to equation 3.2.3, the pdf of the $m$-th observation conditioned on $\mathbf{z}$ can be written as

$$\begin{aligned} p(\mathbf{y}_m|\mathbf{z}) &= \frac{1}{(2\pi\sigma^2)^{\frac{M}{2}}} \exp\left(-\frac{1}{2}\left(\mathbf{y}_m - \mathbf{D}_m\mathbf{z}\right)^T \Sigma^{-1}\left(\mathbf{y}_m - \mathbf{D}_m\mathbf{z}\right)\right) \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{M}{2}}} \exp\left(-\frac{\left(\mathbf{y}_m - \mathbf{D}_m\mathbf{z}\right)^T\left(\mathbf{y}_m - \mathbf{D}_m\mathbf{z}\right)}{2\sigma^2}\right) \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{M}{2}}} \exp\left(-\frac{||\mathbf{y}_m - \mathbf{D}_m\mathbf{z}||^2}{2\sigma^2}\right) \end{aligned} \tag{3.2.4}$$

where $M$ is the width or height of the LR observation (only square images are considered). Substituting equation (3.2.4) in equation (3.2.2)

$$\begin{aligned} \hat{\mathbf{z}}_{ML} &= \arg\max_{\mathbf{z}}[log \sum_{m=1}^{p} \frac{1}{(2\pi\sigma^2)^{\frac{M}{2}}} \exp\left(-\frac{||\mathbf{y}_m - \mathbf{D}_m\mathbf{z}||^2}{2\sigma^2}\right)] \\ &= \arg\max_{\mathbf{z}}[\log \sum_{m=1}^{p} \frac{1}{(2\pi\sigma^2)^{\frac{M}{2}}} + \sum_{m=1}^{p} \log\left(\exp\left(-\frac{||\mathbf{y}_m - \mathbf{D}_m\mathbf{z}||^2}{2\sigma^2}\right)\right)] \end{aligned}$$

The first term in the above equation is independent of $\mathbf{z}$ and hence can be omitted. Then, the above equation becomes

$$\hat{\mathbf{z}}_{ML} = \arg\max_{\mathbf{z}}[\sum_{m=1}^{p}\left(-\frac{||\mathbf{y}_m - \mathbf{D}_m\mathbf{z}||^2}{2\sigma^2}\right)] \tag{3.2.5}$$

A condition for the MLE is that it has to satisfy the following equation, known as the *likelihood equation* [19]

$$\frac{\partial}{\partial \mathbf{z}} \log p(\mathbf{y_1}, \mathbf{y_2}, \cdots, \mathbf{y_p}|\mathbf{z})|_{\mathbf{z}=\hat{\mathbf{z}}_{ML}} = 0 \tag{3.2.6}$$

According to equations (3.2.4) and (3.2.5), the *likelihood equation* for this case can

be written as

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{z}} \sum_{m=1}^{p} \left( -\frac{||\mathbf{y}_m - \mathbf{D}_m \mathbf{z}||^2}{2\sigma^2} \right) |_{\mathbf{z}=\hat{\mathbf{z}}_{ML}} &= 0 \\
\sum_{m=1}^{p} \frac{\partial}{\partial \mathbf{z}} \left( -\frac{(\mathbf{y}_m - \mathbf{D}_m \hat{\mathbf{z}}_{ML})^T (\mathbf{y}_m - \mathbf{D}_m \hat{\mathbf{z}}_{ML})}{2\sigma^2} \right) &= 0 \\
\sum_{m=1}^{p} \frac{\partial}{\partial \mathbf{z}} \left( \mathbf{y}_m^T \mathbf{y}_m - \hat{\mathbf{z}}_{ML}^T \mathbf{D}_m^T \mathbf{y}_m - \mathbf{y}_m^T \mathbf{D}_m \hat{\mathbf{z}}_{ML} - \hat{\mathbf{z}}_{ML}^T \mathbf{D}_m^T \mathbf{D}_m \hat{\mathbf{z}}_{ML} \right) &= 0 \\
\sum_{m=1}^{p} \left( -\mathbf{D}_m^T \mathbf{y}_m + 2\mathbf{D}_m^T \mathbf{D}_m \hat{\mathbf{z}}_{ML} \right) &= 0 \\
\sum_{m=1}^{p} \left( \mathbf{D}_m^T \mathbf{D}_m \hat{\mathbf{z}}_{ML} \right) &= \sum_{m=1}^{p} \mathbf{D}_m^T \mathbf{y}_m \\
\left( \sum_{m=1}^{p} \left( \mathbf{D}_m^T \mathbf{D}_m \right) \right) \hat{\mathbf{z}}_{ML} &= \sum_{m=1}^{p} \mathbf{D}_m^T \mathbf{y}_m
\end{aligned}
\tag{3.2.7}
$$

Solving equation (3.2.7) for $\hat{\mathbf{z}}_{ML}$ gives

$$
\hat{\mathbf{z}}_{ML} = \left[ \sum_{m=1}^{p} \mathbf{D}_m^T \mathbf{D}_m \right]^{-1} \sum_{m=1}^{p} \mathbf{D}_m^T \mathbf{y}_m \tag{3.2.8}
$$

This ML estimate can be directly calculated for *small* images. In Section 4.1, this estimate is calculated for different scenarios using a 16x16 pixel image segment as the HR image. The decimation matrix D is a $M^2 \times N^2$ matrix, where the lexicographically ordered LR image vector $\mathbf{y}$ has the dimension of $M^2 \times 1$, and the HR image to be estimated is a $N^2 \times 1$ vector. Then, the $D^T D$ is a $N^2 \times N^2$ matrix. Therefore, for larger N values (=> 32) due to memory limitations, equation 3.2.8 cannot be evaluated directly. In this case, iterative methods like gradient descent algorithm must be applied to evaluate the ML estimate.

# Chapter 4

# Simulation Results

## 4.1 Maximum Likelihood Estimates of Different Image Combinations

Using equation 3.2.8, this section presents the maximum likelihood (ML) estimates and peak signal-to-noise ratio (PSNR) values gained by a single LR image and combining multiple LR images with the same zoom factor and with different zoom factors. As described at the end of section 3.2, due to memory limitations, equation 3.2.8 cannot be computed directly for large HR images. Therefore, in the following subsections, the ML estimate is calculated for a 16x16 pixel HR image. Instead of displaying gray scaled images, pixel values are shown in the following subsections to represent images.

### 4.1.1 One Low-Resolution Observation

Figure 4.1 shows the intensity values of an image section of size 16x16 pixels, which was considered here as the high-resolution image. Normally distributed noise samples were generated and added to the HR image. This noise represented the additive noise at the receiver. After adding noise to this HR image, intensity values of the noisy image and intensity values of the obtained low-resolution image from it are shown in Figures 4.2 and 4.3, respectively. The LR image size is 4x4, which makes the decimation factor 4. Therefore, the LR image is gained by averaging four neighboring pixels. Figure 4.4 shows the ML estimate of the LR image calculated according to equation (3.2.8). This ML estimate is the

same as if the LR image was expanded by repeating each pixel value of the LR image over the 4x4 neighborhood. Thus, for the case of one LR image, the ML estimate was achieved by expanding it, as described previously. This means that according to equation 3.2.8, if there is only one LR observation, the resolution of the image cannot be enhanced.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 38 | 50 | 50 | 50 | 42 | 38 | 42 | 33 | 33 | 14 | 14 | 24 | 42 | 42 | 38 | 24 |
| 42 | 53 | 42 | 53 | 42 | 33 | 24 | 24 | 33 | 38 | 38 | 33 | 42 | 50 | 42 | 33 |
| 33 | 50 | 33 | 50 | 33 | 33 | 24 | 33 | 38 | 38 | 24 | 24 | 33 | 38 | 33 | 33 |
| 38 | 46 | 42 | 38 | 14 | 14 | 33 | 24 | 33 | 38 | 14 | 4 | 14 | 14 | 24 | 14 |
| 42 | 33 | 42 | 24 | 24 | 24 | 33 | 24 | 14 | 24 | 14 | 4 | 4 | 14 | 14 | 4 |
| 50 | 38 | 38 | 14 | 14 | 14 | 14 | 4 | 33 | 24 | 24 | 4 | 4 | 33 | 33 | 24 |
| 53 | 38 | 42 | 14 | 14 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 24 | 33 | 4 |
| 42 | 46 | 42 | 4 | 4 | 14 | 14 | 4 | 14 | 4 | 4 | 14 | 38 | 50 | 50 | 33 |
| 38 | 42 | 14 | 14 | 24 | 24 | 24 | 24 | 24 | 14 | 14 | 38 | 38 | 46 | 46 | 33 |
| 33 | 42 | 14 | 14 | 14 | 14 | 24 | 46 | 50 | 46 | 24 | 24 | 24 | 14 | 33 | 24 |
| 24 | 33 | 14 | 14 | 14 | 14 | 14 | 33 | 38 | 42 | 42 | 33 | 33 | 14 | 4 | 4 |
| 42 | 24 | 14 | 4 | 14 | 14 | 33 | 38 | 42 | 38 | 33 | 4 | 14 | 24 | 4 | 4 |
| 46 | 14 | 14 | 14 | 24 | 14 | 33 | 33 | 33 | 14 | 33 | 4 | 4 | 24 | 14 | 24 |
| 50 | 14 | 14 | 14 | 14 | 14 | 24 | 24 | 33 | 42 | 38 | 14 | 14 | 14 | 14 | 24 |
| 50 | 14 | 14 | 4 | 14 | 14 | 24 | 33 | 14 | 24 | 24 | 14 | 14 | 4 | 4 | 4 |
| 42 | 14 | 14 | 14 | 24 | 33 | 42 | 33 | 33 | 33 | 14 | 14 | 14 | 4 | 4 | 14 |

