CHAPTER 1
INTRODUCTION

This research project studies a special type of regression model introduced by Adcock in 1877 and further developed by Kendall (1951, 1952) and Fuller (1987) called unreplicated linear functional relationship (ULFR) model. The ULFR model considers the two linearly related explanatory variable X and response variable Y are both subject to errors. The multidimensional version of ULFR (MULFR) model and its coefficient of determination will then be considered and used as a performance indicator for some image processing applications.

Image applications are wide and have great influence on modern lifestyle ranging from the recording of a historical event, illustrating a meaningful story or disseminating useful information to solving selected problems. The demand for a higher quality image as a result of extensive applications of imaging technologies gave rise to the need to reduce errors in applications. As a consequence, there is a need to understand the performance of procedures in an imaging application which may be closely linked to identify the quality of the images involved. One common approach to performance and quality assessment is to evaluate if the reference image and the processed image still remain a high level of similarity. This evaluation process can be easily done visually. However, visual inspection is a highly subjective process and it is depends on the requirements of a given application (Gonzalez and Woods, 1992). Hence inconsistent results may occur because of human fatigue and subjective judgement. Digital image processing is an attempt to make this evaluation process more objective. It includes image acquisition, pre-processing, segmentation, feature representation and description, recognition and interpretation.

In this chapter, various types of digital image are introduced. The fundamental steps in image processing procedures are also briefly discussed. This is followed by the discussion of three problems related to comparing two images. The objective of the study is identified and efforts to achieve this objective are outlined in the last section.

1.1 Type of Images

A digital intensity image or greyscale image is a two-dimensional monochrome image *f*(*x,y*). It has been discretized or digitised both in spatial coordinates (*x,y*) and brightness, *f*, represents its grey level or intensity of the image (Gonzalez et al., 2004). A matrix $f(x,y)$, $x=1,2,\cdots,M$ and $y=1,2,\cdots,N$ can be used to represent a digital image where the elements of such a digital array are called *image element*, *pixel* or *pel*.

An example of the mathematical representation and schematically matrix representation of an image with size $M \times N$ in MATLAB environment is given in the Equation (1.1) and Figure 1.1, respectively. The most commonly use data classes are the double and uint8.

$$f(x,y) \in \begin{cases} [0,1] & \text{for class double} \\ \{0,1,\cdots,255\} & \text{for class uint8} \\ \{0,1,\cdots,65535\} & \text{for class uint16} \\ \{0,1,\cdots,4294967295\} & \text{for class uint32} \end{cases} \tag{1.1}$$

where $x=0,1,\cdots,M-1$ and $y=0,1,\cdots,N-1$.

A special type of digital image only consist of two logical values, 0s and 1s. It is called the Binary image. Although its output only contains black and white colours, an array of greyscale value 0s and 1s from data class ‘uint8’, ‘uint16’ or ‘uint32’ are not considered to be a binary image type.

$f(0,0)$	$f(0,1)$	$\cdots$	$\cdots$	$f(0,N\text{-}1)$
$f(1,0)$	$f(1,1)$	$\cdots$	$\cdots$	$f(1,N\text{-}1)$
$\vdots$	$\vdots$	$f(x,y)$		$\vdots$
$\vdots$	$\vdots$		$\ddots$	$\vdots$
$f(M\text{-}1,0)$	$f(M\text{-}1,1)$	$\cdots$	$\cdots$	$f(M\text{-}1,N\text{-}1)$

Figure 1.1: Matrix representation of an $M \times N$ image.

In the MATLAB image processing environment, a colour image is commonly handled as RGB image or indexed image. However there are other colour spaces such as NTSC, YCbCr, HSV, CMY, CMYK, and HIS (Pratt, 2001). An RGB colour image is represented by the three independent values of colour components which are the red, green, and blue. The RGB colour image may interpreted as a “stack” of three greyscale images with size $M \times N \times 3$. Figure 1.2 gives a schematic formation of RGB colour image where $Z_R, Z_G, Z_B \in [0, 255]$ in data class ‘uint8’. In this case, the corresponding RGB colour image is said to be 24-bits deep (i.e. each uint8 element require 1 byte = 8 bits. Hence the colour image has 8 bits x 3 colour bands = 24 bits) and the number of possible colours is $(2^8)^3$ or 16777216.

