Chapter 1

Introduction

Vision is the most important form of sensual information for human beings. This fact has motivated researchers to look at processing of images by machines. But the underlying complexities have made such an attempt difficult, but a challenging one. Processing of images has a long history; starting from mid 1920 onwards [39], and picking up in mid 60's. Processing of images gained popularity with the availability of low cost, powerful and faster computers and efficient algorithms. For a typical image processing system, one needs [59]:

1. Image acquisition module - like a camera or a scanner
2. A storage module - Magnetic or optic media
3. A powerful processing module - like a computer system
4. Output unit - like a high resolution monitor, a printer or a plotter
5. A good interface to integrate the above and
6. An efficient algorithm

The application spectrum of image processing is very broad. This includes areas like robotics, remote sensing, image storage/retrieval/transmission system, medical imaging, surveillance, entertainment etc. In almost all of these applications, generation of (a) a sharper image void of blur and noise, and/or (b) generation of a high resolution image (zooming/magnification/enlargement) is desired. In this thesis, we address both these tasks.

1.1 Image Restoration

Generally, image degradation is due to blur (motion or out-of-focus) and/or corrupting noise. The goal of image restoration is to recover original image from an observed degraded image. Image restoration techniques are oriented towards modeling these degradations: blur and noise, and then, applying an inverse procedure to obtain an approximation of the original image. Image restoration models normally involve equations which leads to ill-posed problems [41, 131]. Ill-posed problems have one or more of the following properties:

- Solution does not exist. A simple example would be a matrix operator whose determinant is zero, normally, a singular matrix.
- Solution exists, but is not unique. For example, simultaneous equations, where number of equations are less than the number of unknown parameters.
- A unique solution exists, but a small perturbation in the input (noise) will produce a large variation (or oscillations) at the output. For example, an ill-conditioned matrix operator.

Ill-posed problems can be tackled by *regularization* [136]. Regularization imposes smoothness constraint, which reduces the solution space. A typical model for image observation will be

$$y(i,j) = \phi\left\{\sum_m \sum_n h(i,j;m,n)x(m,n)\right\} \odot n(i,j)$$

where ϕ may be a non-linear operator and $\odot$ is an inevitable operator, (for example addition operation). Values m, n correspond to size of the *blurring window.* If the system is linear and time invariant and the noise $n(.)$ is additive, we can rewrite the above equation (for additive noise) as

$$y(i,j) = \sum_m \sum_n h(i-m, j-n)x(m,n) + n(i,j) \tag{1.1}$$

where $h(.)$ is the blurring function, typically called *Point Spread Function* (PSF). It tells us how a point *spreads* due to blurring. For analysis purposes, we can express (1.1) in the matrix form, by lexicographically ordering (row stacked columns) y and x as

$$Y = \mathbf{H}X + N \tag{1.2}$$

where, Y, X and N are lexicographically ordered column vectors of $x(.,.)$, $y(.,.)$ and $n(.,.)$ respectively, and $\mathbf{H}$ is a PSF matrix of $h(.,.)$. Pictorially it is represented in Fig. 1.1. For an image of size $M \times M$, size of X, Y and N vectors will be $M^2 \times 1$ and $\mathbf{H}$ is matrix of size $M^2 \times M^2$. If we consider color images with each pixel represented by red, green and blue component, then size of X, Y and N will be $3M^2 \times 1$ and $\mathbf{H}$ will be $3M^2 \times 3M^2$.

