Enhancing Effective Depth-of-field by Multi-focus Image Fusion using Morphological Techniques

A Dissertation Presented by

Ishita De Ghosh

to

Indian Statistical Institute

in Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

November, 2011

To My Family

Abstract

A scene to be photographed, usually includes objects at varying distances from the camera. *Depth-of-field* of a digital camera is the *range of distance*, all objects within which appear to be sharp in the image. Due to the low depth-of-field of the camera, images acquired by them often suffer from degradation called out-of-focus blur. One way to enhance the effective depth-of-field is to acquire several images of a scene with focus on different parts of it and then combine these images into a single image in such a way that all regions of the scene are in focus. Acquired images are called *multi-focus images* and the process of combination is known as *multi-focus image fusion*. The techniques for multi-focus image fusion belong to the broad categories, pixel-based, block-based and region-based. They concentrate respectively on single pixels, small blocks of size $m \times n$ and arbitrarily shaped regions. Image registration is a necessary pre-requisite for multi-focus image fusion. The thesis presents a new technique for multi-focus image registration and three new techniques for multi-focus image fusion. Among these techniques, the first one is pixel-based, the second one is block-based and the third one is region-based. All of them use mathematical morphological tools. The pixel-based method is a multi-resolution technique that employs morphological wavelet as a tool for signal decomposition and reconstruction. The block-based method uses energy of morphological gradients as a tool for focus measure. Finally the region-based method uses multi-scale morphological tools for obtaining the focused regions from the input images. In this context, existing fusion techniques are studied and categorized. The thesis includes experimental results obtained by applying the proposed methods and other well-known methods on a variety of input data-set. It also includes performance analysis of various methods using standard quantitative evaluation techniques. At the end it presents concluding remarks and a discussion on related future work.

Acknowledgment

I would like to express my deep and sincere gratitude to my academic advisor, Professor Bhabatosh Chanda of ISI, Kolkata. Without his constant understanding, encouragement, and personal guidance, this work would not be possible.

I would like to thank all the members of the ECSU of ISI, Kolkata for their help during my stay in the ECSU laboratory.

I acknowledge the technical support and help provided by Sri Satrajit Ghosh and Dr. Bibhas C. Dhara during different phases of the work.

Special thanks to Dr. Ajoy K. Mukherjee, Principal, Barrackpore R. S. College, for his cooperation and all faculty-members and supporting staff of Department of Computer Science, Barrackpore Rastraguru Surendranath College, for their generosity and logistic support.

Finally, I express my gratitude to my family members including little Iman for their continuous support and encouragement.

(Ishita De Ghosh)

Contents

List of Figures

List of Tables

Abbreviations used in the thesis

Word(s)	Abbreviation
Depth-of-field	DOF
Energy of gradients	EOG
Energy of Laplacian	EOL
Energy of morphologic gradients	EOMG
Focus-measure	FM
Fusion quality index	FQI
Gradient similarity index	GSI
Multi-focus image fusion	MFIF
Mutual information	MI
Modified Laplacian	ML
Multi-resolution	MR
Multi-resolution decomposition	MRD
Multi-scale	MS
Multi-scale decomposition	MSD
Mean-square-error	MSE
Normalized-cross-correlation	NCC
Normalized difference in focus-measure	NDFM
Quad-tree	QT
Root-mean-square-error	RMSE
Structuring element	SE
Spatial frequency	SF
Sum modified Laplacian	SML
Structural similarity index	SSI

Chapter 1

Introduction

Advancements in digital imaging technology have increased the popularity of consumer imaging products such as digital cameras and camcorders. However due to the physical limitations of the imaging systems, images produced by them often suffer from degradations. A scene to be photographed usually includes objects at varying distances from the camera. Sharpness distribution of an image of such a scene is affected by various factors. The object focused by the camera and the objects at the same distance from the camera as the focused object appear to be the sharpest in the image. Sharpness of the objects in front of and behind the focused distance decreases gradually in the image. This sharpness-loss is not significant within a certain range of object distances. This range is called *depth-of-field (DOF)* of the camera [76]. DOF as calculated by par-axial geometric optics model of image formation using a thin convex lens is given in Appendix A of the thesis. DOF depends on various factors such as, the amount of sharpness-loss regarded as acceptable, focal-length of the lens (longer the focal-length, shorter the DOF), distance of the focused object (nearer the object, shorter the DOF) and the aperture used (decreasing the aperture will increase the DOF). The extreme case of decreasing the aperture for maximizing the DOF happens in a pin-hole camera. It has an infinite DOF. Unfortunately, the optical power in the image plane is reduced considerably due to very small aperture. So cameras

with finite DOF are preferred. A finite but large DOF means that objects within a large range (and hence possibly a large number of objects) will appear to be sharp in the photograph. On the other hand, a small DOF means that objects within a small range (and hence possibly a small number of objects) will appear to be sharp in the photograph and all other objects will appear to be out-of-focus in the photograph. Out-of-focus blur is one of the typical degradations which occur in images acquired by digital cameras due to their low DOF [60, 76]. The problem of low DOF is also encountered in microscopy due to increment in magnification and aperture [61, 6, 36].

1.1 Motivation

One way to enhance the effective DOF is to acquire several images of a scene focused on objects at different distances and then integrate these images into a single image in such a way that all regions of the scene are in focus. Acquired images are called *multi-focus images* and the process of combination is known as *multi-focus image fusion (MFIF)*. The process produces an image whose total area-in-focus is more than that of any of the constituent images. Multi-focus images of a scene are acquired one by one either by hand-held cameras or by cameras placed on tripods, in identical environmental conditions in respect to sensor, light, view-direction, orientation and object-contents in the scene. They can be either grey-level or color images. Since each image in a set of multi-focus images has focus on objects at different distances in the scene, an object which is in-focus in the near-focused image may be out-of-focus in other images. Similarly an object which is out-of-focus in the near-focused image may be in-focus in the far-focused image. Hence partial defocusing/blurring is inevitable in this type of images. MFIF produces an image in which blurred regions are deblurred and every area is in focus. The fused image should be better for human viewing as well as for subsequent processing and analysis like segmentation, feature extraction, object recognition etc. Figure 1.1 shows an example of multi-focus images with focus on complementary regions and the fused image with focus on all regions.

(a) Near focused image (b) Far focused image (c) Fused image

Figure 1.1: An example of multi-focus image fusion

The techniques for MFIF belong to the broad categories, pixel-based, block-based and region-based. They concentrate respectively on single pixels, small blocks of size $m \times n$ and arbitrarily shaped regions. It is interesting to study and compare MFIF techniques within a particular category and the ones belonging to different categories.

The ultimate goal of MFIF is to obtain all objects in the final image in focused and identifiable form. *Mathematical morphologic operators* have the capability of handling objects in different shapes and sizes. In this thesis, we explore mathematical morphology as a tool for MFIF and propose new techniques for the same employing this tool. We provide a comparison of results obtained by various techniques and outline some related future work. Image registration is a necessary pre-requisite for MFIF because before fusion the constituent images must be positioned properly with respect to a common coordinate system so that corresponding objects are overlaid properly [41]. We propose a new technique for multi-focus image registration also. In this chapter, a brief review of previous work on MFIF is given in Section 1.2, objective of the thesis is given in Section 1.3, a brief account of mathematical morphologic operators is given in Section 1.4, contribution of the thesis is given in Section 1.5, experimental set-up along with data used for experimentation are given in Section 1.6, evaluation techniques used are given in Section 1.7, and finally organization of the thesis is given in Section 1.8.

1.2 Review of previous work

The fundamental concept behind MFIF is to select the sharply focused regions from the input images to form an image in which all objects are in focus. The basic steps for this are, to divide each input image into overlapping or non-overlapping regions, then measure sharpness of focus for all regions, finally select the best-focused region among all corresponding regions to form the fused image. When all the regions of interest squeeze into single pixels, the approach is called a pixel-based approach, when they are small blocks of size $m \times n$, the approach is called a block-based approach; otherwise it is called a region-based approach. Another categorization is done on whether the technique is based on spatial domain or frequency domain. In spatial domain techniques, input images are fused in spatial domain using physically relevant spatial features. In frequency domain techniques, multi-scale decomposition (MSD) or multi-resolution decomposition (MRD) by pyramid or wavelet transform is required.

An early categorization of frequency domain MRD fusion schemes was given by Zhang and Blum [89]. Piella [65] provided a general framework for these schemes and also proposed a new method for the same. Pajares and Cruz [59] presented a comprehensive tutorial on wavelet-based fusion methods. Goshtasby and Nikolov [30] presented an overview of various fusion techniques. Basic idea of MRD-based fusion schemes is the following. At first each source image is transformed/decomposed up to a level by an MRD scheme. The decomposition gives the scaled image as low frequency coefficients and the detail images as high frequency coefficients. Saliencies of the coefficients are measured by their activity-levels. A selection or decision map is created from the activity-levels of the coefficients from all transformed images. The map is used as a guide to construct the composite representation of the transformed images. Finally fused image is obtained by applying the inverse transform to the composite representation. An MRD fusion scheme is categorized depending on how the activity-levels of MRD-coefficients are measured. If the activity-level is measured for the coefficients related to individual pixels, the method is called pixel-based, if it is measured depending on the coefficients of a small block surrounding the concerned

coefficient, the method is called block-based and finally; if it is measured depending on all coefficients in a region containing the concerned coefficient, then the method is called region-based. Images fused by frequency-domain MRD schemes may lose some information of the source images because of implementation of inverse multi-resolution transform.

The idea of using MRD schemes for image fusion was first proposed by Burt [8] as a model for binocular fusion for human stereo vision. He used *Laplacian pyramid* for MRD and *choose max* rule for coefficient selection. Burt and Adelson [10] later introduced a new approach to image fusion based on *hierarchical image decomposition*. Adelson [2] then used the *Laplacian pyramid* technique for MFIF. Toet [78] proposed the use of *ratio of low-pass pyramids* at successive levels of Gaussian pyramids for fusion of visible and IR images. Burt [9] proposed that fusion within a *gradient pyramid* provides improved stability and noise immunity. Akerman [3] optimized the Laplacian pyramid fusion in respect of multi-sensor fusion. Burt and Kolczynski [11] presented *gradient pyramid fusion* with a local match measure and a window-based saliency measure. Li *et al.* [45] used similar method except that *wavelet transform* is used instead of pyramid transform and consistency verification is done along with window-based activity measure. Wavelet based fusion techniques are proposed later by many other people including Chipman *et al.* [16], Petrovic and Xydeas [63], Scheunders [72], Hill *et al.* [37], Hamza *et al.* [34], De and Chanda [19], Qu and Yan [68] and Lewis *et al.* [44]. Frequency domain techniques in various categories will be discussed in detail in related chapters.

Since multi-focus images of a scene are acquired with focus on complementary regions, focused regions in an image have more contrast than their defocused counter-parts in other images. *Focus-measure (FM)* is a quantity for evaluating the contrast or sharpness of a pixel, block or region [39, 50]. Image variance, image gradients, image Laplacians, energy of image gradients (EOG), energy of image Laplacian (EOL) are traditional FM's employed and validated for applications like autofocusing [76]. Modified Laplacian (ML), Sum modified Laplacian (SML) are modifications of image

Laplacian [58]. Spatial frequency (SF) [25] and Tenengrad [36] were later introduced as focus measures. In spatial domain MFIF techniques, input images are fused in the spatial domain using focus-measure as a physically relevant spatial feature in localized area. Since these techniques emphasize on a specific or desired image area, very little/no change occurs in other areas. Pixel-level weighted averaging is a spatial domain technique in which fusion is done by taking the weighted average of the pixel intensities of the input images. Weights are determined by tools like principle component analysis [71] or adaptive methods [42]. Other spatial domain pixel-level image fusion approaches include, fusion using controllable camera [73], probabilistic methods [5], image gradient method with majority filtering [23].

The basic idea in spatial domain block based fusion methods is to divide the input images into a number of blocks, then measure focus on corresponding blocks and finally select and combine the focused blocks to create the fused image [39]. Often consistency verification is done before creating the final fused image. Spatial domain block based fusion methods are proposed in [47, 48, 55, 29, 27, 87, 21]. Li *et al.* [47] used spatial frequency (SF) as the focus measure. In a subsequent work they [48] used neural network (NN) to select better focused blocks using three features *SF, visibility* and *edge feature*. Miao and Wang [55] used energy of image gradients (EOG) to measure focus in image blocks in an MFIF algorithm based on Pulse Coupled Neural Networks. In the method of Goshtasby [29], focus is measured by the sum of the gradient values of all pixels in the block. In the method of Fedorov *et al.* [27] each image is tiled with overlapping neighborhoods. For each region the tile that corresponds to the best focus (which is measured by ML) is chosen. Zhang and Ge [87] proposed a technique in which focused blocks are detected by measuring their blurriness. De and Chanda [21] introduced a new focus measure called energy of morphologic gradients (EOMG) and used it for image fusion in a block-based MFIF algorithm.

In region-based fusion techniques, among corresponding regions better focused ones are selected and combined to construct the fused image. So block-size is not of any

concern in these methods. Spatial-domain region-based fusion methods are proposed in [53, 57, 22, 49]. Methods described in [53, 57, 22] use multi-scale morphology. Matsopoulos *et al.* [53] used multi-scale morphologic pyramids. Mukhopadhyay and Chanda [57] used morphologic towers instead of morphologic pyramids. De *et al.* [22] proposed multi-focus image fusion techniques using multi-scale top-hat transformation. Li and Yang's technique [49] is a spatial-domain, region-based technique which does not depend on MRD. In this technique, input images are segmented according to the segmentation results of a temporary fused image and better focused regions are selected and stitched to their desired positions to get the final fused image. Spatial domain techniques in various categories will be discussed in detail in related chapters. Objective of the thesis is given now.

1.3 Objective of the thesis

A number of researchers have suggested methods for MFIF as a solution to the problem of low depth-of-field. As discussed before, the techniques belong to the broad categories, pixel-based, block-based and region-based. It is interesting to study and compare MFIF techniques within a particular category and the techniques belonging to different categories. A good algorithm for MFIF should possess some important properties. It should be independent of image content and robust against probable misalignments of input images. It should not produce any unwanted visual effect or artifact. Quality of the fused image should satisfy the requirement for intended application and finally computational complexity should also be affordable. In general, pixel-based techniques are intuitively straightforward, easy to implement and computationally efficient. But they are sensitive to mis-registration of input images. Block-based and region-based techniques are more robust in respect of registration problems though they are more complex in general. Despite the increase in complexity, region-based methods have a number of advantages over pixel-based methods. These include the ability to use more intelligent semantic fusion rules and the ability

to attenuate or accentuate certain properties to the regions [30].

Multi-focus images may contain objects of different shapes and sizes. The ultimate goal of MFIF is to obtain all objects in the final image in focused/deblurred form. Mathematical morphology is a subject which treats an image as an ensemble of sets. Morphologic operators have the capability of handling objects in different shapes and sizes. They have some interesting computational advantages as well. In this thesis we explore morphologic techniques as a tool for MFIF. Algorithms for MFIF proposed in the thesis employ various combinations of morphologic operations.

Given this, the objective of the thesis is to propose and analyze grey-level MFIF schemes employing morphologic operators and having the following desirable properties,

- ability to work on a variety of input images,
- robustness against probable mis-registration of input images,
- extensibility to fuse multi-focus color images,
- low computational cost,
- adaptability to hardware implementation.

Since all algorithms for MFIF proposed in this thesis use morphologic operators, a brief introduction to them is given now.

1.4 Morphologic Operators

Mathematical morphology treats an image as a set of pixels [74, 75]. Morphologic operators work with two sets, the original image to be analyzed and a structuring element (SE). Each SE has a shape and a size and it can be thought of as a parameter

to the operation. Fundamental morphologic operations are morphologic dilation and morphologic erosion. At first we present these two operators for binary images. A two-dimensional binary image signal is a function/mapping from domain D (which is a subset of discrete two-dimensional Euclidean space Z^2) to a binary-set $\{0, 1\}$. Suppose A is the set of points representing the binary-1 pixels of the original binary image and B is the set of points representing binary-1 pixels of the SE. Then *dilation* and *erosion* of A by B, are denoted by $A \oplus B$ and $A \ominus B$ respectively and are defined as

$$A \oplus B = \{b + a \mid \text{for } b \in B \text{ and } a \in A\} \tag{1.1}$$

$$A \ominus B = \{p \mid b + p \in A \text{ for every } b \in B\} \tag{1.2}$$

where '+' denotes the *binary-or* operation. Practically, $A \oplus B$ is the locus of origin of B such that B hits A. Similarly, $A \ominus B$ is the locus of origin of B such that B fits in A.

We now consider the case of grey-scale images. A two-dimensional grey-scale image signal X is a function/mapping from domain D (which is a subset of discrete two-dimensional space Z^2) to the set of grey intensity values $\{g_1, g_2, \ldots, g_n\}$ where each g_i is a nonnegative integer. A grey-scale SE h is a mapping from its domain to the above set of grey values. In this thesis, we use flat SE's that is SE's for which the value of h is always zero. Let (r, c) be a point in domain D, where r and c denote the row and column coordinates respectively. *Dilation* and *erosion* of $X(r, c)$ by $h(r, c)$ are denoted by $(X \oplus h)(r, c)$ and $(X \ominus h)(r, c)$ respectively and are defined as

$$(X \oplus h)(r, c) = \max_{(i,j) \in Domain\ of\ h} (X(r - i, c - j) + h(i, j)) \tag{1.3}$$

$$(X \ominus h)(r, c) = \min_{(i,j) \in Domain\ of\ h} (X(r + i, c + j) - h(i, j)) \tag{1.4}$$

where the maximum and minimum are taken over all (i, j) in the domain of h such that $(r - i, c - j)$ and $(r + i, c + j)$ are in the domain of X. So dilation simply replaces the value at each point of X by the maximum value in the neighborhood defined by the SE when the origin of SE is placed at the point. Similarly erosion replaces the value at each point of X by the minimum value in the neighborhood defined by the

SE when the origin of SE is placed at the point. Other morphologic operators are constructed by combining dilation and erosion. For example, *opening* and *closing* of $X(r,c)$ by $h(r,c)$ are denoted by $(X \circ h)(r,c)$ and $(X \bullet h)(r,c)$ respectively and are defined as

$$(X \circ h)(r,c) = ((X \ominus h) \oplus h)(r,c) \tag{1.5}$$

$$(X \bullet h)(r,c) = ((X \oplus h) \ominus h)(r,c) \tag{1.6}$$

Both opening and closing are increasing operations implying that opening (closing) of an image contains openings (closings) of all its sub-images. Both opening and closing are idempotent operations implying that successive applications of openings (closings) do not further modify the image. Finally, opening is an anti-extensive operation and closing is an extensive operation. In a grey-scale image X, an opening removes all foreground structures in the image that are not large enough to contain the SE. Similarly, a closing removes all background structures in the image that are not large enough to contain the SE. Here foreground structure means an image region of intensity value higher than the surrounding region.

1.4.1 Multi-scale morphologic operators

Extraction of features by mathematical morphology depends on effective use of SE's. Sizes and shapes of SE's play crucial roles here. A morphologic operator with a scalable SE can extract features of various shapes and sizes. A scheme of morphologic operations with a scalable SE is termed as *multi-scale morphology* [15, 52]. For a scalable SE h, size of its domain gets changed. Let B be a set representing the domain of h. Assume that B has a definite shape. Let n be an integer representing the scale-factor of B and let nB denote the scaled version of B at scale n. If B is convex, then nB is obtained by $n-1$ dilations of B by itself.

$$nB = B \underbrace{\oplus B \oplus B \oplus \cdots \oplus B}_{n-1 \text{ times}} \tag{1.7}$$

When $n = 0$, conventionally B is taken to be a disk of unit size so that $nB = \{(0,0)\}$. Let h be a flat-top SE such that its value at every point in its domain nB is zero. Then

a morphologic operation by SE h reduces to an operation by its domain nB. Then multi-scale opening and closing of X by scalable domain nB are defined respectively as

$$(X \circ nB)(r,c) = ((X \ominus nB) \oplus nB)(r,c) \quad (1.8)$$

$$(X \bullet nB)(r,c) = ((X \oplus nB) \ominus nB)(r,c) \quad (1.9)$$

The opening removes all bright/foreground structures in the image X that are not large enough to contain nB. Here foreground structure means an image region of intensity value higher than the surrounding region. Similarly, the closing removes all dark/background structures in the image X that are not large enough to contain nB. These operators are used effectively to detect focused regions which in general have more contrast than corresponding defocused regions.

Given the background and the objective of the thesis and a short introduction to morphologic operators, contribution of the thesis is presented now.

1.5 Contribution of the thesis

It is already discussed that the objective of the thesis is to propose and analyze grey-level MFIF schemes having certain desirable properties. Mathematical morphology is explored as a tool for MFIF and new techniques are presented employing this tool. In addition to a brief review of previous work, the objective of the thesis and a short introduction to morphologic operators, current chapter, viz. Chapter 1 includes the data-set used for experimentation purpose and the quantitative measures used for performance evaluation.

Since registration is a necessary prerequisite for MFIF, a new algorithm for multi-focus image registration is presented in Chapter 2. It is an iterative algorithm for registration of multi-focus images by combining global and local transformation models. In the first step of the algorithm, a global translation is determined by maximizing

the mutual information of the source and the reference images and then it is applied on the source image. In the second step, a block-wise local scaling is applied on the translated source image. The scale-factors are determined by maximizing a similarity measure of two corresponding blocks of the translated source image and the reference image. The global and local transformations constitute a hybrid technique which is iterated to obtain the optimal result. The proposed method is automatic, easy to implement and gives good results. Results obtained by applying the method on different sets of multi-focus images are provided with. Performance of the system is evaluated and is compared with a widely used method.

Chapter 3 presents a pixel-based algorithm for multi-focus image fusion using morphologic wavelets. A *nonlinear morphologic wavelet transform* which preserves the range in the scaled images and involves integer arithmetic only is introduced at first. This transform is employed in a fusion algorithm to fuse a set of grey-scale multi-focus images. The method is computationally efficient and produces good results. Integrated-chip implementations of image processing algorithms are going to become more common in near future. Our method will be useful in this respect. The problem with this algorithm is that being a pixel-based method, it is not robust to mis-registration problem.

Chapter 4 presents a block-based algorithm for multi-focus image fusion using a morphology-based focus measure in a quad-tree structure. Focus-measure is a quantity for evaluating the contrast or sharpness of a pixel, block or region. A new focus-measure called *energy of morphologic gradients (EOMG)* is introduced. It is used for a novel algorithm for MFIF which employs a quad-tree structure for optimal subdivision of input images while selecting the sharply focused blocks. Though the algorithm starts with blocks, it ultimately identifies sharply focused regions in input images. The focus measure EOMG is comparable with other focus measures viz. energy of gradients (EOG) and variance. The algorithm is robust in the sense that it works with any focus measure. It is also robust against pixel mis-registration. But as the algorithm perceives an image as a union of variable-sized blocks, blocking effects

may appear in the boundaries of arbitrary-shaped regions.

Chapter 5 presents a region-based algorithm for multi-focus image fusion using multi-scale morphology. Since multi-focus images of a scene are acquired with focus on the complementary regions, focused regions in an image have more contrast than their de-focused counter-parts in other images. This implies that the focused regions contain larger number of physically relevant features than that contained in corresponding defocused regions. Focused regions are detected by extracting the bright and dark features at various scales by multi-scale top-hat transformation. Since the best-focused regions are detected and copied from one image only, a slight error in registration will have no effect in fusion except in the borders of the focused regions. Hence this region-based method is robust to mis-registration. This method resembles the manual cut-and-paste method of image fusion which is often used for comparison purposes. Thus the fused image obtained by the method is very similar to the ideal fused image. Performance analysis reveals that our method is superior to fusion by a state-of-the-art method.

Chapter 6 presents the conclusion of the thesis including a comparative study of techniques presented in previous chapters. It also presents a discussion on related future work.

In brief, in this thesis

- Chapter 1 presents a brief review on existing literature, the objective and contribution of the thesis, data-set used for experimentation and the quantitative measures used for performance evaluation,
- Chapter 2 presents an iterative algorithm for registration of multi-focus images by combining global and local transformation models,
- Chapter 3 presents a pixel-based algorithm for multi-focus image fusion using morphologic wavelets,

- Chapter 4 presents a block-based algorithm for multi-focus image fusion using a morphology-based focus measure in a quad-tree structure,
- Chapter 5 presents a region-based algorithm for multi-focus image fusion using multi-scale morphology,
- Chapter 6 presents conclusion of the thesis and gives an outline on related future work.

Experimental set-up and the data-set used for experiments are presented now.

1.6 Experimental set-up

Proposed algorithms are implemented using *C language* in *Unix environment*. All programs are executed on a machine with Intel Pentium processor T4400 and 1 GB RAM. Standard algorithms proposed by others have also been implemented in the same environment for comparison purpose.

1.6.1 Data used in experimentation

The algorithms are applied on a large number of multi-focus image-sets which vary in their object-contents and imaging set-up. Object-contents of image-sets vary in number, shape and distance of objects from the camera. Texture of image-sets varies in regularity, density and in combination of micro and macro texture. Some of the image-sets depict indoor scenes whereas others depict outdoor scenes. Images of indoor scenes generally contain human beings, animals and man-made objects. Man-made objects with straight-line edges (for example, book, book-shelf, table, window, door etc.) are helpful to detect artifacts like step-effects generated after processing. Images of outdoor scenes generally contain natural objects like flowers, plants, trees and also buildings. These images in general show irregularity in texture. Registration

of such images is difficult because in addition to other differences temporal changes between shots may occur due to wind. Hence slight mis-registration may be present in this type of images. This may in turn reveal the robustness of the fusion procedure against mis-registration.

Since it is not possible to include all experimental results in the thesis, we have chosen test image-sets in such a way that experiments are validated by different types of images. Twelve representative image-sets are used in the thesis and they are obtained from web-sites [32, 24, 26, 1, 28]. The image-sets named as 'Doll', 'Toy', 'Disk', 'Lab', 'Pepsi', 'Clock', 'Campus', 'Hydrant', 'Garden', 'Rose', 'News' and 'OpenGL' are shown in Fig. 1.2. Among these, the multi-focus 'Doll' images (Fig. 1.2A) are synthetic images generated from the famous painting named 'Las Meninas' by Diego Velázquez kept at 'Museo del Prado' in Madrid. These images have been generated artificially by a modern painter cum art-teacher John Hagan [32]. He has visually estimated the distances of various objects present in the painting. Accordingly different portions of the original image of the painting have been artificially defocused by him to illustrate the concept of 'depth-of-field'. Though the blurring model and the parameters are not known to us, we have used this multi-focus image-set as an ideal synthetic data-set for evaluating the performance of fusion algorithms. Moreover, this image-set contains three multi-focus images, hence it offers better illustration facility than the sets of two images. Image-sets 'Toy', 'Disk', 'Lab', 'Pepsi' and 'Clock' are obtained from web-site [24]; 'Campus' and 'Hydrant' are obtained from web-site [26]; 'Garden' is obtained from web-site [1]; 'Rose', 'News' and 'OpenGL' are obtained from web-site [28]. The characteristics of test image-sets are given now.

