1-1-2015

Development Of A High Performance Mosaicing And Super-Resolution Algorithm

Debabrata Ghosh

Recommended Citation

Ghosh, Debabrata, "Development Of A High Performance Mosaicing And Super-Resolution Algorithm" (2015). *Theses and Dissertations*. 1896.
https://commons.und.edu/theses/1896

DEVELOPMENT OF A HIGH PERFORMANCE MOSAICING AND SUPER-RESOLUTION ALGORITHM

by

Debabrata Ghosh

A Dissertation
Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Grand Forks, North Dakota
December
2015

This dissertation, submitted by Debabrata Ghosh in partial fulfillment of the requirements for the Degree of Doctor of Philosophy from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

Naima Kaabouch, PhD

Saleh Faruque, PhD

William Semke, PhD

Wen-Chen Hu, PhD

Ronald Fevig, PhD

This dissertation is being submitted by the appointed advisory committee as having met all of the requirements of the School of Graduate Studies at the University of North Dakota, and is hereby approved.

Wayne Swisher,
Dean of the Graduate School

Date

PERMISSION

Title	Development Of A High Performance Mosaicing And Super-resolution Algorithm
Department	Electrical Engineering
Degree	Doctor of Philosophy

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work or, in his absence, by the chairperson of the department or the dean of the Graduate School. It is understood that any copying or publication or other use of this dissertation or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my dissertation.

Debabrata Ghosh
07/03/2015

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost I want to thank my advisor Dr. Naima Kaabouch. I appreciate all her contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating. I am also thankful to the other members of my advisory committee: Dr. Saleh Faruque, Dr. William Semke, Dr. Wen-Chen Hu, and Dr. Ronald Fevig for their helpful feedback and support during this process.

Lastly, I would like to thank my parents and brother for all their love and encouragement. I am grateful to my loving, supportive, encouraging, and patient wife Soumita Ghosh whose faithful support during the final stages of this Ph.D. is so appreciated. Thank you.

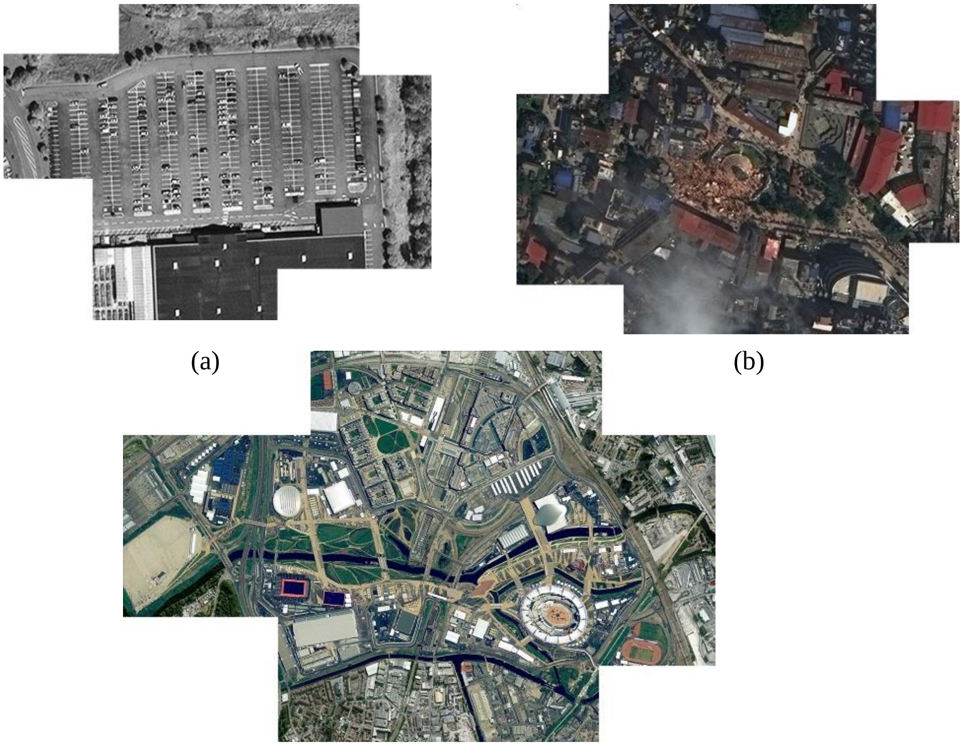
ABSTRACT

In this dissertation, a high-performance mosaicing and super-resolution algorithm is described. The scale invariant feature transform (SIFT)-based mosaicing algorithm builds an initial mosaic which is iteratively updated by the robust super resolution algorithm to achieve the final high-resolution mosaic. Two different types of datasets are used for testing: high altitude balloon data and unmanned aerial vehicle data. To evaluate our algorithm, five performance metrics are employed: mean square error, peak signal to noise ratio, singular value decomposition, slope of reciprocal singular value curve, and cumulative probability of blur detection. Extensive testing shows that the proposed algorithm is effective in improving the captured aerial data and the performance metrics are accurate in quantifying the evaluation of the algorithm.

CHAPTER 1

INTRODUCTION

There are various applications of super-resolution mosaicing, including surveillance, disaster management, and urban mapping. Example results of super-resolution mosaicing relevant to the aforementioned applications are shown in Figure 1.1. The two-fold benefit of super-resolution mosaicing is obvious: 1) instead of processing each and individual frames to analyze a certain scene, this technique gives the advantage of processing a single integral frame, thus it saves significant processing time and 2) higher spatial resolution output of this technique provides the advantage of better content visualization, which is critical in all of the aforementioned applications.



(a) (b)

(c)

Figure. 1.1: Applications of super-resolution mosaicing. (a) Application in surveillance; (b) Application in disaster management; (c) Application in urban mapping.

Higher the quality of the images, better the spatial resolution. The sensor size and detector density primarily determine the spatial resolution of the captured images. Larger the size of the sensor, and/or the higher the density of the detectors, the better the spatial resolution of the acquired images. The most direct hardware-based approach of increasing the spatial resolution is to reduce the detector size or, equivalently, to increase the detector density. Alternatively, the sensor size can also be increased. However, smaller detectors have lower dynamic range, lower fill factor, lower light sensitivity, higher dark signal, higher diffraction sensitivity, and higher non-uniformity [1]. In addition, the hardware cost increases with both the increase of detector density and sensor size. Thus, the aforementioned hardware-based approach often restricts the maximum achievable resolution of the captured images. Besides the sensor-imposed restriction, there are several other factors that limit the quality of the captured images, including lens and atmospheric blurs, finite shutter speed, finite aperture, movement of objects in the scene, sensor noise, and media turbulence [2]. Similarly, the frames lose spatial resolution during video acquisition due to sensor array sampling. Some of these limitations can be overcome by employing computationally intensive image processing algorithms that require high processing time and high-power computers, making them unsuitable for low to medium budget systems.

In order to overcome the limitations inherent in the commercially available affordable electronics, an approach that combines the available optics and a resolution-enhancement algorithm can be used. This technique would overcome the final specification restrictions of the commercial optics. The aim of this dissertation is to develop an efficient, robust and automated super-resolution mosaicing algorithm that is able to run successfully even with limited computational resources. Other than increasing the field of view of commercial imagers, the mosaicing algorithm will add further benefit of eliminating redundant data from overlapping frames, which are required as input for the high-resolution algorithm. The essential steps required for the proposed super-resolution mosaicing are image mosaicing and super-resolution.

Super-resolution reconstruction algorithm (SR) creates a high-resolution (HR) image from a sequence of correlated low-resolution images of the same scene taken from different viewpoints. Since super-resolution increases the spatial resolution by taking advantage of more samples than those found in any single low-resolution image, the presence of motion among the low-resolution images is compulsory for the success of this method. The reconstruction primarily relies on the ability to estimate the aforementioned motion between frames to recover details that are finer than the sampling grid. Simultaneously the effects of blur, noise, and other artifacts are eliminated in the reconstruction process.

Image mosaicing, on the other hand, is the alignment of multiple correlated images into a wider composition. Mosaicing is a special case of scene building where the images are related by planar homography only. This is a reasonable assumption if the images exhibit no parallax effects, i.e. when the scene is approximately planar or the camera purely rotates about its optical center. Using mosaicing, it is possible to extend the field of view of a camera by preserving the original resolution and without introducing undesirable lens deformation.

Combining image mosaicing and super resolution becomes a powerful means to represent all the information on multiple overlapping images and obtain a high-resolution panoramic view of a specific scene. It registers successive frames into a common coordinated system and simultaneously generate mosaic output with an improved spatial resolution. This method is referred to as super-resolution mosaicing. The stability of a super-resolution mosaicing method necessitates that the overlapping images are correlated solely by planar homography, which is fulfilled readily in small satellite applications since the captured images from high altitude do not suffer from parallax effects.

In this dissertation, we will describe a super-resolution mosaicing algorithm and compare its performance with those of other well-known state-of-the-art algorithms. The dissertation is organized as follows: Chapter 2 presents an introduction to image mosaicing framework and reviews the state-of-the-art of image mosaicing techniques. A classification of the techniques is also proposed, highlighting the benefits and drawbacks of different methods. Chapter 3 presents an introduction to super-resolution framework along with the image observation model that has been used in most super-resolution algorithms. A detailed survey of the state-of-the-art super-resolution techniques by classifying them into several categories is also presented. Chapter 4 details the proposed mosaicing technique. All the steps involved, including registration, reprojection, and stitching are described. Finally, some experimental results, based on large datasets, are presented. Chapter 5 presents the proposed super-resolution mosaicing approach in detail. Some experimental results are also discussed and compared to results obtained by other state-of-the-art approaches. Chapter 6 presents the conclusions of this work, and identifies some future research directions.

CHAPTER 2

STATE-OF-THE-ART OF IMAGE MOSAICING METHODS

2.1 Introduction

Image mosaicing is the alignment of multiple overlapping images into a large composition which represents a part of a 3D scene [3]. Mosaicing could be regarded as a special case of scene reconstruction where the images are related by planar homography only [4]. This is a reasonable assumption if the images exhibit no parallax effects, i.e. when the scene is approximately planar or the camera purely rotates about its optical center [5]. Using mosaicing it is possible to extend the field of view (FOV) of a camera by preserving the original resolution and without introducing undesirable lens deformation [6]. There have been a variety of new additions to the classic applications of image mosaicing that primarily aim to augment the FOV. Mosaic construction is finding its practices in many computer vision and computer graphics applications, such as motion detection and tracking [7-9], mosaic-based localization [10,11], resolution enhancement [12-14], augmented reality [15, 16] etc. Furthermore, video compression [17], video indexing [18], and image stabilization [19] are some of the prominent areas where mosaicing is creating significant impacts.

As shown in Figure 2.1, mosaicing involves various steps of image processing: registration, reprojection, stitching, and blending. Registration refers to the establishment of geometric correspondence between a pair of images depicting the same scene. In order to register a set of images, it is required to estimate the geometric transformations, which align the images with

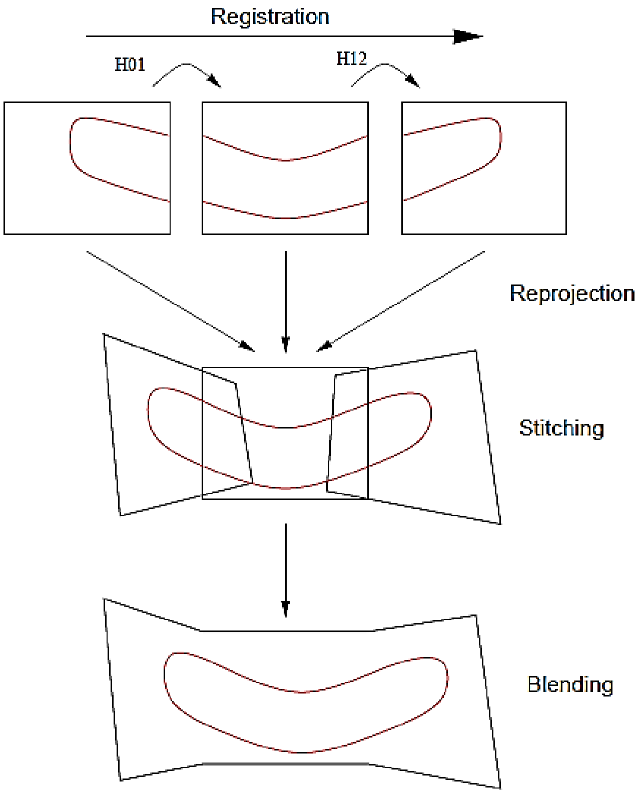


Figure. 2.1: Different steps of image mosaicing. ***Here H are the homography matrices between source images.***

respect to a reference image within that set. The set may consist of two or more images taken of a single scene at different times, from different viewpoints, and/or by different sensors. The most general case of the transformation is the 8 degree of freedom planar homography [3]. The next step, following the registration, is reprojection which refers to the alignment of the images into a common coordinate system using the computed geometric transformations. The goal of the stitching step is to overlay the aligned images on a larger canvas by merging pixel values of the overlapping portions and retaining pixels where no overlap occurs. Errors propagated via geometric and photometric misalignments often result in undesirable object discontinuities and seam visibility in the vicinity of the boundary between two images. Thus, a blending algorithm needs to be used during or after the stitching step in order to minimize the discontinuities in the global appearance of the mosaic.

Image mosaicing is an attractive research area, which has resulted in the development of many algorithms in the literature [12,20-34]. A comprehensive review of the existing algorithms will undoubtedly be a valuable guide to researchers and developers for selecting a suitable image mosaicing method for a specific application. The continuous emergence of new algorithms in recent years further reinforce the necessity of such a comprehensive review. In the following sections, we classify the past and current mosaicing techniques based on image registration as well as image blending. For each of these classifications, we provide a comprehensive review of the major categories of the image mosaicing methods. In addition, this review highlights the evolving paths of those methods by providing the modifications that have been made to those basic methods by different researchers.

Both registration and blending directly influence the performance of image mosaicing. Being the first and last step of image mosaicing, it is almost impossible to build a successful mosaicing algorithm without correctly implementing registration and blending algorithms. Though attempts have been made to overcome the registration errors by utilizing sophisticated blending algorithms, the significance of accurate registration in image mosaicing remains unquestionable. In this chapter, we focus on the classification of the existing image mosaicing algorithms based on their registration methods, as well as based on their blending methods.

2.2 Classification of image mosaicing methods based on registration

Image registration is not only an important step of image mosaicing, but also is the foundation of it. Registration of multi-source images, which are focused on the same target but produced from different sensors, different perspective, and different times, computes the optimal geometric transformation by looking into the correspondences between each pair of images. This process makes the multi-source images aligned into a common reference frame using the estimated geometric transformations. To the extent that corresponding points from multi-source images are aligned together, the registration is successful [41]. The aforementioned correspondences can be established by matching templates between images, or by matching features extracted either from images, or by utilizing the phase correlation property in the frequency domain.

As shown in Figure 2.2, based on image registration methods, image mosaicing algorithms can be spatial domain-based or frequency domain-based. Spatial domain-based image mosaicing can further be grouped into area-based image mosaicing and feature-based image mosaicing. Feature-based image mosaicing can again be subdivided into low-level feature-based image mosaicing and contour-based image mosaicing. Low-level feature-based mosaicing can be divided into four classes: Harris corner detector-based mosaicing, FAST corner detector-based mosaicing, SIFT feature detector-based mosaicing, and SURF detector-based mosaicing. Different classes of image mosaicing algorithms based on the image registration are discussed in details as follows.

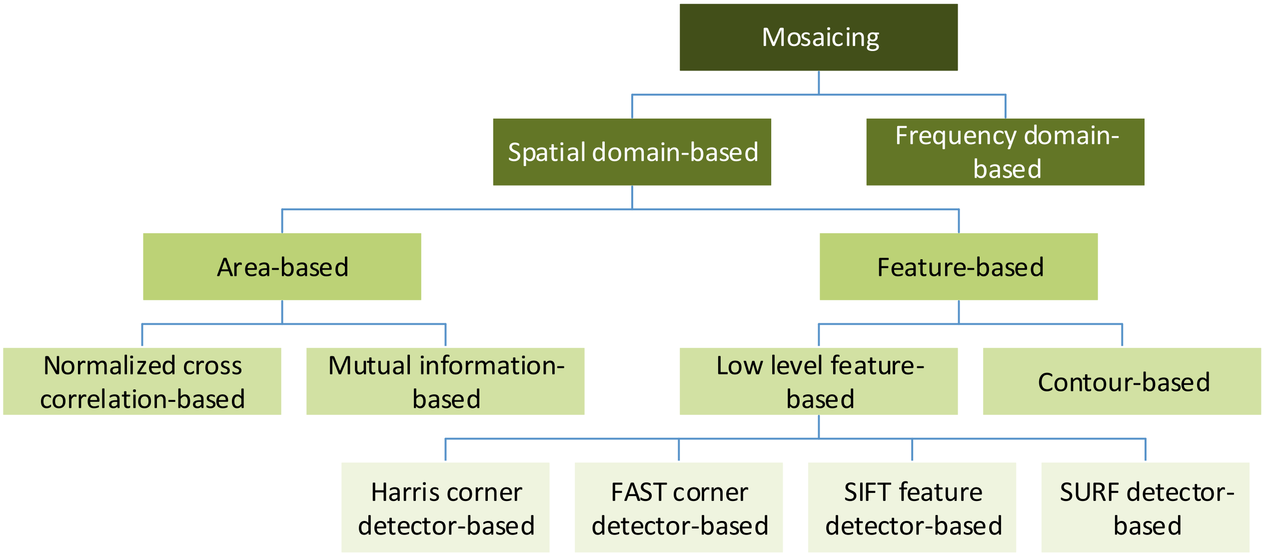


Figure 2.2: Classification of mosaicing based on registration.

2.2.1 Spatial domain image mosaicing methods

Algorithms in this category use properties of pixels to perform registration, and, thus they are the most direct methods of image mosaicing. Majority of the existing image mosaicing algorithms fall into this category. Spatial domain-based image mosaicing can be either area-based or feature-based. Area-based image mosaicing algorithms rely on computation between “windows” of pixel values in the two images, which need to be mosaicked [42]. The fundamental approach is to shift the “windows” of the images relative to each other and see how much the pixels match. Subsequently, transformation parameters are obtained and used to warp and stitch the images. Unlike area-based image mosaicing, feature-based mosaicing methods use feature-to-feature matching in order to compute the geometric transformation between a pair of images. Thus, these methods rely primarily on feature extraction algorithms which can detect salient features from the images. Salient features are subsets of the image domain, often in the form of isolated points, continuous curves or connected regions [49]. Since the features are used as the starting point, the overall algorithm will often be as good as the feature extraction algorithm is.

Two of the most commonly used area-based image mosaicing algorithms are normalized cross correlation-based mosaicing and mutual information-based mosaicing. Based on the types of features extracted, feature-based mosaicing methods can also be classified into low-level feature-based mosaicing and contour-based mosaicing. Again, based on popular low-level feature extraction methods, low-level feature-based mosaicing can be subdivided into the following categories: Harris corner detector-, Features from Accelerated Segment Test (FAST)-, Scale Invariant Feature Transform (SIFT)-, Speeded Up Robust Feature (SURF)-based mosaicing methods. These above-mentioned classes of mosaicing algorithms are described below.

2.2.1.1 Normalized Cross Correlation (NCC)-based mosaicing

This method computes similarity between the "windows" in the two images for each shifts. It is defined as [43]:

$$NCC(u) = \frac{\sum_i \left[U(x_i) - \overline{U}\right]\left[V(x_i + u) - \overline{V}\right]}{\sqrt{\sum_i \left[U(x_i) - \overline{U}\right]^2 \left[V(x_i + u) - \overline{V}\right]^2}} \tag{2.1}$$

where

$$\overline{U} = \frac{1}{N} \sum_i U(x_i) \tag{2.2}$$

$$\overline{V} = \frac{1}{N} \sum_i V(x_i + u) \tag{2.3}$$

where $\overline{U}$ and $\overline{V}$ are the mean images of the corresponding "windows", U and V for the first and second images respectively. N is the number of pixels in the "window", $x_i = (x_i, y_i)$ is the pixel coordinate in the "windows", $u = (u, v)$ is the displacement or shift where NCC coefficient is calculated. The NCC coefficient values are always within the range [-1, 1]. The shift parameter corresponding to the peak NCC value represents the geometric transformation between the two images. Once geometric transformations are obtained between the image pairs, images are warped in the reference frame, and finally stitching is performed to generate the final mosaic. Methods within this category have the advantage of being computationally simple, however, at the cost of being particularly slow. Moreover, they perform accurately only when there are significant overlapping between the source images.

Several techniques [22,44-46] have been proposed to tackle the above mentioned problems. In order to make the computation faster, Berberidis et al. [44] proposed an iterative algorithm for the spatial cross correlation in order to compute the displacements between the source images. Yet another method based on adjusting the correlation windows according to the scale and orientation of extracted interest-points from the source images was proposed by Zhao et al. [45] to increase the computation speed. In order to improve the performance of the algorithm in the presence of non-rigid deformation, Vercauteren et al [46] suggested the use of Riemannian statistics along with a scattered data fitting-based mosaicing. Nasibov et al. [22] employed a brightness correction matrix before the registration step in order to make the algorithm less sensitive to the illumination changes.

2.2.1.2 Mutual Information (MI)-based mosaicing

Unlike NCC, which computes similarity based on image intensity values, mutual information measures similarity based on the quantity of information shared between two images. MI between two images I_1 and I_2 is expressed in terms of entropy as:

$$MI(I_1, I_2) = E(I_1) + E(I_2) - E(I_1, I_2) \tag{2.4}$$

where $E(I_1)$ and $E(I_2)$ are the entropies of I_1 and I_2, respectively. And $E(I_1, I_2)$ represents the joint entropy between the two images. Entropy is a measure of variability of a random variable. Thus variability of I_1 is expressed as:

$$E(I_1) = -\sum_g p_{I_1}(g)\log(p_{I_1}(g)) \tag{2.5}$$

where g are the possible gray level values of I_1, and accordingly $p_{I_1}(g)$ is the probability distribution function of g. Similarly, the joint variability of I_1 and I_2 is expressed as:

$$E(I_1, I_2) = -\sum_{g,h} p_{I_1,I_2}(g,h)\log(p_{I_1,I_2}(g,h)) \tag{2.6}$$

where h indicates the possible gray level values of I_2. $p_{I_1,I_2}(g,h)$ is the joint probability distribution function of g and h. Typically, the joint probability distribution between two images is measured as normalized joint histogram of the gray level values. It is observed that better the alignment between two images, higher the MI between them. Thus, two images are geometrically aligned by a transformation if the MI between them is maximum for that transformation. After the appropriate transformations are obtained between the image pairs, they are reprojected and stitched to get the final mosaic. These mosaicing methods have the advantage of being less sensitive to lighting and occlusion changes between source images. However, similar to NCC-based methods, these techniques have the disadvantages of being computationally slow, and requiring high degree of overlapping between input images.

A number of techniques [24,47,48] have been proposed to address its shortcomings. To increase the computation speed, Dame et al. [47] employed a B-spline function for normalized mutual probability density, in combination with Newton's method for optimizing the MI cost function. Another method presented by Luna et al. [24]uses a stochastic gradient optimization along with MI-based similarity measure in order to make the algorithm faster. Concerning the drawback of MI-based mosaicing algorithms for low overlapping images, Césare et al. [48] proposed a template matching approach capable of explicitly acknowledging the plausibility of similarity between distant neighborhoods, and delaying definite block-to-block association to a step that globally evaluates their collective likelihood.

2.2.1.3 Harris corner detector-based mosaicing

Harris corner detector detects corner points as robust low-level features from source images. Initially a local detection window in an image is chosen. Subsequently the variation in intensity that results by shifting the window by a small amount in different direction is determined as below [39]:

$$E(u,v) = \sum_{x,y} w(x,y) \otimes [I(x+u, y+v) - I(x,y)]^2 \tag{2.7}$$

where $w(x,y)$ is the window function, $I(x,y)$ is the image intensity value at pixel location (x,y), $I(x+u,y+v)$ is the shifted intensity with (u,v) shift, $\otimes$ is the convolution operator. The local texture around pixel (x,y) is expressed as autocorrelation matrix C as below:

$$C = \sum_{x,y} w(x,y) \otimes \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \tag{2.8}$$

where I_x and I_y are the first derivative of $I(x,y)$. Two large eigenvalues for the matrix C corresponds to a corner point. The center point of the window is characterized as a corner point. For more robustness, a "cornerness" measure R is used to eliminate the edge points as below [25]:

$$R = Det(C) - \alpha Tr^2(C) \tag{2.9}$$

where $Tr(C)$ is the trace of C and α is within the range $0.04 \le \alpha \le 0.06$. Corner points are detected as local maxima of R above a predefined threshold T. After the Harris corner points are detected from both the images, correspondences are established either by NCC or by any other Sum of Squared Difference (SDD) method. Subsequently, the geometric motion parameters are calculated and images are warped into a global reference frame in order to stitch them all. Mosaicing algorithms using Harris corner detector are computationally simple and accurate.

One major problem with the Harris corner detector-based mosaicing methods is that large changes in rotation often generates ghosting in the mosaic output. [25] dealt with this by utilizing a luminance center-weighting algorithm which is used following a slope clustering algorithm for Harris corner point matching. Another problem related to the uncertainty in choosing a local detection window was addressed by[51], where the authors used region segmentation and matching in order to limit the search window to potential homologous points. Harris corner detector almost always finds closely crowded feature points. However, this can be overcome by counting the number of feature points in the neighborhood and then accordingly exclude some of the points, as has been done in [26].

2.2.1.4 FAST corner detector-based mosaicing

FAST algorithm is a corner detection algorithm which is computationally more efficient and faster than most of the other low-level feature extraction methods; thus mosaicing methods based on this algorithms are particularly suitable for real-time image processing applications. Initially a circle of sixteen pixels is considered around each corner candidate. According to the FAST algorithm, the candidate is a corner if there exists a set of n contiguous pixels in the circle which are all brighter than the intensity of the candidate pixel plus a threshold, or all darker than the intensity of the candidate pixel minus the threshold, as shown in Figure 2.3. The number n is usually chosen twelve. In order to increase the computational speed of FAST algorithm, a corner response function (CRF) is used. CRF gives the numerical value of the "cornerness" of a corner

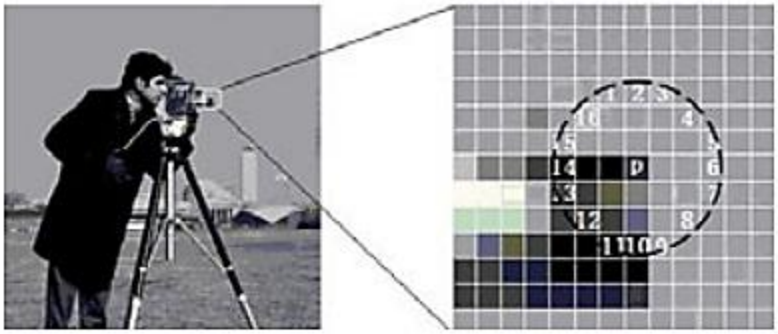


Figure 2.3: Candidate feature detection for FAST algorithm.

point based on image intensities in the local neighborhood [39]. Corners are detected as local maxima for the CRF function computed over the entire image. Following the detection, corner point matching is performed for each pair of frames. Sometimes a Bag-of-Words (BoW) algorithm is used to represent each image as a set of corner descriptors to speed up the matching process as in [27]. Then, homography matrices are computed and finally the images are projected into a common coordinate to get the final mosaic.

Choosing an optimal threshold is often a fundamental challenge of the FAST corner detector-based algorithms. However, it can be addressed by incorporating a robust threshold selection algorithm as in [53]. For matching the corner points from successive frames, they further proposed a threshold learning method together with a region-based gray correlation. Another major issue of the FAST-based algorithms is that they are not particularly robust to increased degree of variations. For that, extending the sampling area beyond the sixteen pixels around each candidate point [52] could be considered as a promising approach, since it gives the FAST corner points more distinctiveness and, in turn, makes them invariant to larger variations.

2.2.1.5 SIFT feature detector-based mosaicing

SIFT algorithm is a low-level feature detection algorithm which detects distinctive features (also called “keypoints”) from images. The SIFT descriptor is invariant to translations, rotations and scaling transformations in the image domain and robust to moderate perspective transformations and illumination variations. SIFT’s operation is based on five major steps: scale-space construction, scale-space extrema detection, keypoint localization, orientation assignment, and defining keypoint descriptors. Initially, a scale space is constructed by convolving an image repeatedly using a Gaussian filter with changing scales and grouping the outputs into octaves as [54]:

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \tag{2.10}$$

where $*$ is the convolution operator, $G(x,y,\sigma)$ is a Gaussian filter with variable scale σ, and $I(x,y)$ is the input image. After the scale space construction is complete, difference-of-Gaussian (DoG) images are computed from adjacent Gaussian-blurred images in each octave as [54]:

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma) \tag{2.11}$$

Following that, candidate keypoints are identified as local extrema of DoG images across the scales. The scale space and DoG scale space construction as well as extrema detection in DoG scale space is illustrated in Figure 2.4. In the next step, low contrast keypoints and edge response points along the edges are discarded using accurate keypoint localization. The keypoints are then assigned one or more orientations based on local image gradient directions as [54]:

$$\theta(x,y) = \tan^{-1}((L(x,y+1) - L(x,y-1)) / (L(x+1,y) - L(x-1,y))) \tag{2.12}$$

where $\theta(x,y)$ represents the gradient direction for $L(x,y,\sigma)$. A set of orientation histograms is formed over the neighborhoods of each keypoint. Finally, a normalized 128-dimensional vector is computed for each keypoint as its descriptor [13]. In order to find the initial matching keypoints

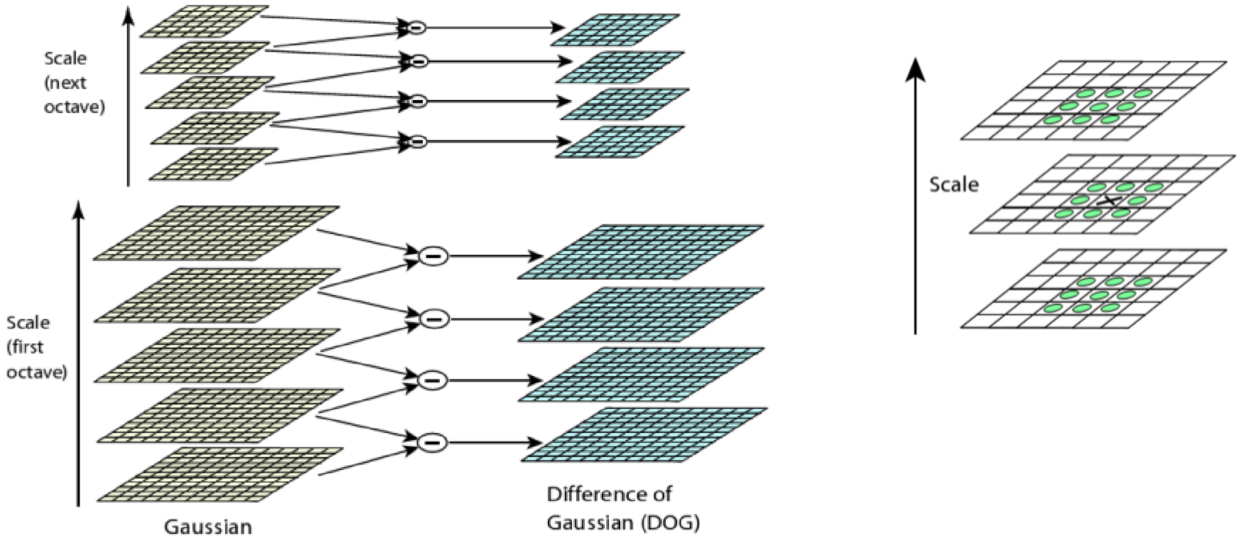


Figure 2.4: Scale space formation and extrema finding. (a) Scale space and DoG scale space construction; (b) Extrema detection in DoG scale space by looking into 26 neighbors.

from two images, nearest neighbor of a keypoint in the first image is identified from a database of keypoints for the second image [54, 55]. Following the initial matching, RANSAC algorithm is used to remove the outliers and to compute the transformation parameters between a pair of frames. Finally, images are warped using the transformation parameters and stitched to generate the mosaic image. SIFT based image mosaicing algorithms are particularly suitable for stitching high resolution images under variety of changes (rotation, scale, affine etc.), however, at the cost of high processing time.

