

Old Dominion University

ODU Digital Commons

Electrical & Computer Engineering Theses &
Dissertations

Electrical & Computer Engineering

Spring 2012

Super-Resolution of Unmanned Airborne Vehicle Images with Maximum Fidelity Stochastic Restoration

Amr Yousef
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/ece_etds



Part of the [Computer Engineering Commons](#), and the [Electrical and Computer Engineering Commons](#)

Recommended Citation

Yousef, Amr. "Super-Resolution of Unmanned Airborne Vehicle Images with Maximum Fidelity Stochastic Restoration" (2012). Doctor of Philosophy (PhD), Dissertation, Electrical & Computer Engineering, Old Dominion University, DOI: 10.25777/hxd9-t888
https://digitalcommons.odu.edu/ece_etds/155

This Dissertation is brought to you for free and open access by the Electrical & Computer Engineering at ODU Digital Commons. It has been accepted for inclusion in Electrical & Computer Engineering Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

**SUPER-RESOLUTION OF UNMANNED AIRBORNE
VEHICLE IMAGES WITH MAXIMUM FIDELITY
STOCHASTIC RESTORATION**

by

Amr Yousef

B.S. May 2001, Alexandria University, Egypt
M.S. December 2006, Alexandria University, Egypt

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

ELECTRICAL AND COMPUTER ENGINEERING

OLD DOMINION UNIVERSITY
May 2012

Approved by:

Dr. Mohammad A. Karim (Director)

Dr. Hussein Abdel-Wahab (Member)

Dr. Jang Li (Member)

Dr. Dumitrie Popescu (Member)

ABSTRACT

SUPER-RESOLUTION OF UNMANNED AIRBORNE VEHICLE IMAGES WITH MAXIMUM FIDELITY STOCHASTIC RESTORATION

Amr Yousef

Old Dominion University, 2012

Director: Dr. Mohammad A. Karim

Super-resolution (SR) refers to reconstructing a single high resolution (HR) image from a set of subsampled, blurred and noisy low resolution (LR) images. One may, then, envision a scenario where a set of LR images is acquired with sensors on a moving platform like unmanned airborne vehicles (UAV). Due to the wind, the UAV may encounter altitude change or rotational effects which can distort the acquired as well as the processed images. Also, the visual quality of the SR image is affected by image acquisition degradations, the available number of the LR images and their relative positions. This dissertation seeks to develop a novel fast stochastic algorithm to reconstruct a single SR image from UAV-captured images in two steps. First, the UAV LR images are aligned using a new hybrid registration algorithm within subpixel accuracy. In the second step, the proposed approach develops a new fast stochastic minimum square constrained Wiener restoration filter for SR reconstruction and restoration using a fully detailed continuous-discrete-continuous (CDC) model. A new parameter that accounts for LR images registration and fusion errors is added to the SR CDC model in addition to a multi-response restoration and reconstruction. Finally, to assess the visual quality of the resultant images, two figures of merit are introduced: information rate and maximum realizable fidelity. Experimental results show that quantitative assessment using the proposed figures coincided with the visual qualitative assessment. We evaluated our filter against other SR techniques and its results were found to be competitive in terms of speed and visual quality.

To my lovely wife, my children, and my parents

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to the various individuals who made this work possible. I would like to show my deepest gratitude to Dr. Mohammad Karim for his constant guidance, understanding, patience, and most importantly, his encouragement during my graduate studies at Old Dominion University. For everything you've done for me, Dr. Karim, I thank you.

Additionally, I am very grateful to my co-advisor, Dr. Jiang Li, for his assistance and support. He has been always there to listen and give advice. I am also thankful to him for reading my reports, commenting on my views and enriching my ideas. Many thanks also to the other committee members, Dr. Hussein Abdel-Wahab and Dr. Dimitrie Popescu, for their time and effort in reviewing this document.

Also, I am indebted to Dr. Zia Rahman who introduced me to this field of research and whose thought-provoking guidance compelled me to produce the best work that I could. Although Dr. Rahman was not present for the completion of this work, I hope that this dissertation constitutes "something [he] and I can be proud of." May Allah shower His Mercy upon him.

My final, and most heartfelt, acknowledgment must go to my wife Asmaa Rashed. Her support, encouragement, companionship and quiet patience have turned my journey through graduate school into a pleasure. For all that, and for being everything I am not, she has my everlasting love.

TABLE OF CONTENTS

	Page
LIST OF TABLES	viii
LIST OF FIGURES	xi
Chapter	
1. INTRODUCTION	1
1.1 OVERVIEW	1
1.2 MOTIVATION	1
1.3 PROBLEM FORMULATION	2
1.4 RECENT RESEARCH	3
1.5 CONTRIBUTION OF THE DISSERTATION	11
1.6 ORGANIZATION OF THE DISSERTATION	14
2. ROBUST REGISTRATION OF UAV CAPTURED IMAGES	16
2.1 SUFFICIENT OVERLAP CRITERION	16
2.2 REGISTRATION BASED ON SPATIAL TRANSFORMATION	20
2.3 AUTOMATED CORRECTION OF UNREGISTERED UAV IMAGES	22
2.4 FREQUENCY-DOMAIN BASED REGISTRATION	28
3. LR IMAGES PROJECTION ONTO HR GRID	44
3.1 OPTIMUM SCENELS HIGH RESOLUTION GRID ALLOCATION	44
3.2 PHASE SHIFT ADJUSTMENT	47
4. STOCHASTIC OPTIMAL RECONSTRUCTION AND RESTORATION	49
4.1 LOW RESOLUTION IMAGE FORMULATION	49
4.2 HR GRID INTERLACING	53
4.3 DERIVED STOCHASTIC WIENER RESTORATION FILTER	54
4.4 ENHANCED CDC MODEL	57
4.5 SPEEDED-UP WIENER FILTER	59
4.6 WIENER CHARACTERISTIC RESTORATION FILTER	61
4.7 WIENER GAUSSIAN ENHANCEMENT FILTER	61
4.8 FIDELITY ANALYSIS	62
4.9 SIMULATIONS AND RESULTS	63
5. INFORMATION THEORETIC APPROACH	84
5.1 INTRODUCTION	84
5.2 INFORMATION RATE	84
5.3 MAXIMUM REALIZABLE FIDELITY	87
6. MULTI-RESPONSE SR RECONSTRUCTED IMAGE	97
6.1 INTRODUCTION	97
6.2 CASE 1: SINGLE WIENER RESTORATION FILTER (SWF)	98
6.3 CASE 2: WIENER FILTERS RESTORATION MATRIX (WFM)	98
6.4 WIENER FILTERS MATRIX DERIVATION	99

	Page
7. CONCLUSIONS AND FUTURE DIRECTIONS	106
7.1 CONCLUSIONS	106
7.2 FUTURE WORK DIRECTIONS	109
 BIBLIOGRAPHY	 111
 VITA	 120

LIST OF TABLES

Table	Page
1 Optimum sampling factors for different image sizes	37
2 Total subpixel registration time.	42
3 Estimated subpixel shifts	76
4 Adjusted Subpixel Shifts	76
5 Fidelity comparison for full-SR reconstructed checkerboard SR images	76
6 Fidelity comparison for partial-SR reconstructed checkerboard SR images	76
7 Computational time for full-SR reconstructed SR images (s)	79
8 Improvements in computational time using speeded-up Wiener filter (s)	79
9 The optimal HR grid locations for the LR scenels (partial-SR)	80

LIST OF FIGURES

Figure	Page
1 SR enhancements hierarchy	5
2 Proposed algorithm pipeline	13
3 UAV rotations: yaw, pitch and roll	17
4 FOV of the camera and the corresponding ground footprint	17
5 Ground footprints with pitch or roll	18
6 Ground footprints with yaw	19
7 Ground footprints of two consecutive images	21
8 Simulated High Resolution Scenes	23
9 Registered images after different rotational parameters	24
10 Generation of scale spaces and DOGs	26
11 Corrected LR images after using the SIFT algorithm (home images)	29
12 Corrected LR images after using the SIFT algorithm (aerial images 1)	30
13 Corrected LR images after using the SIFT algorithm (aerial images 2)	31
14 Upsampled cross correlation with window size 150×150	38
15 Level curves for the upsampled cross correlation	39
16 The estimated NRMSE \hat{E} against the actual one	40
17 The upsampling factor ϵ against the estimation error δs	41
18 Computational time required for the initial estimate	41
19 Computational time for the refinement step	42
20 Total time required for registration	43
21 The effect of optimal and non-optimal allocation on the visual quality	44
22 Complete continuous-discrete-continuous SR reconstruction model	50
23 Pipeline of the speeded up Wiener filter	60
24 Random polygon image	64

Figure	Page
25 Derived Wiener restoration filter against the optical response index (σ).....	65
26 Fidelity against the optical response index (σ).	66
27 Different random polygon reconstructed images.	67
28 Different checkerboard reconstructed images.	68
29 Fidelity comparison for different registration and fusion models.....	69
30 Different Partial-SR reconstructed images ($\sigma = 0.7$ and SNR=32)	71
31 Different full-SR reconstructed images ($\sigma = 0.7$ and SNR=32)	72
32 Fidelity comparison for different full-SR reconstructed images	73
33 Fidelity Components (\mathcal{F}_a , \mathcal{F}_f and \mathcal{F}_n) versus σ	74
34 Fidelity comparison for different partial-SR reconstructed images.....	75
35 Different full-SR reconstructed checkerboard images.	77
36 Different partial-SR reconstructed checkerboard images.	78
37 Different restoration outputs at image-gathering device index $\sigma = 0.8$	81
38 Fidelity comparison for SWR, WCR, and WIGE restoration outputs.	82
39 Fidelity comparison for different restorations with optimal allocations.	83
40 Information rate comparison against optical response index σ	89
41 Information rate comparison against SNR.....	90
42 Information rate comparison against mean spatial detail μ	90
43 Different full-SR reconstructed outputs for various σ	91
44 Different partial-SR reconstructed outputs for various σ	92
45 Different partial-SR reconstructed outputs for various SNR.....	93
46 Different partial-SR reconstructed outputs for various SNR.....	94
47 Different full-SR reconstructed outputs for various mean spatial detail	95
48 Different full-SR reconstructed outputs for various mean spatial detail	96
49 Different spatial frequency responses	97
50 Multi-response SR reconstruction and restoration	98

Figure	Page
51 Comparison between different multi-response outputs.	101
52 Fidelity comparison between different multi-response cases	102
53 Information rate comparison against mean spatial detail μ	102
54 Reconstructed single optical response versus different MSD	103
55 Reconstructed SWF versus different MSD	104
56 Reconstructed WFM versus different MSD	105

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

The process of SR involves reconstructing a single HR image from a set of subsampled, blurred and noisy LR images. SR techniques attempt to improve spatial resolution by incorporating into the final HR result the additional new details that are revealed in each LR image. In super resolution techniques the degraded images are captured either from different cameras looking at the same scene or a single camera capturing a set of successive frames of the scene with sufficient overlap between them [1–7]. These frames are registered with respect to a common set of coordinates, and the inter-subpixel shifts between them are typically unknown, random and don't follow a regular pattern or structure. Consequently, without accurate registration, the visual quality of the reconstructed image is degraded significantly. The spatial resolution of the acquired images can also be improved physically by decreasing the pixel size or increasing the chip size where the latter increases the capacitance, which reduces the data transfer, and the former increases the photo-detector noise [1]. SR has many applications, including improving the resolution of the video printing, high definition television, video surveillance, biometrics, astronomical imaging and medical imaging [8–13].

SR is an effective and inexpensive method for reducing aliasing and increasing spatial resolution. It can be used to reconstruct and restore the image beyond the Nyquist frequency of the acquisition image. Increasing the spatial resolution in SR techniques basically depends on the availability of multiple LR images of the same scene, with slightly different angles of view: i.e., the LR images represent different looks at the same scene. If the shifts between the degraded images are of integer units then these images contain identical information that can't be used to reconstruct an SR image. If the degraded images differ from each other by subpixel amounts then the extra information in every LR image can be used to increase the spatial resolution of the imaging device and produce an SR image [1, 4, 5].

1.2 MOTIVATION

During its flight the UAV experiences roll, pitch and yaw even when it is flying at roughly the same altitude. The change in these parameters changes what falls within the field-of-view (FOV) of the camera [14] which leads to slightly different looks at the same scene. These looks contain similar, but not identical information which can be used to get an SR image of the overlapping area common to all these images. The visual quality of the SR image is affected by many factors, such as the optics blur, the inherent signal-to-noise ratio of the system, quantization artifacts, the number of scenels (scene elements) i.e., the number of overlapped images used for SR reconstruction within the SR grid and their relative arrangement. A technique referred to as microscanning is an effective method for reducing aliasing and increasing spatial resolution. By moving the FOV with predetermined sub-pixel shifts, both aliasing reduction and resolution improvement are realized with increasing effective spatial sampling periods which can be used in obtaining the reconstructed SR image with high quality. However, the LR images may have relative shifts that are unknown. This random pattern of subpixel shifts can lead to unpleasant visual quality, especially at the edges of the reconstructed SR image. Because microscanning is limited by the time required to shift the image gathering sensor horizontally and vertically [15], it can't be done from the UAV. But, by using its concept, it is possible to reduce the artifacts within the reconstructed image [16]. Also, to boost the SR images visual quality and to correct for the acquisition and display degradations, restoration filters such as optimal stochastic Wiener filter can be used to remove artifacts within the reconstructed image [17].

1.3 PROBLEM FORMULATION

The SR problem can be defined as follows:

Given a set of LR images $s_k, k = 1, \dots, K$ each of dimension $M_1 \times M_2$ and an SR factor r , it is required to reconstruct a single HR image of dimension $rM_1 \times rM_2$.

SR is considered an *ill-posed* problem because it is undetermined, as there isn't a sufficient number of LR images, and the blur operator is ill-conditioned. Describing the acquisition of LR images is a key factor in the recovery or the estimation of the original HR scene. Charged coupled device (CCD) cameras contain arrays of photo-detectors or sensors that capture the reflected light from the scene that falls within its FOV. The image gathering

lens blurs the captured images through the spatial convolution between the scene and the spatial response of the camera. The CCD array samples and quantizes the continuous scene into a 2-D digital image. The acquired images are degraded with various types of noises like quantization errors, thermal noise, etc. From the central limit theory, these additive noises can be represented by a single Gaussian random variable [18].

Following Carl et al. [19], the image gathering device transforms the continuous input scene $L(x, y)$ into discrete image $s_i(x, y)$ as defined by:

$$s_i(x, y) = [L(x, y) * \tau_i(x, y) + n_i(x, y)] \underline{\underline{\|}}(x, y) \quad (1)$$

where $\tau_i(x, y)$ is the spatial response of the image acquisition device, $n_i(x, y)$ is the additive photo-detector noise and the symbol $*$ denotes spatial convolution. The sampling function

$$\underline{\underline{\|}}(x, y) = \sum_m \sum_n \delta(x - m, y - n) \quad (2)$$

denotes sampling on a rectangular grid with unit sampling intervals and $\delta(x, y)$ is the Dirac delta function. This formulation is considered as the core of the SR model that will be introduced in section 4.1. This model can be discretized and written as [2, 20]

$$G_i = A_i F + \Upsilon_i \quad (3)$$

where G_i , F and Υ_i are the lexicographical representation of s_i , L and n_i respectively. The operator A_i combines the motion, the blurring and the sampling operators.

1.4 RECENT RESEARCH

Most SR approaches consist of three main steps: registration, reconstruction and restoration. Registration is a process of aligning several images to a reference one. Registration with subpixel accuracy is essential in reconstructing high resolution (HR) images with enhanced visual quality and minimum unwanted artifacts [3]. The subpixel registration techniques can be classified into four types: (1) Correlation interpolation, (2) Intensity interpolation, (3) Differential interpolation and (4) Phase correlation.

In correlation interpolation, a discrete correlation function between two images is calculated and interpolated, and the translation is obtained by searching for the maximum of

the correlation function. In intensity interpolation, parts of the reference image are selected and interpolated according to the subpixel accuracy required and a search is conducted over these parts with the unregistered image [21]. The idea behind differential interpolation is to relate the difference between two consecutive frames to the spatial intensity gradient of the first image [21]. The phase correlation technique can be used to register images with translational shifts and a different scaling [22]. It depends on the idea that the phase of the cross power spectrum between two images contains most of the information about the relative displacement between them [23]. Also, it is known for its high accuracy, low computational complexity, robustness to noise and invariance to lens optical blur [24–26].

Image reconstruction is a method of reconstructing HR images by incorporating the available different registered LR pixels to estimate the missing pixels on the HR grid. In most cases, the registered LR pixels are irregularly distributed over the HR grid. Popular techniques for image reconstruction [12] are nearest neighbor interpolation, bilinear interpolation, cubic spline interpolation and piecewise cubic convolution [27]. Image restoration is a method of correcting the reconstructed HR image from degraded LR images with blurring, aliasing and noise [28]. Popular restoration methods are inverse filters, least square filters and iterative approaches [12, 29].

Developments and challenges in SR algorithms have attracted many researchers in recent years due to the high demand on its many applications. Figure 1 depicts the hierarchy of SR techniques. The earliest work in SR was to enhance the spatial resolution of a single LR image which was extended to cover all the levels in the given hierarchy. The emphasis of this dissertation is the reconstruction and restoration of a single SR image from a set of degraded LR images, which falls at the second level of the given hierarchy. We are going to briefly discuss some of the main SR techniques that works at that level.

SR algorithms can be divided into two categories: spatial domain approaches and spatial frequency domain approaches [3]. Frequency domain approaches include reconstruction via aliasing removal, recursive least squares filters and multichannel sampling theorem based techniques [3]. Spatial domain approaches include non-uniform interpolation, algebraic filtered back projection, probabilistic methods like maximum likelihood (ML) estimation and maximum a posteriori (MAP) based algorithms, projection onto convex set (POCS), hybrid ML/MAP/POCS methods and Tikhonov-Arsenin regularized methods [3]. Frequency domain approaches are simple because the relation between the LR images and the HR images

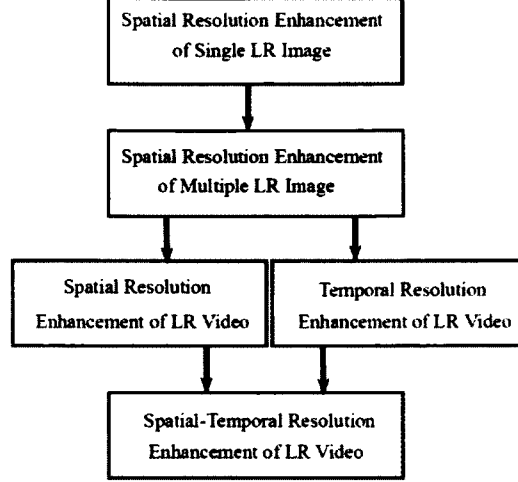


FIG. 1: SR enhancements hierarchy

is easy to understand in the frequency domain. Also, they have low computational complexity but their performance degrades if the motion model is other than a global translation and the blur kernel is not a linear shift invariant (LSI) one. On the other hand, spatial domain approaches can be exploited to work on any arbitrary motion and degradation models such as motion blur spatially varying or invariant blur, but they are computationally intensive approaches [1].

Non-uniform approaches are the simplest ones. Alam et al. [30] proposed a method to reconstruct a HR infrared image from a set of randomly shifted LR infrared frames. They estimated the shifts between the LR frames using a gradient based registration method and used a weighted nearest neighbor approach to estimate missing pixels on the HR grid and finally restore the reconstructed image using the traditional Wiener filter [31]. Ur and Gross [32] used a framework based on the multichannel sampling theorem followed by a deblurring step to reconstruct the SR image. Komatsu et al. [33] proposed a very high definition (VHD) imaging system using multiple cameras with high signal-to-noise ratio, and their approach was based on Lanweber algorithm [34]. Shah and Zakhor [35] developed a method to enhance images extracted from a video sequence by compensating for inaccurate motion estimation by estimating multiple motion vectors for each pixel. Nguyen and Milanfar [36] reconstructed SR images using an interpolation-restoration method based on wavelets theory.

Tsai and Huang [37] proposed one of the earliest frequency domain SR approaches. They

utilized the shift property of the Fourier transform and related the aliased LR images to an ideal image. They relate the degraded LR discrete Fourier transform (DFT) coefficients to the continuous Fourier transform (CFT) of the original scene. Once the DFT coefficients are obtained, they constructed the HR image by means of inverse Fourier transform (IDFT). Their approach is limited to global lateral shifts which are considered its main drawbacks. Kim et al. [38] applied this approach to blurred and noisy images and restored the reconstructed high resolution image using the Tikhonov regularization [39]. Their technique was improved by Kim and Su [40] to incorporate LR images with different blurs. A recursive least square approach was developed by Bose et al. [41] to minimize the registration errors. This method was enhanced by Rhee and Kang [42] to reduce the computational complexity by utilizing the discrete cosine transform (DCT) instead of DFT.

Sauer and Allebach [43] used the POCS method to reconstruct a HR image from LR images based on the assumption that these LR images are not affected by any blur. It estimates the HR image from a convex constrained set that contains all the possible candidates for the HR image. This set is confined by limitations like fidelity, positivity, bounded energy, etc. The HR image is estimated iteratively as given by

$$\hat{F}^{n+1} = P_m P_{m-1} \dots P_1 \hat{F}^n \quad (4)$$

where P_k is the projection operator that maps the i^{th} iteration to the space of the convex set C_k , ($k = 1 \dots m$). Constraints like consistency with a given iteration, bounded energy, and fidelity are given by

$$C_{G_i} = \left\{ \hat{F} : |G_i(x, y) - A_i \hat{F}_i| \leq T_i(x, y) \right\}, \quad (5)$$

$$C_B = \left\{ \hat{F} : v_{min} \leq \hat{F}(x, y) \leq v_{max} \right\} \quad (6)$$

and

$$C_r = \left\{ \hat{F} : \left\| \hat{F} - F_{ref} \right\| \leq \epsilon_r \right\} \quad (7)$$

respectively. $T_i(x, y)$ is a threshold that is selected based on the estimated noise, registration accuracy and number of iterations. v_{min} and v_{max} bound the minimum and the maximum values in the solution space for every iteration. Also, C_r constrains the iteration outputs to be at a certain extent from a target image F_{ref} , that is usually selected as an interpolated

version of the image used in the motion estimation step as a reference image. The main strength and advantage of the POCS method is its robustness to noise and the ability to define a-priori constraints like fidelity, bounded energy, positivity, etc. to regularize the solution space. The main drawback of the POCS is to find the right projection operator that appropriately maps the iteration outputs to the convex solution space. Also, the current realizations are limited by large computational complexity and huge memory requirements and need larger number of iterations to converge to an accepted output [44].

Stark and Oskoui [45] extended the POCS technique to noiseless blurry images and combined interpolation and restoration in one step. Teklap et al. [46] applied the idea of POCS to reconstruct a HR image from a sequence of LR images that are affected by motion blur (non-zero aperture time). An improved technique based on the POCS was demonstrated by Patti et al. [47] to include the space varying blur, optical sensor noise and its physical dimensions. Patti and Altunbasak [48] suggested a POCS technique based on a continuous image formation model to interpolate the reconstructed image with higher orders, and they also improve the quality of the reconstructed images at the edges by minimizing the ringing artifacts. Another approach which is similar to the POCS was suggested by Tom and Katsaggelos [49] to estimate the SR image using ellipsoidal constraint sets. Although, the computational complexity and memory requirements of the POCS techniques are very high, and they reconstruct non-unique SR images, their main advantages are their simplicity and their efficient spatial domain model representation that can include a priori information.

Irani and Peleg [50] developed a SR reconstruction approach based on iterative back projection that is used in tomography to minimize the difference between simulated LR images and actual LR images by defining a back projection kernel that incorporates that difference into the reconstructed HR image. For the i^{th} iteration, the estimated HR image \hat{F}^i is obtained by adding the estimated $(i^{th} - 1)$ image \hat{F}^{i-1} to the back projected difference between the acquired LR frames G_i and the simulated LR images \hat{G}_i as given by:

$$\begin{aligned}\hat{F}^i &= \hat{F}^{i-1} + H_{BP}(G_i - \hat{G}_i) \\ &= \hat{F}^{i-1} + H_{BP}(G_i - A_i \hat{F}^{i-1})\end{aligned}\tag{8}$$

where H_{BP} is the back-projection kernel. Because the back projection approach does not have a step for regularization, the estimated HR image may diverge away from the solution space, which is considered its main drawback. To obtain a smooth estimated image, the

initial estimate may be selected as the average of the registered actual LR images. Later, Picard [51] extended that approach to include a perspective motion model. Then, Irani and Peleg [52] improved the model to cover general motion models.

SR can also be categorized as ill-posed due to the lack of LR images. Deterministic and stochastic regularization techniques can be used to constrain the solution space of the SR ill-posed problem. The deterministic techniques estimate the SR image using the constrained least squares (CLS). The regularization parameter is selected in a way too closely make the effect of the data fidelity and the smoothness of the solution in equilibrium state [28]. Katsaggelos et al. [53, 54] assign a different value for the registration parameter for every iteration using multi-channel iterative CLS approach. The significance of the regularization parameter and its best value is further studied by Bose et al. [55, 56] using the L-curve method.

Stochastic approaches use a Bayesian framework to solve for the SR problem. One of these approaches is the MAP technique, which estimates the SR image by maximizing the *a-posteriori* probability density function (PDF) $P_r(F|G_1, \dots, G_N)$ with respect to F . Using Bayes rule, the estimated image is given by MAP

$$\begin{aligned}\hat{F}_{MAP} &= \arg \max \{P_r(F|G_1, \dots, G_N)\} \\ &= \arg \max \left\{ \frac{P_r(G_1, \dots, G_N|F)P_r(F)}{P_r(G_1, \dots, G_N)} \right\}\end{aligned}\quad (9)$$

By applying the *log* function to Equation (9), the estimated image can be obtained by

$$\hat{F}_{MAP} = \arg \max \{\log[P_r(G_1, \dots, G_N|F)] + \log[P_r(F)]\} \quad (10)$$

The first term accounts for the *log-likelihood* function while the second term accounts for *a-prior* density of the proposed solution. The *log-likelihood* can be described by the joint PDF of the noise $P_{r(\tau)}$ as given by

$$P_r(G_1, \dots, G_N|F) = P_{r(\tau)}\{(G_1 - A_1F), \dots, (G_N - A_NF)\} \quad (11)$$

If the noise is described by independent and identically distributed (IID) Gaussian random variable with zero mean and variance σ^2 then the joint probability density function $P_{r(\tau)}$ can be written as the multiplication between the probability distribution of the individual terms

as given by

$$P_r(G_1, \dots, G_N | F) = \frac{1}{(2\pi)^{(M/2)\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^N \|G_i - A_i F\|^2 \right\} \quad (12)$$

where M is the total number of pixels that the image contains. To solve for the MAP estimator in Equation (9), a Markov random field (MRF) general model based on Gibbs distribution is defined as

$$P_r(F(k)) = \frac{1}{Z} \exp\{-\Omega(k)\} = \frac{1}{Z} \exp \left\{ -\sum_{c \in C} \phi_c(k) \right\} \quad (13)$$

where Z is the partition function and Ω is the energy function that is defined by

$$\Omega(F) = \exp \left\{ -\sum_{c \in C} \phi_c(k) \right\} \quad (14)$$

where C is the set of MRF cliques and ϕ_c is the potential function that is defined on a given clique. Typical prior models are

- $\Omega = \|F\|_2^2$ (l_2 - norm)
- $\Omega = \|\nabla^2 F\|_2^2$ (l_2 - norm of Laplacian)
- $\Omega = |\nabla F|$ (magnitude of gradient)

Applying Gibbs model, the MAP estimator in Equation (9) can be obtained by

$$\hat{F}_{MAP} = \arg \min \left\{ \sum_{i=1}^N \|G_i - A_i F\|^2 + \lambda \Omega(F) \right\} \quad (15)$$

where λ is the regularization parameter that converges the estimated SR image to the solution space. Iterative conjugate descent techniques can be used to minimize Equation (15) that can be differentiated with respect to F as given by

$$E(F) = -2 \sum_{i=1}^N A_i^T (G_i - A_i F) + \lambda \frac{\partial \Omega}{\partial F} \quad (16)$$

and the SR image is estimated by

$$\hat{F}_{k+1} = \hat{F}_k + \mu^k E(\hat{F}^k) \quad (17)$$

where μ^k is the step size that is computed using the steepest descent method as given by

$$\mu^k = \frac{N \|E(\hat{F}^k)\|^2}{\sum_i^N \|A_i E(\hat{F}^k)\|^2} \quad (18)$$

Tom and Katsaggelos [57] developed a ML SR algorithm based on the expectation maximization (EM) algorithm to estimate the subpixel shifts between the LR images, noise contaminated in every frame and the SR image. In [58], Schultz and Stevenson reconstruct an SR image from a sequence of LR video sequence using the MAP algorithm. A MAP approach to estimate both the lateral subpixel shifts between the LR frames and the HR image was discussed by Hardie et al. [59]. Elad and Feuer [2, 60] combine ML, MAP and POCS approaches into a hybrid method to reconstruct SR image from a set of blurred, noisy and undersampled images.

In addition to the above directions for SR research, there are many other approaches. Nguyen et al. [61] used the conjugate gradient method to solve the Tikhonov regularized SR problem by using efficient block circulant preconditions. Farsui et al. [62] used the L_1 norm minimization to reduce errors in blur and inter-LR subpixel shifts estimations. An SR technique based on the adaptive filtering theory and Kalman filter was developed by Elad and Feuer [20] where the HR image is estimated using the steepest descent method. Elad and Hel-Or [63] reconstruct an SR image using two separate fusion and deblurring steps. In their approach, the blur is assumed to be LSI and the subpixel shifts are assumed to be translational only. Hardie [64] reconstructs the HR image using a Wiener sliding window to restore the HR pixels from a weighted sum of LR pixels' that is selected by minimizing the mean square error and is adapted to the relative positions of the surrounding LR pixels.

The main problem of the previous approaches is that they work on a frame by frame basis to reconstruct the HR image and as a result, they don't exploit the available LR pixels spatial structure that can be used efficiently to produce a high quality image with minimum number of artifacts, especially when we have insufficient number of LR images. Also, their approaches depend on a discrete-discrete (DD) model which isn't sufficient to cover the acquisition and display degradations and can't be used to achieve optimal reconstruction and restoration with maximum fidelity between the original HR scene and the reconstructed SR one. Beside this, they use a blind estimation for the inter subpixel shifts between the acquired frames which is not sufficient to cover complex registration models with relatively

large motions. In our proposed approach we have two distinct steps: one for the accurate estimation of the subpixel shifts that is more flexible to cover any motion models and robust to moderate noise and lens optical blur. The other step is for the optimal reconstruction and restoration of the SR image that depends on the fully detailed SR CDC model and it optimally restores the SR image with maximum end-to-end fidelity. We revisit the Wiener filter developed by Carl et al. [19] and extend it for the general SR problem where the subpixel shifts are unknown and random. Although the Wiener filter has been discussed throughout the literature, our formulation for the SR problem is quite different. The earlier developments of this filter are for a single image restoration and are based on the assumption that the CDC model is constrained only by blurring and noise and ignores the insufficient sampling in the image gathering process [31]. Consequently, it will not actually minimize the mean square error of the reconstructed image [16].

1.5 CONTRIBUTION OF THE DISSERTATION

The dissertation seeks to develop a novel fast stochastic algorithm to reconstruct a single SR image from UAV captured images. Figure 2 depicts the pipeline of the proposed SR algorithm consisting of two steps. The first step is a robust registration of UAV LR images, while the second step is the stochastic optimal reconstruction and restoration with information theoretic assessment. The main contributions in the first step are:

- Automatically register the acquired UAV LR with in subpixel accuracy and correct for the distortion in these images. In our problem and due to the wind, the UAV may encounter altitude change or rotational effects such as yaw, pitch and roll, which can distort the acquired as well as the processed images with scaling, shear, tilt or perspective distortions. So, we need to correct for these distortions using spatial transformations such as affine or projective transformations and then align the corrected images within subpixel accuracy. To automate the subpixel registration process, we propose a hybrid algorithm to register the UAV LR images within subpixel accuracy with the following contributions
 - Use scale invariant feature transform (SIFT) to correct the distorted images.
 - Speed-up the performance of one of the most efficient and reliable subpixel algorithms that is robust to noise and lens optical blur. The proposed FFT based method reduces the dimensionality of the Fourier matrix of the cross correlation

and uses a forward and backward search in order to obtain an accurate estimation of the subpixel shifts.

- Optimally map the registered LR images to a denser HR grid with either one of following two methods:
 - * Minimum square distance allocation (MSDA) that finds the constrained optimum transformation to convert the nonuniform subpixel shifts pattern into a uniform one using the Minkowski distance with orders equal 1 or 2.
 - * Mid-point subpixel shifts allocation (MSSA) allows a relaxed allocation on the HR grid depending on the relative shifts and relative structure between the different LR scenels (scene elements).
- Spatially adjust the phase shifts of the interlaced LR frames using discrete Fourier transform shift theorem so that the interlaced LR images will have uniform phase shift differences between them.
- A new fast optimal stochastic minimum square restoration Wiener filter for SR reconstruction and restoration is introduced. This filter is used to boost The visual quality of the SR image that is affected by degradations during the acquisition such as blur due to system optics, aliasing due to insufficient sampling, photodetector noise, registration and fusion error, the number of scenels, i.e., the number of overlapped images used for SR reconstruction within the HR grid and their relative arrangement on the high resolution grid. The following contributions are proposed:
 - Introduce the mathematical derivation that depends on the CDC model to represent most of the degradations encountered during the image-gathering and image-display processes. Also, the end-to-end model is analyzed and the Wiener filter is formulated as a function of the parameters associated with the proposed SR system, such as image gathering and image display response indices, system average signal-to-noise ratio (SNR) and inter-subpixel shifts between the LR images.
 - A new parameter that accounts for LR images registration and fusion errors is added to the SR CDC model in order to improve the overall system fidelity.
 - The filter performance is speeded up by constraining it to work on small patches of the images and can be implemented efficiently in the frequency domain.

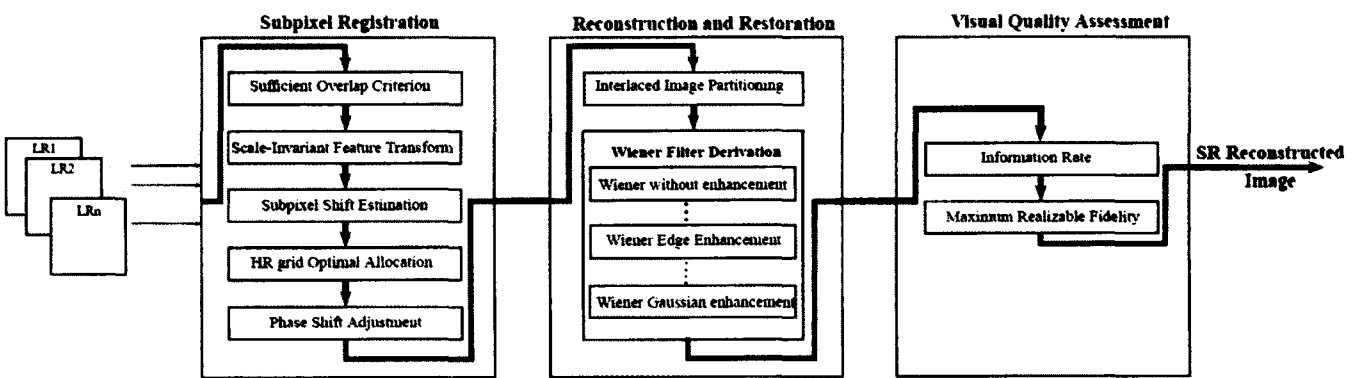


FIG. 2: Proposed algorithm pipeline

- The loss in the end-to-end system fidelity is analyzed and it is separated into three components and every component is related to its corresponding degradation in the proposed system, which enables a good design of the SR model.
- Produce multi- or pyramidal resolution reconstructed images depending on the available number of the LR images and their relative positions. So, it may be possible to produce SR only along a single dimension: diagonal, horizontal or vertical and use interpolation in the orthogonal dimension because there isn't sufficient information to produce a full 2D image. Thus the reconstructed SR image may have areas with full-SR or partial-SR.
- Assess the visual quality of the resultant images with two figures of merits that basically depend on the information theory. The information rate is used to measure the amount of information that the image gathering device is producing while the maximum realizable fidelity is used to measure the closeness between the reconstructed restored output images and the original input HR scenes.
- Reconstruct multi-response SR images by assuming the existence of more than one optical sensor on the UAV with different optical response indices. The reconstruction of SR images is utilized with either a single optimal Wiener filter or a matrix of optimal Wiener filters.

Simulation and experimental results demonstrate that the derived Wiener filter with the optimal allocation of LR images can reduce aliasing and blurring, resulting in a sharper reconstructed image with pleasant visual quality. Throughout simulations and experimental results it is found that quantitative assessment using the proposed figures coincides with the visual qualitative assessment. We evaluate our filter against other SR techniques and its results were found to be competitive.

1.6 ORGANIZATION OF THE DISSERTATION

The rest of this dissertation is organized as follows: In Chapter 2 we describe the robust subpixel registration stage of the proposed technique. Chapter 3 suggests two efficient methods for the optimal allocation of the LR pixel elements to the HR grid. Chapter 4 presents the mathematical development of the stochastic Wiener filter and the analysis of its results and its speeded-up performance. In Chapter 5, we propose two figures of merit based on the

information theory to assess the visual quality of the reconstructed images with its further enhancement using a multi-response reconstruction and restoration is covered in Chapter 6. The conclusions of the work done in this dissertation are summarized in Chapter 7.

CHAPTER 2

ROBUST REGISTRATION OF UAV CAPTURED IMAGES

Image registration is a process of aligning several images to a reference one or to a common reference coordinate grid. Typically, the alignment process brings the input, or the reference image, into alignment with the base image [65]. In our approach, the UAV images will be corrected from distortions using an appropriate spatial transformation like the affine or the projective transformation. Then a speeded-up phase correlation based technique that is implemented efficiently in the frequency domain will align the corrected images within subpixel accuracy. Prior to processing the available LR images, we define a metric to determine if there is enough overlap between them that would allow for SR reconstruction.

2.1 SUFFICIENT OVERLAP CRITERION

It is assumed that a set of LR images is acquired with sensors on a moving platform such as UAVs. During its flight the UAV experiences roll, pitch and yaw even when it is flying at roughly the same altitude. We define yaw as the rotation of the UAV about an axis pointing directly upwards from the body; roll is a rotation about the axis that connects the UAV tail to its nose; and pitch is a rotation about an axis which is orthogonal to the axes of rotation for pitch and yaw. Figure 3 shows the rotations that can affect an UAV during its flight. The change in these parameters changes what falls within the field-of-view (FOV) of the camera, [14] which leads to slightly different looks at the same scene. These looks contain similar, but not identical information which enables the reconstruction of a super-resolution image.