- Image-set 'Doll' depicts an indoor scene with many objects of arbitrary shapes and sizes and placed at different distances.
- Image-set 'Toy' depicts an indoor scene with many objects of regular shapes placed before a large and mostly dark background.
- Image-sets 'Disk' and 'Lab' contain many objects of different sizes and mostly

Table 1.1: Multi-focus images and their sizes

Figure	Size
Doll	384×576
Toy	512×512
Disk	448×576
Lab	448×576
Pepsi	512×512
Clock	256×256
Campus	480×640
Hydrant	480×640
Garden	320×448
Rose	512×704
News	224×320
OpenGL	512×704

of regular geometric shapes.

- Image-set 'Pepsi' and 'Clock' contain large objects all of which have regular geometric shapes.
- Image-sets 'Campus' and 'Hydrant' depict outdoor scenes with objects of mostly irregular texture and at large distances among themselves.
- Image-set 'Garden' depicts an outdoor scene with dense irregular texture.
- Image-set 'Rose' has a large area of regular grid-like structure as background.
- Image-set 'News' contains dense but mostly regular texture.
- Image-set 'OpenGL' contains both micro and macro textures.

As mentioned in Section 1.5, image registration is a necessary pre-requisite before fusion. Among the above image-sets, five (viz. 'Doll', 'Disk', 'Garden', 'Rose' and

A.(i) Doll: Near focused image

A.(ii) Doll: Middle focused image

A.(iii) Doll: Far focused image

Figure 1.2: Multi-focus image data-sets used for experimentation

'News') were not registered and we have registered them. Details of registration are given in Chapter 2. The rest of the images were already registered. Sizes of various image-sets after registration are given in Table 1.1.

1.7 Quantitative Performance Evaluation

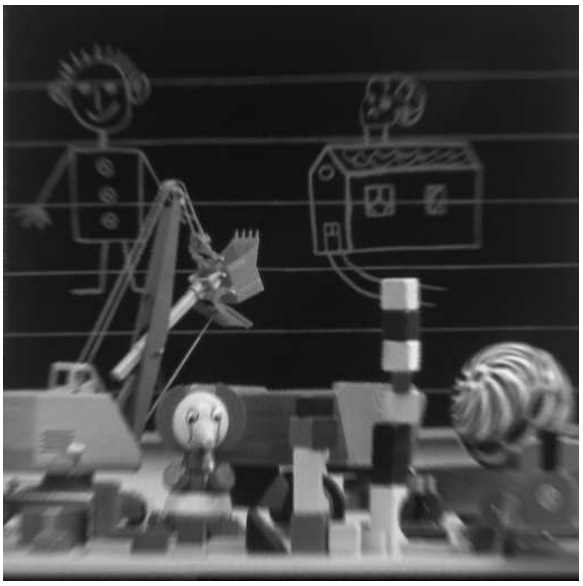
A good fusion algorithm should be able to work on a variety of input images, robust enough to tolerate probable mis-registration of input images and should not produce any unwanted visual effect or artifact. Moreover quality of the fused image should satisfy the requirement for intended application and computational complexity should also be affordable. Quality and time are inter-dependent and they are often related directly, that is, better quality needs more time. So depending on the specific application, one has to compromise/trade off between these two.

There are two types of assessment, subjective or qualitative and objective or quantitative [62]. In qualitative fusion quality assessment, subjects or observers are requested to examine the input image-sets and the output images obtained by various fusion techniques and then rank the output images according to their visual quality [64]. Average of the ranks given by different observers indicates the subjective quality of the techniques under examination. The process is time consuming, laborious and expensive. Moreover the assessment in this process is non-repetitive, that is, for the same set of images the ranking given by an observer may change from time to time. Quantitative fusion quality evaluation overcomes these draw-backs by employing a metric that quantifies the quality of the fused images. The metric should estimate how much information is obtained from the input images because goal of image fusion is to integrate information from multiple sources. In conventional methods, the ideal fused image is used as the reference image and the metrics like mean-square-error (MSE), peak-signal-to-noise-ratio (PSNR) are used to estimate the error between the reference image and the processed image. Since reference images are not available here, we need to use metrics which do not require them.

In this thesis, quantitative evaluation of fusion algorithms is done by using two different metrics. They are based respectively on image gradients and structural similarity index. Each of the metrics yields a numerical value from the input image-set and the fused image. None of them requires any reference image. For both of them, greater

B.(i) Toy: Near focused image

B.(ii) Toy: Middle focused image



B.(iii) Toy: Far focused image

Figure 1.2: Continued

C.(i) Disk: Near focused image

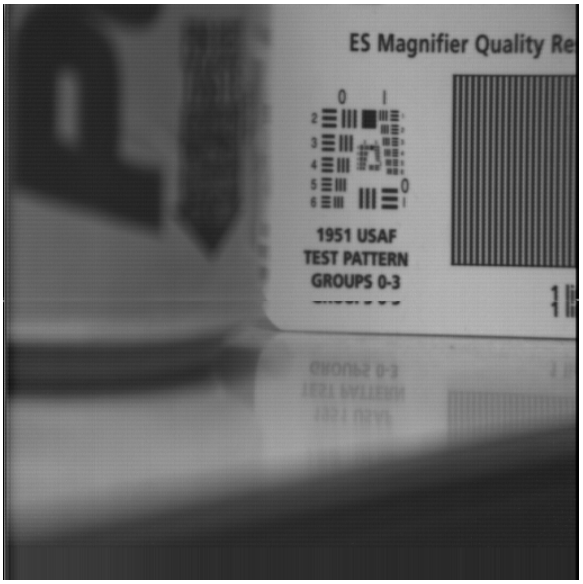
C.(ii) Disk: Far focused image

D.(i) Lab: Near focused image

D.(ii) Lab: Far focused image

Figure 1.2: Continued

E.(i) Pepsi: Near focused image



E.(ii) Pepsi: Far focused image

F.(i) Clock: Near focused image

F.(ii) Clock: Far focused image

Figure 1.2: Continued

G.(i) Campus: Near focused image

G.(ii) Campus: Far focused image

H.(i) Hydrant: Near focused image

H.(ii) Hydrant: Far focused image

Figure 1.2: Continued

I.(i) Garden: Near focused image

I.(ii) Garden: Far focused image

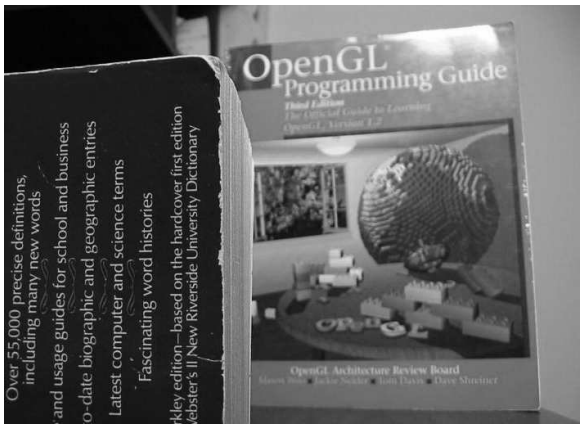
J.(i) Rose: Near focused image

J.(ii) Rose: Far focused image

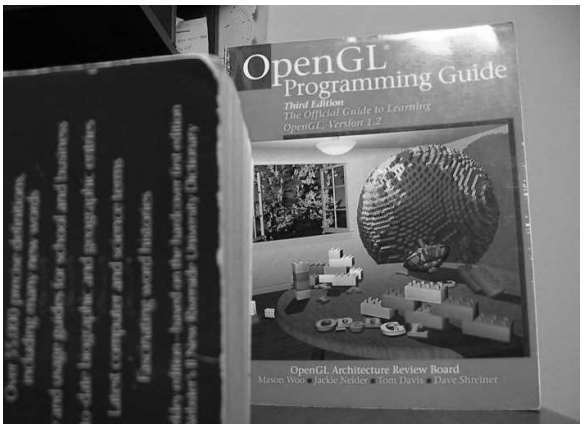
Figure 1.2: Continued

K.(i) News: Near focused image

K.(ii) News: Far focused image



L.(i) OpenGL: Near focused image



L.(ii) OpenGL: Far focused image

Figure 1.2: Continued

value means better fusion. The metrics are described below for two input images, however they can be extended easily to three or more input images.

1.7.1 Gradient similarity index

Gradients operators are useful tools to measure variations in intensity of a pixel with respect to its immediate neighboring pixels [13]. It is observed that a pixel possesses high gradient value when it is sharply focused. So in a set of multi-focus images, pixels of a sharply-focused region possess higher gradient values than pixels of the corresponding out-of-focus region. This observation led to an image fusion performance measure employing image gradients [57, 22]. For two multi-focus input images X_1 and X_2, gradient images G_1 and G_2 are obtained first. Then G_1 and G_2 are combined into G by taking the maximum gradient value at each pixel position (r, c). Therefore

$$G(r,c) = \max\left(G_1(r,c), G_2(r,c)\right) \text{ for all } (r,c) \tag{1.10}$$

Thus only the sharply focused pixels from the constituent images have their contribution in the *maximum gradient image* G. Let $\tilde{G}$ denotes the gradient of the fused or reconstructed image F. It is referred to as the *gradient of fused image*. Then, more similar G and $\tilde{G}$ are, better is the fusion algorithm. Now, following the usual definition of signal-to-noise ratio, a simple objective measure of similarity between two gradient images is calculated as

$$S(G,\tilde{G}) = 1 - \frac{\sqrt{\sum (G(r,c) - \tilde{G}(r,c))^2}}{\sqrt{\sum G^2(r,c)} + \sqrt{\sum \tilde{G}^2(r,c)}} \tag{1.11}$$

We call S the gradient similarity index (GSI). Here, $\sqrt{\sum (G(r,c) - \tilde{G}(r,c))^2}$ determines the error or dissimilarity between the images and it is normalized by the quantity $\sqrt{\sum G^2(r,c)} + \sqrt{\sum \tilde{G}^2(r,c)}$ to make the measure unbiased to overall brightness of the images. So for an ideal fused image S approaches the value 1. For our experimentation, we have calculated the gradients by Robert's gradient operator [70]. For

more than two input images, $G(r, c)$ is calculated as the maximum of the gradients at (r, c) taken over all input images.

1.7.2 Fusion quality index

Structural similarity index (SSI) proposed by Wang and Bovik [83] is an effective metric to measure the quality of an image. For two real-valued sequences $X = (x_1, x_2, \ldots, x_n,)$ and $Y = (y_1, y_2, \ldots, y_n,)$, the metric $Q_0(X, Y)$ defined as

$$Q_0(X, Y) = \frac{4 \times \sigma_{XY} \times \mu_X \times \mu_Y}{(\mu_X^2 + \mu_Y^2) \times (\sigma_X^2 + \sigma_Y^2)} \tag{1.12}$$

measures the structural similarity of X and Y. Here μ_X and μ_Y are the mean values of X and Y; σ_X^2 and σ_Y^2 are the variances of X and Y; and σ_{XY} is the covariance of X and Y. Structural similarity of two images is defined in a similar way. Since image signals are generally non-stationary, it is more appropriate to measure Q_0 over local regions and then combine the different results into a single measure. The authors [83] proposed to use a sliding window approach. Starting from the top-left corner of the two images X_1, X_2, a sliding window of fixed size (with n pixels) moves pixel by pixel over the entire image until the bottom-right corner is reached. For each window w, the local quality index $Q_0(X_1, X_2 \mid w)$ is computed. Finally, the structural similarity index (SSI) Q_0 is computed by averaging all local quality indices.

Piella and Heijmans [66] proposed variants of SSI to measure quality of image fusion. Fusion quality index (FQI) $Q(X_1, X_2, F)$ for input images X_1, X_2 and output image F is defined by them as

$$Q(X_1, X_2, F) = \frac{1}{|W|} \sum_{w \in W} (\lambda_1(w)\, Q_0(X_1, F \mid w) \;+\; \lambda_2(w)\, Q_0(X_2, F \mid w)) \tag{1.13}$$

where $Q_0(X_1, F \mid w)$ is the structural similarity index of X_1 and F over the local window w, W is the family of all local windows, $|W|$ is the cardinality of W and λ_1 and λ_2 are weights obtained from local saliency measures. Local saliency measure $s(X_1|w)$ of input image X_1 should reflect the local relevance of X_1 within the window

w, and it may depend on contrast, sharpness or entropy. Given the local saliencies $s(X_1|w)$ and $s(X_2|w)$, the local weights $\lambda_1(w)$ and $\lambda_2(w)$ is computed. It indicates the relative importance of image X_1 compared to image X_2. A typical choice for $\lambda_1(w)$ is $\frac{s(X_1|w)}{s(X_1|w)+s(X_2|w)}$. In our evaluation, we have taken the window-size to be of 8×8 pixels and the sum of gradient values in the local window to be the local saliency measure. For more than two input images, Q is calculated as the average weighted sum of Q_0's calculated for all images. Here weight for a local window in an image is calculated as the saliency of the window in that image divided by sum of local saliencies for all corresponding windows in all other images.

1.8 Organization of the thesis

Organization of the thesis follows. A survey on multi-focus image registration and an iterative and hybrid method for the same are presented in Chapter 2. Chapter 3 presents a computationally efficient pixel-based algorithm for MFIF using wavelet. Before describing the algorithm, the basic theory and new wavelet called morphologic wavelet is presented. Chapter 4 presents a block-based method for MFIF. It employs a new focus-measure called energy of morphologic gradients. Chapter 5 presents a region-based method for MFIF using multi-scale morphologic operators. In each chapter, after describing the new algorithm, experimental results on data-sets given in Figure 1.2 are presented. Finally, Chapter 6 presents concluding remarks and out-lines future work.

Chapter 2

Multi-focus image registration

2.1 Introduction

Image registration is a necessary pre-requisite for multi-focus image fusion because before fusion the constituent images must be positioned properly with respect to a common coordinate system so that corresponding objects are overlaid properly [41]. In general, the registration techniques may be classified according to two major aspects: methodology and application-area. The methods can be categorized into two types: (i) area-based and (ii) feature-based [92]. A third category has emerged which is a hybrid of area-based and feature-based techniques. Registration techniques may also be classified by their mapping models, that is by examining whether they apply global and/or local mapping models. Global models use information from the entire image to estimate the mapping function parameters. On the other hand, local models treat the image as a composition of blocks/regions and the function parameters are estimated separately for each block/region.

Registration techniques for multi-focus image have been proposed in [41, 90, 91, 18, 22, 29, 27]. Among these, the methods proposed in [41, 90, 91, 18, 22] use global affine transformation models and the ones proposed in [29, 27] use global perspective

transformation models. The technique proposed by Kubota *et al.* [41] is an area-based multi-scale technique. In this technique, from the source and the reference images Gaussian pyramids are obtained at first. At the coarsest level of the pyramids, translation, rotation and magnification parameters are estimated by the minimum MSE between the two images. The parameters are propagated to the next finer level and are further refined. The refinement process continues up to the original resolution level and the parameters obtained there are used to register the source image. Zhang and Blum [90, 91] proposed a hybrid multi-scale scheme using both area-based and feature-based techniques. In this technique also, from the source and the reference images Gaussian pyramids are obtained at first. At the coarsest level of the pyramids, an initial estimation of transformation parameters (mainly rotation and translation) is done by using the edge features. The parameters are updated by iterative refinement of the optical flow estimation. They are propagated to the next finer level and are further refined. The process continues up to the finest level in which the final parameters are obtained and are used to register the source image. De and Chanda [18, 22] described an area-based technique in which at first the source and reference images are divided into equal number of blocks. A source block is swiped over the corresponding reference block to find out the best matching position in the block. Corresponding point-pairs are taken from best-matching blocks. Finally, affine transformation parameters are estimated by the best-matching pairs of points by using the least-square method. These parameters are then used to register the source image. Goshtasby [29] proposed a hybrid registration scheme in which the edge-intersection points are used as unique landmarks. At first, the landmarks in the source image are found. Then corresponding landmarks in the reference image are found by correlation template matching. From the corresponding landmark pairs, the best four satisfying the projective constraints are identified. They are used to calculate the projective transformation parameters. Source image is then registered by using these parameters. Fedorov *et al.* [27] used a hybrid registration scheme in which a number of well-located control-points are extracted globally at first. Preliminary matches of the tie-points are established by identifying the pairs with minimum distance in the descriptor space. Afterwards, the inevitable outliers

are pruned off using RANSAC-like algorithm. Finally perspective transformation parameters are estimated by the matched tie-points using the Normalized Direct Linear Transformation (DLT) algorithm. Source image is then registered by using these parameters.

The methods described above use *global transformation models* and do not apply any *local model* appropriate for registration of multi-focus images. In general, these images are acquired one by one in such a way that each image in the set has focus on objects at a particular distance from the camera. This results in global as well as local variations in the images. In this chapter we explore these variations and present an iterative algorithm for registration of multi-focus images which combines both global and local mapping models [20]. In the first step of the algorithm, a global translation is determined by maximizing the mutual information of the source and the reference images and then it is applied on the source image. In the second step, a block-wise local scaling is applied on the translated source image. The scale-factors are determined by maximizing a similarity measure of two corresponding blocks of the translated source image and the reference image. The global and local transformations constitute a hybrid technique which is iterated to obtain the optimal result. The proposed method is automatic, easy to implement and gives good results. Results obtained by applying the method on different sets of multi-focus images are provided with. Performance of the system is also evaluated and is compared with a widely used method. The chapter is organized as follows. Section 2.2 describes the proposed algorithm. Experimental results and discussion including performance analysis are given in Section 2.3. Finally, concluding remarks are placed in Section 2.4.

2.2 An iterative hybrid registration algorithm

Multi-focus images of a scene are acquired one by one either by hand-held cameras or by cameras placed on tripods, in identical environmental conditions in respect to sensor, light, view direction, orientation and object-content in the scene [22, 27]. Each

image in the set has focus on objects at different distances in the scene. Previous research indicates that when the distance between the scene and the camera is large, it is usually possible to approximate the motion of the scene using an affine transformation [90]. Note that an affine transformation is usually a combination of translation, rotation and scaling (see Appendix B). In reality, for such applications, rotation of the camera relative to the scene is insignificant and hence is not considered here. Global scale-change between images may occur due to changes in focal settings. However in most practical applications, it is less than three percent [76] and hence is not considered here. We consider global (horizontal and/or vertical) translation(s) between images due to accidental camera-pan between shots taken by hand-held cameras and the changes due to variations in focal settings during acquisition.

Focal variations are done intentionally to focus on objects at a particular distance. For example, objects at the background of a scene are farther than those at the foreground and during acquisition, focus at background generates a far-focused image in which the background objects are in focus but the foreground objects are out-of-focus. Similarly focus at foreground generates a near-focused image in which the foreground objects are in focus but the background objects are out-of-focus. Hence partial defocusing/blurring is inevitable in multi-focus images. Partial defocusing affects the images in two ways. Firstly, due to point-spreading, a blurred object appears to be larger in an image when compared to its focused counterpart in some other image [40]. In addition to that the radii-of-blur may vary in near-focused and far-focused images. This results in local scale-change between images. Secondly, the position of an out-of-focus object may be changed when compared to the position of its focused counterpart in some other image. This is shown in Figure 2.1 by a *paraxial geometric optics model of image formation using a thin convex lens*. The focused image of a point-object P is created as a point-image P' on *Plane-2* which is the in-focus image-plane for P. All other image-planes nearer to or farther from the lens than *Plane-2* are out-of-focus image-planes for P. *Plane-1* and *Plane-3* are two such out-of-focus planes. The blurred images of point-object P appear as blur-circles with diameters AB and CD on *Plane-1* and *Plane-3* respectively. Hence the sizes of the

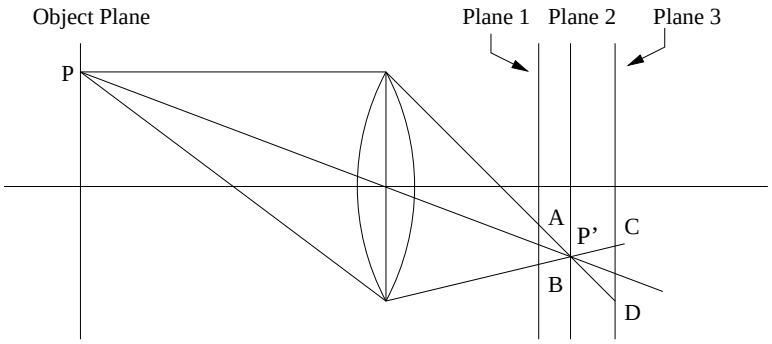


Figure 2.1: Geometric Optics Model of Lens System

focused and blurred images of the point-object do vary. In addition to that, blur-circle is shifted vertically upwards in *Plane-1* and the same is shifted vertically downwards in *Plane-3*. So the focused and blurred images of an object do have position differences as well. Intensity or radiometric differences caused by partial defocusing are not dealt with in this work because they are intrinsic to multi-focus images and we do not intend to change them. We rather concentrate on spatial transformations due to camera pan and partial defocusing. A single global transformation is not adequate to capture all these effects. Considering this fact, a registration technique is presented which works in two steps. To nullify the effects of global translation(s), the source image is translated globally in the direction(s) reverse to that of the camera pan. Once the translation is done, local variations in size and position are corrected by block-wise local scaling. Above two steps are iterated until a certain error criterion is fulfilled. A schematic diagram depicting the iterative steps is shown in Figure 2.2.

In registration of a set of multi-focus images, every image is equally authentic with its coordinate system. One of them is chosen to represent the common coordinate system and is called the reference/target/destination image. Other images are called source images. Source images are then registered to the reference image. Registration is a mapping between two images both spatially and with respect to intensity [7]. If source image X_s and reference/destination image X_d are defined as two-dimensional arrays of intensity values on spatial coordinates (r, c), then mapping between them

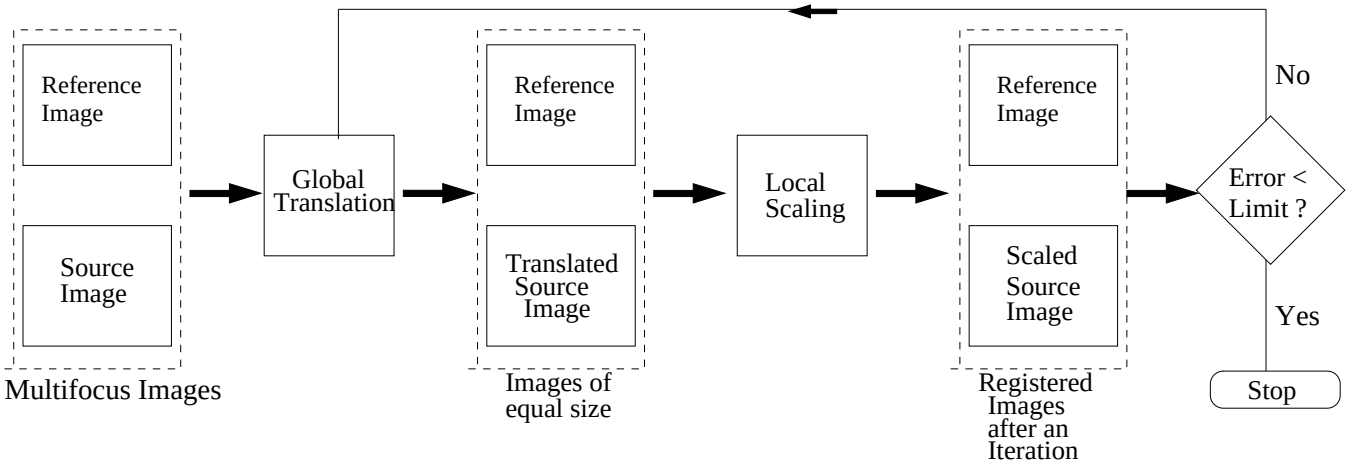


Figure 2.2: Schematic diagram for a hybrid and iterative registration method

can be expressed generally as

$$X'_s(u,v) = F_i(X_s(F_g(r,c))) \tag{2.1}$$

such that

$$e = || X_d(u,v) - X'_s(u,v) ||^2 \tag{2.2}$$

be minimum, where F_i is the mapping for intensity transformation and F_g is the mapping for geometric transformation so that $(u,v) = F_g(r,c)$. The equation may vary depending on the application. In this work, F_g is a space-variant transformation which is a combination of global and local geometric transformations instead of a single global transformation generally used for multi-focus image registration. We describe below the method for registration of source image X_s with reference image X_d.

2.2.1 Global translation

Since multi-focus images are acquired one by one, accidental camera pan during acquisition may happen and this results in global translation(s) of the image in small amounts. This effect can be nullified by translating the image in reverse direction(s). The amount of translation is determined by maximizing the *mutual information* of

X_s and X_d. Mutual information (MI), originating from the information theory, is a measure of statistical dependency between two data-sets [33]. MI between two overlapping images X_s and X_d is given by

$$MI(X_s, X_d) = H(X_s) + H(X_d) - H(X_s, X_d) \tag{2.3}$$

where $H(X_s)$ is the Shannon entropy defined as

$$H(X_s) = -\sum_k p_s(k) \log p_s(k) \tag{2.4}$$

where $p_s(k)$ is the probability of occurrence of grey value k in the image X_s. Similar is the definition for Shannon entropy $H(X_d)$ of image X_d. Joint entropy $H(X_s, X_d)$ of two images X_s and X_d is given by

$$H(X_s, X_d) = -\sum_{(k,l)} p(k,l) \log p(k,l) \tag{2.5}$$

where $p(k,l)$ is the joint probability of occurrences of grey values k and l in images X_s and X_d respectively. Entropy of a probability distribution is low when the distribution has a few sharply defined, dominant peaks and it is maximum when all outcomes have an equal chance of occurring that is, the distribution is uniform. The same is true for joint entropy. It can be seen from Equation (2.3) that a small value of joint entropy leads to a large value of MI. The idea that MI can be used for image registration was pioneered by Collignon *et al.* [17] and Viola and Wells [80]. Both groups used the idea for registration of multi-modal images. It is based on the assumption that if two multi-modal images are properly aligned, then corresponding objects (and hence their respective range of grey values) from two images overlay on one another. This results in a few sharply defined, dominant peaks or ridges in the joint probability distribution of the images. Hence, their joint entropy is minimized and consequently MI is maximized.

The idea is extended to multi-focus image registration. Source image X_s is swiped over reference image X_d in such a way that grids of both images match properly. This is done by applying integral amount(s) of translation(s) along the axes. So no interpolation is employed here. Then for each translation, overlapping sub-images of

the translated source image and the reference image are found. MI of the overlapping sub-images is calculated. Varying the translation-amounts within a range and calculating the MI of overlapping sub-images, the amount of translation which maximizes MI is found. Suppose, the source image X_s after optimum translation(s), is mapped to $X_s(r + T_r, c + T_c)$ where T_r and T_c are respective translations along the row and column axes. After mapping, *only the overlapping portions* of translated source image and reference image are retained and the rest are truncated. So essentially, translated source image and reference image become of same size after truncation. This is important because in next step we need the source and reference images to be of the same size. Henceforth, we shall refer to new source image as X_s and new reference image as X_d.