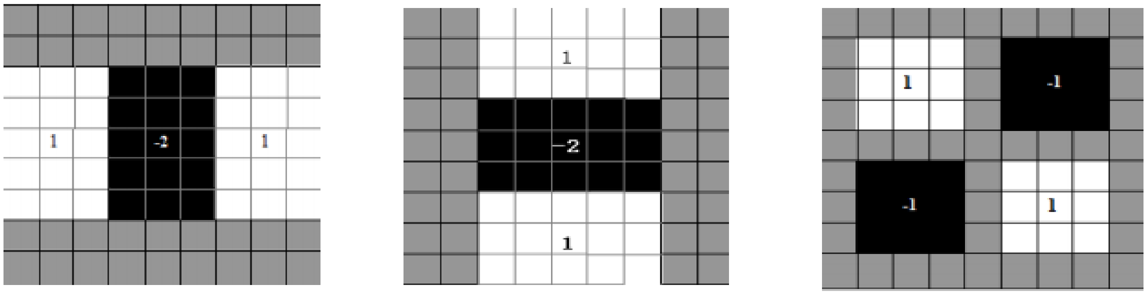
Several researchers have made variations to the above mentioned SIFT-based mosaicing method in order to further improve its performance. For example, in [12] the authors proposed switching between Kanade-Lucas-Tomasi (KLT) tracker and SIFT matching to find the correspondences between successive frames depending on their amount of overlapping. In [28], the author exploited a deformation vector propagation algorithm in the gradient domain to reduce the intensity discrepancy between the mosaiced images. Similarly, a bundle adjustment algorithm along with a modified-RANSAC algorithm capable of developing a probabilistic model is used in [56] to eliminate registration error and make the matching process more accurate.

2.2.1.6 SURF feature detector-based mosaicing

SURF algorithm is a scale and rotation invariant local feature detector. Like SIFT, this algorithm is also based on scale space theory. However, SURF uses Hessian matrix of the integral image to estimate local maxima across different scale spaces [57]. The Hessian matrix of an image I with scale σ at any point $X = (x, y)$ is defined as [58]:

$$H(X,\sigma) = \begin{pmatrix} L_{xx}(X,\sigma) & L_{xy}(X,\sigma) \\ L_{xy}(X,\sigma) & L_{yy}(X,\sigma) \end{pmatrix} \tag{2.13}$$

where $L_{xx}(X,\sigma)$, $L_{yy}(X,\sigma)$, and $L_{xy}(X,\sigma)$ are the convolutions of I in point X with Gaussian second order filters $\frac{\partial^2}{\partial x^2}G(\sigma)$, $\frac{\partial^2}{\partial y^2}G(\sigma)$, and $\frac{\partial^2}{\partial x \partial y}G(\sigma)$ respectively. While computing Hessian matrix at each pixel, the Gaussian filter operations are approximated by operations using box filters as shown in Figure 2.5. The response at each pixel is computed as the determinant of the Hessian matrix. Following that, a thresholding and a 3 x 3 x 3 local maxima detection window are used for non-maxima suppression. The local maxima are then interpolated in scale space to achieve keypoints with their location and scale values. In order to assign orientation for each keypoint, Haar-wavelet responses are calculated within a circular neighborhood around each keypoint. A vector is formed by summing up all the responses within 60-degree window. The longest vector is assigned as orientation to the keypoint. In order to assign descriptor vector to each keypoint, a square neighborhood region around the keypoint is selected. It is then split into smaller sub-regions. Sum of the Haar-wavelet responses from all the sub-regions are then used to generate a 64 dimensional descriptor vector [50]. After finding the matching keypoints from a pair of images, RANSAC algorithm is used to eliminate false matches as well as to calculate the homography matrices. Once homography matrices are achieved, images are warped and stitched to get the final mosaic. SURF based mosaicing techniques are faster than SIFT based techniques. However, they perform poorly under certain variations (particularly color, illumination, some affine transformation).



(a) (b) (c)

Figure 2.5: Approximation of Gaussian second order partial derivatives. (a) Approximation in x direction; (b) Approximation in y direction; (c) Approximation in xy direction.

The process of determining the SURF descriptors as mentioned above has sometimes been modified by some authors. For example, in [59] the local maxima is searched beyond a 3 x 3 x 3 neighborhood in the present scale and two immediately adjacent scales in order to make the feature descriptors more distinctive. In [60], the authors proposed dividing the SURF descriptor window into eight sub-regions while assigning descriptor vector. This technique increases the matching speed at the cost of increased false matches. However, the authors show that the use of RANSAC guarantees elimination of most of those incorrect matches.

Often multiple low-level feature extraction methods are used together in image mosaicing algorithms in order to use their respective benefits. Joshi et al. [61] propose a mosaicing algorithm which uses both Harris corner detector and SURF detector for extracting distinctive features from source images. Feature-based mosaicing algorithm proposed by Bind et al. [62] use both SIFT and SURF based feature detector to detect interest points from images. Kang et al. [63] and Zhu et al. [64] use Harris corner detector and SIFT detector in their feature-based mosaicing algorithm.

2.2.1.7 Contour-based mosaicing

This type of mosaicing algorithms is based on extraction of high-level features from images. Unlike the low-level features, these features are more natural to human perception and therefore they are high-level. High-level feature extraction mostly concerns finding the shapes or textures in an image. Shape extraction implies finding their position, orientation and their size [49]. Usually regions of different structures are extracted as high level image features. Then these features are matched to find correspondences, which are later used to compute the transformation parameters. Different techniques can be used to eliminate the false matches. Finally, warping and blending are performed to generate the mosaic output. The use of high-level features significantly increase the computation in these types of mosaicing algorithms. However, they are particularly suitable to work under larger and complicated motion parameters, and even under multi-layer registration.

Some of the notable contributions in high-level feature-based mosaicing include [65-67]. In [65], the authors used a wide baseline algorithm together with an adaptive region expansion method to achieve robust registration using high-level features. Prescott et al. [66] proposed extracting regions of image structures using a threshold technique and then computing area-based similarity matching for registration. Contour extraction using a segmentation algorithm, followed by finding their centroids for image registration was used in [67].

2.2.2 Frequency domain image mosaicing methods

Unlike spatial domain-based image mosaicing algorithms, methods classified in this category require computation in the frequency domain in order to find the optimal transformation parameters between a pair of images. These algorithms use the property of phase correlation for registering images. Let $I_1(x,y)$ and $I_2(x,y)$ are two images having some overlapping areas. Let's further assume that (x_0 , y_0) is the translation between the images. Thus,

$$I_2(x,y) = I_1(x-x_0, y-y_0) \tag{2.14}$$

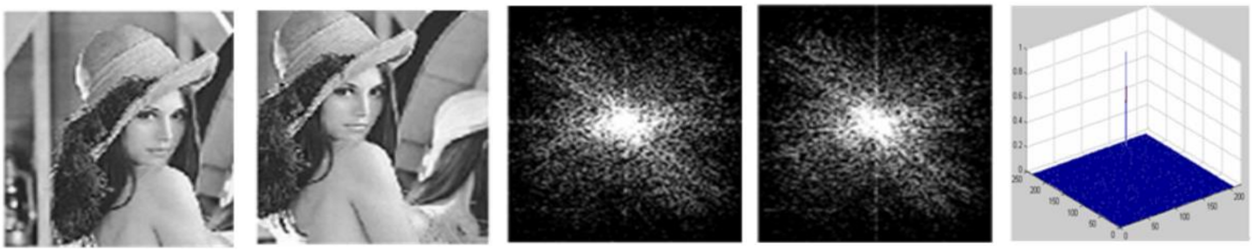
The corresponding Fourier transforms $F_1(u,v)$ and $F_2(u,v)$ are related by:

$$F_2(u,v) = F_1(u,v).e^{-j(ux_0+vy_0)} \tag{2.15}$$

The cross-power spectrum of the two images is defined as: [Ref 54]

$$\frac{F_1^*(u,v)F_2(u,v)}{\left|F_1^*(u,v)F_2(u,v)\right|} = e^{-j(ux_0+vy_0)} \tag{2.16}$$

where $F_1^*(u,v)$ is the complex conjugate of $F_1(u,v)$. The shift theorem guarantees that the phase of the cross-power spectrum is equivalent to the phase difference between the images. (x_0, y_0) could



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

Figure 2.6: Use of cross-power spectrum to detect transformation. (a), (b) Source images with displacement between them; (c), (d) Corresponding spectrum; (e) Impulse function indicating displacement between the images.

be solved in two different ways. One way is to work directly in frequency domain. However, this technique is very sensitive to noise. A better approach is to take inverse Fourier transform of the above equation and get an impulse function $\delta(x-x_0, y-y_0)$, which is approximately zero everywhere except at the displacement (x_0, y_0) as shown in Figure 2.6. With the displacement (translational) parameters the two images are warped and finally stitched to get a mosaic. Mosaicing algorithms based on this technique are usually efficient because of the use of shift property of Fourier transform and the use of Fast Fourier Transform (FFT). However, they suffer from being overly sensitive to noise. Additionally, accurate registration often requires significant overlapping between source images.

The above explained method of image mosaicing has sometimes, as in [30,68,69], been modified to make it suitable for handling transformations other than translation. A two-step method is proposed in [68]. The first step computes the rotation angle by finding the maximum peak by rotating the target image with an incremental angle. Using the computed rotation angle and phase correlation, the second step determines the translational displacement. A log-polar transformation is utilized in [30] to find the scale and translational parameters. In [69], the authors suggested changing the rotation and scale parameters to translational parameters using Fourier-Mellin transform.

The comparative overview of different categories of mosaicing algorithms based on image registration is presented in Table 2.1.

Table 2.1: Comparative overview of different categories of mosaicing methods based on image registration

Category	**Advantages**	**Disadvantages**
NCC-based	No high level structural analysis required, and can be applied directly to image data	Flat similarity due to self-similarity of images, and good only for images with large overlapping
MI-based	Good for multimodal analysis and less sensitive to illumination and occlusion changes	Slow and causes registration error when images have small overlapping
Harris corner detector-based	Simple and accurate computation	Needs prior knowledge of window size and good only for moderate changes in scale and rotation
FAST corner detector-based	Accurate and fast computation	Not robust to high degree of noise, and prior knowledge about threshold required
SIFT feature detector-based	Efficient for high resolution images and offers invariance to various transformations	Computationally expensive
SURF detector-based	Fast computation, good for real-time applications	Poor performance under certain transformations (e.g. color, illumination)
Contour-based	Efficient when large and complicated motion involved	Computationally expensive because of the use of high-level features
Frequency domain-based	Efficient because of FFT	Overly sensitive to noise and accuracy relies on large overlapping

2.3 Classification of image mosaicing methods based on blending

Similar to registration, image blending is also a significant step for successful implementation of mosaicing. Stitching multiple images together to create a seamless mosaic requires the use of a suitable blending algorithm. Blending is often referred to as photometric registration, which is vital to equalize color and luminance appearance in a composite image. There are several reasons (differences in camera exposure, scene illumination, or presence of moving objects between frames or geometric misalignments) which may lead to image inconsistencies in the final mosaic. The visibility of such inconsistencies can be minimized by choosing appropriate blending algorithm. This way, the final mosaic would be visibly free of annoying seams, giving it a more consistent global appearance. Figure 2.7 shows that based on the image blending, mosaicing algorithms can be transition smoothening-based and optimal seam-based. Transition smoothening-based mosaicing can be further grouped into feathering-based, pyramid-based, and gradient-based mosaicing. Different classes of image mosaicing algorithms based on the image blending methods are discussed below.

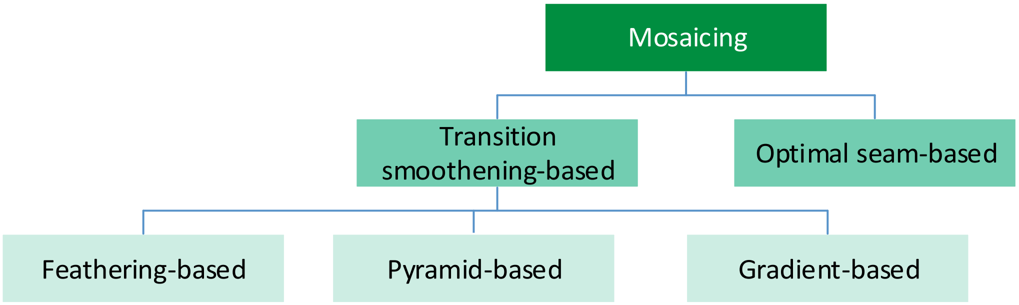


Figure 2.7: Classification of mosaicing based on blending.

2.3.1 Mosaicing methods using transition smoothing-based blending

Mosaicing algorithms within this category attempt to minimize the visibility of seams by smoothing the common overlapping regions of the combined images. The information of the overlapping region between two images is fused in such a way that the boundaries of the images involved become imperceptible. Even though a totally indistinguishable transition may be achieved, the content and coherency of the overlapping region is not guaranteed, as the information is fused without taking into account the content of the scene [70]. Thus, most often, these mosaicing methods generate mosaic with blurry transitions in the boundary regions. Popular methods which use transition smoothing for their blending operation include feathering, pyramid blending, and gradient-based blending. Mosaicing algorithms based on these techniques are discussed briefly as follows:

2.3.1.1 Mosaicing algorithms using feathering-based blending

Mosaicing algorithms within this category perform blending operation by taking an average value in each pixel of the overlapping region. However, the simple average method fails when exposure differences, misalignments, and presence of moving object are very obvious in the input images. A better approach to the averaging method is to use weighted averaging along with a distance map. Pixels near the center of an image are weighted heavily and those near the edges are weighted lightly. This is done by computing a distance map in terms of Euclidean distance of each valid pixel (mask) from its nearest invalid pixel as [43].

$$w_k(x) = \left\| \arg\min_{y} \{ \|yy\| \left| \tilde{I}_k(x+y) \ \textit{is invalid} \right\} \right\| \tag{2.17}$$

where $\tilde{I}_k(x)$ are the warped images and $w_k(x)$ are the weights of the images. Finally, the mosaic image is generated as a weighted combination of the input images. Examples of composite images formed of six color images using simple average blending and feathering are shown in Figure 2.8. Mosaicing algorithms which use the aforementioned technique perform reasonably well under

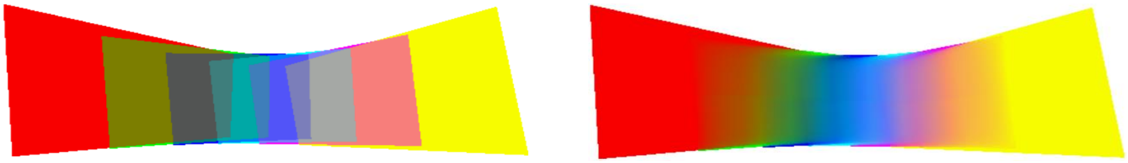

(a) (b)

Figure 2.8: Image blending results. (a) Blending using simple averaging; (b) Blending using feathering.

exposure differences. However, it is difficult in practice to achieve a balance between smoothing out low-frequency exposure differences and preserving sharp enough transitions to prevent blurring. Furthermore, these methods suffer from ghosting effect.

Examples of mosaicing methods using feathering-based blending include [56] [71] and [60]. [56] and [71] used alterations of the above mentioned method for finding the weights of images in the overlapping region. In [56]. the aforementioned weight is measured by computing the distance of the overlapping pixels from the borders of the left and the right images. In [71], the authors used weighted average of the pixel color values in the overlapping region.

2.3.1.2 Mosaicing algorithms using pyramid-based blending

In an attempt to perform the blending operation in a more robust way, these mosaicing algorithms convert the input images into band-pass pyramids as shown in Figure 2.9. Mask image associated with each source image is then created. Mask creation can be made automatic by using grassfire transform as used in [72]. Then the mask image is converted into a low-pass pyramid by using a Gaussian kernel [43]. The resultant blurred and subsampled masks are treated as weights to perform per-level feathering. The final mosaic is then achieved by interpolating and summing the results from per-level feathering as:


(a) (b)

Figure 2.9: Pyramid formation for blending. (a) Low-pass pyramid; (b) Band-pass pyramid.

$$LO(x,y) = GM(x,y) * LI_1(x,y) + (1 - GM(x,y)) * LI_2(x,y) \tag{2.18}$$

where $LI_1(x,y)$ and $LI_2(x,y)$ are the Laplacian pyramids of the warped source images I_1 and I_2. $GM(x,y)$ is the Gaussian pyramid of the mask image $M(x,y)$ and $LO(x,y)$ is the Laplacian pyramid of the output image $O(x,y)$. Sometimes, all the strips are combined in a single blending step when it needs building pyramids for multiple narrow strips as proposed in [31]. Algorithms using the above method achieve reasonable balance between smoothing out low frequency components and preserving sharp enough transitions to prevent blurring [74]. Edge duplication is also eliminated noticeably. However, double contouring and ghosting effects become significant when the registration error is significant.

2.3.1.3 Mosaicing algorithms using gradient-based blending

Another group of transition smoothening method are those based on gradient domain blending. These methods are based on the idea that by suitably mixing the gradient of images, it is possible to mosaic image regions convincingly. In general the gradients across seams are set to zero for smoothing out the color differences. Since humans are more sensitive to gradients than image intensities, mosaicing methods using this technique generate visually more pleasant results compared to the other two techniques discussed before. However, working exclusively in the gradient domain requires higher computational resources to deal with large data sets. Furthermore, for best performance, the alignment of images through registration needs to be almost perfect.

Notable work in this group was developed by [32], [75], and [76]. In [75], the authors used a gradient domain object moving and region filling algorithm to eliminate the visible artifacts arising from moving objects in the scene. Algorithm based on assigning low resolution offset map to each source image followed my Poisson’s blending was proposed by Szeliski et al. [76]. In [32], the authors developed two approaches called GIST (gradient domain image stitching). One of the approaches is based on minimizing a cost function that evaluates the dissimilarity measure between the derivatives of the mosaic and the derivatives of the source images. The other approach is based on inferring a mosaic by optimization over image gradients.

2.3.2 Mosaicing methods using optimal seam-based blending

This type of mosaicing algorithms attempt to minimize the visibility of seams by looking for optimal seams in the joining boundaries between the images. The objective of optimal seam technique is to allocate the optimal location of a seam line by looking into the overlapping region between a pair of images. The seam line placement should be such that it minimizes the photometric differences between the two sides of the line. At the same time the seam line should be able to determine the contribution of each of the images in the final mosaic. Once the placement and the contribution information are obtained, each image is copied to the corresponding side of the seam. When the difference between the two images on the seam line is zero, no seam gradients are produced in the mosaic. Unlike the mosaicing methods using transition smoothing-based blending, optimal seam-based mosaicing algorithms consider the information content of the scene in the overlapping region, allowing to deal with problems like moving objects or parallax. However, no information is fused in the overlapping region, thus the transition between the images can be easily noticeable when there are global intensity or exposure difference between the frames.

Different optimal seam finding methods have been used in mosaicing literature. For example, in [33] a modified region-of-difference method is used.[77] proposed the use of an algorithm based on watershed segmentation and graph cut optimization. Another method based on dynamic programming and grey relational analysis is used in [78].

A general comparison of different categories of mosaicing algorithms based on image blending is presented in Table 2.2.

Table 2.2: Comparative overview of different categories of mosaicing methods based on image blending

Category	**Advantages**	**Disadvantages**
Feathering-based	Fast and good performer under exposure differences	Output often suffer from blur and ghosting effect
Pyramid-based	Good in preventing blur and edge duplication	Suffers from double contouring and ghosting when registration error significant
Gradient-based	Output visually more appealing than other methods	High computation required and registration error must be small for good performance
Optimal seam-based	Good in dealing with moving objects an parallax	Transition obvious when there are exposure differences

CHAPTER 3

STATE-OF-THE-ART OF IMAGE SUPER-RESOLUTION METHODS

3.1 Introduction

Super-Resolution (SR) is the process of achieving a high-resolution (HR) image from a single low-resolution (LR) observation or a sequence of LR observations of a scene taken at different viewpoints. It aims to overcome the limitations of the image capturing devices to produce a high resolution image. In SR context, HR means higher spatial resolution and hence higher information content. HR images are not only visually appealing, but also valuable in several practical applications for extracting additional details. SR has been an active research area over the last two decades and most recently it is gaining growing interests in the image processing community for its potential derivatives. Application areas of SR include but not limited to satellite imaging [79, 80], astronomical image processing [81], medical image processing [82-84], HDR imaging [85], automatic image mosaicing [13], fingerprint and face image enhancement [14], target recognition [86], video surveillance [87], and converting video standards [88].

The sensor size and the density of detectors that form the sensor primarily determines the spatial resolution of the captured images. Larger the size of the sensor, and/ or higher the density of the detectors, better the spatial resolution of the acquired images. The most direct hardware-based approach of increasing the spatial resolution is to reduce the detector size or equivalently increase the detector density. Alternatively, the sensor size can also be increased. However, smaller detectors have lower dynamic range, lower fill factor, worse low light sensitivity, higher dark signal, higher diffraction sensitivity, and higher non-uniformity [89]. Also, the hardware cost increases with the increase of both detector density as well as sensor size. Thus, the above mentioned hardware-based approaches often restricts the maximum achievable resolution of the captured images. Besides the sensor-imposed restriction, there are several other factors that limits the capture of high resolution images, for example lens and atmospheric blurs, finite shutter speed, finite aperture, movement of objects in the scene, sensor noise, and media turbulence. Consequently, a software-based approach (like SR) to obtain images with improved spatial resolution from one or more LR observations becomes an attractive proposition [13].

Single-frame SR increases the spatial resolution by utilizing one or more learning models. In contrast, multi-frame SR increases the spatial resolution by taking advantage of more samples than that found in any single LR observation. Thus, each LR observation must exhibit either sub-pixel shift, or change in illumination, or variation in blur from the other. The physical size of the SR output may be same as the size of one of the LR observations or larger depending on the image interpolation method used [90]. Two closely related techniques of SR are interpolation and restoration. Image interpolation increases the size of an image, however, it does not improve the quality of it. Image restoration, on the other hand, improves (by deblurring and denoising) the quality of an image without changing its physical dimensions. Thus SR must not be confused with either interpolation or restoration, rather it could be seen as a combination of these two techniques. A multi-frame SR technique is illustrated in Figure 3.1.

Interpolated image using single frame

SR output using 5 frames

Figure 3.1: A framework of multi-frame super-resolution.

Being an attractive research area, SR has resulted in the development of numerous algorithms. Thus, it would be extremely difficult for someone interested in this research area to select a suitable method without having a comprehensive survey. In this chapter, we classify the past and the newly emerging SR techniques into several categories. The basics of all the categories are discussed. Furthermore, the improvements over the basic methods made by different researchers are also highlighted. However, before going into the detailed classification, we will discuss an image observation model which is used by almost all reconstruction-based SR methods.

3.2 Image observation model

The first strategic step to understand SR imaging is to formulate an observation model that establishes the relationship between desired HR image and a set of LR images. During the acquisition process, the captured scene undergoes a series of transformations to generate the LR images. For simplicity in the formulation of the observation model, these transformations are limited to the following four operations: 1) geometric transformation, 2) blurring, 3) down-sampling, and 4) addition of white Gaussian noise. Geometric transformation includes global or local translation, rotation, and scaling that are responsible for scene motion. Since these information are usually unknown, a warp operator can be modeled that can estimate the scene motion for each image with reference to one particular image. Blur includes any blurring effect for example optical blur (related to lens and/or sensor), motion blur, atmospheric blur etc. For reconstruction-based SR methods, the characteristics of the blur are assumed to be known. Hence blurs are usually modeled as a point spread function (PSF) kernel. Different downsampling operators can be used to generate LR images of different size. However, for simplicity we would restrict the observation model to generate LR images of same size. Furthermore, we would consider the down-sampling factors for the vertical and horizontal directions to be equal.

To formulate the model, let's assume that x is the desired HR image of size $N_1 \times N_2$, which is derived from a bandlimited continuous scene. Considering q to be the down-sampling factor in both directions, each of the K LR images $(y_k, k = 1\ to\ K)$ is of size $M_1 \times M_2$, where $N_1 = qM_1$ and $N_2 = qM_2$. If the LR images are generated by warping, blurring, down-sampling, and addition of white Gaussian noise to the HR image x, we can represent the observation model as:

$$\begin{aligned} y_k &= DB_kW_kx + n_k \qquad for\ k = 1\ to\ K \\ &= M_kx + n_k \end{aligned} \tag{3.1}$$

where both y_k and x are represented in lexicographically ordered vectors having a size of $M_1M_2 \times 1$ and $N_1N_2 \times 1$, respectively. $D, B_k,$ and W_k are the decimation operator, blur operator, and the warp operator expressed in matrix form. M_k is the matrix which represents all the above mentioned degradation factors. Figure 3.2 shows a graphical representation of the observation model of Eq. (1). Alternation in the order of blur and warp operators in the above mentioned observation model is investigated in [103, 104]. It is explained that when the blur operation is space invariant and motion among the LR images is pure translational, the two aforementioned operators are eligible to commute. However, when the blur operation is space variant, it is more appropriate to use the blur operator first and then the warp operator unlike the one in Eq. (1).

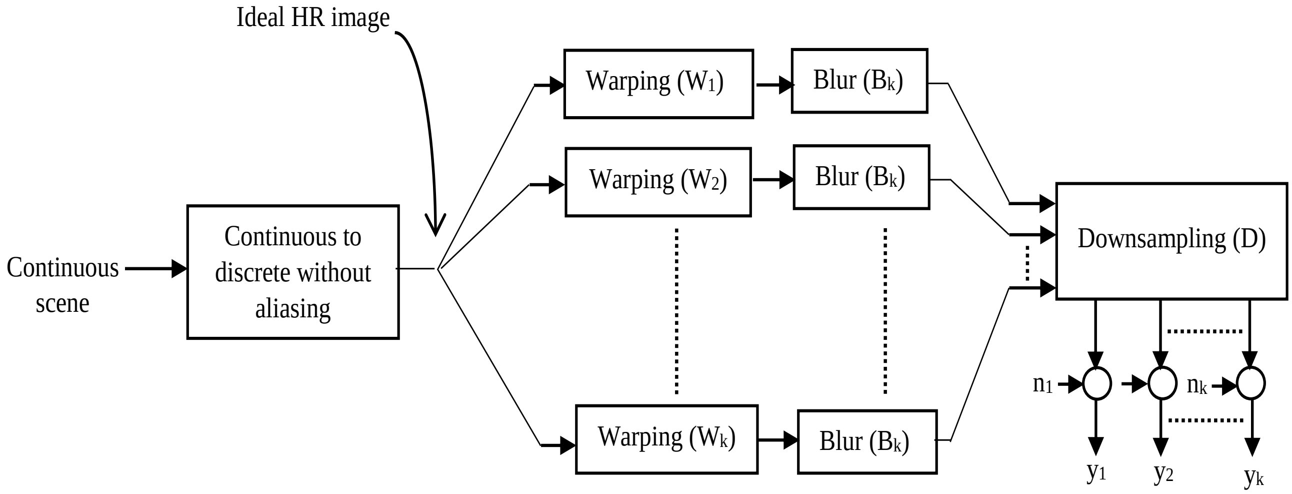


Figure 3.2: Observation model relating LR image with HR image.

The classical observation model on Eq. (1) has been modified by many researchers. In [105], two nonstationary observation models related to the quantization noise are used in the presence of compression. Zhang et al. [106] modifies the observation model to use it with hyperspectral data. In [107] the pointwise interpolation of the conventional observation model is replaced by a technique based on L2 function approximation. [108, 109] propose modification of the model in Eq. (1) by including different zooming in the LR images. [110] Suggest addition of three separate PSF functions for sensor blur, lens blur, and motion blur. The explicit motion parameter of the observation model is replaced by probabilistic motions in [111].

3.3 Classification of image super-resolution methods

SR algorithms can be classified based on several factors, for e.g. number of input LR frames, domains of operation, operating principles, etc. While majority of the previous surveys classified the existing SR methods based on either of these aforementioned factors, it is obvious that these surveys are not comprehensive. In this present survey, the SR algorithms are first classified based on their number of input frames, i.e. multi-frame and single-frame SR techniques. In terms of domain of operation, multi-frame SR algorithms are then grouped into spatial domain, and frequency domain-based methods. Note that almost all of the single-frame SR algorithms are based on spatial domain, thus the classification of the algorithms based on domain of operation is only presented for the multi-frame SR techniques. Single-frame SR algorithms can be classified in terms of their operating principles. The detailed taxonomy used in this survey is shown in Figure 3.3.