In such a case, an SR image can be reconstructed in an area of sufficient overlap between the LR images which generally have a relative shift with respect to each other by unknown subpixel amounts. When sufficient overlap exists, we use this set to reconstruct an SR image. We define a metric that depends on the sensor characteristics that can be used for this purpose. The FOV determines the angular extent of a given scene that is imaged by a camera. The horizontal FOV Φ_H and the vertical FOV Φ_V define the angle that the camera subtends on the ground. The area encompassed by the FOV is, of course, a function of

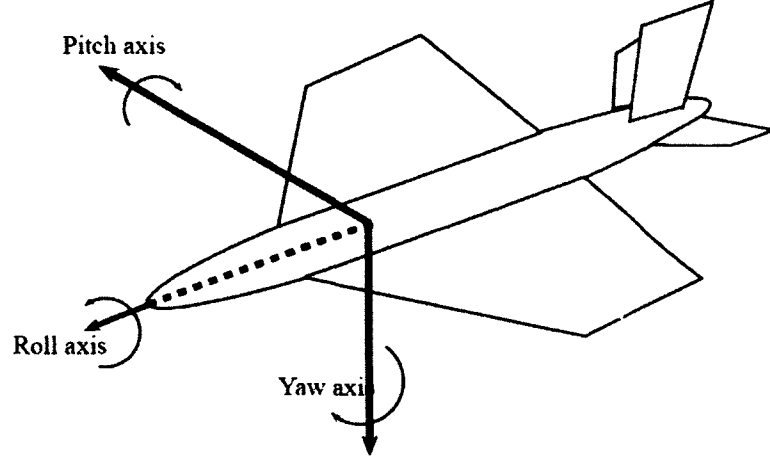


FIG. 3: UAV rotations: yaw, pitch and roll

the camera height above the ground. If the ground footprint due to the FOV of a camera at altitude h is bound by a rectangle with corners at (x_1, y_1) , $(-x_1, y_1)$, $(-x_1, -y_1)$ and $(x_1, -y_1)$, then, using Figure 4, we can derive the relationship between the extent of the footprint, i.e., the projection of the FOV on the ground, in the horizontal and vertical directions, the altitude h and the FOVs as [14]:

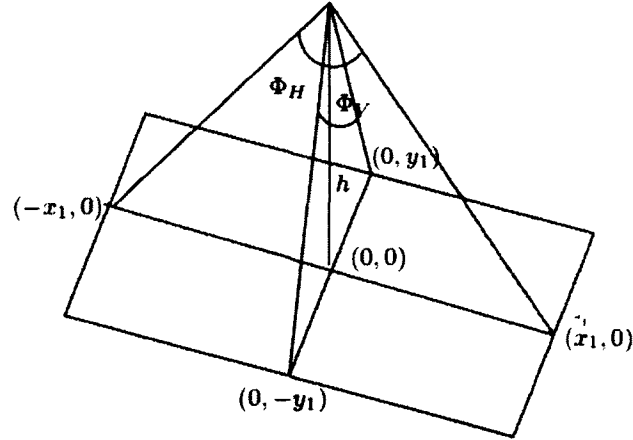


FIG. 4: FOV of the camera and the corresponding ground footprint

$$\begin{aligned} x_1 &= h \tan(\Phi_H/2) \\ y_1 &= h \tan(\Phi_V/2) \end{aligned} \tag{19}$$

Even when the altitude does not change, a change in the rotational parameters causes the camera footprint to change. Figure 5 shows the typical situation when the UAV experiences *pitch* of α° . Ideally this cause a shift in the x -direction only, causing $(x_1, 0)$ to be mapped to $(x_2, 0)$ and $((-x_1, 0)$ to be mapped to $(x_3, 0)$. The new x coordinates are given by:

$$\begin{aligned} x_2 &= h \tan(\Phi_H/2 - \alpha) \\ x_3 &= h \tan(\Phi_H/2 + \alpha). \end{aligned} \quad (20)$$

Similarly, when the UAV experiences a *roll* of β° , ideally only the y coordinates are affected, causing the $(0, y_1)$ coordinate to shift to a new location $(0, y_2)$ and the $(0, -y_1)$ coordinate to shift to $(0, y_3)$:

$$\begin{aligned} y_2 &= h \tan(\Phi_V/2 - \beta) \\ y_3 &= h \tan(\Phi_V/2 + \beta). \end{aligned} \quad (21)$$

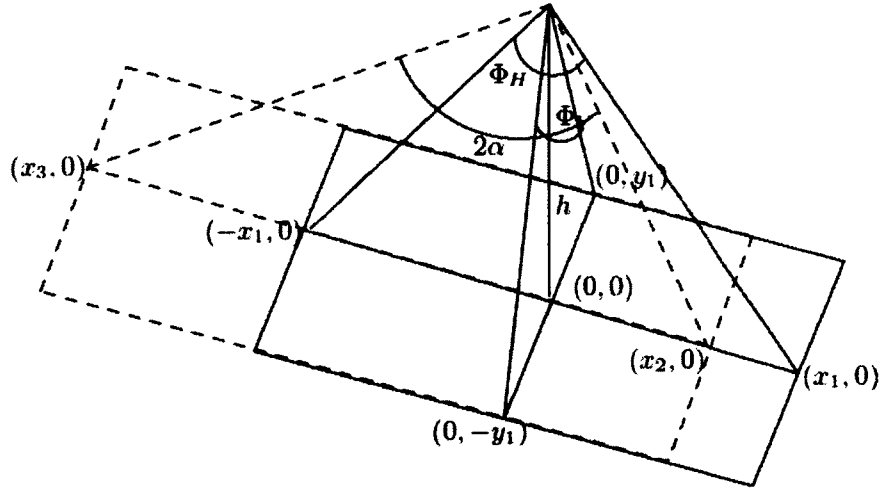


FIG. 5: Change in the x (y) coordinate as the UAV experiences a pitch of α° . Rotating the figure by 90° shows the case for changes due to a roll β° .

Figure 6 shows the typical situation when the UAV experiences a *yaw* of γ° . The yaw operation affects both the x and the y coordinates due to the rotation about the axis normal to the line connecting the tail and the nose of the UAV. The new footprint coordinates in

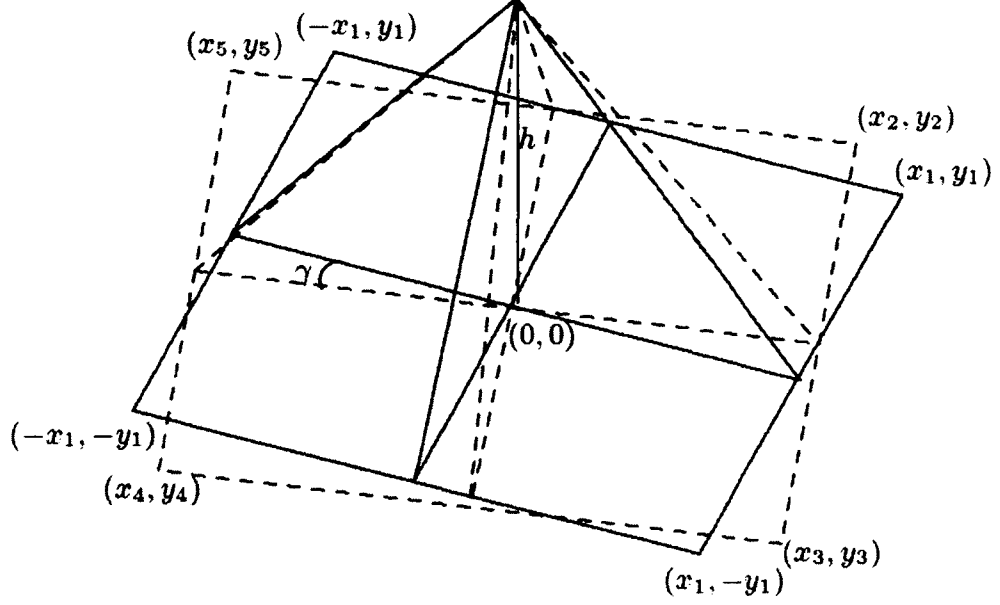


FIG. 6: Change in the x and y coordinates as the UAV experiences yaw = γ°

this case are given by:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} \cos(\gamma) & \sin(\gamma) \\ -\sin(\gamma) & \cos(\gamma) \end{bmatrix} \begin{bmatrix} \pm x_1 \\ \pm y_1 \end{bmatrix}, \quad i = 2, 3, 4, 5, \quad (22)$$

where (x_2, y_2) are obtained by applying the rotation to (x_1, y_1) , (x_3, y_3) from $(x_1, -y_1)$, (x_4, y_4) from $(-x_1, -y_1)$ and (x_5, y_5) from $(-x_1, y_1)$.

Assuming that two frames are captured at time instants t_1 and t_2 , the frame captured at t_1 is treated as the reference frame, i.e., with no rotational changes. All rotational and translational measurements for the frame captured at time t_2 are relative to the reference frame. Additionally, we assume that the change in altitude between the two frames is negligible. So, a metric can be obtained to check if the frames have sufficient overlap so that they can be combined to produce an SR image of the overlapped area. Figure 7 depicts the extreme condition where the two frames have no overlap. This allows us to develop the distance metric d between the two frames where the second frame has experienced roll, pitch and yaw with respect to the first frame. The metric d can be used to determine whether the degree of overlap between the two frames is sufficient to produce an SR image.

In order to determine d , we need first to compute the parameters for the case where the

UAV experiences all three rotational changes. We break this determination into two separate cases: When the UAV experiences pitch and roll, the new coordinates are given by:

$$\begin{aligned}
(x'_1, y'_1) &= (h \tan(\Phi_H/2 - \alpha), h \tan(\Phi_V/2 - \beta)) \\
(x'_2, y'_2) &= (-h \tan(\Phi_H/2 + \alpha), h \tan(\Phi_V/2 - \beta)) \\
(x'_3, y'_3) &= (-h \tan(\Phi_H/2 - \alpha), -h \tan(\Phi_V/2 + \beta)) \\
(x'_4, y'_4) &= (h \tan(\Phi_H/2 - \alpha), -h \tan(\Phi_V/2 + \beta))
\end{aligned} \tag{23}$$

where $x'_1 = x'_4, x'_2 = x'_3, y'_1 = y'_2$, and $y'_3 = y'_4$. Then if it experiences yaw of γ° , the new coordinates will be given by:

$$\begin{bmatrix} x''_i \\ y''_i \end{bmatrix} = \begin{bmatrix} \cos(\gamma) & \sin(\gamma) \\ -\sin(\gamma) & \cos(\gamma) \end{bmatrix} \begin{bmatrix} x'_i \\ y'_i \end{bmatrix}, \quad i = 1, 2, 3, 4 \tag{24}$$

Figure 7 shows two consecutive frames captured from the UAV. The second frame is under the effect of yaw, pitch and roll. The distance d between the centers of the two frames is given by

$$d = x_1 + |x'_2 \cos(\gamma)| + |y'_1 \sin(\gamma)|. \tag{25}$$

Then substituting for x'_2 and y'_1 in Equation (25) and taking into account the sign for pitch and roll, d becomes:

$$d = x_1 + |h \tan(\Phi_H/2 \pm \alpha) \cos(\gamma)| + |h \tan(\Phi_V/2 \pm \beta) \sin(\gamma)| \tag{26}$$

If we let d_n be the minimum distance when the two frames touch each other but there is no overlap, then overlap occurs whenever $d < d_n$. We use $d/d_n \leq 0.50$ as the condition that defines sufficient overlap. In other words, there has to be at least 50% overlap between the two frames to produce a reasonable SR image of the overlap area.

2.2 REGISTRATION BASED ON SPATIAL TRANSFORMATION

Spatial transformations reallocate the coordinates in one image to new coordinates in another image. The registration step depends on the accurate determination of the spatial transformation parameters used in the alignment process of the unregistered images. Usually, when the UAV experiences yaw, pitch and roll the captured image will be affected by shear, or it may be tilted, or it may have perspective distortion. So, spatial transformations like

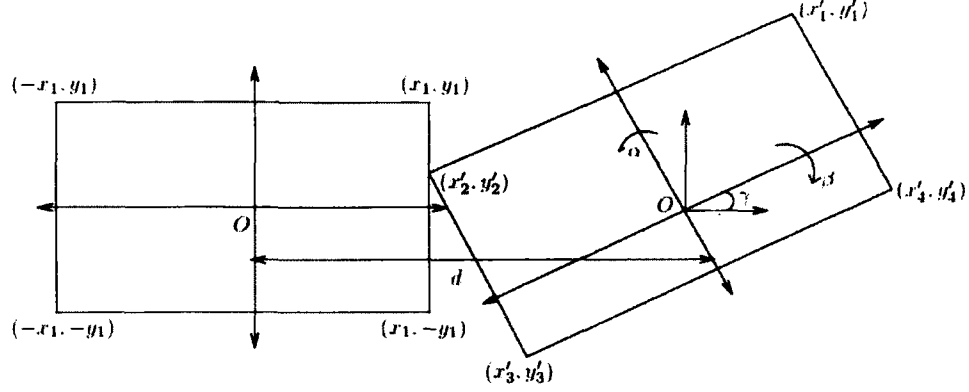


FIG. 7: Ground footprints of two consecutive images under the effect of the rotational parameters

the affine and the projective can be used to correct for those distortions. If the unregistered image is $f(x, y)$, its registered representation $r(x, y) = T(f(x, y))$, where T represents either the affine or the perspective transformation.

2.2.1 AFFINE TRANSFORMATION

The affine mapping is usually utilized when the image is affected by translation, rotation and scaling. If the base image is $b(x, y)$ and the unregistered image is $f(x', y')$ then, given three corresponding points in the two images, the transformation matrix can be calculated from

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix}, \quad i = 1, \dots, 3. \quad (27)$$

where (x_i, y_i) and (x'_i, y'_i) are the coordinates of the control points in the base and the unregistered images respectively; a_{11} , a_{12} , a_{21} and a_{22} control the scale, the rotation and the stretch; and a_{13} and a_{23} are the translation in the x - and y - directions. Since there are 6 unknowns, Equation (27) can be used to compute the transformation matrix T which can then be used to map f to b with 3 keypoint matches between the unregistered image and the reference one.

2.2.2 PROJECTIVE TRANSFORMATION

The projective mapping is usually utilized when the images have perspective distortion

and are tilted. A mapping h is projective if, and only if, there exists a non-singular 3×3 matrix H such that $h(p) = Hp$, $\forall p$, where $p = [x \ y]$ is the vector representation of a point. If the base image is $b(x, y)$ and the unregistered image is $f(x', y')$ then the projective transformation can be written as

$$x'(h_{31}x + h_{32}y + 1) = h_{11}x + h_{12}y + h_{13} \quad (28)$$

$$y'(h_{31}x + h_{32}y + 1) = h_{21}x + h_{22}y + h_{23} \quad (29)$$

These equations are linear in the elements of H . To solve for H , four keypoints matches between the reference image and the unregistered one. These keypoints should be in random pattern with no three points are collinear. After extracting the parameters of H , its inverse will be applied to the unregistered image to correct the perspective distortion.

In our experimental simulations, checkerboard images are used to simulate the UAV captured images. They contain the same scene but with different views. By controlling the camera orientation, the different images will contain the distortions that are expected when the UAV experiences yaw, pitch and roll. Figure 8 shows the results of changing the camera orientation on the base image. Figure 9 shows the results after spatial registration. The “dark” regions show the overlap region that can be used to produce SR images.

2.3 AUTOMATED CORRECTION OF UNREGISTERED UAV IMAGES

The first step of registration is to correct the UAV distorted images using an appropriate spatial transformation (affine or projective) to eliminate the effect of scale change, yaw, pitch or roll. The landmark points that control these spatial transformations are selected manually, which is subject to errors and is not powerful. Accordingly, an automatic selection of these point is done using the scale invariant feature transform (SIFT) [66]. The effectiveness of SIFT comes from its robustness to affine changes, noise, illumination change, and 3D view partial change. It extracts a set of unique features that can be used in image registration. First, the SIFT extracts all the available features from the reference image and store them in a database. Then, it extracts all the available features from the unregistered images and compares them against the stored database by means of nearest-neighborhood search to find the exact matches between the reference image and the unregistered one [67].

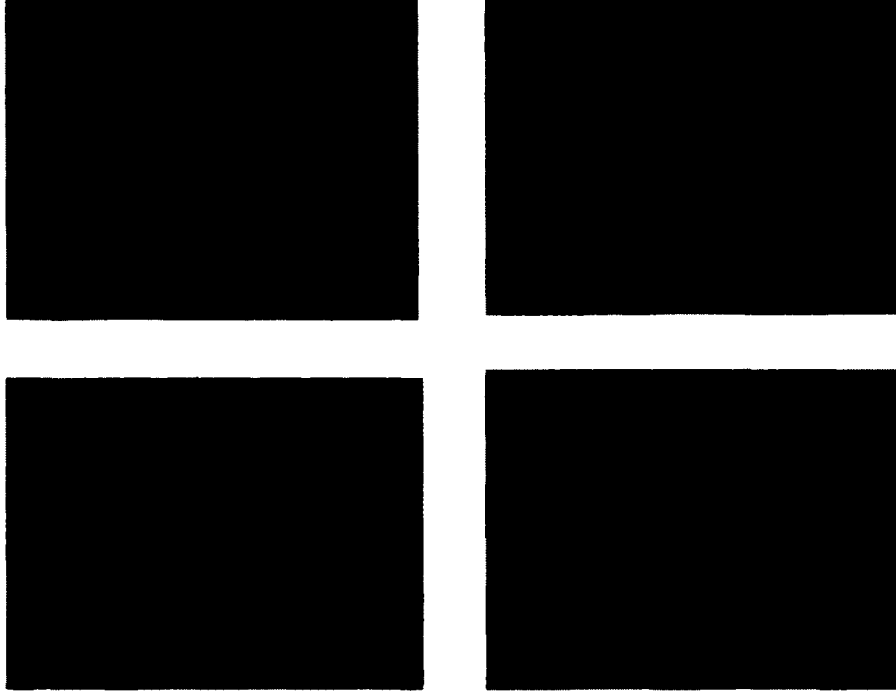


FIG. 8: Simulated high resolution scenes with different camera orientation to simulate yaw, pitch and roll: (top-left) reference image; (top-right) pitch; (bottom-left) yaw; (bottom right) roll

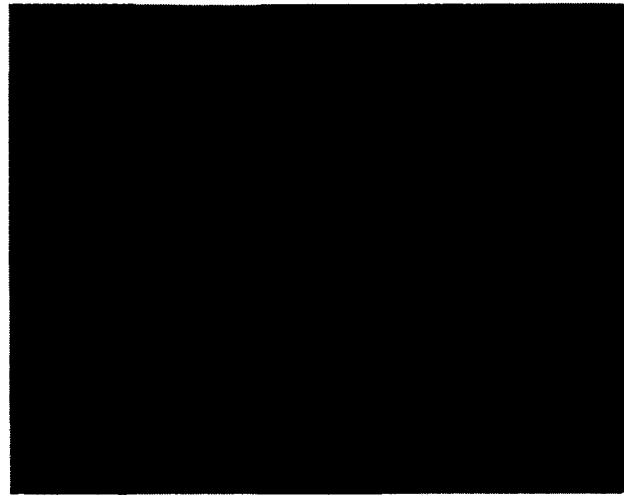
2.3.1 SIFT STAGES

SIFT consists of four major stages [66]: (1) scale-space creation; (2) keypoint allocation; (3) orientation designation; (4) keypoint descriptor. Once the keypoint descriptors have been assigned, they will be used to extract the parameters of the affine or the projective transformation.

Scale-Space Creation

Koendernik [68] and Lindeberg [69] have shown that the reasonable scale space kernel is the Gaussian function. In this sense, the scale space of an images, $SS(x, y, \sigma)$, is defined as:

$$SS(x, y, \sigma) = h(x, y, \sigma) * g(x, y), \quad (30)$$



(a) Registration with roll



(b) Registration with pitch



(c) Registration with yaw

FIG. 9: Registered images after different rotational parameters

where $g(x, y)$ is the image, $*$ refers to spatial convolution and $h(x, y, \sigma)$ is the Gaussian kernel defined by

$$h(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}. \quad (31)$$

To detect the stable features inside the image, SIFT uses difference-of-Gaussian, $DOG(x, y, \sigma)$, which is calculated from two adjacent scales differs from each other by a constant k . The $DOG(x, y, \sigma)$ is defined as:

$$\begin{aligned} DOG(x, y, \sigma) &= [h(x, y, k\sigma) - h(x, y, \sigma)] * g(x, y) \\ &= SS(x, y, k\sigma) - SS(x, y, \sigma) \end{aligned} \quad (32)$$

Figure 10 depicts the process of creating the DOGs at different scales. At the first octave, a set of different scale spaces are computed by convolving the processed image with Gaussian functions that differ in their scales by constant $k = 2^{1/s}$ where s is an integer. Once the scale spaces are created, the DOGs will be calculated by subtracting every scale space from its neighbor scale space. After processing the first octave, the subsequent octaves will be constructed by subsampling the scale spaces within the previous octave by a factor of 2, then the set of DOGs will be calculated similarly as they are computed in the first octave. After processing all octaves, a search for local extrema is conducted. This is done by comparing every point in the DOGs by its eight neighbors in the same scale and the nine neighbors at the above and lower scale and selected only if it is greater than those points.

Keypoint Allocation

After the keypoints have been selected, the next step is to eliminate those points that are insufficiently confined to an edge or have low contrast. The rejection criterion depends on the Taylor expansion of the DOG at the selected point as defined by:

$$DOG(\mathbf{x}) = DOG + \frac{\partial DOG}{\partial \mathbf{x}} \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 DOG}{\partial \mathbf{x}^2} \mathbf{x} \quad (33)$$

where the DOG and its derivatives are evaluated at the point \mathbf{X} that represents the shift from the origin which is set at the keypoint. The location of extremum is found by differentiating Equation (33) and setting it to zero. So, the location of the extremum is given by:

$$\hat{\mathbf{x}} = -\frac{\partial^2 DOG^{-1}}{\partial \mathbf{x}^2} \frac{\partial DOG}{\partial \mathbf{x}} \quad (34)$$

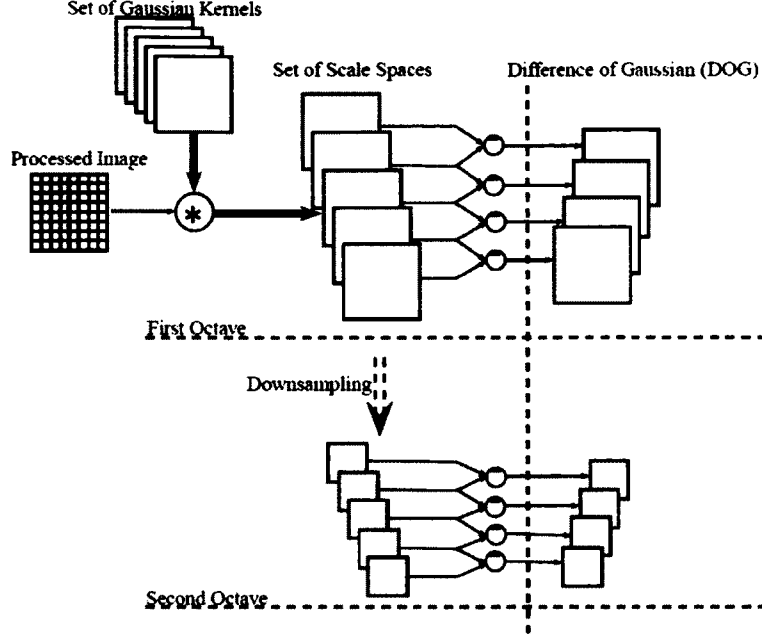


FIG. 10: Generation of scale spaces and DOGs

As suggested by Brown [70], the Hessian matrix and the derivative of the DOG are approximated by the difference of the adjacent points. If the estimated shift is greater than 0.5 then that extremum belongs to a different sample point and it's rejected. Also, to eliminate the unstable extremum with low contrast, the value of the DOG at the estimated shift, $DOG(\hat{x})$, is calculated and it's rejected whenever the value of $|DOG(\hat{x})|$ is less than 0.03. Because the DOG has a strong response at the edges, another elimination of incorrect keypoints is done. A poor peak is characterized by large principal curvature across the edge and a small curvature in the orthogonal direction. The principal curvature can be extracted from the Hessian matrix eigenvalues. The points are rejected whenever the ratio between the square of the Hessian trace to its determinant is greater than 10.

Orientation Designation

Each keypoint is assigned an orientation that will be used to generate the keypoint that is invariant to image rotation. At every keypoint, its scale space SS can be used to generate the gradient magnitude, $G(x, y)$, and the orientation, $\theta(x, y)$, as given by

$$G(x, y) = \sqrt{(SS(x+1, y) - SS(x-1, y))^2 + (SS(x, y+1) - SS(x, y-1))^2} \quad (35)$$

$$\theta(x, y) = \tan^{-1} ((SS(x, y + 1) - SS(x, y - 1)) / (SS(x + 1, y) - SS(x - 1, y))) \quad (36)$$

At each keypoint, an orientation histogram is formed using the available orientations in a region around the examined keypoint. The histogram is divided into 36 bin with every bin covering 10 degree of orientations. The histogram peak corresponds to the local gradient prominent direction and is assigned to be the orientation of the processed keypoint.

Keypoint Descriptor

The orientation designation characterizes each keypoint with image location, scale and orientation. In this step, the descriptors will be generated in such a way to be invariant to other variations such as illumination change and 3D viewpoint. A window of 16×16 samples around each keypoint is broken into 16 windows where the size of each window is 4×4 samples. For each sample of the 16 samples inside each window, the gradient magnitude is calculated and is added to a histogram consisting of 8 bins where their magnitudes are weighted according to the samples distance from the examined keypoint. By repeating this procedure for the 16 windows, each keypoint will be described by a vector of dimension 128. The vector is then normalized to a unit length to make it invariant to illumination change and 3D viewpoint.

Keypoint Matching

First, the sift is applied to the reference image and the extracted descriptors will be saved in a database. Then, it is applied to the unregistered images, and the extracted descriptors will be matched to the saved database through a nearest neighborhood search that minimizes the Euclidean distance between them.

2.3.2 SPATIAL TRANSFORMATIONS PARAMETERS EXTRACTION

In our studies, we need to correct the UAV captured images distortions by means of affine or projective transformations to eliminate the effect of the yaw, pitch, roll and altitude change. Once the match between the reference and the unregistered image has been done, the affine transformation will have 6 degrees of freedom (DOF), so it needs 3 keypoint matches, while the projective transformation has 8 degrees of freedom, so it needs 4 matches to derive the transformation parameters. With 3 keypoints match, the affine transformation given in

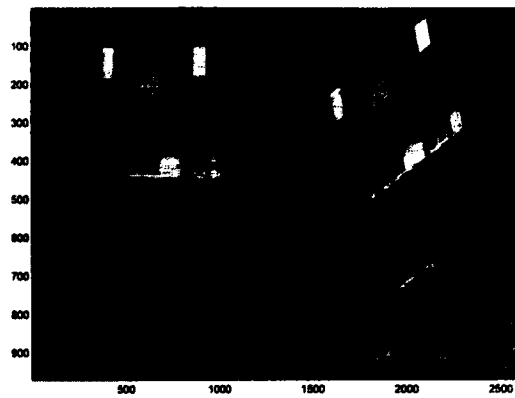
Equation (27) can be rewritten as:

$$\begin{bmatrix} x'_1 & y'_1 & 0 & 0 & 1 & 0 \\ 0 & 0 & x'_1 & y'_1 & 0 & 1 \\ x'_2 & y'_2 & 0 & 0 & 1 & 0 \\ 0 & 0 & x'_2 & y'_2 & 0 & 1 \\ x'_3 & y'_3 & 0 & 0 & 1 & 0 \\ 0 & 0 & x'_3 & y'_3 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \\ a_{13} \\ a_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ x_3 \\ y_3 \end{bmatrix}, \quad (37)$$

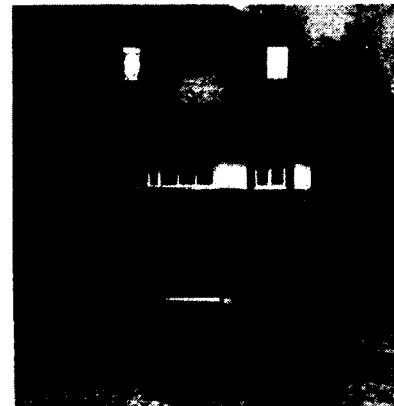
which can be written as a system of linear equations $\mathbf{AX} = \mathbf{B}$. The solution of this system can be obtained through the least square method given by the normal equation $\mathbf{X} = [\mathbf{A}^T \mathbf{A}]^{-1} \mathbf{A}^T \mathbf{B}$ which minimizes the sum of the distance between the corresponding locations in the reference and the unregistered images. Similarly, the projective transformation in Equation (29) can be rewritten as

$$\begin{bmatrix} x'_1 & y'_1 & 1 & 0 & 0 & 0 & -x'_1 x_1 & -y'_1 x_1 \\ 0 & 0 & 0 & x'_1 & y'_1 & 1 & -x'_1 y_1 & -y'_1 y_1 \\ x'_2 & y'_2 & 1 & 0 & 0 & 0 & -x'_2 x_2 & -y'_2 x_2 \\ 0 & 0 & 0 & x'_2 & y'_2 & 1 & -x'_2 y_2 & -y'_2 y_2 \\ x'_3 & y'_3 & 1 & 0 & 0 & 0 & -x'_3 x_3 & -y'_3 x_3 \\ 0 & 0 & 0 & x'_3 & y'_3 & 1 & -x'_3 y_3 & -y'_3 y_3 \\ x'_4 & y'_4 & 1 & 0 & 0 & 0 & -x'_4 x_4 & -y'_4 x_4 \\ 0 & 0 & 0 & x'_4 & y'_4 & 1 & -x'_4 y_4 & -y'_4 y_4 \end{bmatrix} = \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ y_2 \\ \vdots \\ y_4 \end{bmatrix}, \quad (38)$$

which can be solved using the normal equation such as in the affine case. Figure 11 shows the results of applying the SIFT to a set of home images with different orientations to simulate the distortions of shear, tilt and rotations. Also, Figure 12 and 13 show the results of registering two different sets of aerial images. The left column in these three figures shows the corresponding keypoints or landmarks matches between the reference images and the unregistered ones while the right column shows the images after the registration with dark regions showing the areas of overlapping between these processed images that can be used to reconstruct SR images. Once the appropriate transformation is applied to the unregistered image, the reprojected images will be registered within subpixel accuracy using the enhanced SSDFT approach.



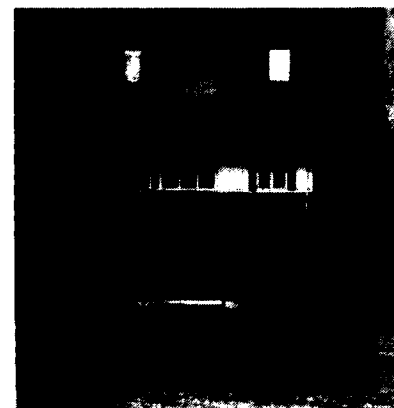
(a) Matched features between base and LR1



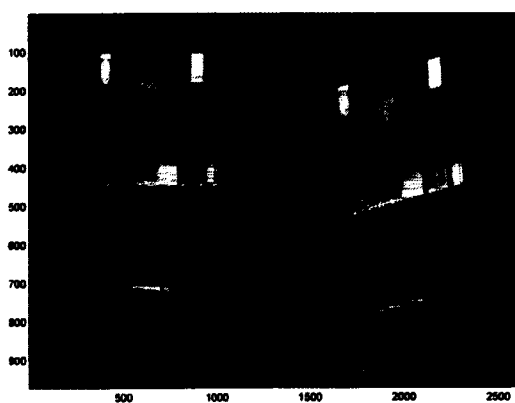
(b) Corrected LR1



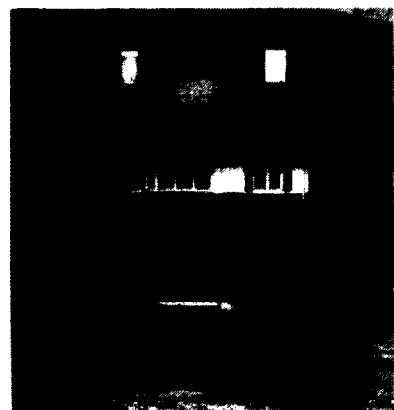
(c) Matched features between base and LR2



(d) Corrected LR2

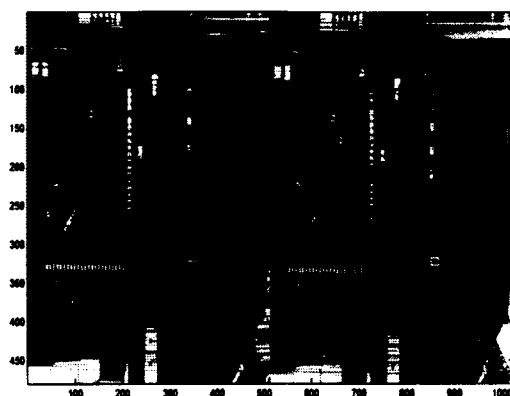


(e) Matched features between base and LR3

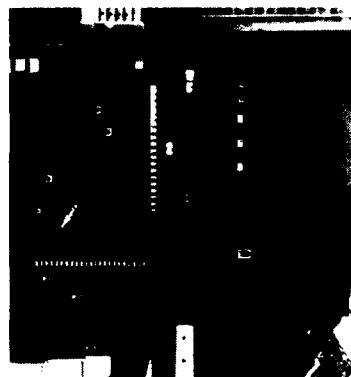


(f) Corrected LR3

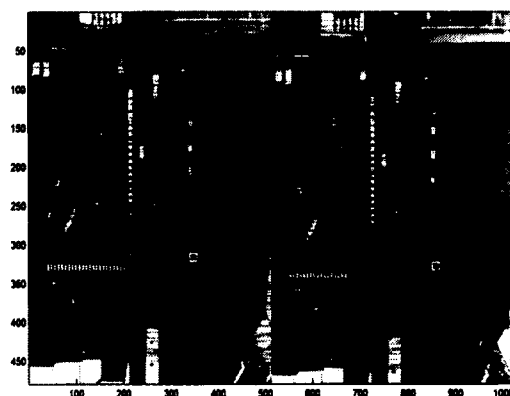
FIG. 11: Corrected LR images after using the SIFT algorithm (home images)



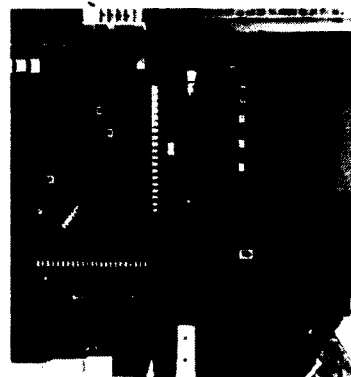
(a) Matched features between base and LR1



(b) Corrected LR1



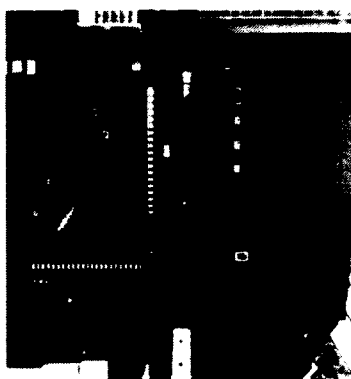
(c) Matched features between base and LR2



(d) Corrected LR2

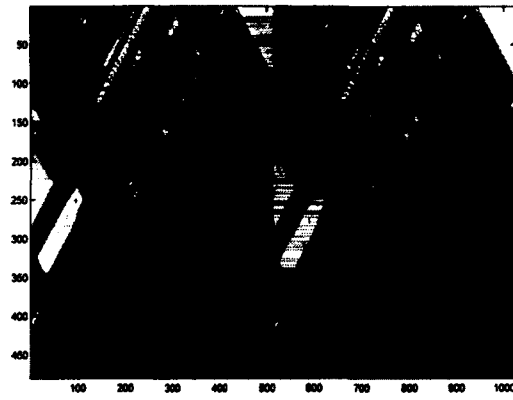


(e) Matched features between base and LR3



(f) Corrected LR3

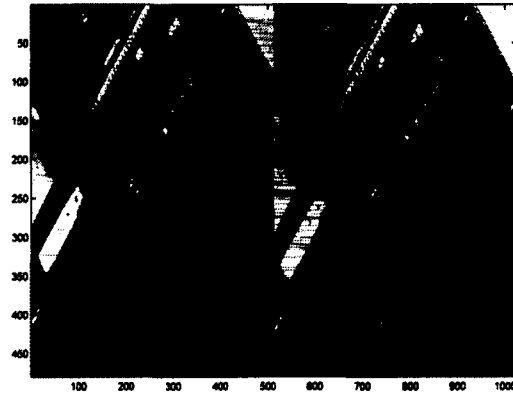
FIG. 12: Corrected LR images after using the SIFT algorithm (aerial images 1)



(a) Matched features between base and LR1



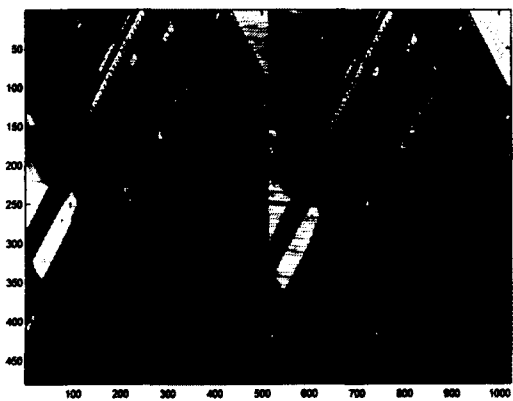
(b) Corrected LR1



(c) Matched features between base and LR2



(d) Corrected LR2



(e) Matched features between base and LR3



(f) Corrected LR3

FIG. 13: Corrected LR images after using the SIFT algorithm (aerial images 2)

2.4 FREQUENCY-DOMAIN BASED REGISTRATION

Up to this point, the UAV images are restored from shear, tilt, and perspective distortions. The following step is to align them within subpixel accuracy, which is considered a crucial step in any SR reconstruction algorithm. Without accurate subpixel registration, severe artifacts will be present within the reconstructed SR images. Many applications require the accuracy of the registration to be within a small portion of a pixel such as in medical imaging, computer vision [50, 71] and remote sensing. In the latter, a pixel in Landsat images measures approximately 80 m on the earth, so 0.1 pixel registration accuracy will lead to a resolution of 8 m [21].

For the case of a translation between two images, the usual technique to address this problem is to compute the cross-correlation between the unregistered and the base images by means of discrete Fourier transform (DFT), and locate its peak [72]. If the image to be registered is $g(x, y)$ and the base image is $f(x, y)$, then the normalized mean square error (NRMSE) E^2 between two images is defined as [73]:

$$\begin{aligned}
 E^2 &= \min_{\alpha, x_0, y_0} \frac{\sum_{x,y} |\alpha g(x - x_0, y - y_0) - f(x, y)|^2}{\sum_{x,y} |f(x, y)|^2} \\
 &= 1 - \frac{\max_{x_0, y_0} |r_{fg}(x_0, y_0)|^2}{\sum_{x,y} |f(x, y)|^2 \sum_{x,y} |g(x, y)|^2}
 \end{aligned} \tag{39}$$

where r_{fg} is the cross-correlation of $f(x, y)$ and $g(x, y)$ is defined by:

$$\begin{aligned}
 r_{fg}(x_0, y_0) &= \sum_{x,y} f(x, y) g^*(x - x_0, y - y_0) \\
 &= \sum_{\mu, \nu} \hat{F}(\mu, \nu) \hat{G}^*(\mu, \nu) \exp \left(i2\pi \left(\frac{\mu x_0}{M_1} + \frac{\nu y_0}{M_2} \right) \right).
 \end{aligned} \tag{40}$$

M_1 and M_2 are the image dimensions; * denotes complex conjugation; and $\hat{F}(\mu, \nu)$ and $\hat{G}(\mu, \nu)$ are the DFTs of $f(x, y)$ and $g(x, y)$, respectively. The 2D-DFT of an image $f(x, y)$ is defined by

$$\hat{F}(\mu, \nu) = \frac{1}{M_1 M_2} \sum_{x=0}^{M_1-1} \sum_{y=0}^{M_2-1} f(x, y) \exp -i2\pi \left(\frac{x\mu}{M_1} + \frac{y\nu}{M_2} \right) \tag{41}$$

The evaluation of the NRMSE requires solving the more general problem of sub-pixel image registration by locating the peak of cross-correlation $r_{fg}(x, y)$. The usual DFT approach to find the cross-correlation peak to within a fraction, $1/\epsilon$, of a pixel is to

1. compute $\hat{F}(\mu, \nu)$ and $\hat{G}(\mu, \nu)$,
2. embed the product $\hat{F}(\mu, \nu)\hat{G}^*(\mu, \nu)$ in a larger array of zeros of dimension $(\epsilon M, \epsilon N)$,
3. compute the inverse DFT to obtain the up-sampled cross-correlation, and
4. locate its peak.