Figure 4.1: Pixel values of an image section of size 16x16 pixels.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 45.68 | 52.36 | 48.77 | 51.49 | 44.56 | 42.62 | 39.30 | 27.46 | 34.14 | 13.92 | 13.90 | 28.27 | 41.86 | 42.16 | 34.13 | 26.49 |
| 48.38 | 47.33 | 42.86 | 58.63 | 48.09 | 33.74 | 26.07 | 23.95 | 34.51 | 38.34 | 36.79 | 39.75 | 43.02 | 48.46 | 42.95 | 31.17 |
| 35.52 | 53.83 | 35.61 | 50.82 | 34.25 | 29.52 | 28.64 | 40.80 | 39.01 | 40.57 | 22.24 | 24.52 | 34.60 | 38.77 | 30.69 | 34.66 |
| 38.09 | 45.80 | 38.90 | 42.77 | 11.27 | 21.63 | 30.43 | 21.88 | 27.06 | 34.80 | 16.94 | -1.39 | 10.74 | 16.27 | 20.38 | 19.28 |
| 43.71 | 31.75 | 41.67 | 25.00 | 31.69 | 22.72 | 28.96 | 24.82 | 16.15 | 20.82 | 9.83 | 4.87 | 4.31 | 14.48 | 9.53 | 6.53 |
| 52.16 | 39.92 | 38.40 | 16.54 | 11.34 | 16.89 | 19.44 | 2.82 | 33.73 | 24.89 | 23.04 | 5.24 | 3.63 | 33.42 | 31.12 | 26.79 |
| 51.13 | 40.03 | 42.19 | 15.83 | 14.88 | 3.57 | 4.32 | 8.16 | 7.91 | 4.91 | -4.22 | 3.68 | 6.17 | 20.81 | 34.64 | 3.29 |
| 41.17 | 49.22 | 43.17 | 9.62 | 6.59 | 18.15 | 11.46 | 1.93 | 14.39 | 3.21 | 6.46 | 14.55 | 43.95 | 54.13 | 45.27 | 33.93 |
| 42.79 | 44.95 | 13.67 | 11.03 | 27.88 | 25.01 | 20.04 | 24.19 | 24.56 | 13.30 | 15.90 | 32.18 | 39.02 | 49.79 | 45.72 | 31.35 |
| 36.18 | 45.88 | 11.79 | 11.10 | 13.79 | 12.49 | 27.91 | 43.67 | 48.08 | 48.84 | 26.98 | 19.24 | 23.69 | 5.84 | 32.96 | 23.50 |
| 24.95 | 32.21 | 12.77 | 15.18 | 16.04 | 14.24 | 18.83 | 32.43 | 34.72 | 39.96 | 38.76 | 35.59 | 33.09 | 7.40 | 2.90 | 3.68 |
| 39.41 | 21.76 | 14.05 | 6.87 | 8.39 | 13.66 | 38.61 | 41.43 | 41.06 | 37.42 | 32.78 | 3.25 | 12.05 | 25.22 | 7.11 | 7.15 |
| 44.44 | 12.12 | 18.32 | 14.48 | 24.18 | 18.48 | 34.99 | 33.43 | 37.60 | 16.27 | 33.25 | -1.15 | -1.47 | 21.27 | 16.03 | 25.37 |
| 52.74 | 13.17 | 15.31 | 13.36 | 11.59 | 16.23 | 24.26 | 23.95 | 38.70 | 42.95 | 32.41 | 12.99 | 15.51 | 10.10 | 23.23 | 23.91 |
| 51.14 | 17.93 | 14.21 | 8.70 | 8.65 | 15.16 | 30.76 | 30.03 | 9.78 | 28.89 | 18.36 | 11.48 | 12.58 | 12.35 | 0.05 | 2.79 |
| 41.74 | 9.10 | 14.93 | 12.03 | 27.49 | 31.09 | 45.99 | 28.32 | 34.22 | 32.86 | 11.91 | 16.18 | 15.22 | 1.14 | 4.49 | 13.23 |

Figure 4.2: Pixel values of the noisy image. For representation convenience only two decimal points are shown.

| | | | |
|---|---|---|---|
| 46.0586 | 31.5185 | 27.7157 | 32.2316 |
| 36.3487 | 14.2378 | 11.8475 | 23.2546 |
| 24.0418 | 23.6678 | 30.7945 | 21.9103 |
| 22.1117 | 25.2918 | 23.5465 | 12.2416 |

Figure 4.3: Pixel values of low-resolution image gained from noisy image. These pixel values are obtained by averaging 4x4 blocks of the HR noisy image.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 46.05 | 46.05 | 46.05 | 46.05 | 31.51 | 31.51 | 31.51 | 31.51 | 27.71 | 27.71 | 27.71 | 27.71 | 32.23 | 32.23 | 32.23 | 32.23 |
| 46.05 | 46.05 | 46.05 | 46.05 | 31.51 | 31.51 | 31.51 | 31.51 | 27.71 | 27.71 | 27.71 | 27.71 | 32.23 | 32.23 | 32.23 | 32.23 |
| 46.05 | 46.05 | 46.05 | 46.05 | 31.51 | 31.51 | 31.51 | 31.51 | 27.71 | 27.71 | 27.71 | 27.71 | 32.23 | 32.23 | 32.23 | 32.23 |
| 46.05 | 46.05 | 46.05 | 46.05 | 31.51 | 31.51 | 31.51 | 31.51 | 27.71 | 27.71 | 27.71 | 27.71 | 32.23 | 32.23 | 32.23 | 32.23 |
| 36.34 | 36.34 | 36.34 | 36.34 | 14.23 | 14.23 | 14.23 | 14.23 | 11.84 | 11.84 | 11.84 | 11.84 | 23.25 | 23.25 | 23.25 | 23.25 |
| 36.34 | 36.34 | 36.34 | 36.34 | 14.23 | 14.23 | 14.23 | 14.23 | 11.84 | 11.84 | 11.84 | 11.84 | 23.25 | 23.25 | 23.25 | 23.25 |
| 36.34 | 36.34 | 36.34 | 36.34 | 14.23 | 14.23 | 14.23 | 14.23 | 11.84 | 11.84 | 11.84 | 11.84 | 23.25 | 23.25 | 23.25 | 23.25 |
| 36.34 | 36.34 | 36.34 | 36.34 | 14.23 | 14.23 | 14.23 | 14.23 | 11.84 | 11.84 | 11.84 | 11.84 | 23.25 | 23.25 | 23.25 | 23.25 |
| 24.04 | 24.04 | 24.04 | 24.04 | 23.66 | 23.66 | 23.66 | 23.66 | 30.79 | 30.79 | 30.79 | 30.79 | 21.91 | 21.91 | 21.91 | 21.91 |
| 24.04 | 24.04 | 24.04 | 24.04 | 23.66 | 23.66 | 23.66 | 23.66 | 30.79 | 30.79 | 30.79 | 30.79 | 21.91 | 21.91 | 21.91 | 21.91 |
| 24.04 | 24.04 | 24.04 | 24.04 | 23.66 | 23.66 | 23.66 | 23.66 | 30.79 | 30.79 | 30.79 | 30.79 | 21.91 | 21.91 | 21.91 | 21.91 |
| 24.04 | 24.04 | 24.04 | 24.04 | 23.66 | 23.66 | 23.66 | 23.66 | 30.79 | 30.79 | 30.79 | 30.79 | 21.91 | 21.91 | 21.91 | 21.91 |
| 22.11 | 22.11 | 22.11 | 22.11 | 25.29 | 25.29 | 25.29 | 25.29 | 23.54 | 23.54 | 23.54 | 23.54 | 12.24 | 12.24 | 12.24 | 12.24 |
| 22.11 | 22.11 | 22.11 | 22.11 | 25.29 | 25.29 | 25.29 | 25.29 | 23.54 | 23.54 | 23.54 | 23.54 | 12.24 | 12.24 | 12.24 | 12.24 |
| 22.11 | 22.11 | 22.11 | 22.11 | 25.29 | 25.29 | 25.29 | 25.29 | 23.54 | 23.54 | 23.54 | 23.54 | 12.24 | 12.24 | 12.24 | 12.24 |
| 22.11 | 22.11 | 22.11 | 22.11 | 25.29 | 25.29 | 25.29 | 25.29 | 23.54 | 23.54 | 23.54 | 23.54 | 12.24 | 12.24 | 12.24 | 12.24 |

Figure 4.4: Maximum likelihood estimate gained from LR image. For representation convenience only two decimal points are shown.

