LINEAR ANALYSIS OF INCIDENCE STRUCTURE: APPLICATIONS IN NON-RIGID OBJECT RECOGNITION

by

Nicholas Wells, MEng.

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

March 2019

Abstract

Experimental developments in object recognition systems are delivered through image analysis. In non-rigid object recognition, view-invariance is a limiting factor in algorithm development. For supervised learning and parametric training data sets, recognition often fails when the statistics of features are not accounted for or are not discriminative. It is therefore imperative to identify, and appropriate salient features, input to the object recognition system. Features undergoing rigid body and projective transformations exhibit nonlinear relationships and as such, interactions between features can be complex to identify and measure. This thesis contains research into incidence geometry applied to non-rigid object recognition. The first part of this work measures correlation accuracy to recover angular displacements. Using a combination of nonuniform sampling and up-sampling, a matched filter analysis reveals a relaxation in interpolation complexity between sampled grid points. Furthermore, the technique identifies application in natural image structure analysis. The second part investigates the accuracy and precision of sub-pixel edge feature measurements. An arbitrary edge direction detection method based on non-integer coefficients and quadratic refinement reveals a precision measurement perhaps applicable for medical and manufacturing image screening applications at the millimetre and micron scale. Lastly, a recognition process utilising the Hough transform to measure and accumulate critical object feature statistics is investigated. Based on the training examples used to test the approach, the recognition error of a front profile was 8.2%. Identified sources of error include the profile measurement point locations and their approximation to characterise a profile. The results and applications of the novel adapted signal processing techniques are examined.

Acknowledgements

I would like to thank my academic supervisor Dr C. W. See for his knowledge, expertise, guidance and time over the years. I thank Dr A. J. Phillips for making it possible to undertake my PhD and Professor M. Mather for research insights.

The research discussed in this thesis was supported through the Engineering and Physical Sciences Research Council.

Contents

Chapter 1 - Introduction

This research examines the incidence and geometry of object features for applications in non-rigid object recognition. The study investigates appropriate linear signal processing techniques to improve nonparametric characterisations of non-rigid features. Focusing on signal pre-processing this work addresses the requirements of a system to acquire and extract invariant critical features (chapters 3, 4 and 5). Secondly, through appropriate geometric structure, the formation of feature sets projected onto linear subspaces (chapter 5) to recognise complex objects. Due to the interdisciplinary nature of signal analysis, the proceeding investigations lend to a wider application base outside the central scope of this thesis. These are presented and discussed accordingly. This chapter in brief introduces the subject area of object recognition with an overview of visual science, historic events and state-of-the-art methods. By examination of the biological motivations behind the modern-day computer vison interface a generic approach for developing an object recognition system is presented. Open areas of current research in object recognition that the reader's attention is brought to are view invariance of the natural world, the recognition of deformable objects and the utilisation of prior knowledge. For discussion of the novel contributions towards these issues the aims and objectives of the work are established, and the thesis chapters are outlined.

1.1 Principles of Object Recognition

Modern-day perspectives of replicating biological vision into algorithms for the computer vision interface are based on multisensory artificial intelligence and machine learning architectures. However, the variability of the world human vision operates in constitute simple and complex limitations that have been forever present for the artificial vision research community to digitally transform. Without consideration for the multi-modal neural plasticity of our five senses, replication of object recognition from biological vision can be characterised as: the identification and recovery of invariant signatures that are representative of the real world and its objects. Before this technological perspective is presented, this thesis first peruses the science of the visual system.

1.1.1 Physiology of the Visual System

The purpose of the biological visual system is to detect and interpret light that is emitted and reflected from object surfaces. This process can be split into two information pathways: anterior (detection) and posterior (interpretation). The eye (optical instrument), illustrated in figure 1.1, is a cascaded system that is responsible for the appropriate formation and transduction of spatiotemporal signals $I(x,y,z,t)$.

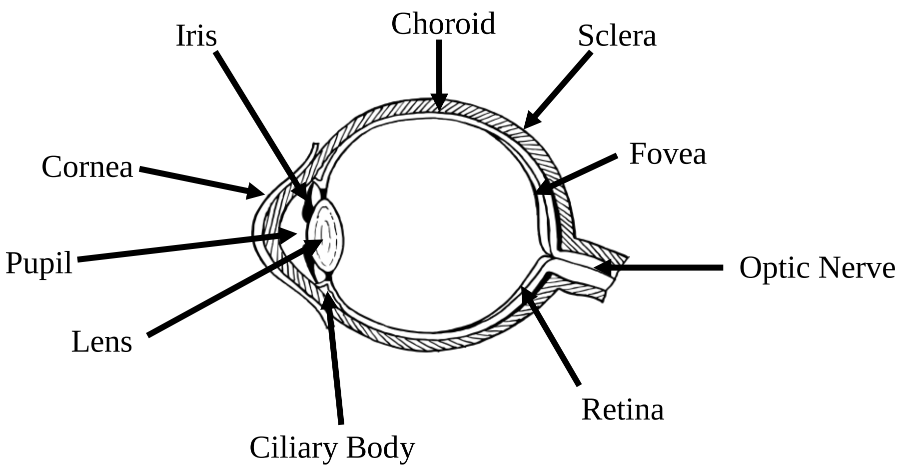


Figure 1.1: Layered illustration of the eye (anterior pathway) [1].

Simplistically, photons travel to the back of the retina through the lens of the eye which are brought into focus at the fovea. Once the signal activates wavelength sensitive photoreceptors there are a cascade of chemical-ion-to-electrical signal transformations through the synapse of retinal neurons spanning the layers of the retina; retinal pigment epithelium, photoreceptors, horizontal cells, bipolar cells, amacrine cells, ganglion cells and optic nerve. Such that approximately 10% of reflected light is detected by a dense population of photoreceptors and the rest is scattered or absorbed in tissue, the eye is regarded as a poor instrument susceptible to aberration and physical damage. This instrument is heavily optimised through a myriad of neural connections which are sensitive to colour, orientation, and spatial frequency. Two important questions that arise from this description that are most critical to this thesis, are: what are the signals detected by the photoreceptors and, what do these signals represent with respect to the form of the input? For readers with interest in a deeper knowledge of eye histology, biophysical perspectives are discussed in [2]–[4].

The posterior pathway of the visual system is a more complex abstraction of a path for information to flow. Signals arriving through the optic nerve travel to the optic chiasm and then to the lateral geniculate nucleus via the optical tract, the signal is then sent to the primary visual cortex, more commonly known as the visual areas, for further processing. Primarily between neuroscience and visual psychophysics, the study of purpose in the visual cortex is a research topic that is currently in debate. Early studies by Hubel and Wiesel [5] of the visual area 1, for which they were awarded a Nobel Prize for in 1981, demonstrated that in mammalian vision sensory receptors are sensitive to orientated spatial patterns. This is seminal with the computer vision community and is one of the main inspirations for advancements in this field of research. From what seems to be of little effort human vision has the ability to discern and perceive information almost instantaneously. Furthermore, digitally implementing visual models has proven to be not so effortless.

1.1.2 Cognitive Science of Vision

To introduce cognitive vision, consider memory as holistically labelled categorisation which enables the perception of the object by recalling visual experience. The neurons (sensory inputs) of the retina are a finite network of signal carriers providing input to stored memory that enables multiple systems of recognition to exist [6]. A memory of an object to correlate detected visual signals must have the ability to change. This means that the tuning of anterior and posterior neurons to specific signals develops, through experience, an order of importance to the signals the object is described by. It is implausible to holistically commit every object to memory, this is why object categorisation becomes taxonomic [7]. For example, infants would not recognise a Persian cat before they have learned to recognise it as a cat or even an animal, whereas, a skilled observer would sub-consciously bypass certain levels of categorisation through their experience of a broad object category.

Our visual world is non-parametric because visual experiences change through time and expand when our eye's gaze over visual information. Gaze is a function where our tuneable visual attention learns to look for information. Therefore, an object can only be holistic and unique when a number of parts representing form have been accounted for. Although an object part can take many forms due to pose or an observer's viewpoint, the part remains statistically the same. Recognition-by-parts [8], [9] is a theory whereby objects can only be recognised by a formation of parts statistically most similar, to what has been learned through experience. To realise the suitability of recognition-by-parts for manging a seemingly infinite memory space of visual data it is important to understand the preluding, although still relevant, theory of the Gestalt psychology.

