A New Approach to Measuring Image Quality via Frequency Analysis

Gian Paolo Lorenzetto

This report is submitted as partial fulfilment of the requirements for the Honours Programme of the Department of Computer Science The University of Western Australia 1999

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

Given that there are few tasks as subjective as that of a human viewer comparing two images, how can image quality be quantified? Two of the most popular error measures are the *root mean square* (RMS) error, and the *signal-to-noise ratio* (SNR). Both these measures are of little use in determining image quality since they are merely tallies of pixel difference. Two images can have disparate pixel values at corresponding points yet still appear identical. Consider two images differing only by a translation. These images have no correspondence in terms of grey values, yet still appear the same.

An extensive amount of work relating to image manipulation routines has been published. Often these articles compare routines in terms of the quality of images produced. However, it is impossible to draw any parallels between two similar routines without a reliable measure of image quality. This is evident in the design of compression algorithms, for which the need to measure the quality of a reconstructed image is essential in determining the compression algorithms performance. In fact many compression algorithms' have their parameters set to minimise RMS error [15].

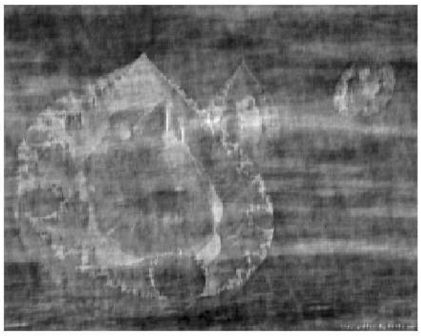
Currently, there is a "severe failure" with image comparison techniques [2]. Two such techniques are the RMS and SNR errors. However, although inherently flawed, there are two reasons the RMS and SNR errors are still in use. First, the simplicity of the RMS error means it has a computational advantage over more complicated approaches, and second it provides a theoretically sound, statistical measure of the mean difference in pixel intensity. That is, the RMS will correctly rank two greyscale images in order of compression. However, the RMS is not useful when applied to binary images, images involving text, or translated images.

There are many existing image comparison measures, ranging from using human viewers to rank images, to models based on the human visual system. However, even the best of these measures can only claim limited success in restricted environments. Chapter 2 gives some background on the types of image comparison measures in use, and their individual strengths and weaknesses.

The following chapters present a new approach to image comparison employing

(a)

(b)



(c)

Figure 1.1: Phase information from image a) and magnitude information from image b) have been combined to create the synthetic image c). The synthetic image clearly corresponds to image a), from which the phase information was extracted.

phase information. The importance of phase information in an image is highlighted via an experiment first performed by Oppenheim and Lim [10]. The experiment involves constructing a synthetic image from two original images, by extracting the phase information from one and the magnitude information from the other. Figure 1.1 illustrates this experiment. The resulting image clearly corresponds to the image from which the phase information was taken. From this result it is evident that phase is important in an image, and consequently a measure based on phase may be used to overcome some of the flaws inherent in RMS type errors.

Daly has stated that one of the problems with current comparison measures is their lack of phase information in image analysis [2]. Moreover, results obtained for the new phase based measure show that, at least partially, some of the problems associated with pixel tally measures are overcome.

A Wavelet approach is employed to perform image analysis. Quadrature pairs of *Log Gabor* filters are convolved with the image at various scales. Log Gabor filters are employed since they allow the calculation of phase and amplitude information localised in both space and frequency. Consequently, frequency information can be calculated local to each point in an image.

A *phase difference* between frequency information extracted from each image is calculated to measure distortion and estimate phase displacement. The estimated phase displacement is used to calculate a measure of translation. A displacement is measured by observing the offset between phase at each point. A constant offset greater than zero implies a translation between images. Distortion is measured by calculating the standard and absolute deviations of the phase difference. Distortion can be calculated irrespective of a translation between images, since the phase difference between identical images is constant; the standard deviation will be zero, or near zero, with or without a displacement.

Comparing displaced images often results in a marginal increase in measured distortion due to two problems. The first is known as the *wrap around* problem, and is caused by a large translation changing the characteristics of the image signal under analysis. This results in an inconsistent phase difference, causing a rise in the standard deviation, and consequently an increased distortion.

By assuming translations of no more than 15 pixels occur the influence of the wrap around problem is minimized.

The second factor effecting the accuracy of the measured distortion is a weak filter response. Given a point in an image, a weak filter response results in spurious phase data at that point. Varying scales of filter are applied to the image in the hope that one filter will give a strong response at every point in the image.

The proposed measure is a relative metric and it can not be used as an absolute number. However, as indicated by the problems with the RMS and SNR measures, image quality cannot be judged on a single number. Rather, a range of image attributes must be considered. Image quality is further discussed in Chapter 3.

A detailed discussion of extracting phase information is presented in Chapter 4. In particular the reasons for choosing wavelet analysis are discussed, calculation of phase and amplitude information is covered, and several methods developed for overcoming weak filter responses are outlined. Chapter 4 also illustrates calculating measures of translation and distortion.

Discussions of phase information as a comparison measure are extremely sparse in the computer vision literature, and consequently the central aim of this work has been to perform a preliminary investigation into using phase information from images as a measure of image difference. The main thrust behind this work is the need for a new image quality measure and the intuitive use of phase information as a technique for quantifying image quality. Results for the new phase based image comparison metric are presented in Chapter 5.

Finally, Chapter 6 summarizes the work presented, and several key areas of research are proposed as further work.

CHAPTER 2

Existing Measures of Image Quality

In general, it is easy for the human visual system (HVS) to assess the quality of two similar images and decide on which one looks better. Many methods have been proposed that attempt to define a metric to emulate a "looks better" comparison. These metrics fall into one of three categories, as proposed by Wilson *et al.* [15]. They are

1. human perception,
2. objective measures based on theoretical models, and
3. subjective measures based on mathematically defined models of the HVS.

The first category of metrics employs a selected group of human viewers to judge which image they believe to be the best quality from a range of selected images [14]. The second category uses proven, mathematically based, theoretical models to take advantage of the fact that images can be represented as a matrix of numerical values. Consequently, standard (or non-standard) transforms can be applied to the matrix. The third category is the most difficult to implement. This is because the basic functional components of the HVS are not well understood. However, it would seem a logical approach to model this system. Unfortunately, to mimic the functional components of the HVS is very difficult, and any such systems are generally only useful in a restricted environment.

A review of each of the three categories will be presented by discussing articles relating to the three different approaches. The measures are in general designed for a specific area of image comparison. However, where possible comparison will be made between similar approaches. The main advantages and disadvantages of each method will also be addressed.

2.1 Human Perception Metrics

2.1.1 Defining the Quality of an Image

A less intuitive approach to image comparison involves predicting the number and type of "dimensions" a viewer uses in rating a series of images. Ahumada and Null describe this approach to human perception metrics in their paper "Image Quality: A Multidimensional Problem", where the term *preference factoring* is used in reference to this method. Preference factoring involves having a range of observers rank a selection of images, and then use the "inter-observer variability to extract dimensions of image quality" [14].

Although effective at extracting the number of dimensions that a viewer uses in rating images, the first method Ahumada and Null describe, does not give any information about each dimension. For example, assume an observer uses two dimensions to rate a selection of images. There is no way to tell what the dimensions are. The first dimension may be "blockiness" and the second "blurriness". The first could be quality and the second near-sightedness (presuming a particular compression method was less visible to near-sighted observers).

A second approach, described by Ahumada and Null and performed by Farrel *et al.* [14], illustrates the type of information gathered by the preference factoring approach. A group of 18 observers ranked a reference image displayed with the reference image compressed and reconstructed. The images are compressed with either the discrete cosine transform or non-uniform sampling and interpolation method, to varying levels. Farrel *et al.* recovered two meaningful dimensions differently weighted by the observers. The first corresponded to compression ratio (bits/pixel) and the second corresponded to the different compression algorithm used.

A more useful approach to preference factoring is a procedure Ahumada and Null call INDSCAL. The authors claim it provides information about the dimensions the observer used in judging the images. This method stresses the need for each observer to rate each display under strict conditions, otherwise the results are rendered meaningless. The INDSCAL procedure also assumes that the effect of the conditions on both observer weights and display weights is equivalent.

Both methods of image comparison described by Ahumada and Null are not a true *Human Metric*, as both methods attempt to predict user responses to images, rather than simply using the responses. It is included in this first category of image comparison types because it is based on using groups of people rating images to then interpret image comparisons. As Lubin [11] notes, if no reliable model is available this is the only way of gathering information about the imaging system.

The main disadvantage of this technique is the high cost involved, both in terms of time and money. It is especially expensive when observers need to be trained before they can be used in studies. Maintaining exactly the same conditions for all observers is also extremely difficult and expensive.

2.1.2 Confirming Image Quality

Hearty's article entitled "Achieving and Confirming Optimum Image Quality" [4] describes his main concern as optimizing compression tools for television images. The task of compressing an image with minimum quality loss is broken into two complementary techniques. These are

1. Perceptual Models and Data. He considers the importance of accounting for shifts in gaze during viewing and examines the feasibility of using predicted gaze patterns to optimize source-coding procedures.

2. Assessments of Image Quality. This should be done under highly controlled conditions and with rigorously applied methods. This corresponds to what Ahumada and Null have suggested in Section 2.1.1: results are meaningless for this type of experiment if tests are not conducted under strict conditions.

Hearty notes that one of the main disadvantages of a perceptual/assessment model is that most successful models have been developed to deal with only a specialized processing or transmission environment. Consequently, each model is limited in its usefulness.

Hearty goes on to suggest that assessment by viewers is the most convincing of methods. He breaks this approach down into three subtasks as follows:

1. Assessment Method. This involves soliciting viewer judgments.

2. Assessment Participants. This involves selecting individuals to judge performance.

3. Assessment Materials. This involves the selection of suitable images used to illustrate the performance of the system.

Hearty makes the point that not only can diagnostic assessments of image quality be used to identify specific areas in which systems should be improved, but also areas in which systems can ease off without appreciable loss. This is important when considering lossless image compression where perceptually equivalent images appear the same to the human visual system.

2.2 Objective Measures

2.2.1 Grey-Scale Image Comparison

Wilson *et al.* [15] define a method based on exploiting topologies in grey-scale images. They claim that in many real images the morphology of the image represented as a function is related to the underlying scene. Consequently, two similar images can be compared by their topologies.

The authors propose a method to overcome some of the faults inherent to the root mean square (RMS) and signal-to-noise ratio (SNR) errors. These errors only incorporate tallies of pixel difference. No information about location or intensity of pixel values is included in the resulting error. Both Hearty [4] and Daly [2] agree, stating that the RMS and SNR errors are inaccurate predictors of perceptual distortion (perceptual distortion is discussed further in Chapter 3).

To overcome these problems, Wilson *et al.* propose a new measure Δ_g based on the Δ measure developed by Baddeley [15] for binary image comparison. The Δ measure is based on a *myopic topology*, generated from the Hausdorff metric.

Wilson *et al.* extend this metric to grey-scale images by defining the distance between any two grey-scale images as the distance between their respective subgraphs. In this way they preserve the theoretical basis upon which it was derived: to create a metric that generates the most appropriate topology when dealing with images.

The authors cite a concern with the high computational cost of the new Δ_g measure. This proved to be problematic, as their experiments indicated their new metric takes substantially longer to calculate than either the RMS or SNR errors.

However, not all results are unfavorable. Given two chess board images, the second having the black and white pixels reversed, the RMS error returns a maximum value of 1. This is understandable as every pixel intensity is different between the two images. However Δ_g returns an error of only 0.249. Moreover, this example corresponds to two images that have little discernible difference to the human eye, indicating a good result in terms of detecting perceptually significant errors.

Wilson *et al.* suggest that for certain types of distortion the response of Δ_g is a significant improvement over the response of RMS. But they add that further improvement can be made and encourage further work in the area.

2.2.2 Phase Based Image Comparison

Lees and Henshaw's [5] paper discusses printed circuit board inspection, and propose an approach to this based on image comparison. The authors cite a concern with the amount of time and fatigue involved in human inspection of images as the main reason for developing a reliable image comparison measure. Inspection by human observers is also very expensive. Non-contact optical inspection methods are desirable, since contact devices such as bed of nails continuity testers can damage traces in the testing process. Lees and Henshaw also state that optical testers are capable of finding defects other than shorts and opens, such as line width errors, pad mouse bites and trace misplacements.

Conventional template matching subtracts a "golden" board image derived from CAD/CAM or image data, from the image of a board under inspection to yield a difference image in which defects are emphasized. While computationally simple this suffers from problems including registration, colour variation, reflectivity variation, and lighting sensitivity. Lees and Henshaw claim a phase only method is light intensity invariant, insensitive to illumination gradients, and has a high tolerance to misregistration of the images to be compared.

Phase-only image comparison has the desirable properties of redundancy removal and edge enhancement. Lees and Henshaw extract phase information using the Fourier transform, then "whiten" the image, and then perform an inverse Fourier transform to produce a map of "significant" image difference. An image is whitened by setting the amplitude component to unity. The authors cite several examples using this technique, illustrating that areas of difference between images result in a high deviation from the original image.

The most interesting property of phase-only image comparison is that of redundancy removal; two similar images can be compared simply by placing them side-by-side to create a composite image, then apply a phase-only transformation to the entire composite image. Lees and Henshaw then suppress the strong periodic component corresponding to similarities between the two images, leaving only the differences.

A second important property of phase-only images involves automated analysis of the processed data. Because the mean and variance are determined by the power spectrum of the image, they are identical for all phase-only images, regardless of the intensity of the starting image. A phase-only image is one having unit amplitude. Lees and Henshaw show for an image with N pixels the mean and variance are

$$\bar{x} = \tfrac{1}{N},$$

$$\sigma^2 = \tfrac{1}{N}.$$

This is important because it means many printed circuit boards' can be tested without worrying about both the brightness of the board (the lighting), and any reflections from the board.

Since the number of pixels will be large, the mean value of the phase-only image will be near zero. The fixed variance of $\frac{1}{N}$ will allow the probability of false detection of defects to be determined for a given threshold, regardless of the input image statistics. Thus many different printed circuit boards' could be tested, without constantly adjusting the thresholds or other parameters of the defect detection process.