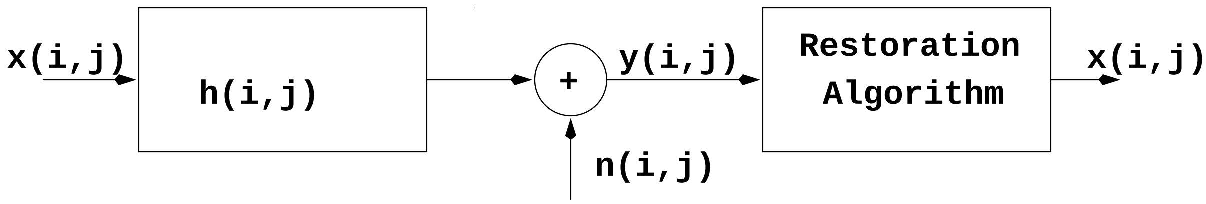

Figure 1.1: Image restoration problem

Most of the image processing algorithms assume $\mathbf{H}$ to be known and that the statistics of N are known. Blurring function decides the kind of the blur: for example, spatial blur or motion blur. Spatial blur can be space variant or space invariant. For a space variant blur, $\mathbf{H}$ is a function of space; for space invariant blur, it will be a constant. Two popular values for space invariant blur are local averaging, where, all the elements of $\mathbf{H}$ matrix will be identical, and Gaussian, where the $h(.)$ matrix is Gaussian distributed. Once $\mathbf{H}$ is known, the problem of estimating X is an inverse problem. These ill-posed problems are solved using constrained optimization methods, smoothness being a general constraint. Different methods are attributed to the modeling and the kind of optimization algorithm adopted. In this thesis we address restoration of color images.

1.2 Image Zooming

Image zooming belongs to a class of techniques, which enhances image *resolution*. Image resolution can be described as the smallest measurable detail in visual perception. Resolution can be classified into three types:

1. *Spatial Resolution*: It is the spacing between adjacent pixels - measured as dots per inch (dpi) or pixels per inch (ppi). Higher dpi or ppi results in a higher resolution and a more detailed image.

2. *Brightness Resolution*: It is the number of brightness levels that can be conveyed by a pixel. Typically it is 8-bits for gray scale images, and 24-bits for color images.

For a color image, each of the RGB (or other color-coordinate) is 8 bits, totaling to 24-bits.

3. *Temporal Resolution*: It is the number of frames per second. Also called frame rate. Typically, it is about 25 to 30 frames per second.

Spatial resolution enhancement or zooming involves insertion of pixel values between pixels. The technique here is to pad each row and columns with zeros and then pass the image through a filter. Type the and order of the filter gives rise to different zooming methodologies, simplest being zero-order hold or pixel-replication. Here adjacent pixels are copied into empty locations. Because of pixel replication, this gives a blocky effect. Other popular method to fill the vacant pixel is by averaging neighboring pixels. Because of averaging, this will tend to smooth the edges. Many methods are proposed in literature to strike a balance between these two. Zooming can also be considered as an ill-posed problem, as there is no *unique* value for the pixel that is being filled.

1.3 Video Zooming

The amount of data needed to represent video is very high [58, 95], hence data compression is needed for efficient storage and transmission. Many standards have been proposed for image and video compression like BTU-T, H.261, H.263, MPEG etc., to cater to a wide range of applications. Compromise is struck between bit-rate, picture quality, complexity, error criterion and delay. A review of these techniques is found in [7, 104] and the references therein. For many applications, one needs to obtain a compressed bit-stream containing the video at a different resolution than the original compressed bit-stream, that too, in *real time*. Typical requirements of such an application are:

1. Browsing remote video database - it is economical to send a low-resolution version of video clip to the user and depending on ones' interest, the user can progressively enhance the resolution.

2. Video transmission over low bit-rate and/or low priority network.

3. Re-sizing is also needed when a viewer is watching two different TV channels on the same TV (picture-in-picture).

4. Conversion between different TV standards or to fit the incoming digital video onto the users' TV screen, like HDTV.

A straight-forward approach of decompressing, carrying out the down-sampling in spatial domain and then re-compressing involves computationally intensive work load. This is more so, because, motion estimation and compensation is often employed during compression, which are compute-intensive operations and account for 60% of the workload of a video encoder [10]. Most of the early research on multimedia applications was focused on compression standards [10], synchronization, storage and software issues [79, 97]. One of the early references for manipulation of video in compressed domain is by Smith and Rowe [130]. They propose techniques to manipulate the compressed image data obtained from run-length coding (RLC). They assume that original data is partitioned into non-overlapping 8×8 blocks. RLC of these blocks are obtained after zig-zag scanning of Discrete Cosine Transform (DCT) coefficients of these blocks. Of late, scenario is changing, with researchers looking at new avenues for down-sampling and other operations on video sequences directly in the compressed domain. These operations do not require decompression and re-compression, making video processing efficient. Dugad and Ahuja [27, 28] have suggested a method to down-sample still images in the DCT domain. A similar problem for video was tackled by Shen *et al* [15]. In this thesis we develop video zooming algorithms which work in the compressed domain.