Gestalt psychology [10] was a movement born out of Germany in the 1920's that changed the prevailing psychological view at the time most popular in America and Bavaria. The then current view was called behaviourism, where sensory inputs were thought to be modified by behaviour to acquire and reinforce sensory learning. The Gestalt's were not satisfied with the behaviourist's explanation [11] and they conceived that sensory learning was also a factor of cognitive processing. To remain brief and relevant, the principles of a Gestalt system are emergence, reification, multi-stability and invariance. Critically, in this system, the perception of an object is invariant to rigid body (translation, rotation and scale) and elastic transformation, and against other objects the strongest stimuli may not be the perception of the object the stimuli belong to. This infers organised perceptual patterns or groupings of a stimulus that varies between perceivers, the Gestalts called this prägnanz: meaning concise and meaningful. For an example of pattern grouping consider an incomplete or occluded object pattern that is still recognised as the object pattern that is familiar. This is made possible because the incomplete pattern of the stimulus is closed-based on the perceivers prior knowledge (experience) of the object. This treatment of perception and object recognition begins to carry tractability for a computational model. However, Gestalt post world war II fell out of favour, not least due to: poor understanding at a neural basis reinforced by neuroscience, formulation of its laws with little quantification, isolated studies of the prägnanz principal and demonstration through subjective reports [12]. It was also thought that popularity waned because when the first and second-generation Gestalts died, so did interest in developing their principles. Regardless, the Gestalt system, based upon holism and the structural representation of primitive shapes (geons), identified suitable constraints to base early computational models of vision on.

One example of an early visual model is that of D. Marr and E. Hildreth [13], who developed an edge detection algorithm modelled on the receptive fields of the retina. Even though the holistic representation of an object in Gestalt theory is not sufficient for a definitive cognitive theory of perception, it, at least, allowed basic object recognition to be digitally transformed into the simplest of cases. This, being the holistic geometric similarity of shape. Before the aims and objectives are given, this introductory chapter discusses computational vision from two perspectives [14]: a top down analysis based on object detection and a bottom up analysis based on visual perception.

1.1.3 Machine/Computer Vison Interface

A second seminal research direction that modern-day computer vision is founded on was by notion, that the biological nervous system was indeed a biological computer. W S. McCulloch and W. Pitts expanded this idea by re-stating the neuron as a logical unit [15], of which they tested that Turing machine programs could be implemented using a finite network of logical units. Known as neural networks, in 1947 network configurations for visual stimuli [16] became an active topic of study with initial success for networks invariant to linear and angular translations. The essence of the artificial neuron is

$$u = \sum_{i=1}^{n} w_i x_i \ , \tag{1.1}$$

where *u* is the weighted sum of all inputs *n*, for a vector containing individual synaptic weights ***W*** and inputs ***X***. The output *y* is a logical decision activated by a threshold value, typically this is

$$y = \begin{Bmatrix} 1, u \geq threshold \\ 0, u < threshold \end{Bmatrix} . \tag{1.2}$$

Configurations of networks based on weighted neurons that are tuned to characteristic features for a predetermined selection of inputs, has progressed understanding of fundamental theory and application of analysis in single and multidimensional natural systems such as speech, vision and multivariate analysis. A key component of this logic model is the activation function, a mathematical expression, which holds the state of the output against a threshold value. In computer vision research the limitation of a binary response to 0,1 places a restriction on the input signals used to train the characteristic weights for identifying regular patterns. This means that any input, not well characterised by features in the training inputs, will break the artificial logic of the finite system. To the extents of pattern recognition, success in the early models of artificial neural networks, this is not akin to what is considered a replica of a biological computer. A common remedy and still widely implemented in modern artificial networks is to increase the size of the training data to account for as much variability, characteristic, of the input as possible. Up until the 1970's increasing the number of inputs became computationally intractable because the computing technology was also not sufficiently advanced at that time. Throughout the 1980's interest in artificial neural networks, as a solution to what was initially thought of as a myriad of applications, faltered. Shortly before this period began, K. Fukushima [17], [18] published and later capitalised on his work on artificial neural networks for character pattern recognition. This seminal work was to be the birth of the convolutional neural network [19], developed by Y. LeCun et al. Computing power from the late 1990's helped regain a resurgence for seeking applications of artificial neural network and machine learning paradigm research, particularly deep learning strategies [20]. Founded in-part from this inspiration, F. F Li [21] launched Image-Net, the largest visual database of training samples used to train state-of-the-art neural networks in computational vision. This has further advanced object recognition because the seemingly infinite experience of human visual stimuli, through "*the internet of things"* and *"the cloud*", could now be practically obtained and stored. However, this is still far less an amount of data that an infant or even new-born would experience. Hence, accounting for an object's variability through a finite experience, especially non-rigid objects, remains a subject of focus in object recognition research.

Going back in time to the when the capabilities of artificial neural networks were beginning to be uncovered, there was, for each type of object to learn, a self-contained network tuned to that object's characteristic features. If a network was required to learn multiple inputs then it would require, at this early stage and in current methodology, multiple (sequential or parallel) artificially connected neural networks. However, this is not correct due to a finite nervous system nor plausible in terms of cognitive processing knowledge. This led to a parallel branch of research in computer vision, whereby characteristic features of an input signal are acquired as opposed to being learned. This resulted in highly constrained systems of pattern recognition which ultimately forced the scientific community to readdress the computational models which had been developed to deal with object invariance. Notable works and algorithms in this school of thought which are founded from D. Gabor's theory of communication [22] include R. O. Duda and P. E Hart's update of the Hough transform [23], [24], D. Lowe's scale invariant feature transform [25], [26], J. Canny's hybrid edge detector [27] and R. K McConnell's Histogram of Gradients feature detector [28].

1.2 Information Processing Issues in Object Recognition Systems

A similar information processing path exists for early and current object recognition algorithms. The key processing steps for a visual input along the generic model path of an object recognition system are signal acquisition, pre-processing, feature extraction and signal classification. Pertaining to each individual process, exists a multitude of roles and solutions for detecting and recognising objects. Scenarios and technological developments against selected techniques and algorithms that are used to carry out these key processing stages are discussed in detail in chapter 2. However, the relationship between each processing step are first outlined. The signal contains information relating to the physical structures within the image, this is usually chromatic or monochromatic pixel intensities. Firstly, the signal is conditioned to enable reliable and reproducible physical interpretations of the signal information. Secondly, low and high-level abstractions of the conditioned input signal are identified and extracted to build a database, or descriptor, relating to the characteristic features of the input signal. Lastly, by grouping the observed extracted features, or patterns of features, categories that identify object classes across class hierarchy can be predicted.

Through visual experience, a viewpoint dependency of an object does not exist for biological vision [29]. Object recognition based on early holistic models using edge/line detection and modern artificial neural networks that learn the characteristics of the objects features, fail if viewpoint dependency is not considered. Often, as stated earlier, coverage and selection of the characteristics in the training samples reduces much uncertainty. Viewpoints, to name but a few, are the object seen at different scales, in-plane angular and linear translations, out of plane rotations, natural light scales (mesopic, scotopic and photopic vision), and occlusion. If the object is not rigid, such as an animal, then an object under motion adds complexity to the system attempting to account for the model description of the object. Translation and rotational view invariance are the first investigations of this thesis. Against the constraints of the developed perspective of object recognition, techniques, to eliminate rigid body translations, linear and angular, are surveyed and built upon.

Take a car as an example, in any instant of time or view point the geometry of the car does not change. Hence, to have enough training examples of a car to compare to a plane requires a low-level account for the variance of the objects form. If the level of recognition is required to differentiate a hatchback to a saloon, then categorisation extends to a deeper level but still, not much additional variance of the object has occurred. The recognition of rigid objects reduces to holistic and local (characteristic feature) boundaries of shape across rigid body translations. Now, consider a non-rigid object such as a cat. Due to the nonlinear interaction of its geometric properties, such as eye separation distance, with a multitude of articulated viewpoint's, the amount of variance the object can exhibit is finitely large. Extending into deeper layers of categorisation for the object class cat, the amount of training data as compared to a rigid object, such as the car, is, assuming there are limits to articulated motion, also finitely large. Multiple non-rigid classes of objects may each constitute similar and non-similar characteristic features. Factoring additional object classes into the formation of a suitable input training set begins to look implausible and a most complex, if not impossible, task to engineer any single solution. As well as non-rigid object recognition being a central issue, the object class cat draws focus to much of the investigations in this thesis.

Modern perspectives of machine learning are directed towards architectures for learning reinforcement and the quantification of reward into artificial units [20], [30]. The purpose of the reinforcement methodology is to ensure artificial units, and the network they comprise, naturally adjust along with; experience of the characteristic and non-characteristic signals of our environments. An experienced visual system has already built up a knowledge of an object and is able to resolve such objects in the most extreme cases and contexts. If an object recognition algorithm has this object experience before it is required to differentiate between a vast amount of input signals, it would relieve some pressures on the desired performance to what is an already complex recognition system. Simply put, knowing what objects you are likely to encounter allows the system to be informed of which sensory inputs to activate or which training sample set to use or which types of characteristic features to detect. This is known as acquiring prior knowledge and it is not limited to objects, but rather the landscape (background) and the object of interest. Prior knowledge of the object class cat through past anatomical studies and a resolve of its most important features pertain to the latter investigations of this thesis in chapter 5.