Lees and Henshaw test the phase-only imaging technique by applying it to digitized printed circuit board images with simulated lighting and registration errors and artificially inserted defects. They use two images, one a composite consisting of a left half with a defect and a master on the right. The second image is the result of phase-only processing of the composite image.

The authors then cite eight examples, all of which give favorable results and many of which could not be done by simple matching and/or subtraction.

Although they have described very favorable results, it must be questioned whether the increase in complexity and decrease in size of circuit boards has made this approach infeasible or less efficient than other approaches. In particular, the authors suggest specific hardware would be required to make the comparison process feasible. However, Daly in the next section also agrees that incorporating phase information into an image is important to detect perceptually important errors.

Both Section 2.2.1 and 2.2.2 attempt to predict the quality of an image by providing mathematically based models. In the next section, a model designed as a functional equivalent to the HVS is presented.

2.3 Human Visual System Models

2.3.1 Assessing Image Fidelity

Daly [2] defines the visual differences predictor (VDP) algorithm as a means of measuring the human visual response. The VDP is motivated by the need to quantitively describe the visual consequences of decisions regarding the design and quality control of imaging products. In essence, what starts as a mathematical difference ends as a visual difference.

Daly states that the problem with many popular techniques is their lack of phase information in analysis, and their simplicity relative to the complexity of the human visual system. Compared to the simplicity of RMS type errors, a measure based on the human visual system is extremely complicated.

The VDP is a relative metric describing the difference between one image in terms of another. Input to the model consists of two images combined with parameters for viewing conditions and calibration. Output consists of a third image that maps the probability of detecting the differences between the two images as a function of their location in each of the original images. Daly summarizes the model as a thresholding technique capable of quantifying the important interactions among threshold differences and suprathreshold image content and structure.

The VDP does not produce a single number measure. This has the advantage that the nature of any differences can be seen and used to improve the design of the imaging product. It can also be used for all types of image distortions including blur, noise, algorithm artifacts, banding, blocking, pixelation, and tone-scale changes.

The VDP contains three components:

1. Calibration Mechanisms. The input images must be calibrated for the HVS model to have any meaning.

2. Human Visual System Model. This model concentrates on the lower-order processing of the visual system. It essentially treats the HVS as a number of processes that limit visual sensitivity.

3. Difference Visualization. Daly proposes two ways to visualize data:

   (a) Free-Field Difference Map. The visible difference predictions appear on a uniform field with a grey value near the system mean. With this method it is sometimes difficult to see where the predicted errors correspond in the input images.

   (b) In-Context Difference Map. This map overcomes the free-field limitation by mapping the output probabilities in colour on the reference image.

Daly comments that the "severe failure" of the common image error metrics (like RMS) is reason enough for a HVS model, such as the VDP, to be developed. Daly also states that the VDPs performance is sufficient to warrant formal psychophysical tests to quantify its degree of success. However no formal results are mentioned.

2.4 Overview

There are many approaches to designing image comparison metrics. These metrics can be classed into three categories as described by Wilson *et al.* [15]. It is often the case that these methods overlap, as is the case with the articles reviewed. Ahumada & Null base their model's prediction on human observer responses, Hearty describes a method of using human viewers to assess image quality, Lees & Henshaw and Wilson *et al.* both propose purely mathematical models, and Daly proposes a metric based on a functional model of the HVS, incorporating mathematical functions to simulate some functional components. All the approaches reviewed claim to have been successful in at least a restricted environment. Moreover, all suggest further study in the area of image comparison metrics is essential.

CHAPTER 3

Image Quality

Presently the most frequently used image comparison measures, such as RMS, do not give a reliable measure of *perceptual distortion* [17, 2, 4], and cannot compare translated images, or images containing text. Perceptual distortion can best be described through the use of an example, namely comparing the relative compressions of the *burst* test pattern image and the *boat* image, both of which are presented in Figure 3.1. Consider for both images the original and 95% compressed (and reconstructed) versions. Perceptually the compressed burst image appears less distorted, with respect to the original *burst* image, than does the compressed *boat* image.

The difference that is observed between perceived distortion and actual compression raises an important issue: does an image comparison technique measure the compression in terms of a percentage, or does it measure the perceptual distortion?

When designing compression algorithms, compression ratios are already known, so a measure of compression is generally not required. Furthermore a measure of compression is meaningless with respect to other types of image manipulation routines in which compression plays no part. It is more useful to measure the type and amount of distortion present: a perceptual measure.

One of the problems with the RMS error is that it fails to provide an indication as to the type of degradation in an image. It is important to understand the difference between two images before making a judgment as to the quality of the image. Consider a translated image. A measure of the images "quality" should be invariant to the translation, opposed to the RMS error that is unable to distinguish between translation and degradation. Moreover, a true measure of image quality is one that can separate the many different aspects of distortion between two images. These aspects include image translation, rotation, compression, and enhancement. The new phase based measure proposed corrects at least one problem common to the RMS error, that is, separating translation from degradation.

Another fault of the RMS error is that all differences between images are given equal perceptual significance. Again, consider a translated image displayed next to the original. The images appear the same, yet the RMS error measure returns a large difference, because of the disparate pixel values.

The new phase based metric is a measure of perceptual distortion, and can be thought of as similar to Daly's VDP [2] in that it produces a measure of relative difference. This result is intuitive as the distortion measure is based on a comparison of the phase between two images, not on a compression ratio or simple tally of pixel difference. Consequently, differences can be weighted. For example, a large translation is not measured as a large distortion. Thus rectifying another of the problems inherent to the RMS.

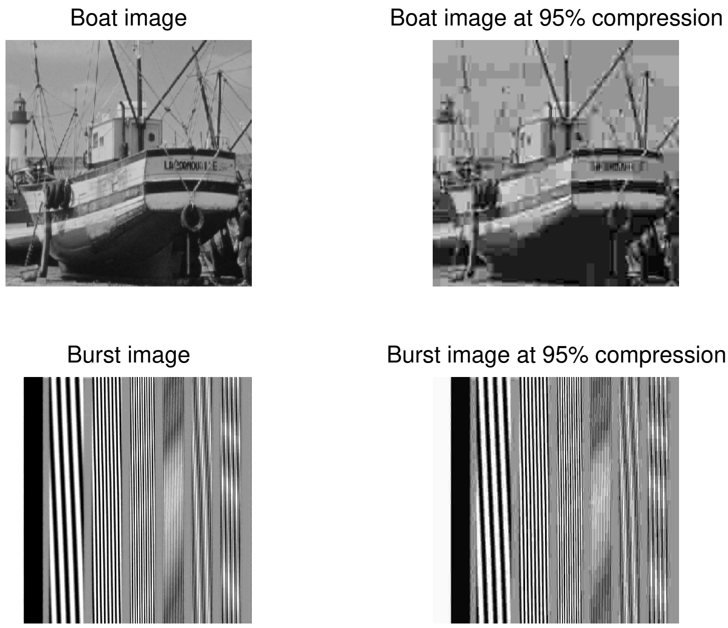


Figure 3.1: The original *boat* and *burst* images, and their respective 95% compressed versions.

Given that the new comparison measure is one based on perceptual distortion, the next question is how can it be validated? The only form of validation available when considering image comparison measures is one discussed by Daly for the testing of the VDP [2]. The method involves comparing images rated by human subjects to the rating given by the phase based measure. Unfortunately, this approach requires extremely strict conditions with rigorously applied methods [4], and makes the assumption that all viewers determine image quality in the same fashion [14]. As Ahumada and Null [14] note, it is difficult to measure the assessed human responses, and although Hearty [4] suggests that assessment by human subjects is the most convincing of methods, it is impossible to check the correctness of any results obtained.

The lack of a formal validation makes the task of assessing the performance of the new phase based measure extremely difficult. It was decided that a correct result was one such that an image of higher compression is given a higher measured distortion. Similarly an image that is modified from the original is also deemed to have been distorted. The criteria for performance rating has the perhaps undesirable effect that a sharpened image, deblurred image, or any similarly enhanced image, is categorized as distorted. Given that the new measure is a relative one, this result is of little importance, and an image, even if appearing better than the original, is still distorted relative to the original. The important factor is whether the image has been distorted, and type of distortion, not whether the image is more aesthetically pleasing.

3.1 Test Images

Testing of the proposed phase based image comparison measure is performed via a series of artificially degraded images, each image compressed by the JPEG still-image compression standard for lossy compression. Lossy image compression, as opposed to lossless image compression, sacrifices image quality for an increase in *compression ratio*. Compression ratio is a measure of the average number of bits required to store each pixel in the image.

Of interest to this work is that as part of the specification of JPEG, greyscale representations of text had to be considered. In fact JPEG is capable of compressing, and reconstructing, greyscale images of text. However, as Pennebaker and Mitchell [12] note, clipping of characters can occur.

The JPEG compression scheme was chosen because of its popularity, and hence wide spread use. A brief history of the JPEG standard, and the goals the JPEG committee set for the scheme, is presented in Appendix C.

The suite of test images consists of a "*de facto* standard" computer vision test image, a computer rendered image, a fractal image, three test pattern style images, and an eye chart. Each image was also compressed to 95%, 85%, 75%, 70%, 55%, 40%, 30%, 15%, and 5% respectively with the JPEG compression tool, as found in the Unix utility program *Xview*. The original and compressed images are displayed together in Appendix B.

A brief description, and "thumbnail", of each image will now be presented.

Standard

The *boat* image can be considered a *de facto* standard in computer vision as it is used throughout the literature as a test image. It was felt such an image was essential for a thorough test of the phase metric.

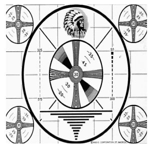
Computer Generated

The *snowface* and *fractal* images were chosen for two reasons. Firstly, both images are typical of those transmitted via the Internet and consequently represent "normal" images that are compressed and reconstructed. The second reason is that both these images exhibit different characteristics to the others in the suite.

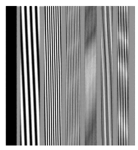
The *fractal* image (above) has a unique structure, that is repeated throughout the image, of very high and very low frequency sections, and the *snowface* image (below) is computer rendered, as opposed to a scanned image of an actual photograph.

Test Patterns

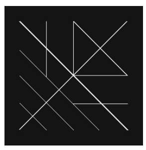
Test pattern images were considered a useful way of quantifying the results of the new phase based comparison measure since the degradation in a test pattern image can be measured more accurately. This is due to its design as a specific tool for testing the amount of distortion present in an image signal. The following test pattern images have been included.



The *indian* image. This image is included because it combines both simple and complex features, making it unique to the other images. It is of interest to observe the performance of the new measure on such an image.



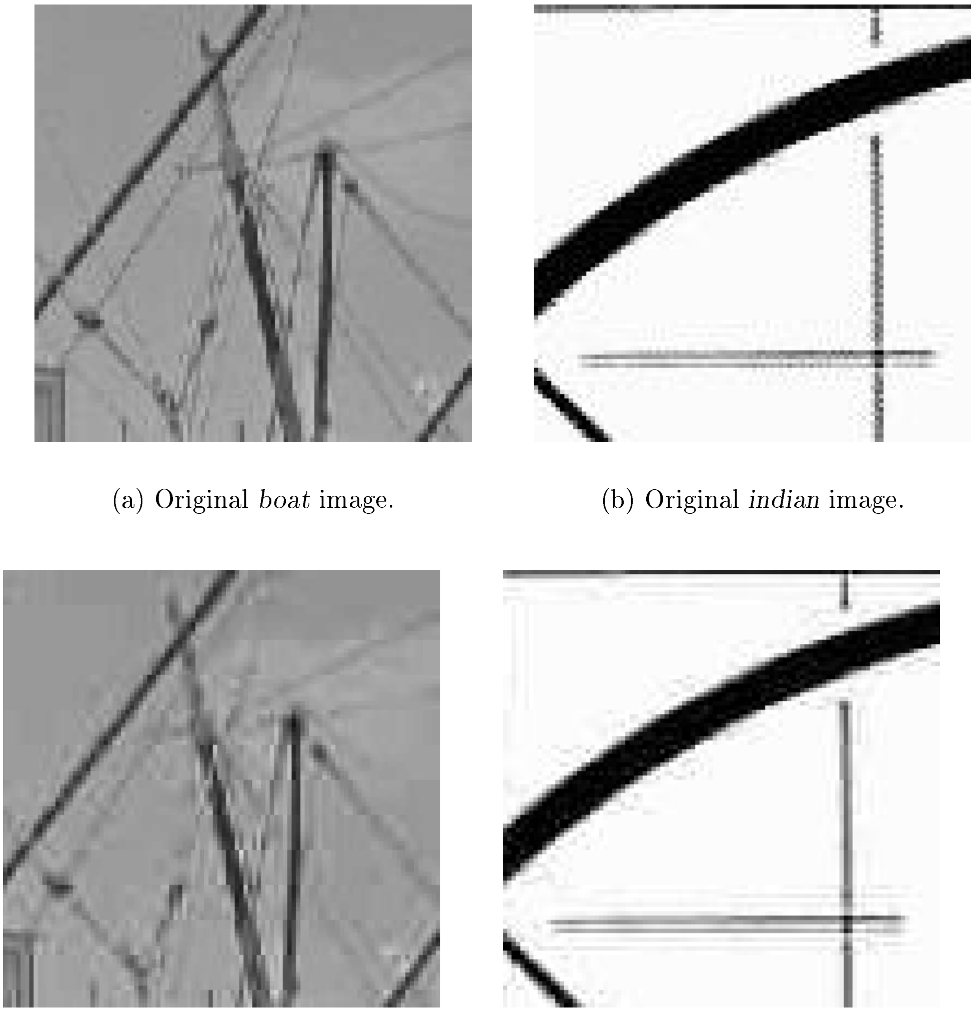
The *burst* image. The *burst* test pattern image is included for the reason that its compressed versions, including very high levels of compression (95%), are almost indistinguishable from the original. It is of interest to test the phase based comparison metric as to how well this relative compression difference is measured.



The *crosses* image. The *crosses* test pattern was chosen since different parts of the image appear distorted at different levels of compression. Although each of the other images exhibit increases in distortion as compression increases, the *crosses* image is less visibly changed when compressed. Given a selection of the compressed *crosses* images, attempting to order the images is virtually impossible.