Although this approach is very accurate and robust to moderate noise, its computational complexity and huge memory requirements make it unrealistic even for small dimension images with large upsampling factors. The computational complexity of this approach is $O\{M_1 M_2 \epsilon [\log_2(\epsilon M_1) + \epsilon \log_2(\epsilon M_2)]\}$ [72] where ϵ is the upsampling factor and M_1 and M_2 are the image dimensions.

Two algorithms were reviewed and compared against the enhanced speedy proposed approach. The first algorithm is non-linear Optimization gradient routine (NLOGR) and the other one is single step discrete Fourier transform (SSDFT). Both algorithms start with an initial estimate for the location of the cross correlation peak by means of usual FFT approach with upsampling factor ϵ_o of 2 which means that the location of the peak is within ± 0.5 pixel accuracy. Most of the time required for the registration process is consumed in this step, which is considered the main drawback of these algorithms. The SSDFT approach is classified as one of the most reliable and efficient subpixel registration algorithms [26], and it is faster and more computationally efficient than NLOGR approach. A speeded-up version of the SSDFT approach is proposed without sacrificing the required accuracy and it is roughly 5X times faster than the original SSDFT approach and can be used efficiently in the case of large dimension images.

2.4.1 NON-LINEAR OPTIMIZATION GRADIENT ROUTINE

The algorithm refines the initial estimate obtained by the usual FFT2X approach using a nonlinear-optimization conjugate-gradient routine to maximize $|r_{fg}(x_0, y_0)|^2$. The partial

derivative of r_{fg} with respect to x_0 is given by

$$\begin{aligned} \frac{\partial |r_{fg}(x_0, y_0)|^2}{\partial x_0} &= 2\mathcal{I} \left(r_{fg}(x_0, y_0) \sum_{\mu, \nu} \frac{2\pi\mu}{M_1} \hat{F}^*(\mu, \nu) \hat{G}(\mu, \nu) \right. \\ &\quad \times \left. \exp \left(i2\pi \left(\frac{\mu x_0}{M_1} + \frac{\nu y_0}{M_2} \right) \right) \right) \end{aligned} \quad (42)$$

with a similar expression for the partial derivative with respect to y_0 . The algorithm iteratively searches for the image displacement (x_0, y_0) that maximizes $r_{fg}(x_0, y_0)$ and can achieve registration precision to within an arbitrary fraction of a pixel. Assuming that the usual FFT2X initial estimate is $\mathbf{X}_{(0)} = (x_{(0)}, y_{(0)})$, then the steps to refine this estimate are:

1. compute the gradient of r_{fg} at $(x_{(0)}, y_{(0)})$ as given by

$$\nabla r_{fg}(\mathbf{X}_{(0)}) = \left(\frac{\partial r_{fg}}{\partial x_0}, \frac{\partial r_{fg}}{\partial y_0} \right) |_{(x_{(0)}, y_{(0)})} \quad (43)$$

2. start with $\mathbf{d}_{(0)} = \mathbf{r}_{(0)} = \nabla r_{fg}(\mathbf{X}_{(0)})$.
3. find $\alpha_{(i)}$ that minimizes $\nabla |r_{fg}(\mathbf{X}_{(i)} + \alpha_{(i)} \mathbf{d}_{(i)})|^2$
4. update:

$$(a) \quad \mathbf{X}_{(i+1)} = \mathbf{X}_{(i)} + \alpha_{(i)} \mathbf{d}_{(i)},$$

$$(b) \quad \mathbf{r}_{(i+1)} = \nabla r_{fg}(\mathbf{X}_{(i+1)})$$

5. calculate $\beta_{(i+1)}$ as given by

$$\beta_{(i+1)} = \frac{\mathbf{r}_{(i+1)}^T \mathbf{r}_{(i+1)}}{\mathbf{r}_{(i)}^T \mathbf{r}_{(i)}} \quad (44)$$

6. calculate $\mathbf{d}_{(i+1)}$ as given by

$$\mathbf{d}_{(i+1)} = \mathbf{r}_{(i+1)} + \beta_{(i+1)} \mathbf{d}_{(i)} \quad (45)$$

7. stop when the maximum iterations exceeded a certain number or $\|\mathbf{r}_{(i)}\| \leq \epsilon \|\mathbf{r}_{(i+1)}\|$ with $\epsilon < 1$.

The parameter $\alpha_{(i)}$ can be obtained using general line search using the Newton-Raphson Method. If we let $\nabla |r_{fg}(\mathbf{X}_{(i)} + \alpha_{(i)} \mathbf{d}_{(i)})|^2 = f(x + \alpha d)$, then the Taylor expansion of the

function f is given by

$$\begin{aligned} f(x + \alpha d) &\approx f(x) + \alpha \left[\frac{d}{d\alpha} f(x + \alpha d) \right]_{\alpha=0} + \frac{\alpha^2}{2} \left[\frac{d^2}{d\alpha^2} f(x + \alpha d) \right]_{\alpha=0} \\ &= f(x) + \alpha [f'(x)]^T d + \frac{\alpha^2}{2} d^T f''(x) d \end{aligned} \quad (46)$$

Differentiating this equation with respect to α yields

$$\frac{d}{d\alpha} f(x + \alpha d) \approx [f'(x)]^T d + \alpha d^T f''(x) d \quad (47)$$

The function $f(x + \alpha d)$ is minimized by setting $\frac{d}{d\alpha} f(x + \alpha d)$ to zero and hence

$$\alpha = -\frac{f'^T d}{d^T f'' d} \quad (48)$$

where $f''(x)$ is the Hessian matrix defined by

$$f''(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} \end{pmatrix} \quad (49)$$

2.4.2 SINGLE STEP DISCRETE FOURIER TRANSFORM APPROACH

The second efficient subpixel algorithm was developed by Guizar-Sicairos et al. [72] to register images with the same accuracy obtained by the usual FFT approach but with a huge reduction in computational time and memory requirements. Their technique was classified as one of the most reliable and best algorithms to register images using phase correlation methods [26]. The SSDFT works on two steps. The first step, which is similar to NLOGR, finds an initial estimate for the location of the cross correlation peak between two images using the usual FFT approach with upsampling factor of $\epsilon_0 = 2$. The usual FFT approach causes a tremendous waste of memory and processing time as it must process the entire zero padded upsampled matrix of dimensions $(\epsilon M_1, \epsilon M_2)$ to get the accurate peak location. On the contrary, the SSDFT approach searches for the accurate peak in a small window around the initial estimate by means of DFT instead of FFT. It utilizes the DFT implementation to obtain an upsampled version of the cross correlation in a small window of size $1.5\epsilon \times 1.5\epsilon$ around the initial estimate without zero padding the product $\hat{F}(\nu, \omega) \hat{G}^*(\nu, \omega)$. This process

is implemented by rewriting Equation (41) as a product of three matrices of dimensions $(1.5\epsilon, M_1)$, (M_1, M_2) , and $(M_2, 1.5\epsilon)$. Then a search for the peak is done over the output matrix of size $(1.5\epsilon, 1.5\epsilon)$. The computational complexity of this approach is $O(M_1 M_2 \epsilon)$, which is a great improvement to the usual FFT approach.

Although the SSDFT approach is an efficient subpixel registration algorithm, its main disadvantage is that most of the time needed for registration is consumed in searching for the initial estimate. This drawback will be tremendously improved in our proposed approach which greatly reduces the time required for locating the initial estimate and also reduces the time required for the refinement step.

2.4.3 SPEEDED-UP SSDFT APPROACH

We enhanced the SSDFT approach by reducing the required time for its two steps, i.e., the time required for the initial estimation of the peak location; and the time needed for the refinement step. Our approach to reduce the computational time for estimating the initial peak location depends on reducing the dimension of the Fourier transform of the cross correlation matrix and by applying the inverse FFT, the initial estimate can be obtained in a faster way than the SSDFT approach.

Suppose we replaced the Fourier transform of both the reference image $\hat{F}_1(\nu, \omega)$ and the unregistered version $\hat{F}_2(\nu, \omega)$ by a sampled version of them; then the right hand side of Equation (41) becomes

$$\sum_{\mu, \nu} \sum_{m, n} \hat{F}(\mu, \nu) \hat{G}^*(\mu, \nu) \delta(\mu - K_1 m) \delta(\nu - K_2 n) \times \exp \left[i 2\pi \left(\frac{d_x \mu}{M_1} + \frac{d_y \nu}{M_2} \right) \right]; \quad (50)$$

where K_1 and K_2 are the sampling factors along the x - and y - directions respectively, and δ is the Dirac delta function. Rearranging the sums and by using the sifting property of the Dirac delta function, the last equation can be written as

$$\sum_{m, n} \hat{F}(K_1 m, K_2 n) \hat{G}^*(K_1 m, K_2 n) \times \exp \left[i 2\pi \left(\frac{d_x m}{M_1/K_1} + \frac{d_y n}{K_2/M_2} \right) \right]; \quad (51)$$

which represents the inverse Fourier transform of the product at the new reduced dimension.

TABLE 1: Optimum sampling factors for different image sizes

Image size	Sampling factor K
128	4
256	8
512	16
1024	32
2048	64
4096	128
8192	256

So, the idea here is to sample the Fourier transform of the two images being registered and to apply the same SSDFT approach in searching for the initial estimate of the peak location. The computational complexity of our enhancement is $O(K_1^{-1}K_2^{-1}M_1M_2\epsilon)$ which is a great improvement over the SSDFT approach.

The selection of the sampling factors K_1 and K_2 will depend on the image size, preferably of base of 2 to gain the full power of the FFT. Table 1 shows the sampling factors for different image sizes. If K_{128} is the sampling factor for an image of size 128×128 , then the sampling factor for an image of size $M \times M$ can be extracted as $K_M = M/128 * K_{128}$. These sampling factors can make the computational time required for the initial estimate of the peak location to be roughly the same.

The second enhancement of the SSDFT approach reduces the number of matrix multiplications required to find the accurate location in the upsampled cross correlation window by minimizing the number of matrix multiplications required to obtain a partial inverse DFT matrix. Consider Equation (41) to be written as a product of three matrices as given by

$$C_{f_1f_2} = A_{1.5\epsilon \times M_1} * B_{M_1 \times M_2} * C_{M_2 \times 1.5\epsilon}; \quad (52)$$

then the SSDFT approach searches for the accurate location by multiplying the whole three matrices and then searching for the peak in the resultant matrix $C_{f_1f_2}$, which consumes sometime as not all the values inside the output matrix are required.

In our approach we overcome this weakness by partially obtaining some values inside the resultant matrix and accurately determining the peak location using a forward and backward search. It depends on the idea that the peak of the cross correlation will be close to one

of the borders of the output upsampled cross correlation matrix (row or column). Figure 14 shows an upsampled cross correlation version of a window of size 150×150 (upsampling factor =100) around the initial peak location, and Figure 15 shows its level curves. Usually, the location of peak will be close to one of the borders of the upsampled cross correlation window. By calculating partial parts of upsampled window, we can speed up the search

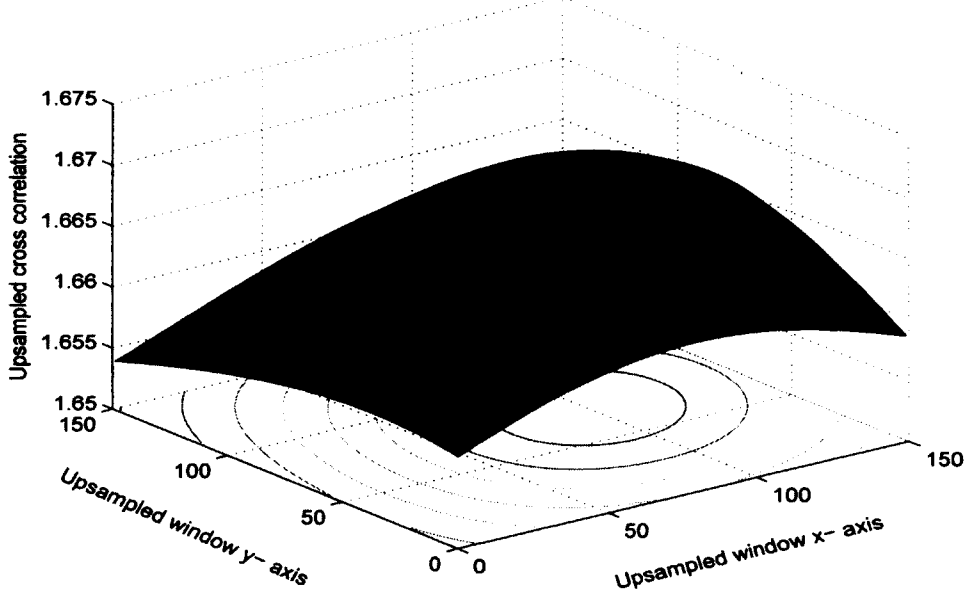


FIG. 14: Upsampled cross correlation with window size 150×150

for the accurate peak location. This can be done using the following steps:

1. calculate the output matrix borders R_1 , R_{M_1} , C_1 , and C_{M_2} as defined by

$$\begin{aligned} R_i &= (A(i,:) * B) * C, \quad i = 1 \text{ or } M_1 \\ C_j &= A * (B * C(:,j)), \quad j = 1 \text{ or } M_2 \end{aligned} \quad (53)$$

where the use of “ $()$ ” specify the order of matrix multiplication,

2. find the max value across these borders and assume for example it is across the first row $R(1,:)$,
3. from $R(1,:)$ the algorithm starts a forward search in steps of σ to find the next max value across the rows $R(\sigma i + 1,:)$ where i refers to the iteration number and it stops

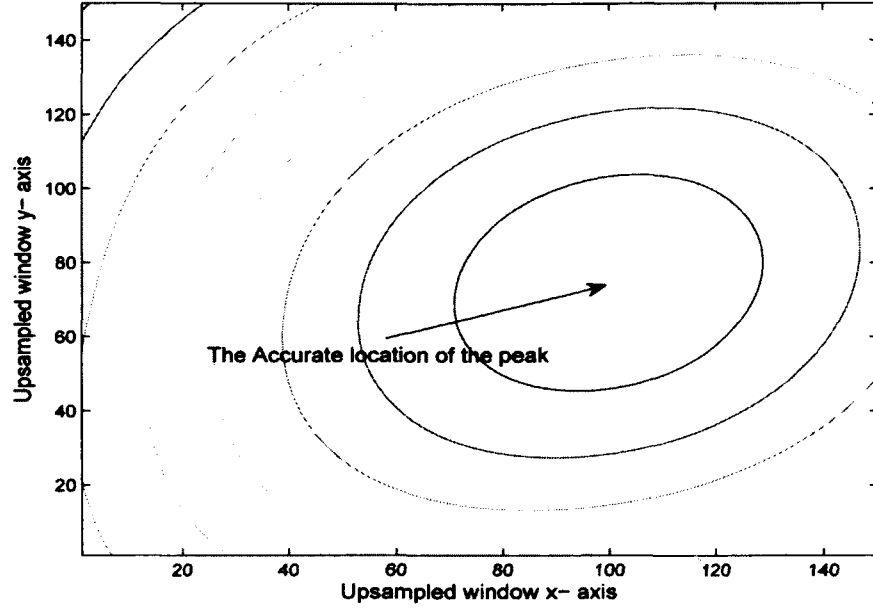


FIG. 15: Level curves for the upsampled cross correlation

when the next maximum value drops below the previous one.

4. from the last scanned row, the algorithm starts a backward search with a decrement of 1 until the next maximum value is less than the previous one.

At the last scanned row, the algorithm accurately finds the maximum peak and its location.

To evaluate our improvements against the SSDFT approach, we compare the performance of the two approaches against different image sizes ranging from 256×256 up to 8192×8192 in a multiple of 2. The images are corrupted by additive white Gaussian noise and blurred by a Gaussian kernel to simulate the optical lens blur. Also, they are shifted by a lateral shifts of $(3.48574, 7.73837)$ in pixels to obtain the unregistered versions of them. The simulations are performed using MATLAB 7.8 Release 2009a program on OPTIPLEX 780 (Intel(R) Core (TM)2 Quad 2.66 GHz CPU, 8.00 GB RAM, MS Windows 7 Professional 2009). The SSDFT technique and the enhanced one register images with accuracy of 0.01 pixel. The estimated shifts for both algorithms are $(3.49, 7.74)$ with estimation error $\delta s = 0.00456$ of a pixel where the estimation error is given by the l_2 -norm between the actual shifts (d_x, d_y)

and the estimated shifts $(\tilde{d}_x, \tilde{d}_y)$ as given by

$$\delta s = \sqrt{(d_x - \tilde{d}_x)^2 + (d_y - \tilde{d}_y)^2} \quad (54)$$

Figure 16 and 17 show the comparison between the SSDFT approaches (the original and the enhanced one) and the NLOGR in terms of the estimated NRMSE \hat{E} and the estimated error δs . The greater the upsampling factor the better is the estimate for the subpixel shifts and the lower value of the NRMSE. A comparison of the computational time required for

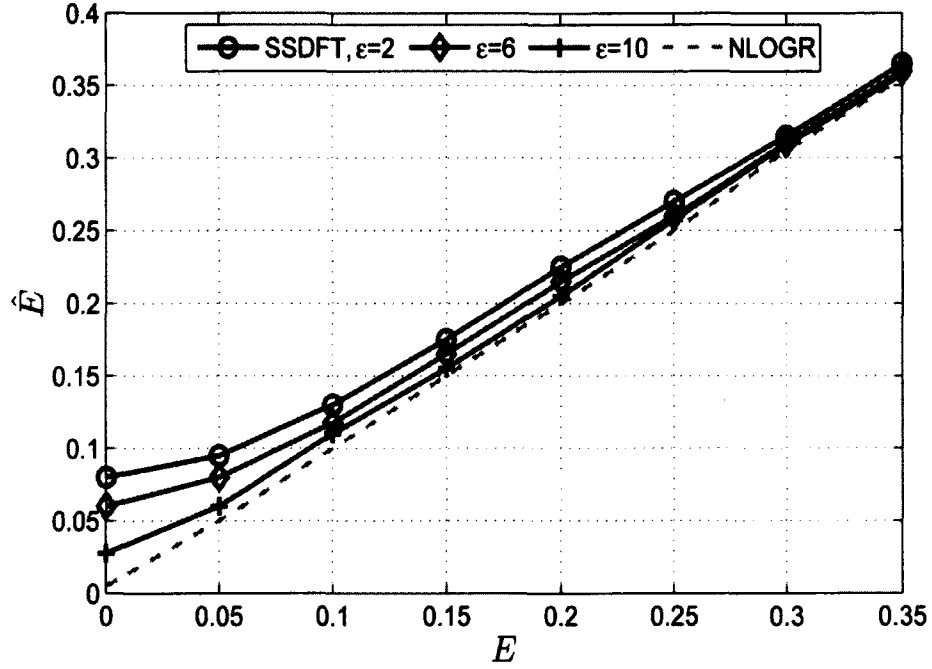


FIG. 16: The estimated NRMSE \hat{E} against the actual one

the initial peak estimation, the refinement step, and the total registration time are shown in Figure 18, Figure 19, and Figure 20, respectively. It can be seen that our approach tremendously reduces the amount of time required to obtain the initial estimate for the peak location compared to the SSDFT approach. Also, our approach makes the time required for this step roughly the same regardless of the image size which can be done by controlling the selection of the sampling factors K_1 and K_2 . For example, for an 8192×8192 image it requires around 2 milliseconds using our approach while it needs around 200 seconds using the SSDFT approach. In the refinement step, our method enhances the performance of the SSDFT approach as can be seen in Figure 19. Throughout simulations, we set the forward

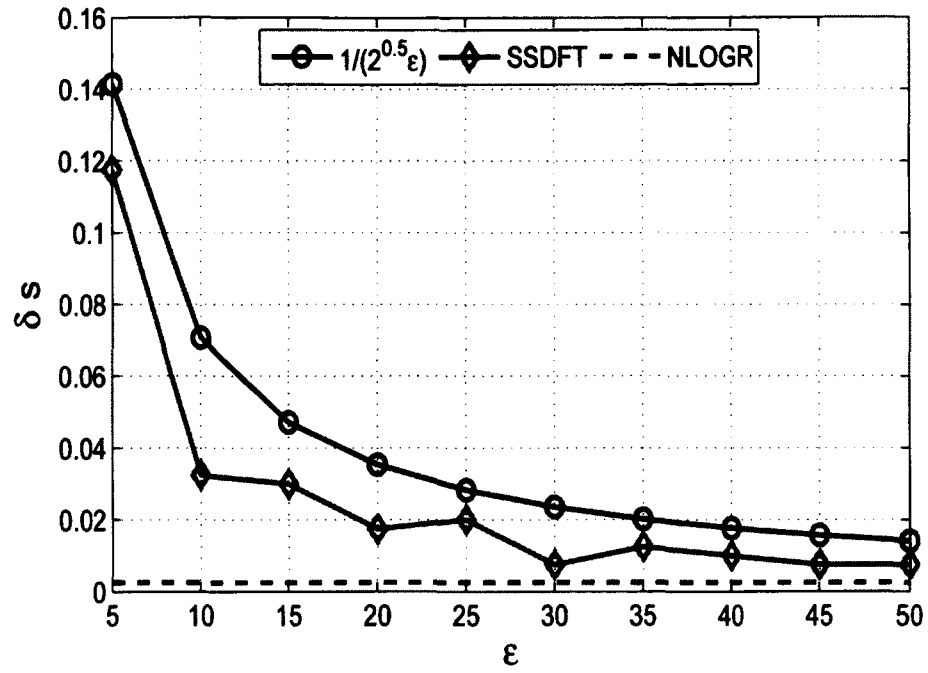


FIG. 17: The upsampling factor ϵ against the estimation error δs

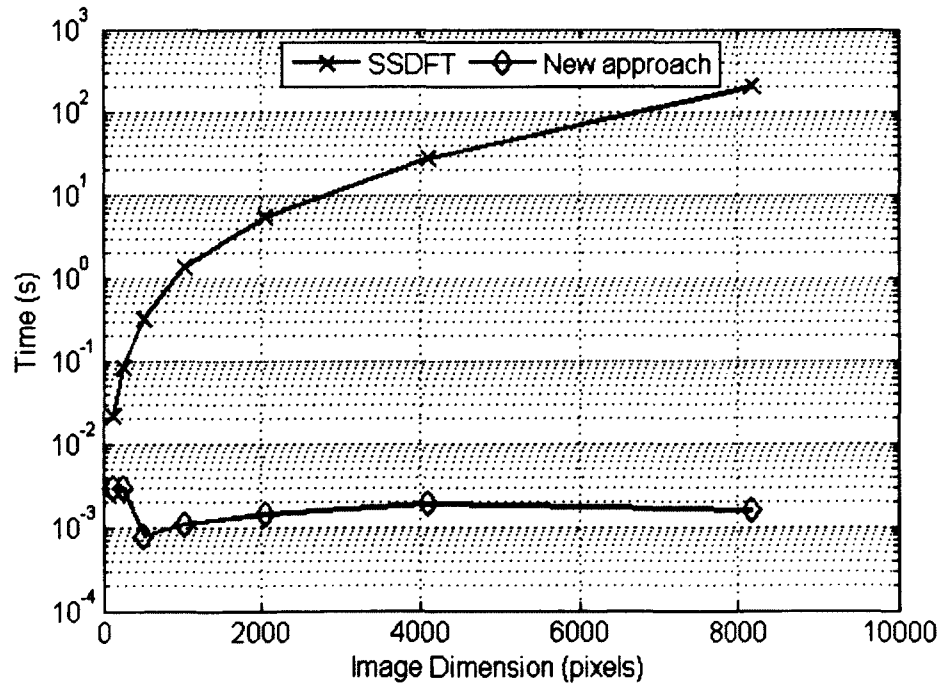


FIG. 18: Computational time required for the initial estimate

TABLE 2: Total time required to register images of dimension 512×512 with different subpixel accuracy.

Subpixel accuracy (pixels)	Enhanced SSDFT (seconds)	SSDFT (seconds)
0.1	0.15	0.54
0.01	0.2	0.63
0.001	1.4	2.98
0.0001	75.13	151.56

step σ to 0.3ϵ . The total computational time required for the whole registration process for both approaches is shown in Figure 20. Over all, our approach is approximately 5X faster than the SSDFT approach.

Also, we test the performance of both algorithms against changing the required subpixel accuracy, which can be seen in Table 2. Our enhancement increases the attainable subpixel accuracy with a large decrease in the computational time.

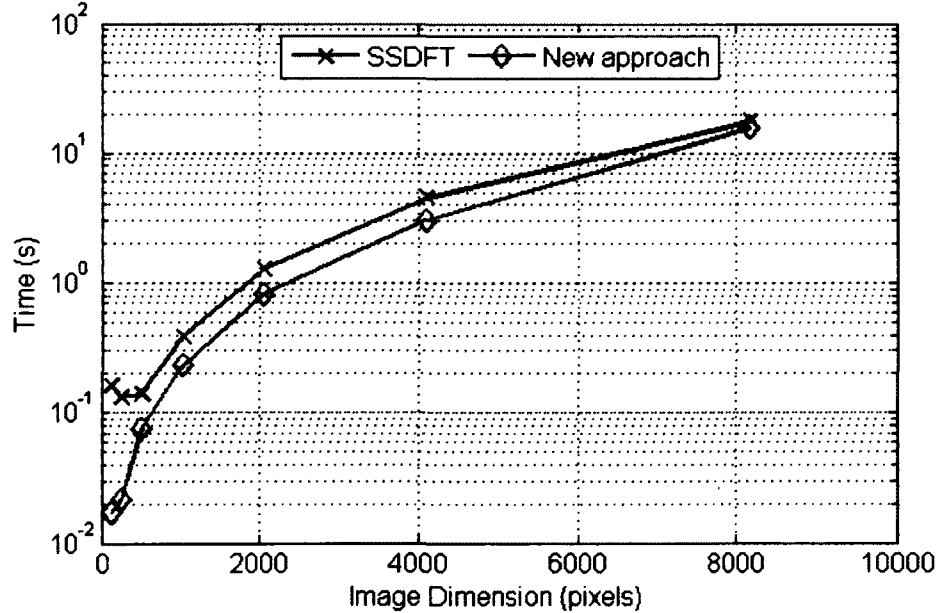


FIG. 19: Computational time for the refinement step

The proposed enhanced approach registers images that differ by translational shifts or scaled by a constant and can be used in the presence of moderate noise. Our enhanced approach offers a great reduction in computational time and memory requirements against the SSDFT and the usual FFT approaches without sacrificing the required accuracy. It

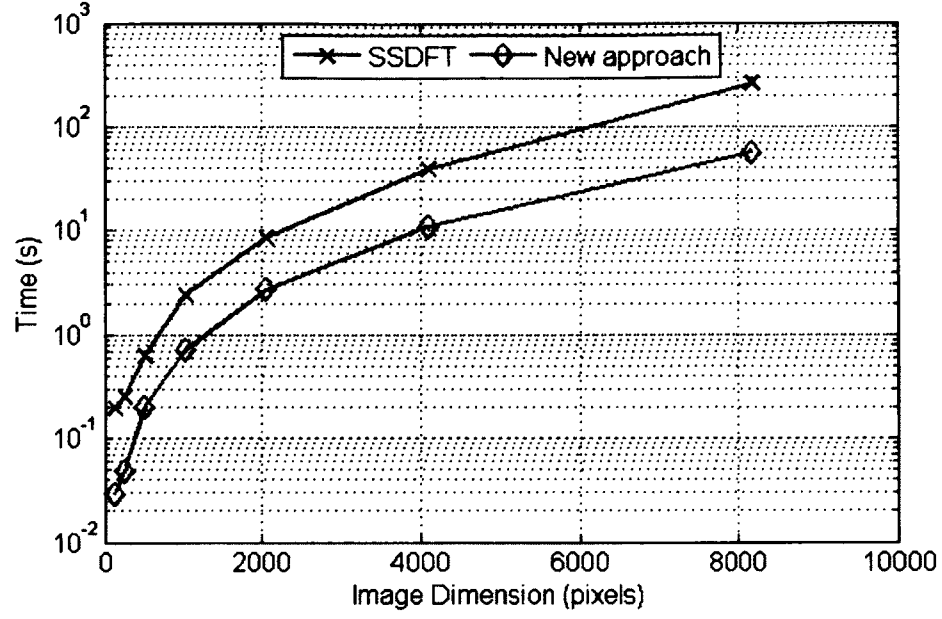


FIG. 20: Total time required for registration

can register very large dimension images of size 8192×8192 in roughly 1 minute compared to 5 minutes using SSDFT approach with subpixel accuracy of 0.01 pixel. Other subpixel registration techniques such as cross correlation surface fitting [71] or stochastic sampling approaches [74] can perform faster but their registration accuracy is not as efficient.

CHAPTER 3

LR IMAGES PROJECTION ONTO HR GRID

3.1 OPTIMUM SCENELS HIGH RESOLUTION GRID ALLOCATION

Like general SR cases, the estimated subpixel shifts are random and don't follow a regular pattern, causing the spacing between the LR scenels to be nonuniform. One approach called nonuniform interpolation converts the nonuniform spaced samples into uniform ones by interpolating the available pixels and using them to estimate the samples or the pixels located at the coordinates of the uniform HR grid, but the reconstructed image using this technique loses some of the frequency components, which in turn affects its visual quality [3, 39].

If the LR scenels are not located correctly to the most correct points of the uniform HR grid coordinates, then the visual quality of the reconstructed images will be affected greatly, which can be seen in Figure 21 that shows the interlaced LR images without reconstruction and restoration. It can be seen with LR scenels optimum allocation, the left image is much better than the right one.

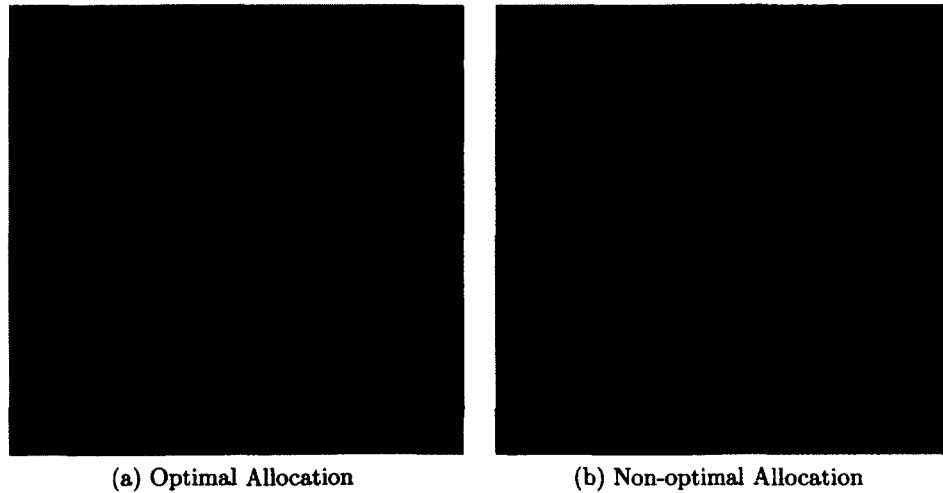


FIG. 21: The effect of optimal and non-optimal allocation on the visual quality

A technique referred to as microscanning is an effective method for reducing aliasing and increasing spatial resolution. The acquisition is performed for each of the microscanned images $s_{k_1 k_2}(x, y)$ for $0 \leq k_1 \leq K_1$ and $0 \leq k_2 \leq K_2$ where K_1 is the number of microscans in the x - direction and K_2 is the number of microscans in the y - direction. It is assumed that the subpixel shifts between scans are equal and constitute a uniform pattern. Thus, for $K_1 \times K_2$ microscanning each scan would differ by a ζ_1/K_1 shift in the x - direction and a ζ_2/K_2 in the y - direction where ζ_1 and ζ_2 are the the photo-detector pitch in the x - and y - directions. Images that have been acquired using a single scan cannot be restored and reconstructed accurately beyond the Nyquist frequency of the acquisition device. Suppose, however, that for a particular input scene, energy is present at frequencies beyond the Nyquist frequency. Without physically adding more detectors to the sampling device, accurate reconstruction of those high frequency features and components could rely on the use of microscanning, which can be used accurately to restore pixel-scale features [75].

Usually, the possible artifacts with the reconstructed image through microscanning are a minimum that results in a reconstructed image with better visual quality. It is desirable to achieve these results in the UAV captured images. The main idea here is to readjust the subpixel shifts of the UAV registered frames that will be combined to form the high resolution image. So, it is required to correctly estimate the subpixel shifts of the frames first and then to optimally allocate them to the HR grid and finally to adjust them to match one of the popular microscanned patterns. Theses modes are 2×2 , 3×3 , and 4×4 scene elements (scenel) per pixel with displacement in the x - and y - directions of $\zeta_1/2$ and $\zeta_2/2$, $\zeta_1/3$ and $\zeta_2/3$, and $\zeta_1/4$ and $\zeta_2/4$, respectively where ζ_1 and ζ_2 are the x and y photo-detector pitches [76].

We propose two approaches that are simple yet effective to optimally allocate the pixels of registered LR frames to the HR grid in order to reduce the visual artifacts in the reconstructed image. The two approaches are: (1) minimum square distance allocation (MSDA) and (2) mid-point subpixel shifts allocation (MSSA). For simplicity, the subpixel shifts in both approaches will be adjusted such that they are located in the first quadrant of \mathbb{R}^2 space. If d_x^{min} and d_y^{min} are the smallest subpixel shifts in the x - and y - directions among all the LR frames, then the adjusted phase shift x'_{k_1} and y'_{k_2} for the $k_1 k_2^{th}$ frame are given by:

$$\begin{aligned} x'_{k_1} &= x_{k_1} + d_x^{min} \\ y'_{k_2} &= y_{k_2} + d_y^{min} \end{aligned} \tag{55}$$

3.1.1 MINIMUM SQUARE DISTANCE ALLOCATION

If the HR grid consists of $M_1 \times M_2$ blocks and every block contains $K_1 \times K_2$ scenels, then the pre-determined uniform HR grid subpixel shifts are $\frac{1}{K_1}$ and $\frac{1}{K_2}$ in the x - and y -directions, respectively, which constitutes a uniform pattern or raster. Unlike the uniform pattern, the subpixel shifts of the registered LR frames form a completely random structure. Consider

$$g = \{(0, 0), (0, \frac{1}{K_2}), \dots, (\frac{K_1-1}{K_1}, \frac{K_2-1}{K_2})\} \quad (56)$$

is the set of the pre-determined subpixel shifts and

$$f = \{(x_0, y_0), (x_0, y_1), \dots, (x_{K_1-1}, y_{K_2-1})\} \quad (57)$$

is the set of estimated subpixel shifts of the registered LR frames. It is required then to find the optimum mapping or transformation $T : f \rightarrow g$ that converts the nonuniform subpixel shifts pattern into a uniform one. The approach here uses the Minkowski distance as a comparison metric to optimally and accurately allocate the elements of the set f to the elements of the set g with minimum distance between the corresponding points in these sets. If $P_{k_1 k_2} = (\frac{k_1}{K_1}, \frac{k_2}{K_2}) \in g$ and $k_i = 0, 1, \dots, K_i - 1$ with $i = 1$ or 2 and $Q_{l_1 l_2} = (x_{l_1}, y_{l_2}) \in f$ and $l_i = 0, 1, \dots, K_i - 1$ with $i = 1$ or 2 then the Minkowski distance of order p between $P_{k_1 k_2}$ and $Q_{l_1 l_2}$ is given by [77]:

$$d(P_{k_1 k_2}, Q_{l_1 l_2}) = \left(\left| \frac{k_1}{K_1} - x_{l_1} \right|^p + \left| \frac{k_2}{K_2} - y_{l_2} \right|^p \right)^{1/p} \quad (58)$$

Typically, the order p is usually set to 1 or 2. To measure the Euclidean distance, we set $p = 2$ and to measure the Manhattan distance, we set $p = 1$. For simplicity and computational complexity purposes we use the Manhattan distance. Assume the HR grid is now empty and we want to fill in the required locations, so we calculate the Manhattan distance between all the points in the set f and only one point in the set g and the point with the minimum distance is set to this location in the HR grid. In other words, for a given point $P_{k_1 k_2} \in g$, the optimum and most close point to it in f is given by:

$$\operatorname{argmin} \{d(P_{k_1 k_2}, Q_{l_1 l_2})\}_{l_i=0}^{l_i=K_i-1} \quad (59)$$

The search for the nearest location should be done in a zigzag scan to avoid misplacing the elements of the set f to the correct locations of the HR grid. If the total number of LR frames is L , then the total number of searches required to allocate all the LR frames to the HR grid is $L!$.

3.1.2 MID-POINT SUBPIXEL SHIFTS ALLOCATION

In the MSDA approach, the allocation of LR frames scenels is constrained by the uniform spacing and structure of the HR grid. So, even with proper adjustment of LR frames phase shift, the reconstructed image will have some undesirable edge artifacts [16]. In the MSSA approach, the scenels will have a relaxed allocation on the HR grid which will be constituted depending on the relative shifts and relative-structure between the different LR scenels, providing a smooth reconstructed image. To construct the HR grid in this approach, we first project the scenels' locations on the x axis and then divide the inter-distance between the projected scenels to obtain the horizontal lines of the HR grid at the mid-points of the x projected scenels' inter-distance. Similarly, the scenels' locations will be projected on the y axis and the vertical lines of the HR grid will be located at the midpoints of the y projected scenels inter-distance. The HR grid will be formed by the intersection between the horizontal and vertical lines and every LR frame scenel is located with respect to its relative location to the other LR frames' scenels.