### 4.1.2 Two Low-Resolution Observations with Different Zoom Factors

This situation, where two observations of a single scene taken at different zoom settings are available, was considered in this thesis. By calculating the ML estimate and observing its pixel values provides a clue as to how to combine the differently zoomed observations. In this section, two LR observations that are related by a zoom factor of 4 are considered. A 16x16 pixel image segment is considered here as the HR image (see Figure 4.5) from which the LR images were obtained. The first LR image was obtained by averaging 4x4 pixel blocks of the HR image and adding Gaussian noise. This LR image contains the entire scene available on the HR image. The second observation was obtained by cropping the HR image to the centered 4x4 pixel block, as shown by the square in Figure 4.5 and adding Gaussian noise. This LR observation can be considered as an image taken by zooming in the HR image. The resolution of this observation is same as the image section shown in the rectangle in the HR image.

Figures 4.6 and Figure 4.7 show the noisy LR observations. Using these observations, the ML estimate of the model considered in this thesis is calculated using equation 3.2.8. The calculated ML estimate is shown in Figure 4.8. As it can be seen in this figure, the calculated ML estimate can be considered as the image obtained by enlarging the first LR observation containing the entire scene to 16x16 pixels and replacing the middle part by the second observation.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 38 | 50 | 50 | 50 | 42 | 38 | 42 | 33 | 33 | 14 | 14 | 24 | 42 | 42 | 38 | 24 |
| 42 | 53 | 42 | 53 | 42 | 33 | 24 | 24 | 33 | 38 | 38 | 33 | 42 | 50 | 42 | 33 |
| 33 | 50 | 33 | 50 | 33 | 33 | 24 | 33 | 38 | 38 | 24 | 24 | 33 | 38 | 33 | 33 |
| 38 | 46 | 42 | 38 | 14 | 14 | 33 | 24 | 33 | 38 | 14 | 4 | 14 | 14 | 24 | 14 |
| 42 | 33 | 42 | 24 | 24 | 24 | 33 | 24 | 14 | 24 | 14 | 4 | 4 | 14 | 14 | 4 |
| 50 | 38 | 38 | 14 | 14 | 14 | 14 | 4 | 33 | 24 | 24 | 4 | 4 | 33 | 33 | 24 |
| 53 | 38 | 42 | 14 | 14 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 24 | 33 | 4 |
| 42 | 46 | 42 | 4 | 4 | 14 | 14 | 4 | 14 | 4 | 4 | 14 | 38 | 50 | 50 | 33 |
| 38 | 42 | 14 | 14 | 24 | 24 | 24 | 24 | 24 | 14 | 14 | 38 | 38 | 46 | 46 | 33 |
| 33 | 42 | 14 | 14 | 14 | 14 | 24 | 46 | 50 | 46 | 24 | 24 | 24 | 14 | 33 | 24 |
| 24 | 33 | 14 | 14 | 14 | 14 | 14 | 33 | 38 | 42 | 42 | 33 | 33 | 14 | 4 | 4 |
| 42 | 24 | 14 | 4 | 14 | 14 | 33 | 38 | 42 | 38 | 33 | 4 | 14 | 24 | 4 | 4 |
| 46 | 14 | 14 | 14 | 24 | 14 | 33 | 33 | 33 | 14 | 33 | 4 | 4 | 24 | 14 | 24 |
| 50 | 14 | 14 | 14 | 14 | 14 | 24 | 24 | 33 | 42 | 38 | 14 | 14 | 14 | 14 | 24 |
| 50 | 14 | 14 | 4 | 14 | 14 | 24 | 33 | 14 | 24 | 24 | 14 | 14 | 4 | 4 | 4 |
| 42 | 14 | 14 | 14 | 24 | 33 | 42 | 33 | 33 | 33 | 14 | 14 | 14 | 4 | 4 | 14 |

Figure 4.5: Pixel values of the 16x16 pixel HR image.

| | | | |
|---|---|---|---|
| 43.0250 | 35.1519 | 28.2279 | 30.2464 |
| 35.7082 | 12.7365 | 9.5603 | 23.1773 |
| 23.7205 | 19.8482 | 33.0121 | 22.1266 |
| 20.3275 | 28.7706 | 26.6518 | 9.5236 |

Figure 4.6: Pixel values of the first LR image obtained by averaging the HR image.

By enlarging the first LR observation, which contains the entire scene to 16x16 pixels, and comparing with the HR image, the calculated peak signal-to-noise ratio was -12.62dB. The calculated PSNR of the ML estimate compared with the HR image was -11.05dB. Therefore, the ML estimate achieved a 1.5dB gain compared to the first LR image.

| | | | |
|---|---|---|---|
| 6.6469 | 2.5341 | 3.3848 | 0.2422 |
| 13.9655 | 4.7029 | 18.9474 | 2.7175 |
| 25.1990 | 24.2388 | 27.3732 | 13.5845 |
| 22.3980 | 50.1327 | 45.6507 | 46.0200 |

Figure 4.7: Pixel values of the second LR image. This observation contains the corresponding image section shown in the rectangle in Figure 4.5.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 43.02 | 43.02 | 43.02 | 43.02 | 35.15 | 35.15 | 35.15 | 35.15 | 28.22 | 28.22 | 28.22 | 28.22 | 30.24 | 30.24 | 30.24 | 30.24 |
| 43.02 | 43.02 | 43.02 | 43.02 | 35.15 | 35.15 | 35.15 | 35.15 | 28.22 | 28.22 | 28.22 | 28.22 | 30.24 | 30.24 | 30.24 | 30.24 |
| 43.02 | 43.02 | 43.02 | 43.02 | 35.15 | 35.15 | 35.15 | 35.15 | 28.22 | 28.22 | 28.22 | 28.22 | 30.24 | 30.24 | 30.24 | 30.24 |
| 43.02 | 43.02 | 43.02 | 43.02 | 35.15 | 35.15 | 35.15 | 35.15 | 28.22 | 28.22 | 28.22 | 28.22 | 30.24 | 30.24 | 30.24 | 30.24 |
| 35.70 | 35.70 | 35.70 | 35.70 | 14.66 | 14.66 | 14.66 | 14.66 | 10.63 | 10.63 | 10.63 | 10.63 | 23.17 | 23.17 | 23.17 | 23.17 |
| 35.70 | 35.70 | 35.70 | 35.70 | 14.66 | 14.66 | 14.66 | 14.66 | 10.63 | 10.63 | 10.63 | 10.63 | 23.17 | 23.17 | 23.17 | 23.17 |
| 35.70 | 35.70 | 35.70 | 35.70 | 14.66 | 14.66 | 6.64 | 2.53 | 3.38 | 0.24 | 10.63 | 10.63 | 23.17 | 23.17 | 23.17 | 23.17 |
| 35.70 | 35.70 | 35.70 | 35.70 | 14.66 | 14.66 | 13.96 | 4.70 | 18.94 | 2.71 | 10.63 | 10.63 | 23.17 | 23.17 | 23.17 | 23.17 |
| 23.72 | 23.72 | 23.72 | 23.72 | 16.30 | 16.30 | 25.19 | 24.23 | 27.37 | 13.58 | 32.96 | 32.96 | 22.12 | 22.12 | 22.12 | 22.12 |
| 23.72 | 23.72 | 23.72 | 23.72 | 16.30 | 16.30 | 22.39 | 50.13 | 45.65 | 46.02 | 32.96 | 32.96 | 22.12 | 22.12 | 22.12 | 22.12 |
| 23.72 | 23.72 | 23.72 | 23.72 | 16.30 | 16.30 | 16.30 | 16.30 | 32.96 | 32.96 | 32.96 | 32.96 | 22.12 | 22.12 | 22.12 | 22.12 |
| 23.72 | 23.72 | 23.72 | 23.72 | 16.30 | 16.30 | 16.30 | 16.30 | 32.96 | 32.96 | 32.96 | 32.96 | 22.12 | 22.12 | 22.12 | 22.12 |
| 20.32 | 20.32 | 20.32 | 20.32 | 28.77 | 28.77 | 28.77 | 28.77 | 26.65 | 26.65 | 26.65 | 26.65 | 9.52 | 9.52 | 9.52 | 9.52 |
| 20.32 | 20.32 | 20.32 | 20.32 | 28.77 | 28.77 | 28.77 | 28.77 | 26.65 | 26.65 | 26.65 | 26.65 | 9.52 | 9.52 | 9.52 | 9.52 |
| 20.32 | 20.32 | 20.32 | 20.32 | 28.77 | 28.77 | 28.77 | 28.77 | 26.65 | 26.65 | 26.65 | 26.65 | 9.52 | 9.52 | 9.52 | 9.52 |
| 20.32 | 20.32 | 20.32 | 20.32 | 28.77 | 28.77 | 28.77 | 28.77 | 26.65 | 26.65 | 26.65 | 26.65 | 9.52 | 9.52 | 9.52 | 9.52 |

Figure 4.8: Maximum likelihood estimate gained from two LR images.