It would be prudent to briefly appreciate where object recognition, in the context of image analysis and computer vision, is and will be having the greatest impacts on everyday life. Industrial processes are very much mechanised and require a labour force that merely exists to maintain and monitor. Including medical science, factory production/inspection processes and transportation technology; versatile and sophisticated object recognition algorithms are required to assist in the complete autonomous development of processes requiring visual intelligence.

For the recognition of the class object cat the level of recognition is application dependent. To recognise a cat from a dog as compared to recognising a breed of cat requires two sets of different abstracted features describing the salient and discriminant features between the object classes in question. For supervised learning the class or classes are known, and the problem is classification, whereas for unsupervised learning the problem is data exploration. This is problematic because similarity measures to identify clusters and between different clusters are unknown. Therefore, the focus of this research lies with the appropriation of object features input to the classification system and the roles features play at core pre-processing steps rather than the classification method itself; such as the popular recurrent/feed-forward/ convolutional neural networks and support vector machines [17], [18], [19].

The primary motivation for this research work is the image processing capabilities of retinal prosthetic devices. In such machine vision systems, the achievable resolutions to discriminate even the simplest of shapes is low (limited by the size of micro/optical arrays safely implanted surrounding critical biological tissue) and the operating environments are limited and largely known (added variance in complex image scenes). Based on associative neuronal firing patterns in biological vision to known objects, open questions exist regarding the extent to which machine vision systems can be implemented into applications where low resolution and systems' hardware size are limiting factors. A key driver for this research is to investigate the enhancement of the recognition range from known rigid objects to known non-rigid objects. A second motivation of this research is the application of measurement sensitive image pre-processing algorithms. One such environment where automated systems require accurate edge feature detection and tracking is in cytometry. Whereby, biological cells can be irregular in shape and exhibit textural nonuniformity, thus the accurate parametrisation of object boundaries become crucial in developing automated classification systems. An additional application of the work prescribed in this research is the recognition of a known object from a known background. For cytometry, both conditions are known, thus image registration and the accuracy of this process becomes critical. A further application is deformation analysis tracked through digital image correlation. Whereby, rigid body displacement as a precursor to image registration can add further distortion to small-scale measurements obtained through correlation that may be sensitive to additive noise.

Biological vision as a reference to base artificial vision on has been of great success to date but the digital implementation to the known standards of the biological visual system has not been so successful. By investigating linear algorithms applicable to non-rigid object recognition, it is hoped that useful insights will be gained for identifying complex object patterns that are hidden and simplified within linear subspaces. Based on what has been discussed thus far, the aims and objectives for the complete works of this thesis now follow.

1.3 Aims and Objectives

The aim of this thesis is to examine the incidence and geometry of object features for applications in non-rigid object recognition. The scope of the developed system is delivered by application of linear pre-processing techniques for recognition of a known non-rigid object (cat) from a known background.

The study begins by inspecting object recognition system processes to determine appropriate signal processing techniques in addressing their critical issues. This aim is delivered through a unique process formulated by the identified pattern recognition tools: correlation, edge detection and Hough transform. To gather understanding for the applicability of the techniques, current state-of-the-art image registration, feature detection, feature extraction and feature characterisation are researched. Emphasis is made on the measurement sensitivity of the techniques and their applications. This provides the investigative narrative for the research direction in implementing a linear signal analysis for characterising non-rigid object features.

An object recognition systems' core processes are feature invariance, feature extraction and classification. The application of correlation, convolution and the Hough transform to resolve these critical issues are empirically studied to deduce their extent of capability as solutions to perform non-rigid object recognition. Algorithm development is carried out by simulation and theoretical analysis. As far as time and the availability of resources extend, the comprehensive aim is to develop and test a fully integrated algorithm. Whereas, the research is presented by demonstration of the developed systems key stages. Due to the interdisciplinary nature of signal processing a wider aim is to identify research and commercial applications of the developed techniques outside the scope of the central aim.

The research objectives are as follows,

- Identify processing steps in object recognition and critically access signal processing techniques to address their capabilities in non-rigid object analysis.
- Develop and validate algorithms to enhance the measurement sensitivity of visual signals; characteristic of the identified object recognition processes and, the non-rigid signals the system is designed to detect.
- Assess the statistics of key object features to formulate salient and discriminate feature datasets appropriate for non-rigid signals.
- Formulate a non-rigid object recognition method and validation routine.

1.4 Structure of the Thesis

The central aim of this thesis investigates the core processes and bottle necks for the design of an object recognition system. For each core issue there is a self-contained chapter that flows from problem analysis, to solution design and implementation study, through to application. For each chapter there is a synopsis of the material to be covered and an introduction containing the subject matter to understand it. The following are brief descriptions of the five chapters that follow.

Chapter 2 contains an introduction and review of linear signal processing techniques that are subject for each self-contained chapter; correlation, edge detection and feature geometry: Hough transform. The motivations for each core chapter are developed based on the techniques sensitivity to distortion and fit for purpose in the context of non-rigid object recognition Wider practicalities and limitations focus on general issues, particularly shape irregularity and the sampling of digital signals.

Chapter 3 investigates view invariance and the application of correlation to register two images displaced by linear and angular units. The investigations are brought into context through discussion of object segmentation and feature alignment. Based on two methods of correlation (phase and normalised) and a simplified noise model, the precision and accuracy of correlation to recover linear displacements is discussed. To recover angular units, the Fourier-Mellin transform is chosen and an algorithm is developed based on nonuniform sampling and a relaxation of interpolation. An extended analysis of nonlinear correlation sensitivity between natural and rigid image structure is investigated.

Chapter 4 investigates procedures to extract object boundary information within images that is beyond the conventional sampling interval limited by a cartesian grid. The investigations are brought into context from the notion that features of interest belonging to a non-rigid object may be orientated in-between discrete angles. To extract boundaries of such objects, a method based on first-order differentiation and the Sobel operator is described to generate arbitrary kernel directions. This algorithm is called Region Maxima (RM) Edge Detection whereby potential pixel gradient directions using local region maxima are evaluated. The analysis begins by distorting a second-order model of a detected edge. The algorithm is then logically progressed via a second iteration of the RM technique to obtain a more precise measurement of a pixel’s gradient direction. This is called iterative RM Edge Detection.

Chapter 5 presents a method of recognition suitable for non-rigid objects. The process is a unique combination of the RM Edge Detection method and the Hough transform applied across a line and plane. The investigation of the recognition method is brought into context by demonstrating, through the obtained training data (appendix A3), two key aspects of the approach: recognition and view/pose tracking by data driven object feature analysis and prior knowledge accumulation. This analysis is progressed to identify and understand the linear/nonlinear interactions between the (appropriate) features used for the demonstrated aspects of the approach. For the purpose of familiarity and view-point dependency the object class cat is used throughout the demonstration.

Chapter 6 concludes on the work in this thesis by summary of the results from each chapter. The achieved objectives are analysed, and further work is discussed. In line with the central scope of the thesis and wider applications, extensions to research are developed.

When going through the main text of this thesis, readers can refer to the flow diagram in figure 1.2 and proceeding method description of the object recognition systems structure. It is intended as a roadmap for the details that follow in chapters 3, 4 and 5.