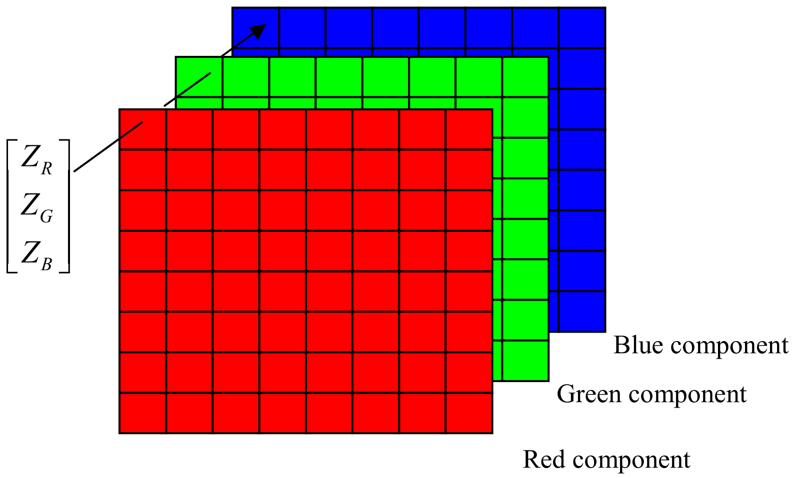

Figure 1.2: Schematic formation of RGB colour image.

An indexed image is a colour image contains two components; a data matrix of integers and a colormap matrix or 'map'. Map is an $m \times 3$ array of class double. Each row of map corresponding to the red, green and blue components of a single colour. The length m is equal to the defined number of colours. For instance, m = 256 if class uint8 is used and m = 65536 for class uint16. The colour of each pixel in an indexed image uses direct mapping where the value of integer matrix is used as a pointer into map. If the integer matrix is of class uint8, all components with value 0 point to the first row of a map, all components with value 1 point to the second row, and so on. Figure 1.3 (Gonzalez et al., 2004) illustrates the concept of direct mapping for a class uint8.

The NTSC or YIQ colour space system is commonly used in United States of America's television system. The NTSC format consists of three components; luminance (Y), hue (I), and saturation (Q). The luminance represents grey scale information and the other two components carry the colour information of a TV signal. The advantage is that the same signal can be used for both colour and monochrome TV sets. Another widely use colour space in digital video is the YCbCr. The luminance information is represented by component Y and the two colours difference components Cb and Cr stored the colour information. Component Cb is the difference between the blue band and a reference value, and component Cr is the difference between the red band and a reference value.

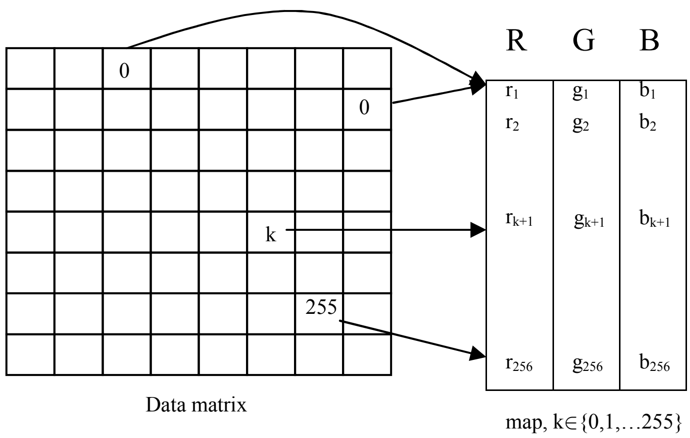

Figure 1.3: Schematic formation of indexed colour image.