## 4.2 Simulation Setup

### 4.2.1 Two LR-Observations with Same Angle

Reconstruction of an HR image using two images of the same scene taken at the same angle but at different zoom settings is illustrated in this section. The high-resolution image is shown in Figure 4.9. Reducing the size of this image to 200x200 pixels and adding noise produced the second LR observation (see Figure 4.10, right hand image). The other low-resolution image was obtained by cropping a 200x200 pixel area of the high-resolution image and adding noise. The cropped area is shown by the red square in Figure 4.9. The high-resolution image was cropped in this manner so that two low-resolution images were obtained, which are related by a zoom factor of two.

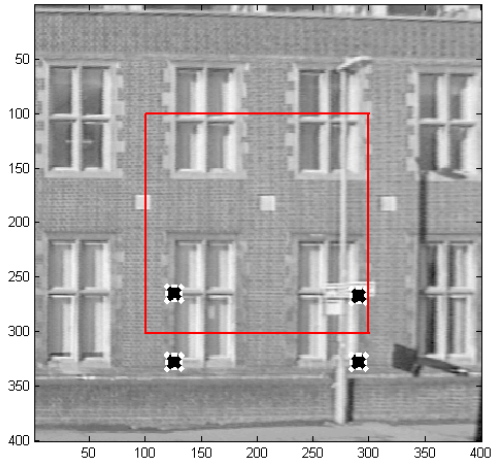


Figure 4.9: High-resolution ground truth image.

The detected feature points of both low-resolution observations, after applying the Harris detector [7], are shown in white squares in Figure 4.11. The Harris detector was applied to the second low-resolution observation, after resizing it to 400x400 pixels. This is legitimate, since the zoom factor (here 2) between the two observations is assumed to be known. All the detected feature points are numbered in blue. The numbers (listed down at a feature point) indicate putative correspondence points in the right-hand image of that

particular feature point in the first low-resolution observation. As shown, there are multiple correspondence points to a single feature point in the first observation. These points were found after specifying the confidence of spatial consistency originally found at all feature points in both images. Then the pairs were selected to have a confidence value defined by equation (2.3.7).

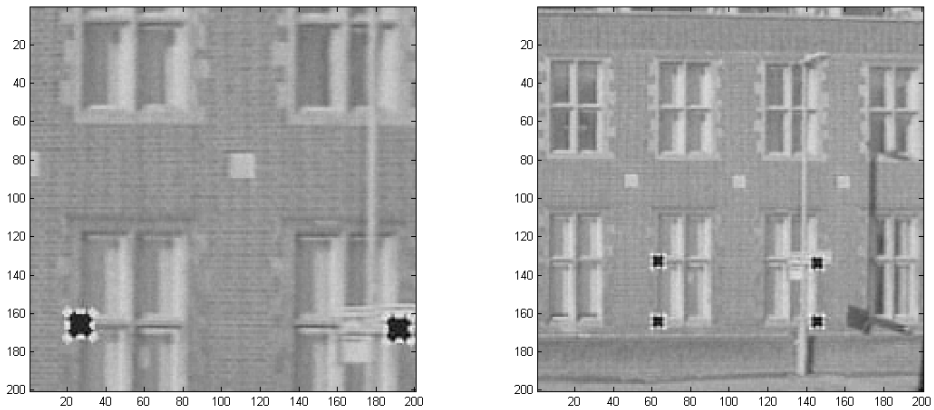

Figure 4.10: Low resolution observations of size 200x200.

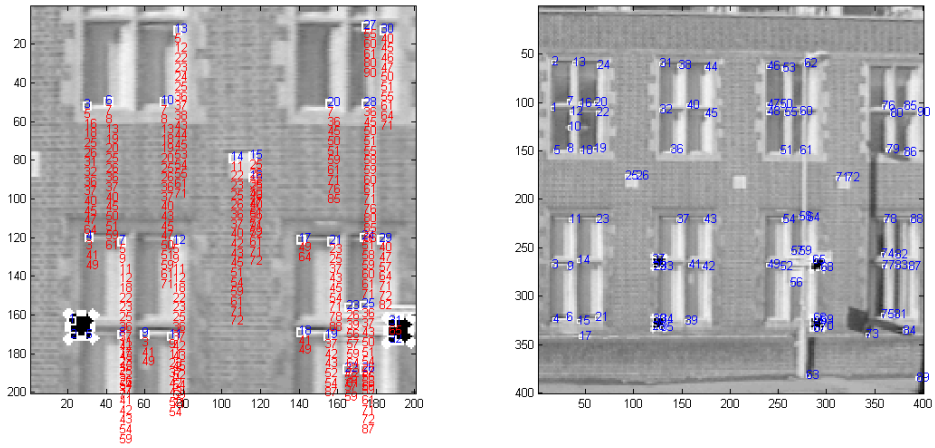

Figure 4.11: Correspondence points found after applying confidence of spatial consistency.

After applying confidence of global smoothness to tentative candidates selected according to equation (2.3.12), the RANSAC algorithm, as explained in Section 2.3, was applied. Finally, correspondence points were obtained, as shown in Figure 4.12. The crosses

represent the finally selected feature points. The numbers on the left image represent the corresponding point in the right image.

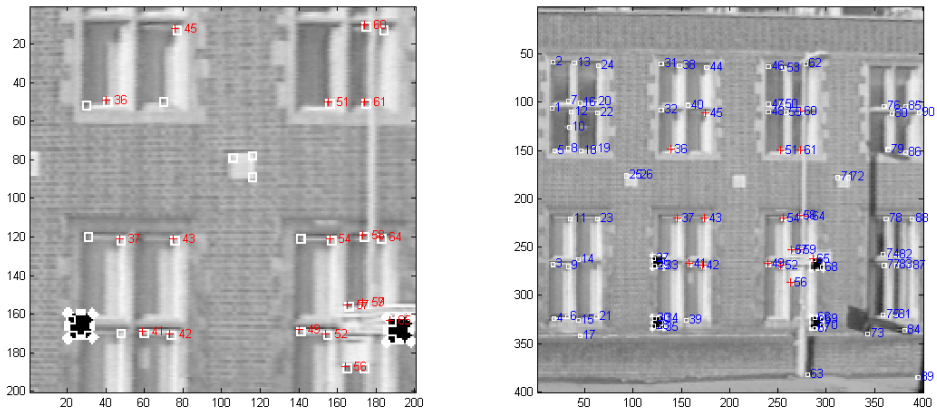

Figure 4.12: Correspondence points after applying global smoothness condition and RANSAC.

Then the geometric transformation (homography matrix as discussed in Section 2.2 was applied to the first low-resolution observation to obtain the aligned image to the enlarged second observation (see Figure 4.13, left hand image).

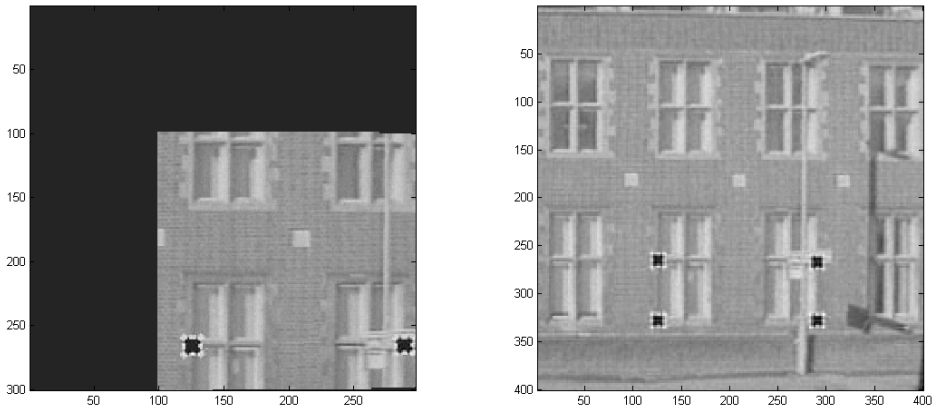

Figure 4.13: After applying the geometric transformation to observation one (left) and enlarged observation two (right).

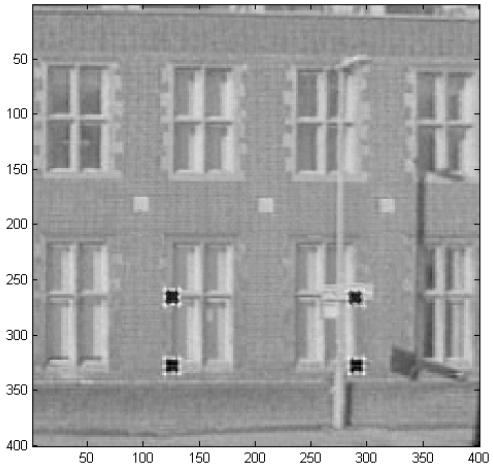


Figure 4.14: HR image obtained by combining images in Figure 4.13.

### Discussion of Simulation Results

Figure 4.14 shows the high-resolution image constructed after averaging the overlapping areas of images shown in Figure 4.13. The overlapping area is highlighted in a square region and shown in Figure 4.15. The calculated PSNR values for the enlarged second LR observation and for the reconstructed high-resolution image are 27.88 dB and 27.86 dB, respectively. There is even a small degradation in the reconstructed image according to those values, as clearly shown in Figure 4.16. The area of the reconstructed HR image (right) is sharper than the same area of the enlarged LR image.