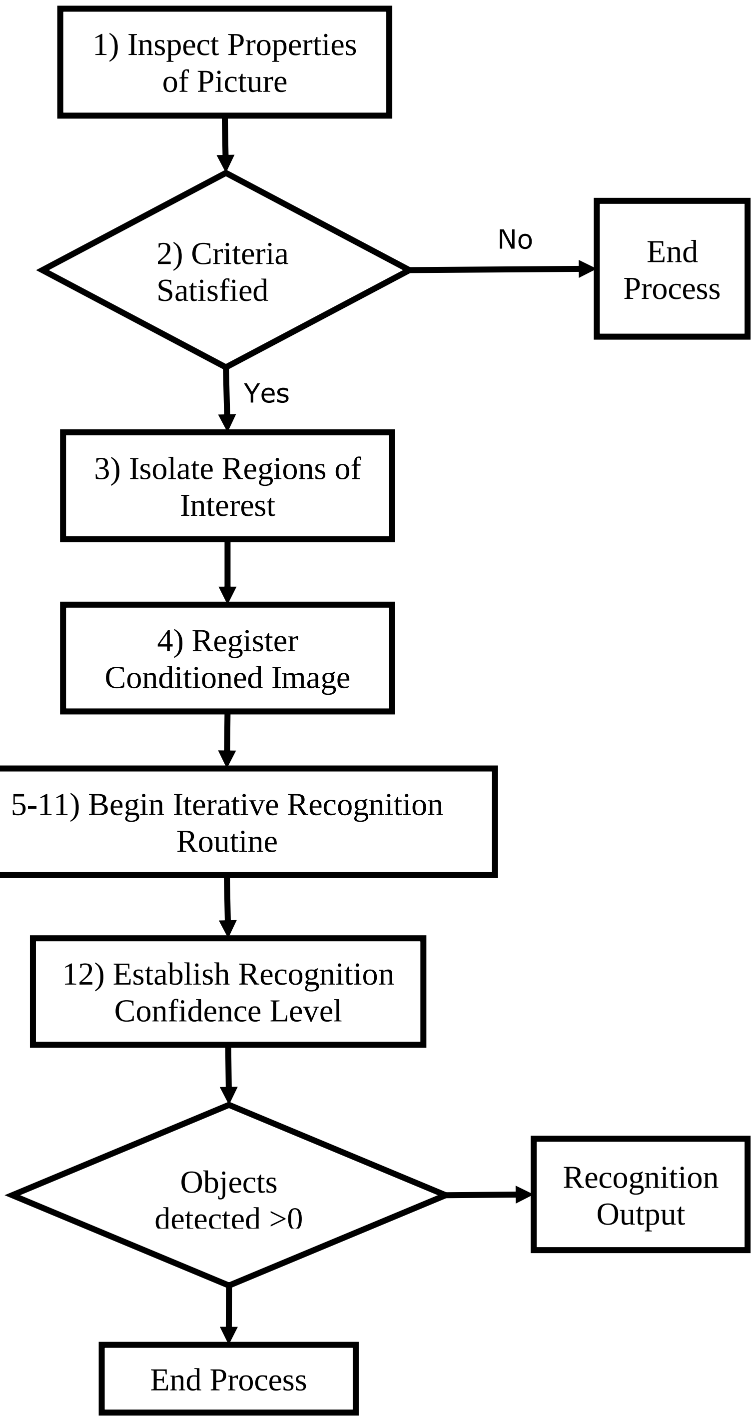


Figure 1.2: Flow diagram of the object recognitions processes

If you are given a picture you may follow the steps below to determine if it contains a particular object:

1) Determine the properties of a picture, including the dimensions (M, N), the average signal strength and the variance of the signal strength. This will allow a quick decision if it is worthwhile to carry on with the rest of the process. If the variance is small compared to the signal strength, the image may not be worth pursuing. ***(Section 3.2-2)***

2) Following a positive outcome from 1) above, a simple transformation removal procedure will be applied. This will be done in conjunction with a reference picture aiming to eliminate scaling, linear translation, and rotation.

3) The methods used to achieve 2) are segmentation, correlation and Fourier-Mellin transform: segmentation isolates an area of interest; correlation removes linear translation; Fourier-Mellin removes rotation and scale. ***(Section 3.4-6)***

4) Once the above pre-processing procedures have been completed, object registration will commence. ***(Section 4 and Section 5)***

5) From the image foreground properties identify appropriate parameters for edge detection: threshold values, feature orientation range. If necessary, store iterations of edge images for cross validation in 9, 10 and 11. ***(Section 4.2)***

6) The method used to achieve 5) edge detection: depending on knowledge of the sample of interest in the image, edge gradient directions across finer intervals can be obtained if required. ***(Section 4.4-5)***

7) Determine the order of the recognition process, this will depend on the object in task. The object class cat is used as the example here, where facial features are known. ***(Section 5.2-2)***

8) Implement the Hough transform to extract line and plane geometry of the first feature: feature data consists of position, length, orientation and radius. If no feature data is detected, repeat over a bank of edge images obtained from 5. ***(Section 5.4)***

9) Select appropriate training template or range of templates to ascertain search regions to detect the next feature in the recognition order. ***(Section 5.5)***

10) Following a positive outcome from 9) above, store the feature geometric data. Look for additional information from the recovered data indicative of the object in task that may assist in the recognition process. ***(Section 5.4-4)***

11) Repeat 8, 9 and 10) for the selected recognition order. Selection of the training examples may change; this will depend on the recovered feature information. If the view is outside of the training data, begin from 7) opting for a different recognition order. Otherwise object recognition process ends. ***(Section 5.5)***

12) Build up a confidence level of the object being recognised: measure the distance away from identified average training profile. If no features can be found due to the extraction method the object recognition process ends. ***(Section 5.4 and 5.6)***

1.5 Novel Work and Contributions

In this thesis, technological reviews, algorithm methodology and analytical approaches have led to new ideas and results. These are summarised as follows.

An object recognition process based on a continuous data driven approach for evaluating geometric inference is proposed and investigated. The object class cat is used throughout this thesis, but application of the recognition method and pattern recognition inquiries are not intended to be restricted to any one complex object.

During this study, a method to track systematic errors caused by angular structure within images is developed. A matched filter analysis to monitor systematic errors not indicative of image structure input into the autocorrelation scan led to preliminary results for the pattern recognition tool. The adopted nonuniform sampling process and look-up table demonstrates an enhanced accuracy using nearest-neighbour approximation of interpolation in-between uniform grid points projected onto a polar grid. A consequence of this technique is that 2D data must be processed as a series of 1D correlations. Secondly, a first-order adapted edge detection technique to measure arbitrary non-rigid object boundary information in between pixels is presented. Symmetry analysis of the reconstructed edge's sampling points reveals a measurement sensitivity that applications requiring accuracy and precision towards the micron scale may benefit from. An example application of these contributions is deformation analysis through digital correlation, whereby accurate tracking of structures and rigid body misalignment, extending to angular units, are crucial. The use of the proposed edge detector input into the Hough transform forms an approach to detect, identify, measure and track complex object features. The connection of these signal processing techniques applied to data driven object recognition is considered as novel. Wider applications are concerned with known irregular circular objects such as automated cytometry, whereby recognition confidence benefits from a data-driven routine and sensitivity driven algorithms.

Chapter 2 - Image Processing Techniques and Review

General approaches and the problematic aspects to devise a computational model of visual recognition have been discussed in chapter 1. As the complexity of multisensory biological visual processes are revealed, through what should be interdisciplinary research, the transformation of psychological and neurophysiological analysis into computational models become less tractable. Descriptively, the restriction to computational model development is driven by a necessity to first identify and reason biological sources, which remain in dispute and unresolved [31]. Since it is accepted that our macro visual world is a regulated and statistically stable space, this chapter explores three naturally inspired linear methods of visual signal computation and analysis: Correlation, Edge detection and shape detection using the Hough Transform. To familiarise the reader's understanding for the subsequent chapters, mathematically presented, each method is discussed in literature with respect to their origin, application, limitations and development in non-rigid object detection and recognition.

2.1 Image Processing, Methods, Challenges and Aims

Classification systems such as neural networks require discriminant input signals that the techniques that this chapter reviews can determine. The purpose of which, is to identify methods to build a feature space representative of the conditioned input signal. Image processing methods (statistical, visual, algebraic and transform based) and aims to extract such features are summarised in figure 2.1, which begins with the operation on a pixel/region or global frame to determine a feature.

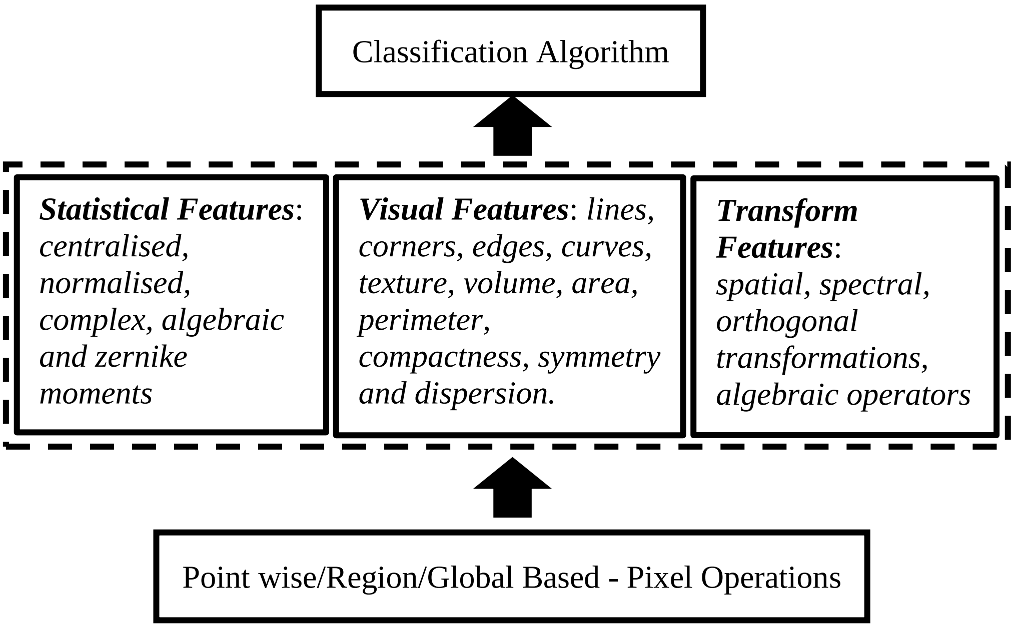


Figure 2.1: Overview of image processing methods in feature extraction.