3.3.1 Multi-frame super-resolution methods

As mentioned earlier, the foundation of multi-frame SR is the availability of multiple LR images of the same scene captured at different view-points. Assuming the LR images are aliased versions of the desired HR scene, there must be subpixel shift present between each pair of the LR images. When the LR images are shifted by integer units, then there exists no additional information that could be exploited to reconstruct the HR image. If the relative scene motion can

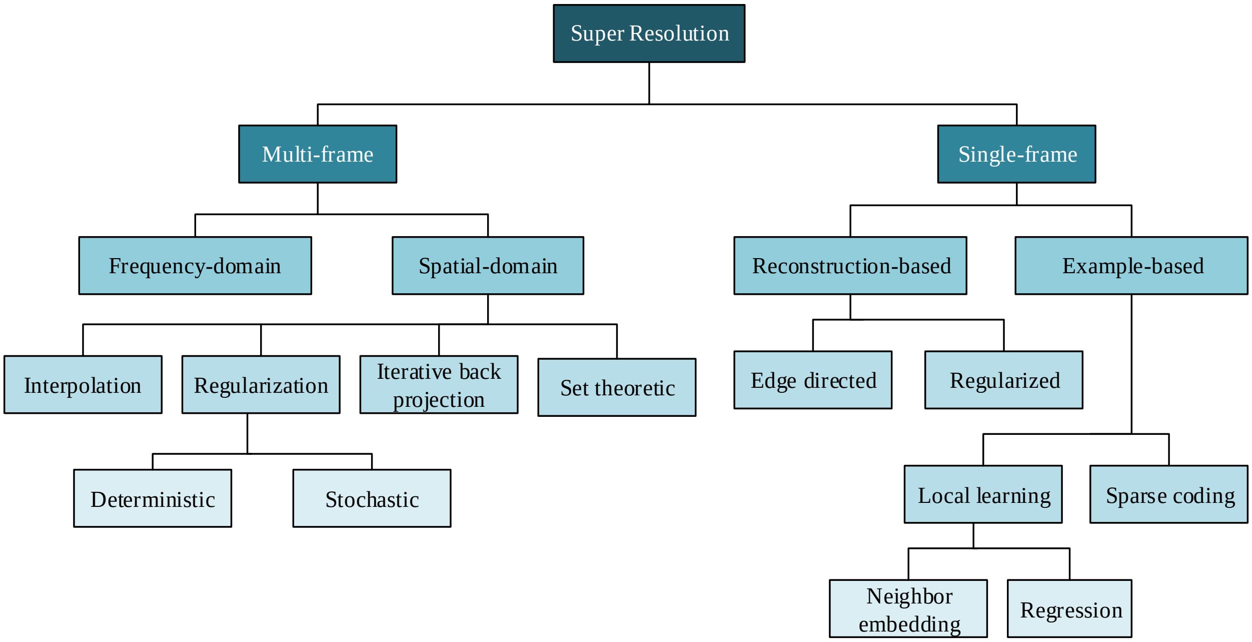


Figure 3.3: Taxonomy of super-resolution algorithms.

be estimated with subpixel precision from different subpixel shifted LR images, it is possible to combine them for SR reconstruction. Based on domain representation, multi-frame SR algorithms can be classified into two classes: frequency-domain SR algorithms and spatial-domain SR algorithms. Though the initial SR work was performed in frequency-domain, majority of the present-day SR research is concentrated in spatial domain because of several advantages including unconstrained inter-frame motion, simplicity in incorporating prior information into the SR solution. Spatial-domain algorithms can be grouped into six categories: interpolation-based methods, deterministic regularization methods, stochastic regularization methods, set theoretic methods, and iterative back projection methods. All the above-mentioned multi-frame SR algorithms are discussed in the following subsections.

3.3.1.1 Frequency-domain super-resolution methods

These algorithms compute the desired HR image by combining multiple subpixel shifted LR images based on shift and aliasing properties of continuous and discrete Fourier transform as proposed in the seminal work by [112]. Let $c(x, y)$ be the continuous HR scene seen by the camera and $c_k(x, y)$ be the kth globally translated scene obtained from $c(x, y)$. Thus $c_k(x, y) = c(x + \Delta x_k, y + \Delta y_k)$, where $(\Delta x_k, \Delta y_k)$ denotes the translation parameter for the kth scene. Let $C(u, v)$ and $C_k(u, v)$ be the continuous Fourier transform (CFT) of the HR scene and the kth translated scene, respectively. Using shifting property of the CFT, $C(u, v)$ and $C_k(u, v)$ can be related as:

$$C_k(u, v) = e^{j2\pi(\Delta x_k u + \Delta y_k v)} C(u, v) \tag{3.2}$$

The translated scenes are sampled with sampling frequencies $1/T_m$ and $1/T_n$ to obtain the LR images, $d_k(m, n) = c_k(mT_m + \Delta x_k, nT_n + \Delta y_k)$ with $m = 0, 1, 2, \ldots\ldots, M - 1$ and $n = 0, 1, 2, \ldots\ldots, N - 1$, where (M, N) is the dimension of each LR image. Let $D_k(u, v)$ be the discrete Fourier transform (DFT) of the kth LR image. Using aliasing property of the DFT,$D_k(u, v)$ and $C_k(u, v)$ are related as:

$$D_k(u, v) = \frac{1}{T_m T_n} \sum_{p=-\infty}^{+\infty} \sum_{q=-\infty}^{+\infty} C_k\left(\frac{2\pi}{T_m}\left(\frac{u}{M} - p\right), (\frac{2\pi}{T_n}\left(\frac{v}{N} - q\right)\right) \tag{3.3}$$

If $c(x, y)$ is assumed to be bandlimited, equation 3.1 and 3.2 can be combined to obtain the relationship between the DFT of the LR images and the CFT of the HR scene as:

$$D = \emptyset C \tag{3.4}$$

where D is a Kx1 column vector (with K be the number of LR images), C is a MNx1 column vector, and $\emptyset$ is a KxMN matrix. Unknown C can be solved using the above equation.

The above framework of frequency-domain SR is extended by several researchers [37-40] in an attempt to make it more acceptable to real world applications by incorporating noise and blur models into the SR formulation. To achieve registration and restoration simultaneously, authors in [117, 118] propose the use of EM algorithm. In [119], the authors use DCT instead of DFT to make the SR computation faster. Vandewalle et al. [120] suggest the use of low frequency (aliasing free) image components only while generating super-resolved image. In spite of all these attempts, frequency-domain-based SR algorithms still remain inappropriate choice for several limitations including extreme sensitivity to model errors, strict requirement of pure translational motion, and linear space invariant blur during image acquisition. Furthermore, the addition of prior information in order to regularize the ill-posed SR problem is considerably difficult using these algorithms.

As an alternative to the above mentioned Fourier transform-based methods, wavelet transform-based algorithms [121-129] have gained much popularity within the family of frequency-domain SR algorithms. Multiresolution analysis in discrete wavelet transform (DWT) generates different frequency sub-bands of a given LR image. Out of these, three high-frequency sub-bands along with the LR image are interpolated and later combined to obtain the HR output by using inverse-DWT. In [122, 123], the authors propose incorporating stationary wavelet transform coefficients to improve the SR result. Demirel et al. [125] use a dual tree complex wavelet transform (DT-CWT) instead of DWT. Their later work [126] extends the work further by introducing an edge directional interpolation following the DT-CWT step. In order to utilize the ease of implementing deconvolution filter using Fourier transform, authors in [127, 128] propose a combined Fourier-wavelet transform.

3.3.1.2 Interpolation-based super-resolution methods

These methods [130-136] construct a HR image from the captured LR images by using a three-step approach: registration, interpolation, and restoration as shown in Figure 3.4. Registration generates a motion compensated single, dense composite image of non-uniformly distributed samples. Non-uniform interpolation produces uniformly spaced samples in the HR image grid. Finally, restoration compensates for various degradations including blur and noise, which are introduced during the image acquisition. Note that single image interpolation cannot reconstruct the lost high frequency components in an image, hence it cannot handle the SR problem.

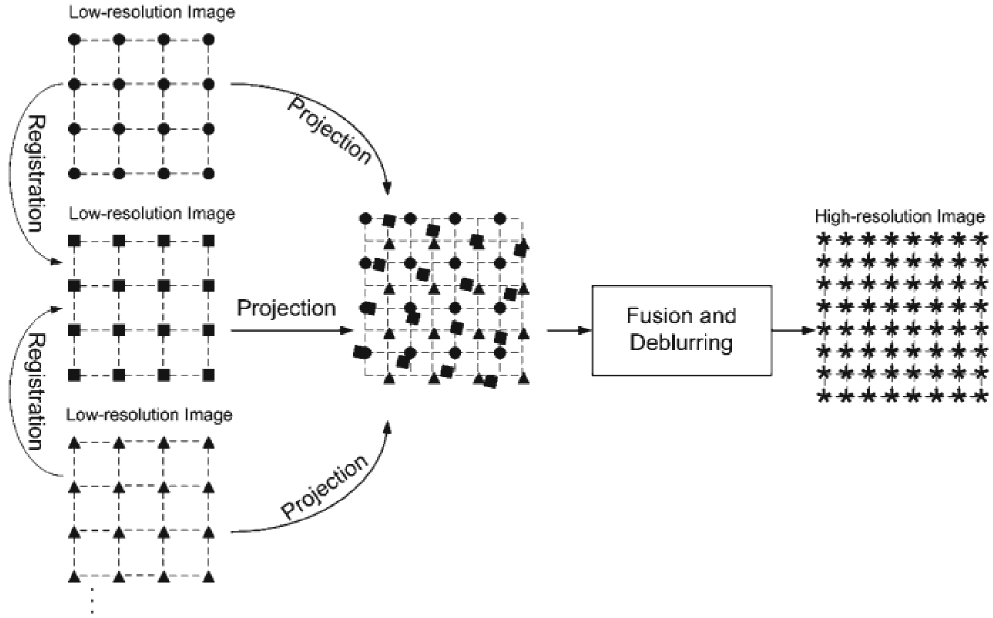


Figure 3.4: Interpolation-based super-resolution reconstruction.

Zhong et al. [130] propose a biharmonic spline interpolation along with frequency domain-based registration for super resolving a sequence of X-Ray images. Based on the idea of generalized multi-channel sampling theorem, Ur et al. [131] perform a non-uniform interpolation on a set of spatially shifted LR images, followed by a deblurring process. Panagiotopoulou et al [132] use a geostatistical interpolation method based on weighted combination of the neighbors, followed by Wiener filtering for deblurring and denoising. In [133], authors present a generalized bilinear interpolation, followed by a deblurring scheme based on back propagation neural networks (BPNN). Addressing the spatial structure information lost while rounding subpixel displacement in the HR grid, authors in [134] propose an interpolation technique based on multisurface the fitting. Gilman et al. [135] propose a least-square optimal interpolation method based on optimizing the resampling filter coefficients using a simple image model in a least square fashion. Hardie [136] suggests a SR algorithm using adaptive Wiener filtering, which combines the non-uniform interpolation and the restoration steps into a single weighted sum operation. Computational simplicity is the advantage of these methods. However, they have several disadvantages including limited degradation models, uncertain optimality, and lack of prior constraints.

3.3.1.3 Deterministic regularization-based super-resolution methods

These methods rely on the fact that by carefully estimating the registration parameters, the observation model can be completely specified [137]. The inverse SR problem is then solved by using prior information about the solution. Among many techniques to impose prior knowledge on the solution space, one common approach is to use the constrained least square (CLS) optimization function. Deterministic methods then seek a super-resolved image $\hat{x}$, which minimizes the following cost functional:

$$\hat{x} = \arg min_x \; [\textstyle\sum_{k=1}^{K} \|y_k - M_k x\|_2^2 + \lambda \|Lx\|_2^2] \tag{3.5}$$

where the operator L is usually a high-pass filter, M_k is the degradation matrix for the LR image y_k. Note that since images tend to be smooth in nature with limited high-frequency components, it is appropriate to implement the prior knowledge as a smoothness constraint as above. The Lagrange multiplier λ is called regularization parameter, which controls the balance between the data fidelity term and the smoothness constraint. When the fidelity of data is low (number of non-redundant low-resolution frames < square of resolution enhancement factor), higher value of λ is used. Conversely, for higher fidelity of data, smaller value of λ leads to good solution. Unique estimate of $\hat{x}$ can be achieved by solving the above optimization problem using any deterministic iterative technique. Using gradient descent optimization, the iterative solution can be expressed as:

$$\hat{x}^{(n+1)} = \hat{x}^n + \beta^{(n)} \, [\textstyle\sum_{k=1}^{K} M_k^T (y_k - M_k \hat{x}^n) - \lambda^n L^T L \hat{x}^n] \tag{3.6}$$

β represents the step size of the gradient descent method. Primary advantage of a deterministic method is that use of a convex and differentiable cost functional guarantees a unique estimate of SR image.

3.3.1.4 Stochastic regularization-based super-resolution methods

Stochastic regularization methods, which treat SR reconstruction as a statistical estimation problem provide a flexible way for the inclusion of prior knowledge necessary for satisfactory solution of the ill-posed SR problem. From the Bayesian perspective, the information that can be extracted from the LR observations about the unknown HR image is contained in the probability distribution of the unknown [92]. Thus, the unknown HR image can be estimated via some statistics of the probability distribution of the unknown HR image. The maximum a posterior (MAP) estimator of x seeks the estimate $\hat{x}$ MAP for which the probability $p(x|y_1, y_2, \ldots., y_k)$ is maximum as:

$$\hat{x}_{MAP} = arg\ max_x\ p(x|y_1, y_2, \ldots., y_k) \tag{3.7}$$

By applying Bayes's rule and taking logarithmic function, the above expression can be re-written as:

$$\hat{x}_{MAP} = arg\ max_x((log\ p(y_1, y_2, \ldots., y_k|x) + log\ p(x)) \tag{3.8}$$

The term $log\ p(y_1, y_2, \ldots., y_k|x)$ is the log-likelihood function and $p(x)$ is the priori density of x. Since the LR images are independent of each other, the above equation becomes:

$$\hat{x}_{MAP} = arg\ max_x(\sum_{k=1}^{K} log\ p(y_k|x) + log\ p(x)) \tag{3.9}$$

Since image is treated as a locally smooth data field, it is common to utilize MRF as prior image model. The priori density can then be written in Gibbs form as:

$$p(x) = \frac{1}{C}\exp\big(-E(x)\big) = \frac{1}{C}\exp(-\sum_{r\in R} P_r(x)) \tag{3.10}$$

where C is a normalizing constant, $E(x)$ is the priori energy function, $P_r(x)$ is the potential energy function that depends on local group of points r, which are called cliques. R is the set of all the cliques. Using MRF prior, the potential energy function is expressed as derivative of the image x.

If the noise is assumed to be an independent identically distributed zero mean white Gaussian noise, the MAP solution becomes

$$\hat{x}_{MAP} = arg\ min_x(\sum_{k=1}^{K} \|y_k - M_k x\|_2^2 + \lambda \sum_{r\in R} P_r(x)) \tag{3.11}$$

where λ is the regularization parameter. When Gaussian prior is used, $P_r(x)$ takes the quadratic form and in the simplest case (Tikhonov regularization [138] the above solution becomes identical to that of the deterministic estimation an in Eq. (5). Tikhonov regularization penalizes the high frequency components severely, resulting in overly smooth final solution. To preserve sharp edges and other discontinuities in the HR image, Huber MRF (HMRF) prior is alternatively used in literature [139-141], which has the following form for the potential energy function:

$$P_r(x) = \begin{cases} x^2, & x \leq \alpha \\ 2\alpha|x| - \alpha^2, & Otherwise \geq 0 \end{cases} \tag{3.12}$$

where α is a threshold parameter which separates quadratic and linear regions.

When the above mentioned clique operation in the MRF solution is approximated over four directional image derivatives, the prior is well known as Gaussian MRF (GMRF) or directional HMRF [13, 142], which uses the following priori energy function:

$$E(x) = \sum_{i=1}^{4} P_r(d_i x) \tag{3.13}$$

When α in the HMRF formulation tends to zero, the prior becomes total variation (TV) prior [143-146] which uses Norn 1 of the image gradient operator as:

$$p(x) = |\nabla x|_1 \tag{3.14}$$

Besides the aforementioned prior models, several other priors have been exploited in the literature, such as, natural image prior [147], soft edge smoothness prior [148], conditional random field (CRF) prior [149], discontinuity adaptive MRF (DAMRF) prior [150], bilateral total variation (BTV) prior [151], principal component analysis based prior [152]. Other than defining the appropriate prior model, another fundamental issue that need to be addressed for these regularized SR approaches is determining the best optimization method to solve the SR problem. Choosing the best optimization often depends on the nature of the regularized cost functional. With convex cost functional, gradient based optimization methods like gradient descent, conjugate gradient are used. Whereas for non-convex cost functional, Bregman method [153, 154], expectation maximization [155], Markov Chain Monte Carlo [147] are used.

Another Bayesian-based approach is maximum likelihood (ML) estimation, which can be derived from the MAP estimation by eliminating the prior term. Due to the ill-posed nature of SR inverse problems, MAP estimation is usually preferred to ML. Robustness and flexibility in modeling noise characteristics and prior knowledge of the solution are the major advantages of the stochastic methods [92]. MAP estimation with convex prior guarantees the uniqueness of the solution. Joint motion estimation and restoration is also possible [156].

3.3.1.5 Set theoretic super-resolution methods

Set theoretic methods [157-165], especially the projection onto convex sets (POCS), utilize an iterative approach for convenient inclusion of prior knowledge into the SR reconstruction process. The desirable SR image characteristics are associated with a set of convex constraint sets in the solution space. Commonly used convex sets include data consistency, bounded energy, positivity, amplitude constraint, and smoothness constraint. Data consistency is modeled as K convex sets as [166]:

$$C_k = \{x| \ \|y_k - M_k x\|^2 \leq \delta^2, 1 \leq k \leq K\} \tag{3.15}$$

where δ is the upper bound of the uncertainty of the model. Similarly, amplitude constraint can be modeled as [159]:

$$C_A = \left\{x \middle| B \leq x_p \leq A, \forall p = 1, \dots ., M\right\} \tag{3.16}$$

where M denotes the number of pixels in x and A and B are the upper and lower bound for the pixel amplitude. For each convex set C_i , a projection operator P_i is defined. Projection operator associated with a particular set projects a point in SR space onto a point in the space of the set. With a group of m convex sets, the space of the SR solution lies in the intersection of these m sets. Using an iterative method, POCS seeks a point within the intersection set (that comply with all the convex constraint sets) given an initial estimate of the unknown HR image as:

$$x^{(n+1)} = P_1 P_2 P_3 \ldots . P_m x^n \tag{3.17}$$

Gevrekci et al. [161] propose a POCS-based restoration algorithm using a constraint set utilizing spatio-intensity neighborhood. In order to address the spatial information loss in the captured remote sensing images, authors in [160] suggest a global weighted POC method. In their method the residual error and gray scale are used as constraint sets during the iterative reconstruction process. To reduce the effect of inaccurate estimation of sub-pixel motion in the SR result, Caner, Tekalp, and Heinzlman [162] propose registration of the multi-views of a dynamic scene. In [163], the authors propose using an adaptive regularization based on noise variance within the POCS framework in order to incorporate the effect of noise energy. The constraint sets defined on edges are modified in [164] to reduce the amount of edge ringing present in the POCS HR estimate. Addressing the issue that traditional constraints yield slow convergence of the POCS method, Brodzik and Mooney [165] suggest a SVD-based constraint consistent with image spectral characteristics.

Advantages of these methods include simplicity in incorporating prior information. However, they have disadvantages like nonuniqueness of the solution because of the variable initial guess, slow convergence, high computation cost. Furthermore, results of these methods will be erroneous if the estimated sub-pixel motion is inaccurate.

3.3.1.6 Iterative back projection-based super-resolution methods

First proposed in [2], IBP methods [167-174] update the estimate of the SR reconstruction by backprojecting the error between simulated LR images and the observed LR images. Having defined the image observation model as in equation 3.5, these methods start by an initial guess for the desired HR image. Such a guess can be obtained by registering the LR images and then interpolating them in the HR grid. A set of K simulated LR images $\{\hat{y}_k, k = 1\ to\ K\}$ is obtained from this initial guess. The residual error between the observed and simulated LR images is then iteratively backprojected to the initial guess until a minimum error-threshold or maximum iteration-number is reached. The residual error (e_r^t) in the tth iteration is given by:

$$e_r^t = \sqrt{\sum_{k=1}^{K} \left\| y_k - \hat{y}_k^t \right\|_2^2} \tag{3.18}$$

The iterative process to obtain HR estimate can be expressed by:

$$x^{(t+1)} = x^t + \frac{1}{K} \sum_{k=1}^{K} (y_k - \hat{y}_k^t) \times H_{BP}^k \tag{3.19}$$

where H_{BP}^k is the backprojection operator that creates a projection from the kth difference images. H_{BP}^k can be modeled as M_k^{-1}. In that case the backprojection operator becomes a combination of inverse warping, upsampling, and deblurring operators. The graphical representation of the IBP method is shown is Figure 3.5.

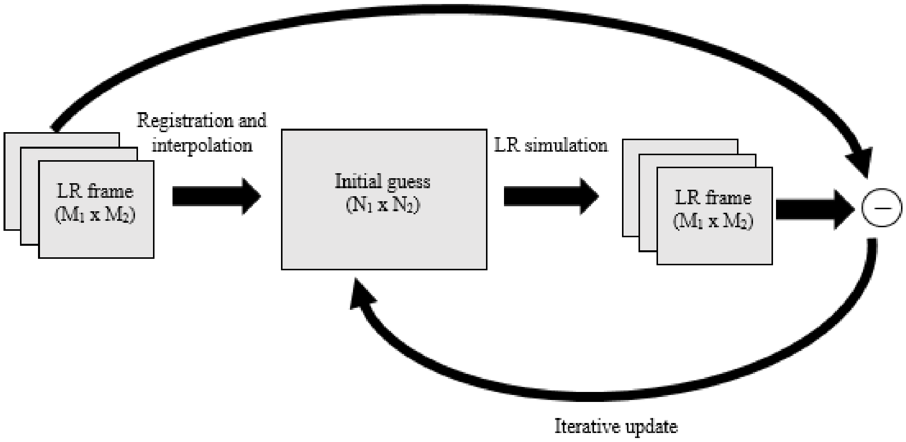


Figure 3.5: Iterative back projection.

One common issue related to the abovementioned IBP methods is that their results suffer from ringing effects around the edges. This is dealt with by employing edge preserving regularization [167, 168], total variation regularization [171]. To address the slow or no convergence of the IBP methods, a meta-heuristic optimization is proposed in [169]. In [172], Qin uses a wavelet-based locally adaptive interpolation for initial value estimation, which results in fast convergence. Choosing the back projection operator is often difficult, since it requires accurate knowledge of geometric transformation and degradation process. Yan et al. [170] propose integrating Papoulis-Gerchberg method with IBP method in order to avoid the requirement of a back projection operator. Even though research has been done [175-177] to add prior knowledge to tackle the ill-posed nature of the inverse SR problem, this process seems significantly difficult compared to POCS and regularized approaches.

A comparative overview of different multi-frame super-resolution algorithms is shown in Table 3.1.

Table 3.1: Comparative overview of different multi-frame super-resolution algorithms

Category	**Advantages**	**Disadvantages**
Frequency-domain	Simple computation	Good for pure translational motion and linear space invariant blur among LR frames
Interpolation	Simple computation	No guarantee about the optimality of the estimation, lack of prior constraints, limited degradation model
Deterministic regularization	Unique solution guaranteed using convex and differentiable cost functional	Final solution often overly smooth due to lack of edge preserving prior
Stochastic regularization	Flexibility in modeling noise characteristics and prior knowledge, easy to implement edge preserving prior	High computation, degradation of performance with large magnification factor
Iterative back projection	Intuitive and easy to understand	Nonunique solution, slow convergence, choice of back projection operator arbitrary, difficult to integrate prior information
Set theoretic	Flexibility in incorporating prior information	Nonunique solution, slow convergence with improper constraint set selection, high computation, performance heavily dependent of accurate motion estimation

3.3.2 Single-frame super-resolution methods

The multi-frame SR techniques, discussed so far, fail to perform accurately if the estimated motion between the LR frames is not precise, i.e., not in subpixel accuracy. Since precise motion estimation becomes less achievable for complex motions of the objects in real world scenes, single-frame SR algorithms offer more attractive solutions for real world applications. However, single-frame SR is inherently an ill-posed problem, as there could be several HR images generated from the same LR input. Accordingly, prior information like those used in reconstruction-based multi-frame SR algorithms is required to approach this problem. But the generic smoothness prior used in case of multi-frame SR algorithms cannot help single-frame SR algorithms in reconstructing the lost high frequencies. Thus, in this case, the prior information is defined either in the explicit form of a class-specific energy functional leading to the reconstruction-based SR, or in the implicit form of example images leading to the example-based SR [178]. Depending on the types of prior used, reconstruction-based methods can be either edge-directed or regularization-based. Depending on the type of dictionary requirement (which is typical in example-based methods), example-based methods can be either neighbor embedding-based, or regression-based, or sparse coding-based. All these methods are discussed in the following subsections.

3.3.2.1 Edge directed super-resolution methods

Motivated by the fact that edges are important primitive image structures that are more robust to image scale changes than other structures, edge-focusing prior are typically used by this category of SR algorithms [148, 179-184]. These algorithms learn prior by looking into the relationship between the edge features present in the LR and the HR images. The learned information is then used to apply an edge-focusing constraint to the reconstruction process. Different features of edges have been used as prior information, including depth and width of an edge [148, 182] and the parameter of the gradient profile [179-181]. In the approach of [182], the authors use statistical edge dependency between two resolutions as the prior information in order to increase image resolution. In [148], given a weighted grid-graph $G = \langle V, E \rangle$, and a curve Cin $\mathbb{R}^2$, a soft-edge smoothness prior is approximated by the following cut metric as follows:

$$|C|_G = \sum_{e \in E_C} w_e \tag{3.20}$$

where E_C the set of edges intersecting with the curve is C, w_e is the edge width, $|C|_G$ is the weight summation of the edges intersecting with C.

Li et al. [181] propose a global non-zero gradient prior while solving an optimization problem to reconstruct the edge components correctly. In [180] the authors show that edge gradient profile in natural images follow the generalized Gaussian distribution as:

$$g(x;\sigma,\lambda) = \frac{\lambda\alpha(\lambda)}{2\sigma\Gamma(\frac{1}{\lambda})} exp\left(-\left(\alpha(\lambda)\left|\frac{x}{\sigma}\right|\right)^{\lambda}\right) \tag{3.21}$$

where $\Gamma(\cdot)$ is the gamma function and $\alpha(\lambda) = \sqrt{\Gamma\left(\frac{3}{\lambda}\right)/\Gamma\left(\frac{1}{\lambda}\right)}$ is the scaling factor which makes the second moment of the above distribution equal to $\sigma^2$and thus allows estimation of σ from the second moment. The shape parameter λ controls the overall shape of the distribution.

Some other algorithms [183, 184] in this category try to reconstruct the image details while focusing on sharpening the edges by utilizing different filters (e.g. bilateral filter, shock filter).

3.3.2.2 Regularization-based super-resolution methods

This group of algorithms [171,185-188] try to solve the ill-posed SR problem by using different regularization methods, e.g. sparsity, total variation etc. Li et al. [186] propose a SR algorithm using two complementary regularization terms for the MAP framework: a steering kernel regression total variation (SKRT) and a non-local total variation (NLTV). In [171], a total variation regularization term is used to guide the iterative back-projection process and minimize the SR reconstruction error. The total variation energy functional is comprised of the total variation norm of an image I and the fidelity of this image to the observational image I_0 as:

$$E_{TV} = \int_{\Omega} \left(|\nabla I| + \frac{\lambda}{2}(I - I_0)^2\right) dxdy \tag{3.22}$$

where $\lambda \in \mathbb{R}$ is a scalar controlling the fidelity of the solution to the input image.

The total variation regularized approaches do not help much in recovering the fine image details and often result in staircase artifacts. Later, several research have been performed in the literature using sparsity-based regularizations with greater success. Using sparsity regularization (constraint that the HR image is sparse in the wavelet domain) along with compressed sensing, the authors in [188] propose a compressive image SR. [185] use a non-local regularization method, which at first estimates the sparse domain of the HR image patches and then utilize a non-local self-similarity constraint to achieve the HR image. Following the similar concept of non-local regularization, Li et al. [187] suggest a dual-sparsity regularized sparse representation in order to tackle the limited capability of the earlier methods. Here a row non-local similarity regularization is introduced along with the conventional column non-local similarity sparse representation model.

3.3.2.3 Neighbor embedding-based super-resolution methods

Example-based single-frame SR method aims at estimating the HR image by employing a dictionary of patch correspondences. One serious consequence of this approach is the necessity of having an enormous dictionary that includes any patches possibly encountered during testing. To overcome the large-dictionary requirement, neighbor embedding-based methods use a manifold learning technique based on local linear embedding (LLE) to estimate the HR image from a single LR image and a set of training patches. The basic assumption is that a patch in the LR input image and the corresponding HR unknown patch share similar neighborhood structures. As a consequence, once the LR patch is expressed as the linear combination of a certain number of its neighbors taken from the dictionary, the output patch can be reconstructed by using the HR patches in the dictionary corresponding to the neighbors selected, and combining them in the same way [193].

Let $L_{test} = \{l_{test}^{p}\}_{p=1}^{N_t}$ be the LR input image represented as collection of N_t overlapping patches and $H_{test} = \{h_{test}^{p}\}_{p=1}^{N_t}$ be the corresponding HR estimate. Let the training set of LR patches be $L_{sample} = \{l_{sample}^{q}\}_{q=1}^{N_s}$ and the corresponding training set of HR patches be $H_{sample} = \{h_{sample}^{q}\}_{q=1}^{N_s}$, where N_s represents the number of training samples. For each LR input patch l_{test}^{p}, these algorithms find the set N_p of k-nearest neighbors in L_{sample} and compute a weight-vector $(\widehat{w})$ for the neighbors that minimizes the error of reconstructing l_{test}^{p} as follows [192]:

$$\widehat{w} = \arg min_{w_s} \| l_{test}^{p} - \sum_{l_{sample}^{q} \in N_p} w_s l_{sample}^{q} \|^2 \qquad (3.23)$$

$$s.t. \sum_{l_{sample}^{q} \in N_p} w_s = 1$$

Finally the same weight-vector is used to reconstruct the HR patch h_{test}^{p} corresponding to the LR patch l_{test}^{p} as:

$$h_{test}^{p} = \sum_{l_{sample}^{q} \in N_p} \widehat{w} h_{sample}^{q}$$

In order to avoid subtractive combination of patches, which often causes unstable behaviors for these algorithms, a semi-nonnegative matrix factorization (SNMF) method is proposed in [193], which replaces the above weight constraint with $w_s \geq 0$

Feature selection plays a crucial role in preserving the neighborhood relationships among the LR and HR patches. In the pioneering work [192] the authors use 1st and 2nd order gradient of the luminance as the feature vectors. In an attempt to reduce the sensitivity of the selected features to noise, and at the same time to preserve the neighborhood in a more robust way, several feature vectors have been utilized. [194] uses a combination of 1st order gradient and norm luminance, [198] uses DCT coefficients of the norm luminance, [196] uses a combination of 1st order gradient and residual luminance, [197]uses norm luminance along with stationary wavelet transform coefficients, [195] uses DCT coefficients of the interpolated patches.

One weakness related to the aforementioned NE-based algorithms is that when they try to preserve the neighborhood of the LR space for the reconstructed HR space, the geometry of the actual HR space is neglected. Since the LR image involves several degradations, the neighborhood relationship of the LR space does not truly reflect the original HR space. In [199-201] the authors propose incorporating a locality constraint into the patch representation objective function as below.

$$\widehat{w} = \arg min_{w_s} \| l_{test}^{p} - \sum_{l_{sample}^{q} \in N_p} w_s l_{sample}^{q} \|^2 + \tau \| dist \ \odot \widehat{w} \|^2 \tag{3.24}$$

where τ is the regularization parameter, $\odot$ denotes the element-wise multiplication, $dist$ is a locality adaptor that gives different freedom for each LR training patch l_{sample}^{q} to its similarity to the LR input patch l_{test}^{p}.