Next, the HSV (hue, saturation, value) colour system is popular in colour selection from a colour wheel or palette such as paints or inks. The CMY colour space use the secondary colours of light i.e. cyan, magenta and yellow instead of the primary colours. Most devices that deposit coloured pigments on paper such as colour printers and copiers require CMY colour space. When a black colour is added to CMY, a ‘four-colour printing’ is created as CMYK colour space. Finally, the HIS (hue, saturation, intensity) colour space decouples the intensity component from the colour which carrying information in a colour image. It is an ideal colour description whereby it is natural and intuitive to humans view.

A video can be viewed as a sequence of images or frames with size $M \times N \times T$ where $M \times N$ is the size of each frame and T is the total number of frames considered (Figure 1.4).

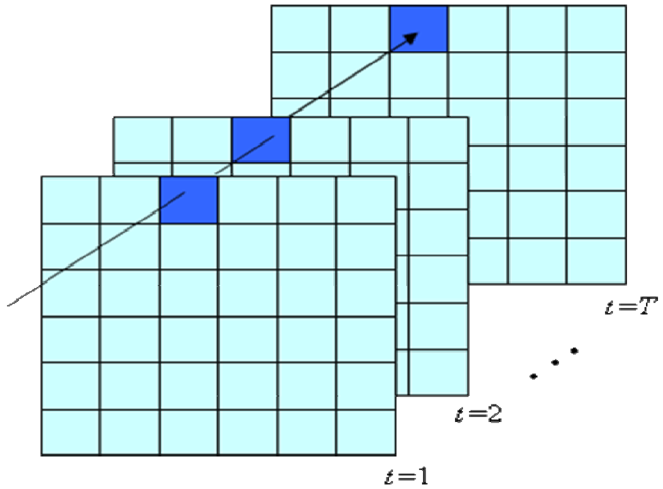

Figure 1.4: Temporal sequence of intensity values for a digital video.

1.2 Image Processing Procedures

Early image processing procedures generally revolve around the task of making images become easier to be interpreted by the human eyes. Image acquisition and image pre-processing are the standard initial procedures. Segmentation of images may then highlight important information leading to the search of useful image features, which in turn may be used for recognition and interpretation purposes (Gonzalez and Woods, 1992). A brief description of the five image processing procedures is as follows:

1. Image acquisition: a digital image acquisition system consists of three hardware components, namely a viewing camera, a frame grabber, and a host computer, on which processing takes place. The input to the viewing camera (charge-coupled device or CCD) is the incoming light, which enters the camera's lens and hits the image plane. Each photo sensor in the image plane converts light energy into voltage. The output of the CCD array is usually a continuous electric signal, and is sent to an electronic device called a frame grabber where it is digitized into a 2-D, rectangular array of $N \times M$ integer values and stored in a memory buffer. Finally, the digitized image is transferred to a host computer for processing.
2. Pre-processing: the key function of pre-processing is to produce an image of better quality (in a pre-defined sense) relative to the original image. Pre-processing typically deals with techniques for image enhancement, image restoration, and compression. Image enhancement techniques such as contrast stretching and histogram equalization are heuristic procedures designed to manipulate an image in order to take advantage of the psychophysical aspects of the human visual system. Image restoration is a process that attempts to reconstruct or recover an image that has been degraded by using some a priori knowledge of the degradation phenomenon such as defocused optics, motion blur, or errors in the transmission of image signals. Image compression devises compact representations for digital images, typically for transmission purposes. Bhaskaran & Konstantinides (1997) and Castleman (1996) referred compression process as the algorithms and architectures for the processing of image to reduce

its size or minimize the bit rate of their digital representation without affecting the quality of the image to an unacceptable level. Compression concentrates on two key policies that are redundant and irrelevant. Contemplating on redundancy, lossless compression may be adopted with the foremost absorption on attaining efficient approaches of encoding data that are exactly reversible with no loss of data in any form. This is especially useful for signal analysis such as to detect faults in machinery from sensor measurements, Electro Encephalo Graph (*EEG)* and other biomedical signals. On the other hand, contemplating on irrelevancy, lossy compression may be adopted with the foremost absorption on simplicity as far as information is concerned. Information that cannot be perceived will be removed to conquer huge file size (Mallat, S.G., 1998). The basis is on applications such as data and image compression and reduction of blur and noise. Contrary to lossless compression, lossy compression eternally eliminates or transforms data that has least significance. Nevertheless, intense compression may be alarming as the discrepancy would be detectable and noticeable which may result to a vastly deteriorated video (Dunn, J.R., 2002).