The main challenges as discussed in section 1.2 to these image processing methods and operations to formulate appropriate inputs into a classification system are: occlusion, illumination, invariance, irregular, deformable and unknown shapes.

2.2 Correlation, A Measure of Statistical Similarity

Through experience the biological visual system purposely responds to an incoming signal by recognition of specific patterns contained within an observation. For visual signals, observations pertain to an object of interest. It is worth considering that in some situations such as fingerprint recognition, the human observer cannot distinguish to the same level that is achieved in face recognition tasks. Hence, for a broad spectrum of verification or recognition applications, machine pattern recognition systems are required to bypass certain limitations found in the biological visual pattern recognition system: speed of action, scalability and depth of the recognition problem (excessive numbers of object classes). A statistical mathematical function that in principle can navigate through complex pattern recognition problems is correlation. The correlation method is simple, and its purpose is to provide an indication of similarity between a created or selected reference signal and an observation.

2.2.1 Correlation Function

Correlation is a linear operation measuring the strength of similarity between two functions: a reference $h(\cdot)$ and an observation $g(\cdot)$. Similarly, covariance also measures the dispersion between two signals, $h(\cdot)$ and $g(\cdot)$. Except that for correlation it is equal to normalised covariance. The correlation between $h(\cdot)$ and $g(\cdot)$ is generated by a third function $c(\cdot)$. To bring this principle description into context for a 2D signal such as those captured by the eye or a static camera, each function is first defined on a grid; usually of a cartesian format The mathematical definition of correlation [32] between a reference image $h(x, y)$ and an observed image $g(x, y)$ is

$$c(x, y) \cong \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y)\, g(x \pm \Delta x, y \pm \Delta y)\, d\Delta x d\Delta y \,. \tag{2.1}$$

Formally, $c(x, y)$ is a function that generates first and second-order statistics directly related to the similarity between two signals; the deviation from the signal mean. Which if equal, has a correlation value of 1. If $h(\cdot)$ and $g(\cdot)$ are not the exact same function, then Eq. (2.1) is referred to as a cross-correlation. For $h(\cdot)$ equal to $g(\cdot)$, the correlation is referred to as autocorrelation

$$c(x,y) \cong \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, g(x \pm \Delta x, y \pm \Delta y) d\Delta x d\Delta y \,. \tag{2.2}$$

The correlation signal is holistically analysed and as such, to successfully apply this technique, distortion within real world observations input to a pattern recognition system must be considered. Distortion includes noise as well as rigid body displacements. Evident by Eq. (2.1 - 2.2), digitally processing double integrals over a finite interval for increases in image pixel resolution $\mathrm{I}:\mathrm{R}^{m\times n}$ incurs increasing computational expense. Since a pattern recognition system and its signal databases are usually designed to be linearly constrained, properties of the Fourier transform and its discrete computational implementation [33] help to reduce memory requirements of intensive processing tasks. Where u and v are spatial frequency coordinates of the Fourier transformed reference $H(\cdot)$ and observation $G(\cdot)$ signals, the correlation; a spatial shift, multiplication and integration of the product $h(\cdot)\times g(\cdot)\ldots$ reduces to

$$C(u,v) = H(u,v)\, G^{*}(u,v)\,. \tag{2.3}$$

To highlight the importance and applicability of the Fourier transform to perform correlation consider zero displacement between an identical reference and observation signal. The autocorrelation integral of Eq. (2.2) is equivalent to the power spectrum of the input signal,

$$\varphi(u,v) = \left|G(u,v)\right|^{2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, g(x \pm \Delta x, y \pm \Delta y) d\Delta x d\Delta y \,. \tag{2.4}$$

For a comprehensive handling of Fourier transforms applied to linear systems, readers can address theoretical text books by E.O Brigham and R. N Bracewell [32], [34]. Since the importance of processing spatial information in the Fourier domain has been established, this discussion now surveys applications of correlation in object and pattern recognition as well as wider image processing tasks.

2.2.2 Applications in Object Recognition

The early model of the visual system by D. Marr and E. Hildreth [13] extracts the object's edge based on the Laplacian of Gaussian filter operator. This feature detector holistically represents, what is regarded as, the critical information pertaining to the object. The object can be described by its whole form or, for a more discriminative pattern recognition system, via finer segmentation a summation of its parts. When no distortion is present, the boundary (edge) under stringent ideal conditions is a function that can be compared to (correlated with) a known boundary of a signal. This is called binary template matching [35], where both the correlation and the extraction of arbitrary closed boundaries [36]–[38] can be formulated and recovered based on obtaining and appropriate formulation of the signals Fourier coefficients. In order to discriminate ideal representations of rigid objects in well-defined systems this identification method is reliable and simple to implement. In practical situations objects may partly be hidden (occluded), not rigid and suffer geometric distortion, hence a holistic representation at the whole or part level may neither be applicable or attainable. Most importantly, without further signal conditioning it is unlikely for a real (rigid and non-rigid) observation to be completely invariant to rigid-body displacements or changes in illumination.

Rigid-body displacements are translations, rotations and uniform scale changes. Either through the observation or the camera system and excluding non-additive noise sources and aberration, these distortions are the extents to deformation that optical instruments capture. The spatial coordinates of image pixel intensity in Fourier space corresponds to a linear phase gradient,

$$F\left(I\left(x,y\right)\right) \cong \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I\left(x,y\right)e^{-j2\pi\left(ux+vy\right)}dxdy \,. \tag{2.5}$$

In addition to the strength of statistical similarity, the correlation between two signals identifies the displacement location given by the phase gradient. It then can therefore be recognised that the magnitude of Eq. (2.5) is invariant to translation. This crucial processing intervention has practical importance when the Fourier representation of a signal is considered to enable a transformation invariant to rotation and scale. Known as the Fourier-Mellin correlation transform this was optically demonstrated in 1976 [39], [40] to preserve simultaneously Euclidean and affine transformation. However, for bandlimited signals a pattern recognition system using the digital implementation of such a transform must consider [41], [42]; the discrete implementation of a continuous function (spectral leakage and sampling), the recovery of the original signal due to removal of the phase component, the effects of distortion in nonlinear transformations due to large translations and the sensitivity of measurement accuracy required by the application. Such issues as those described form the subject and basis for the investigations in chapter 3.

The autocorrelation function (power spectrum) is a powerful tool in statistical pattern recognition. As every signal decomposes into a unique composition of Fourier components it is reasonable to assume that similar signals must form similar spectrums. Hence, the underlying characteristics of discriminative signals are indicators for the types of observations being made. For a well-defined system and application, acquiring prior knowledge at a general level begins the process of elimination that the object recognition system ultimately bases its decision on. From studies in natural image statistics [43]–[45], it has been found that within degrees of error, due to generality, particular combinations of background structure and foreground objects exhibit unique frequency magnitude spectrums [46].

2.2.3 Fast and Accurate Digital Correlation

The application of digital correlation to image registration has been shown to enhance the capabilities of object recognition systems [47]. Obtaining a registration accuracy below the pixel level has been of interest for medical diagnostic systems and analysis [48], [49], mapping and tracking of targets in remote sensing [50]–[52] , object motion estimation in video processing [53], [54] and non-destructive evaluation of deformation [46], [55]. In addition to issues associated with a discrete implementation, the precision of the correlations statistical accuracy may be sensitive and influenced by factors constituent of the environment the measurement is taken from. Subject of this critical précis are robust algorithms that improve the precision of accuracy in digital correlation. In ref [56] four common methods of registering data at a sub-pixel level are discussed. For decreasing orders of accuracy these are; intensity interpolation, differential method, correlation interpolation and the phase correlation. Traditionally, accuracy is traded with speed of computation, where even small data dimension memory requirements can be a burden to the processing system. However, although not a specific topic of focus for this thesis, computational speed is becoming less of an issue for two reasons: computing hardware technology [57], [58].

A differential sub-pixel measurement is based upon modelling optical flow between the spatial derivatives of a reference frame to the temporal derivatives of a sequence of observations [59]. Briefly, the temporal signal may suffer from high frequency perturbations if the discontinuity of data intensity between the frames of a reference sequence are large [60]. Often is the case with many signal processing systems, a low pass filter is usually applied to mitigate large intensity variation providing a trade-off; precision and accuracy due to averaging between samples. An obvious choice of application for this method would be to measure object motion displacement. This method of sub-pixel measurement however, is normally implemented on planar objects with a prior correlation algorithm to recover a pixel level estimation of the displacement field [61]. An advantage of the differential method is that there is no reliance on an interpolation method between discrete samples.