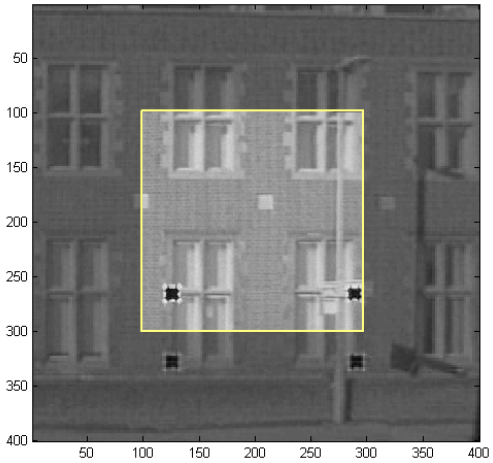


Figure 4.15: Illustration of different areas of HR image.

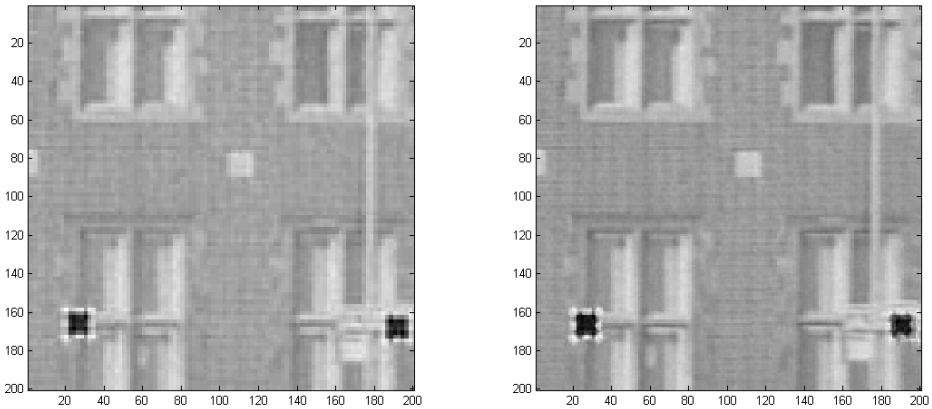


Figure 4.16: Common area cropped from resized LR observation 2 (left) and reconstructed image by averaging.

### 4.2.2 Two LR-Observations with Different Angles and Same Zoom Factor

Reconstruction of an HR image using two images of the same scene taken at different angles but at same zoom settings is illustrated in this section. The high-resolution images are shown in Figure 4.17. Reducing the size of these images to 200x200 pixels and adding noise produced the LR observations (see Figure 4.18).

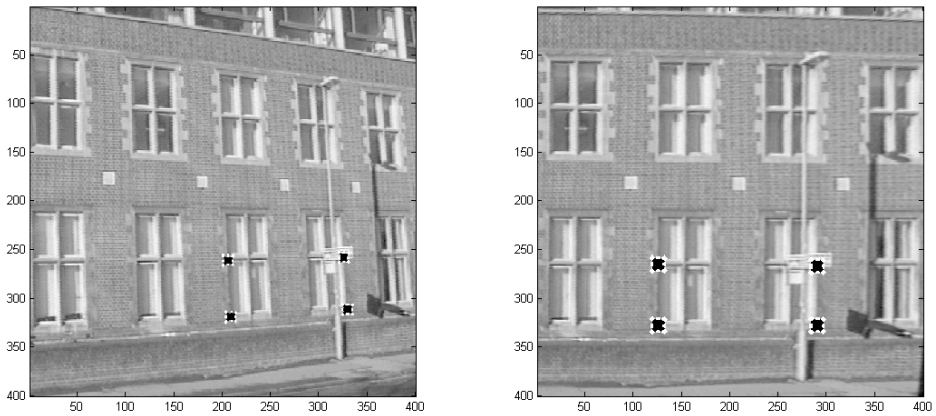


Figure 4.17: High-resolution ground truth images

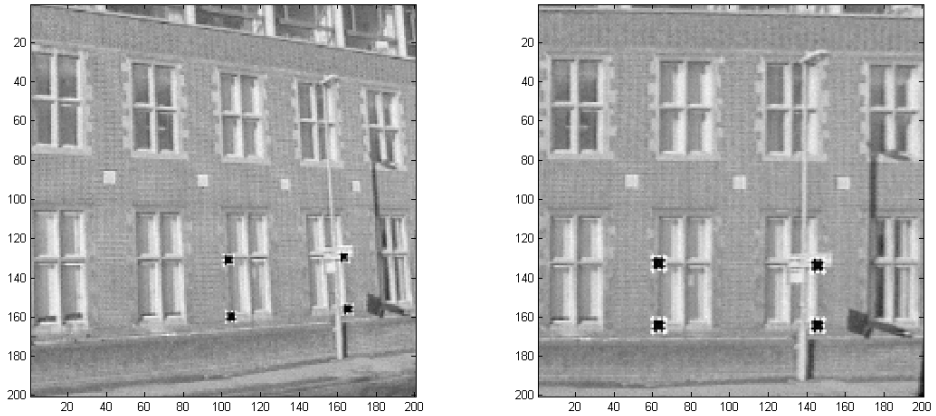


Figure 4.18: Low-Resolution observations of size 200x200

The detected feature points of both low-resolution observations, after applying the Harris detector [7], are shown in white squares in Figure 4.19. All the detected feature

points are numbered in blue. The numbers (listed down at a feature point) indicate putative correspondence points in the right-hand image of that particular feature point in the first low-resolution observation. As shown, there are multiple correspondence points to a single feature point in the first observation. These points were found after specifying the confidence of spatial consistency originally found at all feature points in both images. Then the pairs were selected to have a confidence value defined by equation (2.3.7).

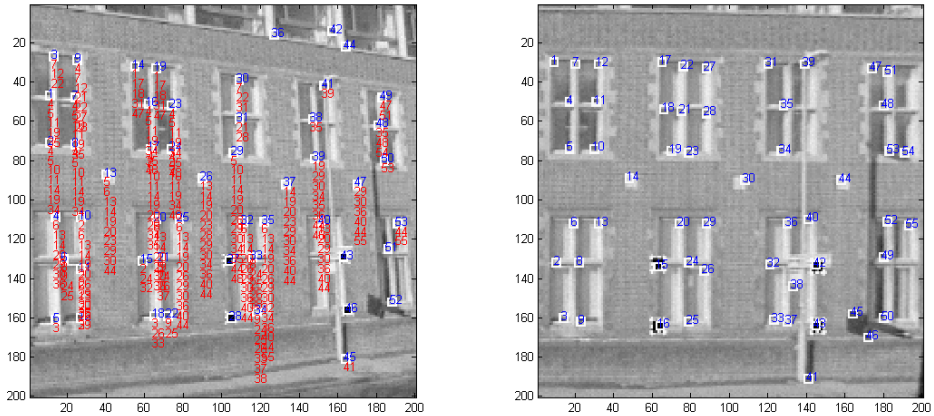

Figure 4.19: Correspondence points found after applying confidence of spatial consistency.

For these putative correspondences, the global smoothness condition was applied, and best correspondence pairs were selected using the RANSAC algorithm. Also, the transformation matrix (homography matrix **H**) was calculated at this step. The remaining correspondence points are shown in Figure 4.20. Corresponding points in both images are shown in crosses, and the corresponding point in image two to a point in image one is shown in a number in image one. As can be seen, the ambiguity of correspondence points has vanished.

Figure 4.21 (left hand image) shows the warped low-resolution observation one using the homography matrix. Since this image contains some parts that are not contained in the second low-resolution image, the resized second observation is shifted accordingly (right). HR image obtained after combining the figures shown in Figure 4.21 is shown in Figure 4.22 (here only the the common area to both the LR images are shown).

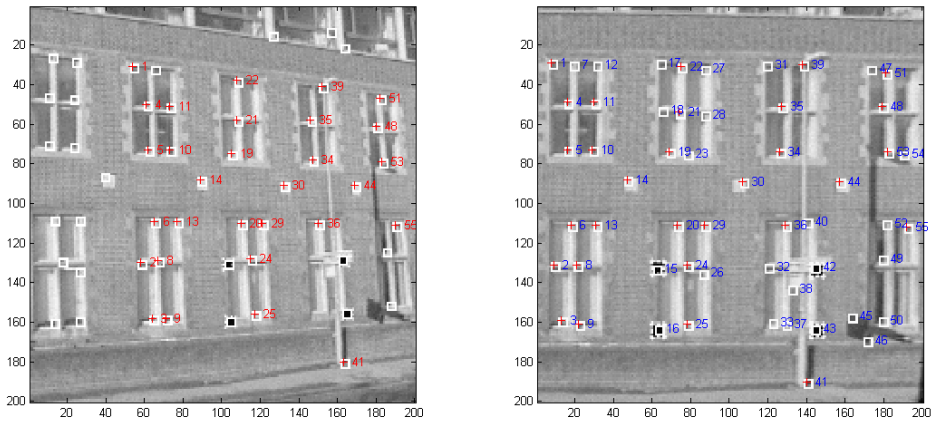

Figure 4.20: Correspondence points after applying global smoothness condition and RANSAC.