3. Segmentation: in image analysis, adequate filtering procedures are applied in order to distinguish the objects of interest from other objects and the background. The output of the segmentation stage usually is raw pixel data, constituting either the boundary of a region or all the points in the region itself. Segmentation algorithms for monochrome images generally are based on one of the two basic properties of gray-level values: discontinuity and similarity. In the first category, the approach is to partition an image based on abrupt changes in gray level. The principal areas of interest within this category are detection of isolated points and detection of lines and edges in an image. The principal

approaches in the second category are based on thresholding, region growing, and region splitting and merging. The concept of segmenting an image based on discontinuity and similarity of the gray-level values of its pixels is applicable to both still and dynamic (time varying) images.

4. Feature representation and description: after an image has been segmented into regions, the resulting aggregate of segmented pixels usually are represented and described in a form suitable for further computer processing. A region of interest can be represented in terms of its external characteristics or in terms of its internal characteristics. This representation is then described by features such as its length, the orientation of the straight line joining the extreme points, and the number of concavities in the boundary. Generally, an external representation is chosen when the primary focus is on shape characteristics. An internal representation is selected when the primary focus is on reflectivity properties, such as colour and texture. The features selected as descriptors should be as insensitive as possible to variations such as changes in size, translation, and rotation.
5. Recognition and interpretation: the last stage in image processing and analysis involves the process that assigns a label to an object based on the information provided by its descriptors (recognition) and assigning meaning to an ensemble of recognized objects (interpretation). Recognition implies that object descriptions or models are already available. The key idea in this model-based recognition is the comparison of image data with a database of models (Trucco and Verri, 1998). When a model is found to correspond to a subset of the data, one says that a match has been found, or that the model matches the data. The matched model is the identity of the object image, in the sense that both data and model represent the same object in the scene.

Closely related to these activities is the task of extracting statistical information from an image. For example image histogram yields the distribution of pixel values and image moments indicate image features such as luminance and contrast. One task frequently touched is the problem of comparing two images. Problems like optimising the algorithms and parameter settings for image acquisition system, assessing image quality of filtered and compressed image, and matching of input data with the database data in pattern recognition require rapid comparison and interpretation of images.

1.3 Problems Related to Comparing Two Images

A number of new problems developed, which do not arise when considering a single image. Three main problems are identified as follows:

1.3.1 No universal similarity measure for all applications

From the literature review, the performance of the existing measures mostly depends on specific image applications. For examples edge-dependent fusion quality index (Piella, 2004) is applied to image fusion problem, mutual information similarity measure (Chen et al., 2003) is used for image registration, and object count agreement (Yasnoff & Bacus, 1984) in segmentation application. The problems that arise will be whenever a new imaging problem exists, there is a need to either create a new similarity measure or choose one of the existing measures. This expensive and time consuming task suggests the need to create a 'multi-purpose' similarity measure.

1.3.2 Conditions of Using Similarity Measure

Before the suitability of a given similarity measure is agreed upon, three issues need to be considered. Firstly the assumption of perfect reference image, i.e. the reference image is noise free. Secondly the use of a single or multiple image quality attributes (example luminance only or luminance with contrast) as their input to calculate the similarity index, and finally computing the global image quality index by calculating the mean values of the local indices.

1.3.3 Levels of Difficulty in Comparing Images

There are many objective image similarity measures (ISM). In view of the availability of the original image, the objective of ISMs can be differentiated into three categories (Carnec et al., 2003; Wang et al., 2002b). They are:

(i) Full Reference (FR) measure: A perfect reference image is available to compare to the distorted image. Comparing the quality of the distorted image with the reference image is the most direct and simplest way of evaluating the quality of the distorted image. Not to mention that it is the most accurate way too.