The phase spectrum of an image's Fourier transform contains the useful information of the input signal: the phase distribution and location of spatial data. The phase correlation between $H(u,v)$ and $G(u,v)$ is

$$C_{\omega}(u,v) = \frac{H(u,v)G^{*}(u,v)}{\left|H(u,v)G(u,v)\right|}, \tag{2.6}$$

where the normalisation removes the reliance on image content and retains the relative displacement found between the phase spectrums of the reference and observation signal. Sub-pixel refinement is obtained by re-sampling a coarsely defined grid surrounding the phase correlation peak, of which, the sampled points can also be interpolated to refine the accuracy; providing the sampling criterion is satisfied. The outcome of the ideal correlation is a delta function at the location of displacement. For any given spectrum of noise, the phase correlation measurement can become unstable. Phase correlations outside of the delta function that should be minimised for real data may be amplified and a real inhibitor in revealing the true correlations peak and location. This can be easily observed when the energy of the auto-correlation peak between two identical signals is neither equal to 1 or located at x_0, y_0. It should be obvious that for real signals in real environments false correlation peaks can occur randomly due to noise and underlying signal characteristics. Most applications [62]–[64] using this correlation method do so because it is assumed that low frequency noise distortion is the dominating signal response. Overall, image content, intensity variation and narrow band noise are minimal contributors to the phase correlation error, and as such phase correlation is a widely adopted algorithm for registering displacements suffering from geometric and intensity distortion.

Either at the pixel or sub-pixel level, the accuracy of a correlation's measurement is heavily influenced by sampling. Precision on the other hand is dictated by the algorithms ability to deliver a stable response over a naturally varying bandlimited signal. Intensity interpolation is a method of up sampling a signal to a much denser sampled grid for an increase in correlation accuracy. For a rectangular grid spacing, an image of dimension $M \times N$. is resampled by εs to $M / \varepsilon s \times N / \varepsilon s$ for an increase in accuracy with a range: $0 < \varepsilon s < 1$. The accuracy of the correlation is therefore restricted to the approximated resampled versions of the original signals. Under the assumption that $H(\cdot)$ and $G(\cdot)$ are bandlimited signals and the original image is sampled at the Nyquist rate [65], a unique recovery of the original signal can be obtained using *sinc* interpolation. In 1D this is

$$\operatorname{sinc}(x) = \frac{\sin(\pi x)}{\pi x}. \tag{2.7}$$

For $\operatorname{sinc}(0,0)=1$ and $\operatorname{sinc}(n,n)=0$ where *n* is a nonzero integer, the sinc function is an ideal low pass filter with a cut-off frequency of 1/2 cycles per pixel. The denser sampling function in the Fourier domain for an accuracy between $0 > \varepsilon s < 1$, is $\operatorname{rect}(u/\varepsilon s, v/\varepsilon s)$. To construct the ideal resampled image, the theoretically infinite filter length of the sinc function is convolved with every data point in the original bandlimited signal. However, this is a time intensive process and as such approximations to the sinc interpolant are usually applied instead [66], [67], examples include: nearest neighbour, bilinear and bicubic interpolation. Variants of the standard Fourier domain approach to correlation using up-sampling of the original functions are discussed in ref [68]. In [60], [69], [70] authors also discuss the extension of interpolating approximations across resampled images in phase correlation. Finally, and perhaps the simplest of interpolation strategies is to fit a curve through discrete points of the correlation peak [71]. An even polynomial will describe a symmetrical correlation peak, whereby a second-order polynomial will suitably locate the peak. Higher order terms may increase the SNR but not the accuracy of the peak location. Applying a high sampling rate to the discretely sampled function may or may not improve the situation.

2.3 Edge Detection and Numerical Differentiation

A question posed in chapter 1 asks, what do the signals detected by the retina represent with respect to the form of the input? Via the layers of the retina, vision develops rapidly from birth through exposure to environments. However, retinal neuronal connectivity is not a completely understood processing system. Readers can acquire comprehensive background knowledge on the synaptic activity of neurons using ref [72] , however to generalise; specific features of visual signals are aggregated through activations that retinal neurons are tuned to. Biological studies in [5] identified that a normal mammalian visual system is sensitive to orientated spatial patterns. In digital images, objects of interest could be described by a perimeter or boundary (primitive sketch) across orientation and scale changes. Thus, an interpretation of human vision and its relationship to image processing is established. The reason for there being such a relationship is that boundaries that can be seen correspond to discontinuities in contrast between an object and its surroundings.

Boundary and perimeter are interchangeable terms and are components of edge analysis. In the ideal scenario edges are a representation of shape which are detected without knowledge of shape. Along with curvature and corners, edges are identified as key low-level feature abstractions in the field of computational vision [73]. Similarly, to object motion, an edge is the position of a rapid change in intensity; logically, this is detected using differentiation. Discussions on edge detection techniques proceed after a brief mathematical dissection of numerical differentiation.

2.3.1 Numerical Differentiation

First consider an operator to detect a change in intensity between two vertical adjacent points. Since the direction of a vertical differencing operator emphasises vertical intensity changes, the vertical difference of a vertical operator is equal to zero. For an image $I(x, y)$ the vertical and horizontal edge operators of two adjacent points are $E_x(x,y)$ and $E_y(x, y)$. In combination these operators detect both horizontal and vertical edges, of which the coefficients of the first order difference kernel $[2 -1 -1]$ are extracted from $E_{x,y}(x, y) = \left|2I(x, y) - I(x+1, y) - I(x, y+1)\right|$. Next consider that for any function the sum over an infinite boundary of the function's derivatives taken at a single point is a Taylor series expansion of that function. This means that any function can be approximated by a finite number of terms with the truncation error $O(\Delta^{n+1})$. Consider the second order 2D Taylor series

$$f(x+\Delta x, y+\Delta y) = f(x, y) + (\Delta x, \Delta y) f'(x, y) + \frac{\Delta x^2, \Delta y^2}{2!} f''(x, y) + O(\Delta x^3, \Delta y^3), \tag{2.8}$$

where f' is the sum partial first derivative $f'_x + f'_y$ and f'' is the sum partial second derivative $f''_x + f''_y$. The first order derivative approximation known as the forward difference method, is

$$f'(x, y) = \frac{f(x+\Delta x, y+\Delta y) - f(x, y)}{\Delta x, \Delta y} - O(\Delta x^2, \Delta y^2). \tag{2.9}$$

Crucially the error of the derivative approximation is confined to the step size. The complexity of functions in terms of a curve, surface or object causes large errors when the coefficients of higher order derivatives are also large; high frequency oscillation and stability of the coefficients. Whilst one error may be limited by the structure of the function, the error due to the step size that the derivative is calculated from can be reduced by taking an additional adjacent point to smooth the approximation. Known as the backward difference approximation, the function of the additional adjacent points is evaluated from the Taylor series expansion of

$$f'(x,y) = \frac{-f(x-\Delta x, y-\Delta y) + f(x,y)}{\Delta x, \Delta y} + O\left(\Delta x^2, \Delta y^2\right). \quad (2.10)$$

By including the additional adjacent points, the first order derivative is obtained by differencing Eq. (2.10) from Eq. (2.9). The average or smoothing action performed by including the extra point reduces the overall error when the step size is considered in the approximation. Hence, the first order numerical derivative is calculated from

$$f'(x,y) = \frac{f(x+\Delta x, y+\Delta y) - f(x-\Delta x, y-\Delta y)}{\Delta x, \Delta y} + O\left(\Delta x^2, \Delta y^2\right). \quad (2.11)$$

The edge operator $E_{x,y}(x,y) = \left| I(x+1,y) - I(x-1,y) + I(x,y+1) - I(x,y+1) \right|$ generates the coefficients of the differencing kernel for the vertical and horizontal edge directions, the coefficients are: $D_x = [1\ 0\ -1]$ and $D_y = [1\ 0\ -1]^T$. Known as the central difference method, Eq. (2.11) forms the fundamental building block of edge detection and analysis for interpreting images.

2.3.2 Edge detection Techniques

The change in intensity of an edge can be varying slowly or rapidly, this means that information that may not be of the boundary (holistically and parts based) within and around the objects of interest will also be detected. For this reason, thresholding is applied to eliminate certain pixels values which potentially reduces the amount of noise (both - signal and scene) detected by the edge detector. In the ideal scenario there is a large contrast between a uniform object and a uniform background. For real images, the pixel intensity distribution is multimodal and the application of thresholding is non-trivial. For a grayscale image Otsu [74] established a threshold method where the separability between pixel classes is maximised, this is particularly suited to bimodal pixel intensity distributions. This can also be extended to handle multimodal distributions by combining or selecting a range of threshold values. In both cases, the statistics of the image pixel distribution establish such threshold values and to generalise between a variety of structure in an image, it is often a trial and error procedure. Thresholding is not covered in any further detail but readers are directed to [75] for a survey of thresholding techniques beyond histogram approaches.