The choice for ranges within which T_r and T_c are varied is an experimental issue. Greater range means better accuracy, but that also means greater time-requirement. We have experimented with various ranges of T_r and T_c and have seen in general that the shifts are within 5 pixels. So we have taken the range of T_r and T_c to be -5 to +5. Mis-alignments greater than 5 pixels are corrected during successive iterations.

2.2.2 Local scaling

Variations in focal settings during acquisition of multi-focus images result in local scale and position differences in focused and defocused images of an object, as explained in the beginning of this section by using Figure 2.1. The problem is addressed by block-wise registration of X_s with X_d. At first, X_s and X_d are divided into n equal-sized non-overlapping blocks. Since X_s and X_d are of same size, their block-sizes are taken to be equal. We have experimented with different values of n and found that $n = 16$ is a reasonably good choice for practical purposes. Each block of X_s is scaled independently by appropriate factors along the axes. Resultant image is then obtained by stitching the scaled blocks at their proper positions.

Best scale-factors for a block

The best scale-factors (along row and column axes) for the *k-th* source block is determined by varying the scale-factors and then finding out the ones which give the best matching with the *k-th* reference block. The range and precision for varying the scale factors are important. We have experimented with three different ranges viz. 0.96-1.04, 0.97-1.03 and 0.98-1.02 with three different increment-values in each range, viz. 0.02, 0.01 and 0.005. It is observed from our experiments that in general increasing the range does not improve the results but finer precision gives better results. The range 0.98-1.02 with precision 0.005 is found to be suitable for our purpose and we have used those values for varying the scale-factors along the axes.

Suppose that *k-th* source block is scaled upon by the scale-factors s_r and s_c respectively along r-axis and c-axis. Depending on the scale-factors, horizontal and vertical dimensions of the scaled block are changed independently. Scaled source block is swiped over *k-th* reference block. Suppose X_s^k and X_d^k respectively are overlapping sub-images of *k-th* source block (after scaling) and *k-th* reference block. To find out the best matching scale-factors of *k-th* source block we need either a similarity or a dissimilarity measure. Small block-size reduces the statistical power of the probability distribution estimation [67]. Hence instead of mutual information, area-based dissimilarity measure *sum of squared differences (SSD)* is used. For each swiping position of the source block, SSD between the overlapping sub-images X_s^k and X_d^k is computed by

$$SSD(X_s^k, X_d^k) = \sum_r \sum_c \{X_s^k(r,c) - X_d^k(r,c)\}^2 \tag{2.6}$$

The best match occurs when the SSD is minimum. The SSD's for best matching positions for 9 different values (in the range 0.98-1.02 with precision 0.005) for each of the scale-factors s_r and s_c are noted. This results in total 81 readings of SSD for the block. The minimum of them gives the best scale factors for the block. The range of scale-factors as stated above is obtained from experiments with a large number of images and is found to satisfy the real-life problem.

Stitching a scaled block

Stitching a scaled block in the resultant image requires additional care. Before scaling all source and reference blocks are of equal size. After scaling, if the scaled source block is smaller in size than the original source block and is stitched to the resultant image, then some blank area will be created. In that case an appropriate larger block surrounding the original source block is scaled and positioned there. If the resized block is larger than the reference block, it is clipped after positioning properly. Essentially, the registered source block and the reference block should be of same size. For clarity, consider the following example.

Let us illustrate the situations which may occur due to local scaling, with a source block of size, say 100×100 pixels. Suppose best scale-factors for the block is 0.98 along both axes. So after scaling, size of the block is 98×98 pixels which is smaller than its target area. Hence a 102×102 block containing the original source block is scaled to obtain a 100×100 block which fits the target area. Now consider another case in which the best scale-factors for the block is 1.02 along both axes. After scaling its size will be 102×102 pixels which is bigger than its target area. The best matching position of it is found by swiping it over the corresponding reference block. After that it is clipped to 100×100 pixels, and then stitched to its proper position. Finally consider the case where the scale-factors are 1.02 along r-axis and 0.98 along c-axis. So the block becomes 102×98 pixels after scaling. In this case, a bigger block of size 100×102 containing the original block is taken, so that it becomes 102×100 after scaling. After finding out the best matching position, the block is clipped to 100×100 pixels and is stitched to the target block. Each source block is registered to the corresponding reference block in this way.

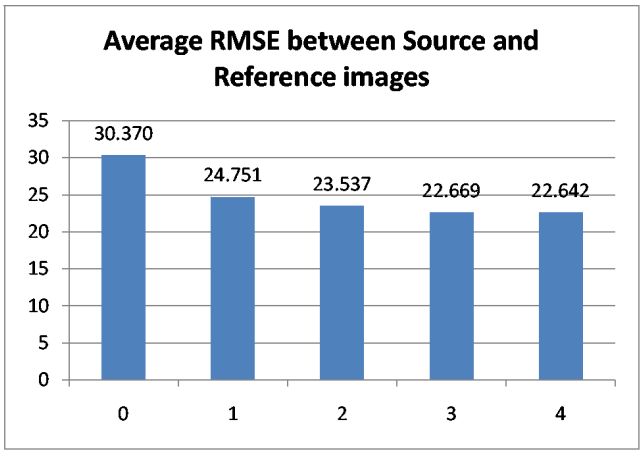


Figure 2.3: Average error between source and reference images as the iteration step number increases

2.2.3 Iteration

As stated above, the proposed registration technique has two distinct steps: (i) global translation and (ii) local scaling. Optimum transformation parameters are determined in these two transformations independently. But when they are combined, independent parameters may not remain optimum any more. Hence we iterate these two steps in the given order to achieve more acceptable result. We expect and experimentally verified to that the transformations F_c and F_i of Equation (2.1) are updated and the error defined in Equation (2.2) is reduced. The iteration is stopped when there is no significant change in error. Root-mean-square-error (RMSE) between the source and reference images is taken as the measure of error for our implementation purpose. Average RMSE between source and reference images (used in Section 2.3) are shown against iteration step number in Figure 2.3. Column-0 indicates RMSE before registration, and for i=1 to 4, Column-i indicates RMSE after i-th step of iteration. It is seen from the Figure 2.3 that RMSE decreases considerably in the first step of iteration, then as the iteration step number increases RMSE decreases, but with gradually slower rate.

Interpolation

In the global translation step, the source image is swiped over the reference image in such a way that grids of both images match properly. Hence no interpolation is required in this step. In the local scaling step, however, grids of the source and the reference blocks do not match in general. Hence, interpolation is required. Bi-linear interpolation is a reasonable choice in terms of ease-of-implementation and time-complexity. But during successive iterations it may reduce the contrast of the images. A higher-order interpolation like bi-cubic interpolation is a better choice in that respect although it takes more time [13]. To reduce the time-requirement, bilinear interpolation is used while estimating the scale-factors for a block, and once the best scale-factors are obtained, the block is reconstructed finally to be a part of the resultant registered image, using bi-cubic interpolation.

2.3 Experimental results and discussion

The proposed algorithm for image registration has been implemented in C language in Unix environment. The global translation step has been implemented by varying the translation-amount from -5 to +5 with unit increment along each axis. In the local scaling step, source and reference images are divided into 16 blocks and for each block the scale-factors along the axes are varied from 0.98 to 1.02 with an increment-value of 0.005. At most three iterations were seen to be enough in each case. Experimental results for five sets of multi-focus images ('Doll', 'Disk', 'Garden', 'Rose' and 'News') are shown in figures 2.4-2.8. In each result, original multi-focus images are followed by registered images by the proposed method. The first image is taken as the reference image in each case. To show the effectiveness of the method, difference images (between the source and the reference image) before and after registration are also provided.

In Fig. 2.4, A.(i), B.(i) and C.(i) respectively are near-focused, middle-focused and

far-focused images before registration; A.(i) is taken as the reference image and B.(i) and C.(i) are taken as source images. After truncation of A.(i); A.(ii) is the new reference image; B.(ii) and C.(ii) are registered versions of B.(i) and C.(i) respectively. In Fig. 2.4, the images D. and E. are difference images between A.(i) and B.(i) and between A.(i) and C.(i) respectively. So they are difference images (between the source and the reference image) before registration. Finally, the images F. and G. are difference images between A.(ii) and B.(ii) and between A.(ii) and C.(ii) respectively. So they are difference images (between the source and the reference image) after registration. As stated in Section 2.2, this work concentrates on position and scale differences due to partial defocusing and does not aim to change the intensity/radiometric differences caused by it. This is evidenced by difference images. In both images A.(i) and B.(i), the background region is out-of-focus and very little or no intensity differences do exist there. Differences in that region are mainly due to position and scale differences. Comparison of the difference images D. and F. shows that the position and scale differences in the background region are reduced considerably in the latter. The differences in other regions (caused by position and scale differences as well as intensity variations due to focusing and defocusing) are not eliminated completely but are reduced which is shown by less bright edges in these regions. Similar is the case for the middle region of difference images E. and G.

Figure 2.5 shows registration of a widely-used set of multi-focus images which depicts an indoor scene containing many objects of different geometric shapes and of different sizes. In this figure, A.(i) and B.(i) are near-focused and far-focused images respectively; A.(i) is taken as the reference image and B.(i) is taken as source image; A.(ii) is the new reference image after truncation of A.(i), and B.(ii) is the registered version of B.(i); finally C. and D. are the difference images between A.(i) and B.(i) and between A.(ii) and B.(ii) respectively. Inspection of difference images inside the clock region reveals thinner and less bright edges in image 2.5.D. which in turn indicate that the proposed registration technique corrects the mis-alignments and local scale variations in the input images.

Figure 2.6 shows the results of registration of multi-focus images of an outdoor scene. In general, registration of images of such scenes is difficult because in addition to other differences temporal changes between shots may occur due to wind. Here the task was more challenging because of the dense and irregular texture of the images. In figure 2.6, A.(i) and B.(i) are near-focused and far-focused images respectively; A.(i) is taken as the reference image and B.(i) is taken as source image; A.(ii) is the new reference image after truncation of A.(i), B.(ii) is the registered version of B.(i); finally C. and D. are the difference images between A.(i) and B.(i), and between A.(ii) and B.(ii) respectively. Comparison of difference images 2.6.C. and 2.6.D. reveals that differences have reduced considerably after registration.

Images of Figure 2.7 have a large area of regular grid-like structure as background. Any mis-alignment can be detected easily in this type of images. The results of registration do not show any such effect. In figure 2.7, A.(i) and B.(i) are near-focused and far-focused images respectively; A.(i) is taken as the reference image and B.(i) is taken as source image; A.(ii) is the new reference image after truncation of A.(i), and B.(ii) is the registered version of B.(i); finally C. and D. are the difference images between A.(i) and B.(i) and between A.(ii) and B.(ii) respectively. Comparison of images 2.7.C. and 2.7.D. shows that edges are less bright in the latter which means mis-alignments have been corrected after registration.

Finally, Figure 2.8 shows the results of registration of images of dense but mostly regular texture. In this figure, A.(i) and B.(i) are near-focused and far-focused images respectively; A.(i) is taken as the reference image and B.(i) is taken as source image; A.(ii) is the new reference image after truncation of A.(i), and B.(ii) is the registered version of B.(i); finally C. and D. are the difference images between A.(i) and B.(i) and between A.(ii) and B.(ii) respectively. Comparison of images 2.8.C. and 2.8.D. shows that dark areas have increased in the latter which means after registration the difference between source and reference images have been reduced.

Careful manual inspection of the results also shows that the proposed registration method does not produce any unwanted visual artifact or aliasing. To show that there

is no block stitching artifact in the registered images, we have provided Figure 2.9 which shows magnification of a portion (where corners of four blocks coincide) of middle-focused image Fig. 2.4.B.(i) and corresponding registered image Fig. 2.4.B.(ii).

A.(i) A.(ii)

B.(i) B.(ii)

C.(i) C.(ii)

Figure 2.4: 'Doll': Original, Registered and Difference images. A.(i) Near-focused image (reference-image), A.(ii) Truncated near-focused image (new reference-image), B.(i) Middle-focused image, B.(ii) Registered middle-focused image, C.(i) Far-focused image, C.(ii) Registered far-focused image, D. Difference between A.(i) & B.(i), E. Difference between A.(i) & C.(i), F. Difference between A.(ii) & B.(ii), G. Difference between A.(ii) & C.(ii)

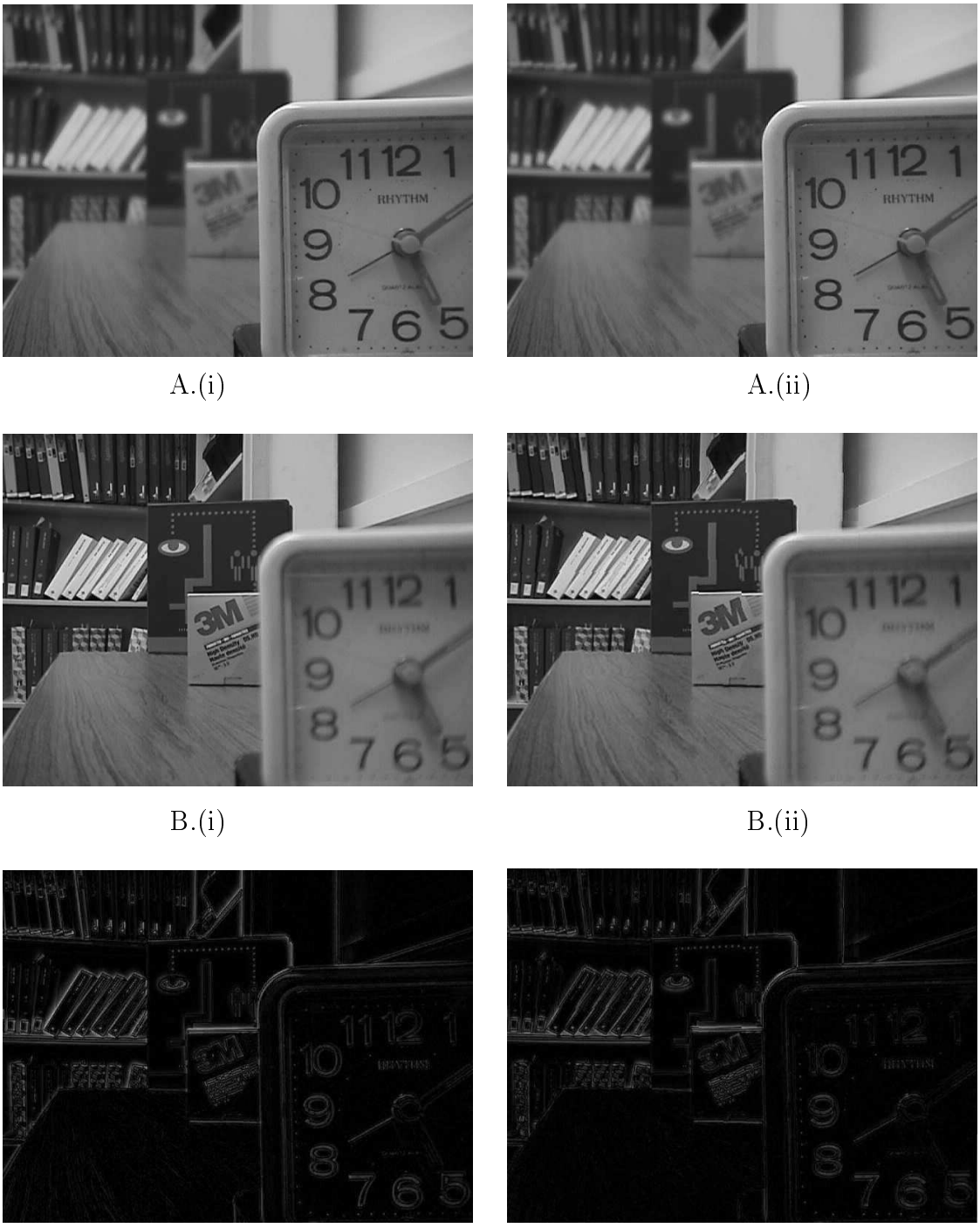
D. E.

F. G.

Figure 2.4: Continued

2.3.1 Quantitative performance evaluation

Comparative study of the proposed registration method with the most widely used one viz. global affine transformation [22] have been done. Since subjective evaluations may not be universally acceptable, we compare the methods by three quantitative measures viz. root-mean-square-error (RMSE), mutual-information (MI) and normalized-cross-correlation (NCC) [13]. Good registration decreases the value of RMSE between source and reference images and increases the value of MI and NCC between the two images. Quantitative results are presented in Tables 2.1-2.3. In each table, the first column gives serial-number of the source and reference image-pairs as shown in figures 2.4-2.8; the second column presents the quantitative-metric values before



A.(i) A.(ii)

B.(i) B.(ii)

C. D.

Figure 2.5: 'Disk': Original, Registered and Difference images. A.(i) Near-focused image (reference-image), A.(ii) Truncated near-focused image (new reference-image), B.(i) Far-focused image, B.(ii) Registered far-focused image, C. Difference between A.(i) & B.(i), D. Difference between A.(ii) & B.(ii)

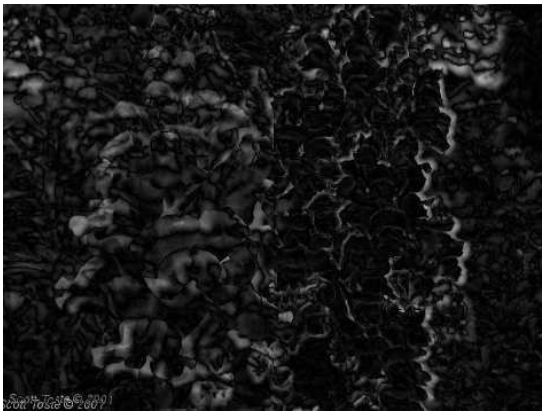
A.(i)

A.(ii)

B.(i)

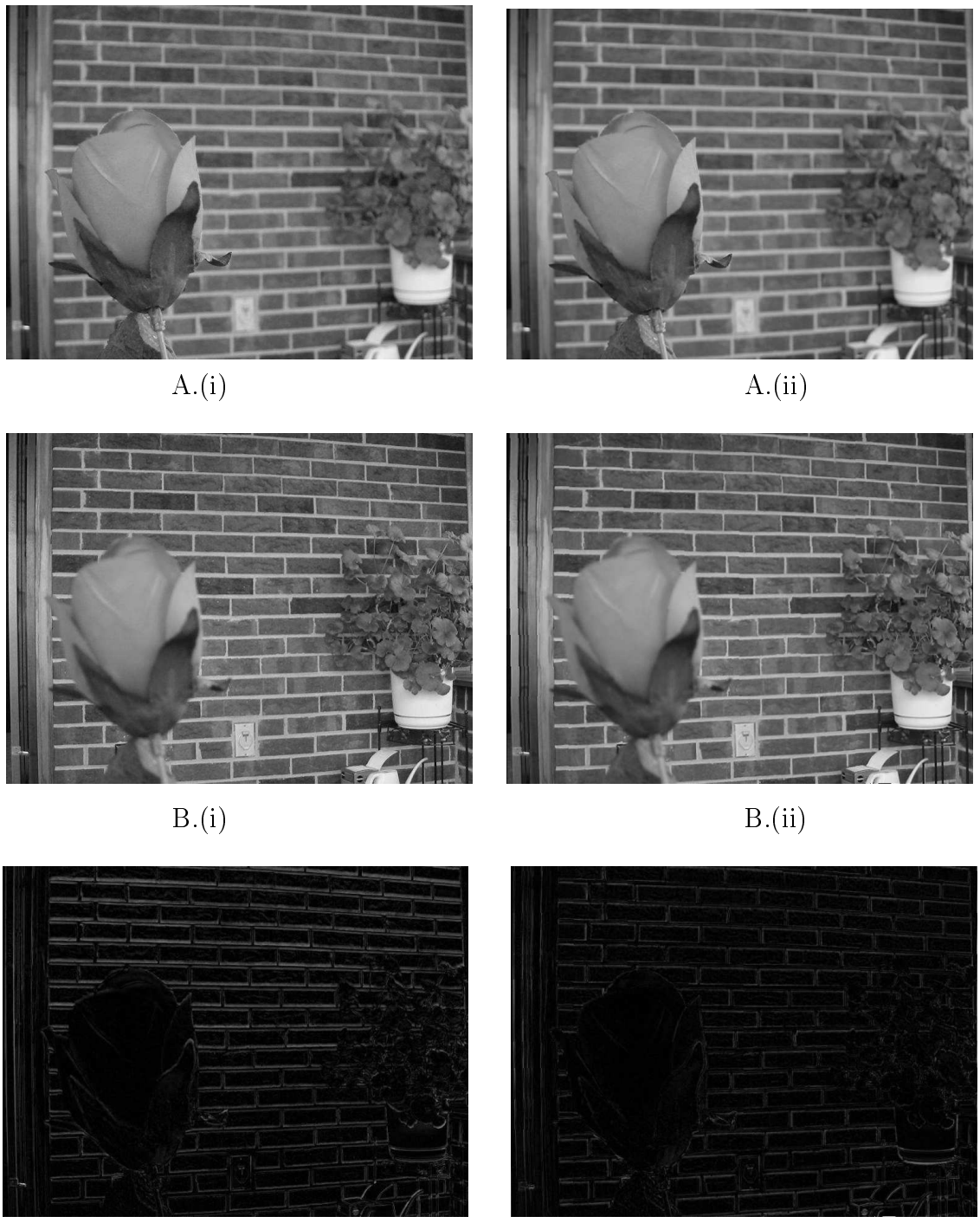
B.(ii)

C.



D.

Figure 2.6: 'Garden': Original, Registered and Difference images. A.(i) Near-focused image (reference-image), A.(ii) Truncated near-focused image (new reference-image), B.(i) Far-focused image, B.(ii) Registered far-focused image, C. Difference between A.(i) & B.(i), D. Difference between A.(ii) & B.(ii)

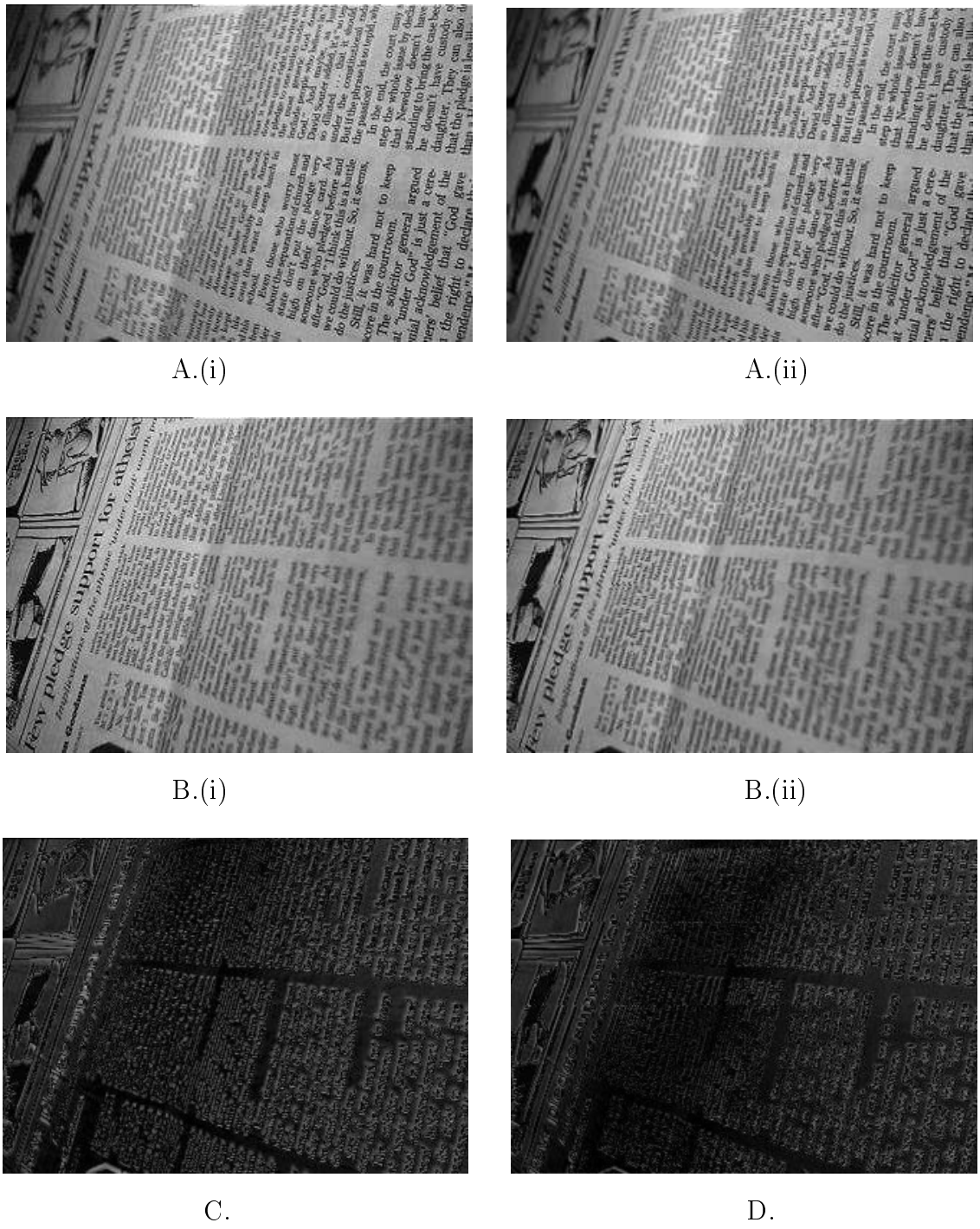


A.(i) A.(ii)

B.(i) B.(ii)

C. D.

Figure 2.7: 'Rose': Original, Registered and Difference images. A.(i) Near-focused image (reference-image), A.(ii) Truncated near-focused image (new reference-image), B.(i) Far-focused image, B.(ii) Registered far-focused image, C. Difference between A.(i) & B.(i), D. Difference between A.(ii) & B.(ii)

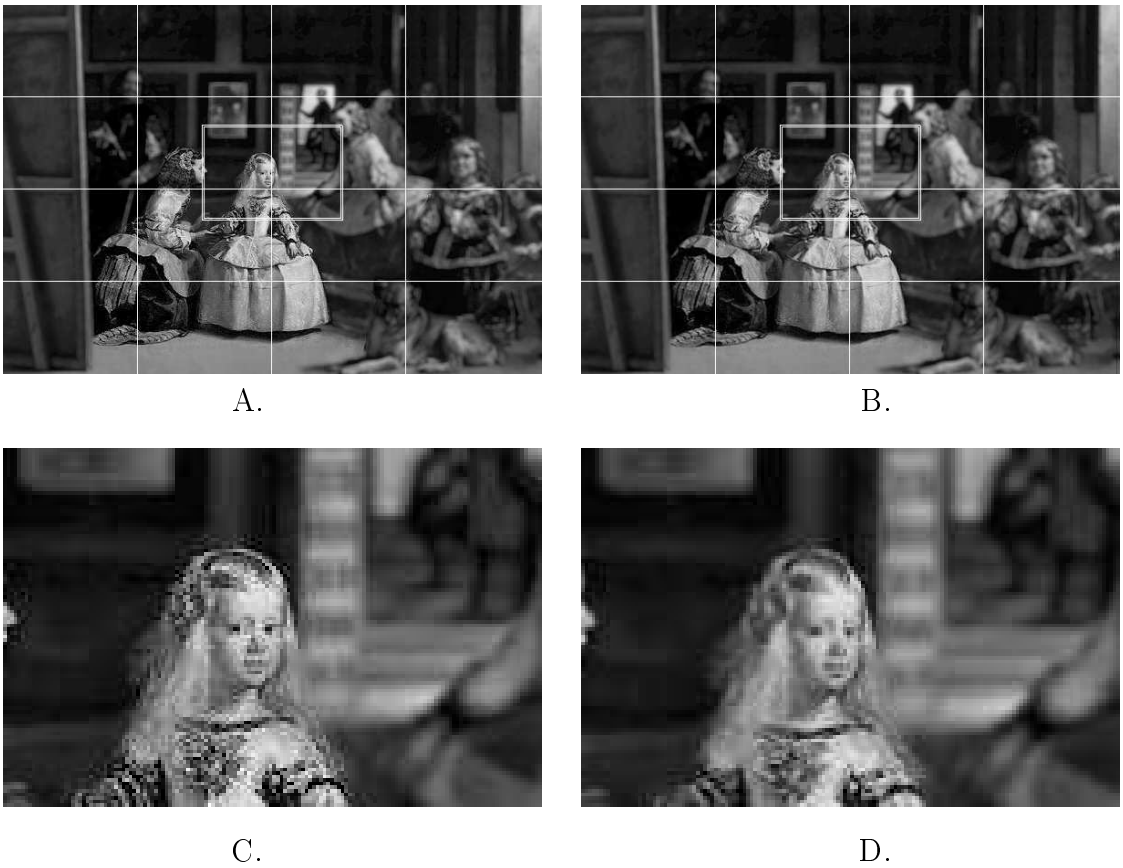


A.(i) A.(ii)

B.(i) B.(ii)

C. D.

Figure 2.8: 'News': Original, Registered and Difference images. A.(i) Near-focused image (reference-image), A.(ii) Truncated near-focused image (new reference-image), B.(i) Far-focused image, B.(ii) Registered far-focused image, C. Difference between A.(i) & B.(i), D. Difference between A.(ii) & B.(ii)



A. B.

C. D.

Figure 2.9: Magnification of selected areas before and after registration. The blocks used for *scaling during registration* and those used for *magnification* are shown in single and double lines respectively. A. Fig. 2.4 B.(i) shown in blocks, B. Fig. 2.4 B.(ii) shown in blocks, C.-D. Corresponding areas bounded by double-lines are shown magnified

registration; the third and fourth columns give the quantitative-metric-values after registration, respectively by global affine transformation method and by the proposed method; and the fifth and sixth columns give percentages of reduction or increment of the metric-values after registration.