Eye charts

The reason for selecting an *eye chart* image is based on the need to quantify perceptual distortion, and an eye chart is a logical choice. Perceptual distortion can be measured by simply observing the image and determining the difficulty in reading each line of the eye chart. This is by no means an absolute scale and is clearly subjective. However it was useful in terms of a relative metric, and the *eye chart* image also shows that the phase based metric can be successfully applied to text based images.



(a) Original *boat* image. (b) Original *indian* image.

(c) Compressed *boat* image. (d) Compressed *indian* image.

Figure 3.2: Two original and compressed images, highlighting the different effect of compression on varying images.

Text based images highlight the different effects compression has on an image, as illustrated in Figure 3.1. Two pairs of images are presented, each pair consisting of an original and compressed section of the *burst* and *indian* images. When applied to the *burst* image, the new measure produced the most favorable results. However the *indian* image exhibits the worst results. Note that the compression of the *burst* image has distorted the image, yet the major features are still the same. However, the distortion present in the *indian* image has created new 'features', or points of high contrast, most evident along the edges of the lines, and against the left edge of the image.

CHAPTER 4

Frequency Analysis

Frequency analysis of an image is based on calculating a *Trigonometric Fourier Series* [3] approximation to a discrete signal. This is accomplished via the summation of harmonically related sinusoidal waves. The Fourier series that approximates a square wave shape, necessary in the analysis of images, can be calculated using Equation 4.1. Note, Equation 4.1 holds for all $t \in \mathbb{R}$. The period of the wave shape is given by $\frac{2\pi}{\omega_0}$.

$$\mathrm{F}(t) \quad = \quad a_0 + \sum_{n=1}^{\infty} a_n \sin n\omega_0 t, \tag{4.1}$$

where,

$$\begin{aligned} w_0 &= \text{frequency of the sine wave, and} \\ a_n &= \begin{cases} 0 & \text{if } n \text{ is even,} \\ 1 & \text{if } n \text{ is odd.} \end{cases} \end{aligned}$$

Figure 4.1 shows the first two sine waves in the series, the sum of the two sine waves, and the sum of the first 50 terms in the Fourier series. Notice how the approximation to a square wave shape becomes increasingly accurate as more terms are added to the Fourier series.

A common approach to frequency analysis is the *Fourier transform* [13]. Although providing the frequencies present in a signal, it unfortunately does not tell us where each frequency occurs, and consequently it is impossible to use the result of the Fourier transform to compare two distinct points in two separate images.

Based on work by Kovesi [9], the extraction of phase and amplitude information is performed via a *wavelet transform*. A wavelet approach is employed as it allows the extraction of phase localised in both frequency and space.

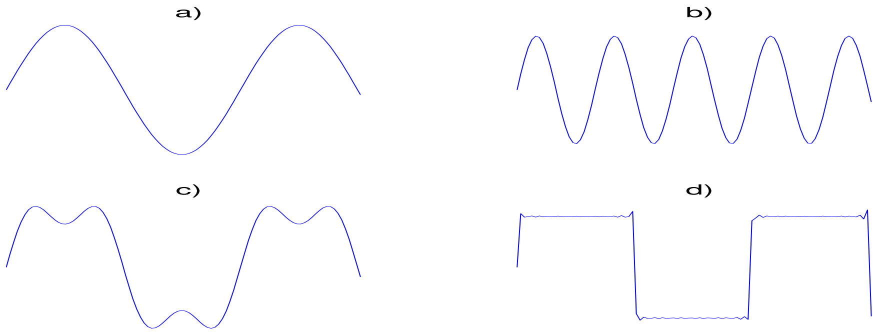


Figure 4.1: a) A standard sine wave. b) A sine wave at 3 times the frequency. c) The sum of the two sine waves. d) The sum of the first 50 sine waves in the Fourier series.

4.1 A Wavelet Approach to Frequency Analysis

Wavelet analysis involves creating a bank of filters in quadrature. Each pair of even and odd symmetric filters is created by geometricly rescaling the respective original wavelet. The bank of filters is then convolved with the image to generate frequency information. The approach taken here, based on work by Kovesi [8, 9], employs *Gabor* wavelets. Gabor wavelets are implemented as complex valued functions, whose real and imaginary components consist of a cosine and sine wave modulated by a Gaussian. Due to the scale of each filter varying geometrically, the filters are seen as having a Gaussian transfer function on a logarithmic scale. This gives rise to the term *Log Gabor* filters. Log Gabor filters are employed because they offer a good compromise between *spatial localisation* and *frequency localisation*.

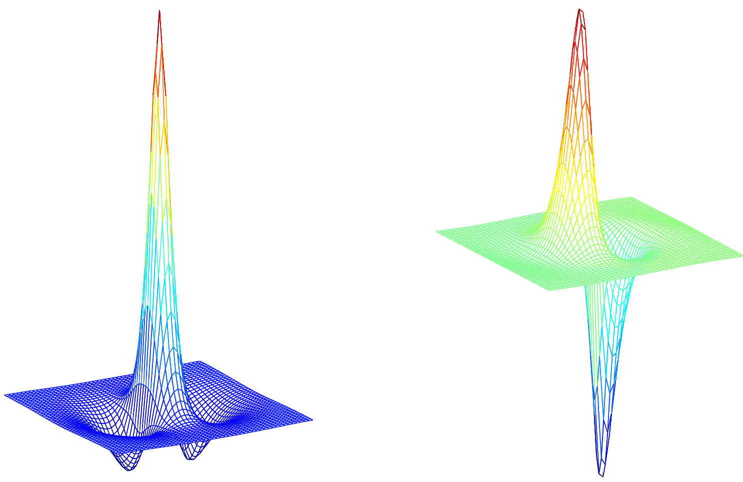


Figure 4.2: Even (right) and odd (left) symmetric Gabor filters.

The width of a filter is controlled by the Gaussian function modulating the sinusoid comprising the Gabor wavelet. The filter is said to be contained in a *Gaussian envelope*. The frequency a filter targets depends on the frequency of the sine/cosine wave and the bandwidth/localisation in space is controlled by the width of the Gaussian envelope. Using filters in quadrature enables the calculation of amplitude and phase data for a particular scale/frequency at a given spatial location. Figure 4.2 illustrates an even symmetric and an odd symmetric Gabor filter respectively.

Log Gabor filters are preferred over other wavelets since they allow arbitrarily large bandwidth filters to be constructed while still maintaining a zero DC component in the even-symmetric filter. A zero DC component cannot be maintained in Gabor functions over one *octave*. By defining the factor between the high and low frequency points, an octave describes the width of the Gaussian envelope used to modulate the filter. This is important as the optimal filter shape was found to have a width of two octaves.

Wavelet frequency analysis decomposes a signal into a series of basis functions representing each frequency recorded. At each point we have phase and amplitude pairs, for each scale of filter. Figure 4.3 illustrates the responses to a signal.

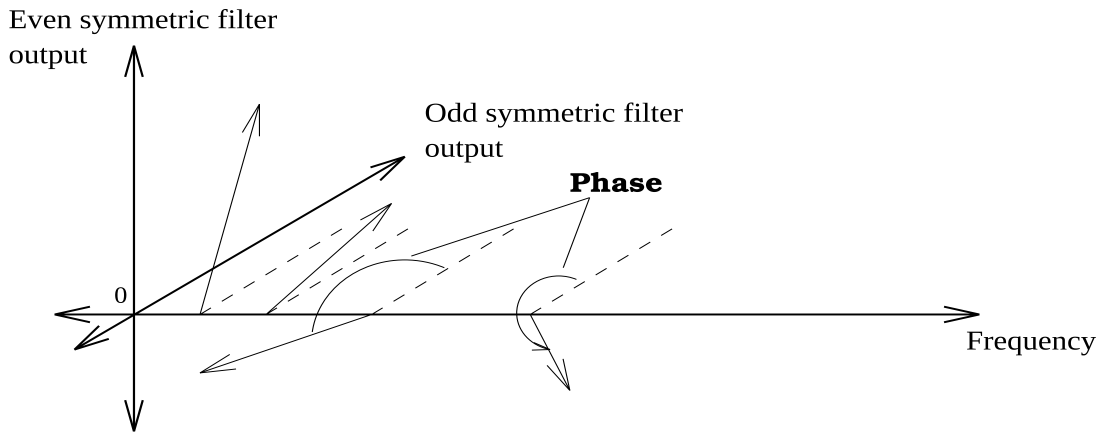


Figure 4.3: An array of filter response vectors, radiating out from the frequency axis, representing a point in the image. The amplitude determines the length of the vector, phase determines the angle.

4.2 Calculating Phase and Amplitude Information

Convolution of a signal with even and odd complex valued Gabor filters results in a complex valued number, at each point in the signal, of the form $a + ib$. We can think of a as the result of convolution between the image and even filter, and b as the result of convolution between the image and odd filter. From this result localised phase information Φ is calculated by

$$\Phi \quad = \quad \mathrm{atan2}(a, b), \tag{4.2}$$

and localised amplitude information A is calculated by

$$A = \sqrt{a^2 + b^2}. \tag{4.3}$$

As a guide, phase data is more accurate for lower frequencies. That is, if a signal contains mainly higher frequency changes (sharp 'edges') then the phase data are likely to be less reliable, and hence less useful. For example, the phase data extracted for the *crosses* image is less accurate in terms of measuring a displacement and distortion, since it contains only high frequency changes: black lines on a white background. Due to the way each wavelet filter is rescaled, a larger scale implies a lower target frequency.

4.3 Frequency Analysis of Images

Analysis of an image is accomplished by convolving each scale of filter with the image, at multiple filter orientations.

Illustrated in Figure 4.4 are the even symmetric filters used in the extraction of phase and amplitude information. Note that only one scale of filter is displayed. In general, six orientations of filter is sufficient to cover an image. Note that a filter orientation of zero radians is equivalent to a filter orientation of π. Consequently filters need only encompass a range of zero to $\pi - (\frac{\pi}{norient})$, where $norient$ is the number of filter orientations.

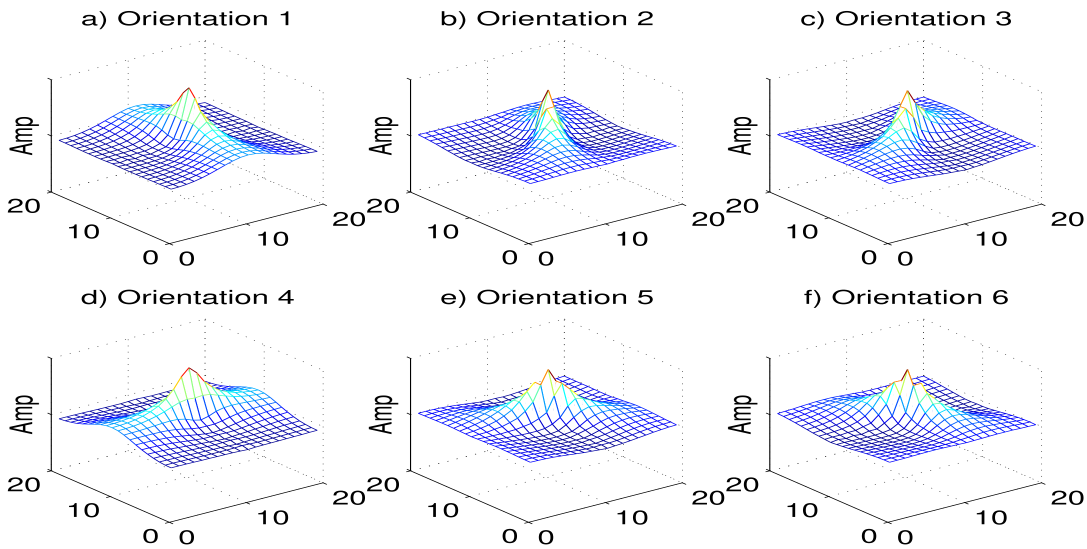


Figure 4.4: Images a) – f) correspond to the even symmetric filters used to extract phase data. The first filter is set to an orientation of π, the last to $\frac{11\pi}{6}$, with increment size $\frac{\pi}{6}$.

4.4 Optimizing the Filter Shape

The reliability of extracted frequency information is strongly influenced by the shape of filter used to analyse an image. An initial study of filter shapes considered only 1-dimensional signal analysis. This is useful because the filters used for extracting 1- and 2-dimensional frequency information are implemented in the same fashion, and are of similar shape.

The problems discussed for the 2-dimensional case, including low amplitude and edge effects, were initially discovered on trialing the 1-dimensional filter. In 1-dimensional signal analysis a filter with a large extent in the spatial domain, but of small wavelength (initial wavelength is 8 pixels) produces the most useful results. Due to the large spatial extent of the filters, some localisation in space is lost. However this increase in spatial width allows the filter to "see" more of the image, producing better results.

If incorrectly set, the width of the filter, determined by the width of the Gaussian envelope, results in an increase in the number of anomalies in the phase information. Consequently, the measured displacement, and in particular, the measured distortion are effected. The distortion measured is extremely sensitive to the shape of the filter; a "good" filter shape can accurately measure distortion in images with translations up to 10 pixels, whereas a "bad" filter shape results in a meaningless measure of distortion between translated images.

Initial results for 2-dimensional analysis were poor, with the wavelet parameters set as for 1-dimensional analysis. Given the similarity between 1- and 2-dimensional filter shape, this result appears incorrect.

Using the same filter shape as for 1-dimensional analysis, the 2-dimensional analysis results were poor. However, another consideration with 2-dimensional filters is the width, or spread, of the Gaussian function. In other words, how little or how much the filters overlap at each orientation. It is desirable to have as little overlap as possible, yet still cover the entire image. Figure 4.5 illustrates increasing the spread of the Gaussian function modulating the filter.

It was found that, as for 1-dimensional analysis, increasing the spatial extent of the filter improves results dramatically.

To measure distortion in general images, a wavelength of 80 pixels produces the most reliable results. The wavelength of the 2-dimensional filter, for a 256×256 image, is on the order of 10 times the size of the 1-dimensional filter.