Noise, false and missing edges are limiting factors in successful edge detection. Generally, there are three types of edge detectors based on a differentiator kernel: first order, second order and hybrid. Every edge detector algorithm ever conceived is not detailed here as there is vast literature on this subject. The readers can follow [76] to gather understanding of an array of techniques. However, of applications requiring edge detection, two approaches are widely implemented: Sobel [77] and Canny [27] edge detection. It is discussed in [78] that empirical studies in the statistical analysis of edges is still primitive in research. The authors elude that whilst computer vision algorithms and applications provide a role for the edge information of images, they do not uncover what the statistics of edge gradients are. To understand the Sobel and Canny edge detectors their fruition and limitations are now presented.

The Sobel operator is a first order edge detector, whereby for a step edge or ramp function $f(\cdot)$ the centre of the edge for an arbitrary direction is located by the maximum of the functions derivative $f'(\cdot)$. By applying a threshold, the location of the pixel at this maximum declares an existence of an edge. In accordance with ref [77] the formulation of the Sobel kernels is easily identified using a vector gradient sum relative to the radial position of the kernel's coefficients on the unit circle. Usually kernels are square and odd so that the centre pixel in the image aligns with the symmetry of the kernel. The configuration of the Sobel operator D extracts edge orientations at $0^{\circ}, 45^{\circ}, 90^{\circ}$ and 135°, these operators are

$$D_{\theta=0^{\circ}} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \qquad D_{\theta=45^{\circ}} = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix},$$

$$D_{\theta=90^{\circ}} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}, \qquad D_{\theta=135^{\circ}} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & -2 \end{bmatrix}. \tag{2.12}$$

Comparing the kernels of Eq. (2.12) at 0° and 90° to the corresponding kernel coefficients obtained from Eq. (2.11), there is an additional vector ($Filter = [1\ 2\ 1]$) representing a low pass filter along the direction perpendicular to the differencing kernel. The outer product of $D_x \otimes Filter^T$ formulates the 3×3 Sobel kernel $D_{\theta=0^{\circ}}$. First order numerical methods are simple to implement but are sensitive to noise. The rate of change for a discretely sampled point of a random signal will not be a constant value and can be significantly larger than the variance of the noise at a point that is distorting an edge. Hence, a triangular smoothing operator $[1\ 2\ 1]$ is applied to the convolution kernel. Two additional concerns with first order methods are; without thresholding, the edges of homogeneous surfaces in images such as sky and water are detected along with the boundaries of interest; the kernel masks of Eq. (2.12) evaluate each pixel using a neighbourhood (additional adjacent points), which results in averaged edges. A final point is that matrices of the Sobel operator can be theoretically designed for larger kernel sizes. However, the kernel coefficients beyond 3×3 are to an extent arbitrary and they can be selected to enhance features such as: suppression of pixel intensity or smoother graduations of high intensity variation.

Second order differentiators offer the ability to better localise true edges because the maximum of the first derivative may not be uniquely identifiable in real image data. For the location of the first derivative's maximum, the second-order derivative yields a zero. This corresponds to the location of a constant rate of change, hence by searching for the zero crossing an edge is detected. The numerical approximation of a 2D second-order central difference method up to the second term is

$$f''(x,y)=\frac{f'(x+\Delta x,y+\Delta y)-2f'(x-\Delta x,y-\Delta y)}{\Delta x,\Delta y}+O\left(\Delta x^2,\Delta y^2\right). \tag{2.13}$$

The generated coefficients of the second order differencing kernel for the vertical and horizontal edge directions, are thus: $D_x^2 = [1\,-2\,1]$ and $D_y^2 = [1\,-2\,1]^T$. The outer product $D_X^2 \otimes D_Y^2$ is equivalent to the well-known Laplacian operator (∇^2), which is defined as the divergence ($\nabla \cdot$) of the gradient operator ∇. In approximating a second order derivative of a function, the Laplacian of an image is a sum of its partial derivatives; $\nabla^2 I(x,y) * D_x^2 + I(x,y) * D_y^2$. Thus, the Laplacian kernel is defined as

$$D_{\theta=0^\circ} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{2.14}$$

The stability of detecting a zero crossing in noisy images is worse than the first order method. Causes include: a zero crossing may not be exactly on top of a pixel but in between and that, the sensitivity of a differential operator increases with the order of D^n. In the same vein as first order edge detection a low pass filtering function is required to optimise the performance of the edge detection. This leads this discussion onto optimal edge detectors; the Canny edge algorithm and the Marr-Hildreth algorithm, of which both employ Gaussian smoothing. Without rewriting the countless studies and explanations of theory of these two algorithms [13], [27], [79]–[82] this discussion will describe Gaussian smoothing and briefly highlight key points of these two approaches.

A 2D Gaussian is a symmetric signal about its origin and analogous to the 2D *sinc* function in frequency domain signal processing, the 2D Gaussian is the ideal response in spatial domain signal processing. Theoretically existing over an infinite boundary, the truncated Gaussian signal response is controlled by the standard deviation. For greater smoothing the standard deviation increases along with the size of the kernel (bandlimited response). Less weight is allocated to the coefficients of the Gaussian signal away from its centre (mean value). This is advantageous for edge detection because an edge kernel places the greatest weight to the image pixel at the centre of the edge kernel. In this sense this type of low pass filter is optimal in both smoothing and localisation. The Canny algorithm convolves a 2D Gaussian with an image to reduce noise, before usually applying the Sobel method to differentiate the image. Whereas to compare, Marr-Hildreth utilises the second derivative of the Gaussian before convolving with an image, hence the term Laplacian of Gaussian. For this approach the edge and Gaussian kernel's size will be smaller and equal in comparison to the image. Hence for ∇^2, separability reduces the computational arithmetic of convolution. Whilst second order differentiators can be applied to the Canny method instead of the Sobel kernel, the main points of the Canny algorithm are: maximal suppression of edges of the gradient image and hysteresis thresholding to link continuous edges.

2.3.3 Applications of Edge Detection

The edges of objects are the initial first guess in computer/machine vision algorithms to base the probability of binary decisions upon. Edges of an object may be the perimeter of object shape or the boundary of parts and detail that make up the object. Whilst object recognition presents itself as the likely application for edge detection, this section first discusses edges within the wider context of digital signal processing.

An edge represents a visual structure, where geometric properties such as length, width, curvature and orientation present features describing the information of that edge. Consider the discussion of correlation applied to image registration in the first section of this chapter. The accuracy and precision of alignment between sets of images can be improved by correlating non-random highly structured visual information. In [83], [84] the authors present a digital correlation technique to measure microscopic deformations through thermal strain on microelectronic devices. Through a deformation free and a deformed image of the sample, thermal strain is tracked through the correlation peak measurement. It is commented that linear displacements add to the sensitivity of focusing error in the microscope and the accuracy of correlation. Tracking changes without knowledge of the object is more useful for fast online detection and inspection. Since edge detection such as Sobel detects edges without knowledge of shape, then with knowledge that structure in the imaged sample exists, correlation of low frequency information can be eliminated. Furthermore, the accuracy of the detected edge and the geometric feature information that can be extracted from the edge, in-particular gradient orientation, can extend the displacement correction sensitivity to angular units.

In modern day object recognition systems based on deep machine learning and convolutional neural networks [85], [86] the edges are the first feature to be extracted using sets of orientated filters. These may include; first and second order differentiators or even the more complex Gabor filters [9], [22], [87] for measuring local spatial frequency and phase information. Whilst this feature set may be one of many that describe objects, generality of object class only extends to the contents of the training datasets used within the supervised learning paradigm of classification.

For well-defined applications with limited ranges of operation, knowledge of the object is known. This may include screening and tracking using medical imaging technology for pap-smear screening [88]–[90] and MRI (magnetic resonance imaging) of tumours [49], [91], [92]. In such precision critical applications, more accurate edge detectors can aide in the tracking of nonuniform shape and texture changes, tumour developments and consequent effects of swelling around healthy tissue, as well as, the automatic classification of tumour types and increases to the speed of diagnostics that provide support in medical decisions. In pap-smear tests the required human element to scan images for chromatin texture and cell irregularity is a time-consuming process [88]. With stringent laws surrounding working patterns for medical analysts the potential for mistaken diagnostics remains. Such that there is a critical need for well-defined and computationally efficient cytometric object detection, tracking and classification algorithms [89].

As one component of post processing steps after the collection of image data, edge detection at the pixel resolution is coarsely defined at either 90° or 45° intervals. Within the structure of images, both rigid and natural, the accuracy of the edge detector is driven by the required sensitivity and accuracy of the application. For non-rigid objects operating in nonstationary environments this especially the case.

2.3.4 Beyond the Pixel Resolution of Edge Detectors

Serving as the critical précis for chapter 4, this section inspects methods for achieving sub-pixel resolution in the detection of edges from digitally sampled images. Sub-pixel edge detectors are generally of three classes; curve fitting, moment based and reconstructive [93].