It is seen from the Tables 2.1-2.3 that RMSE is decreased and MI and NCC are increased in all cases. It is also seen that percentage of decrement in RMSE and percentages of increments in MI and NCC by the proposed method are higher than those by global affine transformation method in all cases except for Figure 2.8 in Table 2.2. In general, there are notable improvements in the results by the proposed method however time-requirement is slightly more in this method.

Table 2.1: Comparison of RMSE (less is better) between source and reference images before and after registration by (i) Global Affine Transformation method and (ii) Proposed Method

Source and reference images	RMSE before Registration	RMSE after Registration		%-reduction in RMSE	
		by (i)	by (ii)	by (i)	by (ii)
2.4.B with 2.4.A	21.746	17.815	16.883	18.07	22.36
2.4.C with 2.4.A	22.970	22.047	17.534	4.01	23.66
2.5.B with 2.5.A	18.469	17.313	15.753	6.25	14.70
2.6.B with 2.6.A	46.793	40.599	33.203	13.23	29.04
2.7.B with 2.7.A	25.722	18.673	14.840	27.40	42.30
2.8.B with 2.8.A	46.519	39.084	37.800	15.98	18.74

Table 2.2: Comparison of MI (more is better) between source and reference images before and after registration by (i) Global Affine Transformation method and (ii) Proposed Method

Source and reference images	MI before Registration	MI after Registration		%-increment in MI	
		by (i)	by (ii)	by (i)	by (ii)
2.4.B with 2.4.A	1.285	1.447	1.487	12.60	15.71
2.4.C with 2.4.A	0.999	1.150	1.329	15.11	33.03
2.5.B with 2.5.A	1.463	1.550	1.594	5.94	8.95
2.6.B with 2.6.A	0.398	0.593	0.708	48.99	77.88
2.7.B with 2.7.A	0.866	1.078	1.311	24.48	51.38
2.8.B with 2.8.A	0.393	0.509	0.469	29.51	19.33

Table 2.3: Comparison of NCC (more is better) between source and reference images before and after registration by (i) Global Affine Transformation method and (ii) Proposed Method

Source and reference images	NCC before Registration	NCC after Registration		%-increment in NCC	
		by (i)	by (ii)	by (i)	by (ii)
2.4.B with 2.4.A	0.903	0.934	0.941	3.43	4.20
2.4.C with 2.4.A	0.892	0.899	0.936	0.78	4.93
2.5.B with 2.5.A	0.918	0.918	0.940	0.00	2.39
2.6.B with 2.6.A	0.588	0.704	0.800	19.72	36.05
2.7.B with 2.7.A	0.790	0.886	0.935	12.15	18.35
2.8.B with 2.8.A	0.402	0.575	0.586	43.03	45.77

2.4 Summary

In this chapter we have proposed an iterative method for registration of multi-focus images by combining global and local transformation models. It is automatic, easy to implement and gives good results. It does not require any manual intervention for feature or ground control point (GCP) selection. Different sets of multi-focus images are registered by the proposed method. We have compared the performance of the method with global affine transformation method in respect of quantitative measures RMSE, MI, NCC and time-requirement. The proposed method is found better than the other method in respect of RMSE, MI, NCC although the time-requirement is more in the proposed method. Generally, registration techniques require interpolation and use of a particular interpolation method affects the results in various ways [92, 67]. In the first step of the proposed algorithm, no interpolation is required since the grids of source and reference images overlay perfectly. In the second step of the algorithm, interpolation is required. Bilinear interpolation is a reasonable choice in terms of ease-of-implementation and time-complexity. But during successive itera-

tions it may reduce the contrast of the images. A higher-order interpolation like bi-cubic interpolation is a better choice in that respect although it takes more time [13]. To reduce the time-requirement, bilinear interpolation is used while estimating the scale-factor for a block, and at the end of the local-scaling step, when a block is reconstructed to be a part of the resultant registered image, bi-cubic interpolation is applied. The proposed image registration technique can be extended for registration of multi-focus color images in RGB format. For this, at first each color image is to be converted to grey-level image by computing the grey-level intensity at each pixel by as $(R+G+B)/3$. Then the global and local transformations are to be determined by applying the proposed technique on the grey-level images. Then the transformations are to be applied on the corresponding color images.

Chapter 3

Pixel-based fusion

3.1 Introduction

Pixel-based multi-focus image fusion techniques concentrate on individual pixels of the images and work either in spatial domain or in frequency/transform domain. In spatial domain techniques, input images are fused in the spatial domain using physically relevant spatial features in localized area. Since they emphasize on a specific or desired image area, very little/no change occurs in other areas. Pixel-level weighted averaging is a spatial domain technique in which fusion is done by taking the weighted average of the pixel intensities of the source images. Weights are determined by tools like principle component analysis [71] or adaptive methods [42]. Weighted averaging often has serious side effects like reduction in the contrast of the fused image. Other spatial domain pixel-level image fusion approaches include, fusion using controllable camera [73], probabilistic methods [5], image gradient method with majority filtering [23]. The method described in [73] depends on controlled camera motion and does not work for arbitrary sets of images. Probabilistic techniques [5] involve huge computation using floating point arithmetic and thus requires a lot of time and memory-space. Image gradient method with majority filtering [23] has the drawback that the defocused zone of one image is enhanced at the expense of focused

zone of others.

An image often contains physically relevant features at many different scales or resolutions. Multi-resolution (MR) techniques for image fusion explore and use that fact. They work either in spatial domain or in frequency/transform domain. A. A. Goshtasby and S. Nikolov [30] pointed out that although pixel-level fusion is a local operation, transform domain algorithms create the fused image globally. By changing a single coefficient in the transformed fused image, all (or a whole neighborhood of) image values in the spatial domain will change. An early categorization of frequency domain MR fusion techniques is found in [89]. The basic idea of these methods is to decompose each source image by an MR transform, then to construct a composite representation from the transformed images and finally to obtain the fused image by applying the inverse transform. The techniques vary in their choice for MR decomposition scheme and in their choice for coefficient-selection rule for making the composite representation. Popular MR decomposition schemes are pyramid transform and wavelet transform. The idea of using MR schemes for image fusion was first proposed by Burt [8] as a model for binocular fusion for human stereo vision. He used *Laplacian pyramid* for MR decomposition and *choose max* rule for coefficient selection. Burt and Adelson [10] later introduced a new approach to image fusion based on *hierarchical image decomposition.* Adelson [2] then used the *Laplacian pyramid* technique for multi-focus image fusion. Toet [78] proposed the use of *ratio of low-pass pyramids at successive levels of Gaussian pyramids* for fusion of visible and IR images. Burt [9] proposed that fusion within a *gradient pyramid* provide improved stability and noise immunity. Akerman [3] optimized the *Laplacian pyramid fusion* in respect of multi-sensor fusion. Burt and Kolczynski [11] presented *gradient pyramid fusion* with a local match measure and a window-based saliency measure. Ranchin and Wald [69] presented one of the first fusion schemes using *wavelet transform.* In their much-referred work, Li *et al.* [45] presented fusion schemes using wavelet transform. In their implementation, the preliminary decision map is generated by window-based activity measure which is then finalized by consistency verification with majority filter. *Wavelet transform* is also considered by Chipman *et al.* [16], Petrovic and Xy-

deas [63], Scheunders [72], Yang *et al.* [86] and Hill *et al.* [37]. Piella [65] provided a general framework for MR image fusion and also proposed a new method for the same. A comprehensive tutorial on wavelet-based fusion methods is found in [59].

Recently, wavelets are considered for image fusion by Wang [81], Wang *et al.* [82], Hamza *et al.* [34], Qu and Yan [68], Lewis *et al.* [44], Tsai and Lee [79] and Yang and Li [85]. Wang [81] presented a pixel-based algorithm employing a multi-wavelet (which is an extension of scalar wavelet) transform with two wavelet functions and two scaling functions. A feature-based fusion rule is used to combine original sub-images. Mutual information is employed for objective evaluation of fusion performance. In the method proposed by Wang *et al.* [82], after decomposition of input images by wavelet transform, images at the lowest resolution are segmented into regions by watershed algorithm. Wavelet-coefficients are tested region-wise for activity-level measurement and match-degree measure. Scaled and detail coefficients are combined respectively by choose-max and weighted average rule. Combined coefficients are inversely transformed to get the final fused image. In the method of Hamza *et al.* [34], input images are decomposed by a bi-orthogonal wavelet transform. Selection-map is created by measuring the activity-level of each wavelet coefficient by Jensen-Renyi divergence. The map is refined into two decision regions according to a threshold. Scaled and detail coefficients are combined either by choose-max and or by weighted average rule depending on the region they belong to. Finally combined coefficients are inversely transformed to get the final fused image. In Qu and Yan's [68] method, after decomposition of input images by a discrete wavelet transform, a pulse-coupled neural network is employed to extract features of the input images in the wavelet domain. Regional firing intensity characteristic is computed and used to combine the coefficients. Finally combined coefficients are inversely transformed to get the final fused image. Lewis *et al.* [44] considered a dual-tree complex wavelet transform (DT-CWT) for segmenting of the features of the input images either jointly or separately to produce a region map. The images are then fused region-wise in the wavelet domain. Tsai and Lee [79] presented a method in which after segmentation of input images into regions, quality of a region is measured from low frequency wavelet bands by adaptive

decomposition algorithm. Then regions with better quality are selected to produce the fused image. Yang and Li [85] proposed a method in which the source images are represented at first by sparse coefficients using an over-complete dictionary. The coefficients are then combined by choose-max fusion rule. Finally the fused image is reconstructed from the combined sparse coefficients and the dictionary.

Wavelet transform can be considered as a special case of pyramid transform but it has more complete theoretical support [51]. One major advantage of wavelet transform is that *spatial* as well as *frequency* domain localization of an image is obtained simultaneously. Another advantage is that it can provide information on sharp contrast changes, and human visual system is especially sensitive to these changes. Wavelet transform is a linear tool in its original form [51]. But *non-linear* extensions of discrete wavelet transform are possible by methods like *lifting scheme* [77] or *morphologic operators* [31, 35]. The problem with linear wavelets like Haar wavelet is that during signal decomposition or analysis the range of the original data is not preserved [35]. Secondly, linear wavelets act as low-pass filters and thus smooth-out the edges. This results in reduction in the contrast in fused images. The nonlinear wavelet introduced by Heijmans and Goutsias [35] overcomes this drawback by using morphologic operators. But it involves division operation and thus either requires floating point arithmetic or introduces truncation error by using integer arithmetic.

In this chapter we present a *nonlinear morphologic wavelet transform* [18, 19] which preserves the range in the scaled images and involves integer arithmetic only. We then use this transform to present a fusion algorithm to fuse a set of grey-scale multi-focus images. The method is computationally efficient and produces good results. Integrated-chip implementations of image processing algorithms are going to become more common in near future. Our method will be useful in this respect. The results obtained by it have been compared with those obtained by using Haar wavelet and the morphologic wavelet suggested by Heijmans and Goutsias [35]. The chapter is organized as follows. Section 3.2 gives the basic theory (without proof) of multi-resolution analysis using wavelets and a brief discussion on morphologic operators. This section

also introduces the proposed wavelet transform based on these operators. Section 3.3 describes the image-fusion algorithm using the new morphologic wavelet. Experimental results and discussion are given in Section 3.4 and the concluding remarks are presented in Section 3.5.

3.2 Basic theory and a new morphologic wavelet

A brief overview of multi-resolution signal decomposition theory using wavelets is given first, followed by the discussion on morphologic operators, and finally a new wavelet transform based on these operators is presented.

3.2.1 Multi-resolution Analysis

The theory of multi-resolution signal decomposition scheme using wavelets can be applied to a wide variety of signals. We are restricted here to two-dimensional grey-scale image signals only. A two-dimensional grey-scale image signal X is a mapping from domain D (which is a subset of discrete two-dimensional space Z^2) to the set of grey values $\{g_1, g_2, \ldots, g_n\}$ where each g_i is a nonnegative integer. Let us consider a set V_0 of such image signals. A multi-resolution signal decomposition scheme on V_0 uses two types of operators, namely, *signal analysis* and *signal synthesis* operators; which are also known as *scaling function* and the *wavelet function* respectively. *Signal analysis* operators $\psi_j^{\uparrow} : V_j \rightarrow V_{j+1}$, map the signal space V_j at level j, to a coarser signal space V_{j+1} and the *detail analysis* operators $\omega_j^{\uparrow} : V_j \rightarrow W_{j+1}$, map V_j to a coarser detail space W_{j+1}. All V_j's and W_j's have the same structure as V_0. Signal analysis operation proceeds by mapping a signal to a level higher in the pyramid structure, thereby reducing information. Details are stored at each level to restore this information loss. If analysis operators are applied j times recursively on an image signal $X \in V_0$, scaled and detail signals at level j are denoted by X^j and Y^j, where

$X^j \in V_j$ and $Y^j \in W_j$. Then we have

$$\psi_j^{\uparrow}(X^j) = X^{j+1},\ X^{j+1} \in V_{j+1} \tag{3.1}$$

$$\omega_j^{\uparrow}(X^j) = Y^{j+1},\ Y^{j+1} \in W_{j+1} \tag{3.2}$$

Signal synthesis or reconstruction is done by *synthesis operator* $\psi_j^{\downarrow} : V_{j+1} \times W_{j+1} \to V_j$, which map a signal to a level lower in the pyramid. To ensure *loss-less* or *perfect reconstruction*, the following condition must be satisfied.

$$\psi_j^{\downarrow}(\psi_j^{\uparrow}(X^j), \omega_j^{\uparrow}(X^j)) = X^j,\ \ X^j \in V_j \tag{3.3}$$

There are two more conditions, namely,

$$\psi_j^{\uparrow}(\psi_j^{\downarrow}(X^{j+1}, Y^{j+1})) = X^{j+1} \tag{3.4}$$

$$\omega_j^{\uparrow}(\psi_j^{\downarrow}(X^{j+1}, Y^{j+1})) = Y^{j+1} \tag{3.5}$$

where $X^{j+1} \in V_{j+1}$ and $Y^{j+1} \in W_{j+1}$. They ensure that the decomposition is *non-redundant* in the sense that repeated applications of these schemes produce the same result. A special case called *uncoupled* wavelet decomposition occurs when there exists a binary operation $\dot{+}$ on V_j and operators $\psi_j^{\downarrow} : V_{j+1} \to V_j$ and $\omega_j^{\downarrow} : W_{j+1} \to V_j$ such that

$$\psi_j^{\downarrow}(X^{j+1}, Y^{j+1}) = \psi_j^{\downarrow}(X^{j+1}) \dot{+} \omega_j^{\downarrow}(Y^{j+1}),\ X^{j+1} \in V_{j+1},\ Y^{j+1} \in W_{j+1} \tag{3.6}$$

Then perfect reconstruction and non-redundancy conditions become

$$\psi_j^{\downarrow}\psi_j^{\uparrow}(X^j) \dot{+} \omega_j^{\downarrow}\omega_j^{\uparrow}(X^j) = X^j,\ X^j \in V_j \tag{3.7}$$

$$\psi_j^{\uparrow}(\psi_j^{\downarrow}(X^{j+1}) \dot{+} \omega_j^{\downarrow}(Y^{j+1})) = X^{j+1},\ X^{j+1} \in V_{j+1},\ Y^{j+1} \in W_{j+1} \tag{3.8}$$

$$\omega_j^{\uparrow}(\psi_j^{\downarrow}(X^{j+1}) \dot{+} \omega_j^{\downarrow}(Y^{j+1})) = Y^{j+1},\ X^{j+1} \in V_{j+1},\ Y^{j+1} \in W_{j+1} \tag{3.9}$$

If a one-dimensional wavelet decomposition scheme can be applied to two and higher dimensions, by applying it to other dimensions sequentially, then this decomposition is called *separable*. A new wavelet transform based on morphologic operators is presented now.

3.2.2 A new morphologic Wavelet

Heijmans and Goutsias introduced a *morphologic variant* of the Haar wavelet by using the morphologic operation *dilation* (*erosion*) [35]. It is an one-dimensional scheme and the multidimensional implementation can be obtained by applying it to other dimensions sequentially. However, a two-dimensional non-separable version of the morphologic Haar wavelet transform has also been defined in [35], which will be used in our experiments for comparison purpose. We, now propose a *non-separable two-dimensional uncoupled morphologic wavelet* decomposition scheme, which will be used for our image-fusion algorithm. Unique analysis operators $(\psi^{\uparrow}, \omega^{\uparrow})$ are used at all levels of the multi-resolution scheme. Similarly, unique synthesis operators $(\psi^{\downarrow}, \omega^{\downarrow})$ are used at all levels. These operators are explained for the lowermost levels 0 and 1.

Let us consider the signal space V_0 of Section 3.2.1. It is our original signal space. Then V_1 and W_1 are the signal and detail spaces at level 1 having the same structure as V_0. Consider an image signal $X \in V_0$. Then X is a mapping of (a subset of) Z^2 to the set of grey-values G and it can be represented by an $M \times N$ matrix, where $M, N \in Z$. Let us assume that M and N both are even. Then X can be divided into consecutive and disjoint 2×2 sub-matrices or blocks, which are total $\frac{MN}{4}$ in number. Four positions of such a block B may be denoted by (r, c), $(r, c+1)$, $(r+1, c)$ and $(r+1, c+1)$ (see Figure 3.1) where r and c denote row and column positions of the image-matrix X. Using quadratic downsampling, the analysis operators $\psi^{\uparrow} : V_0 \rightarrow V_1$ and $\omega^{\uparrow} : V_0 \rightarrow W_1$ are defined as

$$\psi^{\uparrow}(X)(B) = \max\{X(r,c),\, X(r,c+1),\, X(r+1,c),\, X(r+1,c+1)\} \quad (3.10)$$

$$\omega^{\uparrow}(X)(B) = (\, y_v,\, y_h,\, y_d \,) \quad (3.11)$$

where y_v, y_h, y_d represent the *vertical, horizontal* and *diagonal* detail signals respec-

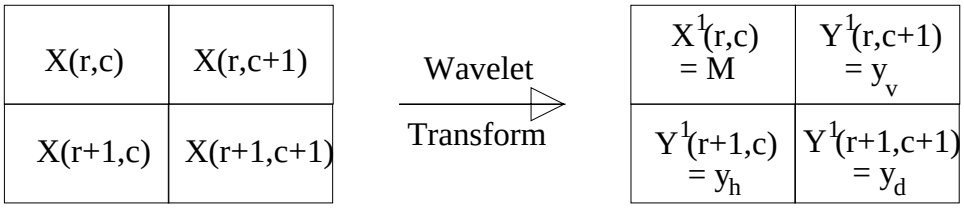


Figure 3.1: Wavelet transform on a 2×2 block

tively. Let $\psi^{\uparrow}(X)(B)$ be denoted by M. Then y_v, y_h, y_d is defined as

$$y_v = \begin{cases} M - X(r, c+1), \text{ if } M - X(r, c+1) \geq 0 \\ X(r, c+1) - M, \text{ otherwise} \end{cases} \tag{3.12}$$

$$y_h = \begin{cases} M - X(r+1, c), \text{ if } M - X(r+1, c) \geq 0 \\ X(r+1, c) - M, \text{ otherwise} \end{cases} \tag{3.13}$$

$$y_d = \begin{cases} M - X(r+1, c+1), \text{ if } M - X(r+1, c+1) \geq 0 \\ X(r+1, c+1) - M, \text{ otherwise} \end{cases} \tag{3.14}$$

The second condition in the last three equations is required to maintain the information on position of the maximum value M as evidenced in the successive example. Scaled signal and detail values obtained above belong to X^1 and Y^1 respectively and they can be stored conveniently in similar positions of another matrix.

The original signal at level 0 is reconstructed by the synthesis operation. Using quadratic upsampling, *synthesized signals* $\widehat{X}$ are given by

$$\widehat{X}(r, c) = \widehat{X}(r, c+1) = \widehat{X}(r+1, c) = \widehat{X}(r+1, c+1) = M \tag{3.15}$$

and *synthesized details* $\widehat{Y}$ are given by

$$\widehat{Y}(r, c) = \min(y_v, y_h, y_d, 0) \tag{3.16}$$

$$\widehat{Y}(r, c+1) = \min(-y_v, 0) \tag{3.17}$$

$$\widehat{Y}(r+1, c) = \min(-y_h, 0) \tag{3.18}$$

$$\widehat{Y}(r+1, c+1) = \min(-y_d, 0) \tag{3.19}$$

where $M = X^1(r, c)$ is the scaled signal at (r, c) and y_v, y_h, y_d are vertical, horizontal and diagonal details respectively. This is an uncoupled decomposition scheme and

the binary operation $\dot{+}$ is the ordinary addition of numbers. Hence the reconstructed signal X' at any point $(u, v) \in \{(r, c), (r, c+1), (r+1, c), (r+1, c+1)\}$ is given by

$$X'(u, v) = \widehat{X}(u, v) \dot{+} \widehat{Y}(u, v) \tag{3.20}$$

Example: Let us consider the 2×2 block B of X with $X(r, c) = T_0$, $X(r, c+1) = T_1$, $X(r+1, c) = T_2$ and $X(r+1, c+1) = T_3$. Let $T_m = \max\{T_0, T_1, T_2, T_3\}$. Then $\psi^{\uparrow}(X)(B) = T_m$ and the details are given by $\omega^{\uparrow}(X)(B) = (T_v, T_h, T_d)$ where

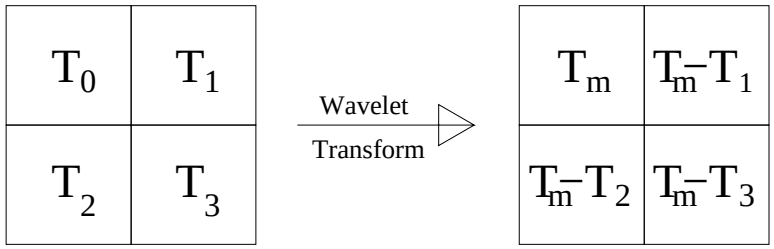
$$T_v = \begin{cases} T_m - T_1, & \text{if } T_m - T_1 > 0 \\ T_0 - T_m, & \text{otherwise} \end{cases}$$

$$T_h = \begin{cases} T_m - T_2, & \text{if } T_m - T_2 > 0 \\ T_0 - T_m, & \text{otherwise} \end{cases}$$

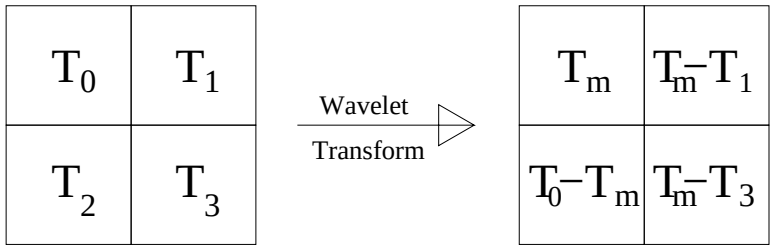
$$T_d = \begin{cases} T_m - T_3, & \text{if } T_m - T_3 > 0 \\ T_0 - T_m, & \text{otherwise} \end{cases}$$

Now T_m may occur at any of the four positions of the block 2×2 submatrix. The situations of T_m occurring at (r, c) and $(r+1, c)$ are illustrated in the figure 3.2. In the first case T_m occurs at position (r, c) and all the detail values are positive. In the second case T_m occurs at position $(r+1, c)$ and the information is preserved by placing the negative value $T_0 - T_m$ as the horizontal detail.

The analysis operator-pair $(\psi_j^{\uparrow}, \omega_j^{\uparrow})$ can be used recursively to decompose a signal up to a desired level $k \geq 1$. Similarly the synthesis operator-pair $(\psi_j^{\downarrow}, \omega_j^{\downarrow})$ can be used recursively to reconstruct a signal from any level to the lowest level 0. It is easy to see that the analysis and synthesis operators satisfy the perfect reconstruction and non-redundancy conditions 3.7 - 3.9 given in Section 3.2.1. The operators $\psi^{\uparrow}$ and $\omega^{\uparrow}$ involve elementary arithmetic operations and one interesting point to note is that the integer values are mapped to integer values only. Another point to note is that, if all values of X belong to the range $[0, R]$, then analyzed signal-values will belong to the range $[0, R]$ and analyzed detail-values will belong to the range $[-R, R]$, irrespective of the number of times the operators are applied [see figure 3.3].