1.4 Contributions

The major contributions of this thesis are:

- *Color Image Restoration:* Here we assume that the degradation is across the Red-Green-Blue (RGB) planes. That is, degradation in the red plane of a color image depends not only on the values of the red plane, but also on the green and blue planes. Our model is robust in the sense that the performance is independent of amount of (partially known) inter-channel degradations. The image is modeled as a Markov Random Field (MRF), and energy minimization is done using simulated annealing (SA). We have considered a linear blur and additive white Gaussian noise for the degraded image. Novelty of the proposed method is that it considers blurring across the channels, and, performs well when this interchannel blurring is unknown.

- *Still Image Zooming:* We propose a novel wavelet and Multiresolution analysis (MRA) based approach for still image zooming. The idea here is to estimate wavelet coefficients at the finer scale. This leads to a sharper zoomed image compared to the ones obtained by sinc or spline interpolation. We also propose MRF modeling and joint MRF and MRA approaches. Comparison is carried out with conventional techniques like pixel replication and linear interpolation. We also compare our method of estimating the wavelet coefficients with the wavelet coefficients obtained by using a scaling function based interpolation, and we observe that wavelet coefficients obtained by both the methods are similar. Comparison with conventional methods show the proposed method is better, both visually and numerically.

- *Video Zooming:* If a motion compensated video is to be up-scaled, then conventional approaches for generating an MPEG, H.261, or H.263 bit-stream requires that the video be decompressed, up-scaled in the spatial domain, motion vectors be re-computed for up-scaled video, followed by re-encoding in a conventional video encoder. This entire process is compute-intensive. We propose to operate directly in the compressed domain, this is achieved by a novel scheme for interpolating motion vectors. Our approach takes significant load off the server with some load added on the client. Manipulating motion vectors gives flexibility to apply our propose method for different compression methods. We justify this by considering the DCT and the Discrete Wavelet Transform (DWT) based compression methods. We consider these two methods because of their popularity in compression methodologies. Current compression standards, like MPEG, use DCT based compression. We expect future video compression standards to use DWT; this statement is motivated by the fact that JPEG2000[1] is based on DWT.

  Simulation results show that DWT based zooming method performs better than the DCT based method. The algorithm is tested on a video stream compressed by Multiresolution Motion Estimation (MRME) method. Performance under such condition is decided by the coder-decoder efficiency. We use the idea of motion vector interpolation to achieve temporal (missing frame) interpolation. The interpolated motion vectors point to a position in the missing frame and the frame in the neighborhood of this location is filled from the previous frame. We compare our method with linear interpolation of pixels in the decompressed video.

1.5 Organization of The Thesis

The thesis is organized into 7 chapters. Each chapter is relatively self-contained, in the sense that the needed mathematical material and some of the published literature

[1] http://www.jpeg.org/JPEG2000.htm

are introduced in the first part of each chapter (*we do not have a separate chapter on literature survey*). Subsequent sections are devoted to the proposed approaches. This is followed by results and discussions. Each chapter concludes with a summary.

Second chapter introduces some mathematical preliminaries. This includes: theory of MRF, wavelets, and MRA. We use MRF for image restoration and zooming, and, wavelets and MRA for image zooming and video zooming applications. We introduce both one and two dimensional DWT.

Third chapter deals with Robust Color Image Restoration. We begin with an overview of some of the techniques for image restoration, and review a few of the methods reported in literature. We then formulate a scheme for color image restoration and show the results for the proposed formulation.

Fourth chapter is on still image zooming. We discuss some of the conventional approaches, as we compare our method with these. We give the formulation of image zooming for MRA and MRF based approaches, and extend it for color images. We show that the proposed method gives better quality of image, both numerically and visually.

Fifth chapter discusses background material for image compression. We look upon zooming as the dual of compression. Since our concern is to work in the compressed domain, discussion on basic image compression techniques is important. Two methods discussed here are Embedded Zerotree wavelet (EZW), and Set Partitioning In Hierarchical Trees (SPIHT).

Sixth chapter is devoted to the proposed video zooming. We show that the interpolation of motion vectors in compressed domain gives good results. Motion vectors are interpolated in DCT or DWT (depending on whether video compression was achieved using DCT or DWT) domain to obtain a zoomed video. Implementation of this algorithm to Multiresolution Motion Estimation (MRME) is discussed. We extend the proposed method for temporal interpolation in the last part of this chapter.

Seventh and the last chapter summarizes and ends with some concluding remarks. It also suggests future directions for the proposed work.

Color coordinate system and a derivation of MRF models are discussed in the appendixes.