3.3.2.4 Regression-based super-resolution methods

Differently than the NE-based methods, regression-based methods [189,202-207] attempt to learn the relationship between the space of the LR patches and the space pf the HR patches by finding a regression function f.Thus, this group of methods are not based on the LLE assumption, which is often doubted. Similar to the NE-based methods, a local LR training set is first chosen according to the similarity between the LR training patch2[150]es and the LR test patch. The local HR training set corresponding to the local LR training set is also identified. A regression function is then learned taking into account the two aforementioned local training sets. If $\left\{l_{sample}^{q}\right\}_{q=1}^{\mathrm{K}}$ and $\left\{h_{sample}^{q}\right\}_{q=1}^{K}$ are the two local training sets from the LR and HR dictionaries (assuming K-nearest neighbors for each test patch are under consideration), the regression function Ꞧ_f can be obtained by minimizing a regularized cost functional as below [202]:

$$\text{Ꞧ}_f = \arg min_{f\in\mathcal{H}} \left[\sum_{q=1}^{K}\left\|h_{sample}^{q} - f(l_{sample}^{q})\right\|_2^2 + \lambda\|f\|_{\mathcal{H}}^2\right] \tag{3.35}$$

where λ denotes the regularization parameter, $\mathcal{H}$ is the Hilbert function space, $\|f\|_{\mathcal{H}}$ is the norm in $\mathcal{H}$. Finally, a unique estimation of the HR patch is generate using the regression function.

Different techniques have been used in literature to obtain the local regression function discussed above. [206,207] use a support vector regression (SVR) based on support vector algorithm, which works in the similar way neural network regression works. On the other hand, [202-205] use kernel ridge regression (KRR), which works on the basis of expanding a kernel function.

3.3.2.5 Sparse coding-based super-resolution methods

Even though the local learning-based methods use smaller training database (by allowing combination of patches) to represent more number of patches, the fixed number of K nearest neighbors for the reconstruction often generates blurring artifacts, due to over- or under-fitting [101]. In order to avoid this problem, another family of example-based method, pioneered by the work of Yang et al. [176], proposes to adaptively choose the relevant number of neighbors based on a sparse representation. Using compressive sensing theory [208], these methods [83, 176, 209-214] assume that the sparse representation of the HR patches can be precisely recovered from their LR counterparts.

Let $H_{sample} \in \mathbb{R}^{nxN_s}$ and $L_{sample} \in \mathbb{R}^{mxN_s}$ be the dictionaries of N_s example HR and LR patches respectively. Each of the HR and LR patches are assumed to be an n-dimensional and a m-dimensional feature vectors respectively. For each input LR image patch l_{test}, its sparse representation can be formulated as [176]:

$$\min\|\alpha\|_0 \qquad s.t. \quad \| L_{sample}\alpha - l_{test}\|_2^2 \leq \epsilon \tag{3.26}$$

where α is a N_s dimensional vector, called coefficient representation vector and $\|\alpha\|_0$ represents the number of non-zero elements in α. Since the l_0-norm is non-convex, it is replaced by l_1-norm to make the above optimization problem convex [210] as:

$$\min\|\alpha\|_1 \qquad s.t. \quad \| L_{sample}\alpha - l_{test}\|_2^2 \leq \epsilon \tag{3.27}$$

which can be rewritten by using Lagrange multiplier as the following optimization problem, well known as Lasso:

$$min_\alpha \| L_{sample}\alpha - l_{test}\|_2^2 + \rho\|\alpha\|_1 \tag{3.28}$$

where ρ is a regularization parameter.

Sparse coding over a large LR patch database directly is computationally expensive. Thus, a joint dictionary learning technique has been used by many researchers in the literature [209, 211, 213, 214]. By concurrently learning the following two dictionaries, the same sparse representations can be maintained with respect to the LR and HR dictionaries.

$$D_l = arg\,min_{D_l,A} \left\| L_{sample} - D_l A \right\|_2^2 + \rho \|A\|_1 \tag{3.29}$$

and

$$D_h = \arg min_{D_h,A} \left\| H_{sample} - D_h A \right\|_2^2 + \rho \|A\|_1 \tag{3.30}$$

where A is the sparse coding vector. The joint learning can then be formulated as below to obtain the new dictionary D_{new}:

$$D_{new} = \arg min_{D,A} \|T - DA\|_2^2 + \hat{\rho}\|A\|_1 \tag{3.31}$$

where $T = \left[\frac{1}{\sqrt{n}} H_{sample};\ \frac{1}{\sqrt{m}} L_{sample}\right]$ and $D = \left[\frac{1}{\sqrt{n}} D_h;\ \frac{1}{\sqrt{m}} D_l\right]$

Once D_{new} is learned, D_{new}^l and D_{new}^h can be obtained by dividing D_{new} into two parts. Using these learned couple dictionaries, sparse representation of the input LR image patch is obtained by optimizing:

$$\hat{\alpha} = min_{\alpha}\ \|D_{new}^l \alpha - l_{test}\|_2^2 + \rho\|\alpha\|_1 \tag{3.32}$$

Finally, the HR image patch h_{test} corresponding to the input LR image patch can be obtained as:

$$h_{test} = D_{new}^h \hat{\alpha} \tag{3.33}$$

Differently than Lasso, several other methods are used in literature to solve the optimization problem like equation 5.11. [215, 216] use KSVD, [211] uses stochastic gradient, [212] uses a combination of K-SVD and Batch-OMP.

A comparative overview of different single-frame super-resolution algorithms is presented in Table 3.2.

Table 3.2: Comparative overview of different single-frame super-resolution algorithms

Category	**Advantages**	**Disadvantages**
Edge directed	Output with high quality edges having proper sharpness and less artifacts	Other high-frequency structures such as textures cannot be reconstructed
Regularization	No necessity for large training dataset, ease of incorporating prior	Performance degrades with large magnification factor
Neighbor embedding	Compact dictionary size, reasonably low computation	Blurring artifacts due to over or under fitting
Regression	Computationally faster than neighbor embedding	Degraded performance compared to neighbor embedding
Sparse coding	Highly compact dictionary size, low computation, no overfitting	No guarantee about the global optimality of the estimation

CHAPTER 4

MOSAICING SYSTEM

This chapter describes the proposed mosaicing system. The first section describes the proposed algorithm. The second section describes the evaluation methodology. The third section talks about result and explanations.

4.1 Methodology

The mosaicing algorithm is composed of Scale Invariant Feature Transform (SIFT), Best Bins First (BBF), Random Sample Consensus (RANSAC), reprojection, and stitching algorithms. Figure 4.1 shows the flowcharts of the proposed mosaicing algorithm.

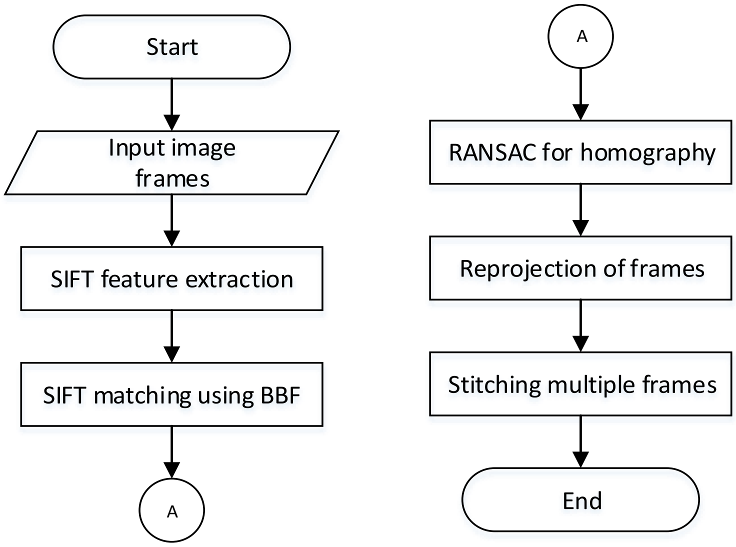


Figure 4.1: Flowchart of the mosaicing algorithm.

As shown in Figure 4.1, the first step of the mosaicing algorithm is the extraction of SIFT features or keypoints from the input frames. Initially, a scale space is constructed using a Gaussian filter with changing scales and an input image. The image is convolved repeatedly with the

Gaussian filters and grouped into octaves. An octave corresponds to doubling the value of the standard deviation of the Gaussian filter that we start the octave with. Then the difference-of-Gaussian (DoG) images are computed from adjacent Gaussian-blurred images in each octave. After the DoG images are obtained, candidate keypoints are identified as local extrema of DoG images across the scales. In the next step, low contrast keypoints and edge response points along the edges are discarded using accurate keypoint localization. The candidate keypoints are then assigned one or more orientations based on local image gradient directions. A set of orientation histograms is formed over 4x4 pixel neighborhoods with 8 bins in each. Since there are 4x4= 16 histograms each with 8 bins, a 128 dimension descriptor vector could be assigned to each keypoint. Since the 128 element keypoint descriptor is represented relative to the orientation(s) assigned to that particular keypoint, the keypoints are invariant to rotation. Furthermore, a very high dimensional descriptor vector makes the keypoints highly distinctive. Figure 2 and Figure 3 show an example dataset and SIFT features extracted from each of the frames in the dataset.

The second step evaluates the best matching keypoints between image pairs. This is achieved by identifying the nearest neighbor of a keypoint in the first image from a database of keypoints for the second image. The nearest neighbor is defined as the keypoint with minimum Euclidean distance from a given descriptor vector. However, because of the high dimensionality of the descriptor vectors, matching the feature points by comparing the descriptor vector one by one will require considerably high computation time. Instead, the use of BBF, which is an approximation algorithm, saves significant computation time at the cost of negligible loss of correct matching. This approximation is achieved by using a parameter BBF NN bins, which indicates when the BBF algorithm cuts off the search while looking for the nearest neighbor candidates for a particular feature vector.

Following the initial-matching-points-searching between a pair of images, the RANSAC algorithm is used to remove the outliers and to compute an optimum homography matrix based on certain homography constraints (geometric distance error, maximum number of inliers, etc.). The homography matrix is a 3x3 matrix which designates several transformation parameters between a pair of images. In order to find the homography matrices for a set of frames, one of the frames is assigned as the reference frame and the current homography matrix is multiplied with all the previous homography matrices until the reference frame is reached. Using the homography matrices, frames are projected into a common coordinate system.

Finally, the reprojected frames are stitched to the reference frame to construct the mosaic output. This stitching is achieved by checking each pixel of the mosaic frame to see if it belongs to the warped frame or the reference frame. Accordingly, that pixel in the mosaic frame is assigned the corresponding pixel value from the warped frame or the reference frame. Once the reference frame and the first reprojected frame is stitched, the result is treated as the reference frame for the next stitching process, and continued until all the frames are stitched.

The mosaicking program is developed in Microsoft Visual Studio C++ platform. A PC with 1.9 GHz Core2 Duo processor and 3 GB RAM is used for all the evaluations.

4.2 Evaluation

Three different categories of datasets: images of 2D surfaces, images of outdoor 3D scenes, and airborne images from an Unmanned Aerial Vehicle, are used to evaluate the performance of the proposed mosaicing algorithm. Each of these datasets has a sequence of ten frames. While the first two categories of datasets were captured by using a handheld camera, the third category of images were collected by an UAV on-board camera. For objective evaluation, four metrics are used: percentage of mismatches, difference of pixel intensities, peak signal-to-noise ratio, and mutual information. All these metrics are used to measure the amount of asymmetry between a mosaic output and a reference image. Thus obtaining the reference image is a crucial step for evaluating the mosaicing algorithm. In the following subsections we will first discuss the evaluation setup followed by the performance metrics and their interpretation.

Initially, from the wide-angle HR image 10 frames are extracted with inter-frame translational motion. This process acts as photographing a scene by multiple shots that cover different areas of the scene. Mosaicing algorithm is then applied to these frames to generate a mosaic output. From the coordinates of the individual frames a mask, imitating the shape of the mosaic, is created. With that appropriate padding (layers of black pixels) is added to give the mask same dimension as that of the mosaic output. A region-of-interest is later extracted from the wide-angle image. This region-of-interest has the same dimension as that of the mosaic frame. The mask created earlier is then used together with the region-of-interest to generate the ground-truth image corresponding to the mosaic output. A pixel-wise multiplication was performed between the mask and the region-of-interest to achieve the ground truth. Figure 4.2 shows the aforementioned procedure using a 2D scene data. Quantitative comparison between the mosaic output and the ground truth is obtained using the following four performance metrics:

Percentage of mismatches: The percentage of mismatches is a measurement of the number of mismatching pixels in two images, an output image and a ground truth image, beyond a threshold intensity value of 15. The lower this value, the higher the similarity between the two images. In the worst case of an 8-bit image, this threshold value of 15 represents only 6% of the highest intensity value (255).

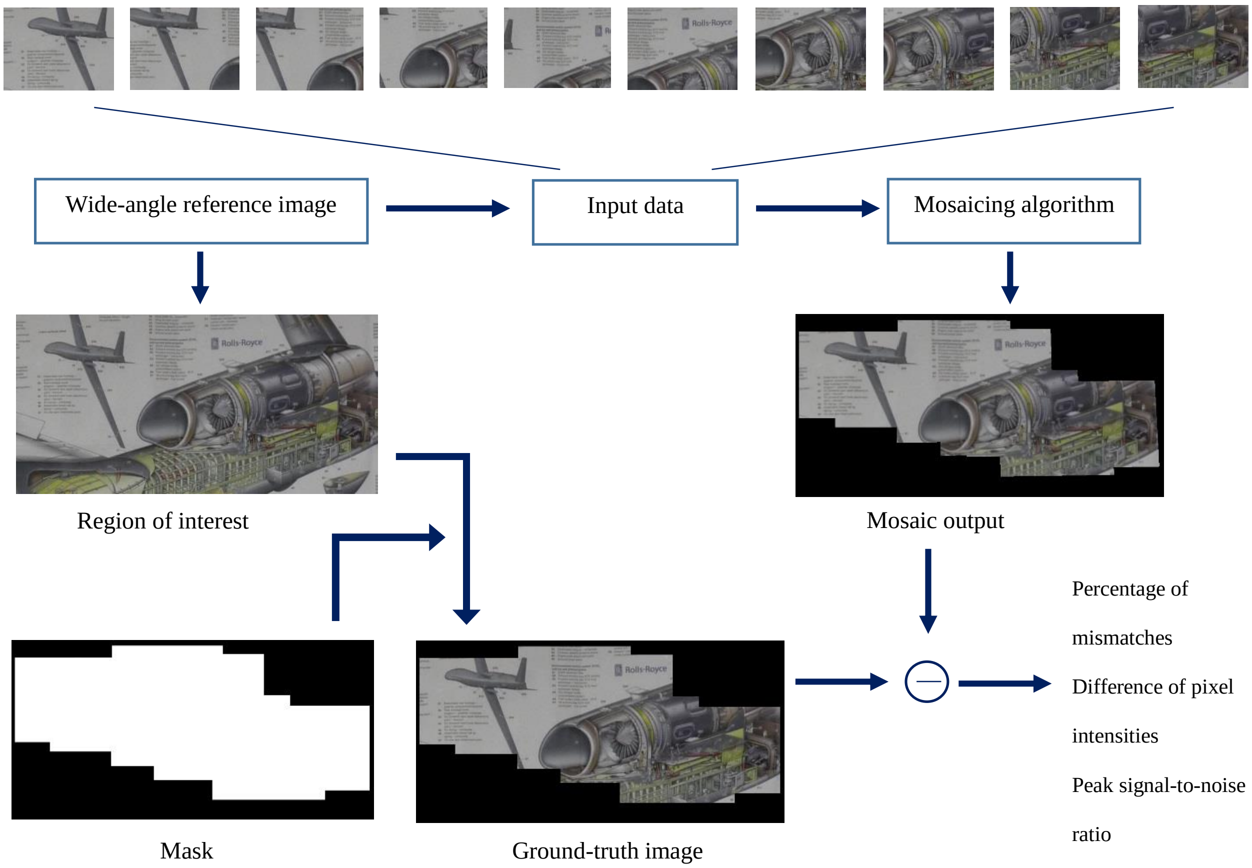


Figure 4.2: Schematic of the mosaicing algorithm's evaluation process.

Average difference of pixel intensities: The average difference of pixel intensities is a measurement of the average error of pixel intensities along with their fluctuation from the mean error value. The lower this value, the higher the similarity between two images, an output image and a ground truth image. The average difference of pixel intensity value can be used to choose the threshold intensity value for calculating the percentage of mismatches.

Peak Signal to Noise Ratio: The peak signal to noise ratio (PSNR) is used as a measurement of the difference between two images. PSNR of corresponding pixel values is defined as:

$$PSNR = \frac{10\log_{10}(\max(G(i,j),O(i,j)))^2}{\mathrm{MSE}} \tag{4.1}$$

where, MSE is the mean square error and G(i,j) and O(i,j) are the (i,j)th pixel values in the ground-truth and the mosaic output respectively. The mean square error is given as

$$MSE = \frac{\sum_{\mathrm{i}}\sum_{\mathrm{j}}(\mathrm{G(I,j)-O(I,j)})^2}{\mathrm{N}} \tag{4.2}$$

where, N is the total number of pixels in each image.

The lower the difference between two the images (and hence lower the MSE), the higher the PSNR between them.

Mutual information: The mutual information (MI) is a measurement of the asymmetry between two images along with its fluctuation from the mean value. MI between two images G(i,j) and O(i,j) is expressed as:

$$I(G,O) \; = \; H(G) + H(O) - H(G,O) \tag{4.3}$$

where, H(G) is the entropy of the image G(i,j), H(O) is the entropy of the image O(i,j),and H(G,O) is their joint entropy, which is expressed as:

$$H(G,O) = -\sum_{i,j} p_{GO}(i,j) log p_{GO}(i,j) \tag{4.4}$$

where $p_{GO}(i,j)$ represents the probability of a single pixel pair from G(i,j) and O(i,j).

Maximizing the mutual information is equivalent to minimizing the joint entropy. The lower the joint entropy (and hence higher the mutual information) between two images, the higher the similarity between them. The advantage of using mutual information over joint entropy is that it includes the entropy of the individual images, which adds offset to pixels that can help in low contrast regions [217].

4.3 Results and discussion

The probability that a nearest neighbor found during SIFT matching is correct can be determined by taking the ratio of distance of the closest neighbor to the distance of the second closest neighbor. Typically a threshold value for the aforementioned ratio is used to eliminate the false or incorrect matches, especially arising from background clutter or noise. Matches for which the ratio is greater than the threshold value are rejected. Lower the value of this threshold, better is the performance of the SIFT matching. Because correct match needs to have the closest neighbor significantly closer than the closest incorrect match. This could be verified from the graph of the

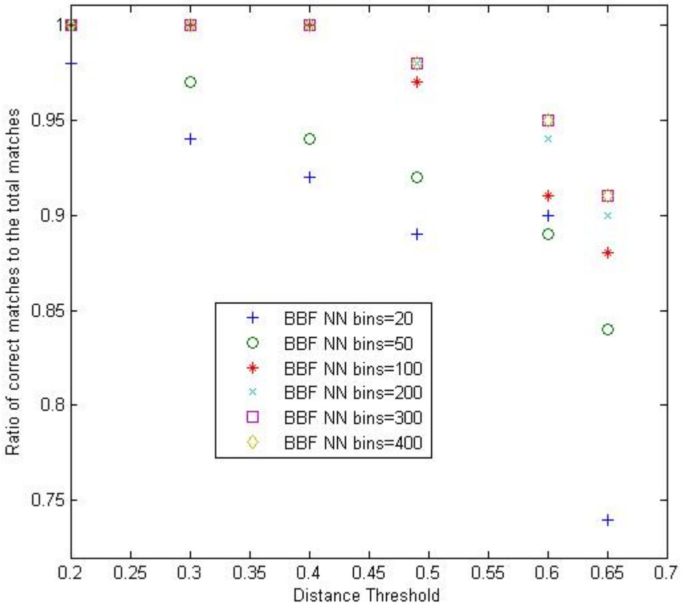


Figure 4.3: Matching performance vs distance threshold at different values of BBF NN bins for an example image pair.

matching performance vs the distance threshold value as shown in Figure 4.3. Note that the plots are drawn for different values of the BBF NN bins, which indicates when the BBF algorithm terminates its search operation. As can be seen from the graph, a distance threshold value close to 0.5 would be ideal. Note that a too low value of this threshold would possibly result in discarding many correct matches.

As previously mentioned, the value of BBF NN bins is chosen such that the search algorithm cuts off the operation while looking for the nearest neighbor candidates for a particular feature vector. Thus, lower the value of this parameter, higher is the speed of the nearest neighbor search. However, a low value could cost losing a high number of correct matches. Figure 4.4 shows that a minimum value of 200 is ideal for BBF bins value in order not to compromise the performance of the algorithm. Note that plots are drawn while varying the aforementioned distance

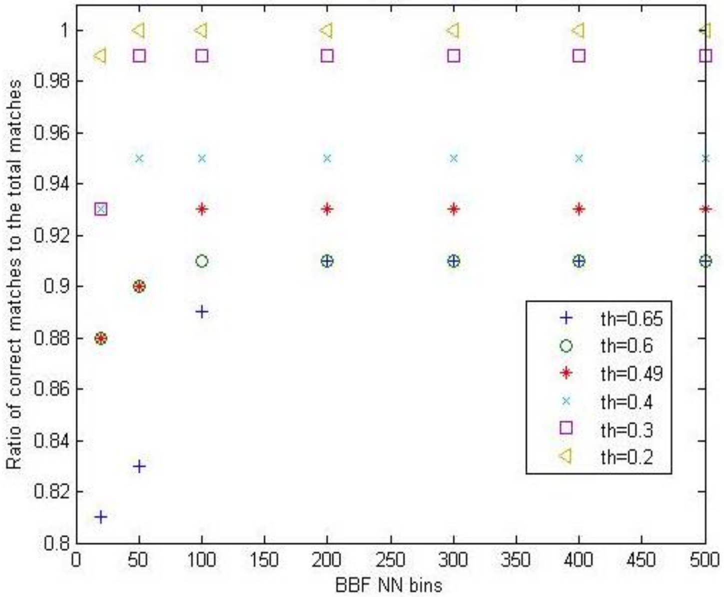


Figure 4.4: Matching performance vs number of BBF NN bins at different distance threshold values for an example image pair.

threshold parameter value. It is clear from the graph that the performance saturates at a BBF NN bins of 200. Thus increasing this number would only result in slowing the entire search process without improving the result.

Probability that the final transformation returned by the RANSAC algorithm is corrupted by outliers is a user-defined parameter which balances the tradeoff between the computation and the performance of the algorithm. Lower the value of this parameter, better is the transformation model, hence better is the performance. However, it costs more computation, since more number of iterations is required. Similarly, distance threshold for considering inliers while computing the final transformation is also a user-defined parameter. Lower the value of this parameter, better is the final transformation model; however at the cost of more computation. Figure 4.5 and Figure

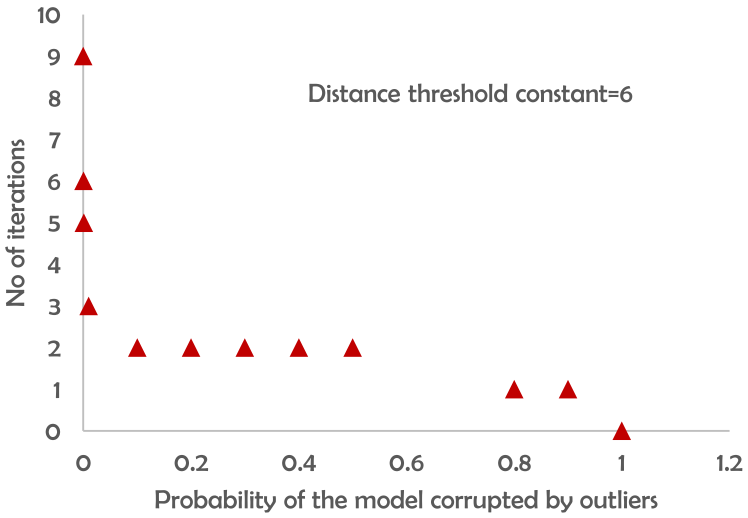


Figure 4.5: Performance vs computation of the RANSAC algorithm at a constant distance threshold value for an example image pair.

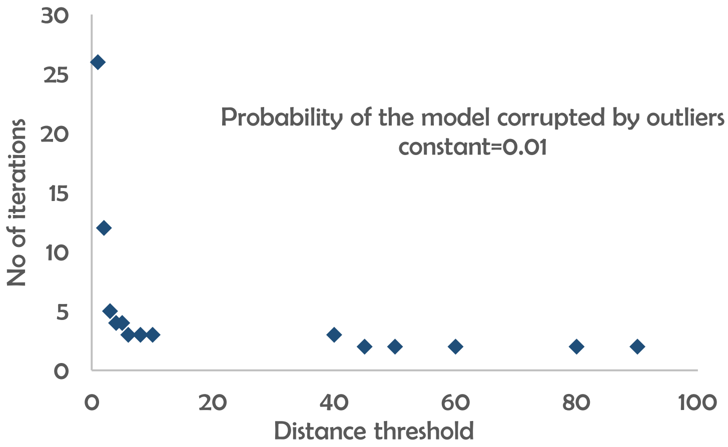


Figure 4.6: Distance threshold vs computation of the RANSAC algorithm at a constant probability of model corruption for an example image pair.

4.6 show the effect of changing these parameters on computation. We have empirically chosen the values of 0.01 and 6 for the two aforementioned parameters.

While evaluating the sensitivity of the mosaicing algorithm to image details, it is seen that the algorithm works even with very few numbers (but not less than 10) of extracted features from the frames. However, a minimum number of 10 matching features is required for the success of the algorithm.

Output of the different steps of the proposed mosaicing algorithm is shown in Figure 4.8-4.11 using an example UAV scene dataset (Figure 4.7). Note that optimal parameter values are chosen based on the aforementioned discussion.

Figure 4.7: Input frames for the mosaicing algorithm.

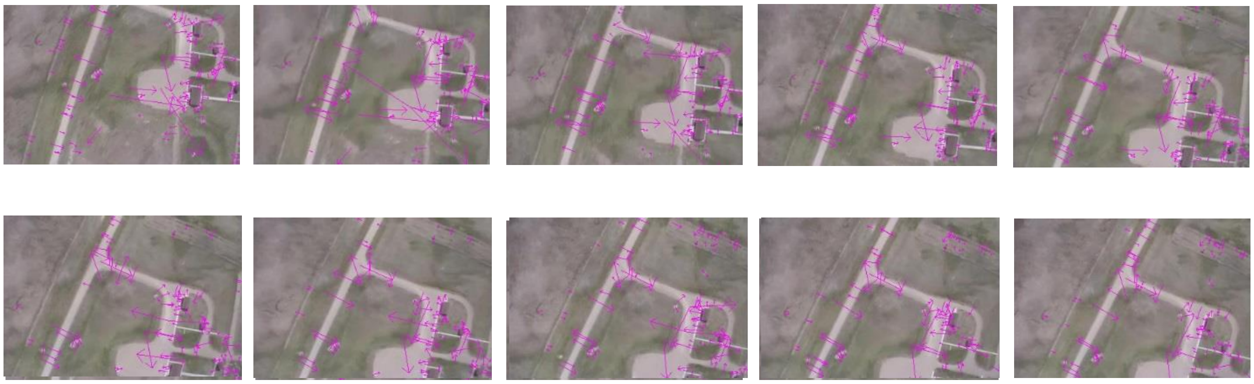


Figure 4.8: SIFT features extracted from the input frames.

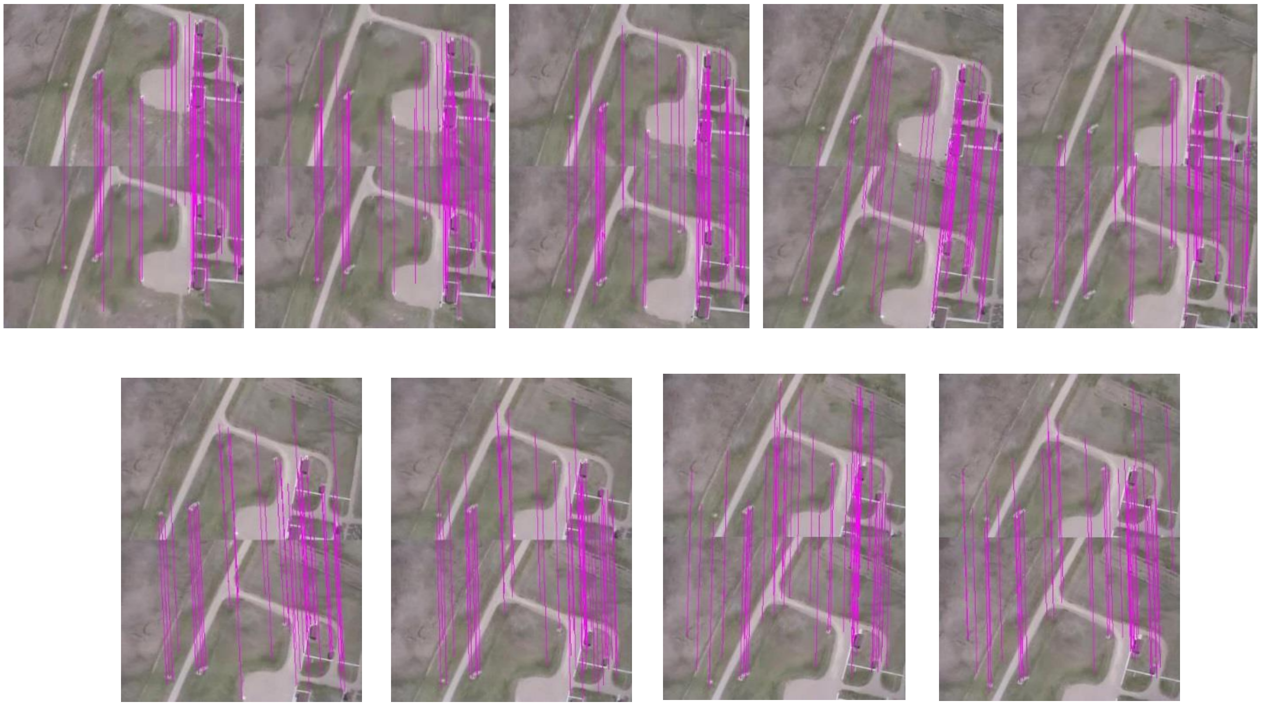


Figure 4.9: SIFT features matching from pair of input frames.