Case 1: Transform when T_0 is maximum

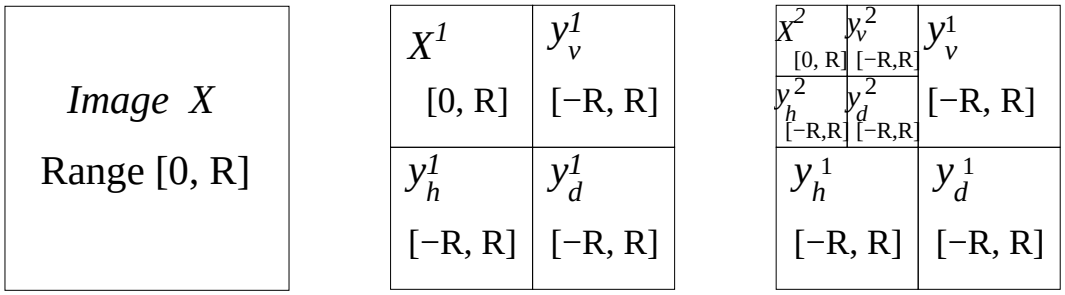


Case 2: Transform when T_2 is maximum

Figure 3.2: Illustration of proposed wavelet transform on a 2×2 block

3.3 Multi-focus image fusion

We now present the image fusion algorithm proposed by us using the morphologic wavelet transform given in Section 3.2.2. Consider n two-dimensional multi-focus images $X_1, X_2, \ldots, X_n$. These images must be registered and of the same size. The proposed analysis operators $\psi^{\uparrow}$ and $\omega^{\uparrow}$, are applied on the n individual images k times recursively. If X_i, $i = 1, 2, ..., n$ are $M \times N$ images, the analysis operators can be applied at most k_{max} times where $k_{max} = \min(\lfloor \log_2 M \rfloor, \lfloor \log_2 N \rfloor)$. After completion of the analysis operation, at the topmost level k, a set of n *scaled images* are obtained. They are denoted by X_i^k, $i = 1, 2, ...n$. A set of *detail images* Y_i^j, $i = 1, 2, ...n$ are also obtained at each level j, $j = 1$ to k. As mentioned in the last section, if the range of greylevels in image X_i is $[0, R]$, then that of the scaled images X_i^k is $[0, R]$ and that of the detail images Y_i^j, $j = 1, 2, ...k$ is $[-R, R]$. While comparing X_i^k, $i = 1, 2, ...n$ position-wise, a higher absolute value corresponds to a brighter pixel and while comparing Y_i^j, $j = 1, 2, ...k$, $i = 1, 2, ...n$ position-wise, a higher absolute value corresponds to sharp-contrast features such as edge, line and region boundaries.



(a) (b) (c)

Figure 3.3: (a) Original signal X, (b) Scaled signal X^1 and details $Y^1 = \{y_v^1, y_h^1, y_d^1\}$ at level 1, (c) Scaled signal X^2 and details $Y^2 = \{y_v^2, y_h^2, y_d^2\}$ at level 2

Based on this observation, scaled images X_i^k, $i = 1, 2, ...n$ are combined by comparing the values at each position (r, c) and choosing the one with the greatest absolute value. Similar operation is applied on corresponding detail images at each level. Thus a single fused image at level k and a detail image at each level j, $j = 1, ..., k$ are obtained. Then the reconstruction phase begins. The image at level $k-1$ is reconstructed by applying the synthesis operators $\psi^{\downarrow}$ and $\omega^{\downarrow}$ (as proposed by us in the previous section) followed by addition. Synthesis operators are applied k times recursively to obtain the image at original domain i.e. at level 0. The algorithm can be summarized as below.

3.3.1 Algorithm

1. **Analysis step:** Apply the analysis operators $\psi^{\uparrow}$ and $\omega^{\uparrow}$, k times recursively, on image X_i, $i = 1, \ldots, n$ and get $\mathbf{X_i} = \{ X_i^k, Y_i^1, Y_i^2, \ldots, Y_i^k \}$, where X_i^k is the scaled image at level k and Y_i^j, $j = 1, \ldots, k$ are the details at levels $1, 2, \ldots, k$ respectively.

2. **Fusion step:** Compare $\{ \mathbf{X_i}, i = 1, 2, \ldots, n \}$ and combine them into $\mathbf{X} = \{X^k, Y^1, Y^2, \ldots, Y^k\}$, where X^k and Y^j are respectively given by

$$\begin{aligned} X^k(r,c) &= \max \{|X_1^k(r,c)|, |X_2^k(r,c)|, \ldots, |X_n^k(r,c)| \} \text{ and} \\ Y^j(r,c) &= \max \{|Y_1^j(r,c)|, |Y_2^j(r,c)|, \ldots, |Y_n^j(r,c)| \} \end{aligned}$$

3. Synthesis step: Reconstruct the fused image X^j at level j, $j = k-1, \ldots, 0$, by applying the synthesis operators $\psi^{\downarrow}$ and $\omega^{\downarrow}$ respectively on X^{j+1} and Y^{j+1} following by addition, i.e.

$$X^j(r,c) \;=\; \psi^{\downarrow}(X^{j+1}(r,c)) + \omega^{\downarrow}(Y^{j+1}(r,c))$$

3.3.2 Illustration

The algorithm is illustrated by using 2×2 sample data A and B taken from the multi-focus images X_1 and X_2 respectively.

Let $A = \begin{bmatrix} a_0 & a_1 \\ a_2 & a_3 \end{bmatrix}$ and $B = \begin{bmatrix} b_0 & b_1 \\ b_2 & b_3 \end{bmatrix}$ where a_i and b_i, $i = 0, 1, 2, 3$ are non-negative integers. Applying the analysis operators $\psi^{\uparrow}$ and $\omega^{\uparrow}$ once, A becomes,

$$A^1 = \begin{bmatrix} a_0^1 & a_1^1 \\ a_2^1 & a_3^1 \end{bmatrix}$$

where

$$a_0^1 = a_{\max} = \max\{a_i,\ i = 0, 1, 2, 3\} \text{ and}$$

$$a_i^1 = \begin{cases} a_{\max} - a_i & \text{if } a_{\max} > a_i \\ -(a_{\max} - a_0) & \text{otherwise} \end{cases}$$

Here a_0^1 is the scaled signal-data and a_i^1, $i = 1, 2, 3$ are the detail-data at level 1.

Similarly, after the analysis operation, B becomes,

$$B^1 = \begin{bmatrix} b_0^1 & b_1^1 \\ b_2^1 & b_3^1 \end{bmatrix}$$

A^1 and B^1 are fused in C^1, by the fusion step, where

$$C^1 = \begin{bmatrix} c_0^1 & c_1^1 \\ c_2^1 & c_3^1 \end{bmatrix} \text{ and } c_i^1 = \begin{cases} a_i^1 & \text{if } |a_i^1| \geq |b_i^1| \\ b_i^1 & \text{otherwise} \end{cases} \quad \text{for } i = 0, 1, 2, 3$$

The fused data C at level 0 is obtained by applying the synthesis operators $\psi^{\downarrow}$ and $\omega^{\downarrow}$ followed by addition. Therefore

$$C = \begin{bmatrix} c_0 & c_1 \\ c_2 & c_3 \end{bmatrix}$$

where

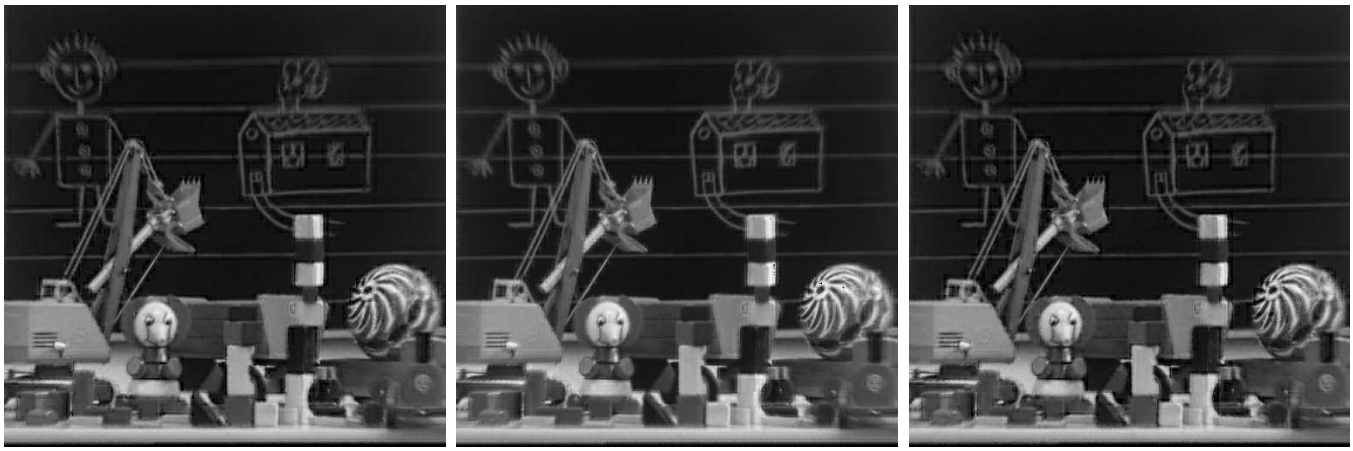
$$c_0 = c_0^1 + \min(0,\, c_1^1,\, c_2^1,\, c_3^1) \text{ and } c_i = \begin{cases} c_0^1 & \text{if } c_i^1 < 0 \\ c_0^1 - c_i^1 & \text{otherwise} \end{cases} \quad \text{for } i = 1, 2, 3$$

We now, claim that, c_i is always less or equal to R, where R is the greatest value of a_i and b_i, $i = 0, 1, 2, 3$. This happens because c_i is obtained by subtracting a non-negative value from $c_0^1 = \max(a_0^1,\, b_0^1)$. However the lower bound of c_i may not remain within the lower bounds of A and B. The method can be applied to the complete images X_1 and X_2 by taking as many 2×2 samples as required.

3.4 Experimental results and discussion

The proposed fusion algorithm was tested on a large number of input images. For comparison purpose, fusion with Haar wavelet and with two-dimensional morphologic wavelet introduced by Heijmans and Goutsias [35] were also implemented. Fusion results for multi-focus image-sets shown in Fig. 1.2 are given in Fig. 3.4. For each input image-set three fused images are shown; the first one is obtained by the proposed wavelet, the second one is obtained by Haar wavelet and the third one is obtained by morphologic wavelet introduced by Heijmans and Goutsias. The fusion is done by decomposing the constituent images up to the third level, in all the cases. Quantitative evaluations by gradient-similarity-index (GSI) and fusion-quality-index (FQI) as explained in Section 1.7 are given respectively in tables 3.1 and 3.2. Note that for both quantitative metrics GSI and FQI, higher the value better is the fusion. Time required in seconds for pixel-based algorithms using different wavelets are given in Table 3.3.

A.(i) A.(ii) A.(iii)



B.(i) B.(ii) B.(iii)

C.(i) C.(ii) C.(iii)

Figure 3.4: Results of multi-focus image fusion by pixel-based methods. In each row, images shown are obtained by applying the proposed algorithm respectively with (i) morphologic wavelet proposed by us, (ii) Haar wavelet and (iii) non-linear wavelet proposed by Heijmans and Goutsias

Table 3.1: Performance evaluation of pixel-based methods by GSI

Figure	Proposed wavelet	Haar wavelet	Heijmans' wavelet
Doll	0.855	0.839	0.847
Toy	0.819	0.830	0.819
Disk	0.873	0.789	0.870
Lab	0.865	0.832	0.864
Pepsi	0.927	0.923	0.928
Clock	0.865	0.890	0.866
Campus	0.794	0.808	0.790
Hydrant	0.869	0.833	0.864
Garden	0.786	0.763	0.778
Rose	0.848	0.847	0.846
News	0.906	0.924	0.902
OpenGL	0.902	0.841	0.902
Average	0.859	0.843	0.856

3.4.1 Discussion

Careful manual inspection of fused images in figure 3.4 reveals that the results obtained by the proposed wavelet are better than that of Haar wavelet and are comparable to that of Heijmans and Goutsias' wavelet [35]. However, artifacts such as blocking effects are noticed in some of the fused images. But this is a common phenomena in all pixel-based image fusion using multi-resolution approach and happens due to the fact that error introduced at the topmost level is amplified during reconstruction [45]. In our case, these effects are found in border regions and in in regions where the data is out of focus in all the source images. For example, one can find such effects along the edges of the clock in Figure 3.4:C and in the middle-right portion in Figure 3.4:A. However, these effects are present in the fused images obtained by the

D.(i) D.(ii) D.(iii)

E.(i) E.(ii) E.(iii)

F.(i) F.(ii) F.(iii)

Figure 3.4: Continued

G.(i)

G.(ii)

G.(iii)

H.(i)

H.(ii)

H.(iii)

I.(i)

I.(ii)

I.(iii)

Figure 3.4: Continued

J.(i)

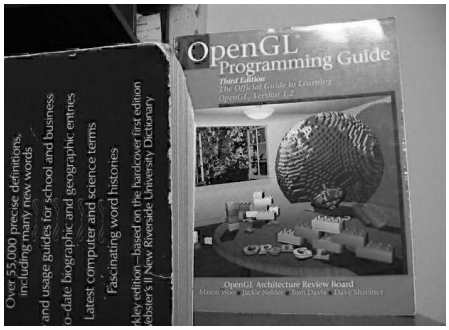
J.(ii)

J.(iii)

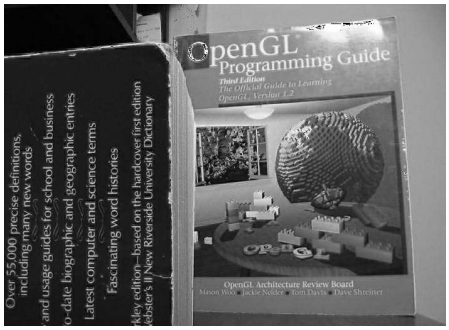
K.(i)

K.(ii)

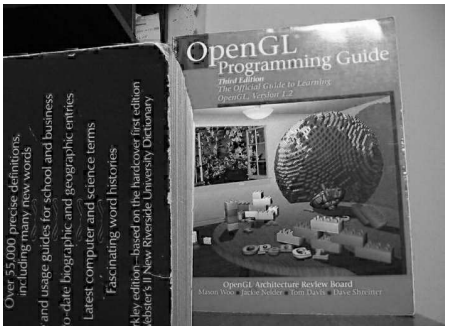
K.(iii)



L.(i)



L.(ii)



L.(iii)

Figure 3.4: Continued

Table 3.2: Performance evaluation of pixel-based methods by FQI

Figure	Proposed wavelet	Haar wavelet	Heijmans' wavelet
Doll	0.805	0.779	0.791
Toy	0.775	0.783	0.760
Disk	0.853	0.851	0.847
Lab	0.851	0.849	0.845
Pepsi	0.869	0.877	0.863
Clock	0.884	0.880	0.882
Campus	0.895	0.875	0.880
Hydrant	0.877	0.875	0.867
Garden	0.791	0.783	0.781
Rose	0.856	0.861	0.849
News	0.867	0.872	0.857
OpenGL	0.857	0.828	0.837
Average	0.848	0.843	0.838

other two wavelets as well. In addition to this, one can find small black spots in fused images obtained by by Haar wavelet method (for example, along the edge of the clock in Figure 3.4:C.(ii)). This happens because fusion by Haar wavelet method generates negative pixel values after reconstruction which are truncated at value zero. A general problem faced by pixel-based methods is sensitivity to mis-registration. Input images 'Garden' as shown in Fig. 1.2:J are dense in texture and difficult to register because being images of outdoor scenes, temporal changes due to wind are present in them along with focus changes. Fused images as shown in Figure 3.4:J are not very good and illustrate the fact that pixel-based methods are sensitive to even slight mis-registration. Table 3.3 shows that the time taken by various wavelets are more or less same and are not significant.

Apart from the quality of the results, the proposed algorithm has some computational

Table 3.3: Time requirement in pixel-based methods

Figure	Proposed 2D morphologic wavelet (seconds)	Haar wavelet (seconds)	Heijmans' morphologic wavelet (seconds)
Doll	10	8	10
Toy	13	11	14
Disk	7	6	9
Lab	10	12	15
Pepsi	11	9	16
Clock	1	1	1
Campus	12	11	16
Hydrant	13	13	18
Garden	7	8	12
Rose	14	12	15
News	1	1	1
OpenGL	11	13	14
Average	9	9	12

advantages as well. Unlike two other wavelets experimented with, our method ensures that integer pixel values are mapped to integer values only during both analysis and synthesis. This is an useful property for lossless data compression [12]. Secondly, irrespective of the number of times the analysis operators are applied, the range of the values in the scaled images will be same as that of the original multi-focus images, say $[0, R]$, and the range of the detail values will be $[-R, R]$. Hence memory-space required during decomposition is fixed. Thirdly, arithmetic operations like addition, subtraction and comparison are the only operations used in the method. Other two methods involve division operation and thus they either requires floating point arithmetic or introduces truncation error by using integer arithmetic. Fourthly, due to the nonlinear nature of the proposed method, important geometric information

(e.g. edges) is well-preserved at lower resolutions. Finally, the method is very fast due to its simplicity. For a set of n, $M \times N$ images, it takes only $O(n \times M \times N)$ computational time. The simplicity of the method and the use of integer arithmetic makes it suitable for chip-level implementation.

Besides this, the nonlinear wavelet proposed by us possesses the following invariance properties. Both analysis and synthesis operators are translation invariant in the spatial domain. In the frequency domain, they are grey shift (multiplication) invariant. That means adding (multiplying) a certain value to all pixel values in the original data will result in adding (multiplying) that value to the scaled signal data during analysis [35]. Also, details will not change in case of addition and will get multiplied by that value in case of multiplication. The wavelets possessing these invariance properties, offer better option for image fusion than those which do not possess them [89].

3.5 Summary

In this chapter we have presented a non-linear wavelet constructed by morphologic operators and also presented a multi-focus image fusion algorithm based on that wavelet. The results are good considering the fact that the computational cost is very low. The use of elementary arithmetic operations makes the method suitable for hardware implementation. However the results may suffer from the problem of blocking effects around the edges and at regions where the data is out of focus in all the source images. Registration error may aggravate the problem. But this is a common problem for other methods experimented with in this chapter. Our method is definitely better than Haar wavelet method and is at par with Heijmans and Goutsias' wavelet method in this respect.

Chapter 4

Block-based fusion

4.1 Introduction

In this chapter we present a block-based method for multi-focus image fusion. Since multi-focus images of a scene are acquired with focus on complementary regions, focused regions in an image have more contrast than their defocused counter-parts in other images. *Focus-measure (FM)* is a quantity for evaluating the contrast or sharpness of a pixel, block or region [76, 38, 50] and can be used effectively for multi-focus image fusion. A focus-measure should possess certain desirable properties [38]. It should be independent of image content, monotonic with respect to blur, unimodal, robust to noise and it should have large variations in values with respect to the degree of blur and should have minimal computational complexity. Image variance, image gradients, image Laplacians, energy of image gradients (EOG), energy of image Laplacian (EOL) are traditional FM's employed and validated for applications like autofocusing [76]. Modified Laplacian (ML), Sum modified Laplacian (SML) are modifications of image Laplacian [58]. Spatial frequency (SF) and *Tenengrad* were later introduced as focus measures [25, 36]. Evaluation of various FM's in MFIF can be found in [38].

A number of block-based fusion methods are available in the literature which employ different focus-measures to distinguish between focused and defocused blocks. Li *et al.* [47] proposed a MFIF technique in which input images are divided into $m \times n$ blocks and better focused ones are selected (by measuring their SF) to produce an initial fused image and the final fused image is produced by majority filtering of the initial result. In a subsequent work [48] they proposed a neural network (NN) to select better focused blocks using three features *SF, visibility and edge feature*. Miao and Wang used EOG to measure focus in image blocks in an MFIF algorithm based on Pulse Coupled Neural Networks in [55]. In Goshtasby's method [29] focus is measured by sum of gradient values of all pixels in the block. Instead of just cutting and pasting the better-focused blocks, entire images are blended with weights that monotonically decrease from block-centers to smooth out the boundary between adjacent blocks. In the method presented by Fedorov *et al.* [27] each image is tiled with overlapping neighborhoods. For each region the tile that corresponds to the best focus is chosen. Selected tiles are seamlessly mosaicked by multi-resolution spline technique to construct the fused image. Zhang and Ge proposed a technique [87] in which focused blocks are detected by measuring their blurriness. Block-maps are created and small isolated blocks are removed. Finally fusion map is constructed and fusion is done accordingly. Block-based techniques presented in [47, 48, 55, 29, 27, 87] are sensitive to block-size. Li *et al.* mentioned that optimal block-size could be chosen by adaptive methods [48]. Goshtasby proposed to determine the optimal block-size by an iterative procedure which is time-consuming [29]. Fedorov *et al.* proposed to constrain the minimum tile-size by use of multi-resolution spline technique [27].

We present an efficient block-based algorithm for MFIF which is not sensitive to block-size. Although it starts with identifying the focused blocks, finally the focused regions in each input image are identified. Hence the results are comparable with region based methods. We also propose a new measure of focus *energy of morphologic gradients (EOMG)* and use it for our purpose. The paper is organized as follows. In Section 4.2, sub-section 4.2.1 describes the quad-tree based algorithm to detect focused blocks, 4.2.2 describes the reconstruction of connected regions and 4.2.3 de-

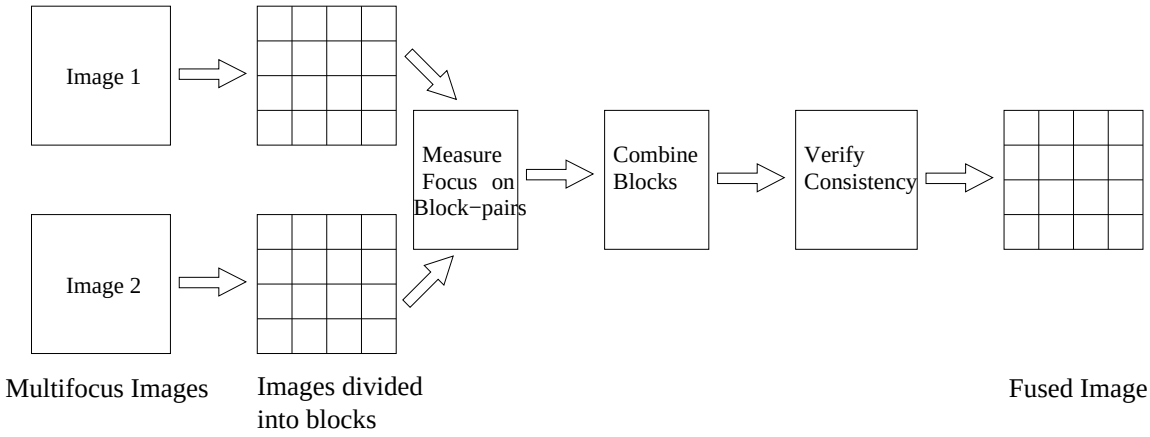


Figure 4.1: A generic schematic diagram for multi-focus image fusion by computing the focus measure on equal-sized blocks

scribes EOMG. Section 4.3 contains experimental results and discussion including performance analysis and finally 4.4 contains summary of the chapter and concluding remarks.

4.2 A new block-based fusion algorithm

A generic schematic diagram of block-based approach for MFIF is shown in Fig. 4.1. The number of blocks n plays a crucial role in this approach [48]. A small value of n means a large size for each block and a large block is more likely to contain portions from both focused and defocused regions. This may lead to selection of considerable amount of defocused regions. On the other hand, a large value of n means small size for each block. This too may lead to selection of some defocused blocks since the relative contrast do not vary much on small and relatively smooth regions. Moreover small blocks are more affected by mis-registration problems. The problem of choosing an ideal n is illustrated in Fig. 4.2. Suppose Image-1 and Image-2 are two multi-focus images of a scene, focus being on complementary regions. Focused regions are shown as shaded regions. If each image is divided into four quadrants, four corresponding block-pairs are created. From each pair, the one with better focus is chosen and

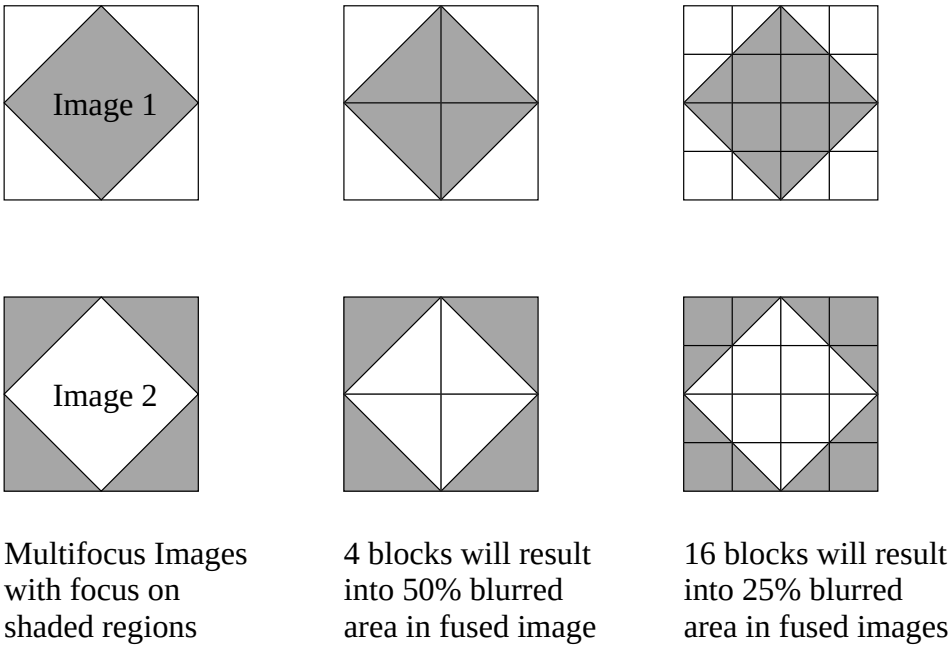


Figure 4.2: Problem of multi-focus image fusion with equal-sized blocks

copied to the resultant image. But in each pair both the blocks contains fifty percent out-of-focus area. So irrespective of which blocks are chosen, total fifty percent area in the fused image will be out of focus. In a similar way, if the images are divided into sixteen blocks each, twenty-five percent of the fused image will be out of focus. A common way to find n that generates the best result is by experimentation and verification with various values of n [48, 29, 27]. This requires a considerable amount of pre-processing time. To overcome these problems, we present an algorithm [21] which do not use fixed number of divisions in any portion of the multi-focus images. Rather it makes use of a quad-tree structure to obtain the optimal subdivision while measuring focus.