3.2 PHASE SHIFT ADJUSTMENT

Once the locations of the LR scenels to the HR grid have been determined using either the MSDA or the MSSA methods, the phase shifts of the interlaced LR frames is spatially shifted, using discrete Fourier transform shift theorem, so that the interlaced LR scenels have uniform phase shift differences between them. If $k_1 k_2^{th}$ LR image is $s_{k_1 k_2}(x - x_{k_1}, y - y_{k_2})$ where x_{k_1} and y_{k_2} are the estimated subpixel shifts in the x - and y - directions then its representation in the spatial frequency domain $\tilde{f}_{k_1 k_2}(\nu, \omega)$ is given by [16]:

$$\tilde{f}_{k_1 k_2}(\nu, \omega) = \tilde{s}_{k_1 k_2}(\nu, \omega) e^{-i2\pi(x_{k_1} \nu + y_{k_2} \omega)} \quad (60)$$

Also, if the determined HR grid locations are (d_{k_1}, d_{k_2}) , then the subpixel shift adjustment can be performed in the spatial frequency domain by re-adjusting the phase of the LR image

$\tilde{f}_{k_1 k_2}(\nu, \omega)$ as given by:

$$\tilde{g}_{k_1 k_2}(\nu, \omega) = \tilde{f}_{k_1 k_2}(\nu, \omega) e^{i2\pi((x_{k_1} - d_{k_1})\nu + (y_{k_2} - d_{k_2})\omega)} \quad (61)$$

So, for example, if we need to double the spatial resolution in the x - and y - directions, we need 4 LR frames and the phase shift adjustments can be performed as follows:

$$\tilde{g}_{00}(\nu, \omega) = \tilde{f}_{00}(\nu, \omega) \exp [i2\pi(x_0 \nu + y_0 \omega)] \quad (62)$$

$$\tilde{g}_{01}(\nu, \omega) = \tilde{f}_{01}(\nu, \omega) \exp [i2\pi(x_0 \nu + (y_1 - \frac{1}{2}) \omega)] \quad (63)$$

$$\tilde{g}_{10}(\nu, \omega) = \tilde{f}_{10}(\nu, \omega) \exp [i2\pi((x_1 - \frac{1}{2}) \nu + y_0 \omega)] \quad (64)$$

$$\tilde{g}_{11}(\nu, \omega) = \tilde{f}_{11}(\nu, \omega) \exp [i2\pi((x_1 - \frac{1}{2}) \nu + (y_1 - \frac{1}{2}) \omega)] \quad (65)$$

CHAPTER 4

STOCHASTIC OPTIMAL RECONSTRUCTION AND RESTORATION

In this dissertation, we revisit the Wiener filter and extend it for the general SR problem where the subpixel shifts are unknown and random. The earlier traditional developments of this filter are for a single image restoration and are based on the assumption that the CDC model is constrained only by blurring and noise and ignores the insufficient sampling in the image gathering process. Consequently, it will not actually minimize the mean square error of the reconstructed image. The proposed work is an extension to the Wiener filter developed by Carl et al. [19] that we extend for addressing the SR problem. Also, our derivation indicates periodic and non-periodic frequency inter-relationships between different CDC parameters in addition to highlighting the decomposed output and fidelity components that result from aliasing, blur and noise encountered during the image-acquisition, intermediate processing and image display processes. We formulate the Wiener filter as a function of the parameters associated with the proposed SR system, such as image gathering and image display response indices, system average signal-to-noise ratio (SNR) and inter-subpixel shifts between the LR images. Also, we derive the loss in system fidelity and separate it into three components and relate every component to its corresponding degradation in the proposed system. Simulation and experimental results demonstrate that the derived Wiener filter with the optimal allocation of LR images can reduce aliasing and blurring, resulting in a sharper reconstructed image with pleasant visual quality.

4.1 LOW RESOLUTION IMAGE FORMULATION

Figure 22 details our Super-resolution CDC based system components. It represents most of the degradations including blur, noise and aliasing that are encountered during the image gathering, image reconstruction and image display processes.

- The blur results from convolving the continuous input scene with spatial shift invariant low pass filter that represents the spatial frequency response (SFR) of the image gathering optical lens. In addition, blur also occurs when the image acquisition device is

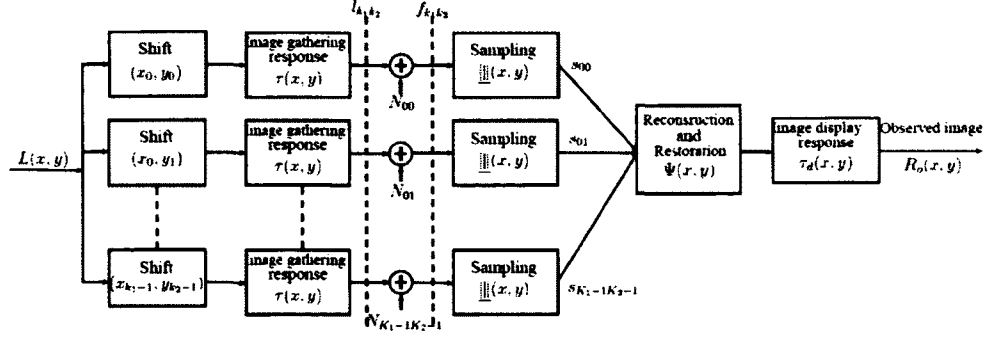


FIG. 22: Complete continuous-discrete-continuous SR reconstruction model

out of focus or there is a relative movement between optical system and the captured scene, and sometimes it is caused by atmospheric turbulence in the case of remote sensing images [29].

- The additive noise results from image gathering photo-detectors or quantization artifacts. Usually, the noise is assumed to be spatially uncorrelated. Sometimes, images and noise are correlated especially when noise is multiplicative instead of additive, or the image gathering is nonlinear. For simplicity, noise is modeled as additive white noise.
- The aliasing results from sampling beyond the Nyquist sampling rate, causing high frequencies in the scene to fold back into the low frequencies band. The aliasing causes certain visual artifacts inside the captured images such as jagged lines, spurious highlights and repeated patterns [29].

According to Carl et al. [19], for the $k_1 k_2^{th}$ LR frame, the image gathering device transforms the continuous input scene $L(x, y)$ into a discrete signal $s_{k_1 k_2}(x, y)$ as defined by:

$$\begin{aligned}
 s_{k_1 k_2}(x, y) &= f_{k_1 k_2}(x, y) \text{|||}(x, y) \\
 &= [L(x - x_{k_1}, y - y_{k_2}) * \tau_{k_1 k_2}(x, y) + N_{k_1 k_2}(x, y)] \text{|||}(x, y)
 \end{aligned} \tag{66}$$

where $\tau(x, y)$ is the spatial response of the image acquisition device, $N_{k_1 k_2}(x, y)$ is the additive

photo-detector noise, and the symbol $*$ denotes spatial convolution. The sampling function

$$\underline{\underline{\text{III}}}(x, y) = \sum_m \sum_n \delta(x - m, y - n) \quad (67)$$

denotes sampling on a rectangular grid with unit sampling intervals and $\delta(x, y)$ is the Dirac delta function. The Fourier transform of Equation (66) gives the spatial frequency representation of the discrete signal $\tilde{s}_{k_1 k_2}$ as defined by:

$$\begin{aligned} \tilde{s}_{k_1 k_2}(\nu, \omega) &= \hat{f}_{k_1 k_2}(\nu, \omega) * \underline{\underline{\text{III}}}(\nu, \omega) \\ &= \sum_m \sum_n \left[\hat{L}(\nu - m, \omega - n) \hat{\tau}_{k_1 k_2}(\nu - m, \omega - n) \exp(-i2\pi(x_{k_1}(\nu - m) \right. \\ &\quad \left. + y_{k_2}(\omega - n))) + \hat{N}_{k_1 k_2}(\nu - m, \omega - n) \right] \end{aligned} \quad (68)$$

where $\hat{L}(\nu, \omega)$ and $\hat{N}_{k_1 k_2}(\nu, \omega)$ are Fourier transforms of the input scene and the photo-detector noise respectively, $\hat{\tau}_{k_1 k_2}(\nu, \omega)$ is the Spatial Frequency Response (SFR) of the image acquisition device, the function

$$\begin{aligned} \underline{\underline{\text{III}}}(\nu, \omega) &= \sum_m \sum_n \delta(\nu - m, \omega - n) \\ &= \delta(\nu, \omega) + \underline{\underline{\text{III}}}_s(\nu - m, \omega - n) \end{aligned} \quad (69)$$

is the Fourier transform of the sampling function and $\underline{\underline{\text{III}}}_s(\nu, \omega)$ accounts for the sampling sidebands. The associated sampling band is defined as:

$$\hat{B} = [(\nu, \omega); |\nu| \leq \frac{1}{2}, |\omega| \leq \frac{1}{2}] \quad (70)$$

The symbol tilde “ \sim ” is used instead of the symbol caret “ $\hat{\cdot}$ ” whenever the corresponding Fourier transformed function is periodic in the spatial frequency domain. Equation (68) can be rewritten as

$$\tilde{s}_{k_1 k_2}(\nu, \omega) = \hat{L}(\nu, \omega) \hat{\tau}_{k_1 k_2}(\nu, \omega) \exp(-i2\pi(x_{k_1} \nu + y_{k_2} \omega)) + \hat{N}_a(\nu, \omega) + \tilde{N}_{k_1 k_2}(\nu, \omega) \quad (71)$$

where

$$\begin{aligned}\hat{N}_a(\nu, \omega) &= \sum_{m \neq 0} \sum_{n \neq 0} \hat{L}(\nu - m, \omega - n) \hat{\tau}_{k_1 k_2}(\nu - m, \omega - n) \\ &\times \exp(-i2\pi(x_{k_1}(\nu - m) + y_{k_2}(\omega - n)))\end{aligned}\quad (72)$$

are the aliased components that insufficient sampling folds back into the sampling passband and $\tilde{N}_{k_1 k_2}(\nu, \omega)$ is Fourier transform of the photo-detector noise. The Fourier components of the wide-sense stationary random fields $\hat{L}(\nu, \omega)$ and $\hat{N}_{k_1 k_2}(\nu, \omega)$ and the co-aliased components of the sampled scene are uncorrelated. These uncorrelated interrelationships can be expressed as

$$E\left\{\hat{L}(\nu - m, \omega - n)\hat{L}^*(\nu - m', \omega - n')\right\} = \hat{\Phi}_L(\nu - m, \omega - n)\delta(m - m', n - n'), \quad (73)$$

$$E\left\{\hat{N}_{k_1 k_2}(\nu - m, \omega - n)\hat{N}_{k_1 k_2}^*(\nu - m', \omega - n')\right\} = \hat{\Phi}_{N_{k_1 k_2}}(\nu - m, \omega - n)\delta(m - m', n - n'), \quad (74)$$

$$E\left\{\hat{N}_{k_1 k_2}(\nu - m, \omega - n)\hat{N}_{i_1 i_2}^*(\nu - m', \omega - n')\right\} = 0, \quad (75)$$

$$E\left\{\hat{L}(\nu - m, \omega - n)\hat{N}_{k_1 k_2}^*(\nu - m', \omega - n')\right\} = 0 \quad (76)$$

The power spectral density (PSD) of the acquired digital image $s_{k_1 k_2}(x, y)$ is defined by:

$$\tilde{\Phi}_{s_{k_1 k_2}}(\nu, \omega) = E\{|\tilde{s}_{k_1 k_2}(\nu, \omega)|^2\} \quad (77)$$

with corresponding variance

$$\sigma_{k_1 k_2}^2 = \iint \tilde{\Phi}_{s_{k_1 k_2}}(\nu, \omega) d\nu d\omega \quad (78)$$

Using Equation (73) through Equation (76), the PSD of the degraded image can be expressed as

$$\tilde{\Phi}_{s_{k_1 k_2}}(\nu, \omega) = \left[\hat{\Phi}_L(\nu, \omega) |\hat{\tau}(\nu, \omega)|^2 + \hat{\Phi}_{N_{k_1 k_2}}(\nu, \omega) \right] * \hat{\Pi}(\nu, \omega) \quad (79)$$

where $\hat{\Phi}_L$ is the PSD of the input scene and is defined by

$$\hat{\Phi}_L(\nu, \omega) = E\left\{\left|\hat{L}(\nu, \omega)\right|^2\right\} \quad (80)$$

with associated variance given by

$$\sigma_L^2 = \iint \hat{\Phi}_L(\nu, \omega) d\nu d\omega \quad (81)$$

and $\hat{\Phi}_{N_{k_1 k_2}}$ is the PSD of the noise associated with the $k_1 k_2^{th}$ LR frame and it is defined by

$$\tilde{\Phi}_{N_{k_1 k_2}}(\nu, \omega) = E\{|\tilde{N}_{k_1 k_2}(\nu, \omega)|^2\} \quad (82)$$

with corresponding variance

$$\sigma_{k_1 k_2}^2 = \iint \tilde{\Phi}_{N_{k_1 k_2}}(\nu, \omega) d\nu d\omega. \quad (83)$$

4.2 HR GRID INTERLACING

After the phase shifts of the individual LR images have been corrected, the reconstruction of the output image is performed by interlacing the pixels of the acquired images into a HR grid with a sampling density equal to $K_1 K_2$ times the sampling density of the individual LR images. Thus, the composite HR image S is given by:

$$S(K_1 m_1 + k_1, K_2 m_2 + k_2) = s_{k_1 k_2}(m_1, m_2), \quad (84)$$

and its Fourier transform is given by:

$$\tilde{S}(\nu, \omega) = \frac{1}{K_1 M_1 K_2 M_2} \sum_{m_1=0}^{K_1 M_1 - 1} \sum_{m_2=0}^{K_2 M_2 - 1} S(m_1, m_2) \exp(-i2\pi (\frac{\nu m_1}{K_1 M_1} + \frac{\omega m_2}{K_2 M_2})) \quad (85)$$

The above equation can be rewritten and simplified as

$$\begin{aligned} \tilde{S}(\nu, \omega) &= \frac{1}{K_1 K_2} \sum_{k_1=0}^{K_1-1} \sum_{k_2=0}^{K_2-1} \frac{1}{M_1 M_2} \sum_{m_1=0}^{M_1-1} \sum_{m_2=0}^{M_2-1} S(K_1 m_1 + k_1, K_2 m_2 + k_2) \\ &\quad \exp\left(-i2\pi \nu \left(\frac{K_1 m_1 + k_1}{K_1 M_1}\right)\right) \exp\left(-i2\pi \omega \left(\frac{K_2 m_2 + k_2}{K_2 M_2}\right)\right) \end{aligned} \quad (86)$$

Rearranging the summation terms

$$\begin{aligned} \tilde{S}(\nu, \omega) &= \frac{1}{K_1 K_2} \sum_{k_1=0}^{K_1-1} \sum_{k_2=0}^{K_2-1} \left[\frac{1}{M_1 M_2} \sum_{m_1=0}^{M_1-1} \sum_{m_2=0}^{M_2-1} s_{k_1 k_2}(m_1, m_2) \right. \\ &\quad \left. \exp\left(-i2\pi \left(\frac{\nu m_1}{M_1} + \frac{\omega m_2}{M_2}\right)\right) \right] \exp\left(-i2\pi \left(\frac{\nu k_1}{K_1 M_1} + \frac{\omega k_2}{K_2 M_2}\right)\right) \end{aligned} \quad (87)$$

Recall that the discrete Fourier transform of the degraded image $\tilde{s}_{k_1 k_2}$ is given by

$$\tilde{s}_{k_1 k_2}(\nu, \omega) = \frac{1}{M_1 M_2} \sum_{m_1=0}^{M_1-1} \sum_{m_2=0}^{M_2-1} s_{k_1 k_2}(m_1, m_2) \exp \left(-i2\pi \left(\frac{\nu m_1}{M_1} + \frac{\omega m_2}{M_2} \right) \right) \quad (88)$$

So, by substituting Equation (88) into Equation (87), the Fourier transform of the reconstructed image $\tilde{S}(\nu, \omega)$ can be expressed as

$$\tilde{S}(\nu, \omega) = \frac{1}{K_1 K_2} \sum_{k_1=0}^{K_1-1} \sum_{k_2=0}^{K_2-1} \tilde{s}_{k_1 k_2}(\nu, \omega) \exp \left(-i2\pi \left(\frac{\nu k_1}{M_1 K_1} + \frac{\omega k_2}{M_2 K_2} \right) \right) \quad (89)$$

If x_{k_1} and y_{k_2} are the subpixel shifts in the x - and y - directions, respectively, and (ν', ω') is the normalized frequency pair, then the last equation can be written as

$$\tilde{S}(\nu', \omega') = \frac{1}{K_1 K_2} \sum_{k_1=0}^{K_1-1} \sum_{k_2=0}^{K_2-1} \tilde{s}_{k_1 k_2}(\nu', \omega') \exp (-i2\pi(\nu' x_{k_1} + \omega' y_{k_2})) \quad (90)$$

Thus, the reconstructed image in the frequency domain is the sum of the phase shifted Fourier transform of the individual scans [75].

4.3 DERIVED STOCHASTIC WIENER RESTORATION FILTER

In this section, we derive the stochastic Wiener filter based on the fully detailed SR CDC model that is given in Figure 22. It is constrained by the periodic and non-periodic inter-relationships between the different frequency components of the proposed SR system. It can be used as reconstruction filter as well as a restoration filter to recover images from the degradations that are introduced during image acquisition and image display. If the number of LR frames is $K_1 K_2$, which is sufficient to produce full SR along the horizontal and vertical dimensions of the reconstructed images, then the derived filter works only as a restoration filter. On the other hand, if the available number of LR images is less than $K_1 K_2$, then the filter works as a restoration and a reconstruction filter to estimate missing pixels on the dense HR grid. In spatial domain, the observed image $R_o(x, y)$ is reconstructed through the spatial convolution of the Wiener filter $\Psi(x, y)$, the spatial response of the image display device $\tau_d(x, y)$ and the interlaced image $S(x, y)$ as given by

$$R_o(x, y) = S(x, y) * \Psi(x, y) * \tau_d(x, y) \quad (91)$$

The spatial frequency representation of this equation is given by

$$\hat{R}_o(\nu, \omega) = \tilde{S}(\nu, \omega) \hat{\Psi}(\nu, \omega) \hat{\tau}_d(\nu, \omega) \quad (92)$$

By substituting Equation (71) and Equation (90) into Equation (92), the output image \hat{R}_o can be decomposed into three components as defined by

$$\hat{R}_o(\nu, \omega) = \hat{R}_f(\nu, \omega) + \hat{R}_a(\nu, \omega) + \hat{R}_n(\nu, \omega) \quad (93)$$

where \hat{R}_f is the filtered component that accounts for the low-pass filtering of the image gathering, image display devices and the restoration filter, and it is given by

$$\hat{R}_f(\nu, \omega) = \frac{1}{K_1 K_2} \sum_{k_1 k_2} \hat{L}(\nu, \omega) \hat{\tau}_{k_1 k_2}(\nu, \omega) \exp(-i4\pi(\nu x_{k_1} + \omega y_{k_2})) \hat{\tau}_d(\nu, \omega) \hat{\Psi}(\nu, \omega), \quad (94)$$

$\hat{R}_n(\nu, \omega)$ is the noise component that accounts for the additive white noise and is given by

$$\hat{R}_n(\nu, \omega) = \frac{1}{K_1 K_2} \sum_{k_1 k_2} \hat{N}_{k_1 k_2}(\nu, \omega) \exp(-i4\pi(\nu x_{k_1} + \omega y_{k_2})) \hat{\tau}_d(\nu, \omega) \hat{\Psi}(\nu, \omega), \quad (95)$$

and $\hat{R}_a(\nu, \omega)$ is the aliasing component that accounts for the frequency folding due to sampling beyond the Nyquist rate, and it is given by

$$\hat{R}_a(\nu, \omega) = \frac{1}{K_1 K_2} \sum_{k_1 k_2} \hat{N}_a(\nu, \omega) \exp(-i4\pi(\nu x_{k_1} + \omega y_{k_2})) \hat{\tau}_d(\nu, \omega) \hat{\Psi}(\nu, \omega), \quad (96)$$

where $\hat{N}_a(\nu, \omega)$ is defined in Equation (72). The Wiener filter minimizes the mean-square restoration error (MSRE) e^2 between the input scene $L(x, y)$ and the output image $R_o(x, y)$, as defined by [19]

$$\begin{aligned} e^2 &= E \left\{ \iint |L(x, y) - R_o(x, y)|^2 dx dy \right\} \\ &= E \left\{ \iint \left| \hat{L}(\nu, \omega) - \tilde{S}(\nu, \omega) \hat{\Psi}(\nu, \omega) \hat{\tau}_d(\nu, \omega) \right|^2 d\nu d\omega \right\} \end{aligned} \quad (97)$$

Using Equation (73) through Equation (76), the MSRE can be expressed as

$$\iint [\hat{\Phi}_L(\nu, \omega) - \hat{\Phi}_{LS^*}(\nu, \omega) \hat{\Psi}^*(\nu, \omega) \hat{\tau}_d^*(\nu, \omega) - \hat{\Phi}_{L^*S}(\nu, \omega) \Psi(\nu, \omega) \tau_d(\nu, \omega) + \tilde{\Phi}_S(\nu, \omega) |\hat{\Psi}(\nu, \omega)|^2 |\hat{\tau}_d(\nu, \omega)|^2] d\nu d\omega \quad (98)$$

where

$$\hat{\Phi}_{LS^*}(\nu, \omega) = (\hat{\Phi}_{L^*S}(\nu, \omega))^* = E\{\hat{L}(\nu, \omega) \tilde{S}^*(\nu, \omega)\} \quad (99)$$

is the cross power spectrum between the input radiance field and the reconstructed image.

The optimal CDC Wiener restoration filter must satisfy [12]:

$$\begin{aligned} \frac{\partial e^2}{\partial \hat{\Psi}} &= -\hat{\Phi}_{L^*S}(\nu, \omega) \hat{\tau}_d(\nu, \omega) + \tilde{\Phi}_S(\nu, \omega) \hat{\Psi}^*(\nu, \omega) |\hat{\tau}_d(\nu, \omega)|^2 \\ &= 0 \end{aligned} \quad (100)$$

Accordingly, the Wiener filter that minimizes the MSRE is given by:

$$\hat{\Psi}(\nu, \omega) = \frac{\hat{\Phi}_{LS^*}(\nu, \omega) \hat{\tau}_d^*(\nu, \omega)}{\tilde{\Phi}_S(\nu, \omega) |\hat{\tau}_d(\nu, \omega)|^2} \quad (101)$$

The cross power spectrum $\hat{\Phi}_{LS^*}(\nu, \omega)$ between the input scene and the reconstructed image is given by:

$$\hat{\Phi}_{LS^*}(\nu, \omega) = \frac{1}{K_1 K_2} \hat{\Phi}_L(\nu, \omega) \sum_{k_1 k_2} \hat{\tau}_{k_1 k_2}^*(\nu, \omega) \exp(i4\pi(\nu x_{k_1} + \omega y_{k_2})) \quad (102)$$

The power spectrum density of the reconstructed image $\tilde{\Phi}_S(\nu, \omega)$ is given by:

$$\begin{aligned} \tilde{\Phi}_S(\nu, \omega) &= \frac{1}{K_1^2 K_2^2} \left(\hat{\Phi}_L(\nu, \omega) * \hat{\mathbb{I}}(\nu, \omega) \sum_{mn} \left| \sum_{k_1 k_2} |\hat{\tau}_{k_1 k_2}^*(\nu', \omega')|^2 \exp(-i2\pi(\nu' x_{k_1} + \omega' y_{k_2})) \right|^2 \right. \\ &\quad \left. + \sum_{k_1 k_2} \hat{\Phi}_{N_{k_1 k_2}}(\nu, \omega) * \hat{\mathbb{I}}(\nu, \omega) \right) \end{aligned} \quad (103)$$

where $\nu' = 2\nu - m$ and $\omega' = 2\omega - n$. If the photo-detector noise is modeled as wide-sense

of the reconstructed image $\tilde{\Phi}_S(\nu, \omega)$ are:

$$\begin{aligned} \tilde{S}(\nu', \omega') &= \frac{1}{K_1 K_2} \sum_{k_1=0}^{K_1-1} \sum_{k_2=0}^{K_2-1} \tilde{s}_{k_1 k_2}(\nu', \omega') \\ &\times \exp(-i2\pi(\nu'(x_{k_1} + \alpha_{k_1}) + \omega'(y_{k_2} + \beta_{k_2}))), \end{aligned} \quad (107)$$

$$\begin{aligned} \hat{\Phi}_{LS^*}(\nu, \omega) &= \frac{1}{K_1 K_2} \hat{\Phi}_L(\nu, \omega) \sum_{k_1 k_2} \hat{\tau}_{k_1 k_2}^*(\nu, \omega) \exp(i4\pi(\nu x_{k_1} + \omega y_{k_2})) \\ &\times E \{ \exp i2\pi(\nu(\alpha_{k_1} + \omega(\beta_{k_2}))) \} \end{aligned} \quad (108)$$

and

$$\begin{aligned} \tilde{\Phi}_S(\nu, \omega) &= \frac{1}{K_1^2 K_2^2} \left(\hat{\Phi}_L(\nu, \omega) * \hat{\mathbb{I}}(\nu, \omega) \sum_{mn} \left| \sum_{k_1 k_2} \hat{\tau}_{k_1 k_2}^*(\nu', \omega') \exp(-i2\pi(\nu' x_{k_1} + \omega' y_{k_2})) \right. \right. \\ &\times \left. \left. E \{ \exp i2\pi(\nu \alpha_{k_1} + \omega \beta_{k_2}) \} \right|^2 + \sum_{k_1 k_2} \hat{\Phi}_{N_{k_1 k_2}}(\nu, \omega) * \hat{\mathbb{I}}(\nu, \omega) \right) \end{aligned} \quad (109)$$

respectively, where α_{k_1} and β_{k_2} are the corresponding registration and fusion errors. Far-siu et al. [79] suggested that the registration and fusion errors are properly described by Laplacian distribution. We derived the $E \{ \exp i2\pi(\nu(\alpha_{k_1} + \omega(\beta_{k_2}))) \}$ term for three different distributions (Laplacian, uniform, and Gaussian) and compare their performance in terms of the maximum realizable fidelity. If the registration and fusion errors are described by a Laplacian probability density function $f_L(\alpha_{k_1}, \beta_{k_2}|b)$ given by

$$f_L(\alpha_{k_1}, \beta_{k_2}) = \frac{1}{2b} e^{-\frac{|\alpha_{k_1}| + |\beta_{k_2}|}{b}} \quad (110)$$

where b is a scale parameter and related to the variance of the error, then the expectation for the error term is given by

$$\begin{aligned} E \{ e^{i2\pi(\nu \alpha_{k_1} + \omega \beta_{k_2})} \} &= \int_{-\infty}^{\infty} e^{i2\pi(\nu \alpha_{k_1} + \omega \beta_{k_2})} f_L(\alpha_{k_1}, \beta_{k_2}) d\alpha_{k_1} d\beta_{k_2} \\ &= \frac{1}{b^2(4\pi^2 b^2 \nu^2 + 1)(4\pi^2 b^2 \omega^2 + 1)} \end{aligned} \quad (111)$$

The proposed second distribution is the Gaussian probability function $f_G(\alpha_{k_1}, \beta_{k_2})$ defined

by

$$f_G(\alpha_{k_1}, \beta_{k_2}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\alpha_{k_1}^2 + \beta_{k_2}^2}{2\sigma^2}} \quad (112)$$

Accordingly, the expectation for the error term is given by

$$\begin{aligned} E \{ e^{i2\pi(\nu\alpha_{k_1} + \omega\beta_{k_2})} \} &= \int_{-\infty}^{\infty} e^{i2\pi(\nu\alpha_{k_1} + \omega\beta_{k_2})} f_G(\alpha_{k_1}, \beta_{k_2}) d\alpha_{k_1} d\beta_{k_2} \\ &= e^{2\pi^2\sigma^2(\nu^2 + \omega^2)} \end{aligned} \quad (113)$$

Likewise if the distribution is the uniform probability distribution function $f_U(\alpha_{k_1}, \beta_{k_2})$ defined by

$$f_U(\alpha_{k_1}, \beta_{k_2}) = \begin{cases} W_x W_y & \text{if } |\alpha_{k_1}| \leq \frac{1}{2W_x}, |\beta_{k_2}| \leq \frac{1}{2W_y} \\ 0 & \text{otherwise} \end{cases} \quad (114)$$

then the corresponding expectation for the error is given by

$$\begin{aligned} E \{ e^{i2\pi(\nu\alpha_{k_1} + \omega\beta_{k_2})} \} &= \int_{-\infty}^{\infty} e^{i2\pi(\nu\alpha_{k_1} + \omega\beta_{k_2})} f_U(\alpha_{k_1}, \beta_{k_2}) d\alpha_{k_1} d\beta_{k_2} \\ &= \text{sinc}\left(\frac{\nu}{W_x}\right) \text{sinc}\left(\frac{\omega}{W_y}\right) \end{aligned} \quad (115)$$

4.5 SPEEDED-UP WIENER FILTER

The main drawback of the proposed SR reconstruction and restoration Wiener filter is its computational time required to prepare the Wiener filter in the frequency domain using Equation (105). The idea here is to reduce the dimensions of the Wiener filter by constraining it to work on small patches on the images. As a result the required time to prepare the filter is reduced in addition to the time required to apply fast Fourier transform and its inverse operations. Back to Equation (91), the interlaced image $S(x, y)$ can be expanded as

$$S(x, y) = \begin{bmatrix} P_{11} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} + \begin{bmatrix} 0 & P_{12} & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} + \dots + \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & P_{p_1 p_2} \end{bmatrix} \quad (116)$$

where P_{ij} represents the ij^{th} partition, 0 is a block zero matrix of dimension $L_1 \times L_2$ and p_1 and p_2 are the number of partitions in the x - and y - directions. respectively. Suppose that $S(x, y)$ and $P_{ij}(x, y)$ are of dimension $M_1 \times M_2$ and $L_1 \times L_2$, respectively; then P_{ij} can

be written as

$$P_{ij} = S(iL_1 : (i+1)L_1 - 1, jL_2 : (j+1)L_2 - 1) \quad (117)$$

As a result of this partitioning, Wiener filter can be applied independently to each of these partitions and because they are dominated by zeros in a well organized structure it can be applied efficiently to its nonzero part. The size of the derived Wiener filter can be constrained to the size of the P_{ij} matrix, and as a consequence, its performance is expected to improve. The main parameters that control the derivation of the Wiener filter are the subpixel shifts of LR images and the mean spatial detail of the acquired scene. After the subpixel shift adjustments, these subpixel shifts constitute a uniform structure and are the same for all partitions. Also, as a good approximation of the mean spatial detail of the scene can be taken as the average spatial detail across all the partitions. As a consequence, only one Wiener filter can be used efficiently in the reconstruction and restoration of all partitions, which enables parallelization of the process and reduces the total time required for the processing of the whole image. Figure 23 shows the pipeline for the implementation of the speeded-up Wiener filter. It can be prepared in the frequency domain with the new smaller dimension and then it is pointwise multiplied with Fourier transform of every partition. Every partition will be recovered independently by means of inverse Fourier transform and finally all the processed partitions can be combined again to construct the final reconstructed SR image. If $M_1 = M_2 = M$ and $L_1 = L_2 = L$, then the new computational complexity of the processing of the Fourier transform operations is $L \log_2 L$ compared to $M \log_2 M$ for the older one.

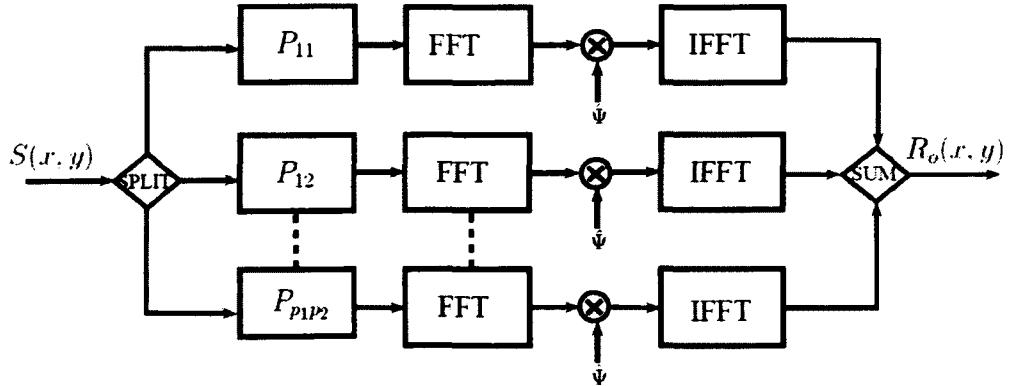


FIG. 23: Pipeline of the speeded up Wiener filter

4.6 WIENER CHARACTERISTIC RESTORATION FILTER

It is sometimes desirable to enhance a specific spatial feature of the input scene which can be formulated by letting $L_c(x, y) = L(x, y) * \tau_c(x, y)$ be the desired representation of the input scene $L(x, y)$, where $\tau_c(x, y)$ is a linear characteristic function. The corresponding MSRE e_c^2 is defined by [27]

$$e_c^2 = E \left\{ \iint \left| \hat{L}_c(\nu, \omega) - \hat{R}_o(\nu, \omega) \right|^2 d\nu d\omega \right\} \quad (118)$$

Thus, the Wiener characteristic filter becomes

$$\hat{\Psi}(\nu, \omega) = \frac{K_1 K_2 \hat{\Phi}_L(\nu, \omega) \hat{\tau}_c^*(\nu, \omega) \hat{\tau}_c^*(\nu, \omega) \hat{\tau}_d^*(\nu, \omega) \exp(i4\pi(\nu x_{k_1} + \omega y_{k_2})) / |\hat{\tau}_d(\nu, \omega)|^2}{\hat{\Phi}_L(\nu, \omega) * \iint(\nu, \omega) \sum_{mn} \left| \hat{\tau}_{k_1 k_2}^*(\nu', \omega') \right|^2 \exp(-i2\pi(\nu' x_{k_1} + \omega' y_{k_2})) \Big|^2 + \sum_{k_1 k_2} (\sigma_L / \sigma_{N_{k_1 k_2}})^{-2}} \quad (119)$$

For example, the characteristic filter $\hat{\tau}_c(\nu, \omega)$ can be the spatial frequency response $\hat{\tau}_e(\nu, \omega)$ of the Laplacian of Gaussian ($\nabla^2 G$) operator commonly used to enhance the input scene transitions for subsequent edge detections [27]. The spatial response and spatial frequency response of this operator, respectively, are

$$\tau_e(x, y) = \frac{1}{\pi \sigma_e^4} \left(1 - \frac{r^2}{2\sigma_e^2} \right) \exp \left[-\frac{r^2}{2\sigma_e^2} \right] \quad (120)$$

and

$$\hat{\tau}_e(\nu, \omega) = (2\pi \rho)^2 \exp [-2(\pi \sigma_e \rho)^2] \quad (121)$$

where $r^2 = x^2 + y^2$, σ_e is the standard deviation of the Gaussian function, and $\rho = \nu^2 + \omega^2$. The selection of the Gaussian standard deviation σ_e normally entails a compromise between high resolution (with a small operator) and suppression of artifacts (with large operator). We choose $\sigma_e = 0.75$ because the spatial response $\tau_e(x, y)$ relative to the unity sampling interval is then the same as that of Marr's model of the smallest operator in early human vision relative to the mean center-to-center distance between photodetectors in the foveal region of the eye's retina [27].

4.7 WIENER GAUSSIAN ENHANCEMENT FILTER

The images that the Wiener filter restores with minimum MSRE normally possess high resolution. However, these images also exhibit visually annoying defects and artifacts due to

aliasing, photodetector noise that is amplified by the high frequency enhancement inherent in the Wiener restoration, and the ringing near sharp edges (Gibbs phenomenon) caused by the steep roll-off in the throughput response of image gathering restoration. So, it is desirable to combine this filter with an interactive control function for enhancing the visual quality. This control is provided by the Wiener- Gaussian enhancement filter [27] as defined by

$$\hat{\Psi}_v(\nu, \omega) = \hat{\Psi}(\nu, \omega) \hat{\tau}_v(\nu, \omega) \quad (122)$$

where

$$\hat{\tau}_v(\nu, \omega) = \exp[-2(\pi\sigma_i\rho)^2] + \zeta \left(\frac{2\pi\sigma_e\rho}{\sigma_o} \right)^2 \exp[-2(\pi\sigma_e\rho)^2] \quad (123)$$

where $\sigma_o = 0.7$. The standard deviation σ_i controls the roll-off of the Wiener filter, while the standard deviation σ_e and the enhancement factor ζ controls the edge enhancement. The ratio σ_e^2/σ_o^2 is included so that the enhancement with $\zeta = 1$ is directly proportional to the change of the intensity at an edge transitions [80]. The resultant WIGE restoration substantially reduces the ringing of the Wiener restoration at the cost of a barely perceptible loss in sharpness.

4.8 FIDELITY ANALYSIS

We analyze and decompose the end-to-end system fidelity and relate every component to its corresponding degradation in the CDC system. Linefoot [81] computed the image fidelity between Fourier transforms of both the continuous input scene and the continuous output scene as

$$\mathcal{F} = 1 - \frac{e^2(\nu, \omega)}{\iint \hat{\Phi}_L d\nu d\omega} \quad (124)$$

By using Eqs. (98), (102) and (103), the fidelity can be expressed as

$$\mathcal{F} = 1 - \mathcal{F}_f - \mathcal{F}_a - \mathcal{F}_n \quad (125)$$

where \mathcal{F}_f is the loss due to the filtration process and is given by

$$\mathcal{F}_f = \frac{\iint \hat{\Phi}'_L(\nu, \omega) \left| 1 - \frac{1}{K_1 K_2} \hat{\tau}_d(\nu, \omega) \hat{\Psi}(\nu, \omega) \sum_{k_1 k_2} \hat{\tau}_{k_1 k_2}(\nu, \omega) \exp(-i4\pi(\nu x_{k_1} + \omega y_{k_2})) \right|^2 d\nu d\omega}{\iint \hat{\Phi}'_L d\nu d\omega}, \quad (126)$$

\mathcal{F}_n is the loss in the fidelity due to noise and is given by

$$\mathcal{F}_n = \frac{\iint \frac{1}{K_1^2 K_2^2} \sum_{k_1 k_2} (\sigma_L / \sigma_{N_{k_1 k_2}})^{-2} \left| \hat{\Psi}(\nu, \omega) \hat{\tau}_d(\nu, \omega) \right|^2 d\nu d\omega}{\iint \hat{\Phi}'_L d\nu d\omega}, \quad (127)$$

and \mathcal{F}_a is the loss in the fidelity due to aliasing and is given by

$$\mathcal{F}_a = \frac{\iint \frac{1}{K_1^2 K_2^2} \sum_{m, n \neq 0} \hat{\Phi}'_L(\nu', \omega') \left| \hat{\tau}_d(\nu', \omega') \hat{\Psi}(\nu', \omega') \sum_{k_1 k_2} \hat{\tau}_{k_1 k_2}(\nu, \omega) \exp(-i2\pi(\nu'' x_{k_1} + \omega'' y_{k_2})) \right|^2 d\nu d\omega}{\iint \hat{\Phi}'_L d\nu d\omega} \quad (128)$$

where $\nu'' = 2\nu - m$, $\omega'' = 2\omega - n$, $\nu' = \nu - m$ and $\omega' = \omega - n$.

To minimize the loss in the fidelity due to filtration process it is necessary that we select

$$\frac{1}{K_1 K_2} \hat{\tau}(\nu, \omega) \hat{\tau}_d(\nu, \omega) \hat{\Psi}(\nu, \omega) \sum_{k_1 k_2} \exp(-i4\pi(\nu x_{k_1} + \omega y_{k_2}))$$

closer to 1 at the frequencies where $\hat{\Phi}'_L(\nu, \omega)$ is large. Also, to reduce the loss in the fidelity due to aliasing, the term

$$\left| \hat{\tau}(\nu', \omega') \hat{\tau}_d(\nu', \omega') \hat{\Psi}(\nu', \omega') \sum_{k_1 k_2} \exp(-i2\pi(\nu'' x_{k_1} + \omega'' y_{k_2})) \right|$$

should be approaching zero at those frequencies where $\hat{\Phi}'_L(\nu', \omega')$ is large. Finally, the loss in the fidelity due to the noise can be reduced by increasing the SNR of the imaging system.

4.9 SIMULATIONS AND RESULTS

To assess and evaluate the proposed approach, we used two sets of images. The first set is the random polygons image shown in Figure 24 to simulate and analyze the Weiner restoration filter results. The other set is the checkerboard images and is used to simulate the UAV captured images. The random polygon consists of regions whose boundaries are distributed according to Poisson probability with a mean separation μ and whose input scene magnitudes are distributed according to independent zero-mean Gaussian statistics of variance σ_L^2 . The mean separation μ is measured relative to the sampling interval of the image-gathering device and treated as the mean spatial detail of the scene [27]. The

checkerboard images contain the same scene but with different views. By controlling the camera orientation, the different images will contain the distortions that are expected when the UAV experiences yaw, pitch and roll. Figure 8 shows the sets of checkerboard images that have been used in simulations.