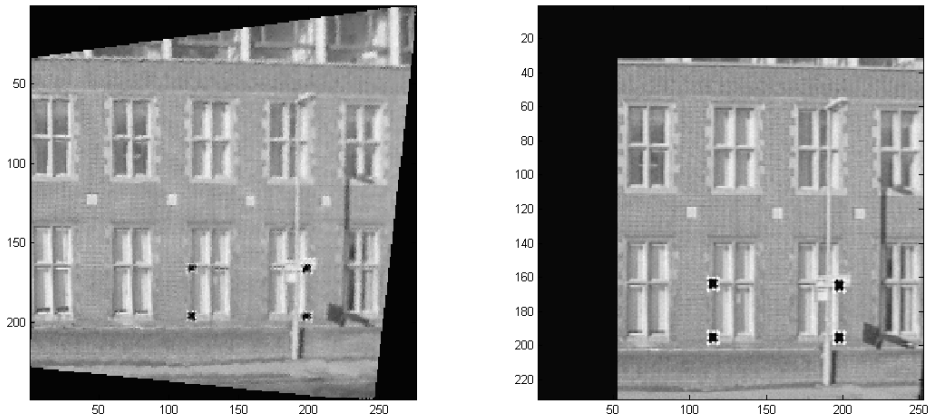

Figure 4.21: After applying the geometric transformation to observation one (left) and enlarged observation two (right).

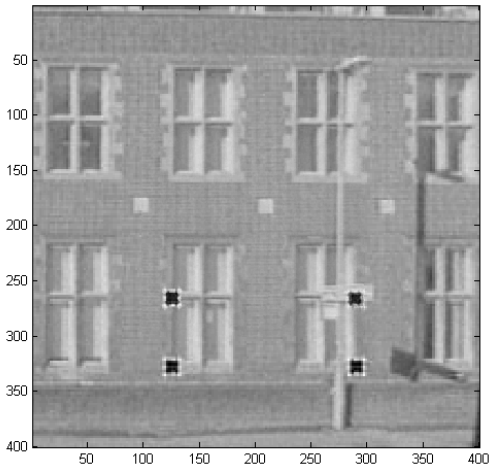


Figure 4.22: HR image obtained by combining images in Figure 4.13.

### 4.2.3 Two LR-observations with Different Angles and Different Zoom Factors

Reconstruction of an HR image using two images of the same scene taken at different angles and at different zoom settings is illustrated in this section. Figure 4.23 shows the high-resolution images from which the low-resolution images were obtained. The first low-resolution image was obtained by cropping the first high-resolution image so that the zoom factor between the low-resolution observation became two. The cropped portion of the high-resolution image one is illustrated by the red square in Figure 4.23 (left). By resizing the second high-resolution image to 200x200 pixels, the second low-resolution observation was obtained. Noise was added to both of these low-resolution images, which are shown in Figure 4.24.

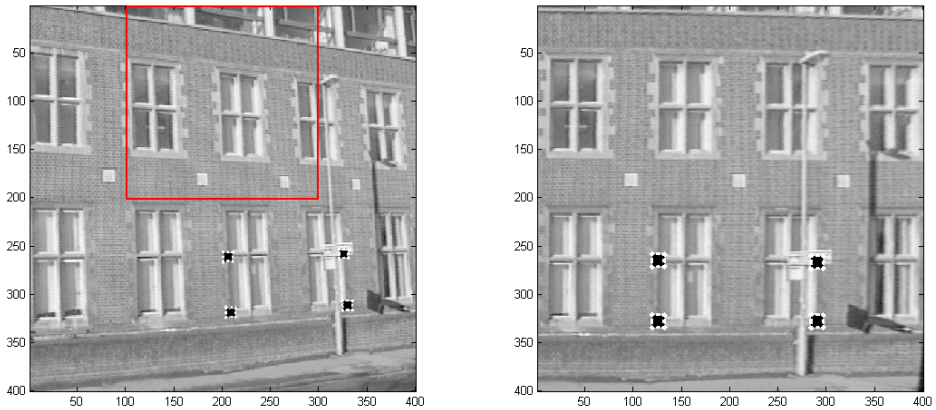


Figure 4.23: High-resolution ground truth images.

The Harris corner detector was applied to observation one, and observation two was resized to 400x400 pixels. Detected feature points on both images are shown in small white squares and numbered from 1 to the maximum number of detected feature points in each image (see Figure 4.25). Local correlation values were calculated and confidence values were assigned to these feature points. After applying the confidence of spatial consistency and selecting those pairs with higher confidence values, as expressed in equation (2.3.7), the calculated putative correspondence points are shown in numbers listed down at selected

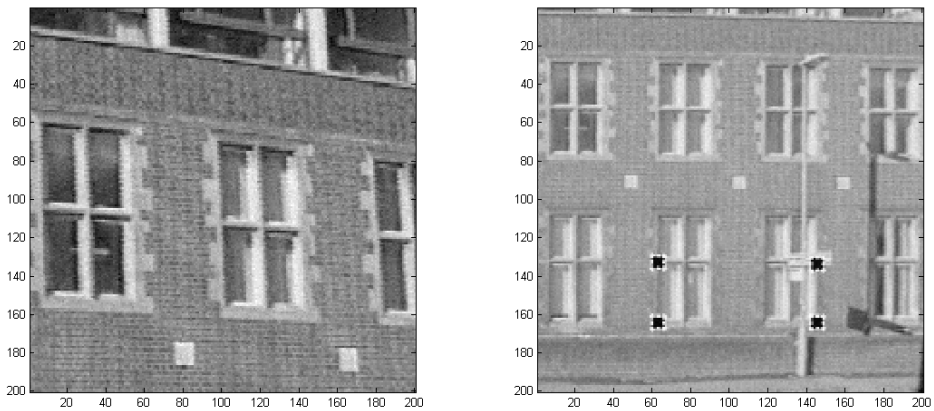

Figure 4.24: Low-resolution observations.

feature points in Figure 4.25 (left). This means a feature point in image one has a number of putative correspondences in image two given by the numbers listed down at a feature point. As shown in Figure 4.25 (left), there are many candidate points in image two for a particular feature point in image one.

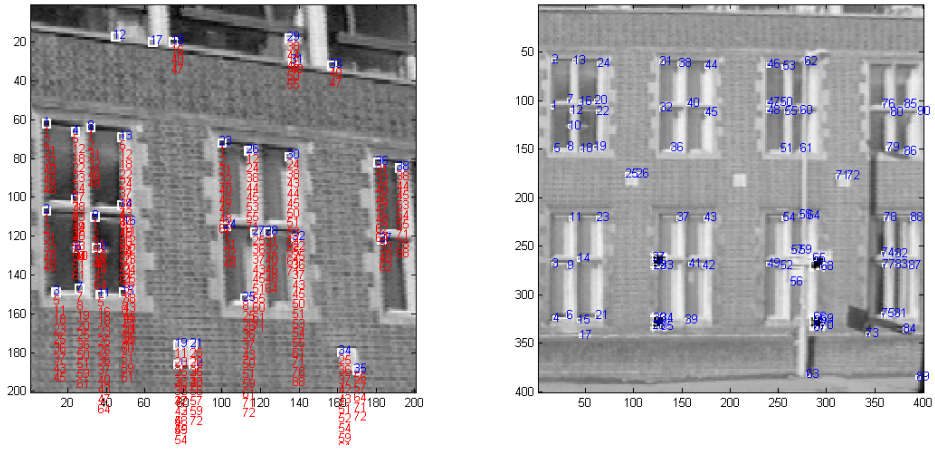

Figure 4.25: Correspondence points found after applying confidence of spatial consistency.

For these putative correspondences, the global smoothness condition was applied, and best correspondence pairs were selected using the RANSAC algorithm. Also, the transformation matrix (homography matrix **H**) was calculated at this step. The remaining correspondence points are shown in Figure 4.26. Corresponding points in both images are shown

in crosses, and the corresponding point in image two to a point in image one is shown in a number in image one. As can be seen, the ambiguity of correspondence points has vanished.

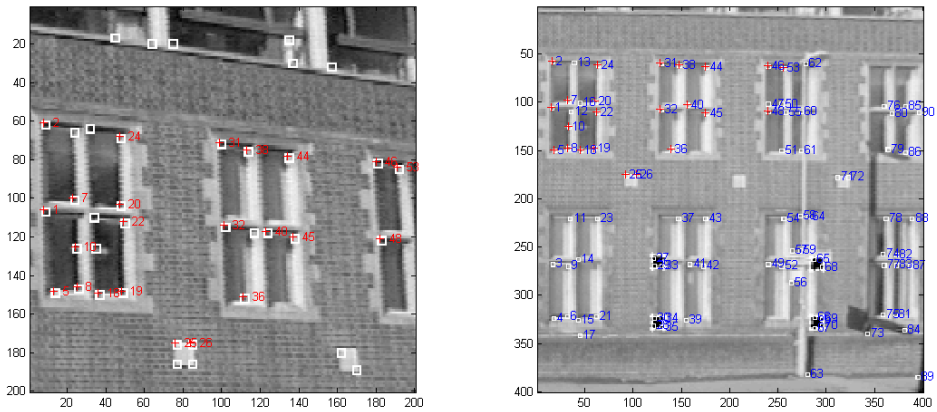

Figure 4.26: Correspondence points after applying global smoothness condition and RANSAC.