4.2.1 Detection of focused blocks in a quad-tree structure

The method is described for two input images and it can be extended easily to three or more input images. Henceforth the words block and node are used interchangeably. Two input images represent the root-nodes at the zero-th level of two quad-trees. Each

input image is divided into *four* quadrants to obtain a quadruple of nodes. Hence four pairs of corresponding blocks are obtained at level one of the quad-tree. For such a pair of blocks, any of the following situations may occur: one is fully focused and the other is fully out-of-focus, both are partially focused or both are fully out-of-focus. To find out the situation, focus-measure is computed on each block of the pair. *Normalized difference in focus-measure (NDFM)* between corresponding blocks is calculated as

$$NDFM = \frac{\text{Absolute difference in focus-measures}}{\text{Sum of focus-measures}} \tag{4.1}$$

NDFM is compared with a threshold then. In first situation, fully focused block has a considerably greater FM and NDFM is greater than the chosen threshold. So the blocks in the pair are not subdivided and the block with greater focus-measure (FM) is copied into the resultant image. In second and third situations, FM's do not vary much on corresponding pair of blocks and NDFM between them is less than or equal to the threshold. So both blocks in the pair are further subdivided into four quadrants. NDFM for corresponding smaller block-pairs are calculated and compared with a threshold again. They are further subdivided if required. The recursive subdivision is stopped if either the block-size becomes very small or NDFM is greater than the threshold at some level. The process is repeated for all four pairs of corresponding blocks obtained after the initial subdivision.

Generally images are combinations of textured and smooth regions. It is seen that variations in focus and hence values of NDFM are greatly influenced by texture and/or grains of the original images. In particular, variations in focus are greater in textured regions than in smooth regions. Values of NDFM obtained at the first two levels of the quad-trees give an initial idea about the distribution of texture/grains in the original images. It is also observed that NDFM between a pair of corresponding blocks at a level are influenced by their immediate ancestor blocks. This is because the former blocks are parts of the later ones. To decide on whether the NDFM is small enough to allow subdivision of a pair of blocks, a global threshold is not effective for blocks at all levels because their size and ancestors are different. The value of threshold for

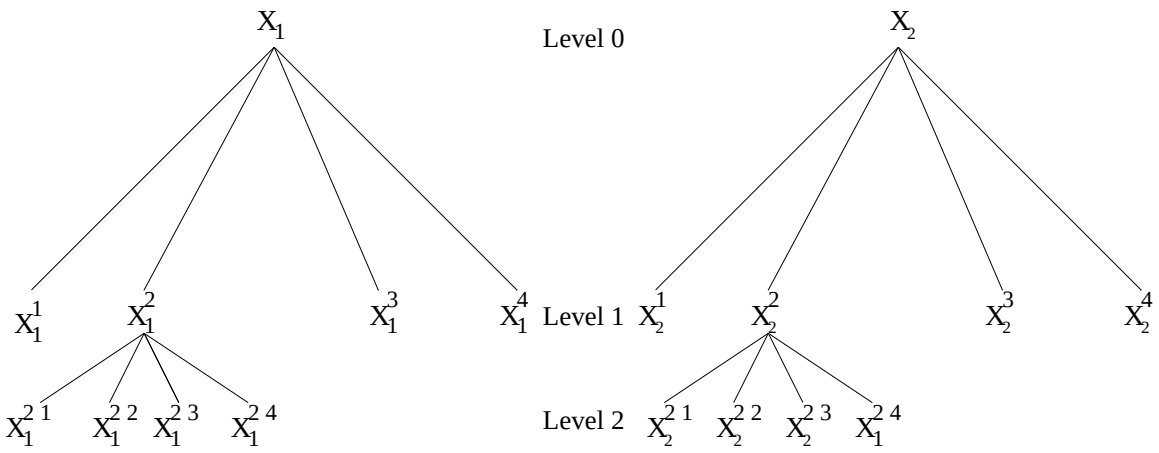


Figure 4.3: Subdivision of images in blocks according to a quad-tree structure

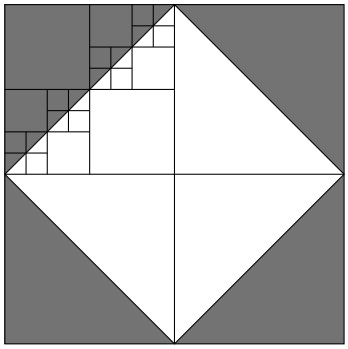
a pair of corresponding blocks at a level is calculated as a constant multiple of the NDFM at their immediate ancestor blocks. So it is dynamically updated for each quadruple of corresponding nodes at a level. In this work the constant multiplier M is calculated as

$$M = \frac{\text{Mean of NDFM at level 1}}{\text{NDFM at level 0}} \times \text{ Standard-deviation of NDFM at level 1} \times 100 \tag{4.2}$$

So it is constant for a set of input images and is derived from them only.

An example of subdivision of two input images in a quad-tree structure is illustrated in Fig. 4.3. Here X_1 and X_2 are input images at level zero. After initial subdivision, X_1^k and X_2^k $(k = 1, \ldots, 4)$ are corresponding pairs of blocks at level 1. NDFM between the root nodes is $\frac{|F(X_1)-F(X_2)|}{F(X_1)+F(X_2)}$. That between X_1^k and X_2^k is $\frac{|F(X_1^k)-F(X_2^k)|}{F(X_1^k)+F(X_2^k)}$ and the threshold for all of them is $T = M \times \frac{|F(X_1)-F(X_2)|}{F(X_1)+F(X_2)}$. At level 1, the second pair of blocks X_1^2 and X_2^2 are subdivided into smaller blocks to create block-pairs X_1^{2k} and X_2^{2k}, $k = 1, \ldots, 4$ at level 2. According to the theory developed, X_1^2 and X_2^2 are subdivided because $\frac{|F(X_1^2)-F(X_2^2)|}{F(X_1^2)+F(X_2^2)} \leq T$. X_1^{2k} and X_2^{2k} will be further subdivided if $\frac{|F(X_1^{2k})-F(X_2^{2k})|}{F(X_1^{2k})+F(X_2^{2k})} \leq M \times \frac{|F(X_1^2)-F(X_2^2)|}{F(X_1^2)+F(X_2^2)}$). Fig. 4.4 illustrates the recursive subdivision of top-left quadrant of *Image 2* in Fig. 4.2. Other quadrants will be subdivided similarly.

If the number of input images is m and $m > 2$, then four sets of m corresponding blocks are created after initial subdivision. For each of the m blocks in the set,



Focus is on shaded regions. After initial division
into four blocks, divide a block if necessary only.

Figure 4.4: Recursive subdivision of upper-left quadrant of Image 2 in Fig. 4.2

focus-measure is computed. Difference of the maximum and the minimum of these m measures divided by sum of these m measures is used as the NDFM for the set. The value of threshold for the set is calculated as a constant multiple of the NDFM at their immediate ancestor blocks. The constant multiplier M is given by the Equation 4.2. Detection of focused blocks from m images is algorithmically presented below.

1. **Read** m input images X_i, $i = 1$ to m.

2. **Divide** each X_i into four quadrants to get four sets of corresponding blocks $\{X_i^k,\ i = 1 \text{ to } m\}$, $k = 1$ to 4 at level 1.

3. **Calculate** constant multiplier M.

4. **Repeat** for each set of corresponding blocks at level i

   (a) **Compute** threshold T for the set.

   (b) **Compute** FM on each block of the set.

   (c) **Find out** their maximum $F_{\max}$ and minimum $F_{\min}$.

   (d) **Calculate** NDFM for the set.

   (e) **If** $NDFM$ greater than threshold then

   Copy the block with greatest FM to the resultant image.

else if size of blocks > minimum size permissible **then**

Subdivide all blocks in the set into *four* smaller quadrants **and**

Repeat the steps in 4 for smaller blocks at level $i = i + 1$

else Copy the block with greatest focus measure to the resultant image **and stop.**

Detected focused blocks of various sizes are merged naturally and focused regions are produced. Fig. 4.5 shows the gradual detection of focused regions as the number of levels in the quad-trees increase. It is seen that as block-size become smaller borders of focused regions are detected more accurately but small spurious defocused blocks appear inside the regions. This is due to noise or small unresolved blocks on which relative contrast do not vary much. So detected regions require reconstruction which is explained in the next subsection.

4.2.2 Reconstruction

It is evident that a focused region must be wider than the dimension of the spurious blocks inside it. The largest connected focused region is constructed by a morphological filter consisting of an alternating sequence of opening and closing with a disk structuring element (SE) of increasing radius [22]. However, opening and closing with a disk SE trims some sharp convex portions and appends some sharp concave portions respectively. In addition to this some unresolved pixels may still be present. As a result the regions obtained from different input images are neither disjoint nor exhaustive. The final fused image is generated as follows. If a pixel belongs to only one region then its value is copied from the corresponding image. If a pixel belongs to no region or more than one region then weighted average of all input-values at that pixel is copied. Related weights are determined by gradient value at the pixel in the corresponding input image divided by sum of the gradient values at the pixel in all input images.

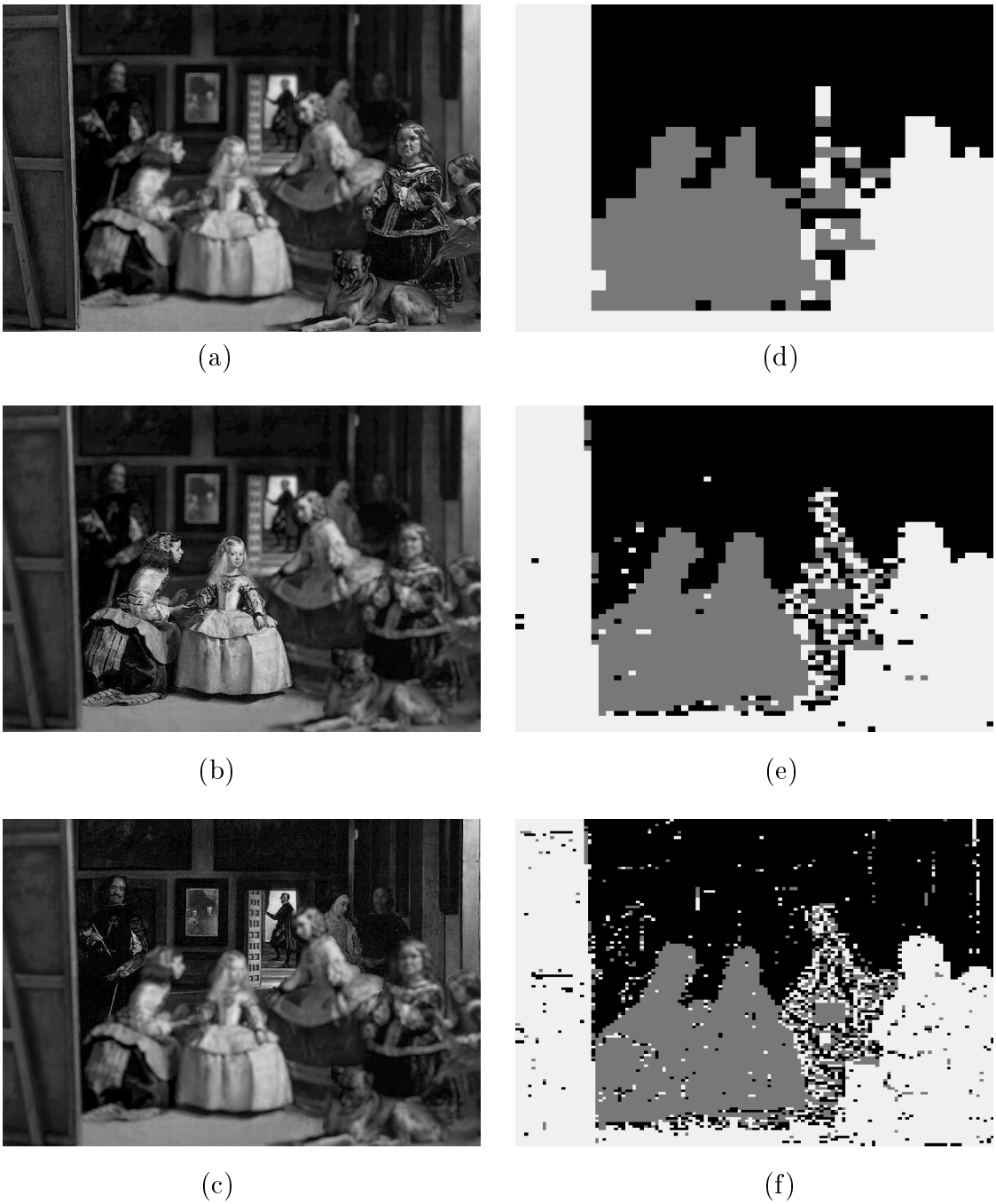


Figure 4.5: Detection of focused regions in a set of multi-focus images up to various levels in a quad-tree. Figures (a), (b) and (c) show input images; (a) Near focused image, (b) Middle focused image, (c) Far focused image. Figures (d), (e) and (f) show detected focused regions respectively up to levels 6, 7 and 8 where white, grey and black blocks are detected from (a), (b) and (c) respectively.

We now present a measure of focus in the next subsection to solve the critical problem of measuring the degree of focus on a block.

4.2.3 Energy of Morphologic Gradients: a new measure of focus

Success of the proposed algorithm depends on how accurately the focus-measure can distinguish between focused and defocused blocks. Edge-strength can be used successfully to identify focused blocks, because they have better contrast and hence more prominent edges than corresponding defocused regions. Edge-strength, in turn is measured by image gradients. Energy of gradients (EOG) is a well-known focus measure theoretically explained and experimentally validated in discrimination of focus quality if the blur function is assumed to be Gaussian or truncated Bessel [76].

A simple method of performing *grey-scale edge detection* by mathematical morphology is to take the difference between an image and its erosion/dilation by a SE [43, 14]. The difference-image is the *edge-strength image*. In general, a grey-scale SE is given by the mapping $h : D \rightarrow \{0, \ldots, 255\}$. *Dilation and erosion* of a grey-scale image $X(r, c)$ by a grey-scale SE $h(r, c)$ are denoted by

$$(X \oplus h)(r, c) = \max_{(i,j) \in Domain\ of\ h} (X(r - i, c - j) + h(i, j)) \tag{4.3}$$

$$(X \ominus h)(r, c) = \min_{(i,j) \in Domain\ of\ h} (X(r + i, c + j) - h(i, j)) \tag{4.4}$$

where the maximum and minimum are taken over all (i, j) in the domain of h such that $(r - i, c - j)$ and $(r + i, c + j)$ are in the domain of X. Most popularly used SE for edge detection is called *rod*. A rod is a grey-scale SE which is flat on top and has a disk-shaped domain with center at $(0, 0)$ [43]. Then the domain of *rod SE of radius 1 (using city-block distance)* is denoted by D_{rod1} and is defined by the set

$$D_{rod1} = \{(0, -1), (0, 1), (0, 0), (-1, 0), (1, 0)\}.$$

Therefore

$$(X \oplus h)(r,c) = \max_{(i,j)\in D_{rod1}} (X(r-i,c-j) + h(i,j)) \quad (4.5)$$

$$(X \ominus h)(r,c) = \min_{(i,j)\in D_{rod1}} (X(r+i,c+j) - h(i,j)) \quad (4.6)$$

Since a rod is flat on top, values of $h(i,j)$ for all $(i,j) \in D_{rod1}$ are taken to be zero. Then

$$(X \oplus h)(r,c) = \max_{(i,j)\in D_{rod1}} (X(r-i,c-j)) \quad (4.7)$$

$$(X \ominus h)(r,c) = \min_{(i,j)\in D_{rod1}} (X(r+i,c+j)) \quad (4.8)$$

Henceforth we denote $(X \oplus h)(r,c)$ and $(X \ominus h)(r,c)$ by $d(r,c)$ and $e(r,c)$ respectively.

Dilation residue edge strength G_d and *erosion residue edge strength* G_e by *rod SE* are obtained as

$$\begin{aligned} G_d(r,c) &= d(r,c) - X(r,c) \\ &= \max_{(i,j)\in D_{rod1}} [X(r-i,c-j)] - X(r,c) \\ &= \max_{(i,j)\in N_4(r,c)} [X(i,j) - X(r,c)] \quad (4.9) \\ G_e(r,c) &= X(r,c) - e(r,c) \\ &= X(r,c) - \min_{(i,j)\in D_{rod1}} X(r+i,c+j) \\ &= \max_{(i,j)\in N_4(r,c)} [X(r,c) - X(i,j)] \quad (4.10) \end{aligned}$$

So morphologic edge operators are local neighborhood operators which take the maximum among the four first differences in directions $0°, 90°, 180°$ and $270°$. Morphologic image gradient $G(r,c)$ at a point (r,c) is calculated as the sum of $G_d(r,c)$ and $G_e(r,c)$

$$G(r,c) = G_d(r,c) + G_e(r,c) \quad (4.11)$$

We define energy of morphologic gradients (EOMG) as

$$EOMG = \sum_r \sum_c (G(r,c))^2 \quad (4.12)$$

A focused block produces larger EOMG than its defocused counterpart because pixels in a focused block are in sharp contrast and hence have greater edge-strength. So EOMG can be used as a measure of focus.

4.3 Experimental results and discussion

The proposed algorithm have been tested on the input images given in Fig. 1.2 with various focus measures, viz. EOMG, EOG, Variance, Tenengrad, EOL, SML, SF etc. Since it is not possible to present all results obtained by various focus measures, we provide the results obtained by EOMG, EOG and Variance in Figure 4.6. For each input image-set three output images are shown; the first one is obtained by EOMG, the second one is obtained by EOG and the third one is obtained by Variance. The fusion is done by allowing the constituent images to be subdivided up to level seven, although that may not be required for all cases. Quantitative evaluations by gradient-similarity-index (GSI) and fusion-quality-index (FQI) as explained in Section 1.7 are given respectively in tables 4.1 and 4.2. The actual run-time in seconds required by the proposed block-based fusion method using three different focus measures viz. EOMG, EOG and Variance are given in Table 4.3.

A.(i) A.(ii) A.(iii)

B.(i) B.(ii) B.(iii)

C.(i) C.(ii) C.(iii)

Figure 4.6: Results of multi-focus image fusion by the proposed block-based method with (i) EOMG, (ii) EOG and (iii) Variance

Table 4.1: Performance evaluation of block-based methods by GSI

Figure	EOMG	EOG	Variance
Doll	0.904	0.902	0.762
Toy	0.822	0.822	0.822
Disk	0.913	0.913	0.877
Lab	0.915	0.910	0.900
Pepsi	0.945	0.945	0.952
Clock	0.885	0.891	0.801
Campus	0.776	0.776	0.760
Hydrant	0.885	0.886	0.868
Garden	0.782	0.777	0.809
Rose	0.882	0.882	0.876
News	0.933	0.932	0.929
OpenGL	0.919	0.914	0.867
Average	0.880	0.879	0.851

4.3.1 Discussion

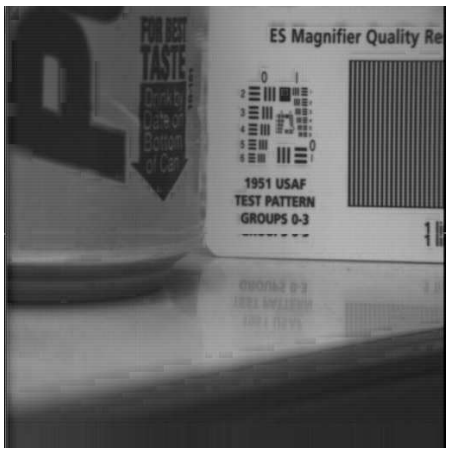
Careful manual inspection of experimental results shows that the proposed focus measure EOMG work equally well on input images which vary widely in their focusing, object-content and in their texture. It also shows that the results obtained by various focus measures are good and do not vary much in their quality. This shows robustness of the algorithm.

Now discussion on quantitative evaluations are given. The tables 4.1 and 4.2 show that for 'Toy' images all the three focus measures yield identical values. This implies that fused images produced by the proposed algorithm with three different focus measures are identical. This happens because quad-trees generated by EOMG, EOG and variance are identical in that particular case. The tables also show that for

D.(i)

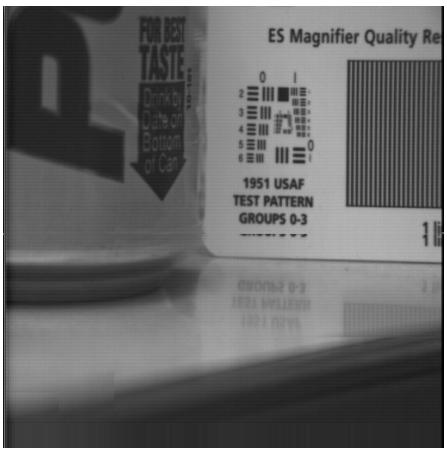
D.(ii)

D.(iii)



E.(i)

E.(ii)



E.(iii)

F.(i)

F.(ii)

F.(iii)

Figure 4.6: Continued

G.(i) G.(ii) G.(iii)

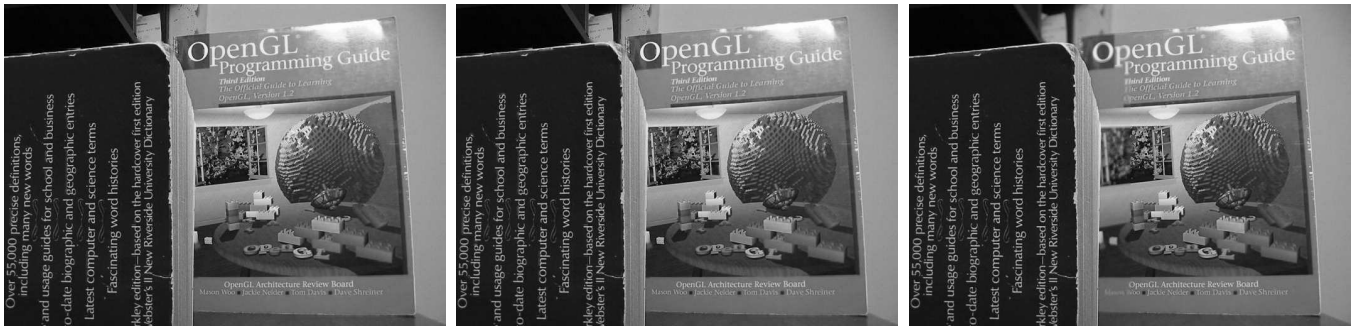
H.(i) H.(ii) H.(iii)

I.(i) I.(ii) I.(iii)

Figure 4.6: Continued

J.(i) J.(ii) J.(iii)

K.(i) K.(ii) K.(iii)



L.(i) L.(ii) L.(iii)

Figure 4.6: Continued

Table 4.2: Performance evaluation of block-based methods by FQI

Figure	EOMG	EOG	Variance
Doll	0.836	0.837	0.795
Toy	0.838	0.838	0.838
Disk	0.877	0.877	0.870
Lab	0.875	0.875	0.873
Pepsi	0.884	0.886	0.894
Clock	0.903	0.903	0.893
Campus	0.909	0.909	0.897
Hydrant	0.914	0.914	0.913
Garden	0.829	0.830	0.820
Rose	0.890	0.890	0.883
News	0.894	0.893	0.890
OpenGL	0.882	0.882	0.871
Average	0.877	0.877	0.869

some images like 'Disk', 'Campus' and 'Rose'; EOMG and EOG produce identical values. Following the argument just stated, it is concluded that quad-trees generated by EOMG and EOG are identical for those images. On average, performance of EOMG is slightly better than EOG and better than variance. Moreover as EOMG works equally well with all input image-sets (which vary widely in content, texture and focusing), we conclude that it fulfills the desirable properties of a focus measure mentioned in section 4.1; viz. ability to measure focus irrespective of image content, monotonicity with respect to blur, unimodality, robustness to noise and capability to produce large variations in values with respect to degree of blur. Regarding the time requirement, average time taken by the proposed algorithm is less than a minute for all the three focus measures EOMG, EOG and variance.

Formal computational complexity of our method is described now. In detection step,

Table 4.3: Time requirement of the block-based algorithm with three different FM's

Figure	EOMG (seconds)	EOG (seconds)	Variance (seconds)
Doll	41	37	45
Toy	59	55	62
Disk	24	27	33
Lab	26	20	21
Pepsi	30	25	39
Clock	19	17	14
Campus	33	30	34
Hydrant	34	28	29
Garden	28	29	25
Rose	31	33	32
News	22	19	23
OpenGL	32	23	27
Average	32	29	32

each input image is subdivided according to a quad-tree structure. Suppose an input image has $M \times N$ pixels and the quad-tree has l levels. Maximum possible value of l depends on the minimum block-size. For example, if the minimum block-size is taken to be 2×2, then for a 512×512 image l is equal to 8. More formally, maximum possible value of l is equal to $\min(\log_2 M - 1, \log_2 N - 1)$. If all levels of the quad-tree have maximum possible nodes then it means all pixels in the image are to be calculated upon up to level l. So computational complexity is $O(M \times N \times l)$ for a single image. For a set of k input images the computational complexity becomes $O(M \times N \times l \times k)$. It is seen in practical cases that maximum possible level is not required always and most of the levels do not have more than half of the maximum possible nodes. So time-requirement in detection step is effectively in order of size of input images. Reconstruction is done by iteratively applying the morphological operators opening

and closing with a disk SE of increasing radius. So time requirement in this step also is in order of size of input images.

4.4 Summary

In this work we have proposed an algorithm to enhance the effective DOF of camera by multi-focus image fusion. It is a simple quad-tree based algorithm which recursively divides and then compares and selects/rejects blocks. For effective subdivision of blocks, NDFM of a set of corresponding blocks is calculated and compared with a threshold. The process of recursive subdivision continues until either NDFM becomes greater than the threshold or the block-size becomes very small. No threshold is supplied manually, rather thresholds are calculated automatically from the input images. Detected focused blocks are merged naturally to form focused regions which are reconstructed then to remove any spurious blocks inside them. So although the proposed method starts with blocks, it finally detects focused regions from each input image. These regions are then copied to create the final fused image. The method is fast and easy to implement. Its time complexity is of the order of size of the input images. We have also proposed a new focus measure EOMG in this work. It satisfies the desirable properties of a focus measure. The proposed algorithm is robust in the sense that any focus measure can be used for its implementation. Since it is a block-based method, it is robust against pixel mis-registration. Moreover, since the algorithm is based on neighbourhood operators, it can be efficiently implemented in machine vision systems having special hardware support for morphologic operations.