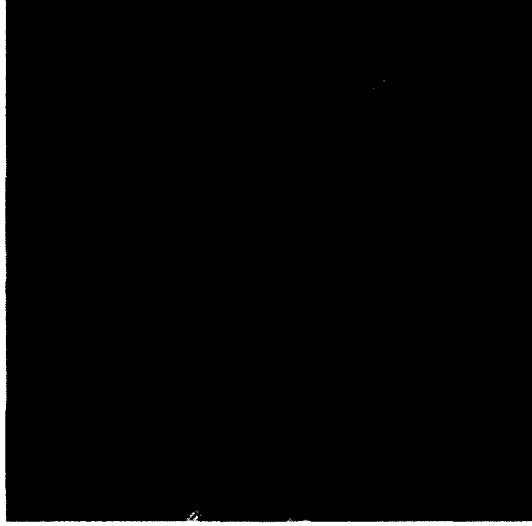


FIG. 24: Random polygon image

4.9.1 WIENER FILTER ASSESSMENT

In our simulations, we start with a simulated high resolution scene that is blurred by a Gaussian low-pass filter defined as:

$$\hat{\tau}(\nu, \omega) = \exp \left[-\frac{\nu^2 + \omega^2}{\sigma^2} \right] \quad (129)$$

where σ is the optical-response index for which $\hat{\tau}(\nu, \omega) \approx 0.37$. This Gaussian filter approximates the SFR of the image-gathering device. Schade[82] and Schreiber[83] concluded that the image-gathering device with SFR $\hat{\tau}(\nu, \omega)$ characterized by $\sigma = 0.8$ provides generally the most favorable trade-off between sharpness and aliasing artifacts without the aid of digital processing. The simulated HR images are down sampled by a factor of 2 and white noise and blurring are superimposed to them such that the blurred signal-to-noise ratio (BSNR) is 30 dB:

$$\text{BSNR} = 10 \log_{10} \left(\frac{\sigma_L^2}{\sigma_N^2} \right) \quad (130)$$

where σ_L^2 is the variance of the blurred image scene and σ_N^2 is the variance of the white noise.

The derivation of the Wiener restoration filter depends on the estimation of the input scene PSD $\hat{\Phi}_L$. Itakura et al. [84] have shown that the PSD of natural scenes can be approximated by:

$$\hat{\Phi}_L(\nu, \omega) = \frac{2\pi\mu^2\sigma_L^2}{[1 + (2\pi\mu\rho)^2]^{3/2}}, \quad (131)$$

where $\rho^2 = \nu^2 + \omega^2$ and μ is the scene mean spatial detail. The HR image is reconstructed and Wiener filter is applied to restore the images from the degradations encountered in the image-gathering process. Figure 25 shows the derived Wiener filters for the random polygon and checkerboard images. The vertical solid line in Figure 25 identifies the amount of the energies that can be restored beyond the Nyquist frequency.

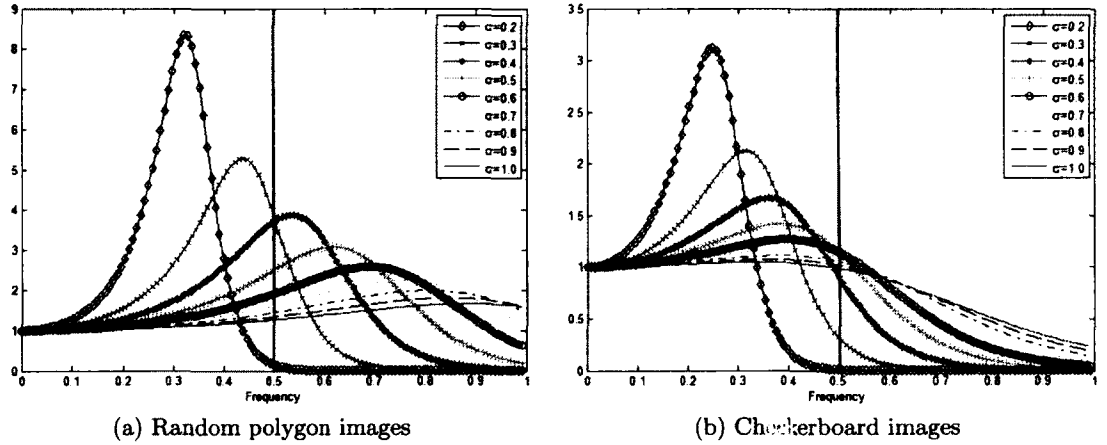


FIG. 25: Derived Wiener restoration filter for different image-gathering optical response index (σ)

As the optical-response index σ of the image-gathering device increases the response of the restoration process becomes better. This is because when σ is small (less than 0.5) the blurring of the image-gathering device is heavy and results in attenuating energies at frequencies in the pass band, and consequently, the restoration can't efficiently restore those energies. The fidelity is used as a comparison metric between the original HR image and the reconstructed one. Figure 26 shows the fidelity between the simulated HR images and the reconstructed ones. The greater the optical-response index σ of the image-gathering device the better the fidelity between the reconstructed and the simulated HR scenes. The reconstructed images for the random polygon and the checkerboard images are shown in Figure 27 and Figure 28, respectively, for different image-gathering optical response index (σ). For the checkerboard images that simulate the UAV captured images, it can be seen

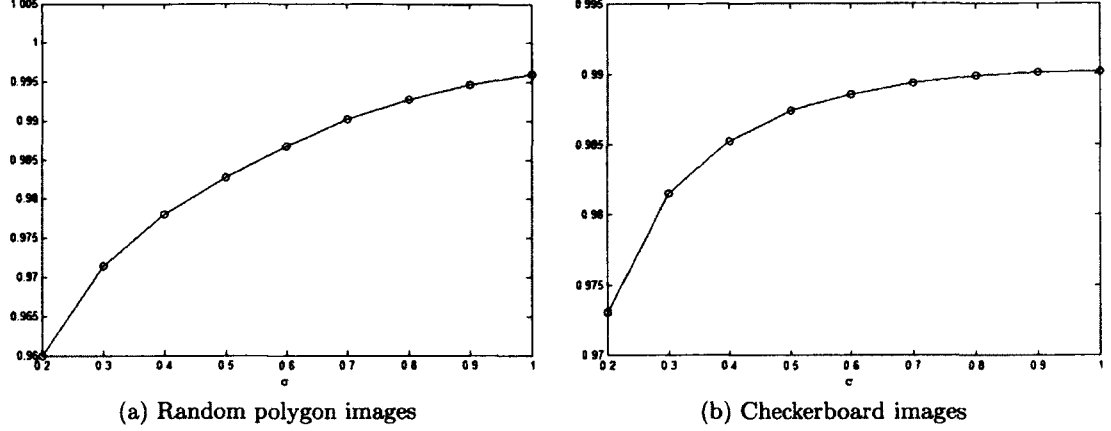


FIG. 26: Fidelity between simulated HR scene and the reconstructed images for different image-gathering optical response index (σ)

that it is possible to reconstruct a HR image from a sequence of degraded images that may be affected with the rotational parameters (yaw, pitch and roll) during the UAV flight, and the reconstructed images have a very good visual quality. Figure 29 shows a comparison between the realizable fidelity for the enhanced CDC model. It can be seen that the maximum realizable fidelity is achieved when the registration and fusion errors are described by Laplacian model which is similar to the assumption proposed by Farsiu et al. [79]. Also, if the error is modeled by Gaussian, it will still have a higher fidelity compared to the cases of using the model without the new error parameter or using the uniform probability model.

4.9.2 MULTI-RESOLUTION SR IMAGE

During our simulations we consider two different cases: (1) LR images are sufficient to reconstruct a full-SR image in the vertical and horizontal dimensions and (2) LR images are insufficient to reconstruct partial-SR image. We compare the performance of our method with some of the well-known SR reconstruction approaches in terms of fidelity, visual quality assessment and the computational time. These techniques are non-uniform interpolation, Papoulis-Gerchberg [85], iterated back projection [50], robust SR [86], POCS [43] and structure-adaptive normalized convolution [87]. All of these algorithms were developed at the Laboratory of Audiovisual Communications (LCAV), École Polytechnique Fédérale de Lausanne (EPFL), Switzerland [88].

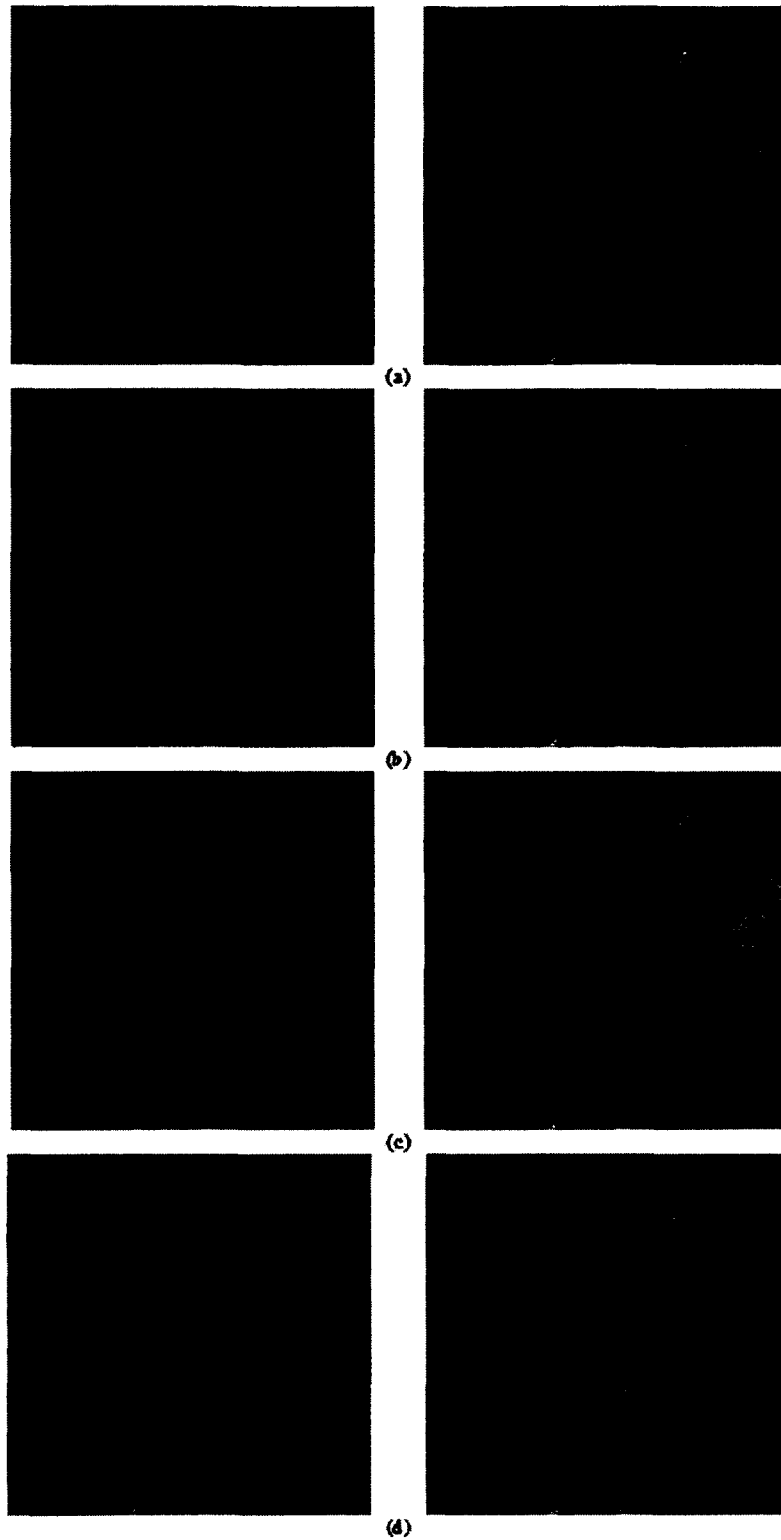


FIG. 27: (left) Degraded low resolution images; (right) random polygon reconstructed HR images for different image-gathering designs. (a) $\sigma = 0.2$; (b) $\sigma = 0.4$; (c) $\sigma = 0.8$; (d) $\sigma = 1$.

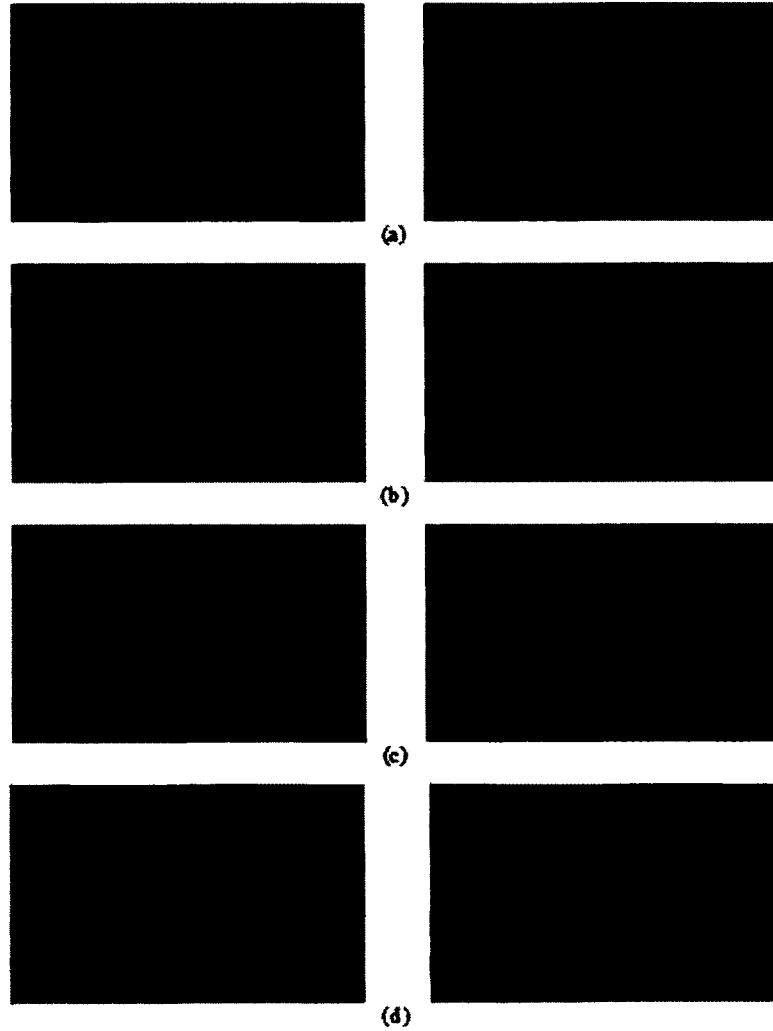


FIG. 28: (left) Degraded low resolution images; (right) checkerboard reconstructed HR images for different image-gathering designs. (a) $\sigma = 0.2$; (b) $\sigma = 0.4$; (c) $\sigma = 0.8$; (d) $\sigma = 1$ (Without HR grid optimal allocation).

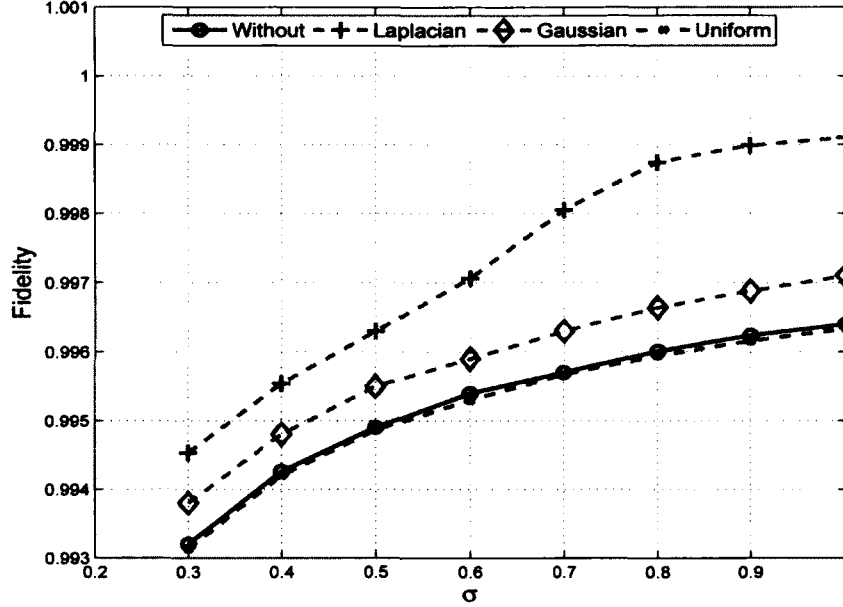


FIG. 29: Fidelity comparison for different registration and fusion models with different image-gathering optical index σ

Case 1: Full-SR

In this case the blurred and noisy image is sampled at a half-pixel location, resulting in 4 images of dimensions 128×128 . These images are used to reconstruct the SR images using the derived Wiener filter in addition to the other mentioned SR reconstruction algorithms. Figure 31 shows the reconstructed SR images using different SR approaches for $\sigma = 0.7$ and SNR=32. It can be seen that all of them are better than the LR image given in Figure 31b. Also, the Wiener reconstructed SR image is very close to the original HR image and is much better than the other reconstructed images. In addition to comparing the visual quality of the reconstructed images, we examine the fidelity between the different reconstructed images and the original simulated scene against either the optical response with a fixed SNR=32 or a SNR with a fixed $\sigma = 1$. The results can be seen in Figure 32a and Figure 32b. The fidelity of the Wiener reconstructed images is greater than the fidelities of the other reconstructed images. Also, The fidelity is improved when the optical response index is increased. This is because when the optical index is decreasing, the reconstructed image will lose some of its high frequency content, which will be reflected in its visual quality and its fidelity.

The loss in the decomposed fidelity components versus the image gathering optical index are plotted in Figure 33 in the case of full-SR and SNR=32. It can be seen that the loss in the fidelity due to aliasing is increasing when σ is increasing while the loss in the filtered component is decreasing. This can be explained as follows: as σ is increasing, the image becomes less blurry and hence the loss in the filtered component will decrease. Conversely, the loss due to aliasing will increase as σ increases, as the aliasing is introduced in the images whenever the image gathering device SFR extends beyond the sampling passband ($\sigma = 0.5$). Also, the more blur in the image the more reduction in the energies of the high frequency components which contain the aliased components; hence, their effect on its corresponding fidelity loss decreases. The loss in the noise is constant as it is assumed white and independent of the optical index of the image gathering device. So, any decrease in the loss of one of the fidelity components will affect the other terms. Over all, the performance of the system in terms of total system fidelity is improved when sigma is greater than 0.6.

Case 2: Partial-SR

Here the blurred and noisy image is sampled at a quarter-pixel location resulting in 16 images of dimensions 64×64 . We select 4 of them to reconstruct the SR images using the same procedure described in the first case. Figure 30 shows the reconstructed SR images using different SR approaches for $\sigma = 0.7$ and SNR=32. It can be seen that all the reconstructed SR images are blurrier than the full-SR reconstructed images due to the interpolation and reconstruction processes. Also, all the reconstructed images are better than the LR image given in Figure 30b. The Wiener reconstructed SR image is close to the original HR image and is much better than the other reconstructed images. Figure 34a and Figure 34b show a fidelity comparison between different SR reconstructed images when changing either the SNR and keeping σ at 0.7 or changing σ and keeping SNR at 32. Similarly, the fidelity of the Wiener reconstructed image is greater than the fidelities of the other reconstructed images, and it is improved when the optical response index is increased.

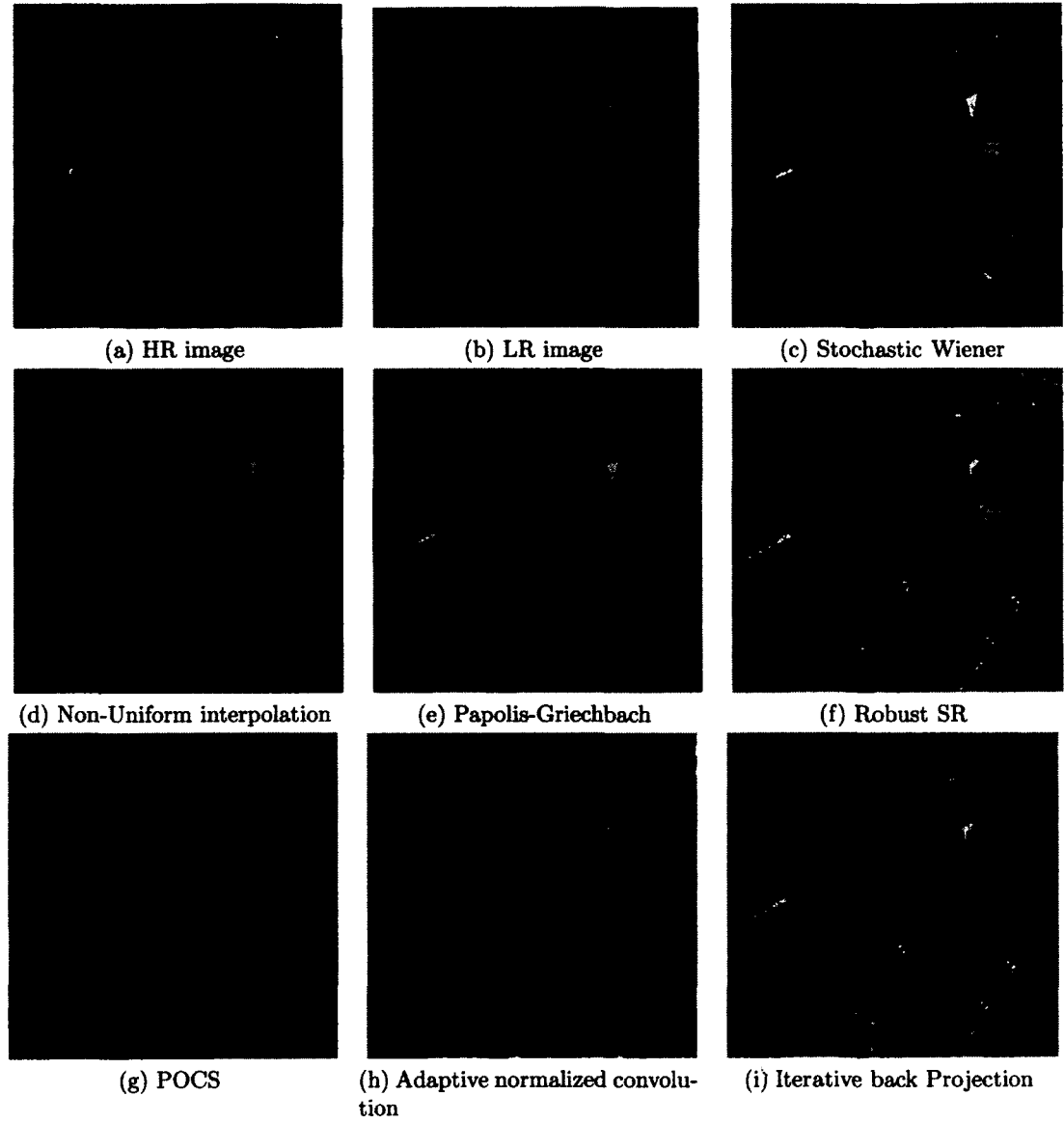


FIG. 30: Different Partial-SR reconstructed images ($\sigma = 0.7$ and $\text{SNR}=32$)

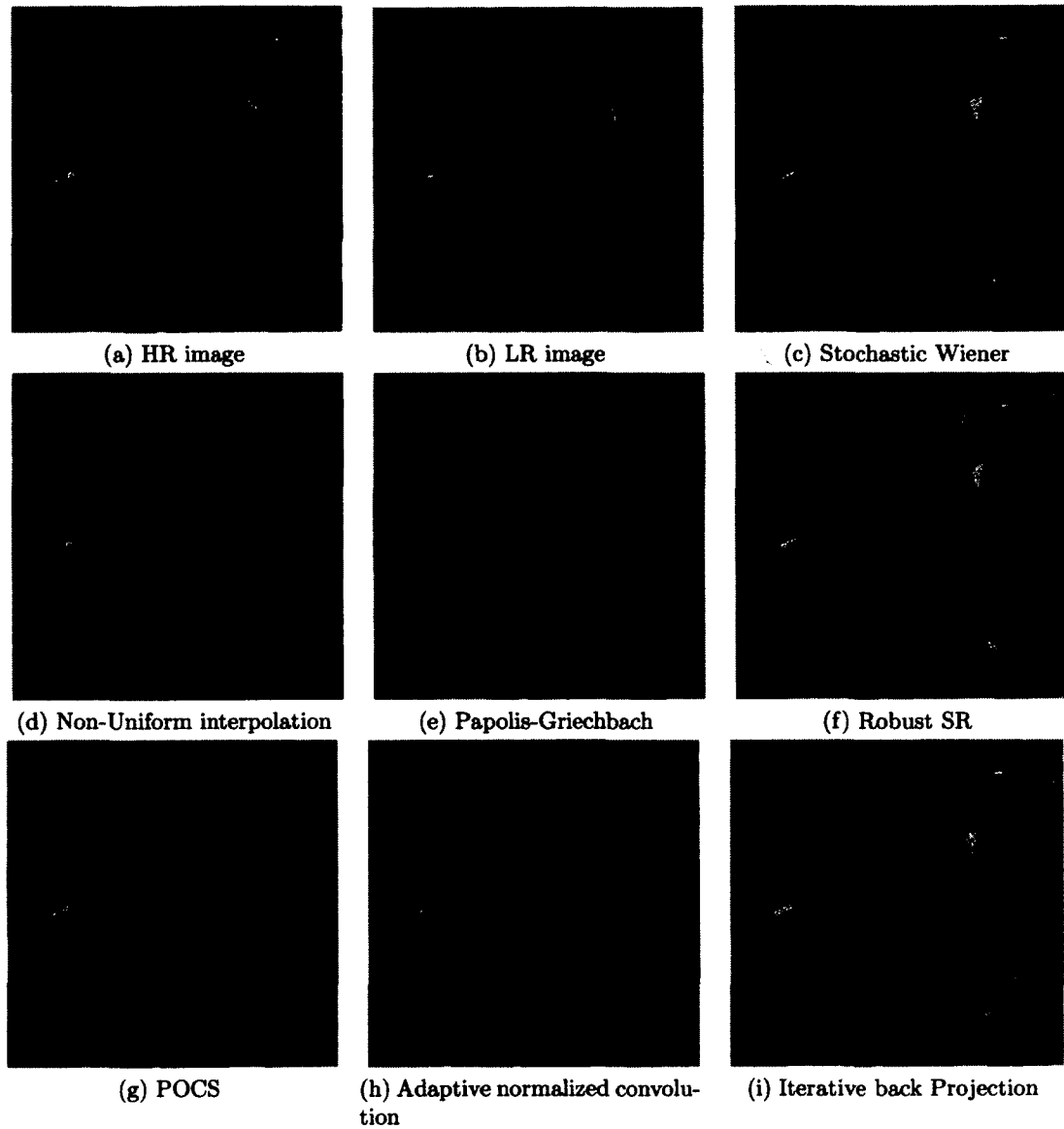


FIG. 31: Different full-SR reconstructed images ($\sigma = 0.7$ and $\text{SNR}=32$)

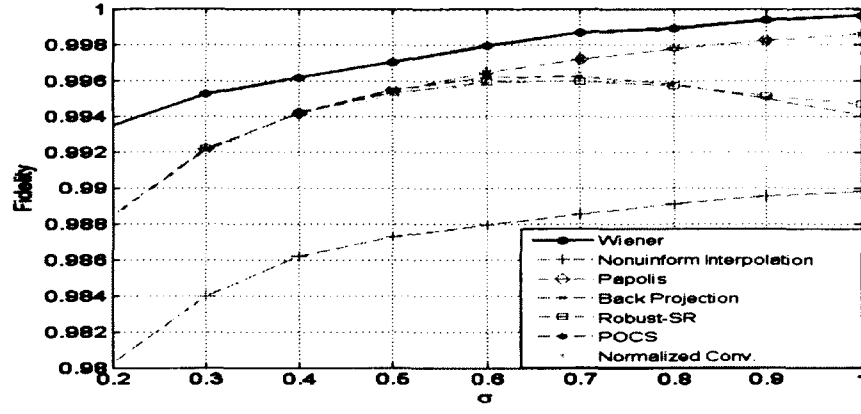
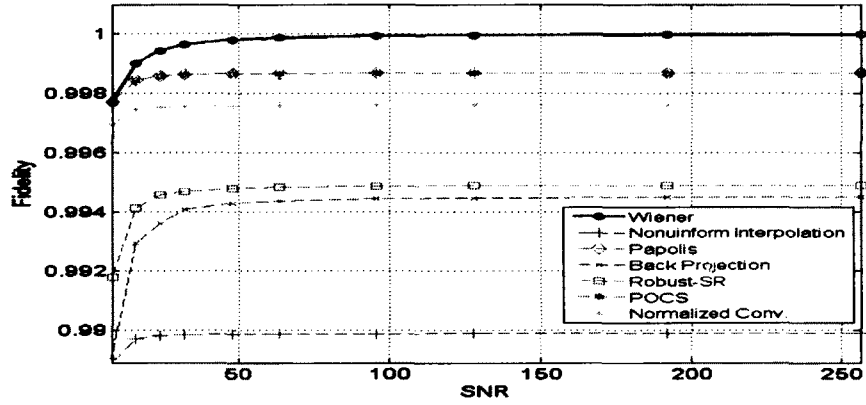
(a) Fidelity versus sigma σ (SNR=32)(b) Fidelity versus SNR ($\sigma = 1$)

FIG. 32: Fidelity comparison for different full-SR reconstructed images

4.9.3 MORE ON EXPERIMENTAL SR

SR reconstructions in real applications have many challenging problems. It is crucial to accurately estimate the subpixel shifts of LR images, as they have a great impact on the reconstruction of SR images. Also, estimating the optical blur caused by system optics and the noise caused by photodetectors greatly affect the restoration process, which in turn affects the visual quality of the reconstructed and restored images. Throughout our work we assume that the blur and noise are already known, and we are concerned with the accurate estimation of the subpixel shifts and their optimal allocation to the HR grid as presented in our recent work [16]. This optimal allocation depends on readjusting the subpixel shifts of LR images to match a uniformly spaced pattern using MSDA approach and Fourier transform

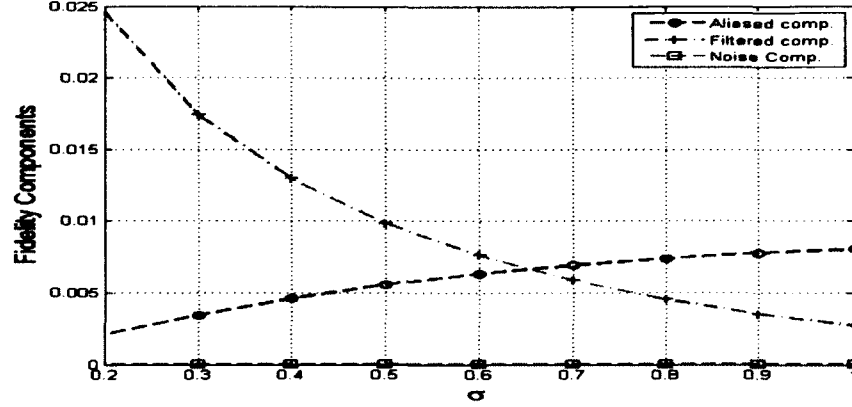


FIG. 33: Fidelity components (\mathcal{F}_a , \mathcal{F}_f and \mathcal{F}_n) versus σ for Wiener reconstructed images (SNR=32)

shift property. Throughout the simulations, we neglect the effect of the camera blur and noise and subsample the acquired images to 256×256 which will be used for the quantitative assessment of SR reconstruction and restoration. Then, degradations are superimposed to the images as we did in the random polygon images. The images will be subsampled at quarter and half pixel locations to examine both partial-SR and full-SR cases, respectively. The degraded images will be registered with respect to a reference image to a subpixel precision. Table 3 lists the estimated subpixel shifts between the different images and the reference one and it can be seen that they're random and don't follow a uniform pattern. They will be adjusted and optimally mapped to a uniform HR grid listed in Table 4. The Wiener filter is applied to the composed HR grid and a comparison between its output and the different SR approaches outputs is displayed in Figure 35 and Figure 36 for the Full and partial-SR cases, respectively. Images with full SR have better visual quality and sharpness than the ones with partial SR. Also, The images with Wiener restoration are much better than the other techniques in both the cases of full and partial SR. Comparisons between the fidelity for the different SR reconstruction techniques are listed in Table 5 and Table 6 for Full-SR and Partial-SR cases, respectively. Images with Wiener restoration have higher fidelity than the other techniques. In addition, full-SR images have higher fidelities than partial-SR images.

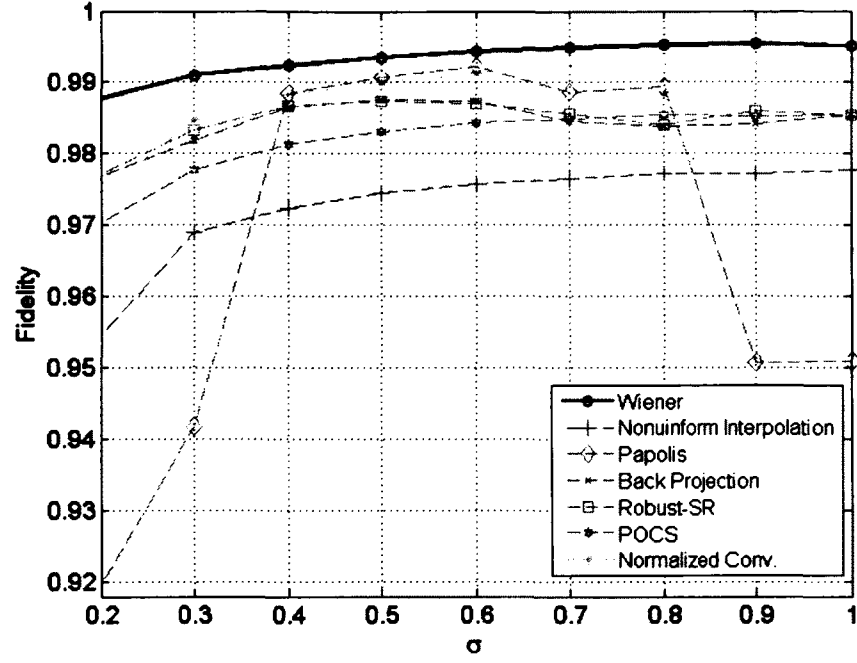
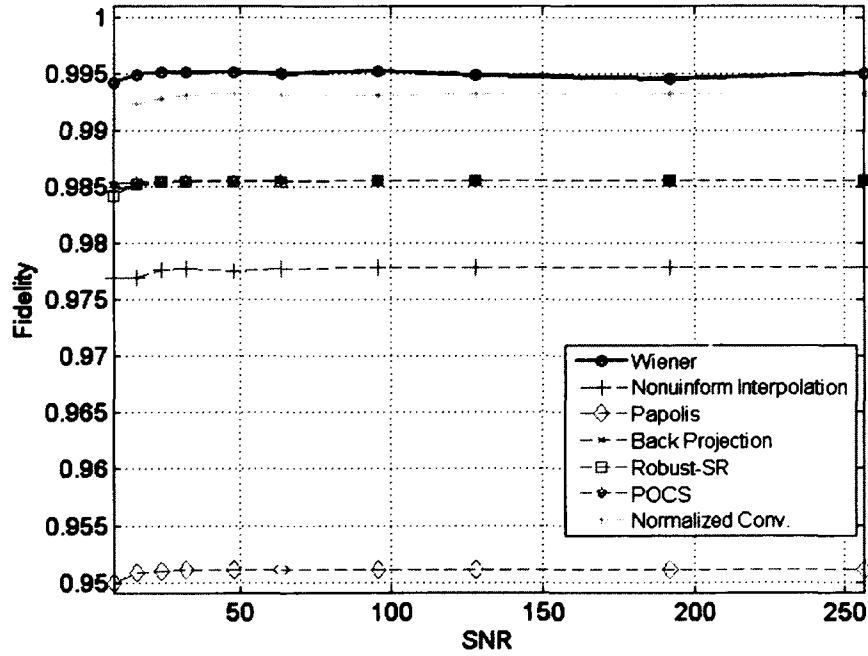
(a) Fidelity versus sigma σ (SNR=32)(b) Fidelity versus SNR ($\sigma = 1$)

FIG. 34: Fidelity comparison for different partial-SR reconstructed images

TABLE 3: Estimated subpixel shifts

	Subpixel Shifts	
	x -direction	y -direction
LR 1	-0.01	-0.06
LR 2	-0.03	-0.03
LR 3	-0.04	-0.03

TABLE 4: Adjusted Subpixel Shifts

Image	New subpixel shift
Reference image	(0.5, 0.5)
LR 1	(0.5, 0)
LR 2	(0, 0.5)
LR 3	(0, 0)

TABLE 5: Fidelity comparison for full-SR reconstructed checkerboard SR images

SR Tech- niques	Stochastic Wiener	Non- Uniform interpo- lation	Robust SR	POCS	Adaptive convolu- tion	Iterative back pro- jection
Fidelity	0.984	0.953	0.963	0.956	0.96	0.964

TABLE 6: Fidelity comparison for partial-SR reconstructed checkerboard SR images

SR Tech- niques	Stochastic Wiener	Non- Uniform interpo- lation	Robust SR	POCS	Adaptive convolu- tion	Iterative back pro- jection
Fidelity	0.927	0.852	0.896	0.875	0.623	0.898

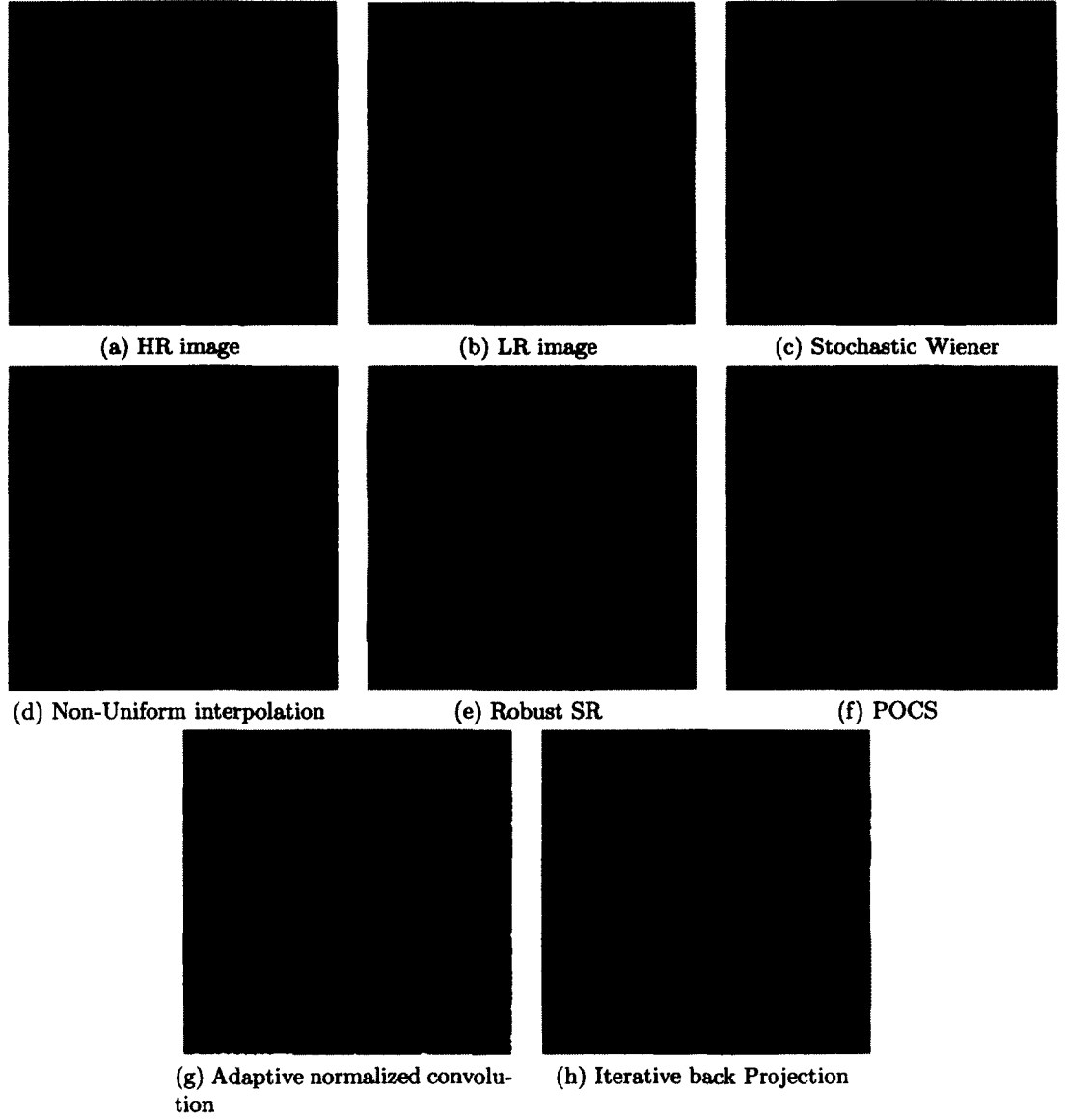


FIG. 35: Different full-SR reconstructed checkerboard images with optimal HR grid allocation ($\sigma = 0.8$ and $\text{SNR}=128$)