Figure 4.27 (left hand image) shows the warped low-resolution observation one using the homography matrix. Since this image contains some parts that are not contained in the second low-resolution image, the resized second observation is shifted accordingly (right). Figure 4.28 shows the mosaic image constructed by combining the images shown in Figure 4.27.

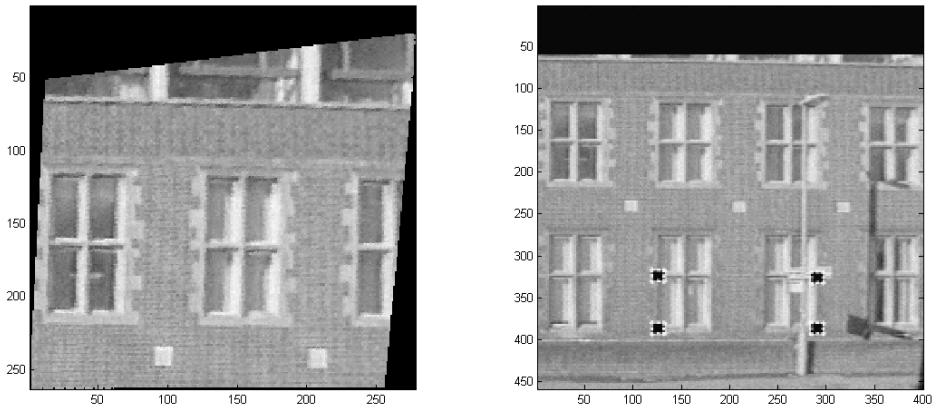

Figure 4.27: LR observation warped to align observation two (left) and shifted observation two (right).

The marked areas (A,B,C,D) in Figure 4.29 are common to both images shown in Figure 4.27. Therefore, for this region, a better peak-signal-to-noise ratio can be achieved by averaging the pixel values. In Figure 4.29, the areas A, B, C, E, D', and D are highlighted for illustration.

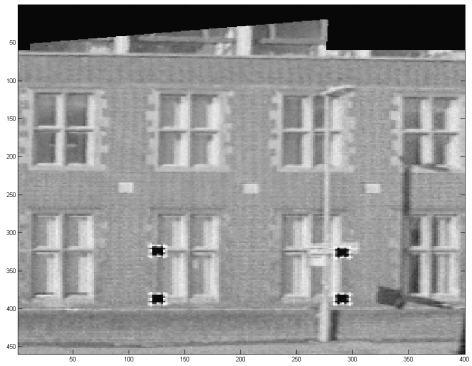

Figure 4.28: Mosaic image obtained by combining images from Figure 4.27

**Discussion of Simulation Results**

The PSNR was calculated for the resized observation and mosaic image. The PSNR value of 28.021 dB was obtained by the enlarged observation two (see Figure 4.27 right), and 27.92 dB was obtained by the mosaic image. To calculate the PSNR value of the mosaic image, the image area that is common to the mosaic image and the enlarged observation was used. Theoretically, this PSNR value should be greater than the value obtained by the enlarged observation two. This imperfection can be caused by not perfectly aligning the two images, that is, by inaccurately estimating the homography matrix. The corresponding area shown as A, B, C, and D in Figure 4.29 is cropped from the enlarged LR observation and the reconstructed HR image, and shown in Figure 4.30 (for simplicity the trapeze area is cropped to a rectangular area). Even in this case, comparing the images shown in Figure 4.30, the area obtained by averaging pixel values is sharper than the same area of the enlarged observation two.

In the reconstructed high-resolution image the area shown by D, C, E, and D' was

not available in the second low-resolution observation. By applying this procedure, higher resolution was gained in the area (A, B,C, D) and new image section (D, C, E, D') in the reconstructed image.

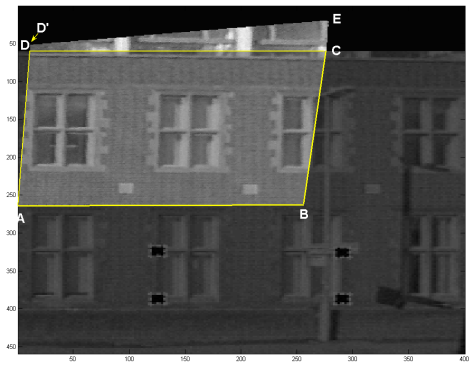


Figure 4.29: Illustration of different areas of mosaic image.

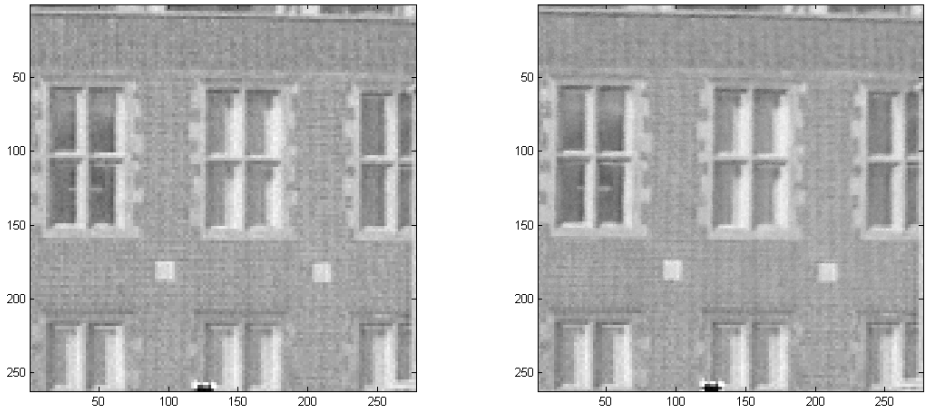


Figure 4.30: Common area cropped from resized LR observation two (left) and reconstructed image by averaging.

Figures 4.31 to Figure 4.36 show the same sequence of images as in the previous simulation results. The difference here is that the feature detection is applied to the LR image of size 200x200 pixels, not after enlarging to 400x400 pixels, as in the previous case. The PSNR value of 28.01 dB was obtained by the enlarged observation two, and 26.80 dB was obtained by the mosaic image.

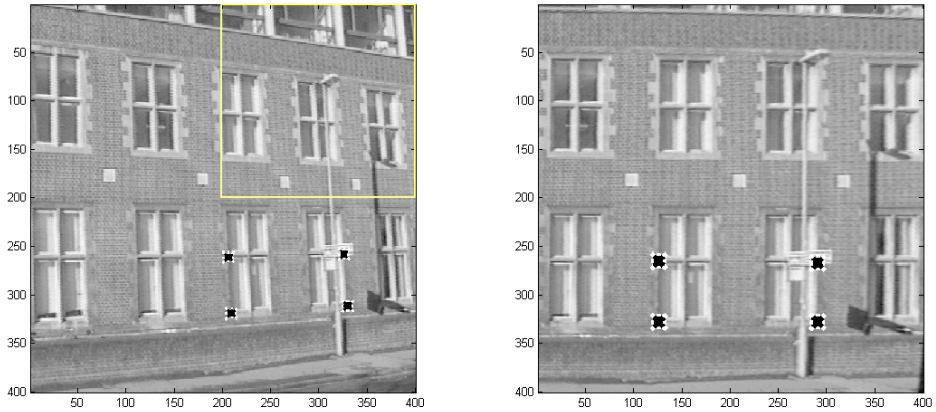

Figure 4.31: High-resolution ground truth images.

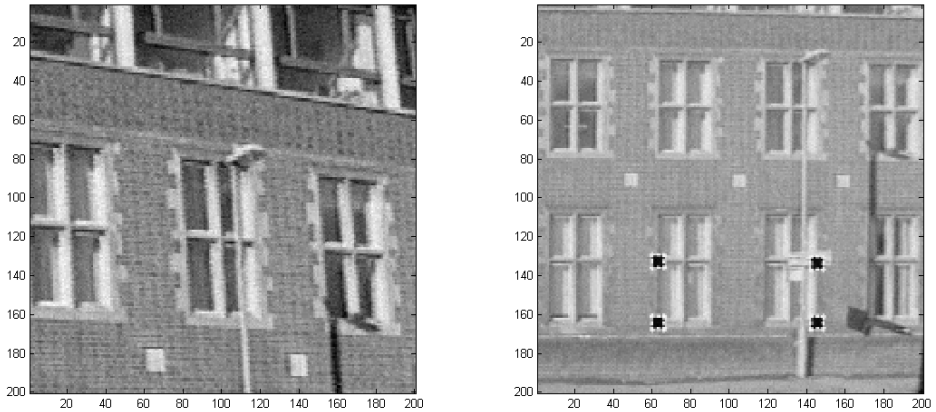

Figure 4.32: Low-resolution observations.