Although a reasonable measure of displacement is obtained using an initial wavelength fixed at 80 pixels, a more accurate measure of displacement results from a filter with initial wavelength $\frac{1}{4}$ the width of the image. The relationship between filter shape and image width was determined after trialing the phase based metric on multiple images, over various displacements.

A filter of fixed wavelength is used since changing the filter shape also changes the

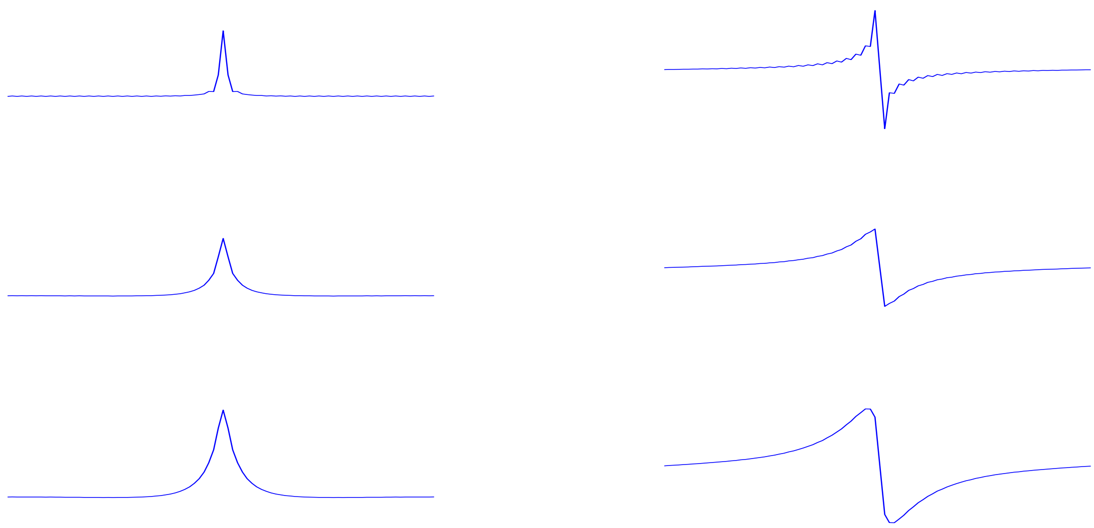


Figure 4.5: Even and odd symmetric filters, created with increasing spatial spreads. The middle pair of filters returns the best results.

scale of measured distortion. This comparison would be useful as a way of quantifying perceptual distortion. However, recall that the new measure is a relative one. Consequently, comparing the distortion measured between two different reference images and their respective modified images is extremely subjective. Moreover there is no method, with the possible exception of human perception measures, to verify results. However, of interest to this work is that the distortion recorded varies by as much as a factor of 10, depending on the type of image. This result is discussed further in Chapter 5.

A fixed initial filter width of 80 pixels does not allow for images smaller than 80×80 pixels. In this case, as a general rule, a wavelength of $\frac{1}{3}$ the image width will give acceptable results, irrespective of translation. This relationship between image size and wavelength was observed to give acceptable results for all image types and sizes, although little testing has been done on images smaller than 200×200.

4.4.1 Calculating a Minimum Reliable Wavelength

If no translation between signals is present then the wavelength of the initial filter is arbitrary: it has little bearing on the quality of phase information generated. However, if a translation has occured, and the wavelength of a filter is less than the translation, all information generated by using that filter is useless. This is intuitive as the filter applied is not wide enough to be able to determine how much of a displacement has occured, and thus cannot distinguish between distortion and displacement. As a general rule, wavelengths of approximately twice the actual displacement are considered reliable [8]. That is, a filter with a wavelength twice that of the translation is able to see enough of the image to differentiate between translation and degradation.

Calculation of a minimum reliable wavelength is necessary for 1-dimensional analysis since the initial 1-dimensional filter wavelength of 8 pixels is less than half the maximum translation. However, recall that the minimum wavelength of the initial 2-dimensional filter is fixed at 80 pixels. Consequently, no translation was greater than $\frac{1}{4}$ the wavelength, and hence all 2-dimensional filters are treated as reliable.

4.5 Calculating Phase Angle Difference

Recall that, at each point in an image, frequency data consists of both a phase angle component and an amplitude component. For purposes of comparison, a *difference* must be calculated between the phase angle data to overcome the wrap around problem. The wrap around problem is caused by the fact that the difference between 359° and 1° is 2°, not 358° as calculated by a simple subtraction. Hence, given a phase angle θ_1, and phase angle θ_2, a *phase difference* $\Delta\theta$ can be calculated based on the sine and cosine difference of angle laws, as

$$\begin{aligned}
\cos(\Delta\theta) &= \cos(\theta_1) \times \cos(\theta_2) + \sin(\theta_1) \times \sin(\theta_2), \\
\sin(\Delta\theta) &= \cos(\theta_1) \times \sin(\theta_2) + \cos(\theta_2) \times \sin(\theta_1), \\
\Delta\theta &= \text{atan2}(\sin(\Delta\theta), \cos(\Delta\theta)).
\end{aligned}$$

The resulting phase difference can now be used to determine the displacement or distortion as required.

4.6 Displacement

An estimate of the translation between two images is calculated using the phase difference, as described in the previous section. Given a point of phase difference $\Delta\theta$, an estimate of the translation between images at that point is given by the phase displacement $\Delta\Phi$, and is calculated by

$$\Delta\Phi = \frac{\Delta\theta}{(2\pi)} \times \lambda_s, \tag{4.4}$$

where λ_s is the wavelength of the filter at scale s.

A displacement is calculated for each point in the signal, for each scale of filter. If the two signals being compared are identical, bar translation, then the measured displacement should be consistent across all points. Consider Figure 4.6, which displays an original and translated signal. The phase difference between the translated and original signals is the same at every point. Unfortunately, in practice the calculated displacement is often irregular.

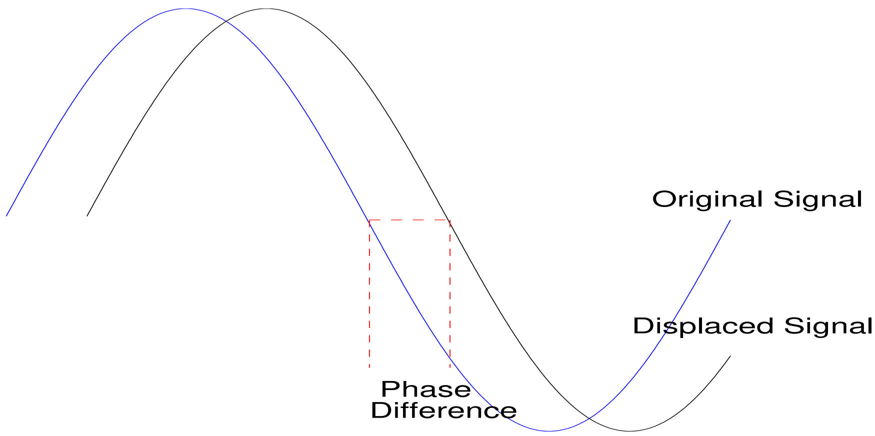


Figure 4.6: The displacement can be calculated between an original and displaced signal by measuring the phase difference at each point.

4.7 Distortion

Calculating the distortion of an image using phase information is based on the premise that the localised estimated displacement at all scales between two identical images is constant, regardless of translation. Moreover, if a standard deviation is calculated over the displacement estimates, the deviation from the mean only increases if there are actual differences between images. Based on this assumption, two methods are proposed for the estimation of distortion: a standard deviation and an absolute deviation [8]. An absolute deviation is implemented along with a standard deviation since taking the absolute value of difference rather than the square root of a squared difference can be more sensitive to minor changes. This increased sensitivity is desirable when measuring slight distortions.

The standard deviation σ, and the absolute deviation α, are calculated as follows:

$$\sigma = \sqrt{\sum_{i}^{N} \frac{(\Delta\Phi_i - \overline{\Delta\Phi})^2}{N}}, \tag{4.5}$$

$$\alpha = \sum_{i}^{N} \frac{|(\Delta\Phi_i - \overline{\Delta\Phi})|}{N}, \tag{4.6}$$

where N is the number of pixels in the image, $\Delta\Phi_i$ is the phase difference calculated at pixel i, and $\overline{\Delta\Phi}$ is the mean phase difference value.

4.7.1 The Low Amplitude Problem

The low amplitude problem corresponds to minimising the effect of phase measurements associated with a zero or near zero amplitude response. This is desirable as a low amplitude component indicates a low response by the filter, which generally implies the phase data is unreliable. Figure 4.7 illustrates the low amplitude problem by displaying two plots of displacement, at varying scales, and the corresponding amplitudes. The displacement anomalies clearly correspond to points of amplitude near or at zero. Such anomalies effect all calculations involving these particular scales. If the phase data are not weighted, frequency information with a weak filter response is given equal weight to frequency information with a strong filter response.

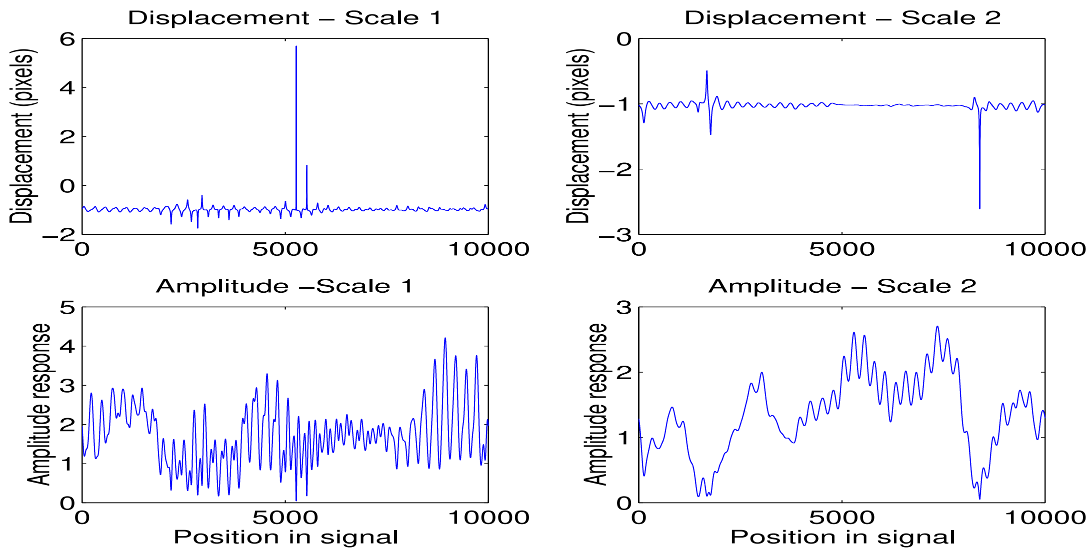


Figure 4.7: Displacement and corresponding minimum amplitude at two varying scales. Note that for points of amplitude near or at zero the corresponding displacement is uncharacteristicly high or low. The second signal differs from the first in that it is translated one position to the right.

Three approaches are taken to overcoming the low amplitude problem. The first involves weighting each point of phase by its respective amplitude. Weighting by amplitude consists of modulating each element of phase data by its respective amplitude component, then taking the summation of the result. This weighting biases the phase data towards scales recording a large amplitude component at the given point. Given a point n in an image, the phase response at n $\phi^s(n)$, is weighted by its amplitude component $A^s_{min}(n)$, for a filter of scale s, as

$$\Phi^s_w(n) = \frac{\sum_{s \in S}(A^s_{min}(n) * \Phi^s(n))}{\sum_{s \in S} A^s_{min}(n)}, \tag{4.7}$$

where S is the number of scales of filter.

Note that a minimum amplitude is used to weight the phase difference. A minimum amplitude, at scale s, is calculated by

$$A^s_{min} = \min(A_1, A_2),$$

where A_1 is the amplitude component from the first image, and A_2 is the amplitude component from the second image.

Recall that the phase data used for comparison is calculated by taking the difference between the phase responses for each image. By using the minimum amplitude the calculated phase difference is weighted towards those points where both images have a high amplitude component.

In some circumstances all scales of filter record a small response at the same point. In this case, weighting by amplitude still fails to relieve the problem of low amplitude, as there is now a point of anomaly across all scales. This problem is overcome by again weighting the phase data by amplitude. However, the amplitude component is now compared to a threshold value, and when calculating displacement or distortion, only phase data with a minimum amplitude component of greater than the threshold is used.

The thresholding method employed takes a *Zscore* [1] of the amplitude responses and ignores all amplitude and phase information with a Zscore below that of the threshold. A Zscore, $Z(x)$, at a point x is simply the number of standard deviations x is away from the mean, and is given by

$$Z(x) = \frac{I(x) - \bar{I}}{\sigma},$$

where I is the signal being analysed, $\bar{I}$ is the signal mean, and σ is the standard deviation of the signal.

A threshold is selected by taking a percentage of the maximum distance any point is from the mean. In practice the threshold is set to 10%. In this manner less useful information is ignored, without the need to set a hard amplitude limit. A hard amplitude limit is undesirable as a low amplitude response for one filter may be equivalent to a high amplitude response for another; the amplitude response is only relative to the scale it was drawn from.

By using a Zscore, only amplitude information with a large negative distance from the mean is ignored. If all amplitude data are close to the mean, then little or no data is ignored.

The two outlined approaches to weighting phase information both involve modulating phase by amplitude. Given the importance of phase information to an image the third method of overcoming low amplitude involves thresholding as just described, but now only raw phase data is considered. The phase data is no longer modulated by its amplitude component, so the amplitude is only used to indicate reliable phase points.

4.8 RMS and SNR Errors

Due to the similar nature and performance of the RMS and SNR errors, only the RMS error is compared to the proposed phase based measure. This is because the RMS is more widely used in the literature. However, both the RMS and SNR errors are defined.

Given two greyscale images f and g, the RMS error [15] is defined as

$$RMS(f,g) = \left[\frac{1}{n(X)} \sum_{x \in X} (f(x) - g(x))^2\right]^{\frac{1}{2}}, \tag{4.8}$$

where $n(X)$ is the number of pixels in the image, and the SNR error [15] is defined as

$$SNR(f,g) = \left[\frac{\sum_{x \in X} g(x)^2}{\sum_{x \in X} (f(x) - g(x))^2}\right]^{\frac{1}{2}} \tag{4.9}$$

CHAPTER 5

Results and Discussion

The aim of this section is to show that the new phase based comparison measure is capable of separating image translation from image distortion. This is a major advantage over both RMS and SNR measures. The results presented below indicate that a phase based measure provides a guide to the "quality" of an image invariant to translation.