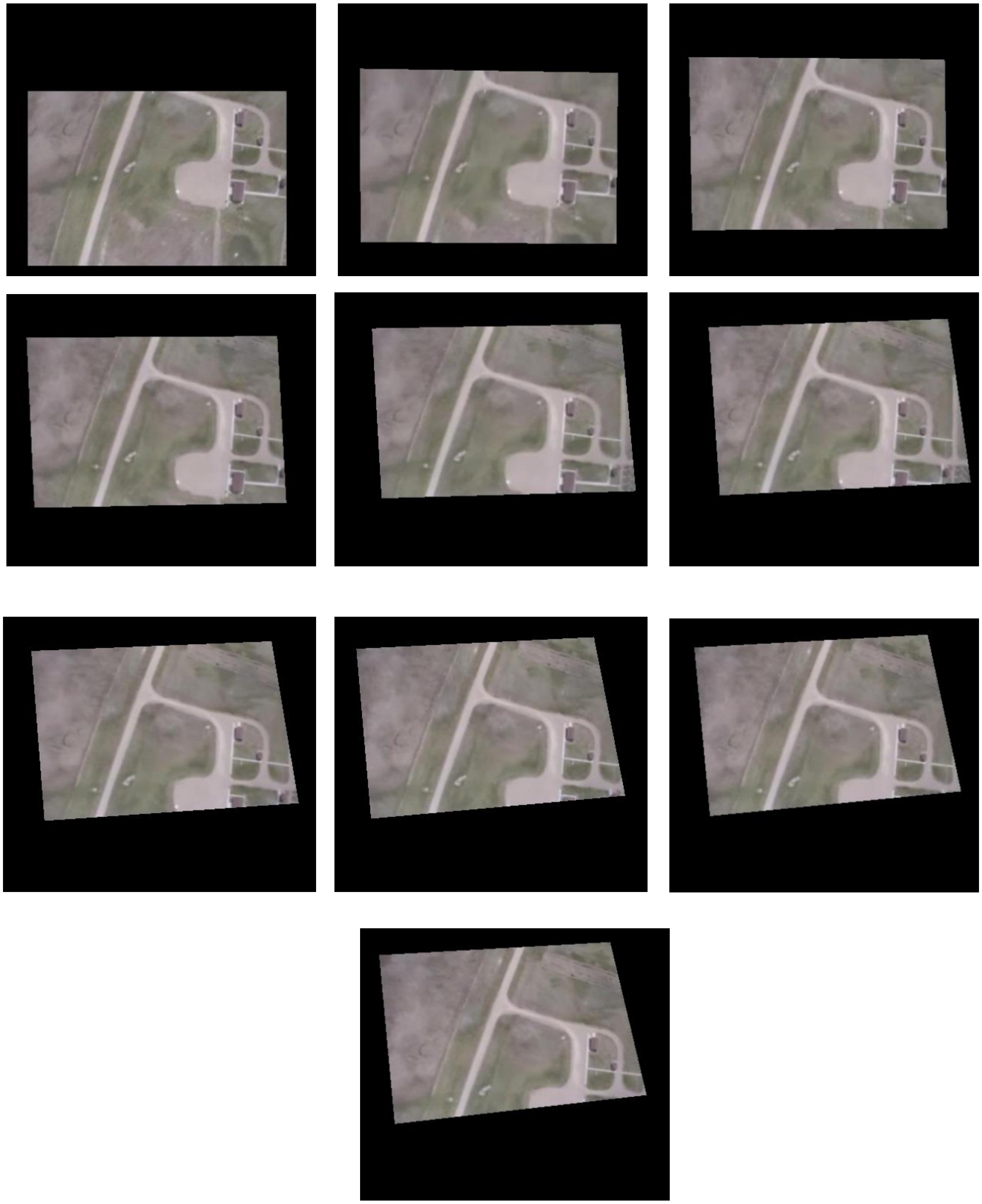


Figure 4.10: Projection of frames into common coordinate.

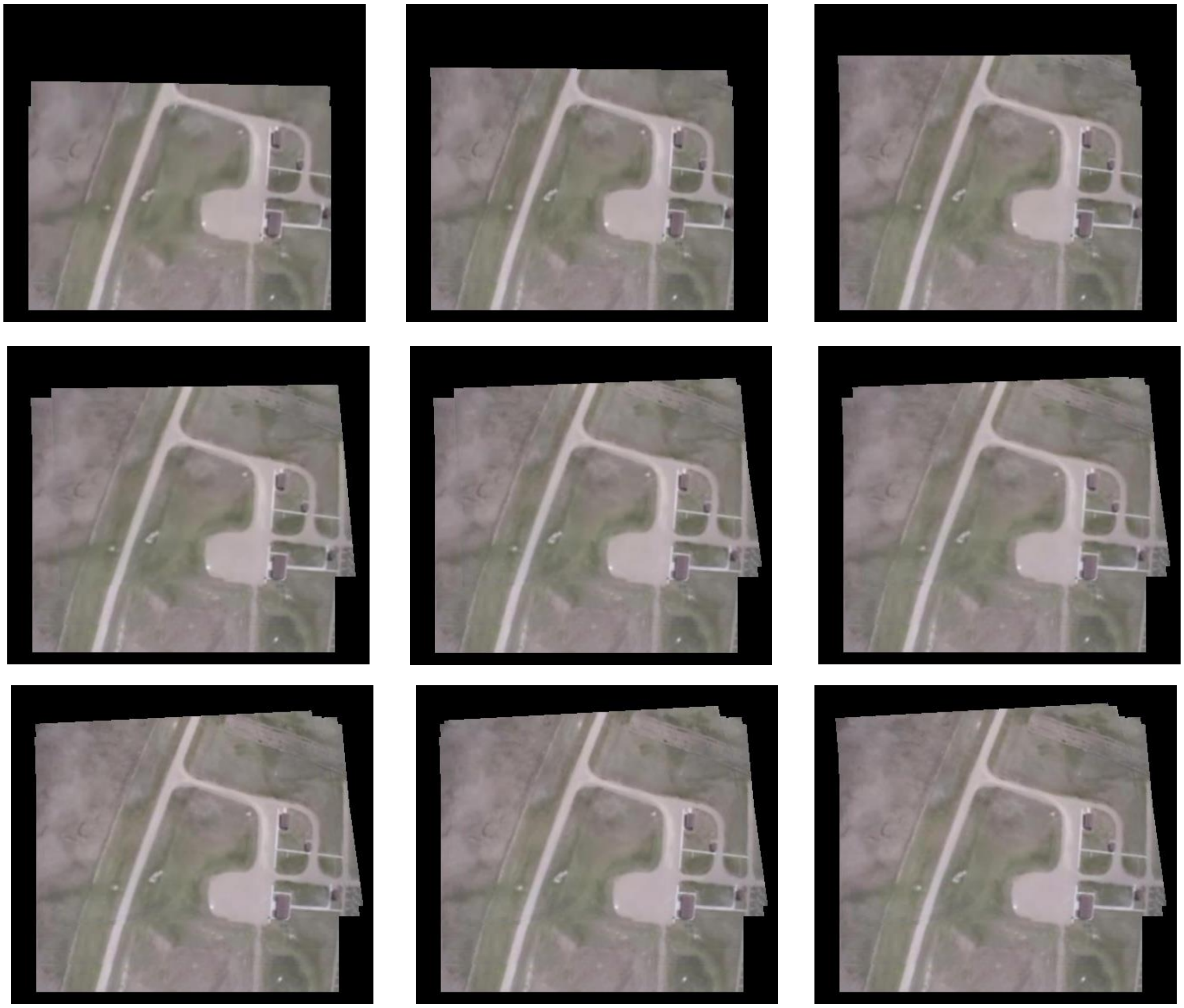
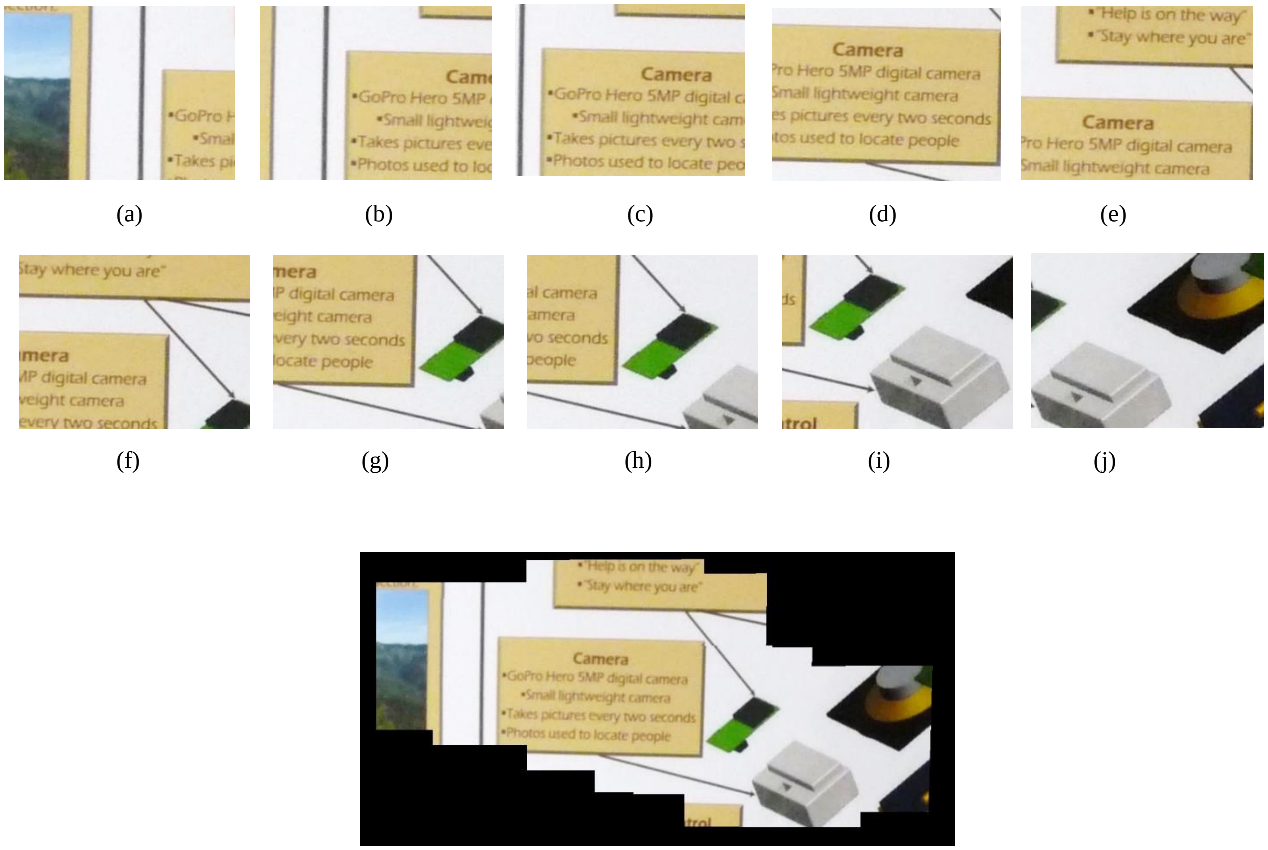


Figure 4.11: Step-by-step stitching process.

Figure 4.12, 4.13 and 4.14 show examples of the mosaicing algorithm using one of the sets of each category of data. Figure 4.12 shows an example of a 2D scene consisting of 10 frames (Figure 4.12 a, b, c, d, e, f, g, h, i, and j) and their corresponding mosaicing output (Figure 4.12 k). Figure 4.13 shows an example of a 3D scene consisting of 10 frames (Figure 4.13 a, b, c, d, e, f, g, h, i, and j) and their corresponding mosaicing output (Figure 4.13 k). Figure 4.14 shows an example of a UAV scene consisting of 10 frames (Figure 4.14 a, b, c, d, e, f, g, h, i, and j) and their corresponding mosaicing output (Figure 4.14 k) obtained during a 2011 University of North Dakota UAV flight test.



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

(f) (g) (h) (i) (j)

(k)

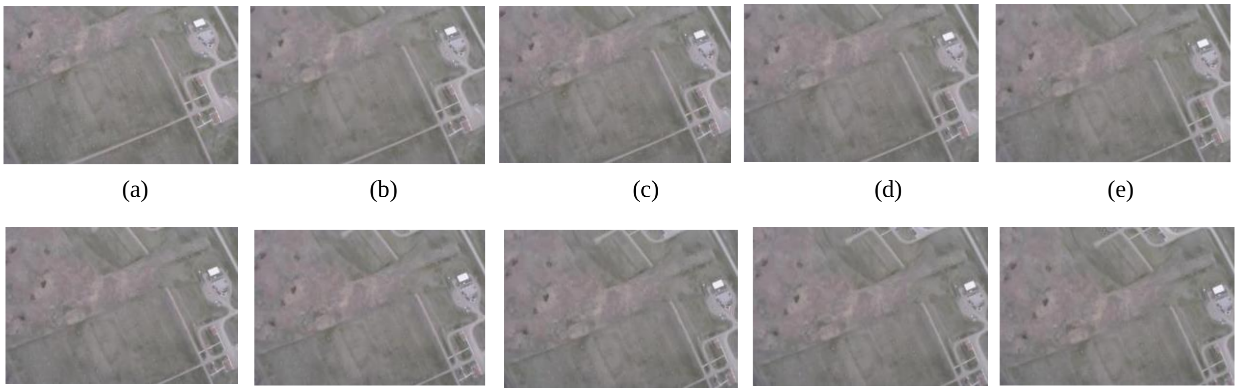
Figure 4.12: Mosaicing using an example 2D scene dataset. (a)-(j) Input frames; (k) Mosaic output.

(f) (g) (h) (i) (j)

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

(k)

Figure 4.13: Mosaicing using an example 3D scene dataset. (a)-(j) Input frames; (k) Mosaic output.



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

(f) (g) (h) (i) (j)

(k)

Figure 4.14: Mosaicing using an example UAV dataset. (a)-(j) Input frames; (k) Mosaic output.

A small value of the average difference of pixel intensities indicates that the mosaicing output is similar to the ground-truth image, and, hence it gives a larger PSNR value. If the mosaicing output is barely similar to the ground-truth (which we have considered as worst case scenarios), the average difference of pixel intensities is of the order of 35, 36 and 20, respectively for the three aforementioned categories. Similarly, the PSNR is of the order of 11dB, 12dB and 15dB for the worst case scenarios.

Table 4.1 shows the results of the assessment using four metrics on the mosaicing outputs of the 36 scenes. The average difference in pixel intensities is of the order of 1, 2 and 3 for the three categories respectively, which gives an average ratio of mismatches for different categories of data ranging from 0.91 to 1.42 using a pixel intensity threshold of 15. The fluctuation of the average difference of pixel intensities is small for all the three categories compared to the highest values of 39.56, 41.47 and 35.2 for the worst cases. PSNR average values are of 22.99 dB, 27.06 dB and 38.50 dB for the three categories of data. Compared to the worst case values of 11 dB, 12 dB and 15 dB, these PSNR average values indicate that the mosaicing algorithm produces outputs similar to their corresponding ground-truth images. Low joint entropy values of 8.01, 8.52 and 5.92, compared to the highest values of 10.16, 10.25 and 7.9 for the three categories of data, designate low mosaicing errors. High mutual information values of 2.8, 3 and 3.46, compared to the lowest values of 1.01, 1.12 and 0.7, show low stitching errors. The average PSNR measures the geometric and photometric error, along with the variation of those errors from the mean error values; accordingly, it is the most appropriate choice for the quantitative evaluation of image mosaicing for the cases, where different input image frames are created from a single reference image.

Table 4.1: Mosaicing algorithm assessment

Categories of data	Average ratio of mismatch (%)	Average difference of pixel intensity	Average PSNR in dB	Joint entropy	Mutual information
Images of 2D surfaces	1.42	3.14+/-14.05	22.99	8.01+/-0.49	2.8+/-0.30
Images of outdoor 3D scenes	1.26	2.83+/-10.58	27.06	8.52+/-0.50	3+/-0.29
Airborne images from UAV	0.91	0.86+/-2.22	38.50	5.92+/-1.00	3.46+/-1.51

CHAPTER 5

SUPER-RESOLUTION MOSAICING SYSTEM

This chapter describes the proposed super-resolution mosaicing system. The first section describes the proposed algorithm. The second section describes the evaluation methodology. The third section talks about result and explanations.

5.1 Methodology

The proposed super-resolution mosaicing method and the two other comparative methods (method based on Tikhonov regularization and Total Variation regularization) are all based on similar concepts of minimizing an error functional using maximum *a posterior* estimates and then solving optimization problems. Thus, these algorithms share similar mathematical backgrounds but utilize different Norms and regularizations. In this section, the common mathematical model using various Norms and regularizations employed by these three algorithms is discussed in detail.

In order to develop a comprehensive understanding of the super-resolution mosaicing algorithm it is often necessary to formulate a linear observation model which relates the acquired low-resolution images to the super-resolution mosaic. Note that this observation model has the similar formulation as that used in Chapter 3, in the state-of-the-art of the super-resolution. The only difference is that the present model includes a mosaicing operator in the formulation. Similar to the earlier model this one incorporates warp, blur (both atmospheric blur and optical blur), noise, and downsampling, since these are the most common degradations and can be modelled fully or partially in different super-resolution mosaicing techniques. Using the same notations used in the

SR observation model, the observation model for super-resolution mosaicing could be expressed as [221]:

$$y_k = DB_kW_kR[x]_k + n_k \qquad \text{for } 1 \leq k \leq K \tag{5.1}$$

Unlike in the observation model for SR, here x is the desired super-resolution mosaic. $R[\quad]$ is the reconstruction operator, that extracts warped images from the super-resolution mosaic.

Since the aim of the super-resolution mosaicing algorithm is to determine an estimate of x given the captured image sequence and the characterization of the imaging process, it is fundamentally an inverse process. Consequently the super-resolution mosaicing algorithm's stability is not solely determined by the availability of multiple low-resolution observations, rather estimation of several other factors like B_k and n_k are also necessary [13]. Clearly, super-resolution mosaic assembly is a large sparse optimization problem which could be solved using iterative methods [222]. However, instead of sparse matrices multiplication, basic image operations (e.g. convolution, warping, down-sampling) could be applied along with gradient computation in order to speed up the required super-resolution computations. Subsequently, an estimate of the super-resolution mosaic $\hat{x}$ could be achieved from equation 5.1 by optimizing a utility function which minimizes the error between the input low-resolution images and the reconstructed ones [223]. A common utility function using the maximum likelihood estimate is expressed as:

$$\hat{x} = \arg min_x \left[\sum_{k=1}^{K} \| y_k - DB_kW_kR[x]_k \|_2^2 \right] \tag{5.2}$$

Note that L2 Norm is used in equation 5.2. Super resolution mosaic is a typical ill-posed problem because of the insufficient number of low-resolution frames and ill-conditioned blur operations

[224]. The minimum number of non-redundant low-resolution frames available must be the square of the resolution enhancement factor. For an under-determined case (number of non-redundant low-resolution frames < square of resolution enhancement factor), there exists an infinite number of solutions for equation 5.2. On the other hand, for over-determined cases (number of non-redundant low-resolution frames ≥ square of resolution enhancement factor) the solution is not stable and becomes extremely sensitive to noise [218]. Thus, considering regularization in super-resolution mosaic problems becomes significant since regularization allows addition of a smoothness constraint to the solution by penalizing gradients or higher-order spatial derivatives in the image. If the smoothness constraints are differentiable and their derivatives can be approximated by image operations, it is possible to include those constraints in the maximum-likelihood estimates in order to achieve maximum a posterior estimators. The most common approach for representing the smoothness constraint is the use of a discrete 2D Laplacian operator, L. In super-resolution literature, this approach is well known as super-resolution with *Tikhonov* regularization, which is one of the most representative algorithms. Using this method, the constrained least square (CLS) formulation for super resolution mosaicing can be written as [219]:

$$\hat{x} = \arg min_x \left[\sum_{k=1}^{K} \| y_k - DB_k W_k R[x]_k \|_2^2 + \lambda \| Lx \|_2^2 \right] \tag{5.3}$$

The first term on the right hand side is the data fidelity term, and the second term is the regularization term. The intention behind the aforementioned regularization method is to enforce spatial smoothness on the solution. For very noisy images, a higher value of λ is desirable since it suppresses the noise components. However, since noise pixels and edge pixels both contain high-frequency components, both of them are subdued in this regularization method, and the final solution becomes overly smoothed, lacking sharp edges and detailed information. A superior method for adding regularization is to use a *Total Variation* (TV) regularization, which reduces the shortcomings of the Tikhonov regularization. The TV regularization term is expressed as:

$$\Gamma(x) = \int_c |\nabla x| dxdy = \sum_c |\nabla x| \tag{5.4}$$

where c denotes a clique (local group of points) in the image domain. Thus the TV regularization term computes gradient operation over the clique.

$$|\nabla x| = \sqrt{|\nabla x|^2} = \sqrt{\nabla x_h^2 + \nabla x_v^2} \tag{5.5}$$

where $\nabla x_h$and ∇x_v are linear operations corresponding to horizontal and vertical first-order differences, respectively. Thus, at pixel (p, q),

$$\nabla x_h(p, q) = x(p + 1, q) - x(p, q) \tag{5.6}$$

and

$$\nabla x_v(p, q) = x(p, q + 1) - x(p, q) \tag{5.7}$$

Based on TV, the CLS formulation for super resolution mosaicing can be written as:

$$\hat{x} = \arg min_x \left[\sum_{k=1}^{K} \|y_k - DB_k W_k R[x]_k\|_2^2 + \lambda \sum_c \sqrt{\nabla x_h^2 + \nabla x_v^2} \right] \tag{5.8}$$

It is noted that the smoothness constraint in the above equation is not differentiable when ∇x=0. Hence, a small positive parameter ε is added to ensure differentiability. Thus the CLS formulation can be rewritten as:

$$\hat{x} = \arg min_x \left[\sum_{k=1}^{K} \|y_k - DB_k W_k R[x]_k\|_2^2 + \lambda \sum_{c} \sqrt{\nabla x_h^2 + \nabla x_v^2 + \varepsilon} \right] \tag{5.9}$$

Although the TV regularization has the advantages of preserving edges and other detailed information, it results in “staircase effects” in the flat regions of the super-resolved image. This can be significantly reduced by adjusting the regularization parameter λ to a large value, but the edge and texture information will be smoothed out. Considering the drawbacks of the TV regularization, I propose a directional Huber-Markov regularization model for the super-resolution mosaicing problem.

Huber-based prior penalizes the edges and other discontinuities less severely and encourages local smoothness in the output image. Using Huber-based regularization, CLS formulation can be written as:

$$\hat{x} = \arg min_x \left[\sum_{k=1}^{K} \|y_k - DB_k W_k R[x]_k\|_2^2 + \lambda f(g) \right] \tag{5.10}$$

The Huber function can be defined as:

$$f(z) = \begin{cases} z^2, & if\ |z| \le \alpha \\ 2\alpha|z| - \alpha^2, & otherwise \end{cases} \tag{5.11}$$

$$f'(z) = \begin{cases} 2z, & if\ |z| \le \alpha \\ 2\alpha\, sgn(z), & otherwise \end{cases} \tag{5.12}$$

Note that Huber function switches between a quadratic and a linear function depending on a threshold α. For smaller z values the function is quadratic, whereas for larger z values the function becomes linear. The variable z often refers to the smoothness measure, thus the Huber function switches between eliminating small scale noise and preserving edges or other discontinuities in the high-resolution image.

For Huber-Markov regularization, the Huber function f(.) is applied over a clique c [225]. Thus, $f(g) = \sum_{g\in c} f(g)$. The operation over cliques is approximated over directional image gradients dx in the super-resolution mosaic x. Conventionally, four directional image gradients are used. Thus, $\sum_{g\in c} f(g)$ can be expressed as $\sum_{i=1}^{4} f(d_i x)$, which becomes directional Huber-Markov regularization.

The Huber Norm has been proven to be more robust with respect to outliers than L1 and L2. Unlike the other two Norms, it uses L1 Norm for the outliers, making this approximation less sensitive to those points. Using directional Huber-Markov regularization along with Huber norm for the data fidelity gives the following CLS approximation:

$$\hat{x} = \arg min_x \left[\sum_{k=1}^{K} h(y_k - DB_k W_k R[x]_k) + \lambda \sum_i f(d_i x)\right] \tag{5.13}$$

where h(.) is another Huber function similar to $f(.)$ discussed above.

Using Tikhonov regularization, TV regularization, and the proposed directional Huber-Markov regularization, the error functional between input low -resolution images and the reconstructed ones can be expressed as:

$$E(x) = \frac{1}{2}\left[\sum_{k=1}^{K} \|y_k - DB_k W_k R[x]_k\|_2^2 + \lambda \|Lx\|_2^2\right] \tag{5.14}$$

$$E(x) = \left[\sum_{k=1}^{K} \|y_k - DB_k W_k R[x]_k\|_2^2 + \lambda \sum_c \sqrt{\nabla x_h^2 + \nabla x_v^2 + \varepsilon}\right] \tag{5.15}$$

$$E(x) = \left[\sum_{k=1}^{K} h(y_k - DB_kW_kR[x]_k) + \lambda \sum_{i} f(d_i x)\right] \tag{5.16}$$

$\nabla E(x) = 0$ is solved in order to find the actual minimizer x of the minimization problem formed by equations 5.3, 5.9, and 5.13. Furthermore, a unique estimate $\hat{x}$ of x can be iteratively achieved by using steepest descent optimization by applying:

$$\hat{x}^{(n+1)} = \hat{x}^n - \beta^{(n)} \nabla E(x) \tag{5.17}$$

where β is the scalar defining the step size of the optimization. Using steepest descent, the iterative solutions for equations 5.14, 5.15 and 5.16 can be expressed as:

$$\hat{x}^{(n+1)} = \hat{x}^n + \beta^{(n)}\{R^T[W_k^T B_k^T D^T(y_k - DB_kW_kR[\hat{x}^n]_k)]_{k=1}^{K} - \lambda L^T L \hat{x}^n\} \tag{5.18}$$

$$\hat{x}^{(n+1)} = \hat{x}^n + \beta^{(n)}\left\{R^T \left[W_k^T B_k^T D^T(y_k - DB_kW_kR[\hat{x}^n]_k)\right]_{k=1}^{K} - \lambda \textstyle\sum_c \, div\left(\frac{\nabla \hat{x}^n}{|\nabla \hat{x}^n|}\right)\right\} \tag{5.19}$$

$$\hat{x}^{(n+1)} = \hat{x}^n + \beta^{(n)}\{R^T[W_k^T B_k^T D^T h'(y_k - DB_kW_kR[\hat{x}^n]_k)]_{k=1}^{K} - \lambda \textstyle\sum_i f'(d_i \hat{x}_n)\} \tag{5.20}$$

where $\nabla \hat{x}^n = \begin{bmatrix} \nabla \hat{x}_h^n \\ \nabla \hat{x}_v^n \end{bmatrix}$ and $|\nabla \hat{x}^n| = \sqrt{\nabla \hat{x}_h^{n\,2} + \nabla \hat{x}_v^{n\,2} + \varepsilon}$ in equation 5.19.

As mentioned earlier, the matrices $W_k, B_k,$ and D can be interpreted as image operations. Therefore, the corresponding transpose matrices W_k^T, B_k^T, and D^T can similarly be interpreted as image operations. If B_k is modeled as point spread function (PSF) kernel, B_k^T could be modeled from those kernels by flipping the columns and rows about the vertical and horizontal axes, respectively. If backward warping (warp from the super-resolution mosaic to the low-resolution frames) is modeled by W_k operator, the corresponding forward warping (warping low-resolution frames to the super-resolution mosaic framework) could be modeled by W_k^T operator. Similarly, D^T would be the interpolation operator corresponding to the decimation operator D.

Instead of using a constant regularization parameter λ, I introduced an adaptive regularization parameter in the super-resolution formulation. Since the super-resolution solution converges to the local minimum, it is desired that the regularization parameter decreases as the iterative procedure progresses. Using the aforementioned adaptive regularization parameter, equation 5.20 can be rewritten as:

$$\hat{x}^{(n+1)} = \hat{x}^n + \beta^{(n)}\{R^T[W_k^T B_k^T D^T h'(y_k - DB_kW_kR[\hat{x}^n]_k)]_{k=1}^K - \lambda^n \textstyle\sum_i f'(d_i\hat{x}_n)\} \quad (5.21)$$

where $\lambda^n = \frac{\sum_{k=1}^K h(y_k - DB_kW_kR[\hat{x}^n]_k)}{K\|\sum_i \mathrm{f}(d_i\hat{x}_n)\|_1}$.

The proposed super-resolution mosaicing method generates mosaic image from a sequence of low-resolution frames and subsequently super-resolves the low-resolution mosaic to produce a high-resolution mosaic. Our method is based on two main algorithms: a mosaicing algorithm and a super-resolution algorithm. The mosaicing algorithm has already been discussed in the earlier chapter. Thus, I will discuss the super-resolution mosaicing algorithm in the following subsection.

Figure 5.1 shows the flowchart of the proposed super-resolution mosaicing algorithm. As shown in the figure, the algorithm first reads ten low-resolution frames and a maximum number of iterations as its inputs. In the following step, the system interpolates those low-resolution frames according to the resolution enhancement factor. Subsequently, the system generates an initial mosaic out of those interpolated low-resolution frames. Then the system reconstructs the low-resolution frames from the initial mosaic using inverse warping (by utilizing the same transformation parameters computed during the registration step of the mosaicing algorithm), then

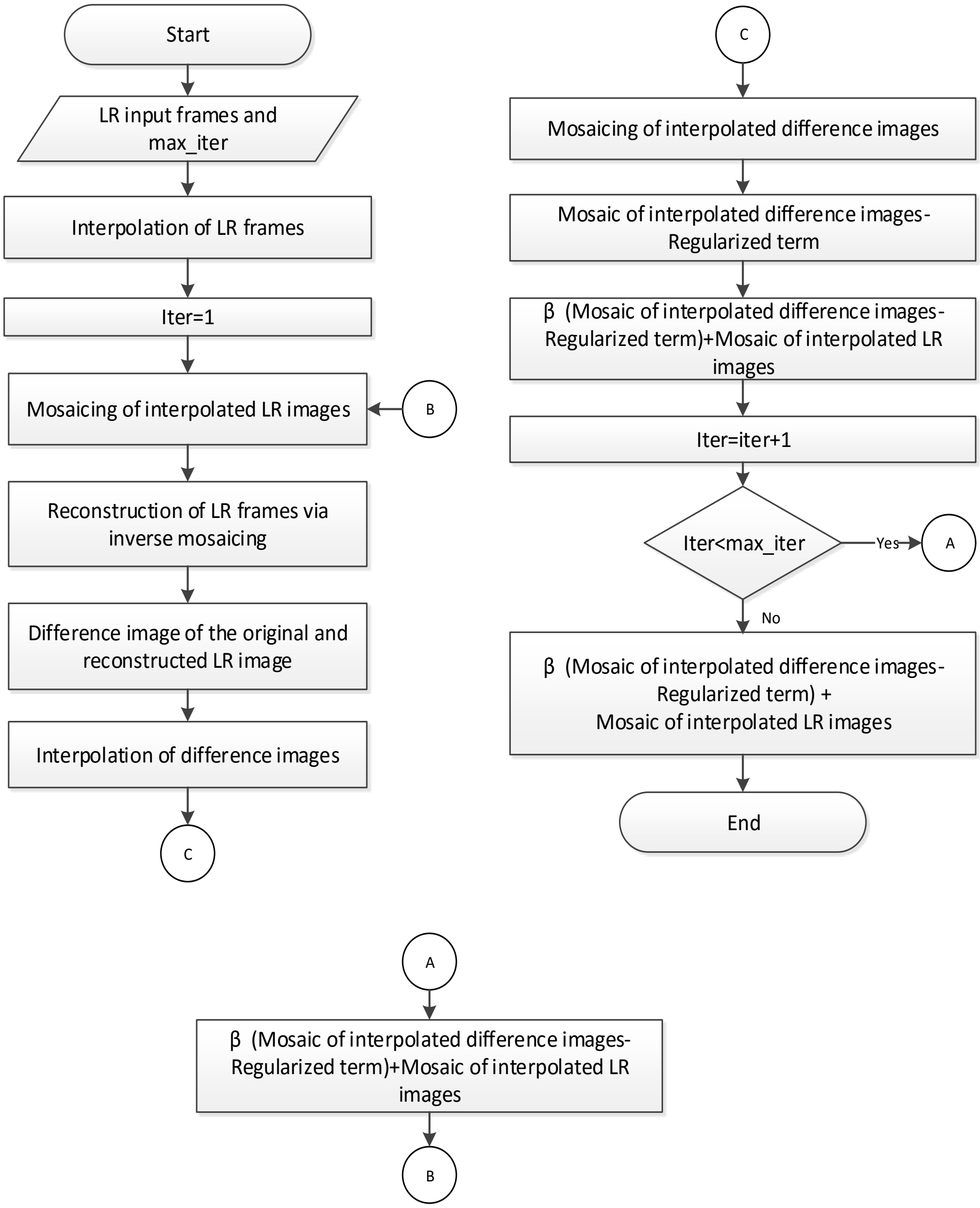


Figure 5.1: Flowchart of super-resolution algorithm.

blurring (by using the flipped kernel of that used while up-sampling the low-resolution frames), and finally down-sampling. The reconstructed low-resolution frames are then used to obtain the difference frames by subtracting them from the input low-resolution frames. Next, a first-order Huber derivative is computed on these difference frames and they are interpolated. These interpolated frames are then used to generate an error mosaic using forward warping and blurring. In the next step, the system subtracts a Huber prior-based adaptive regularization term from the error mosaic. It is then multiplied by the step size of the steepest descent optimization. Finally, this result is used to update the initial low-resolution mosaic as shown in equation 5.13. The algorithm repeats the updating procedure until the maximum number of iterations is reached or until the error between the outputs of two successive iterations falls below a predefined threshold.