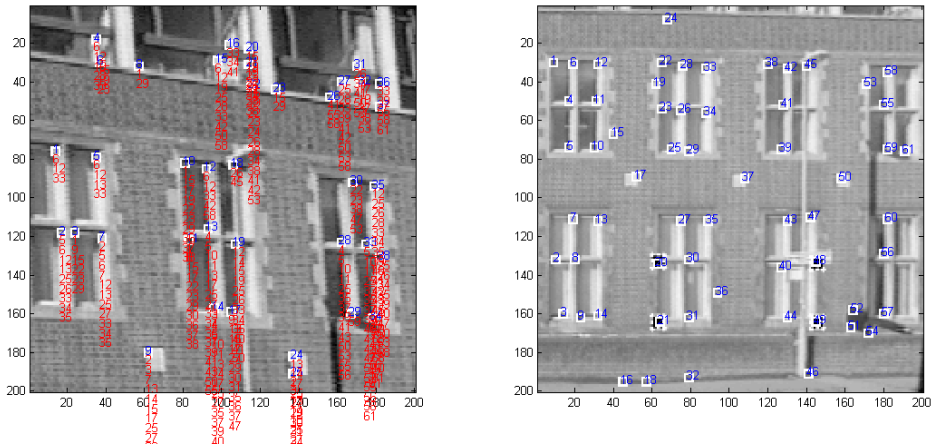

Figure 4.33: Correspondence points found after applying confidence of spatial consistency.

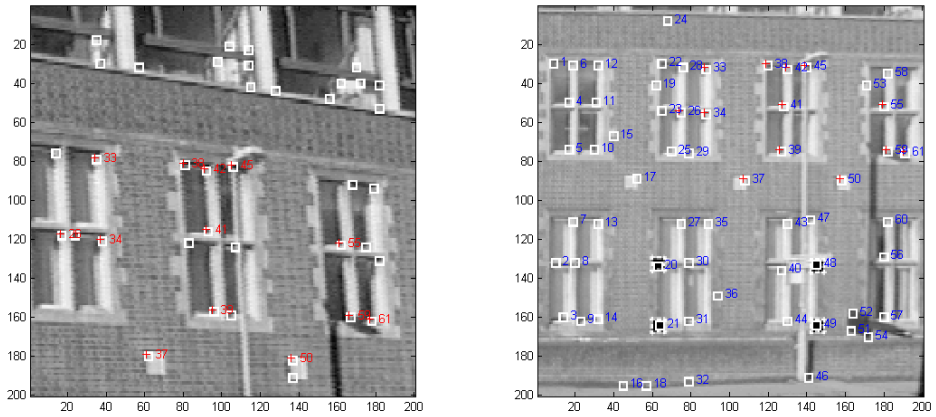

Figure 4.34: Correspondence points after applying global smoothness condition and RANSAC.

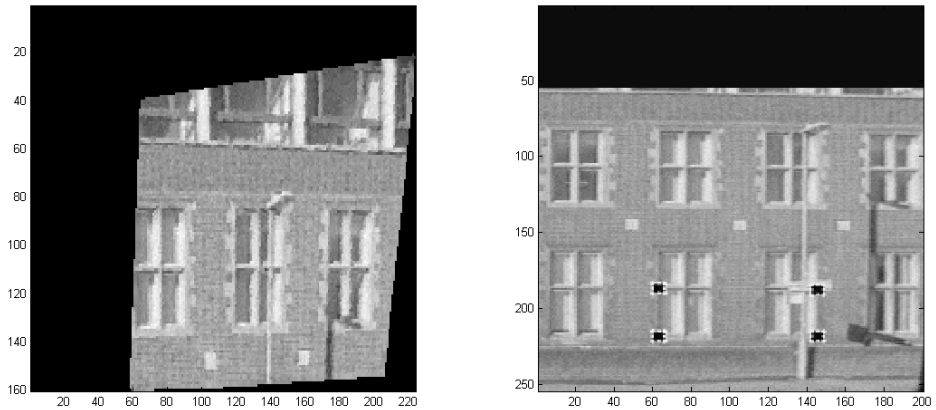

Figure 4.35: LR observation warped to align observation two (left) and shifted observation two (right).

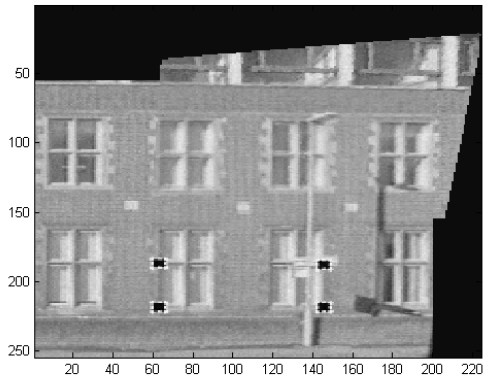

Figure 4.36: Mosaic image obtained by combining images from Figure 4.35.

# Chapter 5

# Conclusions

In this thesis, a fully automated algorithm was implemented to reconstruct high-resolution images by using low-resolution observations taken with different zoom settings and different angles. A point matching algorithm was proposed to deal with low-resolution observations which differ by a zoom factor of two.

Simulations were done with two observations taken at the same angle but with different zoom factors (two). Results show that there was a small degradation in the reconstructed high-resolution image regarding PSNR values. But comparing the area where the pixel values were averaged, the reconstructed high-resolution image was sharper than any of the low-resolution observations.

Simulation results obtained using observations taken at different angles and at different zoom settings also showed the same behavior as in the same angle case. Also, more image content was available in the super-resolved image than in any of the single low-resolution observations.

This algorithm, however, failed when at least 90 percent of the image contents of the highly zoomed image did not have information that was available on the lower-zoomed image. This means that at least 90 percent of information contained in the highly zoomed image should also be contained in the lower-zoomed image.

# REFERENCES

LIST OF REFERENCES


[1] M. V. Joshi and S. Chaudhuri, “Super reolution imaging: Use of zoom as a cue,” in *Proc. ICVGIP*, Aug. 2001, pp. 343–358.

[2] Y. Kanazawa and K. Kanatani, “Robust image matching preserving global consistency,” in *Australia-Japan Advanced Workshop on Computer Vision*, 2003, p. September.

[3] M. Fischler and R. Bolles, “Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography,” *Comm. of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[4] D. Capel and A. Zisserman, “Computer vision applied to super resolution,” *IEEE Signal Processing Magazine*, pp. 75–86, 2003.

[5] O. Landsiedel, K. Wehrle, and S. Goetz, “Accurate prediction of power consumption in sensor networks,” in *Proceedings of The Second IEEE Workshop on Embedded Networked Sensors* (EmNetS-II), Sydney, Australia, May 2005.

[6] S. C. Park, M. K. Park, and M. G. Kang, “Super resolution image reconstruction: A technical overview,” *IEEE*, pp. 21–36, May 2003.

[7] C. Harris and M. Stephens, *“A combined corner and edge detector,” in Proc. 4th Alvey Vision Conf.*, Manchester, UK., August 1988, pp. 147–151.

[8] S. Mann and R. Picard, “Video orbits of the projective group: A simple approach to featureless estimation of paramters,” *IEEE Trans. on Image Processing*, vol. IP-6, no. 9, pp. 1281–1295, Sept. 1997.

[9] M. Kumar, S. Kuthirummal, C. Jawahar, and P. Narayan, “Planar homography from fourier domain representation,” in *Proceedings of the International Conference on Signal Processing and Communications* (SPCOM), Bangalore, India, Dec 2004.

[10] G. Cortelazzo and L. Lucchese, “A new method of image mosaicking and its application to cultural heritage representation,” *Computer Graphics Forum*, vol. 18, no. 3, p. 265, September 1999.

[11] R. Radke, P. Ramadge, T. Echigo, and S. Iisak, “Efficiently estimating projective transformations,” in *International Conference on Image Processing*, vol. 1, Vancouver, BC, Canada, 2000, pp. 232–235.

[12] J. Denton and J. R. Beveridge, “Two dimensional projective point matching,” in Fifth IEEE Southwest Symposium on Image Analysis and Interpretation, April 2002, pp. 77–81.

[13] M. Zuliani, C. Kenney, and B. Manjunath, “The multiransac algorithm and its application to detect planar homographies,” in IEEE International Conference on Image Processing,Genova, Italy, Sep 2005.

[14] R. Hartley and A. Zisserman, Eds., *Multiple View Geometry in Computer Vision*, 2nd ed., Cambridge University Press, 2003.

[15] M. G. Kang and S. Chaudhuri, “Super resolution image reconstructing,” *IEEE Sig. Pro. Mag*., pp. 19–20, May 2003.

[16] J. Besag, “Spatial interaction and the statistical analysis of lattice systems,” *Journal of Royal Statistical Society*, Series B, vol. 36, pp. 192–236, 1974.

[17] R. C. Dubes and A. K. Jain, “Random field models in image analysis,” *Journal of Applied Statistics*, vol. 16, no. 2, pp. 131–164, 1989.

[18] D. Ranjan and S. Chaudhuri, “An MRF-based approach to generation of super - resolution images from blured observations,” *Journal of Mathmatical Imaging and Vision*, pp. 5–15, 2002.

[19] H. V. Poor, *An Introduction to Signal Detection and Estimation* (2nd ed.)., NY, USA: Springer-Verlag Inc., 1994.

[20] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*. N.Y,: Addison-Wesley, 1989.