Shown also is that a phase based measure can accurately record the amount by which an image has been translated, with respect to the original.

A comparison between the new phase measure and the RMS is presented to once again highlight the improved performance of the new measure over that of the RMS. Note that results for only the *snowface*, *indian*, *burst*, and *eye chart* images are included. Results for the *boat* and *fractal* images were very similar to those for the *snowface* image. Similarly, results for the *crosses* and *eye chart* images are almost identical.

It should be noted that the most favorable results are obtained with a thresholding approach to the weighting of phase information by amplitude. Although recording similar results, the thresholding approach whereby phase data are not modulated by amplitude proved most reliable. Consequently, the results presented in the next section are all calculated from unweighted, thresholded phase data, as discussed in Section 4.7.

5.1 Measuring Distortion

This section presents results for both the new phase based measure and the RMS error, when applied to the test images noted above. Both measures are presented with pairs of images, each pair consisting of an original reference image and the reference image compressed and reconstructed. This is done for increasing levels of distortion, and zero, one, three, and six pixel translations between images. These translations are designed to highlight the fact that for small translations (three or less pixels) the phase based measure performs well, but that for larger translations (six or more) the measured distortion becomes much larger.

The distortion recorded for the *snowface* image, presented in Figure 5.1, is typical of the results on general images. Note that even for a translation of six pixels the effect of the translation is minimal.

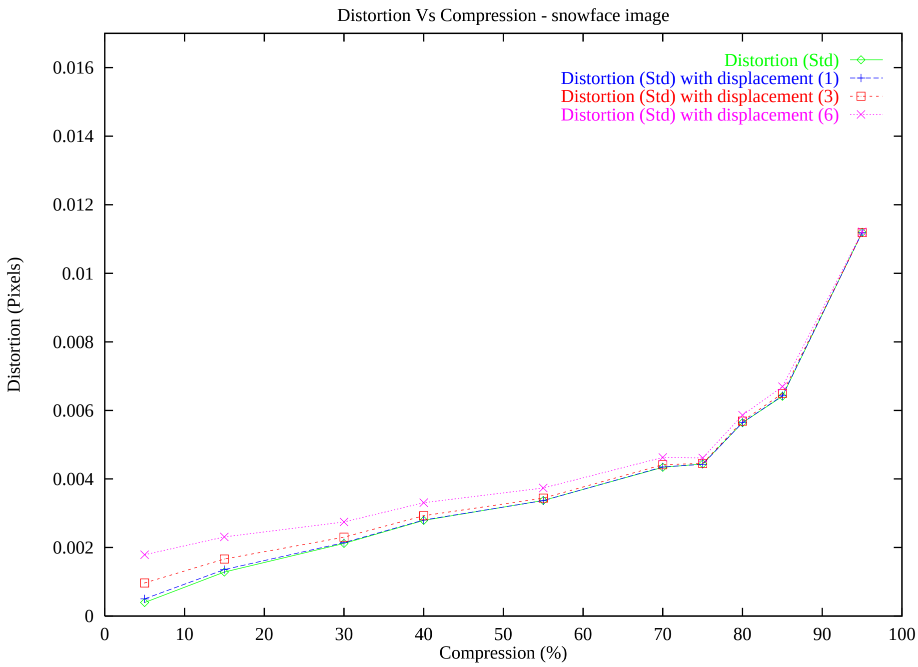


Figure 5.1: Measured distortion for the *snowface* image, with a displacement of 0, 1, 3, and 6 pixels respectively.

Figure 5.2 illustrates the results for the RMS error when applied to the *snowface* image. The effect of translation on the RMS is clearly evident. In particular, note that the RMS error decreases as compression increases for a translation of one pixel. Note also the extreme increase in measured distortion for any translation between images. Figure 5.2 is typical of the results produced by the RMS error, and as such no further RMS results are displayed in this section. However, results for the RMS error applied to the *indian*, *crosses*, and *eye chart* images can be seen in Appendix D.

The phase based measure recorded the poorest performance when applied to the *indian* image, illustrated in Figure 5.3. The reason for this is the varying effect of compression on different image types, as discussed in Section 3.1. In particular, the *indian* image when compressed and reconstructed displays new "features". Consequently, the measured distortion shows a large increase as the translation increases. However the RMS error records worse results than the phase based measure (see Appendix D).

Results for the *eye chart* image are included to show that the new phase based measure can be reliably applied to text only images. Figure 5.4 displays the recorded distortion for the new phase measure. As for the other images, the

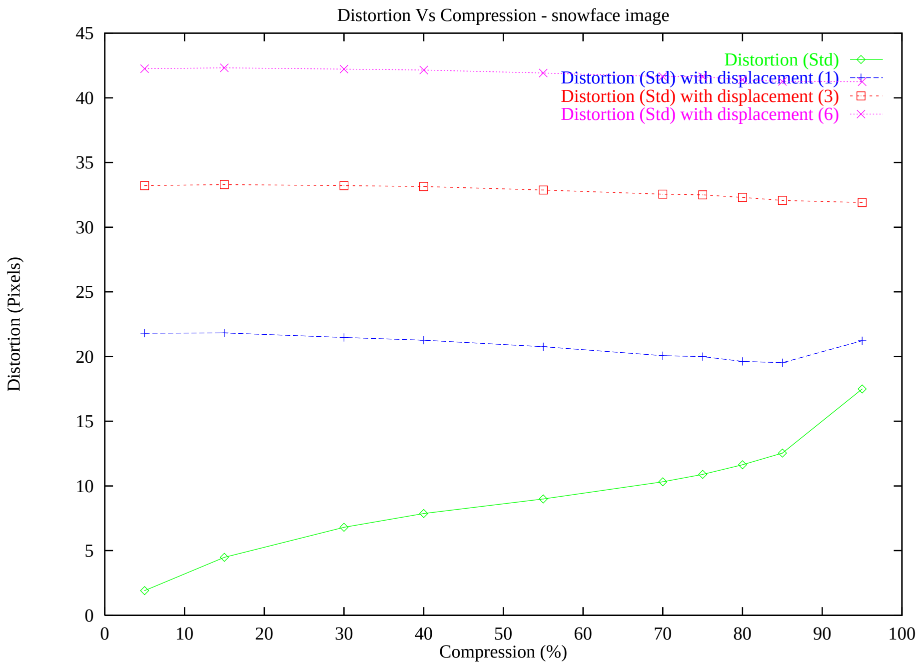


Figure 5.2: The distortion returned by the RMS error when applied to the *snowface* image for displacements of 0, 1, 3, and 6 pixels respectively.

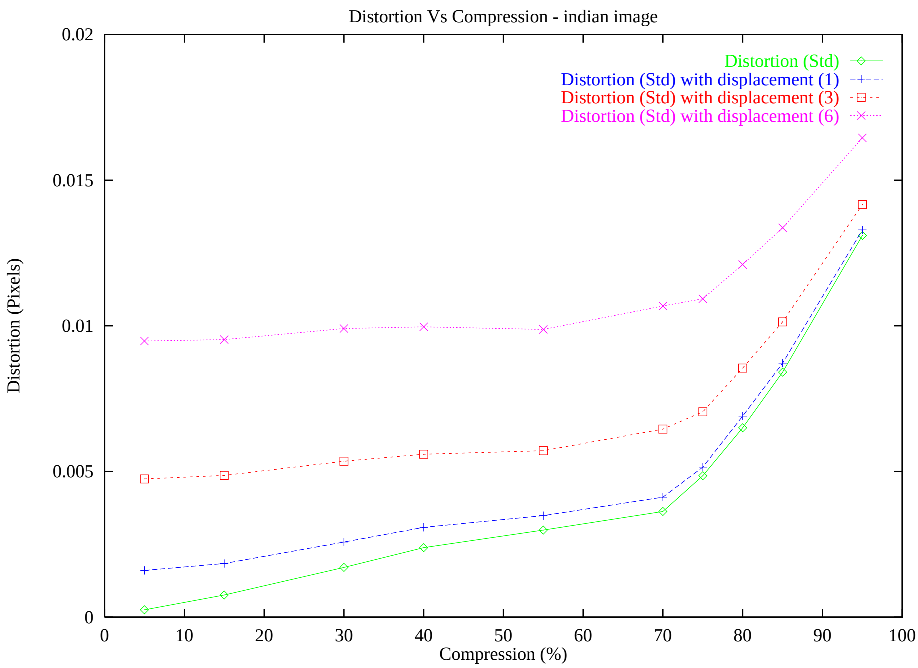


Figure 5.3: Measured distortion for the *indian* image, with a displacement of 0, 1, 3, and 6 pixels respectively.

results are most favorable.

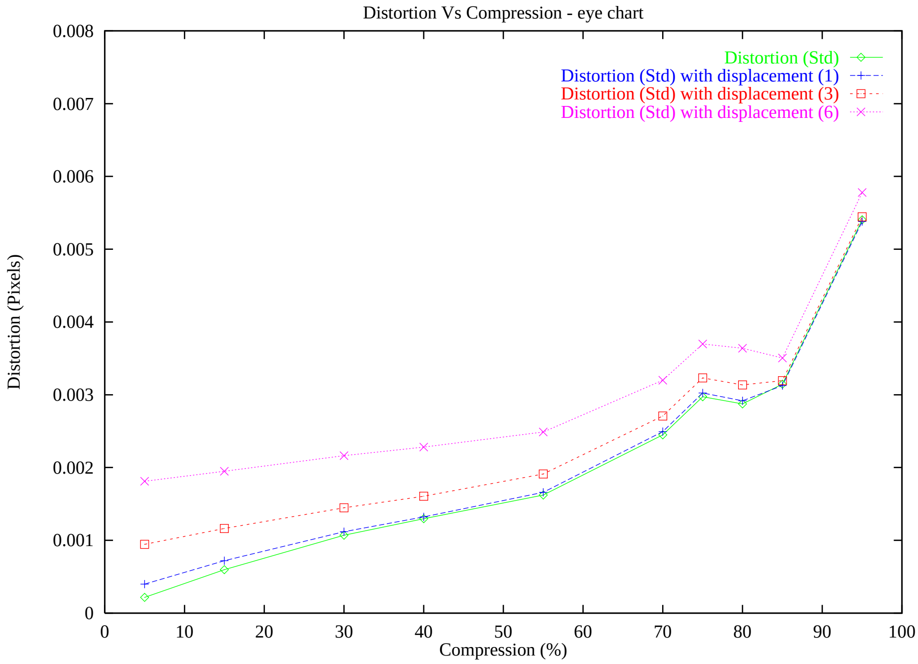


Figure 5.4: Measured distortion for the *eye chart* image, with a displacement of 0, 1, 3, and 6 pixels respectively.

Figure 5.5 displays the results recorded by the phase based measure when applied to the *burst* image. This result is clearly the most favorable, the distortion is uneffected by translation. When applied to the *burst* image the RMS performs poorly, particularly between translated images (see Appendix D).

Note that for all images the recorded distortion converges with increasing compression. The reason for this convergence is unknown. However, a possible explanation is the increase in distortion, caused by increasing compression, has a "blurring" effect on the image such that the distortion caused by translation is only a small factor of the actual distortion.

If we consider the convergence from the point of two identical images, any translation will cause a small amount of distortion due to the imperfect extraction of frequency data. However, the distortion increase with translation is not constant, and thus the effect of translation is reduced in some fashion with increasing distortion.

Table 5.1 lists the distortion recorded for each image at 95% compression, for the standard deviation distortion measure. No translation is present between images. The table is designed to highlight the different levels of distortion measured for varying image type. Note that there is no apparent correlation between distortion measured and perceived difference, or image type. This highlights an important point: the new measure is intended as a relative measure, and must not be used to compare entirely different images. Moreover, it highlights the need for the

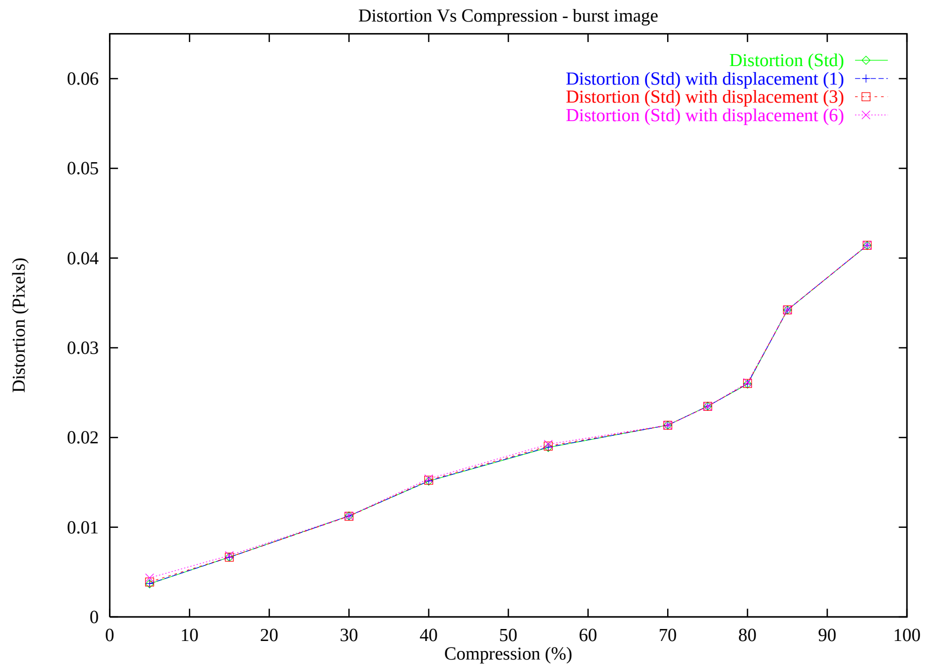


Figure 5.5: Measured distortion for the *burst* image, with a displacement of 0, 1, 3, and 6 pixels respectively.

separation of the type of degradation present in an image. Although the new measure provides a ranking of images in terms of distortion, it still provides no information as to the type of distortion.