Chapter 5

Region based fusion

5.1 Introduction

In this chapter we present a region-based method for multi-focus image fusion (MFIF). Generally regions convey more semantic information than single pixels and small blocks. So region-based fusion approaches can avail more intelligent semantic fusion rules than pixel-based and block-based methods. A number of region-based fusion techniques are found in the literature. They are based either on spatial domain or on frequency/transform domain. Spatial domain fusion methods may depend on multi-resolution or multi-scale decomposition (MSD). Frequency domain fusion techniques do depend on MSD. They use either pyramid or wavelet transform for MSD. Region-based fusion techniques using pyramid or wavelet transform have been proposed by many researchers [88, 86, 65, 64, 82, 34, 85]. Some key points of region-based MSD image fusion approach [88] are each pixel is considered as part of object or region of interest, image features such as edges and regions are used to guide the fusion, both spatial and frequency information are retrieved from the coefficients. Basic steps of MSD image fusion techniques are the following. At first each source image is transformed/decomposed by an MSD scheme like pyramid transform or wavelet transform. Low and high frequency coefficients forming respectively scaled and detail images are

obtained from the transform. Regions representing image features are then extracted from the transform coefficients by an image segmentation method. The regions are then fused based on region characteristics. The final fused image is obtained by applying the inverse transform to the composite representation. Experimental results of these methods are encouraging. However, the images fused in this way may still lose some information of the source images because of the implementation of the inverse transform. There are region-based fusion methods using combination of wavelet transform and artificial neural networks [55, 46, 68, 39, 84]. These methods are motivated by fusion of different sensor signals in biological systems and use multi-layer perceptron neural networks or pulse-coupled neural networks. There are transform domain methods which use Independent Component Analysis (ICA) and Topographic Independent Component Analysis bases in image fusion [56].

Spatial-domain region-based fusion methods may or may not depend on multi-scale decomposition (MSD). Methods for image fusion using multi-scale morphology are described in [53, 57]. In method of Matsopoulos et al. morphological filters with structuring elements of varying size are used to construct a morphological pyramid. Such pyramids are constructed for each input image. Then morphological difference pyramids are constructed for each of the above pyramids. After that, an intermediate pyramid is constructed by combining information at each level from the above difference pyramids. Finally, reconstruction is done by using appropriate morphological operations on the intermediate pyramid to produce the final fused image. This method can be used for multi-focus images as well, but it was not mentioned whether the method can be applied to more than two input images. Since in the third step the difference pyramids are combined by choosing the maximum at each pixel, this method is sensitive to the problem of mis-registration as mentioned before. Mukhopadhyay and Chanda proposed a similar method in [57] except that they have used morphological towers instead of morphological pyramids. They have used their method for fusion of multi-focus images. But their method involves processing and storing of scaled data at various levels which are of the same size as that of the original images. This results in a huge amount of memory and time requirement.

Li and Yang's technique [49] is a spatial-domain, region-based technique which does not depend on MSD. In this technique, a temporary fused image is obtained at first by taking the average of all input images. The temporary image is segmented by normalized-cut algorithm. Input images are segmented according to the segmentation results of the temporary image. Focus measure of corresponding regions are measured by spatial frequency and better focused regions are selected and stitched to their desired positions to get the final fused image. The technique is time-consuming and depends greatly on the performance of the segmentation algorithm.

In this chapter a spatial domain, region-based fusion method is presented. Multi-scale morphological filters are employed to identify focused regions from input images. Focused regions from various images are then stitched at their proper positions to create the final fused image. Since best-focused regions are selected and copied from one image only, a slight error in registration will have no effect in fusion except in the border of the regions. Prior segmentation is not required in the method. Manual cut-and-paste of focused regions from multi-focus images is considered to be the best and it is often used for comparison purposes [45]. The proposed method is a close approximation to this and produces good results. The results have been compared with those obtained by Li and Yang's technique [49]. The chapter is organized as follows. Section 5.2 describes the proposed method in detail. Subsections 5.2.1, 5.2.2 and 5.2.3 present the methods for multi-scale top-hat transformation, detection of focused regions and image reconstruction respectively. Experimental results and discussion including performance analysis are given in Section 5.3. Finally, summary of the chapter and concluding remarks are placed in Section 5.4.

5.2 Fusion by multi-scale morphology

The objective of region-based fusion methods is to detect focused regions from every input image, then to stitch detected focused regions to their proper positions in the fused image. Since multi-focus images of a scene are acquired with focus on comple-

mentary regions, focused regions in an image have more contrast than their defocused counter-parts in other images. This implies that focused regions contain larger number of physically relevant features of different shapes and sizes. Extraction of features by mathematical morphology depends on effective use of structuring elements. Sizes and shapes of structuring elements play crucial roles here. A morphological operator with a scalable structuring element can extract features of various shapes and sizes. A scheme of morphological operations with a structuring element of varying scale is termed as *multi-scale morphology* [15, 52]. We use such a scheme for our purpose. The scheme is known as multi-scale top-hat transformation. We describe now the transformation and detection of focused regions by using it.

5.2.1 Multi-scale top-hat transformation

A two-dimensional grey-scale image signal X is a function/mapping from domain D (which is a subset of discrete two-dimensional space Z^2) to the set of grey values $\{g_1, g_2, \ldots, g_n\}$ where each g_i is a nonnegative integer. A grey-scale structuring element (SE) 'h' is a mapping from its domain to the above set of grey values. For a scalable SE 'h', size of the domain get changed. Let B be a set representing the domain of 'h'. Assume that B has a definite shape. Let n be an integer representing the scale-factor of B and let nB denote the scaled version of B at scale n. If B is convex, then nB is obtained by $n-1$ dilations of B by itself.

$$nB = B \underbrace{\oplus B \oplus B \oplus \cdots \oplus B}_{n-1 \text{ times}} \tag{5.1}$$

When $n = 0$, conventionally B is taken to be a disk of unit size so that $nB = \{(0,0)\}$.

Let 'h' be a flat-top SE such that its value at every point in its domain nB is zero. Then a morphologic operation by 'h' reduces to an operation by its domain nB. Then multi-scale opening and closing of X by scalable domain nB are defined respectively

as

$$(X \circ nB)(r,c) = ((X \ominus nB) \oplus nB)(r,c) \quad (5.2)$$

$$(X \bullet nB)(r,c) = ((X \oplus nB) \ominus nB)(r,c) \quad (5.3)$$

Opening removes all bright/foreground structures in the image X that are not large enough to contain nB. Hence $X \circ nB$ essentially contains all bright/foreground structures of X having size greater than or equal to nB. Similarly, closing removes all background structures in the image X that are not large enough to contain nB. Hence, $X \bullet nB$ contains all dark/background structures of X having size greater than or equal to nB. Here foreground structure means an image region of intensity value higher than the surrounding region.

Top-hat transformation for opening and closing filters are defined respectively as:

$$d_o^{(n)}(X(r,c)) = (X \circ (n-1)B)(r,c) - (X \circ nB)(r,c) \quad (5.4)$$

$$d_c^{(n)}(X(r,c)) = (X \bullet nB)(r,c) - (X \bullet (n-1)B)(r,c) \quad (5.5)$$

Thus, $d_o^{(n)}(X)$ contains all the bright features that have size greater than or equal to $(n-1)B$ but less than nB. Similarly, $d_c^{(n)}(X)$ contains all the dark features within the same range of size. Hence the feature image defined as

$$D^{(n)}(X(r,c)) = \max\{d_o^{(n)}(X(r,c)), d_c^{(n)}(X(r,c))\} \quad (5.6)$$

contains all the image features having size within the range $[(n-1)B, nB)$. Hence image features are thus sieved out based on their size and stored in corresponding $D^{(n)}(X)$.

5.2.2 Detection of focused regions

It is evident from the previous discussion that if a particular feature (bright or dark) of an image is sharply focused it is sieved out in relatively lower scale. Let $X_j, j = 1, 2, \ldots, k$ be a set of multi-focus images and let $D_j^{(n)}$ denote the feature image of X_j

at scale n. Now if a pixel (r, c) is sharply focused in the image X_i, then at lower scale, $D_i^{(n)}(r, c)$ should be greater than $D_j^{(n)}(r, c)$ for all $j \neq i$. Thus, the focused region of each image can be identified and marked. Let $F_j^{(n)}$ denote the focused region at scale n of image X_j and let $F^{(n)}$ denote the union of focused regions at scale n from all images $X_j, j = 1, 2, \ldots, k$. Then detection of focused regions at various scales $F_j^{(n)}$ can be algorithmically presented as

Step-1: $F_j^{(0)}(r, c) = 0 \quad$ for all j

Step-2: $n = 1$

Step-3: Calculate $D_j^{(n)}(r, c)$ for all images X_j

Step-4: $F_j^{(n)}(r, c) = 1$, if $D_j^{(n)}(r, c) > \max\{D_i^{(n)}(r, c)\}$ for all $i \neq j$

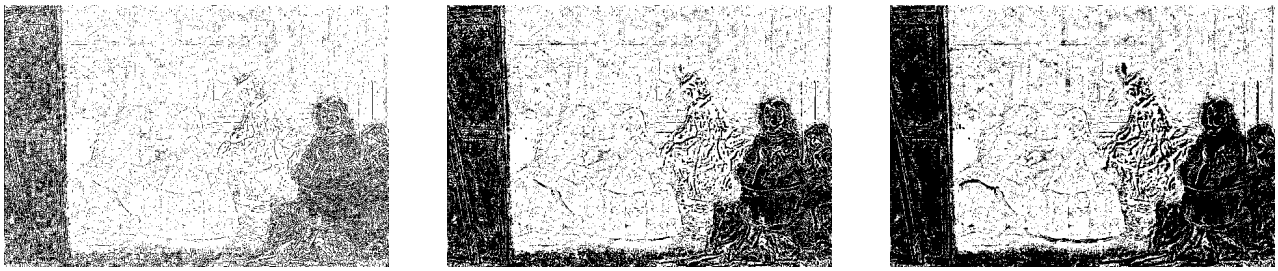
Step-5: $F^{(n)}(r, c) = F_1^{(n)}(r, c) \vee F_2^{(n)}(r, c) \vee \cdots \vee F_k^{(n)}(r, c)$

Step-6: If all pixels of $F^{(n)}$ are not equal to 1, increase n by 1 and go to Step-3

Hence, the focused regions or, more specifically, the focused pixels in the image X_j are marked by 1 in $F_j^{(n)}$. In practice, we terminate this algorithm when at least p-percent pixels of $F^{(n)}$ become 1 or no further change occurs in $F^{(n)}$. Rest unresolved pixels where $F^{(n)}(r, c) = 0$ either belong to smooth regions or belong to boundary of focused regions and are taken care of at the subsequent stage. Binary images corresponding to focused regions detected at various scales for the near-focused 'Doll' image in Fig. 1.2 are shown in Fig. 5.1.

5.2.3 Reconstruction

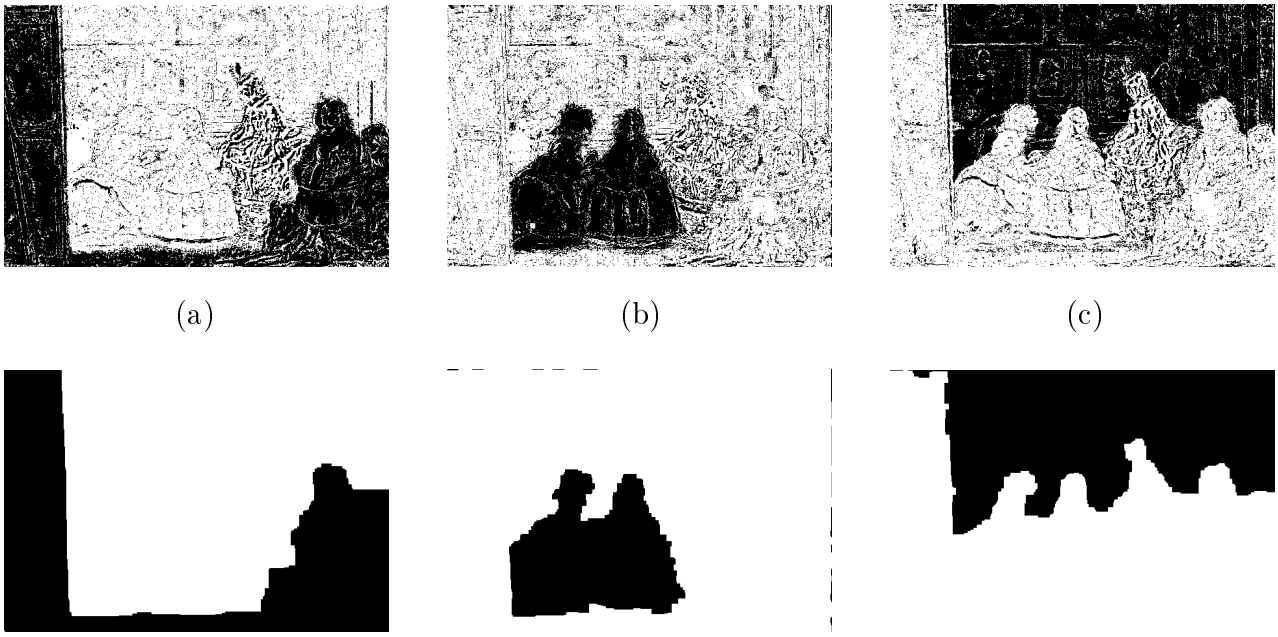
Image of focused region $F_j^{(n)}$ for j-th input image may appear to contain spurious white spots in sharply focused region (shown here as black colored) and black spots in the out-of-focus region (shown here as white colored). This phenomenon can be observed in Fig. 5.1. It is evident that a focused object or region must be wider



(a) (b) (c)

Figure 5.1: Binary images corresponding to focused regions detected at various scales for the near-focused 'Doll' image in Fig. 1.2. Focused regions obtained by using (a) 2×2 SE. (b) 4×4 SE. (c) 8×8 SE.

than the dimension of these spurious spots. Then these spots can be treated as additive and subtractive noise that are introduced due to salt-and-pepper noise present in the out-of-focus regions and also due to pixels that remain unresolved after the previous processing. It is well-known that opening and closing filter respectively can remove additive and subtractive noise efficiently. Hence, an alternating sequential filter formed by concatenating opening and closing with a small disk structuring element is applied on each binary image $F_j(r, c)$ to obtain $R_j(r, c)$ consisting solid black blob(s). So the largest connected regions R_j mark the final focused regions in X_j. Binary images corresponding to focused regions (of multi-focus 'Doll' images Fig. 1.2) detected at the third iteration and the corresponding largest connected regions are shown in Fig. 5.2. Now the image where all regions are properly focused may be reconstructed by putting together the pixels of X_j's corresponding to marked (black) regions of R_j's. However, it should be mentioned here that opening and closing with a disk structuring element trims some sharp convex portions from the blob and appends some sharp concave portions to the blob respectively. In addition to this some unresolved pixels may still be present. As a result R_j's are neither disjoint nor exhaustive. That means neither $R_i \wedge R_j, i \neq j$ produces a blank (or white) image nor $\bigvee_j R_j$ produces a filled (or black) image. Hence, the resultant fused image $\tilde{X}(r, c)$ is



(a) (b) (c)

(d) (e) (f)

Figure 5.2: Detected focused regions and the corresponding largest connected regions for multi-focus images

generated by non-linear superposition of $X_j(r,c)$ depending on $R_j(r,c)$ as follows.

$$\tilde{X}(r,c) = \begin{cases} X_j(r,c), \text{ if } R_j(r,c) = 1 \text{ and } R_i(r,c) = 0 \text{ for all } i \neq j \\ \text{avg } \{X_j(r,c) \,|\, R_j(r,c) = 0 \text{ for all } j\} \\ \text{avg } \{X_j(r,c) \,|\, R_j(r,c) = 1 \text{ for more than one } j\} \end{cases} \tag{5.7}$$

The function $avg(.)$ stands for pixel-wise average from a set of images. Experimental results and discussions are presented now.

5.3 Experimental results and Discussion

The proposed algorithm has been tested on the input images given in Fig. 1.2. It is seen in the experiments that three iterations are sufficient to detect focused regions in all images. So top-hat transformation is applied on input images at three different scales. Disk structuring elements of three different sizes viz. (a) 2×2, (b) 4×4, (c) 8×8 are employed for this purpose. For comparison purpose, we have implemented

the region-based fusion technique presented by Li and Yang [49]. Experimental results by the proposed algorithm and by Li's method are shown in figure 5.3. In each row, the fused images obtained by the proposed method are given first, followed by fused images by Li and Yang's method. Objective evaluations by gradient-similarity-index (GSI) and fusion-quality-index (FQI) as explained in Section 1.7 are given respectively in tables 5.1 and 5.2. Actual time in seconds required by the proposed algorithm and Li's method are shown in table 5.3.

Table 5.1: Performance evaluation of region-based methods by GSI

Figure	Proposed method	Li's method
Doll	0.907	0.799
Toy	0.808	0.805
Disk	0.912	0.869
Lab	0.924	0.800
Pepsi	0.942	0.945
Clock	0.864	0.870
Campus	0.782	0.799
Hydrant	0.862	0.843
Garden	0.857	0.784
Rose	0.891	0.873
News	0.926	0.853
OpenGL	0.906	0.863
Average	0.882	0.840

A.(i)

A.(ii)

B.(i)

B.(ii)

C.(i)

C.(ii)

Figure 5.3: Results of multi-focus image fusion by region-based methods. In each row, images shown are obtained by applying (i) proposed region-based algorithm (ii) Li' algorithm

Table 5.2: Performance evaluation of region-based methods by FQI

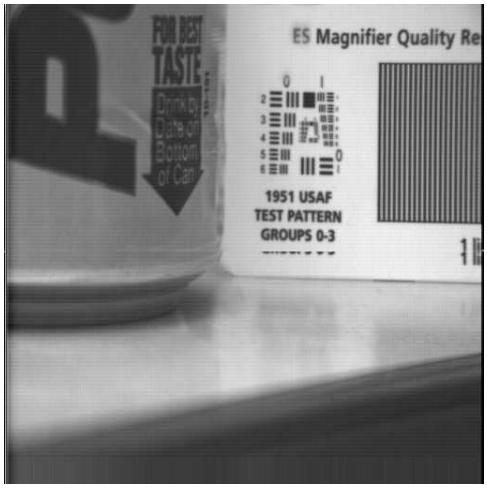
Figure	Proposed method	Li's method
Doll	0.840	0.818
Toy	0.846	0.833
Disk	0.873	0.871
Lab	0.877	0.842
Pepsi	0.896	0.892
Clock	0.910	0.906
Campus	0.916	0.912
Hydrant	0.917	0.906
Garden	0.838	0.812
Rose	0.900	0.883
News	0.897	0.873
OpenGL	0.877	0.863
Average	0.882	0.867

5.3.1 Discussion

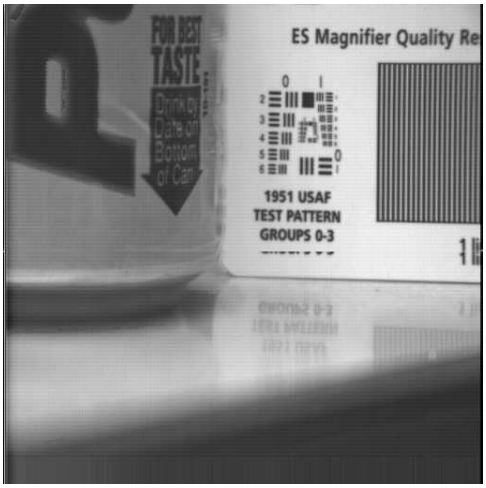
Careful manual inspection of images in figure 5.3 reveals that the results obtained by the proposed fusion method is better than fusion by Li's method. Inspection of the crane in the fused image 'Toy', the edge of the table in the fused image 'Pepsi', the rose in fused image 'Rose' and the letter 'O' in fused image 'OpenGL' show that it produces better results than Li's method. In both the tables 5.1 and 5.2, the average values produced by our method is better than that produced by Li's method. The average time taken by the proposed method is less than one and a half minute, however the average time taken by Li's method is more than six minutes. So also in respect of time requirement, the proposed method is better.

D.(i)

D.(ii)



E.(i)



E.(ii)

F.(i)

F.(ii)

Figure 5.3: Continued

G.(i)

G.(ii)

H.(i)

H.(ii)

I.(i)

I.(ii)

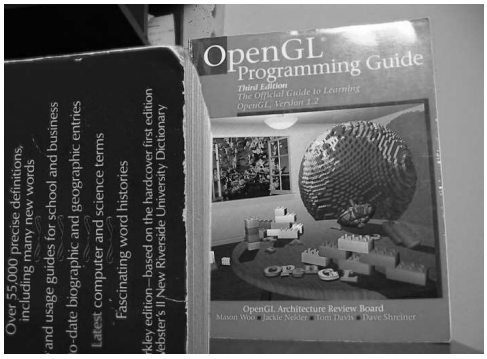
Figure 5.3: Continued

J.(i)

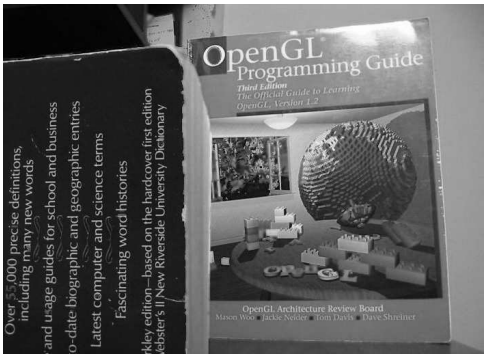
J.(ii)

K.(i)

K.(ii)



L.(i)



L.(ii)

Figure 5.3: Continued

Table 5.3: Time requirement in region-based methods

Figure	Proposed method (seconds)	Li's method (seconds)
Doll	71	221
Toy	89	505
Disk	78	339
Lab	76	342
Pepsi	80	490
Clock	37	119
Campus	96	479
Hydrant	89	470
Garden	78	214
Rose	81	628
News	38	102
OpenGL	82	587
Average	75	375

5.4 Summary

In this work we have proposed a region-based method for multi-focus image fusion. In general, region-based methods do depend on prior segmentation of input images. Therefore, performance of the segmentation algorithm, both in respect of time and quality, affects the performance of the fusion algorithm. Generally, number of regions produced by segmentation process is much larger than actual number of focused/defocused regions. It means more processing time is required during evaluation of focus-quality of corresponding regions. Moreover, it is difficult to evaluate the focus quality for small regions, which means distinction between corresponding focused and defocused counterparts is difficult which ultimately may lead to selection of a defocused region. The proposed method does not need any prior segmentation.

Rather, it employs multi-scale morphological filters to detect features in focus at various scales from each input image. Union of them constitutes focused regions from the image. Detected regions from all input images are put together to reconstruct the fused image. Since the best-focused regions are detected and copied from one image only, a slight error in registration will have no effect in fusion except in the borders of the focused regions. Hence this region-based method is robust to mis-registration. This method resembles the manual cut-and-paste method of image fusion which is often used for comparison purposes. Thus the fused image obtained by the method is very similar to the ideal fused image. Performance of the algorithm is compared with a region-based algorithm proposed by Li and Yang. Performance analysis reveals that our method is superior to fusion by Li and Yang's method.

Chapter 6

Conclusion and future work

Multi-focus image fusion (MFIF) is a way to enhance effective depth-of-field of a digital camera. Techniques for MFIF can be divided into broad categories, pixel-based, block-based and region-based. It is interesting to study and compare techniques within a particular category and techniques belonging to different categories. The thesis (i) surveys extensively on existing literature for MFIF methods and classifies the methods according to the above categories, (ii) proposes a method for multi-focus image registration, (iii) proposes new methods for MFIF, one in each of the categories mentioned above, (iv) presents experimental results for proposed methods on a large data-set, (v) compares the results with those obtained by other well-known methods and (vi) does performance analysis using standard quantitative evaluation techniques. All techniques for MFIF proposed in the thesis use mathematical morphologic tools.

Image registration is a necessary pre-requisite for image fusion. The thesis presents a method for multi-focus image registration in Chapter 2. The method is compared with a widely used registration technique and is found to produce better results than the latter. Chapter 3 proposes a non-linear tool morphological wavelet and presents a pixel-based algorithm for MFIF using the same. The algorithm is comparable with other standard pixel-based techniques. Interesting mathematical properties of the wavelet used makes the algorithm hardware implementable. Chapter 4 proposes a fo-

Table 6.1: Performance summery of the (i) Pixel-based algorithm by the proposed 2D morphologic wavelet, (ii) Block-based algorithm by proposed focus-measure EOMG and (iii) Proposed region-based algorithm

Figure	Pixel-based		Block-based		Region-based	
	Time (seconds)	FQI	Time (seconds)	FQI	Time (seconds)	FQI
Doll	10	0.805	41	0.836	71	0.840
Toy	13	0.775	59	0.838	89	0.846
Disk	7	0.853	24	0.877	78	0.873
Lab	10	0.851	26	0.875	76	0.877
Pepsi	11	0.869	30	0.884	80	0.896
Clock	1	0.884	19	0.903	37	0.910
Campus	12	0.895	33	0.909	96	0.916
Hydrant	13	0.877	34	0.914	89	0.917
Garden	7	0.791	28	0.829	78	0.838
Rose	14	0.856	31	0.890	81	0.900
News	1	0.867	22	0.894	38	0.897
OpenGL	11	0.857	32	0.882	82	0.877
Average	9	0.848	32	0.877	75	0.882

cus measure based on mathematical morphology and presents a block-based algorithm for MFIF using the same. The algorithm is fast, easy-to-implement, and produces good results. Finally, Chapter 5 presents a region-based algorithm for MFIF using mathematical morphology. This algorithm properly selects the focused regions from multi-focus input images and then copies and pastes them to form the final fused image. It resembles the manual cut-and-paste method for MFIF often used to produce image all-in-focus, for testing purposes [45]. Results produced by this algorithm are the best amongst all the three proposed MFIF algorithms. For comprehensive assessment of the proposed pixel-based, block-based and region-based methods, actual

run-time in seconds and fusion quality index (FQI) for the test images are given in Table 6.1.

After comparing the results of various techniques for MFIF, we find the following.

- Pixel-based and region-based methods are respectively the least and the most time-consuming amongst the methods in the three categories mentioned above; and block-based methods take more time than pixel-based methods but less time than region-based methods.
- Performance-wise pixel-based and region-based methods are respectively the worst and the best methods among the three categories, and block-based methods are mid-way between pixel-based and region-based methods.
- Pixel-based and region-based methods are respectively the least and the most robust in respect of mis-registration problem which means slight registration-error results into severe performance degradation in pixel-based methods whereas that error does not affect region-based methods so much.
- Finally, pixel-based and region-based methods are respectively the least and the most difficult in respect of implementation and usage.
- Block-based methods maintain a good trade-off in terms of time-complexity and performance; and they are moderate also in respect of implementation and usage. Finally they are moderately affected by mis-registration problem. So they are often chosen for practical purposes.

The proposed MFIF techniques can be used to integrate multi-focus color images. For that, the input multi-focus colour images (in R,G,B format) may be represented in intensity-chromaticity format. The proposed MFIF methods may then be applied to the intensity images to get the fused intensity image. Input chromaticity images are combined following the same steps as the ones used in case of intensity images. Finally, the fused intensity and chromaticity images are combined to get the fused color image. We now outline some future work as extension of the work done.

6.1 Future work

6.1.1 Fusion by area morphology

Basic tools of morphologic filters are opening and closing. Morphologic (also called structural) opening is an increasing, idempotent and anti-extensive operation defined as an erosion by a SE followed by a dilation by the reflected SE. An operation having the same properties, but that cannot be written as a unique erosion followed by a dilation, is called an *algebraic opening* [75]. *Area opening* is an algebraic opening. It preserves the connected sets in the foreground having areas greater than a given threshold value and removes all other sets. Here foreground means an image region of intensity value higher than the surrounding region and area is measured in number of pixels. The dual operation of area opening is *area closing*. It is an algebraic closing which preserves the connected sets in the background having areas greater than a given threshold value and removes all other sets. This information can be used successfully to extract focused regions in multi-focus images.