The super-resolution program is developed in Microsoft Visual Studio C++ platform. A PC with 1.9 GHz Core2 Duo processor and 3 GB RAM is used for all the evaluations.

5.2 Evaluation

Sixteen datasets are used to assess the performance of the proposed super-resolution mosaicking method. The first eight experiments are simulated experiments with known high-resolution images, and the following eight experiments are the real data experiments with no access to the high-resolution images.

To evaluate the performance of the proposed algorithm, six metrics are used: mean square error (MSE), peak signal-to-noise ratio (PSNR), singular value decomposition (SVD) based measure, structural similarity (SSIM) based metric, metric based on slope of reciprocal singular value (RSV) curve, and cumulative probability of blur detection (CPBD). MSE, PSNR, SVD, and SSIM-based metrics are used to measure the difference between the ground truth and the output of the proposed algorithm. Slope of RSV curve and CPBD-based metrics are used to measure the amount of distortion (especially blurriness) present in the output of the super-resolution mosaicking algorithm. Note that the first four metrics are fully-reference image quality-measurement metrics, whereas, the last two metrics are no-reference image quality measurement metrics. The definitions and interpretations of the aforementioned performance metrics are demonstrated below:

MSE metric: MSE measures the sum of squared differences between the super-resolution mosaic and the original high-resolution mosaic (ground-truth) divided by the number of pixels in each image as [13]:

$$MSE = \frac{\sum_m \sum_n (O(m,n) - GT(m,n))^2}{N} \tag{5.22}$$

where $O(m,n)$ and $GT(m,n)$ are the $(m,n)^{th}$ pixel values of super-resolution mosaic and original high-resolution mosaic. N is the total number of pixels in each of those two images. Lower the value of this metric, better the image quality, and hence better the performance of the super-resolution mosaicking algorithm.

PSNR metric: PSNR measures the ratio of the maximum pixel value between the super-resolution mosaic and the original high-resolution mosaic to the MSE as [4]:

$$PSNR = \frac{10\ log_{10}(\max(O(m,n), GT(m,n)))^2}{MSE} \tag{5.23}$$

Higher the value of this metric, better the performance of the super-resolution mosaicking algorithm.

SVD-based metric: Hypostatic information, which has good stability, can be expressed by SVD of an image. Thus SVD can be used as one of the primary feature of an image. Any m x n rectangular matrix, A, can be decomposed into the product of three matrices, a m×m orthogonal matrix, U, a m×n diagonal matrix, S, and the transpose of a n×n orthogonal matrix, V, as:

$$A = USV^T \tag{5.24}$$

Where $U^T U = I = V^T V$. The columns of U are the eigenvectors of AA^T, the columns of V are the eigenvectors of A^TA and S is a diagonal matrix containing the square roots of eigenvalues from AA^T and A^TA in descending order.

SVD-based metric measures the square root of the sum of squared differences between the corresponding singular values of the super-resolution mosaic and the original high-resolution mosaic as [4]:

$$D_{SVD} = sqrt\left[\sum_{i=1}^{p}(S_{O_i} - S_{GT_i})^2\right] \tag{5.25}$$

where S_{O_i} and S_{GT_i} are the i^{th} singular values of the super-resolution mosaic and the original high-resolution mosaic. p is the total number of singular values is each of those images. Lower the value of this metric, better the image quality, and hence better the performance of the super-resolution mosaicking algorithm.

SSIM metric: SSIM index measures the similarity between the super-resolution mosaic and the original high-resolution mosaic by combining luminance, contrast, and structure comparison functions. The simplified form of SSIM index can be expressed as:

$$SSIM(O, GT) = \frac{(2\mu_O\mu_{GT} + C_1)(2\sigma_{O,GT} + C_2)}{(\mu_O^2 + \mu_{GT}^2 + C_1)(\sigma_O^2 + \sigma_{GT}^2 + C_2)} \tag{5.26}$$

where $\mu_O, \mu_{GT}, \sigma_O, \sigma_{GT},$ and $\sigma_{O,GT}$ are the local means, variances, and cross-covariance for the super-resolution mosaic and the original high-resolution mosaic. Higher the value of this metric, better the performance of the super-resolution mosaicking algorithm.

Slope of RSV curve-based metric: This metric uses the computed singular values of an image to assess the amount of distortion in it, thus it requires no reference image for its computation. The RSV curve is generated by plotting the singular values of an image against the index of the singular vectors. The rate of fall-off of the curve becomes less with larger degrees of blurriness [41]. In contrast, with lower degrees of noise, the rate of fall-off of the curve becomes less. Thus, the rate of fall-off of the RSV curve of an image characterizes the degree of distortion in the image. A higher slope of the RSV curve in the trough region indicates better performance of the proposed algorithm in the presence of blur, and lower slope value of the reciprocal singular value curve in the trough region indicates better performance of the algorithm in the presence of noise.

CPBD-based metric: As SR mosaicing fundamentally increases the resolution of the initial LR mosaic, which requires the use a no-reference image blur metric for quantitative evaluation. Because of its closeness to human blur perception, a metric that calculates the probability of blur detection is necessary such as cumulative probability of blur detection. This metric postulates that the blur around an edge is more or less visible based on the local contrast around that edge. Based on this idea, the probability of blur detection at each edge is computed and then pooled over the entire image to obtain a final quality score.

For a given contrast, the probability of blur detection at an edge $\mathrm{e_i}$ takes the form of a psychometric function as discussed in [32]. This probability is expressed as:

$$p = \sum_{i} p(e_i) = 1 - \exp(-\left|\frac{\omega(e_i)}{\omega JNB(e_i)}\right|^{\beta}) \tag{5.27}$$

where p is the total probability, $p(e_i)$ is the probability of blur detection at an edge e_i, $\omega(e_i)$ is the width of the edge e_i, $\omega JNB(e_i)$ is the width of just noticeable blur (JNB) when the local contrast is C. β is dependent on global contrast of the image. JNB is defined as the minimum amount of perceived blurriness around an edge. β is a parameter which depends on the global contrast of the image. wJNB is measured to be 5 when $C \leq 50$ and 3 when $C \geq 51$. If the actual width of an edge is the same as the JNB width, then $P(e_i) = 63\%$, below which blur is considered to be undetectable. CPBD corresponds to the percentage of edges at which the probability of blur detection is below 63%, i.e. the percentage of edges at which blur cannot be detected. Thus, a higher metric value corresponds to better performance of the algorithm.

5.3 Results and discussion

To assess the relative merits of the proposed methodology, I compare the proposed algorithm with Tikhonov regularization, and TV regularization-based algorithms. Furthermore, for comparison purpose, I use the mosaics without super-resolution that are obtained by mosaicking the bicubic interpolated low-resolution observations.

The algorithms are tested extensively using video data captured by on-board cameras fitted in a high-altitude balloon and a UAV payload. Both the cameras had a frame rate of 30 fps. Initially, still frames are extracted from the video streams. Once still frames are extracted, they are cropped near the borders in order to eliminate the border-abnormalities (e.g. shadows from appending payload or black pixels along the boundaries). Following that, testing datasets are formed by choosing successive frames having moderate (≥ 50%) overlapping. This overlapping criterion guarantees the best performance of the SIFT-based registration algorithm.

The cropped frames mentioned above are treated as high-resolution observations for the simulated experiments. In order to simulate the real low-resolution observations, frames in each dataset are degraded using a series of operations: decimation, convolution with Gaussian blurring kernel, contamination with additive white Gaussian noise. Decimation factor of 1/2 is used for all the datasets. For both types of datasets, the size of the blurring kernel is selected as 3 x 3 pixels and the standard deviation of the kernel is varied over the range [0.4-2.0]. Similarly, for both categories of datasets, noise with zero mean and variance changed within the range [0.0002-0.001] are used. Eight datasets, each containing ten frames, are generated in the aforementioned method to use them for simulated experiments. Other eight datasets, without degrading them in the aforementioned method, are used for the real data experiments.

The atmospheric blur matrix B_k^a is assumed to be an identity matrix I, and the optical blur matrix is chosen as the space-invariant linear kernel : [0.101, 0.117, 0.101; 0.117, 0.125, 0.117; 0.101, 0.117, 0.101]. The resolution enhancement factor is set to be 2 in all the experiments. The value of BBF NN bins is set to 200 to speed up the SIFT matching process without compromising the performance. The threshold parameters of the Huber Norms in our model, the regularization parameters in the Tikhonov and TV models are all adjusted until the best super-resolution mosaicking results are obtained. For the simulated experiments, the best result is selected to be the one with highest SSIM value, whereas for the real data experiments, the best result is selected to be the one with highest CPBD value.

Maximum number of iterations is set to be the termination condition of the steepest descent procedure. With the increase in number of iterations, the algorithm enhances the sharpness of the

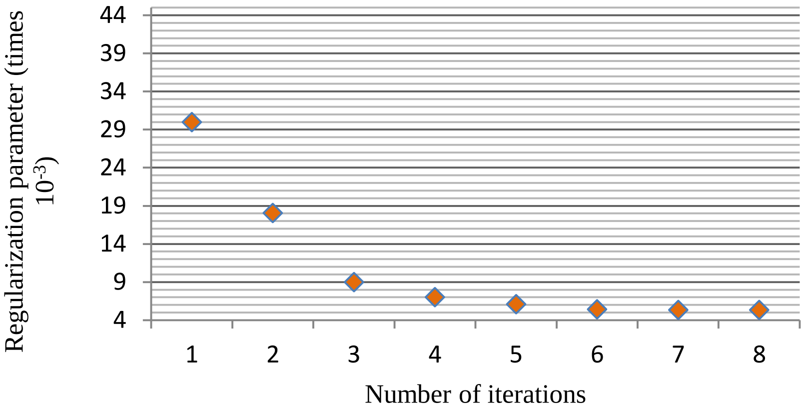
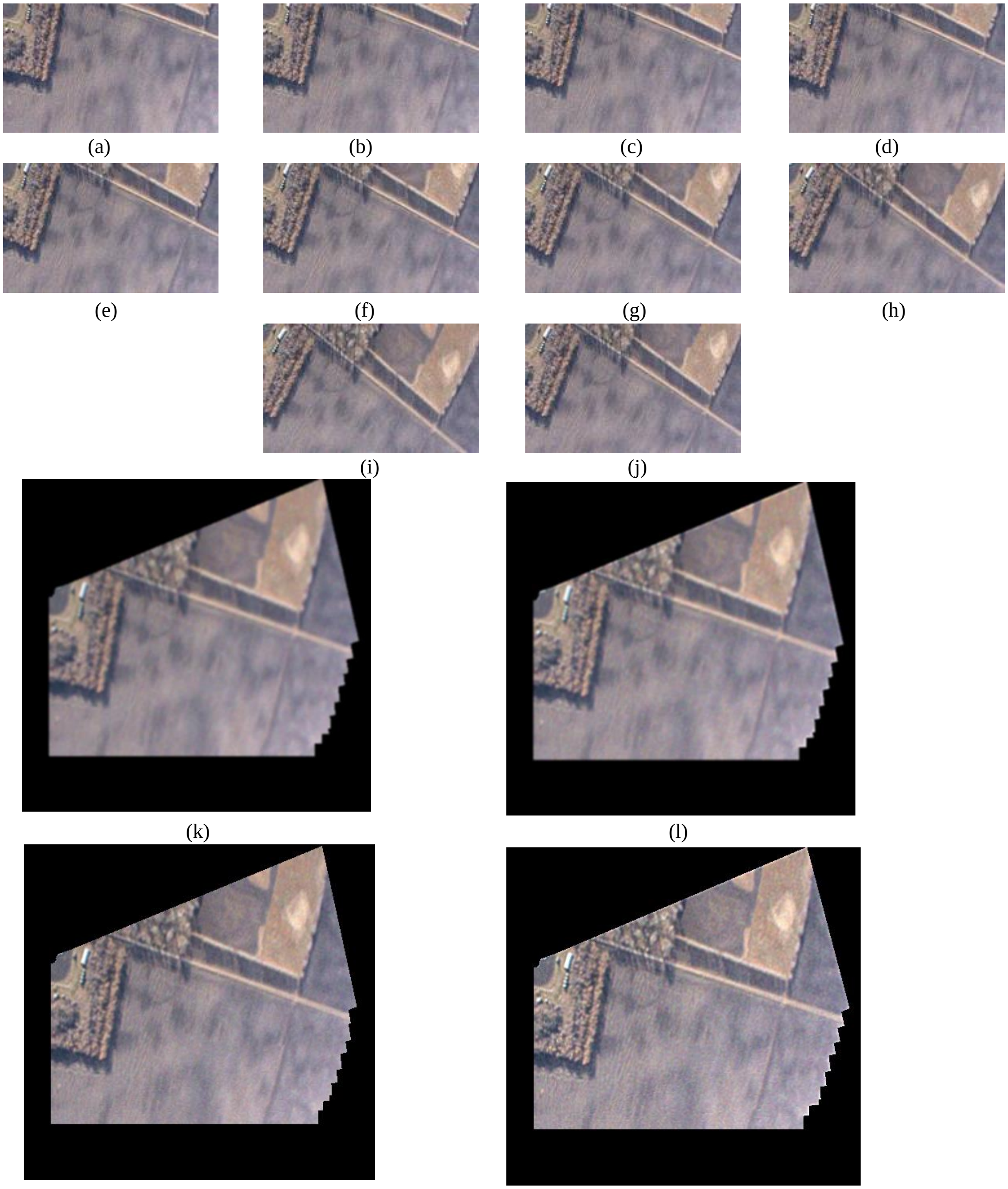


Figure 5.2: Regularization parameter vs number of iterations.

reconstructed image. Extensive testing shows that the quality does not improve after 7 iterations. Thus the maximum number of iterations was set to 7 throughout the testing. As stated in the methodology section, the decrease of the regularization parameter with the algorithm iteration indicates that the algorithm is moving toward a better solution. This was verified on different datasets and an example of results is shown in Figure 5.2. As one can see the regularization parameter has a value of 30x10-3 for one iteration and decreases exponentially as the number of iterations increases. Note that for iteration 7, λ value does not change significantly from that of iteration 6. This is further confirmed at iteration 8, when the value of λ is almost the same as its value at iteration 7. Therefore, in our algorithm I limited the number of iterations to 7.

5.3.1 Results of Simulated Data Experiments

In the simulated experiments, five high-altitude balloon datasets and three UAV datasets are used. The super-resolution mosaicking results using a high-altitude balloon dataset (Figs. 5.3a-j), obtained by contaminating high-resolution images by a Gaussian noise (variance = 0.0008), are presented in Figs. 5.3l- n. Fig. 3k is the mosaic with bicubic-interpolated low-resolution frames.



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

(e) (f) (g) (h)

(i) (j)

(k) (l)

(m) (n)

Figure 5.3: SR result using simulated balloon data. (a)-(j) LR input frames; (k) Mosaic without SR; (l) Mosaic with Tikhonov regularization-based SR; (m) Mosaic with TV regularization-based SR; (n) Mosaic with proposed algorithm

Fig. 5.3l, 5.3m, 5.3n represent the super-resolution mosaicking results of the Tikhonov

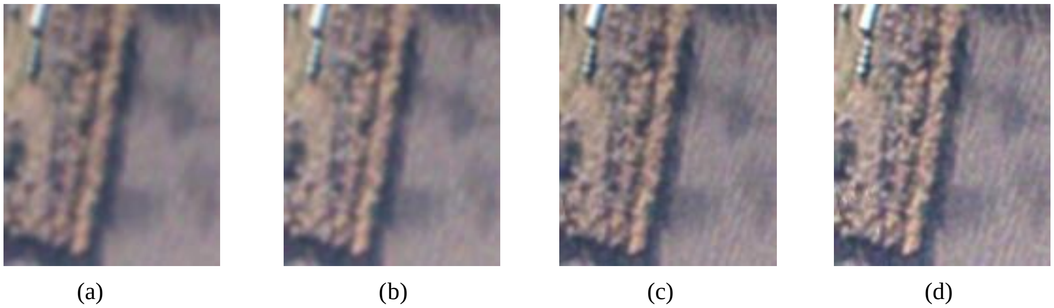


Figure 5.4: Detailed regions cropped from SR results using simulated balloon data. (a) Region from Figure 5.3k; (b) Region from Figure 5.3l; (c) Region from Figure 5.3m; (d) Region from Figure 5.3n.

regularization model, TV regularization model, and the proposed double Huber regularization model, respectively. The detailed regions cropped from Figs. 5.3k-n are presented in Figs. 5.4a, 5.4b, 5.4c, and 5.4d, respectively.

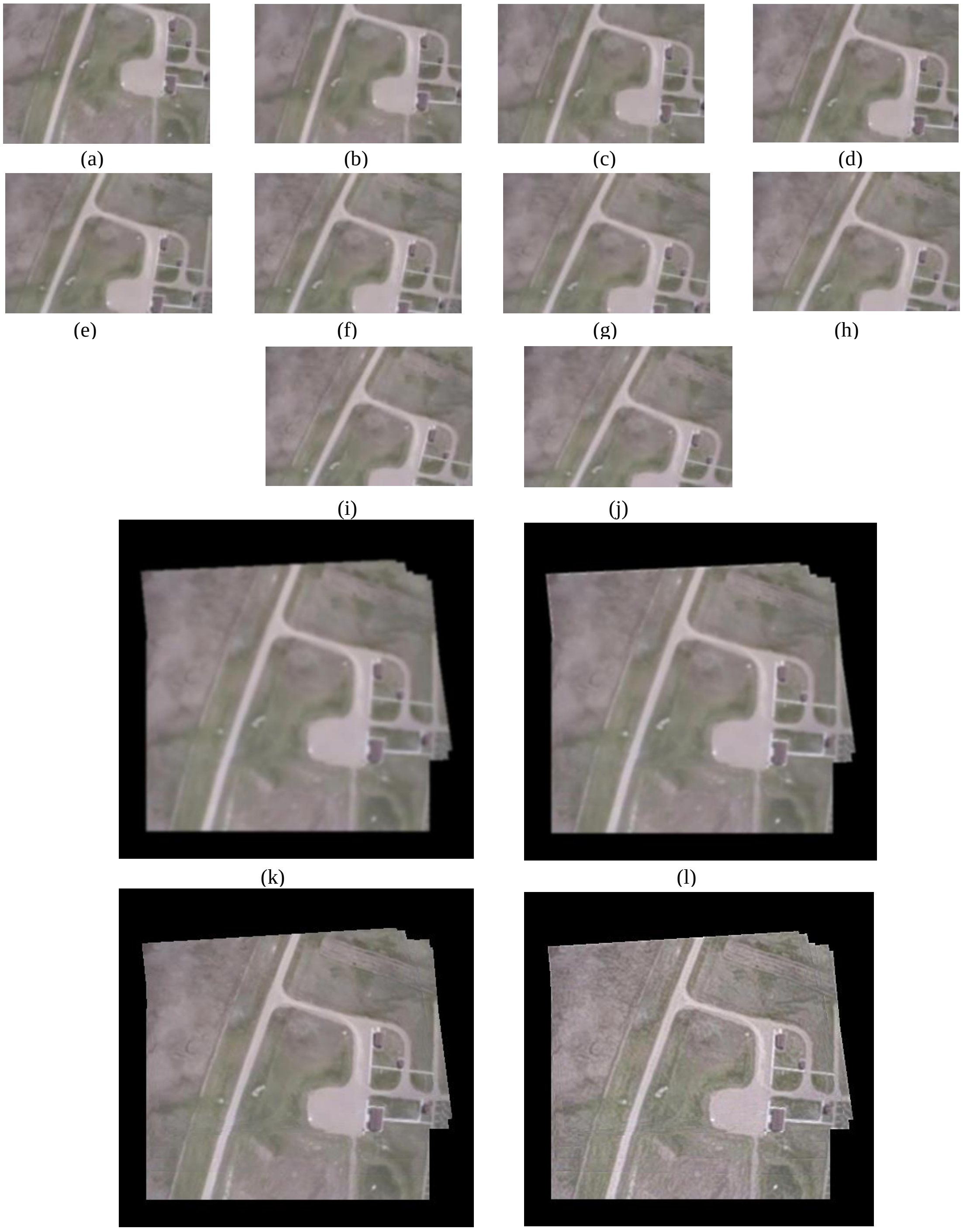
Among the three super-resolution mosaicing algorithms, it is clear that the proposed method produces better visually enhanced result. In the edge area, the detailed information is well preserved. However, the other two comparable methods tend to produce output with overly smooth edges. It can be visualized especially in the detailed regions presented in Fig. 5.4. Since the Tikhonov regularization-based method does not consider the local spatial property in the image while applying the smoothness constraint, it fails to preserve the edges. By using the local image spatial characteristics in the smoothness constraint, TV regularization-based method performs better in preserving edges, and denoising. However, edge blurring is still noticeably present in the output. Clearly, the proposed double Huber-based method performs best in terms of maintaining the tradeoff among preserving edges, deblurring and denoising.

The superior performance of the proposed method is illustrated quantitatively by MSE, PSNR, SVD, and SSIM metrics presented in Table 5.1. It is shown that the proposed method produces results with lowest MSE, SVD values, and highest PSNR, SSIM values, which indicate that our method generates a better super-resolution mosaicking result, close to the original high-resolution mosaic.

Table 5.1: MSE, PSNR, SVD, and SSIM values of different super-resolution mosaicking results using noisy high-altitude balloon frames

	MSE	PSNR (dB)	SVD	SSIM
Bicubic interpolation	22.32	34.64	1511.30	0.3403
Tikhonov	19.66	35.19	1066.20	0.372
TV	17.04	35.82	909.59	0.3968
Proposed method	**14.62**	**36.48**	**651.23**	**0.4071**

The super-resolution mosaicking results using a UAV dataset (Figs. 5.5a- j), obtained by contaminating high-resolution images by a 3x3 Gaussian blurring kernel (SD = 1.2), are presented in Figs. 5.5l- n. Fig. 5.5k is the mosaic with bicubic-interpolated low-resolution frames. Fig. 5.5l, 5.5m, 5.5n represent the super-resolution mosaicking results of the Tikhonov regularization model, TV regularization model, and the proposed double Huber regularization model, respectively. The detailed regions cropped from Figs. 5.5k-n are presented in Figs. 5.6a, 5.6b, 5.6c, and 5.6d, respectively. Noticeably, the Tikhonov regularization-based method generates mosaic with more details compared to the mosaic obtained without using any super-resolution technique. However, significant amount of blurry edges are still present as shown in Fig. 5.6b. TV regularization-based method certainly suppresses a substantial amount of edge blurs; however, blurring artifacts are still present. As can be seen from Fig. 5.6d, edge information is best preserved using the proposed



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

(e) (f) (g) (h)

(i) (j)

(k) (l)

(m) (n)

Figure 5.5: SR result using simulated UAV data. (a)-(j) LR input frames; (k) Mosaic without SR; (l) Mosaic with Tikhonov regularization-based SR; (m) Mosaic with TV regularization-based SR; (n) Mosaic with proposed algorithm.

method when compared to the other two methods. Table 5.2 shows the superiority of the proposed


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

Figure 5.6: Detailed regions cropped from SR results using simulated UAV data. (a) Region from Figure 5.5k; (b) Region from Figure 5.5l; (c) Region from Figure 5.5m; (d) Region from Figure 5.5n.

algorithm. Highest PSNR, SSIM values, and lowest MSE, SVD values are obtained using the proposed algorithm.

Table 5.2: MSE, PSNR, SVD, and SSIM values of different super-resolution mosaicing results using blurry UAV frames

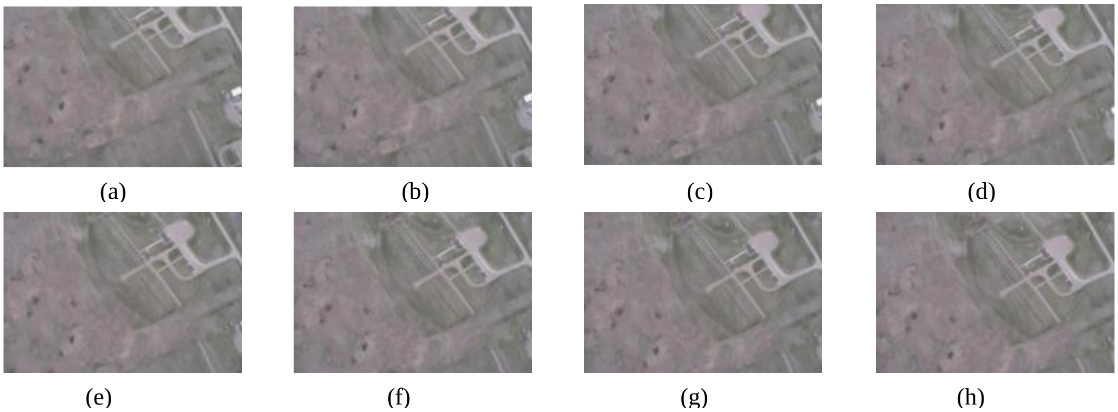
	MSE	PSNR (dB)	SVD	SSIM
Bicubic interpolation	3.56	42.62	1334.8	0.8798
Tikhonov	3.42	42.79	607.66	0.9125
TV	1.80	45.57	472.05	0.9276
Proposed method	**1.43**	**46.57**	**294.37**	**0.9335**

5.3.2 Results of real data experiments

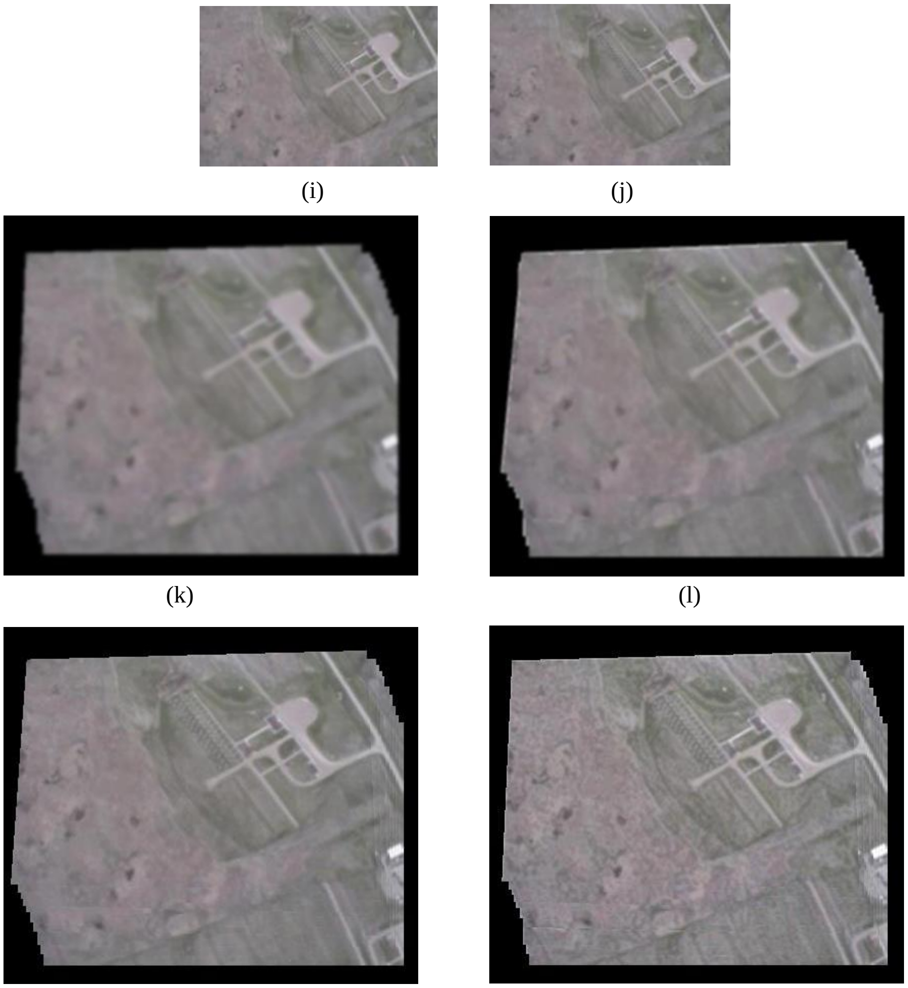
In the real data experiments, five high-altitude balloon datasets and three UAV datasets are used to evaluate the proposed algorithm. The super-resolution mosaicking results using a high-altitude balloon dataset (Figs. 5.7a- j) are shown in Figs. 5.7l- n. Fig. 5.7k is the mosaic output without using super-resolution. Fig. 5.7l, 5.7m, 5.7n represent the super-resolution mosaicking results of the Tikhonov regularization model, TV regularization model, and the proposed double

Huber regularization model, respectively. The detailed regions cropped from Figs. 5.7k-n are presented in Figs. 5.8a, 5.8b, 8c, and 5.8d, respectively.

From these figures, it is obvious that resolution of the mosaic output has certainly increased after using super-resolution techniques, compared to the bicubic-interpolation result. The output of Tikhonov regularization-based method suffers from blurry edges. Although the edge blurring is suppressed to some extent in the TV regularization-based method, details are not well preserved, as can be seen from Fig. 5.8c. Clearly, the edge and other details are best preserved using the proposed algorithm. The quantitative assessment results are presented in Table 5.3. It can be seen that the proposed double Huber-based method generates output with highest CPBD and RSV slope values. Note that the visible degradation present in the real data is blur (both optical blur and motion blur). The amount of noise contamination is insignificant compared to the blur. Thus, output with higher value for the RSV curve's slope essentially indicates lower degradation, and hence better quality of the image.


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

(e) (f) (g) (h)



(i) (j)

(k) (l)

(m) (n)

Figure 5.7: SR result using real UAV data. (a)-(j) LR input frames; (k) Mosaic without SR; (l) Mosaic with Tikhonov regularization-based SR; (m) Mosaic with TV regularization-based SR; (n) Mosaic with proposed algorithm.