5.2 Measuring Translations

Beyond translations of 15 pixels, both the 1- and 2-dimensional analysis techniques are only useful as a relative metric: large translations cause measured distortions to increase considerably. However, as illustrated in Figure 5.6, the estimate of the translation is accurate to two pixels. Note that three estimates of translation are given. These correspond to the various phase weighting approaches. The most favorable results are obtained by the two thresholding approaches.

5.3 Constant Greylevel Modulation

Modifying the greyscale values of an image by a constant amount had no impact on the measures of distortion and displacement. Figure 5.7 illustrates the effect of a constant greyscale modulation to an image. This result is expected since a constant change in no way effects the ratio of pixel intensities across the image. This experiment raises the question of whether the two images are identical or

Image	Distortion (Standard Deviation) 95%
boat	0.00312
burst	0.05276
fractal	0.00465
snowface	0.01399
indian	0.01329
crosses	0.02173
eye chart	0.00616

Table 5.1: Distortion measured between a reference image, and the reference image compressed to 95% and reconstructed. The reconstructed image is also translated by three pixels.

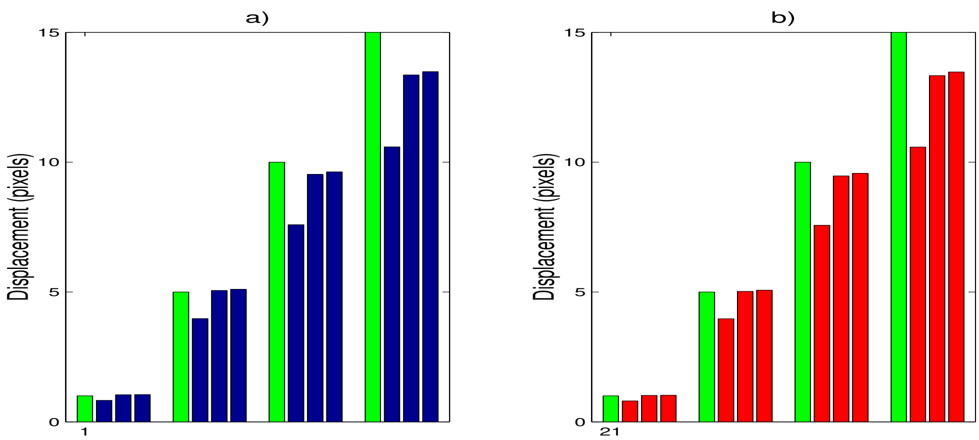


Figure 5.6: Measured displacement for the *snowface* image. Actual displacement is indicated by the green bar. Minimum wavelength of filter is 64 pixels. The original image is compared with itself in a), graph b) compares the original image and 85% compressed version.

not. It is clear that the two images are perceptually different. However, it is perhaps more meaningful to know that, although perceived to be different by the human eye, the images are in fact identical in terms of greyscale intensity changes across the image.

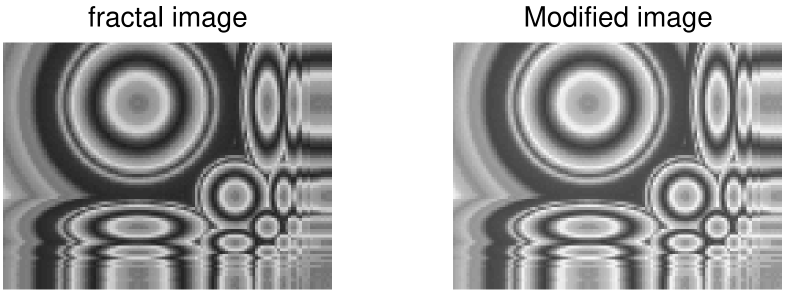


Figure 5.7: The original and modified *fractal* images. The modified image is the result of applying a constant amplitude modulation.

5.4 Noisy Images

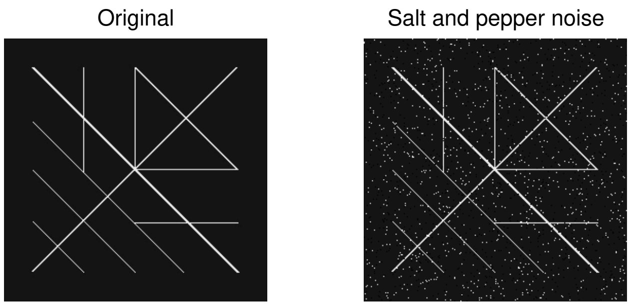


Figure 5.8: Result of adding 'salt and pepper' noise to the *crosses* image.

Given the problem the new phase based measure has when measuring distortion for the translated *indian* image, one way to test the theory that the new "features" present in the image are the cause of some, if not all, of the problem is to add noise to the image. Adding noise effectively adds new features, or points of contrast, to an image. The results obtained for the phase based measure support the theory that adding new features to an image cause large increases to the distortion measured between translated images.

Figure 5.9 highlights the increase in distortion between an original and translated, "noisy" image. Figure 5.8 illustrates the effect of adding "salt and pepper" noise to an image.

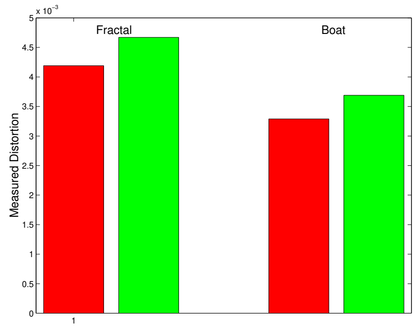


Figure 5.9: Measured distortion after adding "salt and pepper" noise to the *boat* and *crosses* images. The red bars indicate distortion measured between the original and "noisy" images. The green bar indicates the distortion measured between the original and translated "noisy" image.

CHAPTER 6

Conclusions and Further Work

The following section outlines the key areas of work required for the new phase based comparison metric to progress. A summary of the work is then presented.

6.1 Further Work

The key area of research is the further analysis of the wavelet transform used to extract frequency information from an image. A more extensive examination of the phase information extracted, for multiple scales and orientations of filter, is required to determine the most reliable set of wavelet control parameters. Moreover, the initial wavelength of filter is currently fixed. An investigation into the relationship between image size and filter wavelength is necessary to allow the analysis of variable sized images.

Although successful at measuring distortion invariant to small translations, continued work relating to the separation of image difference factors is essential. In particular, methods for distinguishing between

- translation,
- distortion,
- level of distortion, and
- type of distortion,

need to be developed. The last factor is important because it would not only allow the designer of an image manipulation routine to measure the distortion present in the modified image, but it would also give a measure of what has happened to the image to cause that distortion. Consequently, the designer has more feedback as to how the routine in use has modified the image.

Before any new metric can be used, it must be shown to be correct. Unfortunately, testing an image comparison metric can only be performed via human inspection.

Consequently, a psychophysical experiment to test the new measure, against a human viewer, needs to be developed.

As highlighted by the results, the behavior of the phase based measure is not fully understood. In particular the tendency to converge to the same level of maximum distortion, regardless of translation, needs to be examined. New methods for improving the accuracy of the measured displacement are also essential.

Finally, a more thorough examination of the new comparison measure against existing image comparison techniques is required.

6.2 Conclusion

Comparing images is a highly subjective task, and one that is especially hard to show correct. A new image comparison metric based on using frequency information has been presented and shown to alleviate some of the problems inherent with the RMS and SNR errors.

In particular, the new measure is able to separate translation from distortion. However, much work is still required in understanding the behavior of the wavelet analysis technique. Moreover, the filter shape is not optimal, and more work is required in understanding the effect of the filter shape on the reliability of frequency information extracted.

A brief background was presented on the various categories of image comparison techniques, along with each techniques strengths and weaknesses. However, a theme present throughout the discussion is that image quality is extremely difficult to measure. Moreover, to date there is no satisfactory measure, or definition, of image quality.

This thesis has highlighted the issues involved in measuring image quality. The new phase based measure presented has solved the problem of separating translation and distortion, but poses many more. However, the new phase based measure, although in the early stages of development, shows great potential and further work is strongly encouraged.

References

1. Lincoln L. Chao. *Statistics: Methods and Analyses.* McGraw-Hill, Sydney, 1969.

2. Scott Daly. The visible differences predictor: An algorithm for the assessment of image fidelity. In Andrew B. Watson, editor, *Digital Images and Human Vision*, pages 179–206. MIT Press, Cambridge, 1993.

3. Roger E. Ziemer, William H. Tranter & D. Ronald Fannin. *Signals and Systems: Continuous and Discrete*, chapter 3. Maxwell Macmillan International, 3rd edition, 1993.

4. Paul J. Hearty. Achieving and confirming optimum image quality. In Andrew B. Watson, editor, *Digital Images and Human Vision*, pages 149–162. MIT Press, Cambridge, 1993.

5. David E. B. Lees & Philip D. Henshaw. Printed circuit board inspection — a novel approach. In *Proceedings of the International Society for Optical Engineering*, volume 730, pages 164–173. SPARTA, 21 Worthern Road, Lexington, MA, 02173, 1987.

6. Peter Kovesi. Calculating frequency information local to a point in an image. Lecture notes from the unit Computer Vision, Department of Computer Science, The University of Western Australia, 1994.

7. Peter Kovesi. The Fourier transform. Lecture notes from the unit Computer Vision, Department of Computer Science, The University of Western Australia, 1994.

8. Peter Kovesi. *Invariant Measures of Image Features From Phase Information.* PhD thesis, The University of Western Australia, 1996.

9. Peter Kovesi. Symmetry and asymmetry from local phase. In *Tenth Australian Joint Conference on Artificial Intelligence.* AI'97, December 1997.

10. Alan V. Oppenheim & Jae S. Lim. The importance of phase in signals. In *Proceedings of the IEEE*, volume 69, pages 529–541, May 1981.

11. Jeffrey Lubin. The use of psychophysical data and models in the analysis of display system performance. In Andrew B. Watson, editor, *Digital Images and Human Vision*, pages 163–178. MIT Press, Cambridge, 1993.

12. William B. Pennebaker & Joan L. Mitchell. *JPEG Still Image Data Compression Standard.* Van Nostrand Reinhold, NewYork, 1993.

13. N. Morrison. *Introduction to Fourier Analysis.* John Wiley & Sons, Brisbane, 1994.

14. Albert J. Ahumada Jr. & Cynthia H. Null. Image quality: A multidimensional problem. In Andrew B. Watson, editor, *Digital Images and Human Vision*, pages 141–148. MIT Press, Cambridge, 1993.

15. Dale L. Wilson, Adrian J. Baddeley & Robyn A. Owens. A new metric for grey-scale image comparison. *International Journal of Computer Vision*, 24(1):5–17, 1997.

16. D. C. Munson Jr. & J. L. C. Sanz. Phase-only image reconstruction from offset Fourier data. In *Proceedings of the Society for Optical Engineering*, volume 556, pages 227–235, 1985.

17. Andrew B. Watson, editor. *Digital Images and Human Vision*, chapter 3, pages 139–140. MIT Press, Cambridge, 1993.

APPENDIX A

Original Research Proposal

Title: Phased Based Image Comparison Metrics

Author: Gian Paolo Lorenzetto

Supervisor: Dr Peter Kovesi

Keywords: Frequency Domain, Spatial Domain, Local Phase, Perceptual Distortion, Image Error Measures, Image Compression

Background

When presented with two images, a human viewer is able to make a judgment as to which looks better. Many methods have been proposed that either attempt to simulate this human visual system rating, or generate some mathematically based comparison. All of these methods have been successful to various degrees. It is essential that a metric is available that can be relied on to give an accurate comparison between two images for several reasons. For instance, take the design of lossy image compression algorithms. A lossy image compression algorithm is one that attempts to maximise compression while sacrificing image quality. Given a compression ratio of 1000:1, how does the designer of the algorithm know how well the algorithm has maintained image quality? If the result of the comparison is unreliable, changing control parameters for the image compression algorithm is meaningless.

As for image compression algorithms, image enhancement algorithms require reliable feedback on the quality of a reconstructed image compared to the original.

The measure currently used most often is the root mean square error. A criticism of this and other similar comparison methods is that they are inaccurate measures of perceptual distortion. For example, shifting each pixel in an image one place to the left makes little perceptual difference, but to the root mean square error this shift represents a very large difference. The root mean square error also cannot be meaningfully applied to the reconstruction of binary images nor images that contain both text and images.

This lack of error measures that focus on both differences in pixel intensities and perceptual distortion is a serious problem in the evaluation of image compression and enhancement algorithms. Work done in the Department of Computer Science, in collaboration with the Department of Psychology, has shown that perceptually significant features in images are found at points of local phase congruence. Oppenheim and Lim [10], in 1981, first demonstrated the importance of phase information in an image, with a now famous (in Computer Vision) experiment. Thus errors in the phase information of a reconstructed image may not only indicate pixel differences, but more importantly highlight perceptual distortion. If this distortion could be measured, it may prove to be a more accurate indication of the loss of image quality in a compressed then reconstructed image.

Aim

Oppenheim and Lim [10], using the Fourier Transform, combined the phase information from one image and the magnitude information from another, to produce a synthetic image. This synthetic image was then back-transformed to produce a new image, the features of which, although distorted, clearly corresponding to those features found in the image from which the phase information was obtained. This result is easily re-produced, and work by Kovesi [8], Lees and Henshaw [5], Munsen and Sanz [16], not to mention many others has followed in the study of the importance of phase in images.

It is my proposal to develope a new image error measure defined in the frequency(Phase) domain rather than the spatial(Magnitude) domain. In particular it is intended to use the phase information local to each point in an image, to extract the necessary data to detect pixel differences, perceptual distortion, and translation.

Phase information can be used in both a general and specific manner. By analyzing phase characteristics of the whole image general frequency trends can be seen. For instance if an image is translated left one pixel the displacement of the phase information would indicate a shift of the original picture as opposed to a different or distorted image [5].

The second stage of analysis is to generate phase information local to each pixel for multiple orientations and scales. This would provide large amounts of data for small sections in the image, and it is hoped that this information will lead to the determination of whether perceptually important features have been distorted.