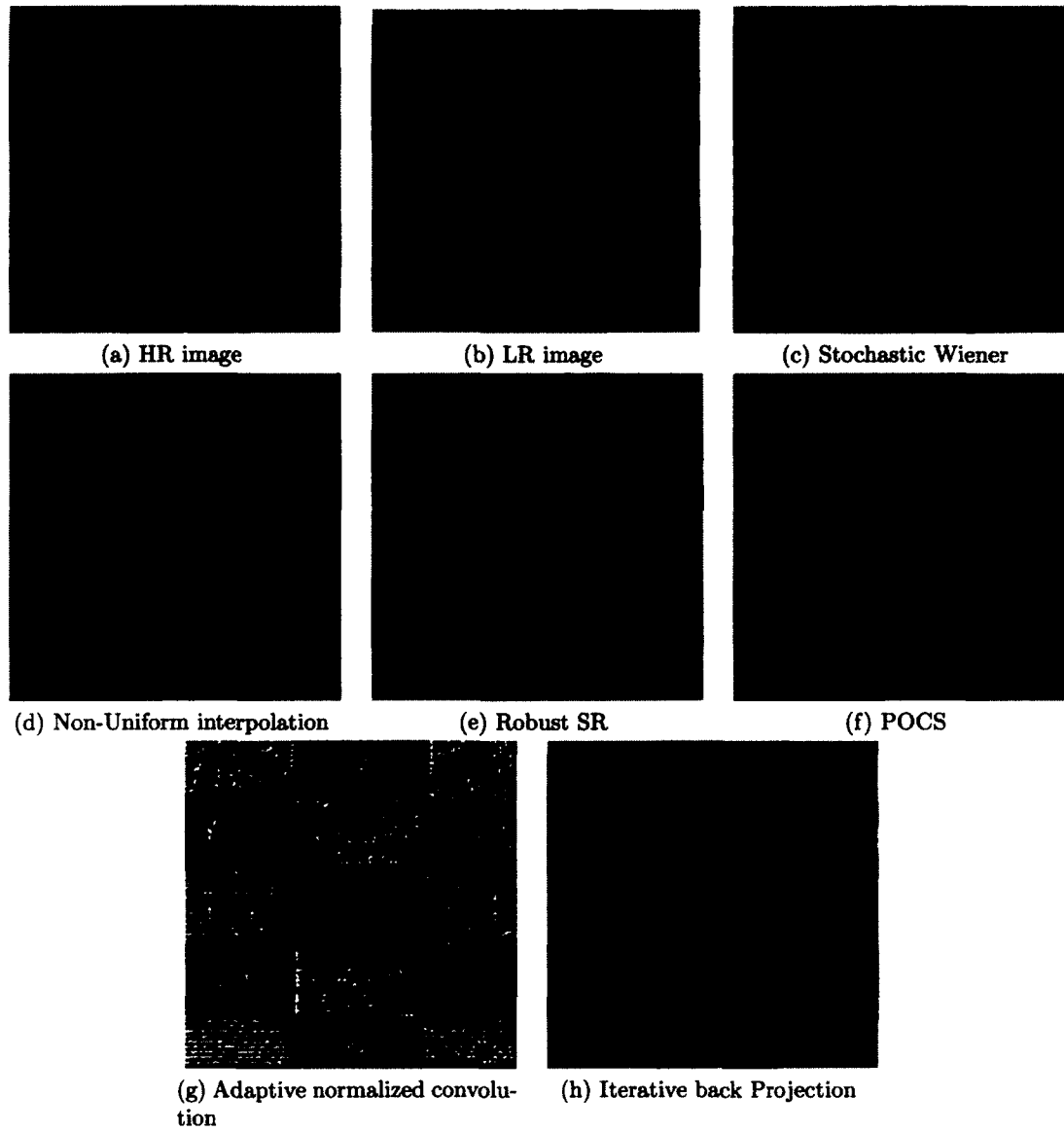


FIG. 36: Different partial-SR reconstructed checkerboard images with optimal HR grid allocation($\sigma = 0.8$ and $\text{SNR}=128$)

TABLE 7: Computational time for full-SR reconstructed SR images (s)

SR Techniques	Stochastic Wiener	Non-Uniform interpolation	Robust SR	POCS	Adaptive convolution	Iterative back projection
Full-SR	0.92	2.839	36.56	13.665	11.716	13.713
Partial-SR	0.434	0.96	35.319	12.616	7.176	5.756

TABLE 8: Improvements in computational time using speeded-up Wiener filter (s)

SR Techniques	Normal Wiener	Wiener with 4 partitions	Wiener with 16 partitions
Full-SR	0.92	0.3978	0.1326
Partial-SR	0.434	0.257	0.089

4.9.4 COMPUTATIONAL COSTS

The computational costs of the different SR techniques are listed in Table 7. The simulations are performed using MATLAB 7.8 Release 2009a program on OPTIPLEX 780 (Intel(R) Core (TM)2 Quad 2.66 GHz CPU, 8.00 GB RAM, MS Windows 7 Professional 2009). The performance of the different SR techniques in case of partial SR is much faster than that in case of full SR. Also, Wiener computational time is much smaller than the other SR techniques. Most of Wiener computational time is consumed in a preprocessing step (0.355 s in case of partial SR and 0.78 s in case of full SR) which is considered as the main demerit of this approach. The results of the speeded-up Wiener filter are shown in Table 8. As can be seen from the results, we enhanced the performance of the filter in terms of its speed and it is faster than the original one. Also, by increasing the number of partitions we can parallelize the process and reduce both the computational time and memory requirements.

4.9.5 MSDA AND MSSA ASSESSMENTS

To evaluate the performance when using either the MSDA or the MSSA along with full-SR or partial-SR reconstruction, we need to examine the available number of LR frames on the reconstructed image along with the two other types of Wiener filter. So, if we have 4 LR frames, we can obtain full SR in both dimensions by using 4 scenels per pixel, and

by using the MSDA method we can optimally allocate the scenels to the HR grid. Table 4 shows the locations of LR scenels after using the MSDA method. Also, if we have only two of these scenels and suppose they are LR 1 and LR 2 then we can use either the MSDA or the MSSA methods to allocate them to the HR grid and obtain a partial SR in a certain dimension and interpolate the missing pixels on the other dimension. Table ?? shows the results of using the MSDA and the MSSA methods. The MSSA makes the allocation of the scenels more relaxed to the HR grid than the MSDA as the latter restricts the scenels to a pre-defined points on the HR grid. After the LR frames are allocated to the HR grid using

TABLE 9: The optimal HR grid locations for the LR scenels (partial-SR)

	LR 1	LR 3
MSDA	(0, 0.5)	(0.5, 0.5)
MSSA	(0, 0)	(0.5, 0.5)

the MSDA or the MSSA, they are sub-pixel shifted using Equation (61) to readjust their phase shifts. Then we use three different Wiener restoration filters introduced in Sections 4.3, 4.6 and 4.7 to enhance and restore the reconstructed images from the degradations encountered during the image-gathering process. We examine the cases of sufficient and insufficient scenels along with using the MSDA and MSSA methods to produce full-SR or partial-SR where the fidelity is used as a comparison metric between the original HR image and the reconstructed ones. The WIGE filter parameters used in this simulation are: $\zeta = 0.2$, $\sigma_o = 0.7$, $\sigma_i = 0.3$ and $\sigma_e = 0.8$. Figure 37 shows the different reconstructed images using different enhancement filters and the effect of allocating the LR scenels using either the MSDA and MSSA. Also, Figure 38 and Figure 39 show a fidelity comparison between the different reconstructed images and the original simulated HR scene. It can be seen that the full-SR MSDA reconstructed images are better than the ones that have partial-SR. Also, the images with MSSA allocation are much better than the the ones with partial-SR MSDA allocation. In terms of their fidelity, images with MSDA full-SR reconstruction have higher fidelity than the ones with partial-SR reconstruction using either MSDA or MSSA methods. When comparing the performance of different restoration filters along with using the different allocations methods, it can be seen that both the SWR and WCR filters perform nearly the same and they are much better than the WIGE filter reconstructed images. The WIGE filter outputs are blurrier than the outputs produced by either the SWR or WCR filters.

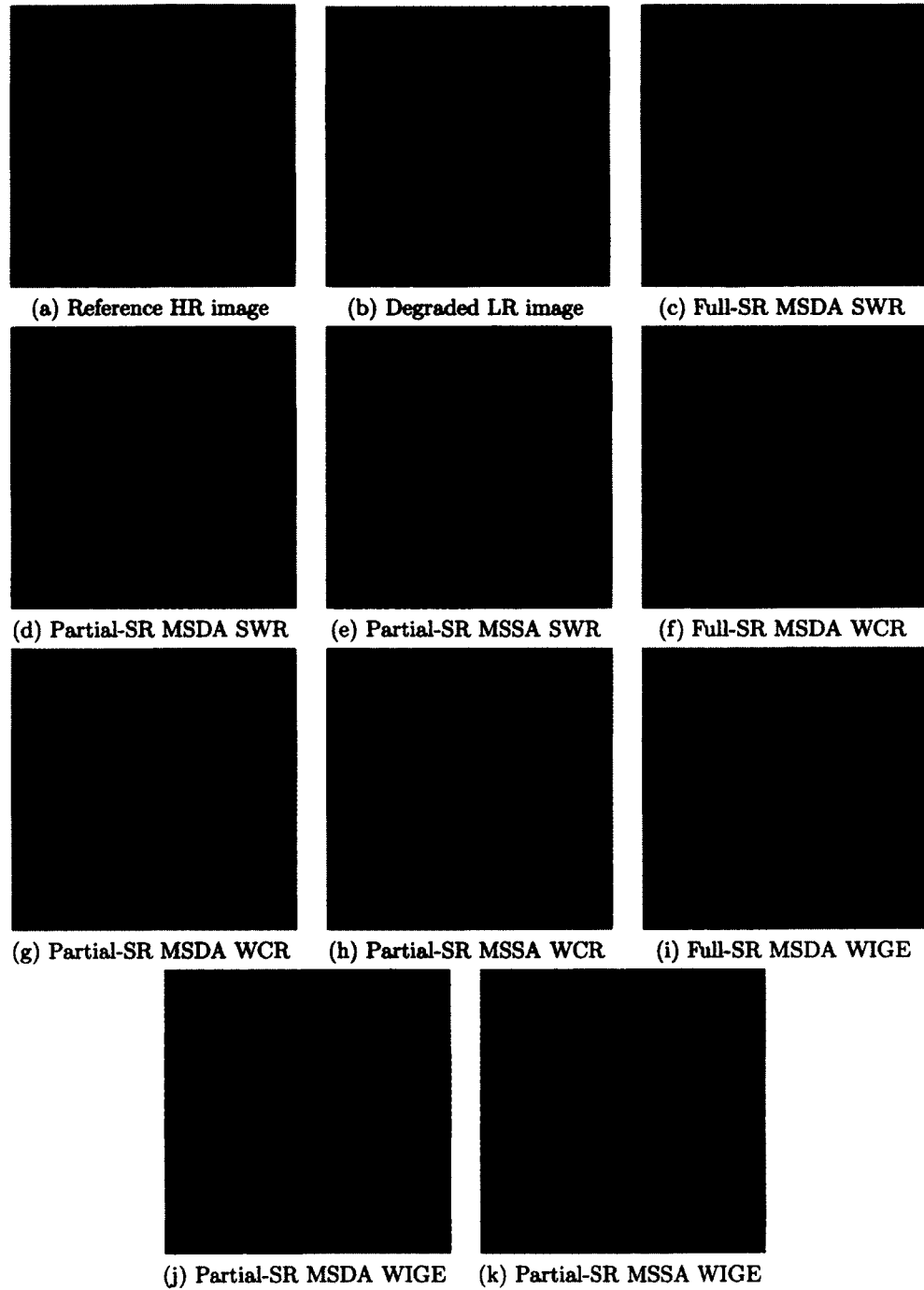
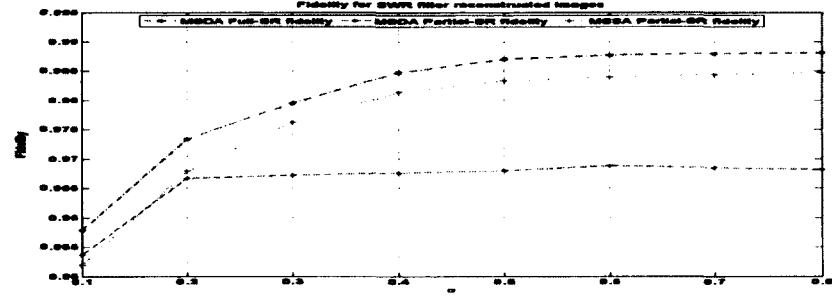
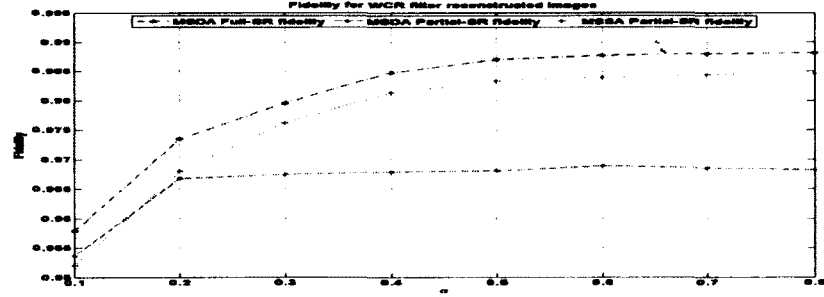


FIG. 37: Different restoration outputs at image-gathering device index $\sigma = 0.8$

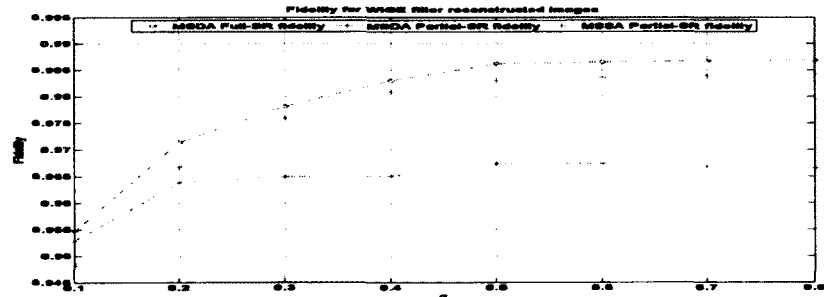
When comparing the fidelity of the reconstructed images through the different restoration filters, it can be concluded that the SWR reconstructed images are little higher than the WCR reconstructed images while the fidelity of SWR and WCR reconstructed images are much higher than the ones with WIGE reconstruction.



(a) SWR filter reconstructed images' fidelity

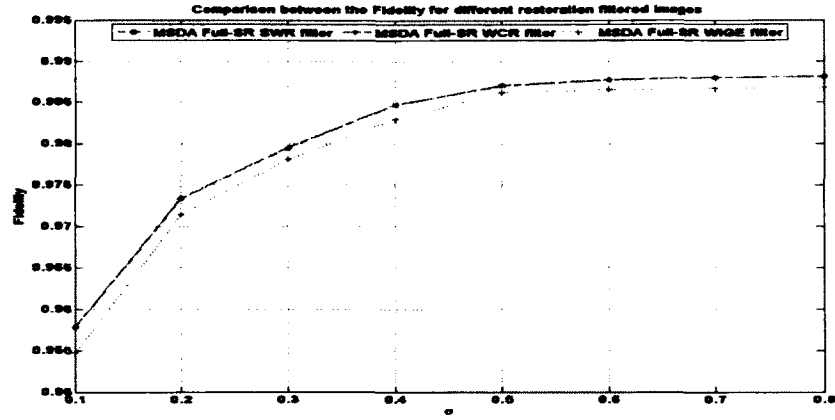


(b) WCR filter reconstructed images' fidelity

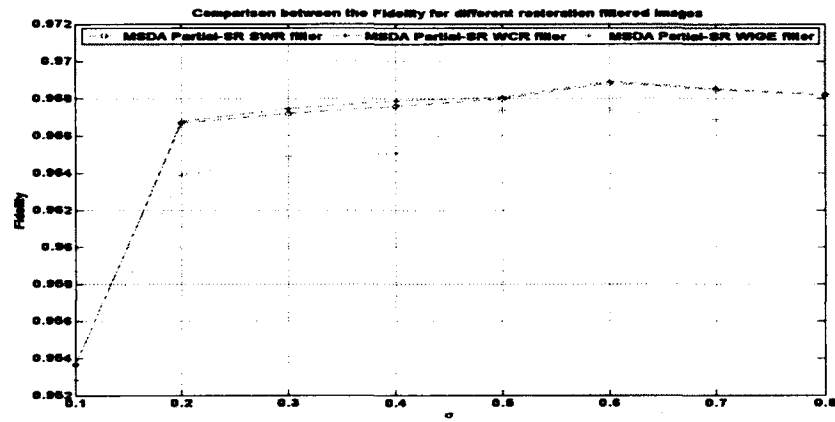


(c) WIGE filter reconstructed images' fidelity

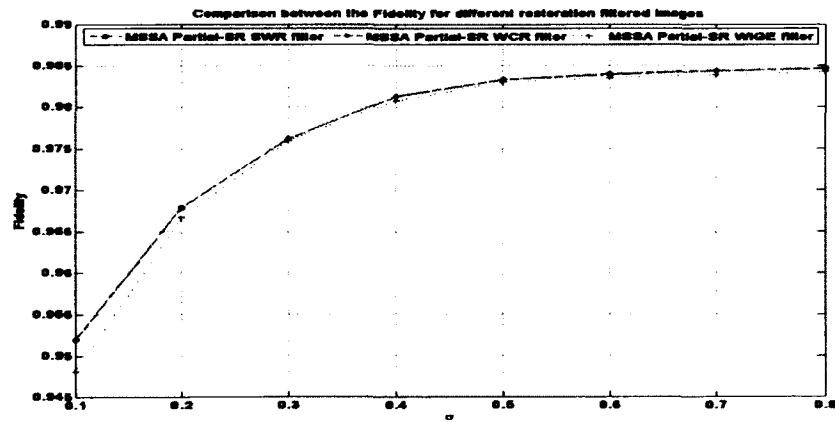
FIG. 38: Fidelity comparison for SWR, WCR, and WIGE restoration outputs with different image-gathering optical index σ



(a) MSDA different full-SR restoration filtered images fidelity comparison



(b) MSDA different partial-SR restoration filtered images fidelity comparison



(c) MSSA different partial-SR restoration filtered images fidelity comparison

FIG. 39: Fidelity comparison for different restorations with optimal allocation methods against different image-gathering optical index σ

CHAPTER 5

INFORMATION THEORETIC APPROACH FOR SR VISUAL QUALITY ASSESSMENT

5.1 INTRODUCTION

In this chapter we incorporate the information theory into the visual quality assessment of the reconstructed SR images. We emphasize the close correlation between the information rate and the visual quality of the images. The higher the information rate, the better the visual quality. The proposed model contains two major transformations: (i) the continuous-to-discrete transformation of the continuous input image $L(x, y)$ into discrete interlaced image $S(x, y)$ and (ii) the digital-to-continuous transformation of the discrete interlaced image $S(x, y)$ into the continuous output image $R_o(x, y)$. We propose two figures of merit: (i) information rate \mathcal{H} and (ii) maximum realizable fidelity \mathcal{F} that can be used to assess these transformations. The information rate assesses the first transformation while the maximum realizable fidelity assesses the end-to-end CDC transformation. In addition, the information rate is used to measure the amount of information that the image gathering device is producing, while the maximum realizable fidelity is used to compare the closeness between the reconstructed restored output images and the original input HR scenes.

5.2 INFORMATION RATE

Shannon [89, 90] defined the information rate that is produced by the image acquisition device or in other words the mutual information between the continuous input scene L and the discrete interlaced image S as

$$\mathcal{H} = \mathcal{E}[S(x, y)] - \mathcal{E}[S(x, y)|L(x, y)] \quad (132a)$$

$$\mathcal{H} = \mathcal{E}[\tilde{S}(\nu, \omega)] - \mathcal{E}[\tilde{S}(\nu, \omega)|\hat{L}(\nu, \omega)] \quad (132b)$$

where $\mathcal{E}[\cdot]$ refers to the entropy of the interlaced image $S(x, y)$ defined in the spatial and frequency domains, respectively, and $\mathcal{E}[\cdot|\cdot]$ is the conditional entropy of the same interlaced

image given the input scene. The composite image $\tilde{S}(\nu, \omega)$ in Equation (90) can be rewritten as

$$\tilde{S}(\nu, \omega) = \frac{1}{K_1 K_2} \sum_{k_1 k_2} \hat{L}(\nu, \omega) \hat{\tau}_{k_1 k_2}(\nu, \omega) \exp(-i4\pi(x_{k_1}\nu + y_{k_2}\omega)) + \hat{n}(\nu, \omega) \quad (133)$$

where $\hat{n}(\nu, \omega)$ represents the composite noise that results from aliasing and the photodetector noise and it is given by

$$\hat{n}(\nu, \omega) = \frac{1}{K_1 K_2} \sum_{k_1 k_2} [\hat{N}_a(\nu, \omega) + \tilde{N}_{k_1 k_2}(\nu, \omega)] \exp(-i2\pi(x_{k_1}\nu + y_{k_2}\omega)) \quad (134)$$

If the aliased noise components and the photo-detector noise are assumed to be independent additive Gaussian noise, then the information rate can be simplified to [90]

$$\mathcal{H} = \mathcal{E}[S(x, y)] - \mathcal{E}[n(x, y)] \quad (135a)$$

$$\mathcal{H} = \mathcal{E}[\tilde{S}(\nu, \omega)] - \mathcal{E}[\hat{n}(\nu, \omega)] \quad (135b)$$

Equation (135a) measures the mutual information between the input scene and the interlaced image $S(x, y)$. Thus it gives the obtainable amount of information of the image gathering device minus the noise component. If the Gaussian probability density $P_S[\tilde{S}(\nu, \omega)]$ of the composite image $\tilde{S}(\nu, \omega)$ and the Gaussian probability $P_n[\hat{n}(\nu, \omega)]$ of the noise are given by

$$P_S[\tilde{S}(\nu, \omega)] = \frac{1}{\pi \tilde{\Phi}_S(\nu, \omega)} \exp \left[- \left| \tilde{S}(\nu, \omega) \right|^2 / \tilde{\Phi}_S(\nu, \omega) \right] \quad (136)$$

and

$$P_n[\hat{n}(\nu, \omega)] = \frac{1}{\pi \hat{\Phi}_n(\nu, \omega)} \exp \left[- \left| \hat{n}(\nu, \omega) \right|^2 / \hat{\Phi}_n(\nu, \omega) \right] \quad (137)$$

respectively, then the information rate \mathcal{H} can be written as [91]

$$\begin{aligned} \mathcal{H} &= \frac{1}{2} \iint_{\hat{B}} \log_2 \frac{\tilde{\Phi}_S(\nu, \omega)}{\hat{\Phi}_n(\nu, \omega)} d\nu d\omega \\ &= -\frac{1}{2} \iint_{\hat{B}} \log_2 \frac{\hat{\Phi}_n(\nu, \omega)}{\tilde{\Phi}_S(\nu, \omega)} d\nu d\omega \end{aligned} \quad (138)$$

where the PSD $\tilde{\Phi}_S(\nu, \omega)$ is given by

$$\begin{aligned} \tilde{\Phi}_S(\nu, \omega) = & \frac{1}{K_1^2 K_2^2} \left[\hat{\Phi}_L(\nu, \omega) * \hat{\mathbb{I}}(\nu, \omega) \sum_{mn} \left| \sum_{k_1 k_2} \hat{\tau}_{k_1 k_2}(\nu - m, \omega - n) \right. \right. \\ & \times \left. \exp(-i2\pi((2\nu - m)x_{k_1} + (2\omega - n)y_{k_2})) \right|^2 \\ & + \left. \sum_{k_1 k_2} \hat{\Phi}_{k_1 k_2}(\nu, \omega) * \hat{\mathbb{I}}(\nu, \omega) \right] \end{aligned} \quad (139)$$

and the PSD $\hat{\Phi}_n(\nu, \omega)$ is given by

$$\begin{aligned} \hat{\Phi}_n(\nu, \omega) = & \frac{1}{K_1^2 K_2^2} \left[\hat{\Phi}_L(\nu, \omega) * \hat{\mathbb{I}}_s(\nu, \omega) \sum_{mn} \left| \sum_{k_1 k_2} \hat{\tau}_{k_1 k_2}(\nu - m, \omega - n) \right. \right. \\ & \times \left. \exp(-i2\pi((2\nu - m)x_{k_1} + (2\omega - n)y_{k_2})) \right|^2 \\ & + \left. \sum_{k_1 k_2} \hat{\Phi}_{k_1 k_2}(\nu, \omega) * \hat{\mathbb{I}}(\nu, \omega) \right] \end{aligned} \quad (140)$$

Accordingly, $\tilde{\Phi}_S(\nu, \omega)$ can be written as

$$\begin{aligned} \tilde{\Phi}_S(\nu, \omega) = & \hat{\Phi}_n(\nu, \omega) + \frac{1}{K_1^2 K_2^2} \hat{\Phi}_L(\nu, \omega) \left| \sum_{k_1 k_2} \hat{\tau}_{k_1 k_2}(\nu, \omega) \right. \\ & \times \left. \exp(-i4\pi(\nu x_{k_1} + \omega y_{k_2})) \right|^2 \end{aligned} \quad (141)$$

By using Equation (140) and (141) into Equation (138), the information rate \mathcal{H} can be written as

$$\begin{aligned} \mathcal{H} = & -\frac{1}{2} \iint_{\hat{B}} \log_2 \left[1 - \frac{1}{K_1^2 K_2^2} \frac{\hat{\Phi}_L(\nu, \omega)}{\hat{\Phi}_S(\nu, \omega)} \left| \sum_{k_1 k_2} \hat{\tau}_{k_1 k_2}(\nu, \omega) \right. \right. \\ & \times \left. \left. \exp(-i4\pi(\nu x_{k_1} + \omega y_{k_2})) \right|^2 \right] \end{aligned} \quad (142)$$

From Equation (101), $\tilde{\Phi}_S(\nu, \omega)$ can be written as

$$\tilde{\Phi}_S(\nu, \omega) = \frac{1}{K_1 K_2} \frac{\hat{\Phi}_L(\nu, \omega)}{\hat{\Psi}(\nu, \omega)} \sum_{k_1 k_2} \hat{\tau}_{k_1 k_2}^*(\nu, \omega) \exp(i4\pi(\nu x_{k_1} + \omega y_{k_2})) \quad (143)$$

Substituting Equation (143) into Equation (142) yields

$$\mathcal{H} = -\frac{1}{2} \iint_{\hat{B}} \log_2 \left[1 - \frac{\hat{\Psi}(\nu, \omega)}{K_1 K_2} \sum_{k_1 k_2} \hat{\tau}_{k_1 k_2}(\nu, \omega) \exp(-i4\pi(\nu x_{k_1} + \omega y_{k_2})) \right] \quad (144)$$

Thus, the information rate produced by the imaging gathering device depends on its optical response index, the design of Wiener filter, and the inter-subpixel shifts between the individual LR frames. From Equations (97) and (101), the minimum MSRE error $\hat{\epsilon}_{\min}^2(\nu, \omega)$ can be expressed as

$$\hat{\epsilon}_{\min}^2(\nu, \omega) = \hat{\Phi}_L(\nu, \omega) \left(1 - \hat{\Gamma}(\nu, \omega) \right) \quad (145)$$

where $\hat{\Gamma}$ is given by

$$\hat{\Gamma}(\nu, \omega) = \frac{\hat{\Psi}(\nu, \omega)}{K_1 K_2} \sum_{k_1 k_2} \hat{\tau}_{k_1 k_2}(\nu, \omega) \exp(i4\pi(x_{k_1}\nu + y_{k_2}\omega)) \quad (146)$$

By comparing Equations (144), (145) and (146), the information rate can be written as

$$\mathcal{H} = -\frac{1}{2} \iint_{\hat{B}} \log_2 \left[1 - \hat{\Gamma}(\nu, \omega) \right] d\nu d\omega \quad (147a)$$

$$\mathcal{H} = \frac{1}{2} \iint_{\hat{B}} \log_2 \frac{\hat{\Phi}_L(\nu, \omega)}{\hat{\epsilon}_{\min}^2} d\nu d\omega \quad (147b)$$

$$\mathcal{H} = \frac{1}{2} \iint_{\hat{B}} \log_2 \hat{\Phi}_L(\nu, \omega) d\nu d\omega - \frac{1}{2} \iint_{\hat{B}} \log_2 \hat{\epsilon}_{\min}^2 d\nu d\omega \quad (147c)$$

The first term in Equation (147c) represents the information rate that the image gathering device can produce without any degradations during the acquisition process and the second term refers to the loss in the information rate caused by the CDC system degradations.

5.3 MAXIMUM REALIZABLE FIDELITY

Linefoot [81] defined as the fidelity the similarity between the continuous input scene and the reconstructed output scene as given by

$$\mathbf{F} = 1 - \sigma_L^{-2} \iint \hat{\epsilon}^2(\nu, \omega) d\nu d\omega \quad (148)$$

By substituting the minimum MSRE cost function given by Equation (145) into Equation

(148), the maximum realizable fidelity \mathcal{F} is given by

$$\mathcal{F} = \sigma_L^{-2} \iint \hat{\Phi}_L(\nu, \omega) \hat{\Gamma}(\nu, \omega) d\nu d\omega \quad (149)$$

From Equation (147), the information rate spectral distribution $\hat{\mathcal{H}}(\nu, \omega)$ is given by

$$\hat{\mathcal{H}}(\nu, \omega) = -\log_2 \left[1 - \hat{\Gamma}(\nu, \omega) \right] \quad (150)$$

Accordingly the maximum realizable fidelity \mathcal{F} is given by

$$\mathcal{F} = \sigma_L^{-2} \iint \hat{\Phi}_L(\nu, \omega) \left[1 - 2^{-\hat{\mathcal{H}}(\nu, \omega)} \right] d\nu d\omega \quad (151)$$

It can be concluded that whenever the information rate spectral distribution $\hat{\mathcal{H}}(\nu, \omega)$ is high over all the spectrum, \mathcal{F} hits its maximum value for a given image gathering device.

In our simulations and to assess the close correlation between the visual quality and both the information rate and the maximum realizable fidelity, we discussed three cases: (a) for a given MSD and SNR of 3 and 32 respectively, we compare the change in the information rate with the optical response index σ that controls the design of the image gathering device; (b) Similar to the first case, we compare the change in the information rate with the total SNR of the system when σ is 0.8 and for a given input scene with MSD of 3; (c) finally, we discuss the change in the information rate with the MSD of the input scenes while keeping both the SNR at 32 and for a given image gathering device with optical response index of $\sigma = 0.8$. Figure 40, 41, and 42 show the results of these three cases. The information rate produced by the image gathering device improves with the increase in its spatial response index that controls the trade off between the aliasing and blur and also it improves with the enhancement in the system SNR that reflects the reduction in the effect of the total noise in the system. Also, the amount of information produced by the image device is affected with the mean spatial detail of the scene. The smaller the MSD, the higher the information contained in the output scene. Finally, the performance in full-SR case is better than partial-SR case as more information is contained in the former case. Figures 43-48 show the reconstructed and restored outputs for different image gathering device designs, different SNR and different input scenes for full-and partial-SR cases. The reconstructed images in both full-SR and partial-SR are given in Figure 43 and 44 for various image designs. It can be seen that the visual quality improves as the optical response index of the image gathering

device increases, and the quality of the images in full-SR is better than partial-SR. Also, there's a close correlation between the visual quality and the information rate. The higher the information rate the better the visual quality. It can be concluded from Figure 45 and 46 that the visual quality and the information rate are also high whenever the SNR is high. Figure 47 and 48 depict the reconstructed outputs for different input scenes. The smaller the MSD of the input scene, the higher the information rate. Also, the full-SR reconstructed outputs are sharper than the partial-SR cases as the information contained in full-SR images is much higher than the partial-SR images.

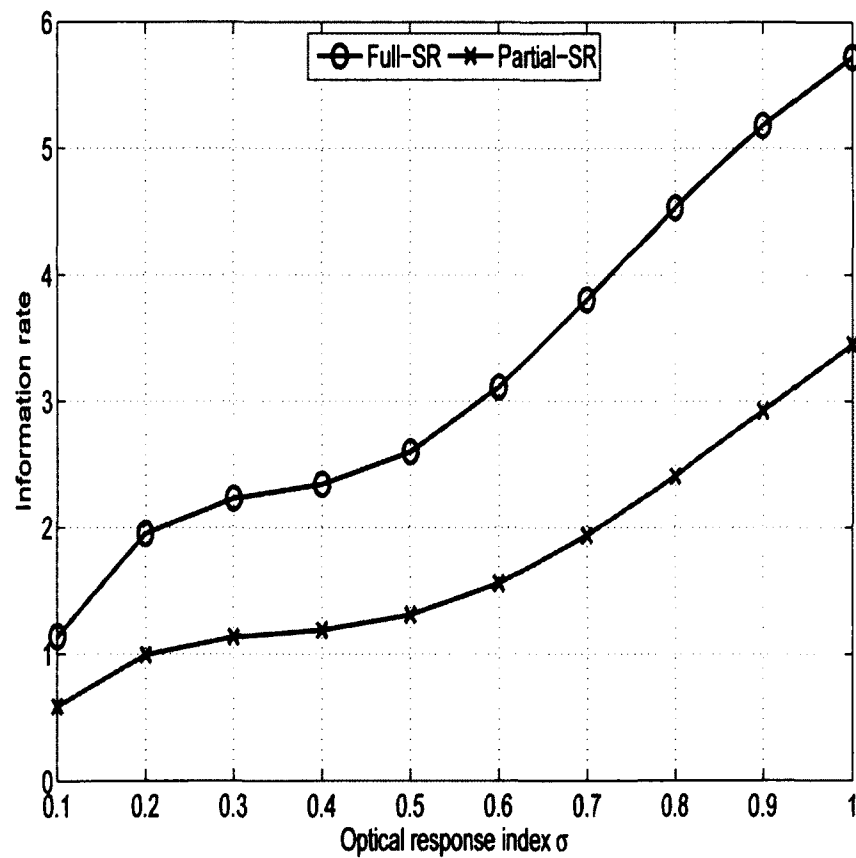


FIG. 40: Information rate comparison against optical response index σ

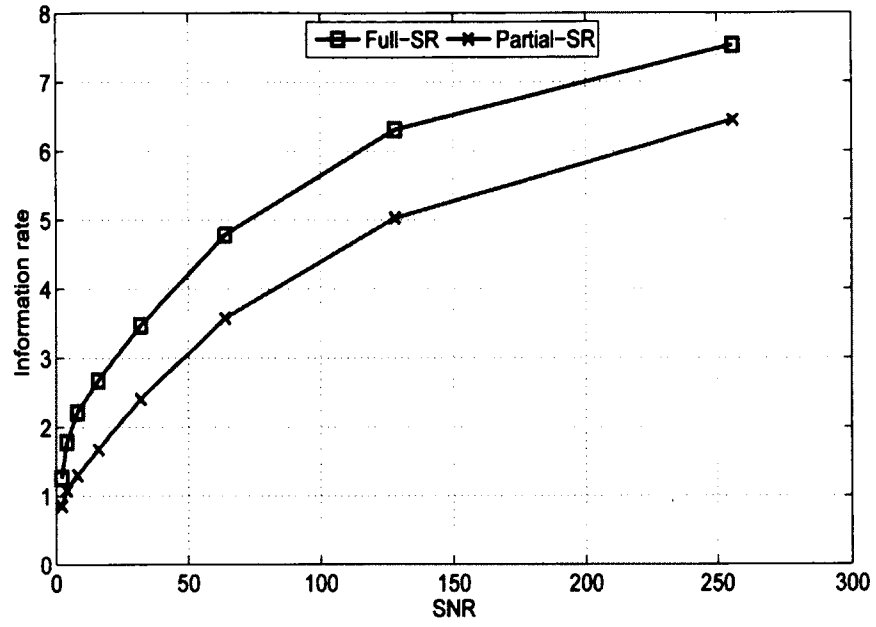
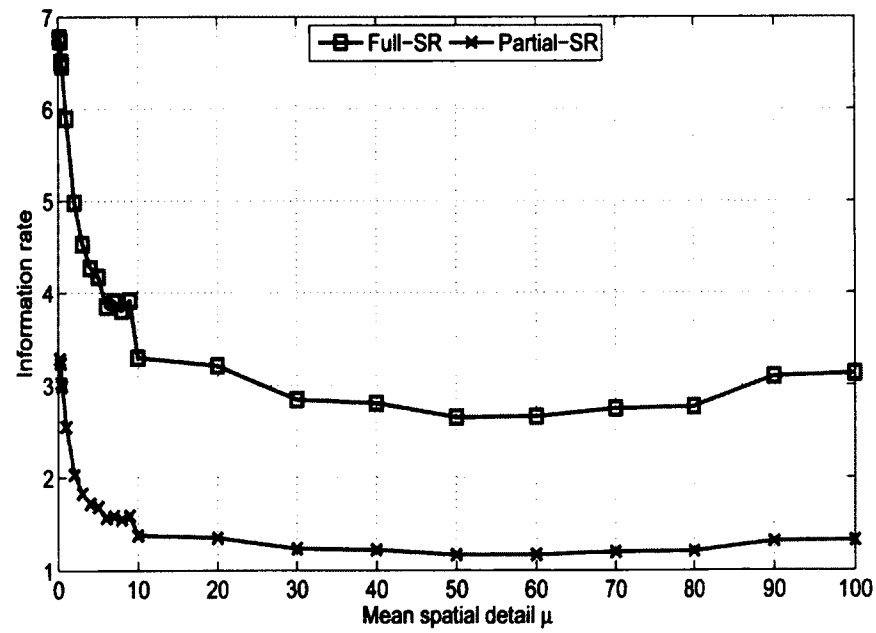