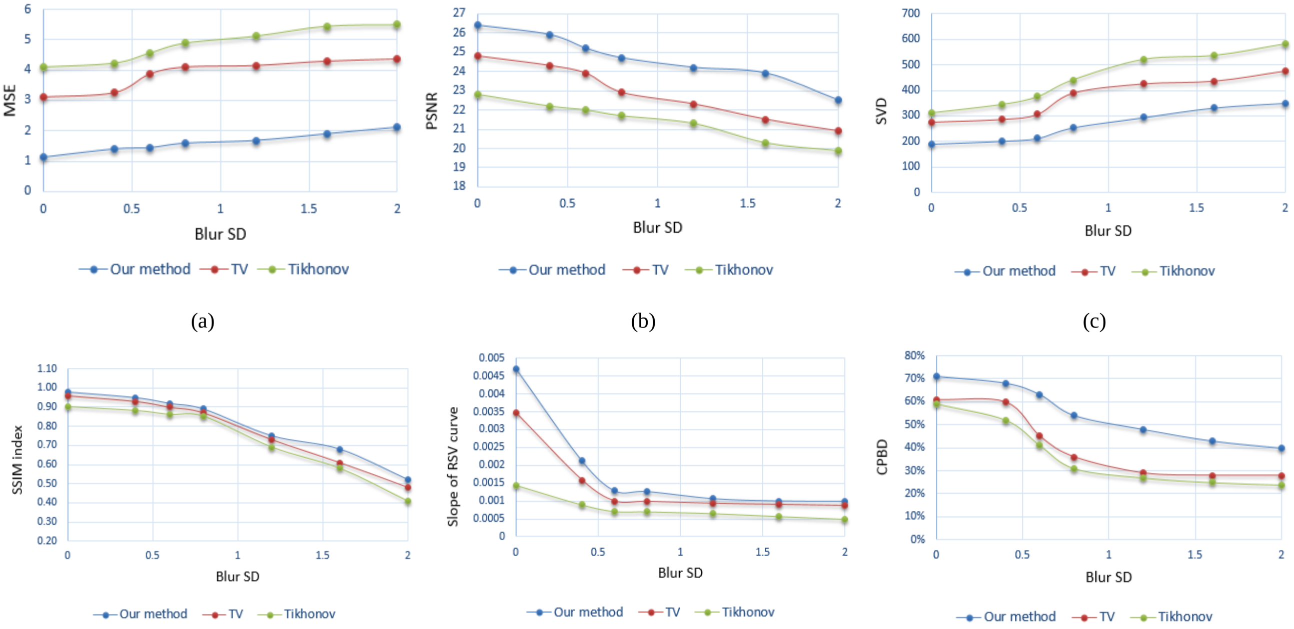
(a) (b) (c) (d)

Figure 5.8: Detailed regions cropped from SR results using real UAV data. (a) Region from Figure 5.7k; (b) Region from Figure 5.7l; (c) Region from Figure 5.7m; (d) Region from Figure 5.7n.

Table 5.3: CPBD and RSV slope values of different super-resolution mosaicing results using real UAV frames

	CPBD	RSV curve slope
Simple interpolation	13.36%	0.01%
Tikhonov	33.81%	0.09%
TV	48.72%	0.42%
Our method	**56.47%**	**0.56%**

Metrics' behaviors with varying standard deviation of the blurring kernel for a single dataset are plotted for Tikhonov regularization, TV regularization, and the proposed method. Typical results are illustrated in Fig. 5.9a, Fig. 5.9b, Fig. 5.9c, Fig. 5.9d, Fig. 5.9e, and Fig. 5.9f for MSE, PSNR, SVD, SSIM, slope of RSV curves, and CPBD, respectively. As the standard deviation of the blurring kernel increases, MSE, and SVD values monotonically increase as shown in Fig. 5.9a and Fig. 5.9c. PSNR, SSIM, slope of RSV curve, and CPBD values behave conversely as the standard deviation of the blurring kernel increases, as shown in Fig. 5.9b, Fig. 5.9d, Fig. 5.9e, and Fig. 5.9f. Thus, behaviors of these plots characterize the consistency of the performance metrics. As lower values of MSE, SVD, and higher values of PSNR, SSIM, slope of RSV curve, CPBD are indicative of the superior performance of a super-resolution mosaicking algorithm, the proposed method clearly outperforms the other two methods since it records lowest values for MSE, SVD and highest values for PSNR, SSIM, RSV curve slopes, CPBD for any given dataset. The TV regularization-based method is seen to outperform the Tikhonov regularization-based technique.



(a)

(b)

(c)

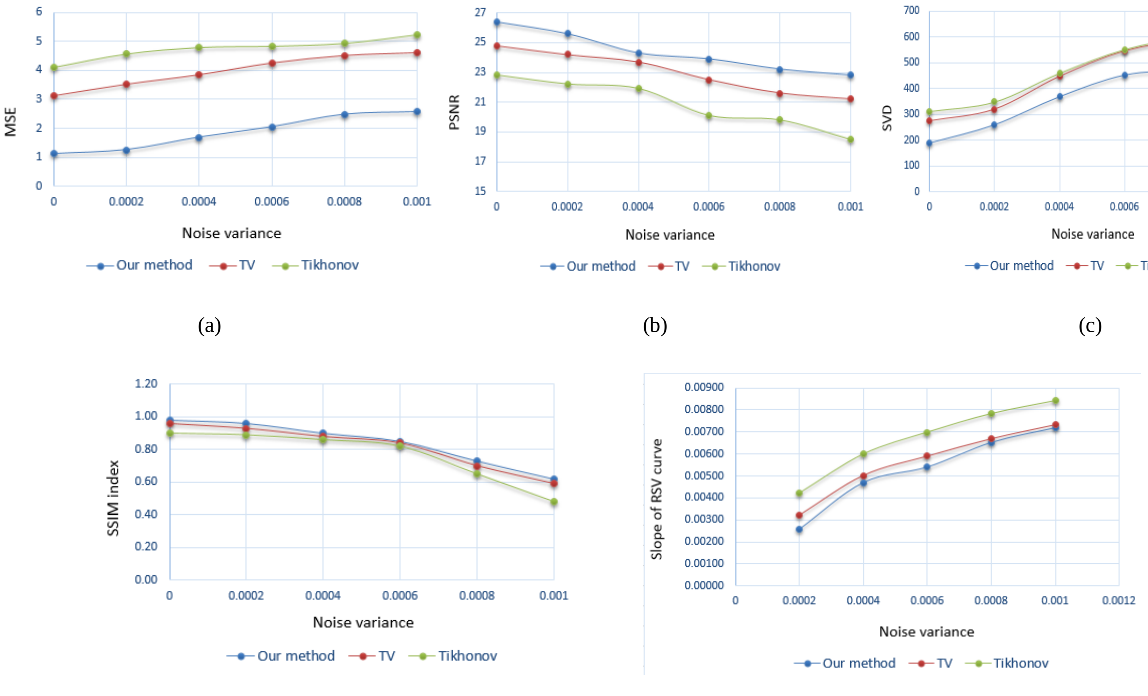
(d)

(e)

(f)

Figure 5.9: Behavior of performance metrics in the presence of blur. (a) MSE behavior; (b) PSNR behavior; (c) SVD behavior; (d) SSIM behavior; (e) behavior of RSV curve's slope; (f) CPBD behavior for a single dataset.

Fig. 5.10 presents the metrics' behaviors for the three methods in the presence of additive Gaussian noise. Fig. 5.10a, Fig. 5.10b, Fig. 5.10c, Fig. 5.10d, and Fig. 5.10e show the behaviors for MSE, PSNR, SVD, and slope of RSV curves, respectively. As the variance of the additive white Gaussian noise increases, MSE, SVD, and slope of RSV curve values monotonically increase, as shown in Fig. 5.10a, Fig. 5.10c, and Fig. 5.10e, respectively. PSNR and SSIM values behave conversely with the increase in variance of the additive white Gaussian noise, as shown in Fig. 5.10b and Fig. 5.10d, respectively. Again, behaviors of these plots show the consistency of the performance metrics. It can be observed from the graphs that the proposed method is superior to the other two, since it generates output with smallest MSE, SVD, slope of RSV curve values, and highest PSNR and SSIM curve values.



(a) (b) (c)

(d) (e)

Figure 5.10: Behavior of performance metrics in the presence of noise. (a) MSE behavior; (b) PSNR behavior; (c) SVD behavior; (d) SSIM behavior; (e) Slope of RSV curve's behavior for a single dataset

The CPBD values of the super-resolution mosaicking algorithm's output (in the presence of Gaussian blur with SD=1) are plotted for bicubic interpolation, Tikhonov regularization, TV regularization, and proposed models as shown in Fig. 5.11. The CPBD values are shown for all the real high-altitude balloon datasets. It is observed that the least amount of blur is perceived using the proposed method, followed by TV regularization, Tikhonov regularization, and bicubic interpolation. Thus, the proposed algorithm based on using the Huber Norm for data fidelity in combination with directional Huber-Markov regularization performs better than the other methods. Also, note that, the TV regularization-based technique predominates over the Tikhonov regularization-based technique. Fig. 5.12 shows the reciprocal singular value curves for a single dataset (in the presence of Gaussian blur with SD=1.2) using the aforementioned four methods. It can be observed that the fastest fall-off of the curve occurs using the proposed method, followed by TV regularization-based method, Tikhonov regularization-based method, and interpolation-based method. The corresponding slope values in the trough regions are 0.014, 0.008, 0.002, and 0.001 for the four methods, which indicates that the proposed method performs best in terms of preserving details in the final output.

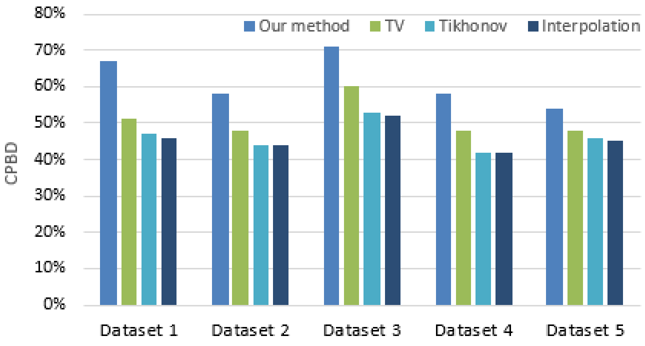


Figure 5.11. Output CPBD values for five real datasets using different super-resolution mosaicking algorithms.

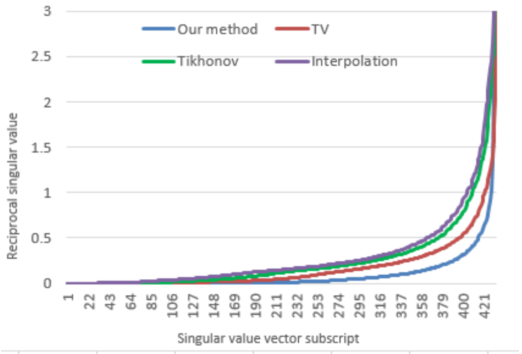


Figure 5.12. Reciprocal singular value curves for a single dataset using different super-resolution mosaicking.

The behavior of the MSE, PSNR, SSIM, and SVD metrics show some inconsistency while evaluating all our datasets. Six percent (6%) inconsistency is observed over total observations for each MSE, PSNR, and SSIM metric, whereas 15% inconsistency is observed over total observations for the SVD metric. However, the CPBD metric values and the slopes of RSV curve values remain consistent throughout the experiments when different distortion types and different distortion levels within a distortion type are evaluated. Furthermore, both of them, being no-reference performance metrics, are preferred choice for several real world applications where ground truths are not available.

Since, a future direction of this research is to implement the proposed algorithm in a small satellite's imaging platform, measuring its processing time and comparing it with the existing methods becomes significant. Table 5.4 shows the processing times of different super-resolution mosaicing methods using a sequence of 10 100 x 67 frames. Note that the amount of blur, noise, and similarity with ground truth are also shown in the table to realize the relative significance of processing time on system performance. The processing time of the proposed algorithm is highest among the three methods. However, the slightly high processing time of the proposed method is paid off by lowest amount of blur, noise, and maximum amount of similarity with ground truth of the output. The amount of blur and the amount of noise in the table indicate CPBD and PSNR measurements, respectively as discussed in section 5.2.

Table 5.4: Comparative processing time of different super-resolution mosaicing methods

	Tikhonov method	Total variation method	Our method
Processing time (lower is better)	**5.5998 sec**	6.143 sec	6.312 sec
Amount of blur (higher is better)	32.8%	42.3%	**57.9%**
Amount of noise (higher is better)	40.72 dB	44.85 dB	**47.19 dB**
Similarity with ground truth (higher is better)	90%	92%	**96%**

CHAPTER 6

CONCLUSIONS

In this dissertation a novel approach for combing multiple overlapping views of a scene into a single, high-resolution still image has been proposed. The technique is intended to be used in small satellite platforms. Thus the aim is to generate high-resolution mosaics over large areas, allowing the broad scale monitoring of the underneath environment for surveillance, urban mapping, and remote sensing purposes, among others.

Hardware cost and sensor-imposed restrictions often limit the imager selection for UAV or satellite’s payload. The contributions of this dissertation concern all the super-resolution mosaicing steps that can generate high quality panorama by accepting the low quality images captured by small satellites’ imagers. The proposed approach investigates two main areas: image mosaicing and super-resolution.

Image mosaicing is the stitching of multiple correlated images into a larger composition. In Chapter 2, the common steps of image mosaicing were described. A comprehensive state-of-the-art on image mosaicing techniques has also been presented in this chapter. Grouping of image mosaicing techniques can be performed based on image registration and image blending. Based on image registration, there are four main categories of image mosaicing methods in the literature: area-based methods, low level feature-based methods, contour-based methods, and frequency domain-based methods. Area-based methods are slow and often require images with significant overlapping for accurate mosaicing output. Low level feature-based methods rely primarily on low level feature extraction algorithms, such as Harris, FAST, SIFT, SURF detectors. These methods offer invariance to several transformations. They are fast and accurate in computation. Contour-based methods are good for large and complicated motions between frames. However, they are computationally expensive since they require finding high level features. Frequency domain-based methods are fast, but require large overlapping among images. Thus, low level feature-based methods become the most adequate strategy to perform image mosaicing, especially when dealing with images captured by satellite's uncontrolled camera motion. Image blending technique is only important when captured images exhibit substantial parallax, which is particularly insignificant for images captured from a high altitude.

Super-resolution is a technique of achieving a high-resolution image from one of multiple low-resolution observations. A comprehensive review of the state-of-the-art super-resolution techniques has been presented in Chapter 3. Two main groups of super-resolution methods in literature are multi-frame methods and single-frame methods. Since this dissertation concerns at developing a multi-frame super-resolution algorithm, I investigated the categories of multi-frame super-resolution methods: frequency domain, interpolation, deterministic regularization, stochastic regularization, IBP, and set theoretic methods. Frequency domain methods are simple but only suitable for translational motion between images. Interpolation and IBP methods are simple but offer no easy way to incorporate prior constraints, which are inevitable to tackle ill-posed nature of super-resolution problem. Deterministic regularization methods often generate overly smooth output. Stochastic regularization methods give easy way to implement edge preserving prior constraint. Set theoretic methods give flexibility to incorporate edge preserving prior constraint; however, they suffer for nonuniqueness of the solution and slow convergence.

Thus, stochastic regularization methods become the best approach to perform multi-frame super-resolution, especially when the aim is to generate output with great deal of edge information.

This dissertation contributes to the state-of-the-art super-resolution mosaicing methods, focusing on the detail enhancement of the generated image mosaics and the computational complexity of the algorithm. The proposed framework includes two major steps mosaicing and super-resolution mosaicing, which are presented in Chapter 4 and Chapter 5, respectively. In Chapter 4, a low level feature-based mosaicing method is proposed, which uses SIFT feature detector. A five step SIFT feature detection algorithm is initially employed: scale-space construction, scale-space extrema detection, keypoint localization, orientation assignment, and defining keypoint descriptors. Once the SIFT features are extracted from a set of input frames and saved into databases, the BBF algorithm (a modified version of the k-d tree) is used to estimate the initial matching points between image pairs. RANSAC algorithm is later used to discard false matches and to estimate an optimum homography matrix based on homography constraints. Using the homography matrices, images are warped into a common coordinate frame. Finally, stitching is employed to obtain the final mosaicing output. The developed mosaicing algorithm is then extensively tested using 36 datasets falling into three categories: images of 2D surface, images of outdoor 3D scenes, airborne images from UAV. In order to quantify the performance of the proposed algorithm, four metrics are used: percentage of mismatches, difference of pixel intensities, peak signal-to-noise ratio, and mutual information. Evaluation shows that the proposed performance metrics are effective in evaluating the quality of mosaicing outputs.

The proposed super-resolution mosaicing approach is presented in Chapter 5. This approach combines a SIFT-based image mosaicing method and a stochastic regularization-based super-resolution method. The low-resolution input frames are initially interpolated. An initial mosaic is then generated using the proposed mosaicing algorithm. After that low-resolution frames are reconstructed from the initial mosaic. An error mosaic is then generated using the difference frames between the input low-resolution frames and the reconstructed low-resolution frames. Finally this error mosaic is used to iteratively update the initial mosaic using a Huber prior-based adaptive regularization. A gradient descent based optimization is used to solve the iterative process. Maximum number of iterations is used as terminating criterion for the iterative procedure. To achieve maximum robustness to the outliers I further proposed the use of Huber Norm for the data fidelity term of the super-resolution formulation. The proposed adaptive regularization guarantees that its value decreases as the iteration proceeds. The developed algorithm has been tested using 16 image sequences falling into two categories: UAV images and high-altitude balloon images. To evaluate the performance of the proposed algorithm, five metrics are used: mean square error (MSE), peak signal-to-noise ratio (PSNR), singular value decomposition (SVD) based measure, metric based on slope of reciprocal singular value (RSV) curve, and cumulative probability of blur detection (CPBD). The results obtained by the proposed approach have been compared with the results obtained by some of the existing super-resolution methods. Our method has demonstrated to be better than the existing techniques.

In this dissertation, I have proposed a super-resolution mosaicing framework. Its functionality and superiority over the other methods are well demonstrated. I consider the future work should include testing my algorithm in small satellite environment, i.e. in limited computing and power resources. The evaluation performed here required feeding the input low-resolution frames manually to the algorithm. Thus the future work must include adding a piece of code so that it automatically triggers the small satellite imaging payload and let it capture low-resolution frames and fed them to the proposed algorithm. This would make the proposed method fully automatic to be used in small satellite environment.

Within the work it was mentioned that blending is not necessary considering the nature of the input frames. However, in some extreme cases when the low-resolution frames are overly degraded by noise and blur and/ or successive frames have huge illumination differences, blending needs to be exploited. Though, such input frames are unlikely to appear frequently, extension of this work should include a geometric and photometric blending algorithm into it.

In order to address the ill-posed nature of the super-resolution problem, we have proposed a directional Huber-Markov regularization. Though, this regularization performs significantly well in recovering edge and other high frequency details, it might be too much computation if the input frames are not severely degraded in quality. Thus, future work could incorporate a model-based regularization, which would analyze the input frames first and then based on the amount of degradation it would select which regularization to use. For low to medium degraded input frames, simple regularization (for example: Tikhonov) could be used. On the other hand, for severe degraded input frames, our proposed regularization could be employed.

Future work would also include investigating the effect of altitude and speed of the imaging payload. At different altitude, the captured frames would have different spatial resolution.

Similarly, the speed of the payload would have direct impact on quality of the captured frames. The results of the proposed super-resolution mosaicing algorithm can be evaluated using frames captured at different altitude as well as with different speed of the payload. The results could then be analyzed to show the super-resolution mosaicing result as a function of altitude as well as the speed of the payload.

REFERENCES

[1] J. Farrell, F. Xiao, and S. Kavusi, "Resolution and light sensitivity tradeoff with pixel size," Electronic Imaging, pp. 60690N-60690N-8, 2006.

[2] P. Milanfar, Super-resolution imaging: CRC Press, 2010.

[3] D. P. Capel, "Image Mosaicing and Superresolution," 2004.

[4] D. Ghosh, S. Park, N. Kaabouch, and W. Semke, "Quantitative evaluation of image mosaicing in multiple scene categories," in Electro/Information Technology (EIT), 2012 IEEE International Conference on, 2012, pp. 1-6.

[5] P. Azzari, L. Di Stefano, and S. Mattoccia, "An Evaluation Methodology for Image Mosaicing Algorithms," in Advanced Concepts for Intelligent Vision Systems. vol. 5259, J. Blanc-Talon, S. Bourennane, W. Philips, D. Popescu, and P. Scheunders, Eds., ed: Springer Berlin Heidelberg, 2008, pp. 89-100.

[6] A. Bevilacqua and P. Azzari, "A Fast and Reliable Image Mosaicing Technique with Application to Wide Area Motion Detection," in Image Analysis and Recognition. vol. 4633, M. Kamel and A. Campilho, Eds., ed: Springer Berlin Heidelberg, 2007, pp. 501-512.

[7] K. S. Bhat, M. Saptharishi, and P. K. Khosla, "Motion detection and segmentation using image mosaics," in Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on, 2000, pp. 1577-1580.

[8] S. T. Sreyas, J. Kumar, and S. Pandey, "Real time mosaicing and change detection system," in Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing, 2012, p. 53.

[9] M. Vivet, B. Martínez, and X. Binefa, "Real-time motion detection for a mobile observer using multiple kernel tracking and belief propagation," in Pattern Recognition and Image Analysis, ed: Springer, 2009, pp. 144-151.

[10] A. E. S. Lucas, "Mosaic-Based Localization And Mapping."

[11] T. Suzuki, Y. Amano, and T. Hashizume, "Vision based localization of a small UAV for generating a large mosaic image," in SICE Annual Conference 2010, Proceedings of, 2010, pp. 2960-2964.

[12] A. Nemra and N. Aouf, "Robust invariant automatic image mosaicing and super resolution for UAV mapping," in Mechatronics and its Applications, 2009. ISMA '09. 6th International Symposium on, 2009, pp. 1-7.

[13] D. Ghosh, N. Kaabouch, and R. A. Fevig, "Robust Spatial-Domain Based Super-Resolution Mosaicing of CubeSat Video Frames: Algorithm and Evaluation," Computer and Information Science, vol. 7, p. p68, 2014.

[14] D. Ghosh, N. Kaabouch, and W. Semke, "Super-Resolution Mosaicing of Unmanned Aircraft System (UAS) Surveillance Video Frames," International Journal of Scientific & Engineering Research, vol. 4, 2013.

[15] P. Azzari, L. Di Stefano, F. Tombari, and S. Mattoccia, "Markerless Augmented Reality Using Image Mosaics," in Image and Signal Processing. vol. 5099, A. Elmoataz, O. Lezoray, F. Nouboud, and D. Mammass, Eds., ed: Springer Berlin Heidelberg, 2008, pp. 413-420.

[16] J. Seokhee and G. J. Kim, "Mosaicing a Wide Geometric Field of View for Effective Interaction in Augmented Reality," in Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on, 2007, pp. 265-266.

[17] M. Irani, S. Hsu, and P. Anandan, "Video compression using mosaic representations," Signal Processing: Image Communication, vol. 7, pp. 529-552, 1995.

[18] C. Yun-Hee, S. Yeong Kyeong, and C. Tae-Sun, "Image mosaicing with automatic scene segmentation for video indexing," in Consumer Electronics, 2002. ICCE. 2002 Digest of Technical Papers. International Conference on, 2002, pp. 74-75.

[19] M. Ramachandran and R. Chellappa, "Stabilization and Mosaicing of Airborne Videos," in Image Processing, 2006 IEEE International Conference on, 2006, pp. 345-348.

[20] L. Miao and Y. Yue, "Automatic document image mosaicing algorithm with hand-held camera," in Intelligent Control and Information Processing (ICICIP), 2011 2nd International Conference on, 2011, pp. 1094-1097.

[21] H. Shejiao, H. Yaling, C. Zonghai, and J. Ping, "Feature-Based Image Automatic Mosaicing Algorithm," in Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on, 2006, pp. 10361-10364.

[22] A. Nasibov, H. Nasibov, and F. Hacizade, "Seamless image stitching algorithm using radiometric lens calibration for high resolution optical microscopy," in Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control, 2009. ICSCCW 2009. Fifth International Conference on, 2009, pp. 1-4.

[23] T. Vercauteren, A. Perchant, X. Pennec, and N. Ayache, "Mosaicing of confocal microscopic in vivo soft tissue video sequences," in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2005, ed: Springer, 2005, pp. 753-760.

[24] R. Miranda-Luna, C. Daul, W. C. Blondel, Y. Hernandez-Mier, D. Wolf, and F. Guillemin, "Mosaicing of bladder endoscopic image sequences: Distortion calibration and registration algorithm," Biomedical Engineering, IEEE Transactions on, vol. 55, pp. 541-553, 2008.

[25] G. Gao and K. Jia, "A New Image Mosaics Algorithm Based on Feature Points Matching," in Innovative Computing, Information and Control, 2007. ICICIC '07. Second International Conference on, 2007, pp. 471-471.

[26] K.-i. Okumura, S. Raut, Q. Gu, T. Aoyama, T. Takaki, and I. Ishii, "Real-time feature-based video mosaicing at 500 fps," in Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on, 2013, pp. 2665-2670.

[27] T. Botterill, S. Mills, and R. Green, "Real-time aerial image mosaicing," in Image and Vision Computing New Zealand (IVCNZ), 2010 25th International Conference of, 2010, pp. 1-8.

[28] L. Yao, "Image mosaic based on SIFT and deformation propagation," in 2008 IEEE International Symposium on Knowledge Acquisition and Modeling Workshop, 2008, pp. 848-851.

[29] G. Jun-Hui, Z. Jun-Hua, A. Zhen-Zhou, Z. Wei-Wei, and L. Hui-Min, "An approach for X-ray image mosaicing based on Speeded-Up Robust Features," in Wavelet Active Media

Technology and Information Processing (ICWAMTIP), 2012 International Conference on, 2012, pp. 432-435.

[30] F. Yang, L. Wei, Z. Zhang, and H. Tang, "Image mosaic based on phase correlation and Harris operator," Journal of Computational Information Systems, vol. 8, pp. 2647-2655, 2012.

[31] M. Vivet, S. Peleg, and X. Binefa, "Real-Time Stereo Mosaicing using Feature Tracking," in Multimedia (ISM), 2011 IEEE International Symposium on, 2011, pp. 577-582.

[32] A. Levin, A. Zomet, S. Peleg, and Y. Weiss, "Seamless image stitching in the gradient domain," in Computer Vision-ECCV 2004, ed: Springer, 2004, pp. 377-389.

[33] M. El-Saban, M. Izz, A. Kaheel, and M. Refaat, "Improved optimal seam selection blending for fast video stitching of videos captured from freely moving devices," in Image Processing (ICIP), 2011 18th IEEE International Conference on, 2011, pp. 1481-1484.

[34] S. Park, D. Ghosh, N. Kaabouch, R. A. Fevig, and W. Semke, "Hierarchical multi-level image mosaicing for autonomous navigation of UAV," in IS&T/SPIE Electronic Imaging, 2012, pp. 830116-830116-9.

[35] S. Z. Kovalsky, G. Cohen, and J. M. Francos, "Registration of joint geometric and radiometric image deformations in the presence of noise," in Statistical Signal Processing, 2007. SSP'07. IEEE/SP 14th Workshop on, 2007, pp. 561-565.

[36] D. Vaghela and K. Naina, "A Review of Image Mosaicing Techniques," arXiv preprint arXiv:1405.2539, 2014.

[37] P. Jain and V. K. Shandliya, "A review paper on various approaches for image mosaicing."

[38] R. Abraham and P. Simon, "Review on mosaicing techniques in image processing," in Advanced Computing and Communication Technologies (ACCT), 2013 Third International Conference on, 2013, pp. 63-68.

[39] H. Joshi and M. K. Sinha, "A survey on image mosaicing techniques," 2013.

[40] M. H. M. Patel, A. P. P. J. Patel, and A. P. M. S. G. Patel, "Comprehensive Study And Review Of Image Mosaicing Methods," in International Journal of Engineering Research and Technology, 2012.

[41] J. M. Fitzpatrick, D. L. Hill, and C. R. Maurer Jr, "Image registration."

[42] S. Ghannam and A. L. Abbott, "Cross Correlation versus Mutual Information for Image Mosaicing," International Journal, 2013.

[43] R. Szeliski, "Image alignment and stitching: A tutorial," Foundations and Trends® in Computer Graphics and Vision, vol. 2, pp. 1-104, 2006.

[44] K. Berberidis and I. Karybali, "A new efficient cross-correlation based image registration technique with improved performance," in Proceedings of the 11th European Signal Processing Conference (EUSIPCO 2002), pp. 3-6.

[45] F. Zhao, Q. Huang, and W. Gao, "Image matching by normalized cross-correlation," in Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on, 2006, pp. II-II.

[46] T. Vercauteren, A. Meining, F. Lacombe, and A. Perchant, "Real time autonomous video image registration for endomicroscopy: fighting the compromises," in Biomedical Optics (BiOS) 2008, 2008, pp. 68610C-68610C-8.

[47] A. Dame and E. Marchand, "Video mosaicing using a mutual information-based motion estimation process," in Image Processing (ICIP), 2011 18th IEEE International Conference on, 2011, pp. 1493-1496.

[48] C. de Cesare, M.-J. Rendas, A.-G. Allais, and M. Perrier, "Low overlap image registration based on both entropy and mutual information measures," in OCEANS 2008, 2008, pp. 1-9.

[49] M. B. Islam and M. M. J. Kabir, "A New Feature-Based Image Registration Algorithm."

[50] V. S. Bind, "Robust Techniques for Feature-based Image Mosaicing," National Institute of Technology Rourkela, 2013.

[51] E. Zagrouba, W. Barhoumi, and S. Amri, "An efficient image-mosaicing method based on multifeature matching," Machine Vision and Applications, vol. 20, pp. 139-162, 2009/04/01 2009.

[52] X. Wang, J. Sun, and H.-Y. Peng, "Efficient Panorama Mosaicing Based on Enhanced-FAST and Graph Cuts," in Recent Advances in Computer Science and Information Engineering. vol. 128, Z. Qian, L. Cao, W. Su, T. Wang, and H. Yang, Eds., ed: Springer Berlin Heidelberg, 2012, pp. 757-762.

[53] J. Jiao, B. Zhao, and S. Wu, "A speed-up and robust image registration algorithm based on fast," in Computer Science and Automation Engineering (CSAE), 2011 IEEE International Conference on, 2011, pp. 160-164.

[54] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International journal of computer vision, vol. 60, pp. 91-110, 2004.

[55] D. Liqian and J. Yuehui, "Moon landform images fusion and Mosaic based on SIFT method," in Computer and Information Application (ICCIA), 2010 International Conference on, 2010, pp. 29-32.

[56] Y. Li, Y. Wang, W. Huang, and Z. Zhang, "Automatic image stitching using sift," in Audio, Language and Image Processing, 2008. ICALIP 2008. International Conference on, 2008, pp. 568-571.

[57] Y. Lei, W. Xiaoyu, Z. Jun, and L. Hui, "A research of feature-based image mosaic algorithm," in Image and Signal Processing (CISP), 2011 4th International Congress on, 2011, pp. 846-849.