Once this phase information is extracted there is no reason why it cannot be used for other means. Oppenheim and Lim [10] discuss many other applications in which the phase information of an image is of great importance. Areas such as image alignment [5], blind deconvolution, image coding using Fourier techniques and the restoration of images degraded by additive noise are some examples. All of these are linked by the importance of phase, and as such the more research done on phase information the further these areas will move.

Method

The first task to be completed is a review of literature on image compression algorithms. This is necessary to understand the different effects that image compression algorithms have on image quality. The next step will be to examine techniques for the extraction of phase information from images; by looking at lecture notes on the Fourier transform of an image [7] and lecture notes on calculating the frequency information local to a point in an image [6]. A review of papers on related subjects will also be of great help in noting possible problems and their solutions before encountering them myself. This will give me more familiarity with a frequency based representation of an image, helping in the interpretation of graphs involving phase information.

Extracting phase information from the image will be approached in two ways. The first is a global representation of the phase information by taking the Fourier transform of the image. Preferably the Fast Fourier transform will be used over the standard Fourier transform, as it is much cheaper computationally. However to make best use of the Fast Fourier transform the image may need to be padded with zeros so that its matrix dimensions are a power of two; if the dimensions are powers of two then the Fourier transform can be calculated as quickly as possible. The second approach is to extract local phase information in an image (on a pixel by pixel basis) using convolution of the signal with Gabor wavelets. This will be done in the frequency domain so convolution between the signal (the image) and the Gabor function can be done by multiplication according to the Theorem of Convolution [13].

Once all the phase data have been extracted from an image, it is my intention to devise several phase difference metrics, in both the global and local frequency representations of an image. These metrics can then be used to compare the phase information of the original image and the phase information of the compressed then reconstructed image. I will then evaluate the performance of my error measures on a compression algorithm. Most likely the JPEG format will be used in an image processing utility such as `xview`. Using the JPEG format to compress images to varying degrees I hope to evaluate the quality of the image in respect to the original. It will then be possible to make a comparison between my measures and the root mean square error measure, the Wilson Baddeley Owens [15] error measure, and possibly other error measures. This comparison will be done on a compression versus image quality basis.

I hope to have completed all coding requirements by the end of first semester to allow ample time to test and modify the program. Possible modifications are changing the wavelet function and changing the range of basis frequencies over which the program calculates phase data.

Software and Hardware Requirements

The machines in Lab2.1, Lab2.7, and the Pentiums in LabG11 of the Department of Computer Science adequately cover all hardware requirements. The Departments licensed copy of Matlab Version 5.0 meets all software requirements.

APPENDIX B

Test Images

The following are the test images used in this work.

For a description of why these images in particular have been chosen as test images see Section B.

B.1 Boat Image

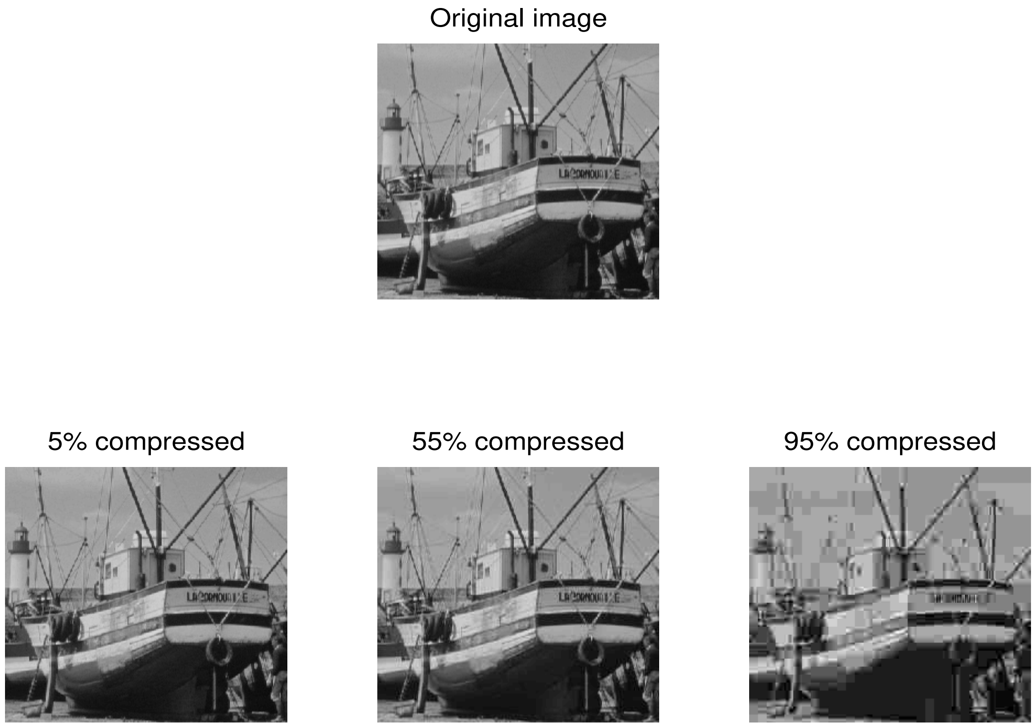


Figure B.1: The *boat* image is displayed, along with the same image reconstructed from various levels of compression.

B.2 Indian Test Pattern

Table B.1 lists the major features of the *indian* image, and the reasons for these features. These guidelines can also be applied to the other images when assessing levels of distortion.

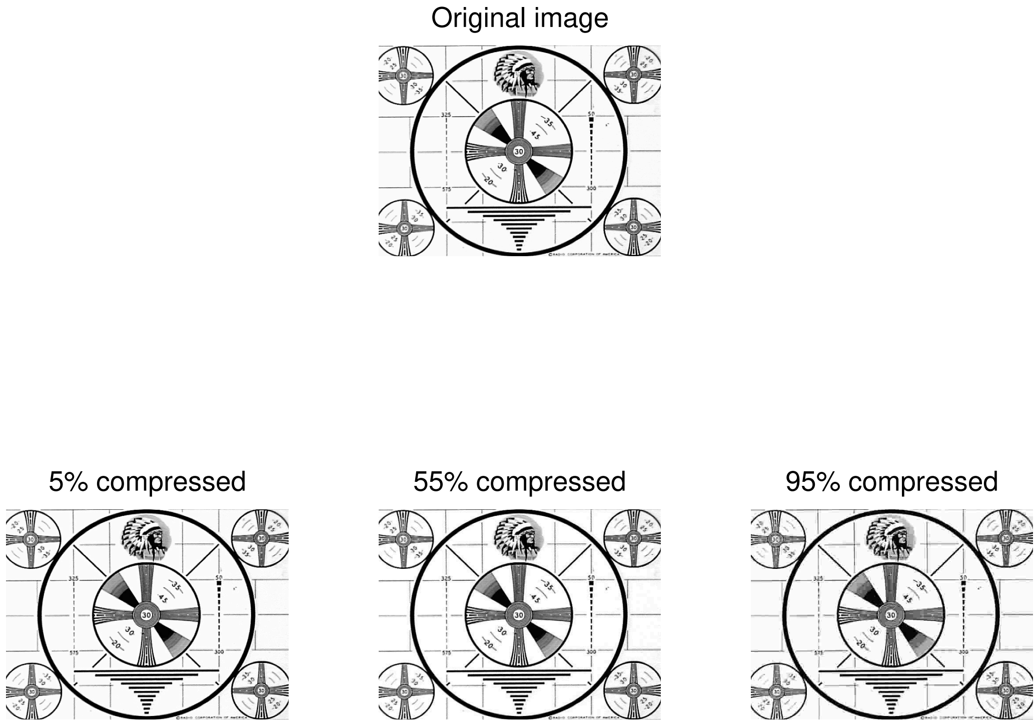


Figure B.2: The *indian* image is displayed, along with the same image reconstructed from various levels of compression.

Design	Reason
Crosshatch Pattern	Check overall linearity
Corner Circles and Focus Text	Check on corner linearity and focus.
Black and White Bars on a Gray Background	Check for streaking and ghosting in the image.
Resolution Line Pattern	Check on sharpness and resolution.
Points, Circles, and Dither Patterns	Check for defocusing, blooming, and halos.
Gray and White Text	Check on the sharpness, contrast and visibility of text.

Table B.1: The major features of the *indian* image, and how to interpret those features, when determining the amount of degradation present.

B.3 Burst Test Pattern

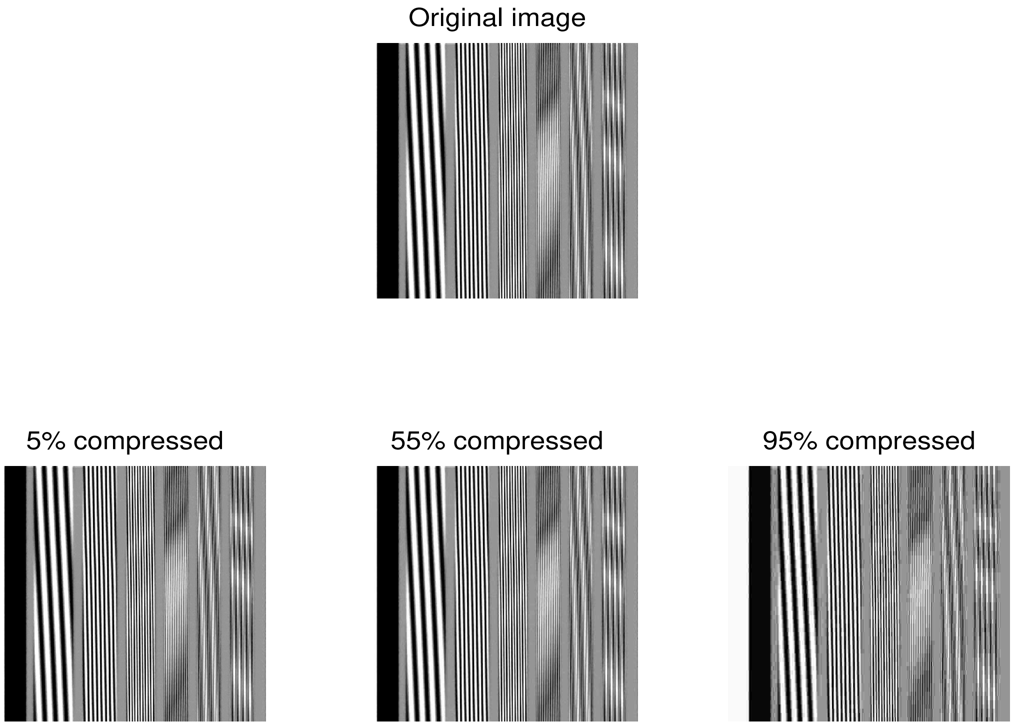


Figure B.3: The *burst* image is displayed, along with the same image reconstructed from various levels of compression.

B.4 Crosses Test Pattern

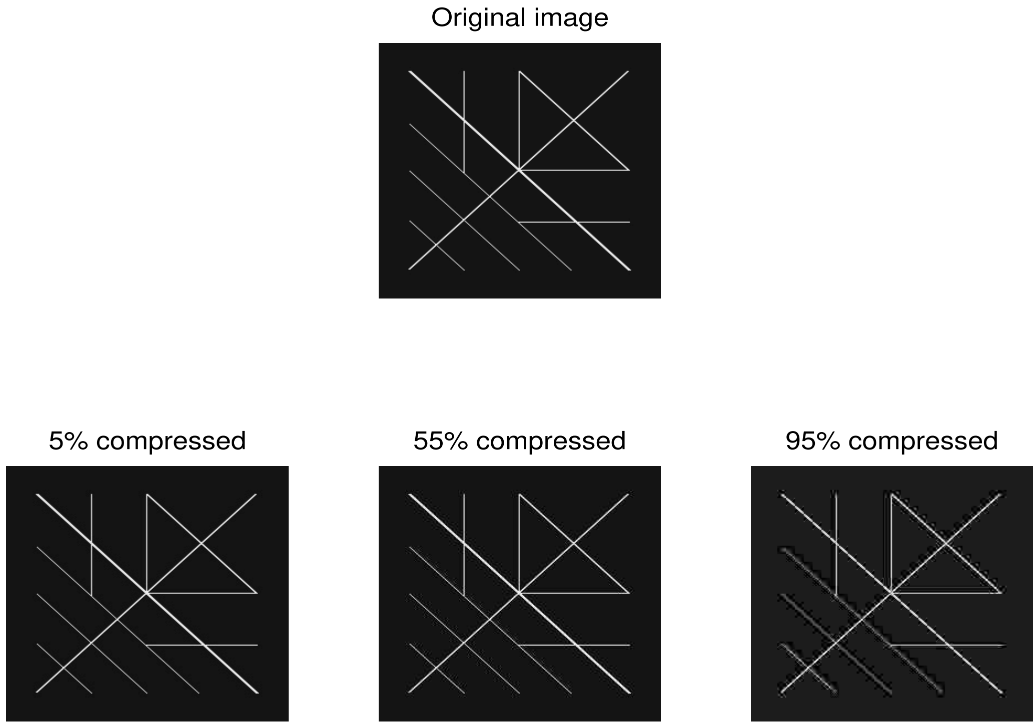


Figure B.4: The *crosses* image is displayed, along with the same image reconstructed from various levels of compression.