FIG. 41: Information rate comparison against SNR

FIG. 42: Information rate comparison against mean spatial detail μ

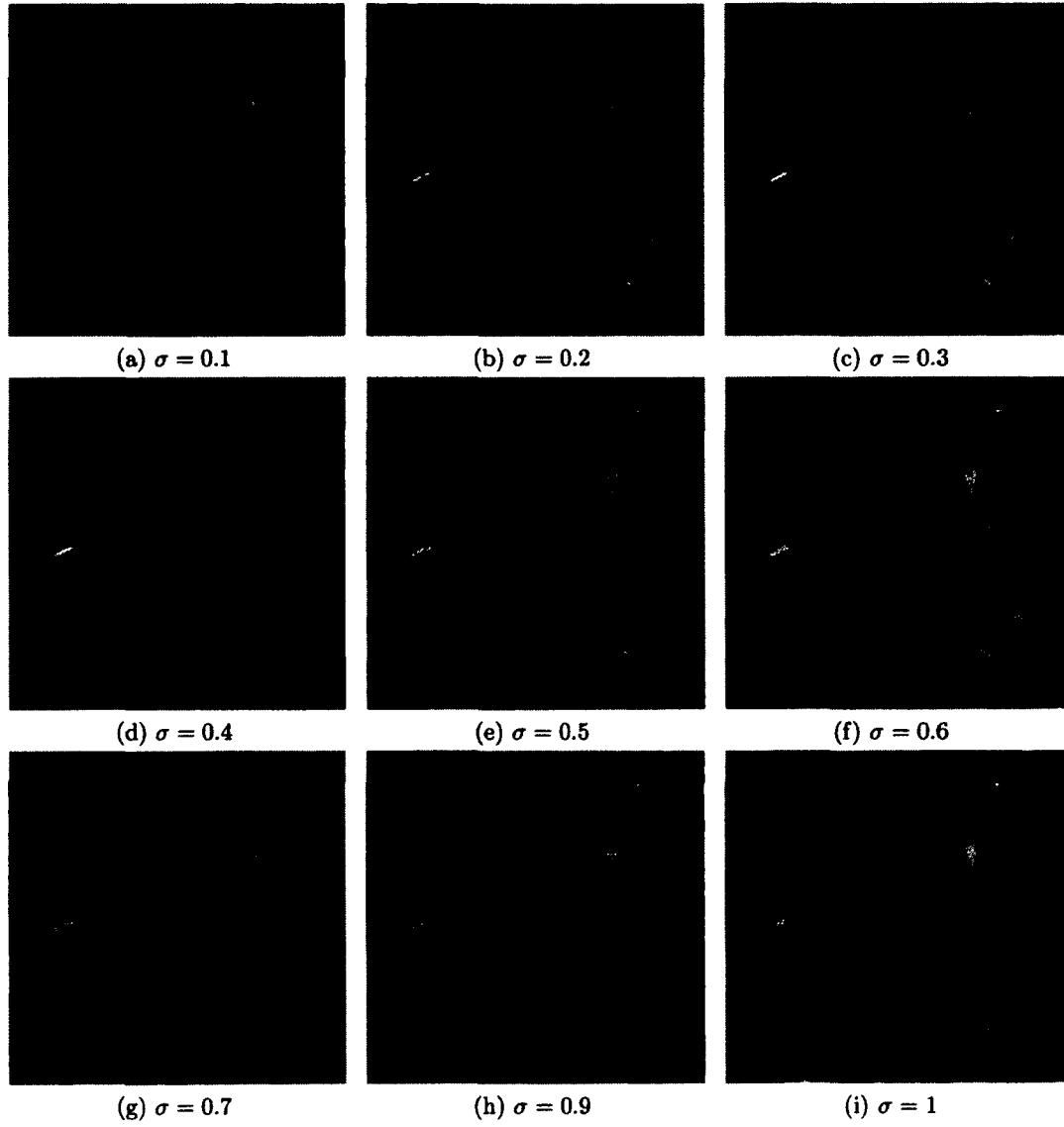


FIG. 43: Different full-SR reconstructed outputs for various σ

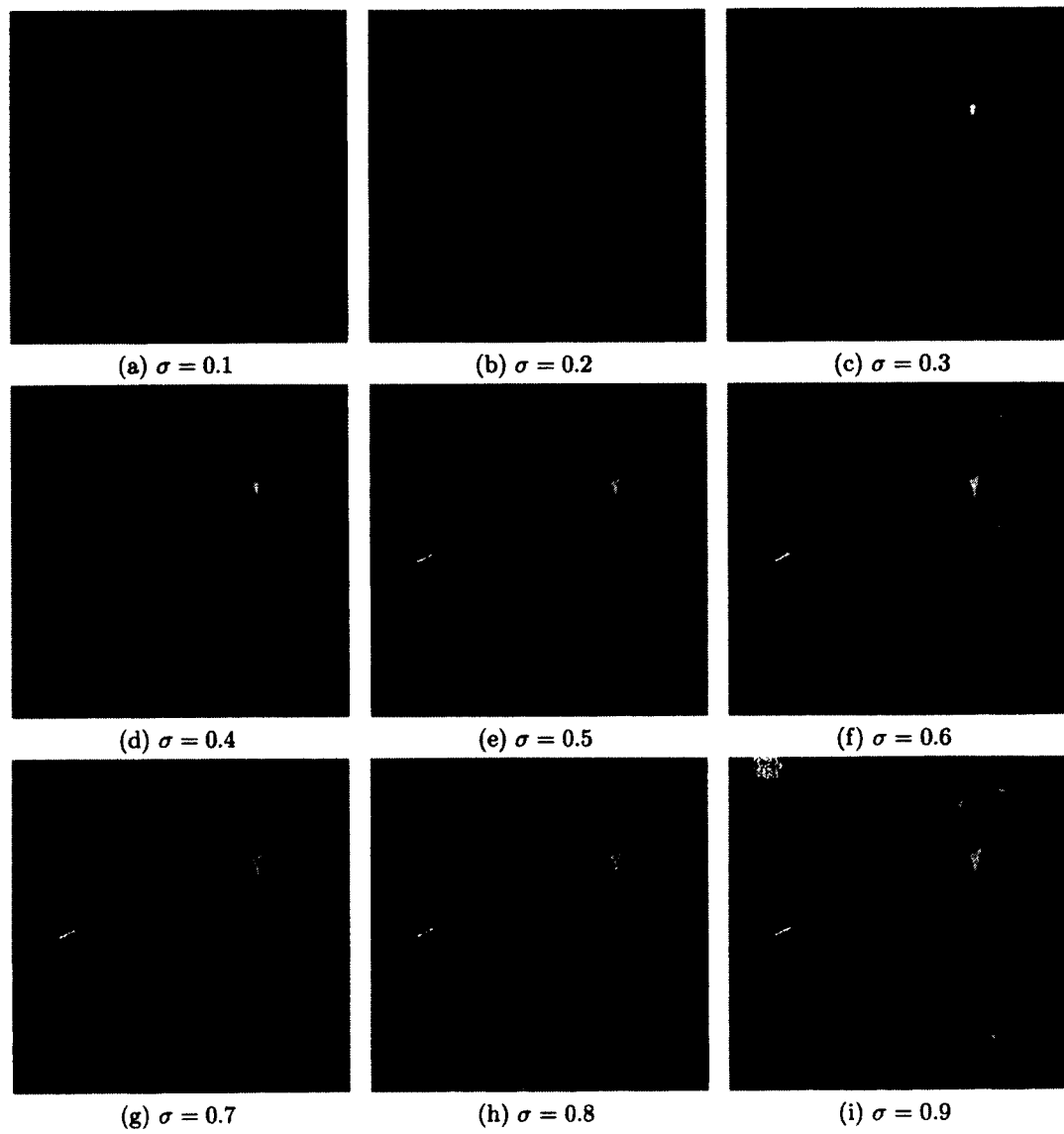


FIG. 44: Different partial-SR reconstructed outputs for various σ

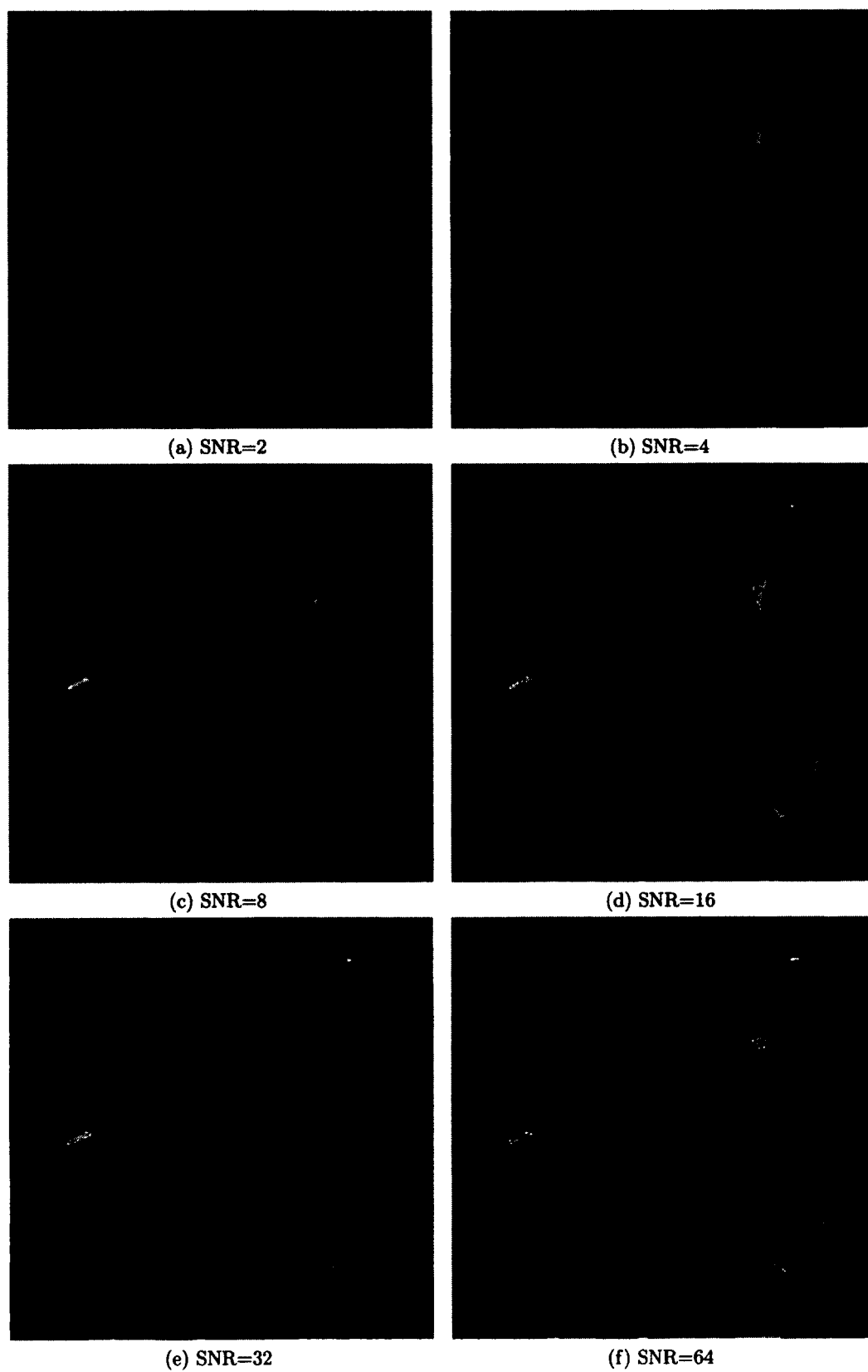


FIG. 45: Different partial-SR reconstructed outputs for various SNR

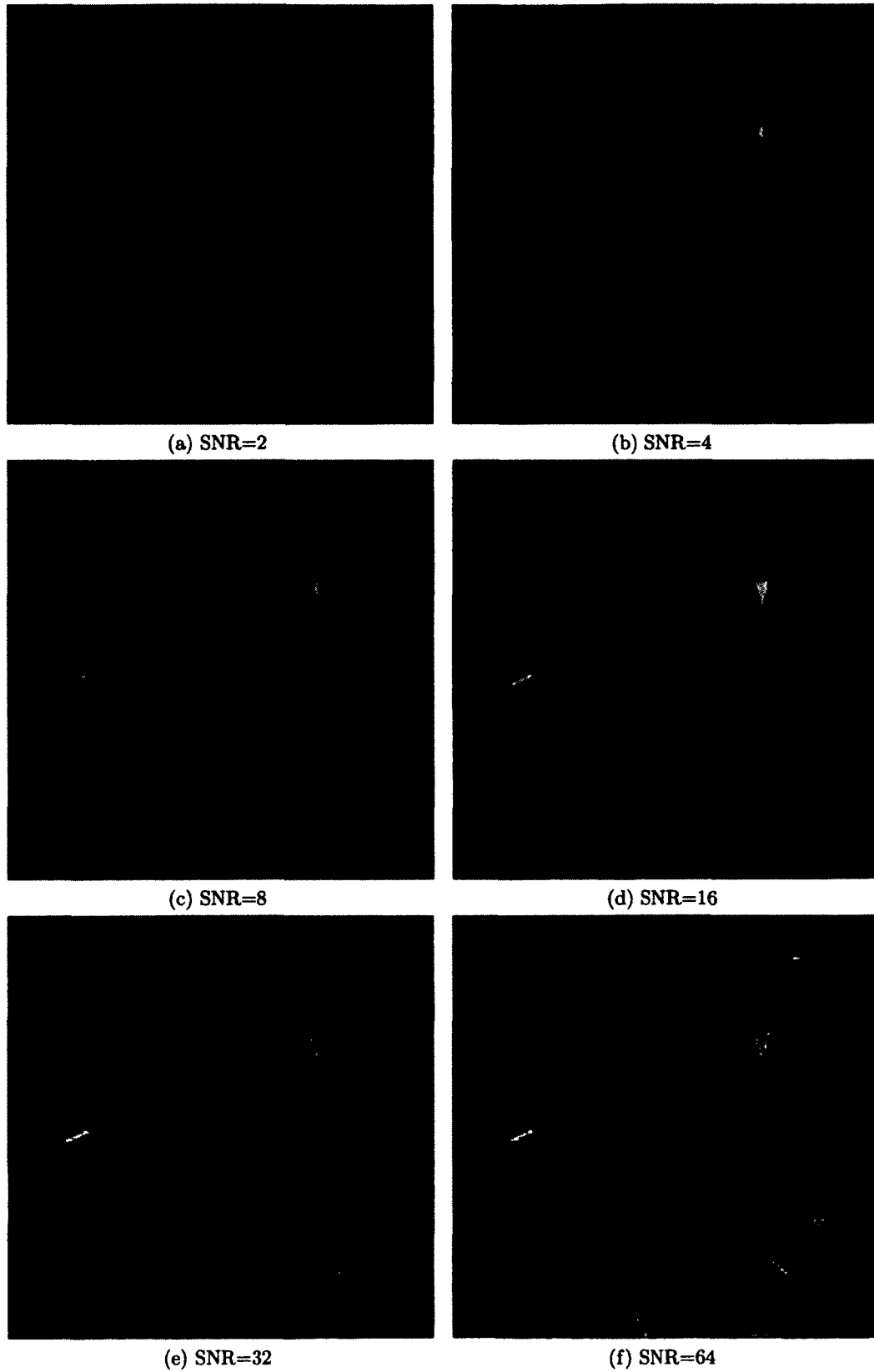


FIG. 46: Different partial-SR reconstructed outputs for various SNR

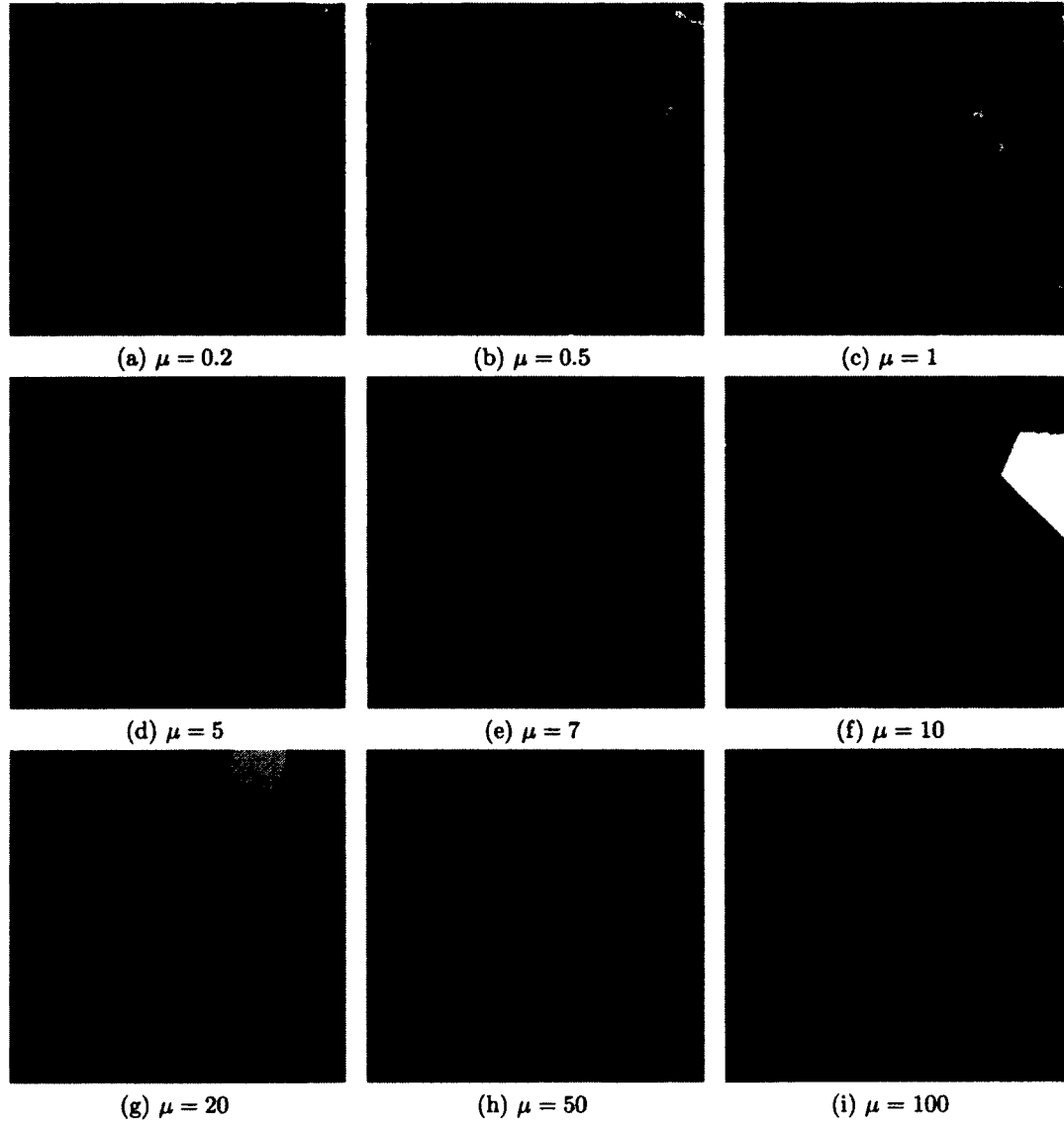


FIG. 47: Different full-SR reconstructed outputs for various mean spatial detail

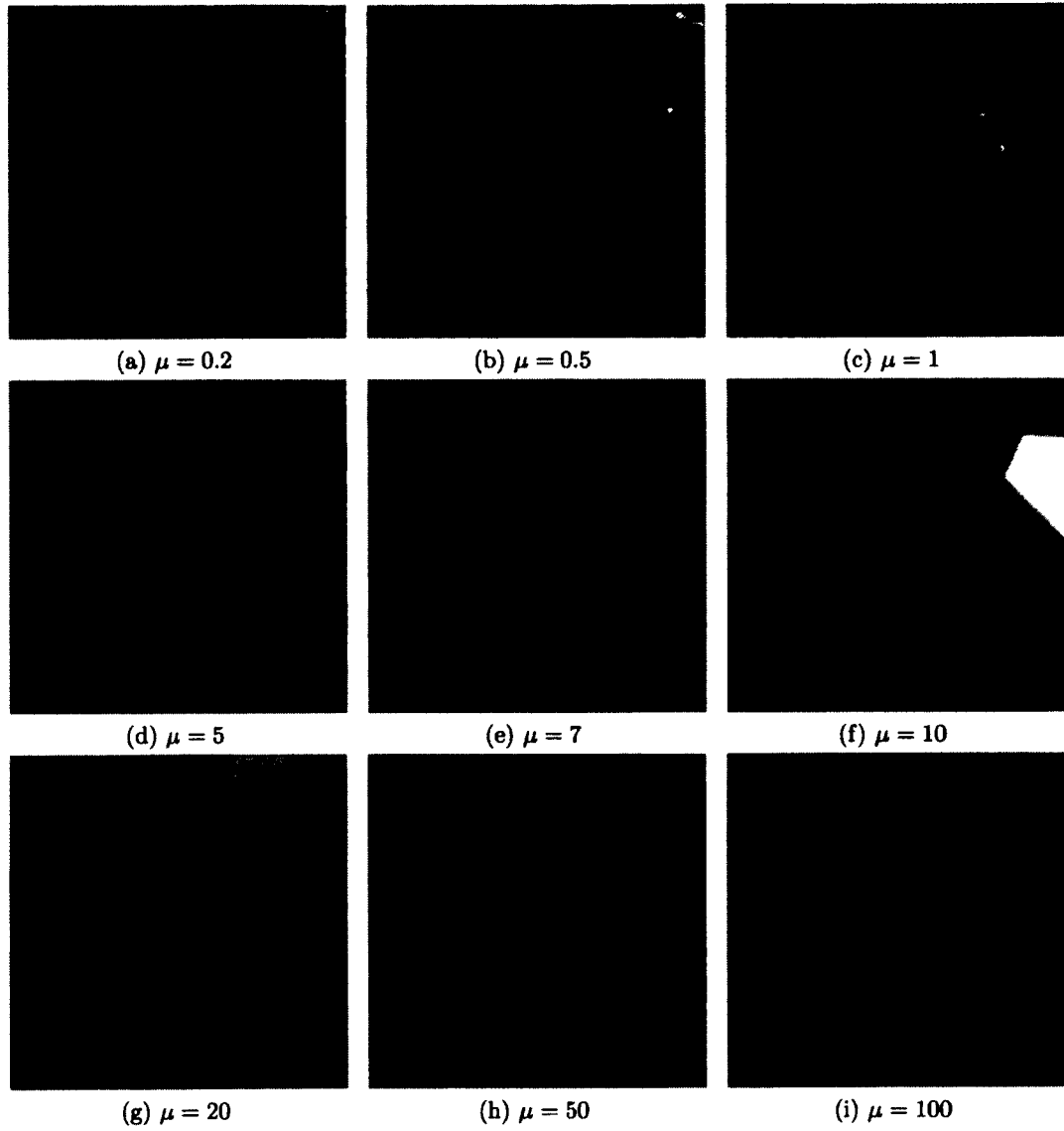


FIG. 48: Different full-SR reconstructed outputs for various mean spatial detail

CHAPTER 6

MULTI-RESPONSE SR RECONSTRUCTED IMAGE

6.1 INTRODUCTION

It is assumed that more than one optical sensor is mounted on the UAV. Every sensor acquires a set of degraded LR images that have subpixel shifts between them. The idea here is to incorporate the extra information obtained from the spatial response of different sensors to construct a single SR image that will contain more details than the case of using only one sensor. Figure 49 represents different image gathering spatial frequency responses. The vertical solid line designates the sampling passband of every sensor. The more the spatial response extends beyond the the sampling passband the more information is gathered by the sensor. Huck et al. [27] discuss the multi-response restoration in case of single image

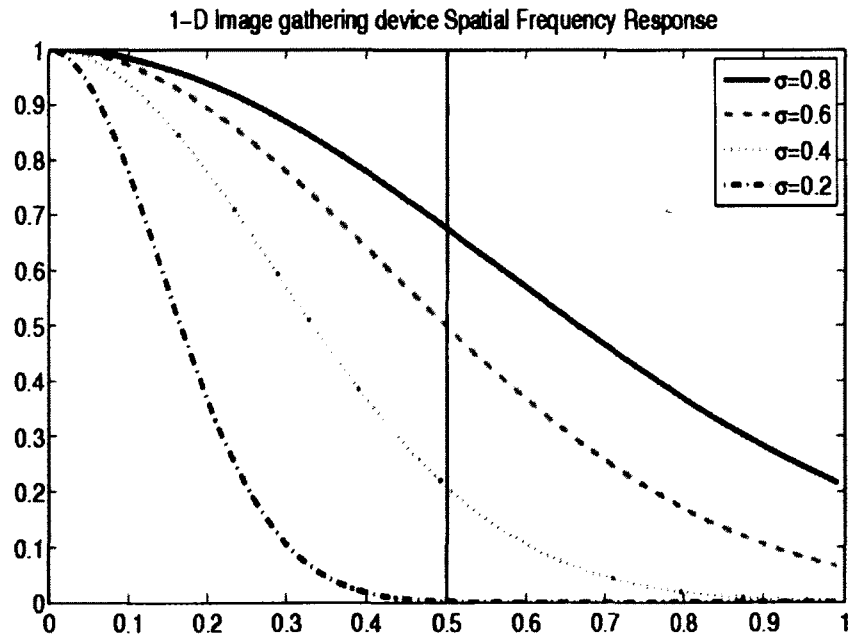


FIG. 49: Different spatial frequency responses

reconstruction and restoration without increasing its resolution. They assume that every LR image is captured with a different sensor, and they restore every single image and then

combine the processed images to get one restored image. Their approach is incomplete as they don't account for the subpixel shifts between the acquired frames. Accordingly, we extend their approach in the field of SR problems by assuming two cases. In both cases, every sensor captures a set of LR images which will be registered with respect to each other and with respect to the other sensors' LR images.

6.2 CASE 1: SINGLE WIENER RESTORATION FILTER (SWF)

The degraded LR images from all the available sensors will be mapped to a single HR grid and only one Wiener filter will be applied to reconstruct and restore a single SR image. Similar to the derivation given in Chapter 4, the Wiener filter for reconstruction and restoration of SR image is given by

$$\hat{\Psi}(\nu, \omega) = \frac{K_1 K_2 \hat{\Phi}_L(\nu, \omega) \hat{\tau}^*(\nu, \omega) \hat{\tau}_d^*(\nu, \omega) \exp(i4\pi(\nu x_{k_1} + \omega y_{k_2})) / |\hat{\tau}_d(\nu, \omega)|^2}{\hat{\Phi}_L(\nu, \omega) * \hat{\Pi}(\nu, \omega) \sum_{mn} \left| \sum_{k_1 k_2} \left| \hat{\tau}_{k_1 k_2}^*(\nu', \omega') \right|^2 \exp(-i2\pi(\nu' x_{k_1} + \omega' y_{k_2})) \right|^2 + \sum_{k_1 k_2} (\sigma_L / \sigma_{N_{k_1 k_2}})^{-2}} \quad (152)$$

6.3 CASE 2: WIENER FILTERS RESTORATION MATRIX (WFM)

In this case, we will consider groups of LR images that belongs to one sensor. Wiener filters matrix will be derived from these groups and every derived Wiener filter sub-matrix will be applied to its corresponding group to produce its corresponding SR image that will be combined with other SR images to reconstruct a single SR image, $\hat{R}_A(\nu, \omega)$, as given by

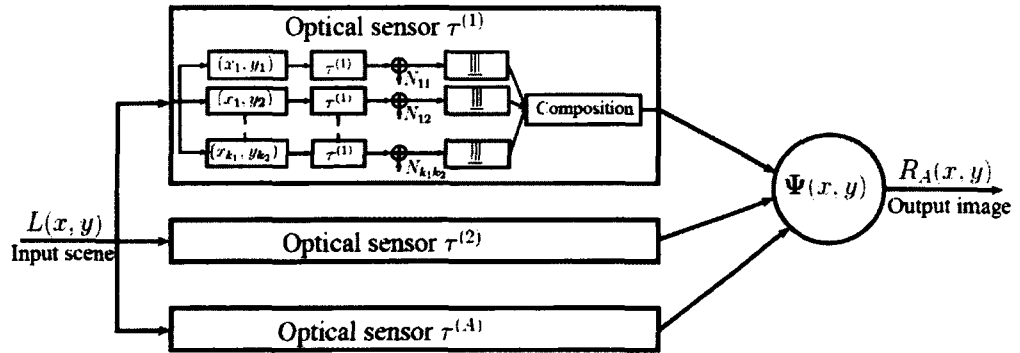


FIG. 50: Multi-response SR reconstruction and restoration

$$\hat{R}_A(\nu, \omega) = \frac{1}{A} \sum_{\alpha=1}^A \hat{\Psi}_\alpha(\nu, \omega) \tilde{S}_\alpha(\nu, \omega) \quad (153)$$

We will derive a closed form for Wiener filters matrix and compare the performance and the visual quality of the obtained SR images in terms of the produced information rate and the maximum realizable fidelity. For a given image gathering device characterized by an optical response index $\hat{\tau}^{(\alpha)}$, the acquired LR image for a given optical sensor $\tilde{s}_{k_1 k_2}^{(\alpha)}$ is given by

$$\begin{aligned} \tilde{s}_{k_1 k_2}^{(\alpha)} &= \sum_{mn} \left[\hat{L}(\nu - m, \omega - n) \hat{\tau}^{(\alpha)}(\nu - m, \omega - n) \exp(-i2\pi((\nu - m)x_{k_1}^{(\alpha)} \right. \\ &\quad \left. + (\omega - n)y_{k_2}^{(\alpha)}) + N_{k_1 k_2}^{(\alpha)}(\nu - m, \omega - n) \right] \end{aligned} \quad (154)$$

and its corresponding interlaced image \tilde{S}_α is given by

$$\tilde{S}_\alpha = \frac{1}{K_1 K_2} \sum_{k_1 k_2} \tilde{s}_{k_1 k_2}^{(\alpha)} \exp(-i2\pi(\nu x_{k_1}^{(\alpha)} + \omega y_{k_2}^{(\alpha)})) \quad (155)$$

6.4 WIENER FILTERS MATRIX DERIVATION

The MSRE e_A^2 between the input scene L and the reconstructed output R_A is given by

$$e_A^2 = E \left\{ \iint \left| \hat{L}(\nu, \omega) - \hat{R}_A(\nu, \omega) \right|^2 d\nu d\omega \right\} \quad (156)$$

which can be expanded as

$$\begin{aligned} e_A^2 &= \iint \left[\hat{\Phi}_L(\nu, \omega) - \sum_{\alpha=1}^A \hat{\Psi}_\alpha(\nu, \omega) E\{\tilde{s}_\alpha(\nu, \omega) \hat{L}^*(\nu, \omega)\} - \sum_{\alpha=1}^A \hat{\Psi}_\alpha^*(\nu, \omega) \right. \\ &\quad \left. \times E\{\tilde{s}_\alpha^*(\nu, \omega) \hat{L}(\nu, \omega)\} + \sum_{\alpha=1}^A \sum_{\beta=1}^A \hat{\Psi}_\alpha(\nu, \omega) \hat{\Psi}_\beta^*(\nu, \omega) E\{\tilde{s}_\alpha(\nu, \omega) \tilde{s}_\beta^*(\nu, \omega)\} \right] \end{aligned} \quad (157)$$

The optimal Wiener matrix filter must satisfy

$$\begin{aligned} \frac{\partial e^2}{\partial \Psi_\alpha} &= -E\{\tilde{s}_\alpha(\nu, \omega) \hat{L}^*(\nu, \omega)\} + \sum_{\beta=1}^A \hat{\Psi}_\beta^*(\nu, \omega) E\{\tilde{s}_\alpha(\nu, \omega) \tilde{s}_\beta^*(\nu, \omega)\} \\ &= 0 \end{aligned} \quad (158)$$

Accordingly, the Wiener filter matrix can be extracted from

$$\sum_{\beta=1}^A \hat{\Psi}_{\beta}(\nu, \omega) E\{\tilde{s}_{\alpha}^*(\nu, \omega) \tilde{s}_{\beta}(\nu, \omega)\} = E\{\tilde{s}_{\alpha}^*(\nu, \omega) \hat{L}(\nu, \omega)\} \quad (159)$$

For $\alpha = 1, \dots, A$, Equation (159) can be written as

$$\begin{bmatrix} E\{\tilde{s}_1^* \tilde{s}_1\} & E\{\tilde{s}_1^* \tilde{s}_2\} & \dots & E\{\tilde{s}_1^* \tilde{s}_A\} \\ E\{\tilde{s}_2^* \tilde{s}_1\} & E\{\tilde{s}_2^* \tilde{s}_2\} & \dots & E\{\tilde{s}_2^* \tilde{s}_A\} \\ \vdots & \vdots & \ddots & \vdots \\ E\{\tilde{s}_A^* \tilde{s}_1\} & E\{\tilde{s}_A^* \tilde{s}_2\} & \dots & E\{\tilde{s}_A^* \tilde{s}_A\} \end{bmatrix} \begin{bmatrix} \hat{\Psi}_1 \\ \hat{\Psi}_2 \\ \vdots \\ \hat{\Psi}_A \end{bmatrix} = \dots \begin{bmatrix} E\{\tilde{s}_1^* \hat{L}\} \\ E\{\tilde{s}_2^* \hat{L}\} \\ \vdots \\ E\{\tilde{s}_A^* \hat{L}\} \end{bmatrix} \quad (160)$$

where $E\{\tilde{s}_{\alpha}^* \hat{L}\}$ and $E\{\tilde{s}_{\alpha}^* \tilde{s}_{\beta}\}$ are given respectively by

$$E\{\tilde{s}_{\alpha}^* \hat{L}\} = \frac{1}{K_1 K_2} \sum_{k_1 k_2} \hat{\Phi}_L(\nu, \omega) \hat{\tau}_{k_1 k_2}^{(\alpha)*}(\nu, \omega) \exp(i4\pi(\nu x_{k_1}^{(\alpha)} + \omega y_{k_2}^{(\alpha)})) \quad (161)$$

and

$$\begin{aligned} E\{\tilde{s}_{\alpha}^* \tilde{s}_{\beta}\} &= \frac{1}{K_1^2 K_2^2} \left(\hat{\Phi}_L(\nu, \omega) * \hat{\Pi}(\nu, \omega) \sum_{mn} \sum_{k_1 k_2} \hat{\tau}^{(\alpha)*}(\nu - m, \omega - n) \right. \\ &\quad \times \hat{\tau}^{(\beta)}(\nu - m, \omega - n) \exp -i2\pi \left[(2\nu - m)(x_{k_1}^{(\beta)} - x_{k_1}^{(\alpha)}) \right. \\ &\quad \left. \left. + (2\omega - n)(y_{k_2}^{(\beta)} - y_{k_2}^{(\alpha)}) \right] + \sum_{k_1 k_2} \hat{\Phi}_{N_{k_1 k_2}}(\nu, \omega) * \hat{\Pi}(\nu, \omega) \delta(\alpha, \beta) \right) \end{aligned} \quad (162)$$

Equation (160) can be rearranged in a compact matrix notation as

$$\mathbf{E}_{s_{\alpha}^* s_{\beta}} \mathbf{\Psi} = \mathbf{E}_{s_{\alpha}^* L} \quad (163)$$

The Wiener filter matrix can be obtained through the least square solution given by the normal equation $\mathbf{\Psi} = [\mathbf{E}_{s_{\alpha}^* s_{\beta}}^T \mathbf{E}_{s_{\alpha}^* s_{\beta}}]^{-1} \mathbf{E}_{s_{\alpha}^* s_{\beta}}^T \mathbf{E}_{s_{\alpha}^* L}$. Once the Wiener filters matrix is obtained, Equation (153) is used to find the multi-response reconstructed and restored SR image.

Figure 52 compares the original HR input scene to the reconstructed outputs using single optical sensor, SWF, and WFM cases. It can be seen WFM case is most close to the original input scene and it has the highest visual quality compared to the other two cases. This is because WFM processes more information and it also reassembles the aliased components in the sampling passband, so it decreases the effect of aliasing and reduce the artifacts within

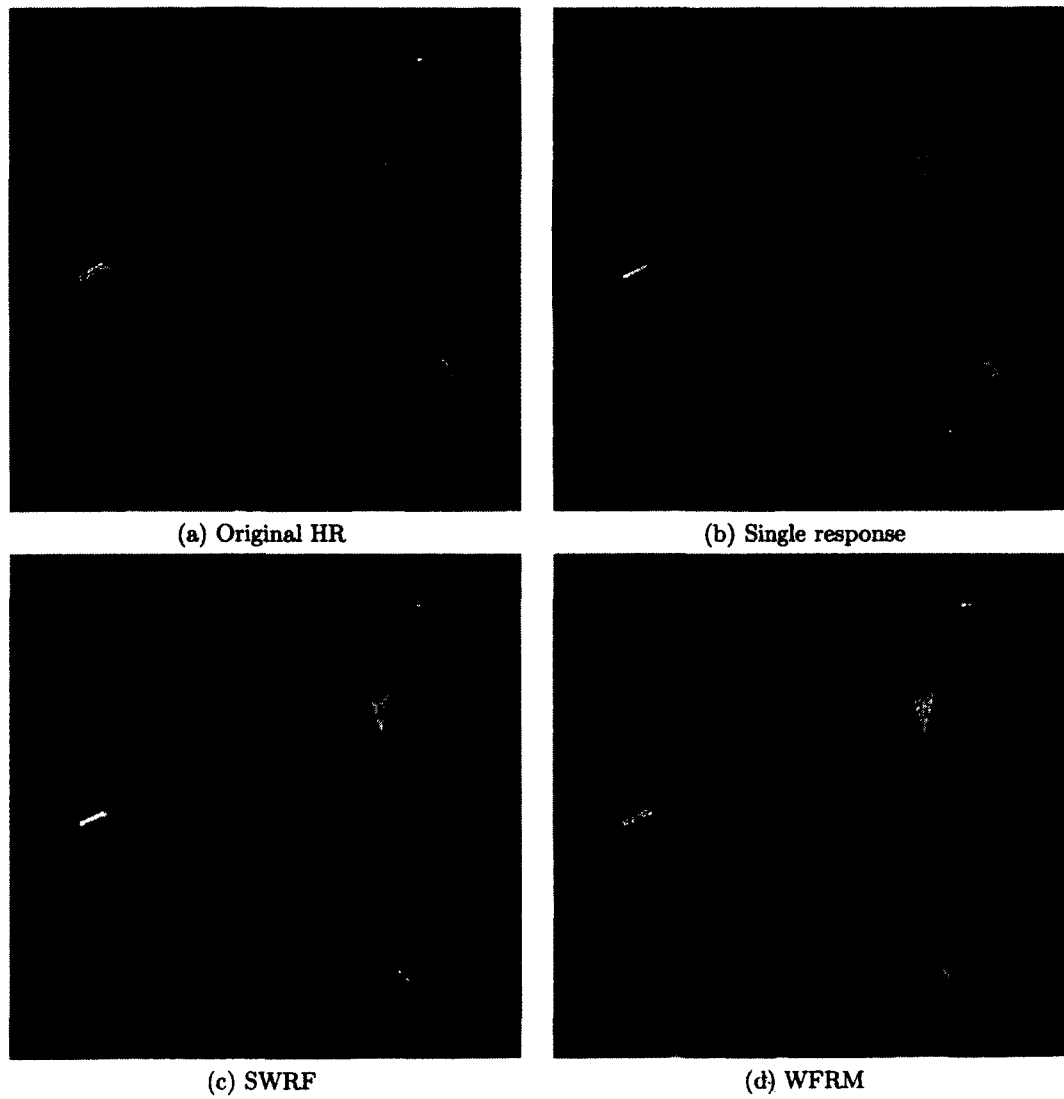


FIG. 51: Comparison between different multi-response outputs.

the reconstructed image. More outputs are considered in Figures 54, 55, and 56. A comparison between the three cases in terms of the fidelity and the information rate are given in Figure 52 and 53, respectively. It can be seen the highest fidelity and information rate are obtained when restoring the output using WFM, and the performance of SWF is better than restoring the output using a single optical response.