[58] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in Computer Vision–ECCV 2006, ed: Springer, 2006, pp. 404-417.

[59] N. Geng, D. He, and Y. Song, "Camera Image Mosaicing Based on an Optimized SURF Algorithm," TELKOMNIKA Indonesian Journal of Electrical Engineering, vol. 10, pp. 2183-2193, 2012.

[60] R. Wen, C. Hui, L. Jiaju, X. Yanyan, and R. Haeusler, "Mosaicing of microscope images based on SURF," in Image and Vision Computing New Zealand, 2009. IVCNZ '09. 24th International Conference, 2009, pp. 271-275.

[61] H. JOSHI and K. SINHA, "NOVEL TECHNIQUE IMAGE MOSAICING BASED ON IMAGE FUSION USING HARRIS AND SURF."

[62] V. S. Bind, P. R. Muduli, and U. C. Pati, "A Robust Technique for Feature-based Image Mosaicing using Image Fusion," 2013.

[63] K. Peng and M. Hongbing, "An automatic airborne image mosaicing method based on the SIFT feature matching," in Multimedia Technology (ICMT), 2011 International Conference on, 2011, pp. 155-159.

[64] J. Zhu and M. Ren, "Image Mosaic Method Based on SIFT Features of Line Segment," Computational and mathematical methods in medicine, vol. 2014, 2014.

[65] J. Xiao, Y. Zhang, and M. Shah, "Adaptive region-based video registration," in Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1. Seventh IEEE Workshops on, 2005, pp. 215-220.

[66] J. Prescott, M. Clary, G. Wiet, T. Pan, and K. Huang, "Automatic registration of large set of microscopic images using high-level features," in Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on, 2006, pp. 1284-1287.

[67] M. Deshmukh and U. Bhosle, "A survey of image registration," International Journal of Image Processing (IJIP), vol. 5, p. 245, 2011.

[68] V. R and A. R, "Implementation of FFT Based Automatic Image Mosaicing," International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, vol. 2, 2013.

[69] C. Wang, Y. Cheng, and C. Zhao, "Robust subpixel registration for image mosaicing," in Pattern Recognition, 2009. CCPR 2009. Chinese Conference on, 2009, pp. 1-5.

[70] R. Prados Gutiérrez, "Image blending techniques and their application in underwater mosaicing," University of Girona, 2013.

[71] D. K. Jain, G. Saxena, and V. K. Singh, "Image Mosaicing Using Corner Techniques," in Communication Systems and Network Technologies (CSNT), 2012 International Conference on, 2012, pp. 79-84.

[72] Y. Xiong and K. Turkowski, "Registration, calibration and blending in creating high quality panoramas," in Applications of Computer Vision, 1998. WACV'98. Proceedings., Fourth IEEE Workshop on, 1998, pp. 69-74.

[73] P. Liang, X. Zhiwei, and D. Jiguang, "Joint edge detector based on Laplacian pyramid," in Image and Signal Processing (CISP), 2010 3rd International Congress on, 2010, pp. 978-982.

[74] A. Pandey and U. C. Pati, "A novel technique for non-overlapping image mosaicing based on pyramid method," in India Conference (INDICON), 2013 Annual IEEE, 2013, pp. 1-6.

[75] Y. Xiong, "Eliminating ghosting artifacts for panoramic images," in Multimedia, 2009. ISM'09. 11th IEEE International Symposium on, 2009, pp. 432-437.

[76] R. Szeliski, M. Uyttendaele, and D. Steedly, "Fast Poisson blending using multi-splines," in Computational Photography (ICCP), 2011 IEEE International Conference on, 2011, pp. 1-8.

[77] P. Nyman, "Image Stitching using Watersheds and Graph Cuts."

[78] H. Wen and J. Zhou, "An improved algorithm for image mosaic," in Information Science and Engineering, 2008. ISISE'08. International Symposium on, 2008, pp. 497-500.

[79] C. H. Genitha and K. Vani, "Super resolution mapping of satellite images using Hopfield neural networks," in Recent Advances in Space Technology Services and Climate Change (RSTSCC), 2010, 2010, pp. 114-118.

[80] H. Zhang, Z. Yang, L. Zhang, and H. Shen, "Super-resolution reconstruction for multi-angle remote sensing images considering resolution differences," Remote Sensing, vol. 6, pp. 637-657, 2014.

[81] R. Willett, I. Jermyn, R. Nowak, and J. Zerubia, "Wavelet-based superresolution in astronomy," in Astronomical Data Analysis Software and Systems (ADASS) XIII, 2004, p. 107.

[82] D. Kouame and M. Ploquin, "Super-resolution in medical imaging: An illustrative approach through ultrasound," in Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on, 2009, pp. 249-252.

[83] D.-H. Trinh, M. Luong, F. Dibos, J. Rocchisani, C. Pham, and T. Q. Nguyen, "Novel example-based method for super-resolution and denoising of medical images," 2014.

[84] Y.-H. Wang, J.-B. Li, and P. Fu, "Medical image super-resolution analysis with sparse representation," in Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on, 2012, pp. 106-109.

[85] T. Bengtsson, T. McKelvey, and I.-H. Gu, "Super-Resolution reconstruction of High Dynamic Range images with perceptual weighting of errors," in Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, 2013, pp. 2212-2216.

[86] B. Li, H. Chang, S. Shan, and X. Chen, "Low-resolution face recognition via coupled locality preserving mappings," Signal Processing Letters, IEEE, vol. 17, pp. 20-23, 2010.

[87] N. H. Farhat, "Microwave diversity imaging and automated target identification based on models of neural networks," Proceedings of the IEEE, vol. 77, pp. 670-681, 1989.

[88] T. Yoshida, T. Takahashi, D. Deguchi, I. Ide, and H. Murase, "Robust face super-resolution using free-form deformations for low-quality surveillance video," in Multimedia and Expo (ICME), 2012 IEEE International Conference on, 2012, pp. 368-373.

[89] M. Sugie, S. Gohshi, H. Takeshita, and C. Mori, "Subjective assessment of super-resolution 4K video using paired comparison," in Intelligent Signal Processing and Communication Systems (ISPACS), 2014 International Symposium on, 2014, pp. 042-047.

[90] J. Farrell, F. Xiao, and S. Kavusi, "Resolution and light sensitivity tradeoff with pixel size," in Electronic Imaging 2006, 2006, pp. 60690N-60690N-8.

[91] J. Tian and K.-K. Ma, "A survey on super-resolution imaging," Signal, Image and Video Processing, vol. 5, pp. 329-342, 2011.

[92] R. S. Babu and K. S. Murthy, "A survey on the methods of super-resolution image reconstruction," International Journal of Computer Applications, vol. 15, pp. 1-6, 2011.

[93] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," Signal Processing Magazine, IEEE, vol. 20, pp. 21-36, 2003.

[94] S. Borman and R. L. Stevenson, "Super-resolution from image sequences-a review," in Circuits and Systems, Midwest Symposium on, 1998, pp. 374-374.

[95] A. J. Shah and S. B. Gupta, "Image super resolution-A survey," in Emerging Technology Trends in Electronics, Communication and Networking (ET2ECN), 2012 1st International Conference on, 2012, pp. 1-6.

[96] R. Scaria and J. Yomas, "A Survey of Single Image and Multi Image Super Resolution Techniques."

[97] R. Macwan, N. Patel, P. Prajapati, and J. Chavda, "A Survey on Various Techniques of Super Resolution Imaging," International Journal of Computer Applications, vol. 90, pp. 19-22, 2014.

[98] P. Hardeep, P. B. Swadas, and M. Joshi, "A survey on techniques and challenges in image super resolution reconstruction," Int. J. Comput. Sci. Mob. Comput, vol. 2, pp. 317-325, 2013.

[99] K. Shah, J. Pandya, and S. Vahora, "A Survey On Super Resolution Image Reconstruction Techniques," in International Journal of Engineering Research and Technology, 2013.

[100] N. Panchal, B. Limbasiya, and A. Prajapati, "Survey On Multi-Frame Image Super-Resolution," International Journal of Scientific & Technology Research, vol. 2, 2013.

[101] J. Van Ouwerkerk, "Image super-resolution survey," Image and Vision Computing, vol. 24, pp. 1039-1052, 2006.

[102] E. Karimi, K. Kangarloo, and S. Javadi, "A Survey on Super-Resolution Methods for Image Reconstruction," International Journal of Computer Applications, vol. 90, pp. 32-39, 2014.

[103] K. Nasrollahi and T. B. Moeslund, "Super-resolution: a comprehensive survey," Machine vision and applications, vol. 25, pp. 1423-1468, 2014.

[104] Z. Wang and F. Qi, "Analysis of multiframe super-resolution reconstruction for image anti-aliasing and deblurring," Image and Vision Computing, vol. 23, pp. 393-404, 2005.

[105] F. Humblot and A. Mohammad-Djafari, "Super-resolution using hidden Markov model and Bayesian detection estimation framework," EURASIP Journal on Applied Signal Processing, vol. 10, p. ID, 2006.

[106] S. P. Belekos, N. P. Galatsanos, and A. K. Katsaggelos, "Maximum a posteriori super-resolution of compressed video with a novel multichannel image prior and a new observation model," in Proc. EUSIPCO, 2011, pp. 293-297.

[107] H. Zhang, L. Zhang, and H. Shen, "A super-resolution reconstruction algorithm for hyperspectral images," Signal Processing, vol. 92, pp. 2082-2096, 2012.

[108] G. Rochefort, F. Champagnat, G. Le Besnerais, and J.-F. Giovannelli, "An improved observation model for super-resolution under affine motion," Image Processing, IEEE Transactions on, vol. 15, pp. 3325-3337, 2006.

[109] P. Gajjar and M. Joshi, "Zoom based super-resolution: A fast approach using particle swarm optimization," in Image and Signal Processing, ed: Springer, 2010, pp. 63-70.

[110] Y. Tian and K.-H. Yap, "Multi-frame super-resolution from observations with zooming motion," in Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, 2012, pp. 1257-1260.

[111] A. Agrawal and R. Raskar, "Resolving objects at higher resolution from a single motion-blurred image," in Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, 2007, pp. 1-8.

[112] S. Basavaraja, A. S. Bopardikar, and S. Velusamy, "Detail warping based video super-resolution using image guides," in Proceedings of IEEE International Conference on Image Processing, Hong Kong, 2010, pp. 2009-2012.

[113] R. Tsai and T. S. Huang, "Multiframe image restoration and registration," Advances in computer vision and Image Processing, vol. 1, pp. 317-339, 1984.

[114] S. Kim, N. K. Bose, and H. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 38, pp. 1013-1027, 1990.

[115] N. Bose, H. Kim, and H. 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, 1993, pp. 269-272.

[116] N. Bose, H. Kim, and B. Zhou, "Performance analysis of the TLS algorithm for image reconstruction from a sequence of undersampled noisy and blurred frames," in Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference, 1994, pp. 571-574.

[117] W. Yu-Su and S. P. Kim, "High-resolution restoration of dynamic image sequences," International Journal of Imaging Systems and Technology, vol. 5, pp. 330-339, 1994.

[118] B. C. Tom, A. K. Katsaggelos, and N. 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, 1994, pp. 553-557.

[119] N. A. Woods, N. P. Galatsanos, and A. K. Katsaggelos, "Stochastic methods for joint registration, restoration, and interpolation of multiple undersampled images," Image Processing, IEEE Transactions on, vol. 15, pp. 201-213, 2006.

[120] S. Rhee and M. G. Kang, "DCT-based regularized algorithm for high-resolution image reconstruction," in Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on, 1999, pp. 184-187.

[121] P. Vandewalle, S. Sü, and M. Vetterli, "A frequency domain approach to registration of aliased images with application to super-resolution," EURASIP Journal on Advances in Signal Processing, vol. 2006, 2006.

[122] G. Anbarjafari and H. Demirel, "Image super resolution based on interpolation of wavelet domain high frequency subbands and the spatial domain input image," ETRI journal, vol. 32, pp. 390-394, 2010.

[123] F. Arif and T. Sarwar, "Super-Resolution Using Edge Modification through Stationary Wavelet Transform," in Information Visualisation (IV), 2014 18th International Conference on, 2014, pp. 357-360.

[124] G. Sneha and T. R. T Ramakrishnaiah, "A DWT-SWT based Image Super Resolution with Multi Surface Fitting," International Journal of Computer Applications, vol. 103, pp. 9-13, 2014.

[125] A. Garg, S. V. Naidu, H. Yahia, and D. Singh, "Wavelet Based Resolution Enhancement for Low Resolution Satellite Images," in 9th IEEE International Conference on Industrial and Information Systems (ICIIS2014), 2014.

[126] H. Demirel, G. Anbarjafari, C. Ozcinar, and S. Izadpanahi, "Video resolution enhancement by using complex wavelet transform," in Image Processing (ICIP), 2011 18th IEEE International Conference on, 2011, pp. 2093-2096.

[127] S. Izadpanahi and H. Demirel, "Multi-frame super resolution using edge directed interpolation and complex wavelet transform," 2012.

[128] L. Liyakathunisa and C. Kumar, "A novel wavelet based super resolution reconstruction of low resolution images using adaptive interpolation," in Advances in Recent Technologies in Communication and Computing (ARTCom), 2010 International Conference on, 2010, pp. 146-150.

[129] M. D. Robinson, C. A. Toth, J. Y. Lo, and S. Farsiu, "Efficient Fourier-wavelet super-resolution," Image Processing, IEEE Transactions on, vol. 19, pp. 2669-2681, 2010.

[130] S. A. Devi and A. Vasuki, "Image Super Resolution using Fourier-Wavelet transform," in 2012 International Conference on Machine Vision and Image Processing (MVIP).

[131] M. Zhong and Z. Xinbo, "Super-resolution from multiframe X-ray images," in Complex Medical Engineering, 2009. CME. ICME International Conference on, 2009, pp. 1-6.

[132] H. Ur and D. Gross, "Improved resolution from subpixel shifted pictures," CVGIP: Graphical Models and Image Processing, vol. 54, pp. 181-186, 1992.

[133] A. Panagiotopoulou and V. Anastassopoulos, "Super-resolution image reconstruction employing Kriging interpolation technique," in Systems, Signals and Image Processing, 2007 and

6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services. 14th International Workshop on, 2007, pp. 144-147.

[134] Y. Xuefeng, L. Jinzong, L. Dongdong, and Z. Bing, "A BPNN based two-step image super-resolution reconstruction method," in Signal Processing Systems (ICSPS), 2010 2nd International Conference on, 2010, pp. V1-595-V1-598.

[135] F. Zhou, W. Yang, and Q. Liao, "Interpolation-based image super-resolution using multisurface fitting," Image Processing, IEEE Transactions on, vol. 21, pp. 3312-3318, 2012.

[136] A. Gilman, D. G. Bailey, and S. Marsland, "Least-squares optimal interpolation for fast image super-resolution," in Electronic Design, Test and Application, 2010. DELTA'10. Fifth IEEE International Symposium on, 2010, pp. 29-34.

[137] R. Hardie, "A fast image super-resolution algorithm using an adaptive Wiener filter," Image Processing, IEEE Transactions on, vol. 16, pp. 2953-2964, 2007.

[138] M. Bevilacqua, "Algorithms for super-resolution of images and videos based on learning methods," Rennes 1, 2014.

[139] X. Zhang, E. Y. Lam, E. X. Wu, and K. K. Wong, "Application of Tikhonov regularization to super-resolution reconstruction of brain MRI images," in Medical Imaging and Informatics, ed: Springer, 2008, pp. 51-56.

[140] Y. Park and S. I. Yoo, "Isotropic Huber MRFS for structure super-resolution," in Image Processing (ICIP), 2011 18th IEEE International Conference on, 2011, pp. 1137-1140.

[141] H. He and L. P. Kondi, "MAP based resolution enhancement of video sequences using a Huber-Markov random field image prior model," in Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on, 2003, pp. II-933-6 vol. 3.

[142] R. R. Schultz and R. L. Stevenson, "A Bayesian approach to image expansion for improved definition," Image Processing, IEEE Transactions on, vol. 3, pp. 233-242, 1994.

[143] T. Katsuki and M. Inoue, "Posterior mean super-resolution with a compound Gaussian Markov random field prior," in Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, 2012, pp. 841-844.

[144] M. K. Ng, H. Shen, E. Y. Lam, and L. Zhang, "A total variation regularization based super-resolution reconstruction algorithm for digital video," EURASIP Journal on Advances in Signal Processing, vol. 2007, 2007.

[145] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," SIAM Journal on Imaging Sciences, vol. 1, pp. 248-272, 2008.

[146] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation-based image restoration," Multiscale Modeling & Simulation, vol. 4, pp. 460-489, 2005.

[147] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Total variation super resolution using a variational approach," in Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, 2008, pp. 641-644.

[148] H. Zhang, Y. Zhang, H. Li, and T. S. Huang, "Generative Bayesian image super resolution with natural image prior," Image Processing, IEEE Transactions on, vol. 21, pp. 4054-4067, 2012.

[149] S. Dai, M. Han, W. Xu, Y. Wu, and Y. Gong, "Soft edge smoothness prior for alpha channel super resolution," in Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, 2007, pp. 1-8.

[150] D. Kong, M. Han, W. Xu, H. Tao, and Y. Gong, "A conditional random field model for video super-resolution," in Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, 2006, pp. 619-622.

[151] K. V. Suresh and A. N. Rajagopalan, "Robust and computationally efficient superresolution algorithm," JOSA A, vol. 24, pp. 984-992, 2007.

[152] V. Patanaviji and S. Jitapunkul, "An iterative super-resolution reconstruction of image sequences using a Bayesian approach with BTV prior and affine block-based registration," in Computer and Robot Vision, 2006. The 3rd Canadian Conference on, 2006, pp. 45-45.

[153] B. V. Kumar and R. Aravind, "Computationally efficient algorithm for face super-resolution using (2D) 2-PCA based prior," IET image processing, vol. 4, pp. 61-69, 2010.

[154] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," SIAM Journal on Imaging Sciences, vol. 2, pp. 323-343, 2009.

[155] P. Purkait and B. Chanda, "Super resolution image reconstruction through Bregman iteration using morphologic regularization," Image Processing, IEEE Transactions on, vol. 21, pp. 4029-4039, 2012.

[156] S. Harmeling, S. Sra, M. Hirsch, and B. Scholkopf, "Multiframe blind deconvolution, super-resolution, and saturation correction via incremental EM," in Image Processing (ICIP), 2010 17th IEEE International Conference on, 2010, pp. 3313-3316.

[157] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," Image Processing, IEEE Transactions on, vol. 6, pp. 1621-1633, 1997.

[158] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," JOSA A, vol. 6, pp. 1715-1726, 1989.

[159] Z. Fan and Z. Qi-dan, "Super-resolution image reconstruction for omni-vision based on POCS," in Control and Decision Conference, 2009. CCDC'09. Chinese, 2009, pp. 5045-5049.

[160] M. Hafner, M. Liedlgruber, and A. Uhl, "POCS-based super-resolution for HD endoscopy video frames," in Computer-Based Medical Systems (CBMS), 2013 IEEE 26th International Symposium on, 2013, pp. 185-190.

[161] Z. Zhang, X. Wang, J. Ma, and G. Jia, "Super resolution reconstruction of three view remote sensing images based on global weighted POCS algorithm," in Remote Sensing, Environment and Transportation Engineering (RSETE), 2011 International Conference on, 2011, pp. 3615-3618.

[162] M. Gevrekci, B. K. Gunturk, and Y. Altunbasak, "POCS-based restoration of Bayer-sampled image sequences," in Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on, 2007, pp. I-753-I-756.

[163] G. Caner, A. M. Tekalp, and W. Heinzelman, "Super resolution recovery for multi-camera surveillance imaging," in Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on, 2003, pp. I-109-12 vol. 1.

[164] S. Panda, M. Prasad, and G. Jena, "POCS based super-resolution image reconstruction using an adaptive regularization parameter," arXiv preprint arXiv:1112.1484, 2011.

[165] A. J. Patti and Y. Altunbasak, "Artifact reduction for POCS-based super resolution with edge adaptive regularization and higher-order interpolants," in Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on, 1998, pp. 217-221.

[166] A. K. Brodzik and J. M. Mooney, "Convex projections algorithm for restoration of limited-angle chromotomographic images," JOSA A, vol. 16, pp. 246-257, 1999.

[167] M. Irani and S. Peleg, "Improving resolution by image registration," CVGIP: Graphical models and image processing, vol. 53, pp. 231-239, 1991.

[168] R. Nayak, S. Harshavardhan, and D. Patra, "Morphology based iterative back-projection for super-resolution reconstruction of image," in Emerging Technology Trends in Electronics, Communication and Networking (ET2ECN), 2014 2nd International Conference on, 2014, pp. 1-6.

[169] X. Liang and Z. Gan, "Improved non-local iterative back-projection method for image super-resolution," in Image and Graphics (ICIG), 2011 Sixth International Conference on, 2011, pp. 176-181.

[170] R. Nayak, S. Monalisa, and D. Patra, "Spatial super resolution based image reconstruction using HIBP," in India Conference (INDICON), 2013 Annual IEEE, 2013, pp. 1-6.

[171] Z. Yan and Y. Lu, "Super resolution of MRI using improved IBP," in Computational Intelligence and Security, 2009. CIS'09. International Conference on, 2009, pp. 643-647.

[172] R. Lai, Y.-t. Yang, H.-x. Zhou, H.-l. Qin, and B.-J. Wang, "Total variation regularized iterative back-projection method for single frame image super resolution," in Signal Processing (ICSP), 2012 IEEE 11th International Conference on, 2012, pp. 931-934.

[173] F. Qin, "An improved super resolution reconstruction method based on initial value estimation," in Image and Signal Processing (CISP), 2010 3rd International Congress on, 2010, pp. 826-829.

[174] M.-K. Chang, M.-C. Su, G.-S. Lin, and F.-T. Chien, "A content-adaptive image enlargement scheme based on iterative back-projection," in Information Science, Electronics and Electrical Engineering (ISEEE), 2014 International Conference on, 2014, pp. 1104-1109.

[175] P. Rasti, H. Demirel, and G. Anbarjafari, "Improved Iterative Back Projection for video super-resolution," in Signal Processing and Communications Applications Conference (SIU), 2014 22nd, 2014, pp. 552-555.

[176] S. Dai, M. Han, Y. Wu, and Y. Gong, "Bilateral back-projection for single image super resolution," in Multimedia and Expo, 2007 IEEE International Conference on, 2007, pp. 1039-1042.

[177] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1-8.

[178] E. Faramarzi, V. R. Bhakta, D. Rajan, and M. P. Christensen, "Super resolution results in PANOPTES, an adaptive multi-aperture folded architecture," in Image Processing (ICIP), 2010 17th IEEE International Conference on, 2010, pp. 2833-2836.

[179] K. I. Kim and Y. Kwon, "Example-based learning for single-image super-resolution," in Pattern Recognition, ed: Springer, 2008, pp. 456-465.

[180] Y.-W. Tai, S. Liu, M. S. Brown, and S. Lin, "Super resolution using edge prior and single image detail synthesis," in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, 2010, pp. 2400-2407.

[181] J. Sun, Z. Xu, and H.-Y. Shum, "Image super-resolution using gradient profile prior," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1-8.

[182] J. Li, W. Gong, W. Li, and F. Pan, "Single-image super-resolution reconstruction based on global non-zero gradient penalty and non-local Laplacian sparse coding," Digital Signal Processing, vol. 26, pp. 101-112, 2014.

[183] R. Fattal, "Image upsampling via imposed edge statistics," in ACM Transactions on Graphics (TOG), 2007, p. 95.

[184] Y. Fan, Z. Gan, Y. Qiu, and X. Zhu, "Single image super resolution method based on edge preservation," in Image and Graphics (ICIG), 2011 Sixth International Conference on, 2011, pp. 394-399.

[185] Q. Zhou, S. Chen, J. Liu, and X. Tang, "Edge-preserving single image super-resolution," in Proceedings of the 19th ACM international conference on Multimedia, 2011, pp. 1037-1040.

[186] W. Dong, G. Shi, L. Zhang, and X. Wu, "Super-resolution with nonlocal regularized sparse representation," in Visual Communications and Image Processing 2010, 2010, pp. 77440H-77440H-10.

[187] L. Li, Y. Xie, W. Hu, and W. Zhang, "Single image super-resolution using combined total variation regularization by split Bregman Iteration," Neurocomputing, vol. 142, pp. 551-560, 2014.
[188] J. Li, W. Gong, and W. Li, "Dual-sparsity regularized sparse representation for single image super-resolution," Information Sciences, vol. 298, pp. 257-273, 2015.

[189] P. Sen and S. Darabi, "Compressive image super-resolution," in Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on, 2009, pp. 1235-1242.

[190] J. Yang, Z. Lin, and S. Cohen, "Fast image super-resolution based on in-place example regression," in Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, 2013, pp. 1059-1066.

[191] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in Computer Vision, 2009 IEEE 12th International Conference on, 2009, pp. 349-356.

[192] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," Computer Graphics and Applications, IEEE, vol. 22, pp. 56-65, 2002.

[193] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, 2004, pp. I-I.

[194] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L. Morel, "Neighbor embedding based single-image super-resolution using Semi-Nonnegative Matrix Factorization," in Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, 2012, pp. 1289-1292.

[195] T.-M. Chan, J. Zhang, J. Pu, and H. Huang, "Neighbor embedding based super-resolution algorithm through edge detection and feature selection," Pattern Recognition Letters, vol. 30, pp. 494-502, 2009.

[196] M. Cao, Z. Gan, and X. Zhu, "Super-resolution algorithm through neighbor embedding with new feature selection and example training," in Signal Processing (ICSP), 2012 IEEE 11th International Conference on, 2012, pp. 825-828.

[197] D. Mishra, B. Majhi, and P. K. Sa, "Neighbor Embedding based Super-Resolution using Residual Luminance," in India Conference (INDICON), 2014 Annual IEEE, 2014, pp. 1-6.

[198] X. Liao, G. Han, Y. Wo, H. Huang, and Z. Li, "New feature selection for neighbor embedding based super-resolution," in Multimedia Technology (ICMT), 2011 International Conference on, 2011, pp. 441-444.

[199] T. Yu, Z. Gan, and X. Zhu, "Novel neighbor embedding super resolution method for compressed images," in Image Analysis and Signal Processing (IASP), 2012 International Conference on, 2012, pp. 1-4.

[200] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Face Super-Resolution via Multi-layer Locality-constrained Iterative Neighbor Embedding and Intermediate Dictionary Learning," 2014.

[201] S. Qu, R. Hu, S. Chen, L. Chen, and M. Zhang, "Robust face super-resolution via position-patch neighborhood preserving," in Multimedia and Expo Workshops (ICMEW), 2014 IEEE International Conference on, 2014, pp. 1-5.

[202] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Noise robust face hallucination via locality-constrained representation," 2014.

[203] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L. Alberi Morel, "Single-image super-resolution via linear mapping of interpolated self-examples," 2014.

[204] Y. Tang, P. Yan, Y. Yuan, and X. Li, "Single-image super-resolution via local learning," International Journal of Machine Learning and Cybernetics, vol. 2, pp. 15-23, 2011.

[205] L. Xiaoqiang, Y. Haoliang, Y. Yuan, Y. Pingkun, L. Luoqing, and L. Xuelong, "Local learning-based image super-resolution," in Multimedia Signal Processing (MMSP), 2011 IEEE 13th International Workshop on, 2011, pp. 1-5.

[206] K. I. Kim and Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, pp. 1127-1133, 2010.

[207] D. Li and S. Simske, "Example based single-frame image super-resolution by support vector regression," Journal of Pattern Recognition Research, vol. 1, pp. 104-118, 2010.

[208] K. S. Ni and T. Q. Nguyen, "Image superresolution using support vector regression," Image Processing, IEEE Transactions on, vol. 16, pp. 1596-1610, 2007.

[209] R. Baraniuk, "Compressive sensing," IEEE signal processing magazine, vol. 24, 2007.

[210] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," Image Processing, IEEE Transactions on, vol. 19, pp. 2861-2873, 2010.

[211] W. Dong, D. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," Image Processing, IEEE Transactions on, vol. 20, pp. 1838-1857, 2011.

[212] J. Li, J. Wu, S. Yang, and J. Liu, "Dictionary learning for image super-resolution," in Control Conference (CCC), 2014 33rd Chinese, 2014, pp. 7195-7199.

[213] Z. Zhihui, W. Bo, and S. Kang, "Single Remote Sensing Image Super-Resolution and Denoising via Sparse Representation," in Multi-Platform/Multi-Sensor Remote Sensing and Mapping (M2RSM), 2011 International Workshop on, 2011, pp. 1-5.

[214] Y. Tang, Y. Yuan, P. Yan, and X. Li, "Single-image super-resolution via sparse coding regression," in Image and Graphics (ICIG), 2011 Sixth International Conference on, 2011, pp. 267-272.

[215] N. A. Zamani, A. Zahamdin, S. N. H. S. Abdullah, and M. J. Nordin, "Sparse representation super-resolution method for enhancement analysis in video forensics," in Intelligent Systems Design and Applications (ISDA), 2012 12th International Conference on, 2012, pp. 921-926.

[216] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," Image Processing, IEEE Transactions on, vol. 15, pp. 3736-3745, 2006.

[217] J. P. W. Pluim, J.B.A. Maintz, and M.A. Viergever, “Mutual information based registration of medical images: a survey,” IEEE Trans Med Imaging, vol. 22(8), pp. 986 – 1004, August 2003.

[218] S. Farsiu, M. Elad, and P. Milanfar, "A practical approach to superresolution," in Electronic Imaging 2006, 2006, pp. 607703-607703-15.

[219] M. Pickering, G. Ye, M. Frater, and J. Arnold, "A transform-domain approach to super-resolution mosaicing of compressed images," in Journal of Physics: Conference Series, 2008, p. 012039.

[220] R. Ferzli and L. J. Karam, "A No-Reference Objective Image Sharpness Metric Based on the Notion of Just Noticeable Blur (JNB)," Image Processing, IEEE Transactions on, vol. 18, pp. 717-728, 2009.

[221] S. Wenzhong, T. Yan, and L. Jian, "A fast super-resolution reconstruction from image sequence," Wuhan University Journal of Natural Sciences, vol. 11, pp. 399-404, 2006/02/01 2006.

[222] A. Zomet and S. Peleg, "Super-Resolution from Multiple Images Having Arbitrary Mutual Motion," in Super-Resolution Imaging. vol. 632, S. Chaudhuri, Ed., ed: Springer US, 2002, pp. 195-209.

[223] H. Su, Y. Wu, and J. Zhou, "Super-resolution without dense flow," Image Processing, IEEE Transactions on, vol. 21, pp. 1782-1795, 2012.

[224] W. Qing and S. Xiaoli, "Joint image registration and super-resolution reconstruction based on regularized total least norm," in Image Processing (ICIP), 2009 16th IEEE International Conference on, 2009, pp. 1537-1540.

[225] L. Zhang, H. Zhang, H. Shen, and P. Li, "A super-resolution reconstruction algorithm for surveillance images," Signal Processing, vol. 90, pp. 848-859, 2010.

[226] Q. Sang, X. Wu, C. Li, and A. C. Bovik, "Blind image quality assessment using a reciprocal singular value curve," Signal Processing: Image Communication.