B.5 Fractals

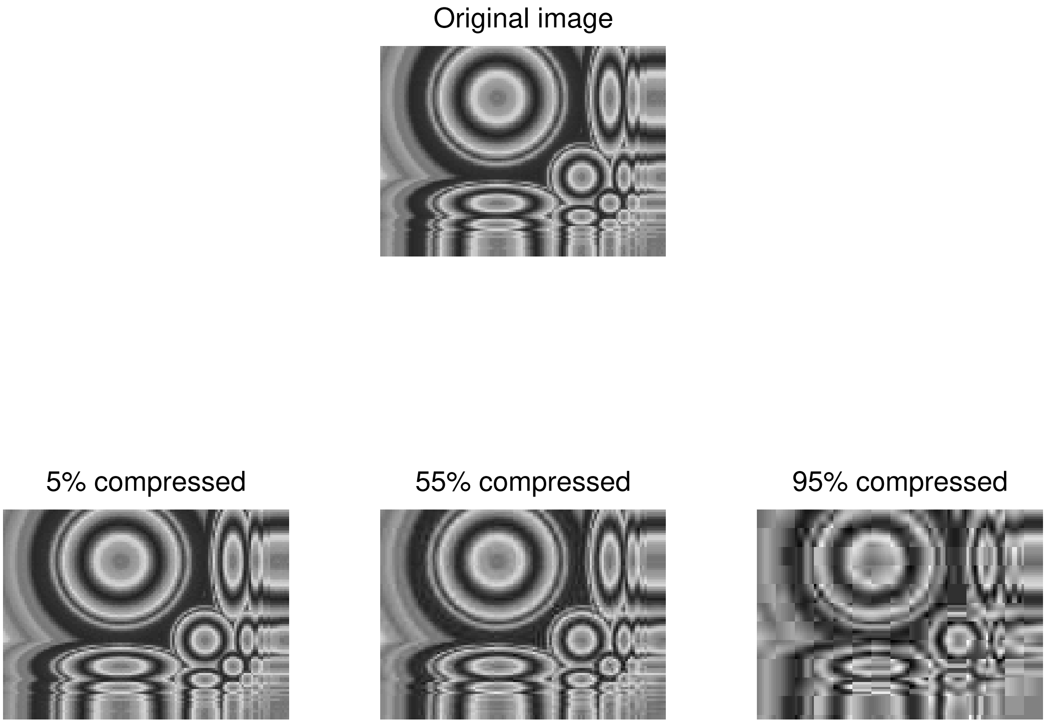


Figure B.5: The *fractal* image is displayed, along with the same image reconstructed from various levels of compression.

B.6 Snowface

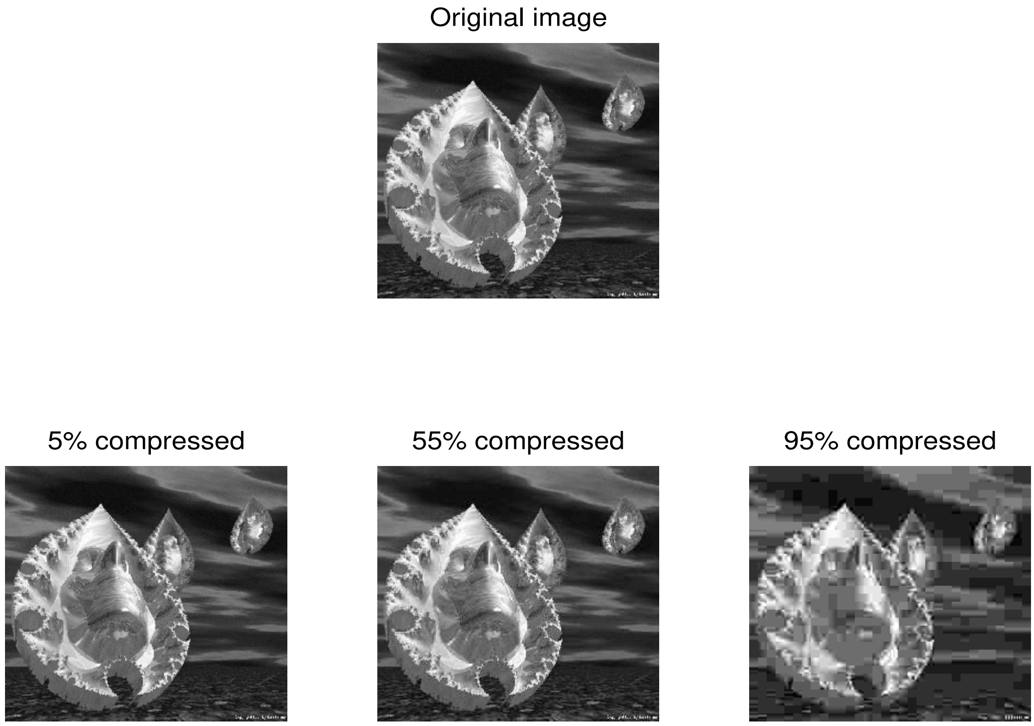


Figure B.6: The *snowface* image is displayed, along with the same image reconstructed from various levels of compression.

B.7 Eye Chart

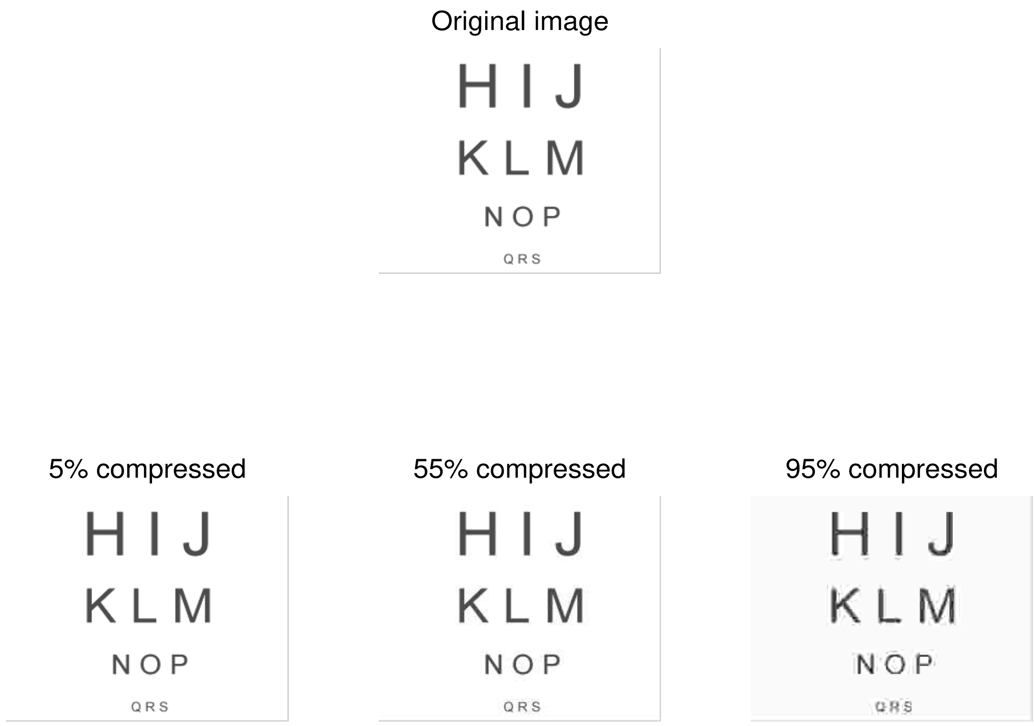


Figure B.7: The *eye chart* image is displayed, along with the same image reconstructed from various levels of compression.

APPENDIX C

The JPEG Still-Image Compression Standard

The Joint Photographic Expert Group (JPEG) was formed in 1982, with original goals including

- compression of still, continuous-tone greyscale and color images,
- cover the widest range of applications, including lossy coding, and
- to meet targets for compression ratio/image quality as defined in Table C.1.

Researchers interested in photovideotex began the JPEG in anticipation of the need for a standard progressive data compression scheme, that would operate at ISDN (64 Kbits/sec) rates [12]. In 1986 the International Telegraph and Telephone Consultative Committee Study Group VIII's Special Rapporteur Group on New Forms of Image Communication joind the JPEG.

In January 1988 the three techniques deemed by the group most likely to meet the goals outlined were tested at a new compression ratio/image quality target rate as shown in Table C.2. These new target rates were selected to 'stress test' the final contenders.

Eventually an Adaptive Discrete Cosine transform was selected as the best approach to meet the needs originally proposed by the JPEG standards committee. In use the scheme developed is referred to as the JPEG compression scheme, or simply JPEG.

Bits/pixel	Image quality
0.25	Recognizable
0.75–1.0	Excellent
4	Indistinguishable from original

Table C.1: Original JPEG compression ratio/image quality targets

Bits/pixel	Image quality
0.083	Recognizable
0.25	Useful
0.75	Excellent
2.25	Indistinguishable from original

Table C.2: Modified JPEG compression ratio/image quality targets

APPENDIX D

Further Results for the RMS Error Measure

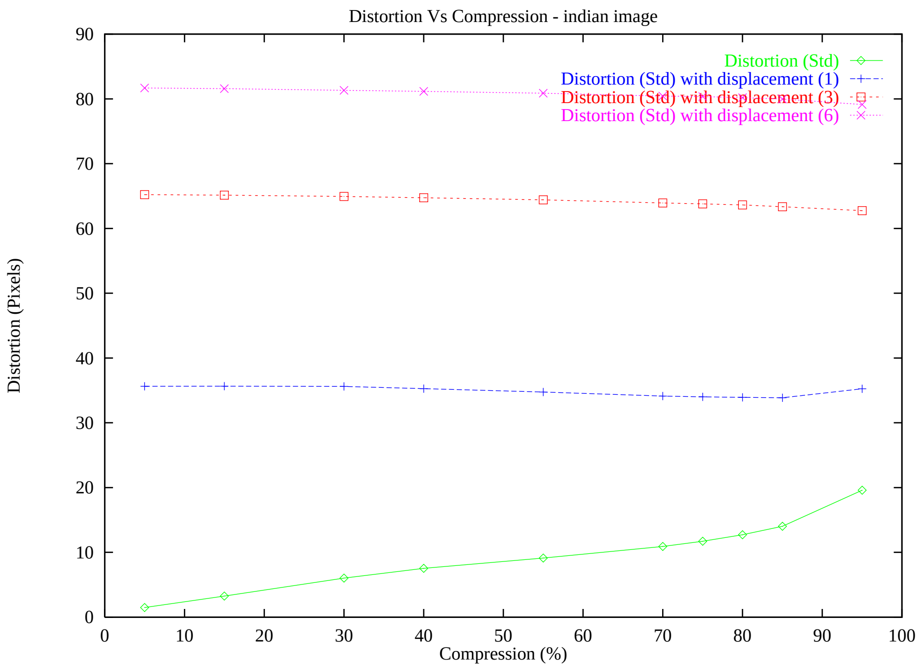


Figure D.1: The distortion returned by the RMS error when applied to the *indian* image for displacements of 0, 1, 3, and 6 pixels respectively.

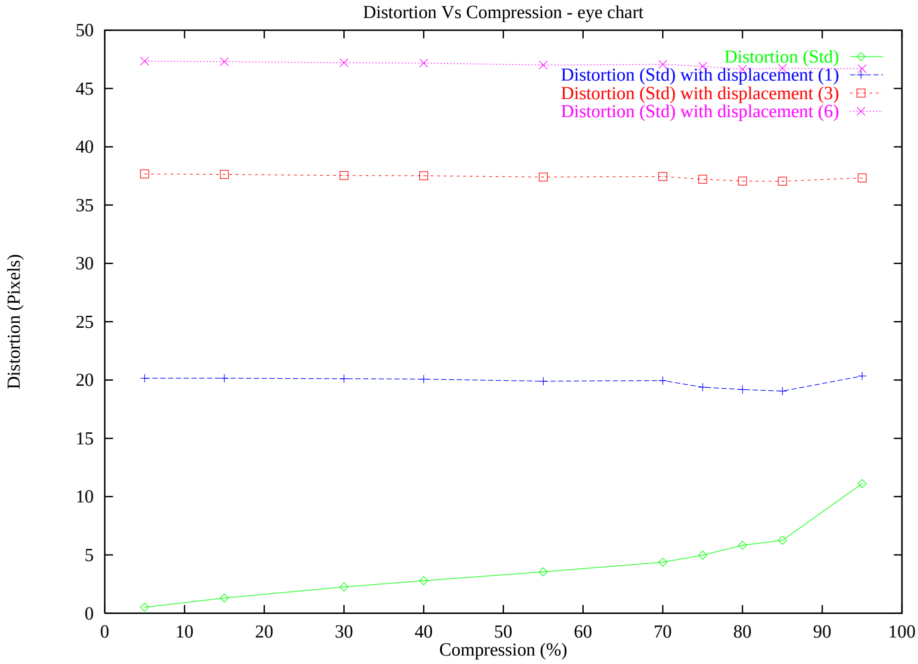


Figure D.2: The distortion returned by the RMS error when applied to the *eye chart* image for displacements of 0, 1, 3, and 6 pixels respectively.

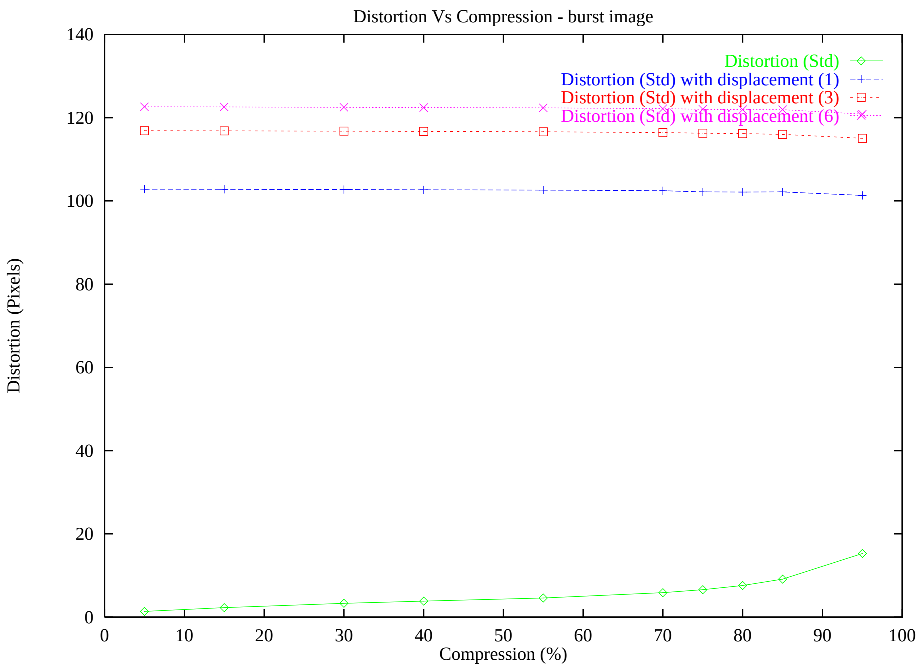


Figure D.3: The distortion returned by the RMS error when applied to the *burst* image for displacements of 0, 1, 3, and 6 pixels respectively.