(ii) Reduced Reference (RR) measure: Perfect reference image is not available to compare to the distorted image. In practice, it is difficult to obtain an original image as the Full Reference. A reduced reference image contains only partial information of the original image or a set of extracted features made available as side information.

(iii) No Reference (NR) measure: No original image or any information is available to compare with the distorted image. This is the most general case as most of the time reference image is not available. No reference measure can be used to assess the quality of the distorted image without having to compare with other images. However, this kind of measure is usually very complex and its result is not promising. It is noted that this thesis would not discuss further on the NR measure.

When the original reference image is available, FR metrics often yielded a good and robust measurement of image quality. However, in some applications, for example, transmitting the whole original reference image is not realistic. Therefore, RR metrics is a good alternative when the reduced reference size is not too large. The reduced reference can be coded and embedded in the bitstream and constitutes with a practical approach to quality evaluation. Finally, the NR metrics are the ideal solution since no extra data is added to the bitstream. Correia and Pereira (2002) considered the quality evaluation of video segmentation into two classes; standalone evaluation with no reference image and relative evaluation with a ground truth image.

1.4 Objectives of the Study

The main objective of this study is to develop a new statistical-based similarity measure for comparing two images where both are subject to errors. This study focuses on FR statistical-based similarity measures as a performance indicator. The proposed similarity measure allows a one-to-one comparison between the two images, namely reference image and distorted image.

To achieve this objective, the following will be done:

i) To survey existing statistical-based similarity measures in the literature and understand its limitations, and suggest improvement where possible.

ii) To develop a correlation-based similarity measure using the functional linear relationship model. This similarity measure gives a measure of relationship between two images that are both subject to errors. Errors may be due to pre-processing and/or post-processing methods.

iii) To investigate the properties of the new similarity measure. The robustness of the similarity measure to the model assumptions will also be studied.

iv) To apply the new similarity measure to selected imaging applications as a performance indicator. In particular an image compression illustrates a quality issue while the character recognition problem identifies a performance indicator.

v) Recommendation on the use or application of the proposed similarity measure.

1.5 Thesis Structure and Organization

This study is organized in the following manner:

Chapter One describes the main ingredient of the research topic. Definition and notation of digital images are provided. The background of digital image processing, in particular, image processing procedures are briefly discussed, and the problems of comparing images are highlighted. The objective of the study is to develop a new similarity measure based on correlation was made.

In Chapter Two, a literature review of the applications of existing statistical image similarity measures is given. The chronological remark for these statistical measures will also be discussed. Three image problems that have not been solved simultaneously in FR-SISMs will be highlighted and selected FR-SISMs will be discussed in details.

Chapter Three briefly illustrates the reasons why the conventional regression models and correlation methods are not used to solve the three image problems identified. This leads to the introduction of the one-dimensional unreplicated linear functional relationship (ULFR) model and its coefficient of determination (COD) as an initial solution to solve the three image problems. A review on ULFR model and the derivation of its COD will be discussed.

In Chapter Four, we discuss the multivariate version of the ULFR model with single slope, labelled as MULFR. The formation and properties of MULFR are illustrated. Its corresponding coefficient of determination is also derived as the new statistical-based similarity measure $\left(R_P^2\right)$ for some image processing problems.

The sampling properties of the proposed statistical-based similarity measure are investigated in Chapter Five. The performance of R_P^2 for various combinations of parameters is investigated. The robustness of R_P^2 will also investigate when the underlying assumptions are violated.

The next two chapters, Chapter Six and Chapter Seven attempt at illustrating the potential of the similarity measure. The successful application of R_P^2 in two very different problems, firstly character recognition in Chapter Six and secondly data compression in Chapter Seven is an attempt to illustrate the potential universal use of the proposed similarity measure.

Chapter Eight ends the study with a general discussion, recommendation and conclusion.