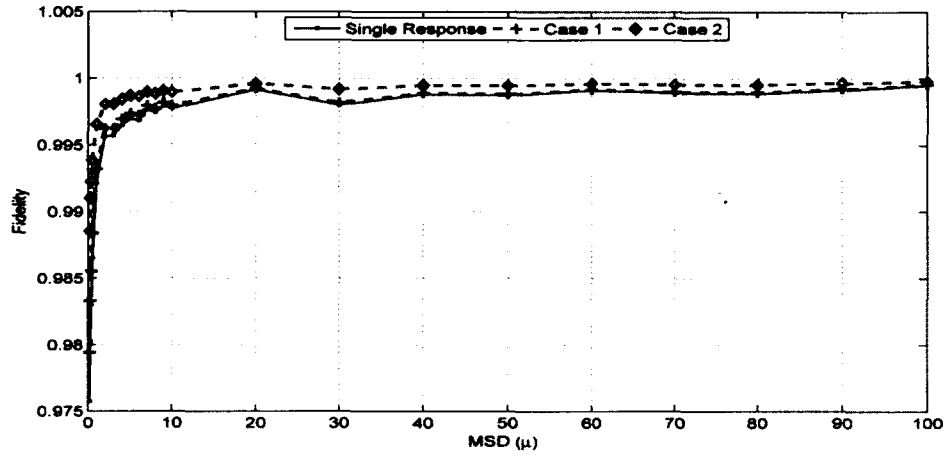


FIG. 52: Fidelity comparison between different multi-response cases

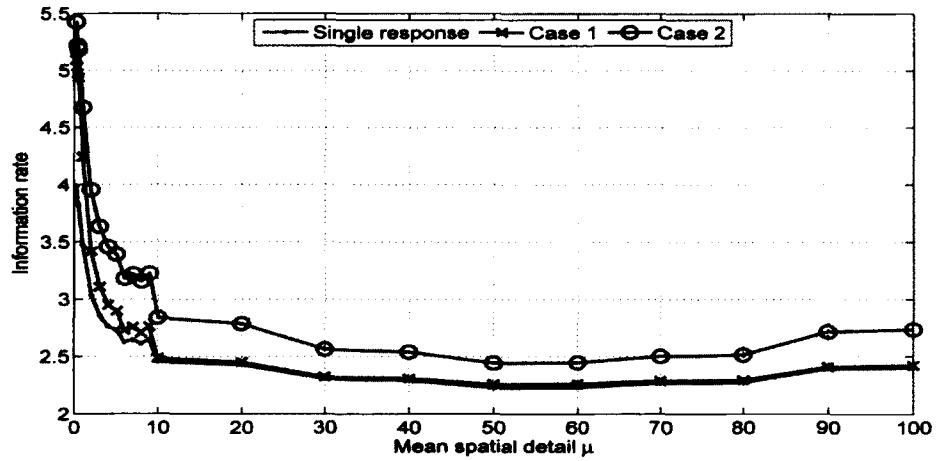


FIG. 53: Information rate comparison against mean spatial detail μ

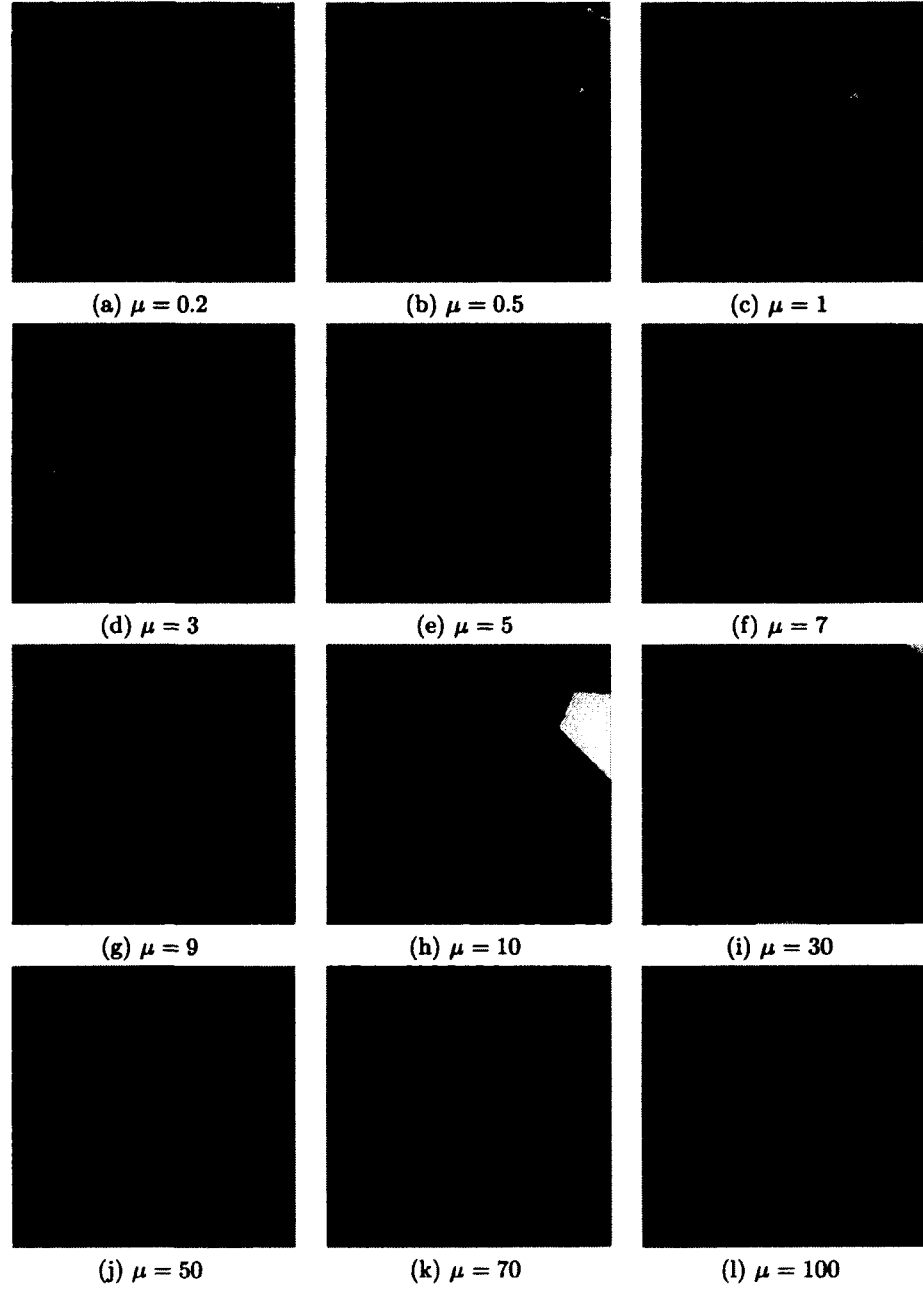


FIG. 54: Reconstructed single optical response versus different MSD

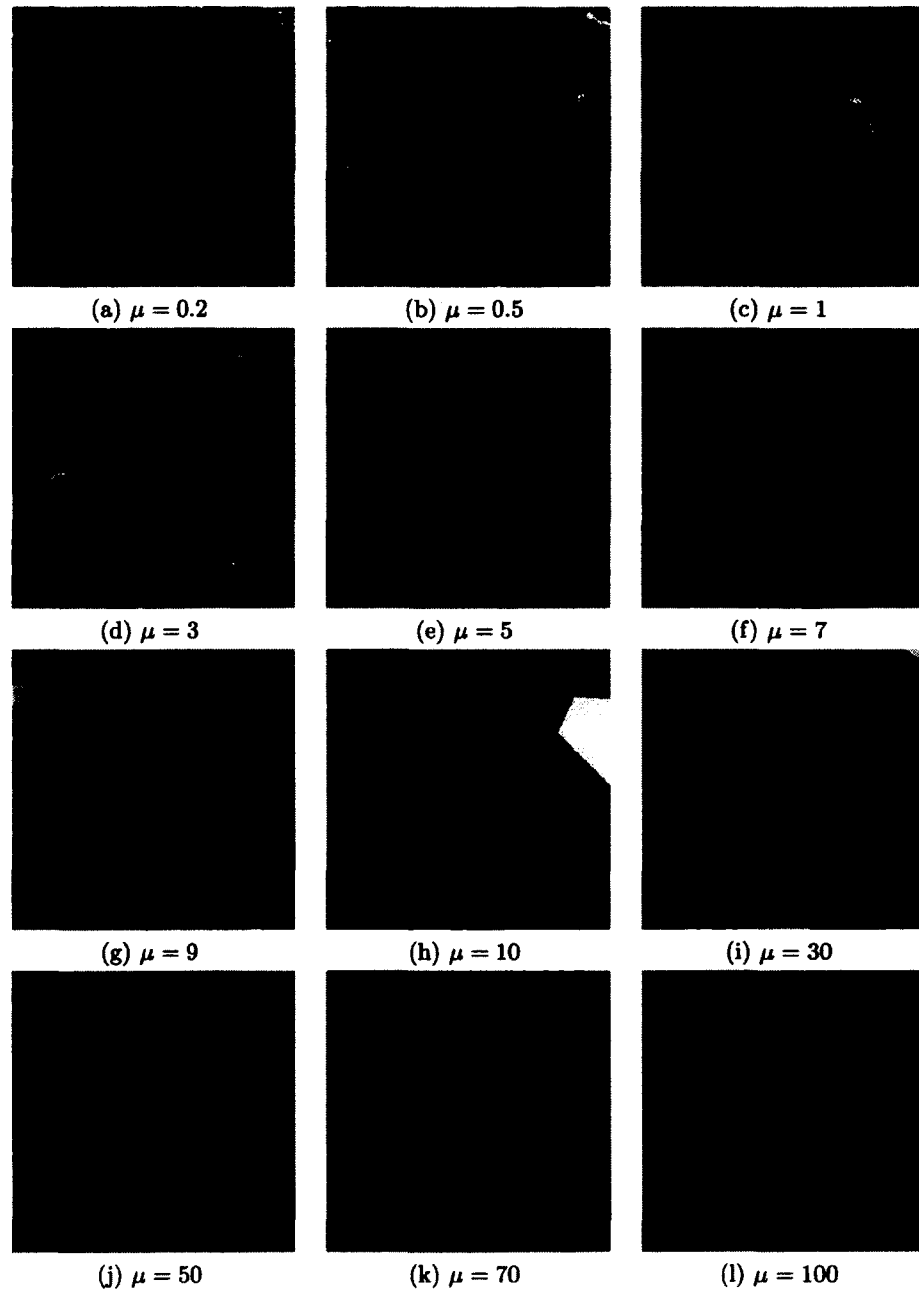


FIG. 55: Reconstructed SWF versus different MSD

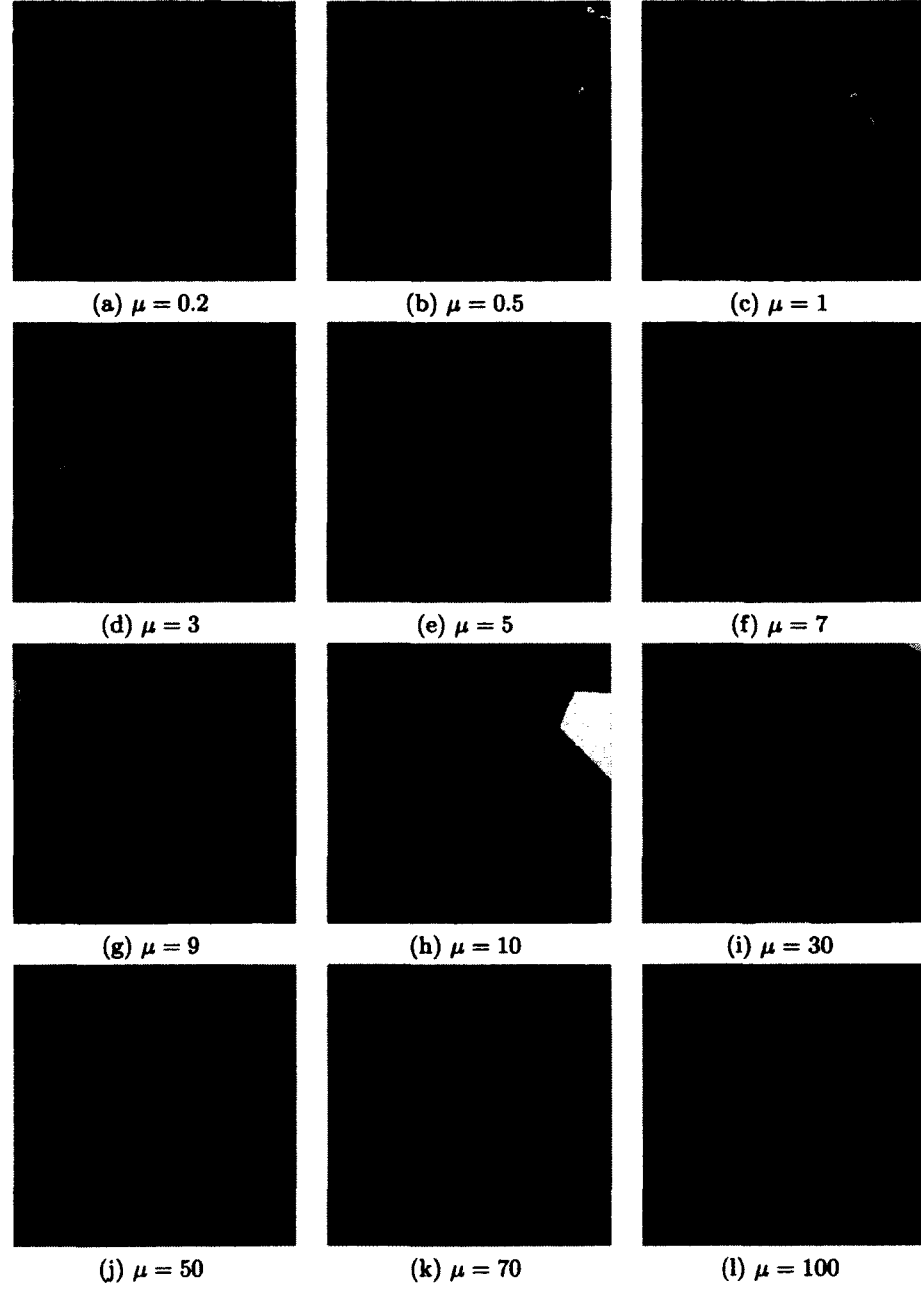


FIG. 56: Reconstructed WFM versus different MSD

CHAPTER 7

CONCLUSIONS AND FUTURE DIRECTIONS

7.1 CONCLUSIONS

In this dissertation, we developed a new stochastic algorithm to reconstruct and restore a single SR image from UAV captured images in two steps. The first one is for the accurate estimation of the subpixel shifts, and it is more flexible to cover any motion models and robust to moderate noise and lens optical blur. The other one is for the optimal reconstruction and restoration of the SR image that depends on the fully detailed SR CDC model, and it optimally restores the SR image with maximum end-to-end fidelity. The main problem of the previous research is that they work on a frame by frame basis to reconstruct the HR image and as a result, they don't exploit the available LR pixels spatial structure that can be used efficiently to produce a high quality image with minimum number of artifacts, especially when we have insufficient number of LR images. Also, their approaches depend on the DD model, which isn't sufficient to cover the acquisition and display degradations and can't be used to achieve optimal reconstruction and restoration with maximum fidelity between the original HR scene and the reconstructed SR one. Besides this, they have a blind estimation for the inter subpixel shifts between the acquired frames, which is not sufficient to cover complex registration models with relatively large motions.

The first step in our algorithm is the robust fast subpixel registration described in Chapter 2. The proposed approach is both immune to moderate noise and invariant to image features, lighting conditions and camera orientation. Because, the UAV captured images may have some acquisition distortions like shear and tilt that results the UAV rotational parameters such as encounters yaw, pitch, roll, or altitude change, we used the SIFT to correct for these distortions. Then, a more accurate registration algorithm was used to register the image with in subpixel accuracy. We enhanced one of the most efficient subpixel registration algorithms proposed by Guizar-Sicairos et al. [72]. Our enhancements are based on reducing both the dimensionality of the Fourier matrix of the cross-correlation matrix and the number of matrix multiplications required to find the accurate estimation of the subpixel shifts by

using a backward and forward search.

Before reconstructing the SR image, the registered LR pixel elements are optimally mapped to a uniform HR grid as described in chapter 3 in order to reduce the visual artifacts within the reconstructed image. The mapping is done with either the MSDA or the MSSA depending on the available number of LR images. With a sufficient number of LR frames, the MSDA method performs better than the MSSA method.

The second step described in Chapter 4 is to filter the composite image using new fast optimal stochastic minimum square restoration Wiener filter for SR reconstruction and restoration that basically depends on the fully detailed SR CDC model depicted in Figure 22. This filter is used to boost the visual quality of the SR image that is affected by degradations during the acquisition, such as blur due to system optics, aliasing due to insufficient sampling, photodetector noise, registration and fusion error, the number of scenels, and their relative arrangement on the high resolution grid. We revisited the Wiener filter that was developed by Carl et al. [19] and extended it for the general SR problem where the subpixel shifts are unknown and random. We introduced its mathematical derivation and analyze the end-to-end model and formulated the Wiener filter as a function of the parameters associated with the proposed SR system such as image gathering and image display response indices, system average signal-to-noise ratio (SNR) and inter-subpixel shifts between the LR images. In addition, we added a new parameter that accounts for LR images registration and fusion errors in order to improve the overall system performance. Also, we speeded-up the filter performance by constraining it to work on small patches of the images, and consequently, it can be implemented efficiently in the frequency domain. Also, we analyzed the loss in the end-to-end system fidelity and separated into three components and every component is related to its corresponding degradation in the proposed system, which enables a good design of the SR model.

As discussed in Chapter 5, we proposed two figures of merit: The information rate and the maximum realizable fidelity. Both are based on the information theory to asses the visual quality of the reconstructed images. The information rate is used to measure the amount of information produced by the imaging gathering device, while the maximum realizable

fidelity is used to measure the closeness between the reconstructed output and the original HR scene. It is shown that SR images can be reconstructed with both a high information rate and maximum realizable fidelity.

Finally, we discussed the reconstruction of multi-response SR images in Chapter 6 to enhance the visual quality of the reconstructed image. By assuming the existence of more than one optical sensor on the UAV with different optical response indices, we can balance and reduce the effect of both the blur and aliasing that exist in the reconstructed images with a single response. The reconstruction of SR images is utilized with either a single optimal Wiener filter or a matrix of optimal Wiener filters.

The research presented in this dissertation demonstrated that the derived Wiener filter with the optimal allocation of LR images can reduce aliasing and blurring, resulting in a sharper reconstructed image with pleasant visual quality. Throughout simulations and experimental results it is found that quantitative assessment using the proposed figures coincides with the visual qualitative assessment. We evaluated our filter against other SR techniques and its results were found to be competitive.

7.2 FUTURE WORK DIRECTIONS

The research in the area of SR reconstruction and restoration is both rich and challenging. The work explored in this dissertation is expected to lead to future research and considerations including the following avenues:

- The CDC model considered in this work is solely constrained by LSI blur and additive white Gaussian noise. So, the model can be revised to include other spatially variant blur models, such as out-of-focus or atmospheric turbulence blur. A blur identification stage may need to be included to find the best approximation for the blur operator that should be then fed back into the SR algorithm. In addition, the noise herein is modeled by an additive Gaussian random variable. The proposed model can be generalized to include other sources of noises such as impulsive noise and salt-and-pepper noise.
- SR techniques attempt to increase the spatial resolution of the imaging device by reconstructing HR images with sharp and crisp visual quality. Due to the smoothness effect of the stochastic Wiener filter, some of the image information at the edges and the sharp transitions areas of the image are lost and consequently the reconstructed images tend to be a bit blurry. So, it is desirable to combine the proposed Wiener filter approach with edge preserving techniques such as anisotropic diffusion in order to reconstruct high quality SR images with sharp edges and features. By including a term that depends on the differential structure of the image, the smoothness can be limited at edges in the reconstructed images.
- The performance of the proposed multi-response SR reconstruction and restoration is limited because of its computational time and high memory requirement which would limit its use in real-time applications. The proposed system can be accelerated by using possibly parallel and multi-thread computations. Each group of LR images can be processed independently and all the optical sensors can be synchronized so the reconstruction and the fusion of single SR image can be speeded-up.
- Shadows in the acquired LR images degrade the registration process and reduce significantly the quality of the reconstructed images. Techniques for shadow removal should be considered to eliminate it from LR images before applying the subpixel registration step.

- SR images have high resolution on behalf of the image size and the bandwidth consumed in transmission. So, it may be desirable to reconstruct SR compressed images without sacrificing the visual quality. The analysis of the information rate and the image entropy can be used to reduce the size of the reconstructed image using lossless compression and obtaining high compression ratios.

BIBLIOGRAPHY

- [1] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, 2003.
- [2] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1646–1658, 1997.
- [3] S. Borman and R. Stevenson, "Spatial resolution enhancement of low-resolution image sequences: A comprehensive review with directions for future research," tech. rep., Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana, USA, July 1998.
- [4] R. Alter-Gartenberg, C. Fales, F. Huck, Z.-U. Rahman, and S. Reichenbach, "Multiresponse imaging: Information and fidelity," *Multidimensional Systems and Signal Processing*, vol. 3, pp. 189–210, 1992.
- [5] R. Alter-Gartenberg, C. L. Fales, F. O. Huck, Z.-U. Rahman, and S. E. Reichenbach, "Multiresponse imaging system design for improved resolution," vol. 1605, pp. 745–756, SPIE, 1991.
- [6] M. K. Ng and N. K. Bose, "Mathematical analysis of super-resolution methodology," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 62–74, 2003.
- [7] J. Vanouwerkerk, "Image super-resolution survey," *Image and Vision Computing*, vol. 24, no. 10, pp. 1039–1052, 2006.
- [8] S. Chaudhuri and D. Taur, "High-resolution slow-motion sequencing: how to generate a slow-motion sequence from a bit stream," *IEEE Signal Processing Magazine*, vol. 22, pp. 16 – 24, March 2005.
- [9] J. Cui, Y. Wang, J. Huang, T. Tan, and Z. Sun, "An iris image synthesis method based on pca and super-resolution," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 4, pp. 471 – 474 Vol.4, August. 2004.

- [10] P. Cheeseman, B. Kanefsky, R. Hanson, and J. Stutz, "Super-resolved surface reconstruction from multiple images," *Maximum Entropy and Bayesian Methods*, no. FIA-94-12, pp. 293–308, 1996.
- [11] R. Peeters, P. Kornprobst, M. Nikolova, S. Sunaert, T. Vieville, G. Malandain, R. Deriche, O. Faugeras, M. Ng, and P. Van Hecke, "The use of superresolution techniques to reduce slice thickness in functional MRI," *International Journal of Imaging Systems and Technology (IJIST), Special Issue on High Resolution Image Reconstruction*, vol. 14, no. 3, pp. 131–138, 2004.
- [12] J. Shi, S. E. Reichenbach, and J. D. Howe, "Small-kernel superresolution methods for microscanning imaging systems," *Appl. Opt.*, vol. 45, no. 6, pp. 1203–1214, 2006.
- [13] M. K. Ng and A. C. Yau, "Super-resolution image restoration from blurred low-resolution images," *J. Math. Imaging Vis.*, vol. 23, no. 12, pp. 262–271, 2005.
- [14] Z.-U. Rahman, G. D. Hines, and M. J. Logan, "Detecting changes in terrain using unmanned aerial vehicles," vol. 5817, pp. 53–63, SPIE, 2005.
- [15] J. Fortin and P. C. Chevette, "Realization of a fast microscanning device for infrared focal plane arrays," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* (G. C. Holst, ed.), vol. 2743 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pp. 185–196, 1996.
- [16] A. H. Yousef, Z.-U. Rahman, and M. Karim, "On the restoration of the microscanned images captured from unmanned airborne vehicles," vol. 8056, p. 80560D, SPIE, 2011.
- [17] A. Yousef, J. Li, and M. Karim, "Mathematical model development of super-resolution image wiener restoration," *Journal of Optical Engineering* (Accepted, December 2011).
- [18] G. Holst, *CCD Arrays, Cameras, and Displays*. Spie P. Series, JCD Publishing, 1998.

- [19] C. L. Fales, F. O. Huck, J. A. McCormick, and S. K. Park, "Wiener restoration of sampled image data: end-to-end analysis," *J. Opt. Soc. Am. A*, vol. 5, no. 3, pp. 300–314, 1988.
- [20] M. Elad and A. Feuer, "Superresolution restoration of an image sequence: adaptive filtering approach," *IEEE Transactions on Image Processing*, vol. 8, pp. 387–395, mar 1999.
- [21] Q. Tian and M. N. Huhns, "Algorithms for subpixel registration," *Comput. Vision Graph. Image Process.*, vol. 35, pp. 220–233, August 1986.
- [22] E. Vera and S. Torres, "Subpixel accuracy analysis of phase correlation registration methods applied to aliased imagery," *EURASIP*, 2008.
- [23] H. Foroosh and M. Balci, "Sub-pixel registration and estimation of local shifts directly in the fourier domain," in *Image Processing, 2004. ICIP '04. 2004 International Conference on*, vol. 3, pp. 1915 – 1918 Vol. 3, October 2004.
- [24] H. Foroosh, J. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Transactions on Image Processing*, vol. 11, pp. 188–200, mar 2002.
- [25] S. Leprince, S. Barbot, F. Ayoub, and J.-P. Avouac, "Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, pp. 1529–1558, June 2007.
- [26] R. A. Reed, "Comparison of subpixel phase correlation methods for image registration," final report, Aerospace Testing Alliance, Arnold Air Force Base, Tennessee, USA, April 2010.
- [27] F. O. Huck, C. L. Fales, and Z.-U. Rahman, *Visual communications: An information theory approach*. Boston, Mass.: Kluwer Academic Publishers, 1st ed., 1997.
- [28] A. Katsaggelos, *Digital image restoration*. Springer series in information sciences, Springer-Verlag, 1991.
- [29] A. Bovik, ed., *Handbook of Image and Video Processing*, ch. Basic methods for image restoration and identification, pp. 167–173. Academic Press, 1st ed., 2000.

- [30] M. Alam, J. Bogner, R. Hardie, and B. Yasuda, "Infrared image registration and high-resolution reconstruction using multiple translationally shifted aliased video frames," *IEEE Transactions on Instrumentation and Measurement*, vol. 49, no. 5, pp. 915–923, 2000.
- [31] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Prentice Hall, 2nd ed., January 2002.
- [32] H. Ur and D. Gross, "Improved resolution from subpixel shifted pictures," *CVGIP: Graph. Models Image Process.*, vol. 54, pp. 181–186, March 1992.
- [33] T. Komatsu, K. Aizawa, T. Igarashi, and T. Saito, "Signal-processing based method for acquiring very high resolution images with multiple cameras and its theoretical analysis," *IEE Proceedings I Communications, Speech and Vision*, vol. 140, pp. 19–24, feb. 1993.
- [34] L. Landweber, "An iteration formula for fredholm integral equations of the first kind," *American Journal of Mathematics*, vol. 73, no. 3, pp. 615–624, 1951.
- [35] N. Shah and A. Zakhor, "Resolution enhancement of color video sequences," *IEEE Transactions on Image Processing*, vol. 8, no. 6, pp. 879–885, 1999.
- [36] N. Nguyen and P. Milanfar, "A wavelet-based interpolation-restoration method for superresolution (wavelet superresolution)," *Circuits, Systems, and Signal Processing*, vol. 19, pp. 321–338, 2000. 10.1007/BF01200891.
- [37] T. S. Huang, ed., *Advances in computer vision and image processing*, vol. 1, ch. Multiframe image restoration and registration. JAI Press, 1984.
- [38] S. Kim and W. Su, "Recursive high-resolution reconstruction of blurred multiframe images," *IEEE Transactions on Image Processing*, vol. 2, no. 4, pp. 534–539, 1993.
- [39] H. E. Sankaran, "Super-resolution using non-uniform to uniform resampling in spline spaces," master of science, Tamere University of Technology, Tampere, Finland, March 2007.
- [40] S. Kim, N. Bose, and H. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframe," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, no. 6, pp. 1013–1027, 1990.

- [41] N. Bose, H. Kim, and H. Valenzuela, "Recursive implementation of total least squares algorithm for image reconstruction from noisy, undersampled multiframe," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, 1993. ICASSP-93.*, vol. 5, pp. 269–272 vol.5, April 1993.
- [42] S. Rhee and M. G. Kang, "Discrete cosine transform based regularized high-resolution image reconstruction algorithm," *Optical Engineering*, vol. 38, no. 8, pp. 1348–1356, 1999.
- [43] K. Sauer and J. Allebach, "Iterative reconstruction of bandlimited images from nonuniformly spaced samples," *IEEE Transactions on Circuits and Systems*, vol. 34, no. 12, pp. 1497 – 1506, 1987.
- [44] S. Borman, *Topics in Multiframe Superresolution Restoration*. PhD thesis, University of Notre Dame, Notre Dame, Indiana., April 2004.
- [45] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," *J. Opt. Soc. Am. A*, vol. 6, pp. 1715–1726, November 1989.
- [46] A. Patti, M. Sezan, and A. Murat Tekalp, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Transactions on Image Processing*, vol. 6, no. 8, pp. 1064–1076, 1997.
- [47] A. Patti, M. I. Sezan, and A. M. Tekalp, "Super resolution video reconstruction with arbitrary sampling lattices and non-zero aperture time," *IEEE Transactions on Image Processing*, vol. 6, pp. 1064–1076, 1997.
- [48] A. Patti and Y. Altunbasak, "Artifact reduction for set theoretic super resolution image reconstruction with edge adaptive constraints and higher-order interpolants," *IEEE Transactions on Image Processing*, vol. 10, pp. 179–186, January 2001.
- [49] B. Tom and A. Katsaggelos, "An iterative algorithm for improving the resolution of video sequences," in *Proc. SPIE: Visual Communications and Image Processing*, vol. 2727, pp. 1430–1438, Mar. 1996.
- [50] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graph. Models Image Process.*, vol. 53, pp. 231–239, April 1991.

- [51] S. Mann and R. Picard, "Virtual bellows: constructing high quality stills from video," in *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference*, vol. 1, pp. 363–367 vol.1, November 1994.
- [52] M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion, and transparency," *Journal of Visual Communication and Image Representation*, vol. 4, pp. 324–335, 1993.
- [53] M.-C. Hong, M. G. Kang, and A. K. Katsaggelos, "Regularized multichannel restoration approach for globally optimal high-resolution video sequence," vol. 3024, pp. 1306–1316, SPIE, 1997.
- [54] M.-C. Hong, M. G. Kang, and A. K. Katsaggelos, "An iterative weighted regularized algorithm for improving the resolution of video sequences," in *Proc. 1997 Int. Conf. Image Processing*, pp. 474–477, 1997.
- [55] N. Bose, S. Lertrattanapanich, and J. Koo, "Advances in superresolution using l-curve," in *Circuits and Systems, 2001. ISCAS 2001. The 2001 IEEE International Symposium on*, vol. 2, pp. 433–436 vol. 2, May 2001.
- [56] P. C. Hansen and D. P. O'Leary, "The use of the l-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Comput.*, vol. 14, pp. 1487–1503, November 1993.
- [57] B. Tom and A. Katsaggelos, "Reconstruction of a high-resolution image by simultaneous registration, restoration, and interpolation of low-resolution images," in *International Conference on Image Processing*, vol. 2, pp. 539–542 vol.2, October 1995.
- [58] R. Schultz and R. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Transactions on Image Processing*, vol. 5, pp. 996–1011, June 1996.
- [59] R. Hardie, K. Barnard, and E. Armstrong, "Joint map registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Transactions on Image Processing*, vol. 6, pp. 1621–1633, December 1997.
- [60] M. Elad and A. Feuer, "Super-resolution reconstruction of an image," in *Nineteenth Convention of Electrical and Electronics Engineers in Israel*, pp. 391–394, 1996.

- [61] N. Nguyen, P. Milanfar, and G. Golub, "A computationally efficient superresolution image reconstruction algorithm," *IEEE Transactions on Image Processing*, vol. 10, pp. 573–583, Apr. 2001.
- [62] S. Farsiu, M. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image Processing*, vol. 13, pp. 1327–1344, October 2004.
- [63] M. Elad and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur," *Image Processing, IEEE Transactions on*, vol. 10, pp. 1187–1193, August 2001.
- [64] R. Hardie, "A fast image super-resolution algorithm using an adaptive wiener filter," *IEEE Transactions on Image Processing*, vol. 16, pp. 2953–2964, December 2007.
- [65] M. D. Pritt, "Image registration with use of the epipolar constraint for parallel projections," *J. Opt. Soc. Am. A*, vol. 10, pp. 2187–2192, October 1993.
- [66] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, pp. 91–110, November 2004.
- [67] R. Szeliski, *Computer Vision: Algorithms and Applications*. New York: Springer, 1st ed., 2010.
- [68] J. Koenderink, "The structure of images," *Biological Cybernetics*, vol. 50, pp. 363–370, 1984. 10.1007/BF00336961.
- [69] T. Lindeberg, "Scale-space theory: A basic tool for analysing structures at different scales," *Journal of Applied Statistics*, pp. 224–270, 1994.
- [70] M. Brown and D. Lowe, "Invariant features from interest point groups," in *In British Machine Vision Conference*, pp. 656–665, 2002.
- [71] A. Wade and F. Fitzke, "A fast, robust pattern recognition system for low light level image registration and its application to retinal imaging," *Opt. Express*, vol. 3, pp. 190–197, August 1998.
- [72] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," *Optics Letters*, vol. 33, pp. 156–158, 2008.

- [73] J. R. Fienup, "Invariant error metrics for image reconstruction," *Appl. Opt.*, vol. 36, pp. 8352–8357, November 1997.
- [74] P. Viol and W. M. W. III, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, p. 137154, 1997.
- [75] M. R. Idema, *Sub-Pixel Techniques to Improve Spatial Resolution*. PhD thesis, William & Mary College, Williamsburg, Virginia, December 2001.
- [76] X.-R. Wang and J.-Q. Zhang, "Quantitative characterization of the limiting resolution of a microscanning imager," *J. Opt. Soc. Am. A*, vol. 23, no. 8, pp. 1865–1869, 2006.
- [77] A. R. Webb, *Statistical Pattern Recognition*. Wiley, 2nd ed., 2002.
- [78] O. A. Omer and T. Tanaka, "Image superresolution based on locally adaptive mixed-norm," *JECE*, vol. 2010, pp. 8:1–8:7, January 2010.
- [79] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Robust Shift and Add Approach to Super-Resolution," in *Proc. of the 2003 SPIE Conf. on Applications of Digital Signal and Image Processing*, 2003.
- [80] S. E. Reichenbach and S. K. Park, "Small convolution kernels for high-fidelity image restoration," *IEEE Transactions on Signal Processing*, vol. 39, no. 10, pp. 2263–2274, 1991.
- [81] E. H. Linfoot, "Transmission factors and optical design," *J. Opt. Soc. Am.*, vol. 46, pp. 740–747, September 1956.
- [82] S. O. H. Schade, "Image gradation, graininess and sharpness in television and motion-picture systems," *J. Soc. Motion Pict. Telev. Eng.*, vol. 56, no. 2, pp. 137–174, 1951.
- [83] W. F. Schreiber, *Fundamentals of Electronic Imaging Systems*. Berlin: Springer-Verlag, 3rd ed., 1993.
- [84] Y. Itakura, S. Tsutsumi, and T. Takagi, "Statistical properties of the background noise for the atmospheric windows in the intermediate infrared region," *Infrared Physics*, vol. 14, no. 1, pp. 17–29, 1974.

- [85] P. Chatterjee, S. Mukherjee, S. Chaudhuri, and G. Seetharaman, "Application of papoulis-gerchberg method in image super-resolution and inpainting," *Comput. J.*, vol. 52, pp. 80–89, January 2009.
- [86] A. Zomet, A. Rav-Acha, and S. Peleg, "Robust super-resolution," *Proceedings international conference on computer vision and pattern recognition (CVPR)*, 2001.
- [87] T. Pham, L. van Vliet, and K. Schutte, "Robust fusion of irregularly sampled data using adaptive normalized convolution," *EURASIP Journal on Applied Signal Processing*, vol. 2006, no. 8, pp. 1–12, 2006.
- [88] P. Vandewalle, K. Krichane, and P. Zbinden, "Superresolution- graphical user interface." <http://lcav.epfl.ch/software/superresolution>, 2006. [Online; accessed 14 April, 2011].
- [89] C. Shannon, N. Petigara, and S. Seshasai, "A mathematical theory of," *Communication, Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.
- [90] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana: University of Illinois Press, 1st ed., 1964.
- [91] F. O. Huck, C. L. Fales, R. Alter-Gartenberg, S. K. Park, and Z.-U. Rahman, "Information-theoretic assessment of sampled imaging systems," *Optical Engineering*, vol. 38, no. 5, pp. 742–762, 1999.

VITA

Amr Yousef
 Department of Electrical and Computer Engineering
 Old Dominion University
 Norfolk, VA 23529

Born in Alexandria, Egypt, September 11, 1979. Entered Alexandria University in 1995. Graduated cum laude from Alexandria University with a Bachelor of Science in Electrical Engineering, Communications and Electronic Section. Joined the Engineering Mathematics Department with teaching assistantship in 2003. Obtained a Master of Science in Engineering Mathematics in 2006. Joined Old Dominion University with a research assistantship under the guidance of Dr. Zia Rahman. Supervised by Dr. Mohammad Karim after Dr. Zia Rahman passed away.

LIST OF PUBLICATIONS

Journal Papers

- Yousef, A. H., Li, J., Karim, M. Mathematical Model Development of Super-resolution Image Wiener Restoration, *Optical Engineering*. (Accepted, December 2011)
- Yousef, A. H., Li, J., Karim, M. High Speed Efficient Image Registration Algorithm with Subpixel Accuracy. *Optics Letters*. (Submitted, February 2012)

Conference Papers

- Yousef, A. H., Li, J., Karim, M. Toward Automatic Subpixel Registration of Unmanned Airborne Vehicle Images In: *Visual Information Processing XX on SPIE Defense, Security and Sensing Conference*, April 23 - 27, 2012. (To be Published, April 2012)
- Yousef, A. H., Li, J., Karim, M. Fast Stochastic Wiener Filter for Super-resolution Images Restoration with Theoretic Visual Quality Assessment In: *Visual Information Processing XX on SPIE Defense, Security and Sensing Conference*, April 23 - 27, 2012. (To be published, April 2012)
- Yousef, A. H., Li, J., Karim, M., On the visual quality enhancement of super-resolution images in *Applications of Digital Image Processing XXXIV*, edited by Andrew G. Tescher, Proceedings of SPIE Vol. 8135 (SPIE, Bellingham, WA 2011) 81350Z.
- Yousef, A. H., Rahman, Z., Karim, M., On the restoration of the microscanned images captured from unmanned airborne vehicles in *Visual Information Processing XX*, edited by Zia-ur Rahman, Stephen E. Reichenbach, Mark A. Neifeld, Proceedings of SPIE Vol. 8056 (SPIE, Bellingham, WA 2011) 80560D.
- Yousef, A. H., Rahman, Z., Super-resolution reconstruction of images captured from airborne unmanned vehicles in *Visual Information Processing XIX*, edited by Zia-ur Rahman, Stephen E. Reichenbach, Mark A. Neifeld, Proceedings of SPIE Vol. 7701 (SPIE, Bellingham, WA 2010) 77010H.