In multi-focus images, focused regions have more contrast and hence larger number of small features or grains than their defocused counter-parts. This implies that when compared with the corresponding defocused region, a focused region has (i) greater range of grey values and (ii) greater number of pixels with very high and very low intensity values in the range. These pixels can be extracted by using area opening and area closing respectively. Subsequently, focused regions can be detected by using these pixels. In general, region-based fusion methods are more complex and time-consuming than pixel-based and block-based methods. We plan to work for a simple region-based multi-focus image fusion method by using the operations, area opening and closing. Use of efficient algorithms [54] for these operations may reduce the time-requirement for the method.

6.1.2 Extension to multi-modal images

The proposed MFIF techniques can be used to integrate multi-modal images. In a multi-sensor data acquisition system, the image data of an object consists of information acquired by different sensors from different perspectives and possibly at different resolutions. The clarity of the object features may be different in different imaging modalities. For example, in the area of biomedical imaging, two widely used modalities, namely the Magnetic Resonance Imaging (MRI) and the Computer Tomographic (CT) scan do not reveal all types of tissue structure with equal clarity. CT scan is especially suitable for imaging bones or hard tissues, whereas the MR images are much superior in depicting the soft tissues. These two imaging modalities are thus complementary in many ways and no one alone is sufficient in terms of required information content. The propsed methods may be extended for fusing such multi-modal images.

6.1.3 Hardware embedding

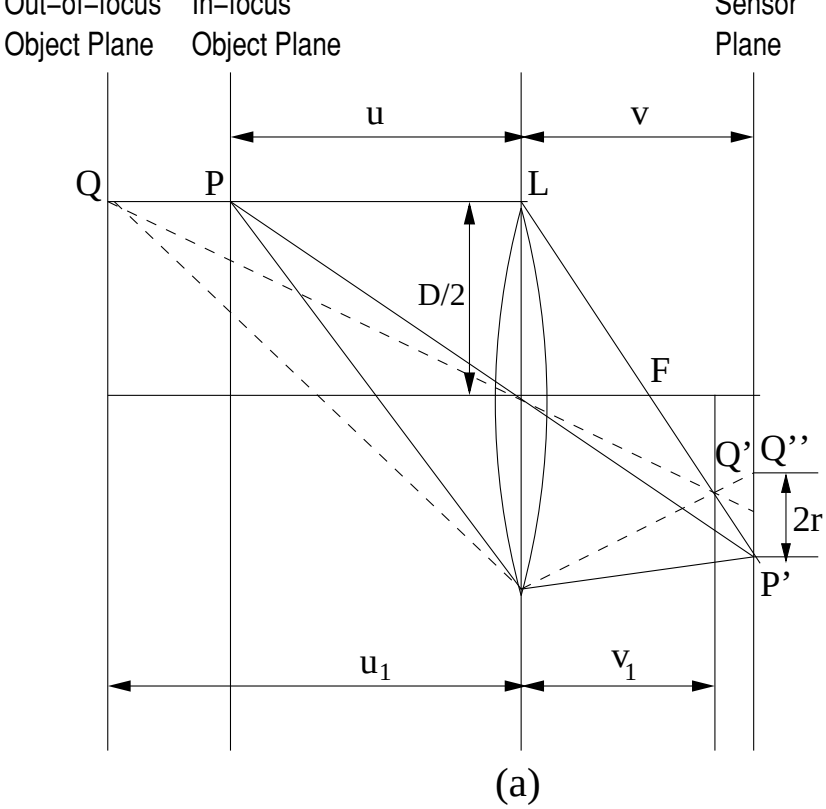
Tomorrow's computing and communication technology will rely on extensive use of embedded software. There are previous work on special purpose hardware design for mathematical morphologic algorithms [4]. Proposed algorithms can be embedded in hardware using efficient gate-arrays. Embedded hardware design primarily depends on use of parallel operations. Graph-theoretic design approaches like precedence graph and interval graph may be applied on proposed algorithms to explore their inherent parallelism and hence their potential for hardware embedding. Simpler of the proposed algorithms can be efficiently implemented in machine vision systems having special hardware support for morphologic operations.

Appendix A

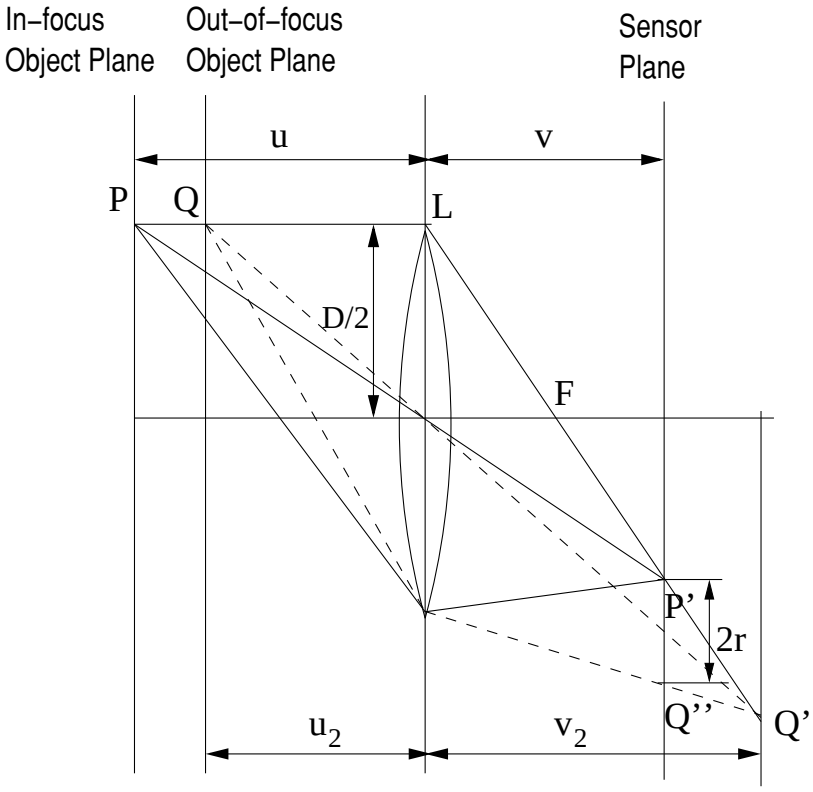
Depth of field

Depth of field can be calculated by par-axial geometric optics model of image formation using a thin convex lens [76]. Figures A.1(a) and A.1(b) illustrate two different situations using such a model. In both figures, P and Q are two point-objects, L is the lens, F is the focal point and D is the diameter of aperture of the lens (assumed to be circular in this case). Point-object P on object plane at distance u from the lens is perfectly focused as point-image P' on sensor plane at distance v from the lens. Well-known lens equation $\frac{1}{f} = \frac{1}{u} + \frac{1}{v}$ relates the position of these two points, u and v, with that of the focal length f of the lens. Point-object Q is taken in such a way that it is further from the lens than P in fig. A.1(a) and nearer to the lens than P in fig. A.1(b). The distance of Q from the lens is u_1 and u_2 in figures A.1(a) and A.1(b) respectively where $u_1 > u$ and $u_2 < u$. Focused images Q' of Q are formed at distances v_1 and v_2 in respective figures where $v_1 < v$ and $v_2 > v$. So focused images of Q are formed in front of and behind the sensor plane in respective figures. In both cases, blurred circular images of Q with diameter $P'Q''$ is formed on the sensor-plane. We can estimate the blur-circle radius r in fig. A.1(a) using similar triangles,

$$\begin{aligned} \frac{2r}{D} &= \frac{v - v_1}{v_1} \\ r &= (v - v_1)\frac{D}{2v_1} \end{aligned}$$



(a)



(b)

L Lens D Aperture Diameter

F Focal Point r Blur−circle radius

Figure A.1: Par-axial geometric optics model of image formation

Similarly we can estimate the blur-circle radius r in fig. A.1(b),

$$\begin{aligned} \frac{2r}{D} &= \frac{v_2 - v}{v_2} \\ r &= (v_2 - v)\frac{D}{2v_2} \end{aligned}$$

It is readily seen in both cases that, larger aperture generates larger blur circle. Using the above relations for the blur circles, we can derive *depth-of-field (DOF)* for a lens system, where r now becomes the radius of largest acceptable blur circle in the resultant image, which can be chosen based upon sensor resolution and human visual acuity limits. Note that the lens equation is also satisfied by u_1, v_1 pair and u_2, v_2 pair. We can estimate the DOF from fig. A.1 using similar triangles,

$$\begin{aligned} \frac{2r}{D} &= \frac{v - v_1}{v_1} \\ \frac{v}{v_1} &= 1 + \frac{2r}{D} \\ \frac{1}{v_1} &= \frac{1}{v}(1 + \frac{2r}{D}) \\ \frac{1}{u_1} &= \frac{1}{f} - \frac{1}{v}(1 + \frac{2r}{D}) \\ &= \frac{1}{f} - (\frac{1}{f} - \frac{1}{u})(1 + \frac{2r}{D}) \\ &= \frac{(f-u)(1 - \frac{4r^2}{D^2}) + u(1 - \frac{2r}{D})}{uf(1 - \frac{2r}{D})} \end{aligned}$$

Therefore,

$$u_1 = \frac{uf(1 - 2\frac{r}{D})}{f - 2\frac{r}{D}u} \text{ if } 4r^2 << D^2 \tag{A.1}$$

Similarly from fig. A.1(b),

$$u_2 = \frac{uf(1 + 2\frac{r}{D})}{f + 2\frac{r}{D}u} \text{ if } 4r^2 << D^2 \tag{A.2}$$

And,

$$DOF = u_1 - u_2 \tag{A.3}$$

where u_1, u_2 are the distances to the nearest and the furthest object planes with blur circles having radii less than or equal to the chosen r. As D tends to infinity, u_1 and

u_2 tend to u and DOF tends to zero. This result agrees with the common knowledge that reducing the aperture-size increases DOF and increasing the same reduces it.

In summary, DOF depends on the following factors, the amount of sharpness-loss regarded as acceptable, the aperture used (decreasing the aperture will increase the DOF), the focal length of the lens (longer the focal length, shorter the DOF) and the distance of the focused object (nearer the object, shorter the DOF). An extreme case of decreasing the aperture for maximizing the DOF happens in case of pin-hole camera. It has an infinite DOF. Unfortunately, the optical power in the image plane is reduced considerably due to infinite DOF. So cameras with finite DOF are preferred. But the problem is that they cannot generate the images of all objects at various distances from the camera with equal clarity.

Appendix B

Affine transformation

Affine transformation is a common technique for image registration. It is based on the assumption that only geometric transformations possible between a related pair of images are translation, rotation, isotropic scaling and shear (non-uniform scaling in some direction). Properties like parallelism, ratio of lengths of collinear or parallel segments, ratio of areas, linear combination of vectors are invariant under affine transformation [13]. The equation for affine transformation is given now.

Let (x, y) be a point in two-dimensional Cartesian coordinate system. In homogeneous coordinate system the point is represented by $(x, y, 1)$. Let the point be represented by $(X, Y, 1)$ in homogeneous coordinate system after transformation. Then affine transformation of $(x, y, 1)$ to $(X, Y, 1)$ is represented by the equation

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

where the 3×3 coefficient matrix is a composite form of the transformation functions for translation, rotation, isotropic scaling and shear.

Bibliography

[1] Website: http://www.photozone.de/4Technique/compose/dof.htm. Last accessed: 2005.

[2] E. H. Adelson. Depth-of-focus imaging process method. *United States Patent*, (4,661,986), 1987.

[3] A. Akerman. Pyramid techniques for multisensor fusion. *Proceedings of SPIE*, 1828(5):124–131, 1992.

[4] D. Baumann and J. Tinembart. Designing mathematical morphology algorithms on fpgas: An application to image processing. *Lecture Notes in Computer Science*, 3691:562–569, 2005.

[5] I. Bloch. Information combination operators for data fusion: a review with classification. *IEEE Transactions on SMC : Part A*, 26:52–67, 1996.

[6] S. Bradburn, W. T. Cathey, and E. R. Dowski (Jr.). Applications of extended depth of focus technology to light microscope systems. www.colorado.edu/isl/papers, 1998.

[7] L. G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24:325–376, December 1992.

[8] P. J. Burt. The pyramid as structure for efficient computation. In A. Rosenfeld, editor, *Multiresolution image processing and analysis*, pages 6–35. Springer-Verlag, Berlin, Germany, 1984.

[9] P. J. Burt. A gradient pyramid basis for pattern selective image fusion. In *Society for Information Displays (SID) International Symposium Digest of Technical Papers*, volume 23, pages 467–470, 1992.

[10] P. J. Burt and E. H. Adelson. Merging images through pattern decomposition. *Proceedings of SPIE*, 575:173–181, 1985.

[11] P. J. Burt and R. J. Kolczynski. Enhanced image capture through fusion. In *Fourth International Conference on Computer Vision*, pages 173–182, Berlin, Germany, 1993.

[12] A. R. Calderbank, I. Daubechies, W. Sweldens, and B. L. Yeo. Wavelet transforms that map integers to integers. *Applied and Computational Harmonic Analysis*, 5:332–369, 1998.

[13] B. Chanda and D. DuttaMajumdar. *Digital Image Processing and Analysis*. Prentice Hall of India, New Delhi-110001, 2000.

[14] B. Chanda, M. K. Kundu, and Y. Vani Padmaja. A multi-scale morphologic edge detector. *Pattern Recognition*, 31:1469–1478, 1998.

[15] M. Chen and P. Yan. A multiscaling approach based on morphological filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:694–700, 1989.

[16] L. Chipman, T. Orr, and L. Graham. Wavelets and image fusion. *Proceedings of SPIE*, 2569:208–219, 1995.

[17] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multi-modality image registration based on information theory. In Y. Bizais, C. Barillot, and R. Di Paola, editors, *Information Processing in medical Imaging*, pages 263–274. Kluwer, The Netherlands, 1995.

[18] I. De and B. Chanda. Multi focus image fusion using morphological wavelets. In *Workshop on Computer Vision, Graphics and Image Processing (WCVGIP06)*, pages 7–12, Hyderabad, 2006.

[19] I. De and B. Chanda. A simple and efficient algorithm for multifocus image fusion using morphological wavelets. *Signal Processing*, 86:924–936, 2006.

[20] I. De and B. Chanda. An iterative algorithm for registration of multi-focus images by combining global and local transformation models. *International Journal of Image and Graphics*, 2012. Revised and Communicated.

[21] I. De and B. Chanda. Multi-focus image fusion using a morphology-based focus measure in a quad-tree structure. *Information Fusion*, 2012. Accepted.

[22] I. De, B. Chanda, and B. Chattopadhyay. Enhancing effective depth-of-field by image fusion using mathematical morphology. *Image and Vision Computing*, 24:1278–1287, 2006.

[23] H. A. Eltoukhy and S. Kavusi. A computationally efficient algorithm for multi-focus image reconstruction. *Proceedings of SPIE Electronic Imaging*, pages 332–341, June 2003.

[24] Electrical Engineering and Computer Science Department of Lehigh University. Investigations of image fusion. Website: http://www.ece.lehigh.edu/spcrl. Last accessed: 2005.

[25] A. M. Eskicioglu and P. S. Fisher. Image quality measures and their performance. *IEEE Transactions on Communications*, 43(12):2959–2965, 1995.

[26] D. Fedorov. Website: http://vision.ece.ucsb.edu/tileframework. Last accessed: 2010.

[27] D. Fedorov, B. Sumengen, and B. S. Manjunath. Multi-focus imaging using local focus estimation and mosaicking. In *IEEE International Conference on Image Processing (ICIP06)*, pages 2093–2096, Atlanta, GA, USA, October 2006.

[28] A. A. Goshtasby. Image fusion systems research. Website: http://www.imgfsr.com/sitebuilder/images. Last accessed: 2009.

[29] A. A. Goshtasby. Fusion of multi-focus images to maximize image information. *Proceedings of SPIE*, 6229:1–10, 2006.

[30] A. A. Goshtasby and S. Nikolov. Image fusion: Advances in the state of the art. *Information fusion*, 8:114–118, 2007.

[31] J. Goutsias and H. Heijmans. Nonlinear multiresolution signal decomposition schemes. part 1: Morphological pyramids. *IEEE Transactions on Image Processing*, 9(11):1862–1876, November 2000.

[32] J. Hagan. Website: http://www.scribd.com/doc/55364147/17/DEPTH-OF-FIELD. Last accessed: 2012.

[33] R. W. Hamming. *Coding and information theory.* Prentice Hall, New Jersey, 1980.

[34] A. B. Hamza, Y. He, H. Krim, and A. Willsky. A multiscale approach to pixel-level image fusion. *Integrated Computer-Aided Engineering*, 12:135–146, 2005.

[35] H. Heijmans and J. Goutsias. Nonlinear multiresolution signal decomposition schemes. part 2: Morphological wavelets. *IEEE Transactions on Image Processing*, 9(11):1897–1913, November 2000.

[36] F. S. Helmli and S. Scherer. Adaptive shape from focus with an errer estimation in light microscopy. In *Second International Symposium on Image and Signal Processing and Analysis (ISAP2001)*, pages 188–193, 2001.

[37] P. Hill, N. Canagarajah, and D. Bull. Image fusion using complex wavelets. In *Thirteenth British Machine Vision Conference*, Cardiff, UK, 2002.

[38] W. Huang and Z. Jing. Evaluation of focus measures in multi-focus image fusion. *Pattern Recognition Letters*, 28:493–500, 2007.

[39] W. Huang and Z. Jing. Multi-focus image fusion using pulse coupled neural network. *Pattern Recognition Letters*, 28:1123–1132, 2007.

[40] A. Kubota and K. Aizawa. Reconstructing arbitrarily focused images from two differently focused images using linear filters. *IEEE Transactions on Image Processing*, 14(11):1848–1859, November 2005.

[41] A. Kubota, K. Kodama, and K. Aizawa. Registration and blur estimation methods for multiple differently focused images. In *IEEE International Conference on Image Processing (ICIP'99)*, volume 2, pages 447–451, Kobe, Japan, October 1999.

[42] E. Lallier and M. Farooq. A real time pixel-level based image fusion via adaptive weight averaging. In *International Conference on Information Fusion*, France, 2000.

[43] J. S. J. Lee, R. M. Haralick, and L. G. Shapiro. Morphologic edge detection. *IEEE Transactions on Robotics and Automation*, 3:140–156, 1987.

[44] J. J. Lewis, R. J. O'Callaghan, S. G. Nikolov, D. Bull, and C. mCanagarajah. Pixel and region-based image fusion using complex wavelets. *Information Fusion*, 8(2):119–130, 2007.

[45] H. Li, B. S. Manjunath, and S. K. Mitra. Multisensor image fusion using the wavelet transform. *Graphical Models and Image Processing*, 57(3):235–245, 1995.

[46] M. Li, W. Cai, and Z. Tan. A region-based multi-sensor image fusion scheme using pulse-coupled neural network. *Pattern Recognition Letters*, 27:1948–1956, 2006.

[47] S. Li, J. T. Kwok, and Y. Wang. Combination of images with diverse focuses using the spatial frequency. *Information Fusion*, 2:169–176, 2001.

[48] S. Li, J. T. Kwok, and Y. Wang. Multifocus image fusion using artificial neural networks. *Pattern Recognition Letters*, 23:985–997, 2002.

[49] S. Li and B. Yang. Multifocus image fusion using region segmentation and spatial frequency. *Image and Vision Computing*, 26:971–979, 2008.

[50] A. S. Malik and T. S. Choi. A novel algorithm for estimation of depth map using image focus for 3d shape recovery in the presence of noise. *Pattern Recognition*, 41:2200–2225, 2008.

[51] S. Mallat. *A Wavelet tour of Signal Processing*. Academic Press, London, 1998.

[52] P. Maragos. Pattern spectrum and multiscale shape representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:701–716, 1989.

[53] G. K. Matsopoulos, S. Marshall, and J. N. M. Brunt. Multiresolution morphological fusion of mr and ct images of the human brain. *IEEE Proceedings on Vision, Image and Signal Processing*, 141(3):137–142, 1994.

[54] A. Meijster and M. Wilkinson. A comparison of algorithms for connected set openings and closings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:484–494, 2002.

[55] Q. Miao and B. Wang. A novel adaptive multi-focus image fusion algorithm based on pcnn and sharpness. *Proceedings of SPIE*, 5778:704–712, 2005.

[56] N. Mitianoudis and T. Stathaki. Pixel-based and region-based image fusion schemes using ica bases. *Information Fusion*, 8:131–142, 2007.

[57] S. Mukhopadhyay and B. Chanda. Fusion of 2d gray scale images using multiscale morphology. *Pattern recognition*, 34:1939–1949, 2001.

[58] S. K. Nayar and Y. Nakagawa. Shape from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):824–831, 1994.

[59] G. Pajares and J. M. de la Cruz. A wavelet-based image fusion tutorial. *Pattern Recognition*, 37:1855–1872, 2004.

[60] A. P. Pentland. A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):523–531, July 1987.

[61] P. Perlman. *Basic Microscope Techniques*. Chemical Publishing Company, New York, 1971.

[62] V. S. Petrovic. Subjective tests for image fusion evaluation and objective metric validation. *Information fusion*, 8(2):208–216, 2007.

[63] V. S. Petrovic and C. S. Xydeas. Multiresolution image fusion using cross band feature selection. *Proceedings of SPIE*, 3719:319–326, April 1999.

[64] V. S. Petrovic and C. S. Xydeas. Gradient-based multiresolution image fusion. *IEEE Transactions on Image Processing*, 13(2):228–237, February 2004.

[65] G. Piella. A general framework for multiresolution image fusion: from pixels to regions. *Information Fusion*, 4:259–280, 2003.

[66] G. Piella and H. Heijmans. A new quality metric for image fusion. In *IEEE International Conference on Image Processing*, pages 173–176, Barcelona, Spain, 2003.

[67] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging*, 22:986–1004, August 2003.

[68] Xiaobo Qu and Jingwen Yan. Multi-focus image fusion algorithm based on regional firing characteristic of pulse coupled neural networks. In *Second International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA07)*, pages 62–66, Zhengzhou, 2007.

[69] T. Ranchin and L. Wald. The wavelet transform for the analysis of remotely sensed images. *International Journal of Remote Sensing*, 14:615–619, 1993.

[70] L. G. Roberts. Machine perception of three dimensional solids. In J. T. Tippet, editor, *Optical and Electro-optical Information Processing*. MIT Press, Cambridge, Massachusetts, 1965.

[71] O. Rockinger and T. Fechner. Pixel-level image fusion: The case of image sequences. *Proceedings of SPIE*, 3374:378–388, 1998.

[72] P. Scheunders. Multiscale edge representation applied to image fusion. *Proceedings of SPIE*, 4119:894–901, 2000.

[73] W. Seales and S. Dutta. Everywhere-in-focus image fusion using controllable cameras. *Proceedings of SPIE*, 2905:227–34, November 1996.

[74] J. Serra. *Mathematical Morphology and Image Analysis*. Academic Press, London, 1982.

[75] P. Soille. *Morphological Image Analysis*. Springer, Berlin, 2003.

[76] M. Subbarao, T. S. Choi, and A. Nikzad. Focusing techniques. *Optical Engineering*, 32(11):2824–2836, 1993.

[77] W. Sweldens. The lifting scheme: A new philosophy in biorthogonal wavelet constructions. In A.F. Laine and M. Unser, editors, *Wavelet Applications in Signal and Image Processing*, volume III, pages 68–79. Proceedings of SPIE 2569, 1995.

[78] A. Toet. Hierarchical image fusion. *Machine Vision and Applications*, 3:3–11, 1990.

[79] Y. H. Tsai and Y. H. Lee. Wavelet-based image fusion by adaptive decomposition. In *International Conference on Intelligent Systems Design and Applications (ISDA08)*, pages 283–287, Kaohsiung, 2008.

[80] P. Viola and W. M. Wells. Alignment by maximization of mutual information. In E. Grimson, S. Shafer, A. Blake, and K. Suguhara, editors, *International Conference Computer Vision*, pages 16–23, Los Alamitos, CA, 1995.

[81] H. H. Wang. A new multiwavelet-based approach to image fusion. *Journal of mathematical imaging and vision*, 21:177–192, 2004.

[82] Rong Wang, Li-Qun Gao, Shu Yang, Yu-Hua Chai, and Yan-Chun Liu. An image fusion approach based on segmentation region. *International Journal of Information Technology*, 11(7):92–100, 2005.

[83] Z. Wang and A. C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, March 2002.

[84] Z. Wang, Y. Ma, and J. Gu. Multi-focus image fusion using pcnn. *Pattern Recognition*, 43(6):2003–2016, 2010.

[85] B. Yang and S. Li. Multifocus image fusion and restoration with sparse representation. *IEEE Transactions on Instrument Measurement*, 59(4):884–892, April 2010.

[86] X. Yang, W. Yang, and J. Pei. Different focus points images fusion based on wavelet decomposition. In *Third International Conference on Information Fusion*, volume 1, pages 3–8, 2000.

[87] Y. Zhang and L. Ge. Efficient fusion scheme for multi-focus images by using blurring measure. *Digital Signal Processing*, 19:186–193, 2009.

[88] Z. Zhang and R. S. Blum. A region-based image fusion scheme for concealed weapon detection. In *30th Annual Conference on Information Sciences and Systems, Johns Hopkins University*, pages 168–173, Baltimore, MD, March 1997.

[89] Z. Zhang and R. S. Blum. A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application. *Proceedings of IEEE*, 87:1315–1326, August 1999.

[90] Z. Zhang and R. S. Blum. A hybrid image registration technique for a digital camera image fusion application. *Information Fusion*, 2:135–149, June 2001.

[91] Z. Zhang and R. S. Blum. Image registration for multi-focus image fusion. *Proceedings of SPIE*, 4396:279–290, 2001.

[92] B. Zitova and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21:977–1000, November 2003.

List of Publications Related to the Thesis

- Ishita De and Bhabatosh Chanda. Multi focus image fusion using morphological wavelets. In *Workshop on Computer Vision, Graphics and Image Processing (WCVGIP06)*, pages 7–12, Hyderabad, 2006.

- Ishita De and Bhabatosh Chanda. A simple and efficient algorithm for multi-focus image fusion using morphological wavelets. *Signal Processing*, 86:924-936, 2006.

- Ishita De, Bhabatosh Chanda, and Buddhajyoti Chattopadhyay. Enhancing effective depth-of-field by image fusion using mathematical morphology. *Image and Vision Computing*, 24:1278-1287, 2006.

- Ishita De and Bhabatosh Chanda. An iterative algorithm for registration of multi-focus images by combining global and local transformation models. *International Journal of Image and Graphics*, 2012. Revised and communicated.

- Ishita De and Bhabatosh Chanda. Multi-focus image fusion using a morphology-based focus measure in a quad-tree structure. *Information Fusion*, 2012. Accepted.