For curve fitting methods [94], [95], an estimation of a discretely sampled border is required to fit a curve. This method's accuracy is determined from the gradient orientation resolution of the edge detector; providing the prior knowledge of the edge points in the curve. Moment based approaches [96]–[98] calculate intensity and spatial moments in a pixel neighborhood representing a vector corresponding to the center pixel. The main disadvantage of this approach is that there is a moment vector for every pixel and all pixels defined as an edge can only be reliably associated to the edge in a small neighborhood [93]. Typically, this approach is considered for edge refinement. Reconstructive methods [99]–[101] generate a continuous gradient function obtained by interpolation of discretely sampled edge gradients. As with curve fitting methods, the discrete edge samples are obtained using either first order, second order or hybrid edge detectors. Usually, estimation of an edge within pixel error is found from the standard differentiator kernels, where the distance that the interpolator bridges is fixed by the pixel accuracy of the kernel.

It has been established through certain applications and operational situations, there is a requirement for an increased accuracy of edge gradients to detect edge features in irregular and deforming shapes. The primary deficiency, in the discussed methods, is that gradient accuracy is restricted by the resolution of the coarse differentiator kernel. Based upon estimating sub-pixel kernels formulated from the reduced complexity and accuracy of first order differentiation; the Sobel operator coefficients, kernel size and kernel arrangement formulate the motivations of the investigations in chapter 4.

2.4 Feature Representation Through Geometry

Relationships between localised pixel information within an image can be characterised by spatial locations and intensities with the aim of portraying similar or congruent information. In this section motivation is developed for connecting this type of visual data to an object's critical features. This is accomplished by utilising the detection of boundaries and correlation; as discussed in sections 2.1 and 2.2 in this chapter. Whilst a boundary conveys a holistic representation of an object, the object itself can be deconstructed into objects parts, each described by either a uniform or non-uniformly shaped boundary. The simplification of object parts into geometric structural components is atomistic and aids correlation in providing reliable analysis of image data. A brief overview of the Gestalt system [10] in chapter 1 introduces a system of simple 2D and 3D forms called geons [8]. Whereby, the forms of a geon establishes a set of simple structural shapes that are analogous to representing simple parts of an object. Whilst those geonic shapes will not be the topic for implementation and analysis in this thesis, particularly the investigations in chapter 5, this section presents linear planar geometry applied to the description of simplified object features. Following this, multivariate statistical data analysis techniques are discussed with reference to the notion of visual information projected onto holistic and atomistic subspaces. These topics direct this summary towards the Hough Transform [102], a signal processing technique that extracts linearly constrained object feature information confined to a point, line, plane and volume. Although, problems explored in this thesis are bounded within a plane only.

2.4.1 The Euclidean System, Features of a Line and Plane

Around two thousand three hundred and seventeen years ago the Greek mathematician Euclid of Alexandria wrote the mathematical text book series Elements [103]. The discourse in part consists of 13 books defining, postulating and proofing the mathematical system of plane and solid geometry. Known as the father of geometry, society commonly refers to his knowledge as Euclidean Geometry [104] and communicates it through algebraic language. Euclidean space is an interchangeable term that is also referred to as linear geometry.

A vector field $\Re$ containing a finite sequence of n ordered tuples in cartesian space is represented as $\Re^n = \Re \times \Re \times ... \Re$. A vector x consists of coordinates $x_1, x_2, ... x_n$ where n is the dimension of the space. In general, a field of real numbers comprising the vector are called scalers. These can also be represented as a point p in the vector field $\Re^n$. In this thesis the dimension of space is restricted to $\Re^2$, this is represented as a vector (x_1, y_1) or $p(x, y)$. Describing $\Re^2$ using $\Re^3$ is possible and may be useful under situations where a relative observation of a plane is required , for $p(x, y, z)$ this is achieved by setting z to zero. For a linear system, the distance d between two points in $\Re^n$ at an intersection between two lines is preserved, under Pythagorean law these metrics are: $d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$, $\theta = \arctan(1/2)$ and the Euclidean norm $d(x, y) = \|x - y\|$. For nonlinear geometries such as ellipses or hyperbolas, distances increase and decrease away from the intersecting point. The study of parallelism is called affine geometry [105] and the metrics describing nonlinear geometry extend from differential geometry. Applications of non-Euclidean geometry include the analysis of topological manifolds [106]. Whereby linear metrics, which are not restricted to the cartesian domain, estimate vector space relationships that are confined to points that lie at distances on the curvature of a manifold. Readers can learn about this topic from the introductory text by J. Lee [107].

2.4.2 Subspaces, Techniques and Applications

The last section concluded by discussion of non-Euclidean vector spaces, particularly the distortion of parallelism between lines due to exponential increases and decreases in distance between two lines. For an image displaying a planar object containing multiple planer object parts there exists an $\Re^2$ vector space of the image and a subspace $\Re^n$ pertaining to the image components. For clarity, a subspace in linear algebra is defined as a subspace of a higher dimensional vector space when the following conditions are met; the zero vector is in the subspace, the sum of two individual elements of the higher dimensional vector subspace is also an element of the subspace and, the product of a scaler value in higher dimensional space and an element of the vector subspace is an individual element of the subspace. In this definition and application, a planer object can either be represented as one holistic component or as many components which follow the atomistic relationship of the whole. This mathematical tool enables the representation of Biederman's recognition by parts [8] assuming there exists Euclidean plane geometry of the object parts. A challenge in the linear generalisations of problems at this level of statistical investigation, is to identify the interactions between linear components of the whole and the parts that are operated on by different scaler amounts. Thus, mirroring the inherent non-linearity a real non-rigid object would exhibit at any instance of time.

Explorative data analysis techniques can be used to analyse vector-spaces and subspaces. The discussion that follows looks at the purpose of these techniques to uncover underlying structural observations of linearly posed problems and how they evolve to help encompass and identify complex relationships. The primary technique discussed is a special case of projection pursuit analysis (PPA) [108], [109] known as principal component analysis (PCA) [110], [111]. To begin this discussion the necessary statistics are defined; standard deviation, variance and covariance. Standard deviation is a measure of how spread out data sets are, this means that the standard deviation (σ) is the rms distance from the mean (μ) of the dataset. Closely related is variance (σ^2), an equivalent statistical measure from the normal distribution relating the spread of data from its mean value. In 2D both the standard deviation and variance are 1D operators. To analyse the spread of data in multiple dimensions a separate calculation for each dimension is required. To provide information about the dispersion relationship between datasets the covariance between two variables is used. In matrix algebra the covariance matrix, eigenvectors and eigenvalues along with the second-order statistic covariance encompass the essence of PCA. Briefly, the covariance matrix stores the covariance entries between independent variables in a square matrix. The eigenvector known as a characteristic vector, is a linear transformation of a nonzero vector space by a scaler amount onto a vector-subspace. The eigenvalues, known as the characteristic values, are the scaler amounts in this linear transformation. Readers can follow the tutorial on principal components in [112], [113] for an introduction of these quantities and processes.

The purpose of PCA is to transform a large data set, containing potentially correlated variables, into a smaller quantity of variables called principal components.

In doing so, it allows analysts to identify patterns and trends, or not, in large data sets by using a projection of the larger dataset onto its principal components (smaller dataset). Of which, retaining the potentially correlated behaviour of the variables that exhibit maximum variance. This multivariate technique is important for linear regression, data compression and signal separation applications [114]–[117]. Perhaps the most commonly known application of PCA is in face recognition [118]–[120], for an up to-date discussion and detailed review on facial recognition capabilities readers can follow [121], [122]. From obtaining a dataset to conduct PCA, the procedure is summarised as follows: *1) subtract the mean across each dimension to reduce mean bias in the data set described by the first principal component; 2) calculate the covariance matrix; 3) calculate eigenvectors and eigenvalues of the covariance matrix; 4) organise or rank eigenvalues from highest to lowest (maximum variance); 5) form feature vector to retain eigenvectors explaining a percentage of the dataset.* The stark reality of data analysis is in the interpretation of the results. PCA is unsupervised in the sense that the model does not care and does not need any prior information about the input to perform the analysis. Therefore, it is non-parametric and in deciding whether to apply this technique, the primary assumption made in the algorithm's set-up must be considered. The primary assumption is that the dataset needs to be decorrelated from second-order statistics [44]. This is because each characteristic vector or new vector space must be orthogonal to present clear distinction (distortion mitigation) of the principal components. In literature a simple example of this is demonstrated by applying PCA to non-Gaussian data [112]. If maximum variance is the desired optimal response, and the dataset can be equally predicated by its most significant components, the true goal is error minimisation of the reduced dataset. When the dataset set is decorrelated there may exist higher-order terms in the dataset that have not been removed. This fundamentally restricts the application of PCA because the error cannot be reliably minimised without some knowledge of the dataset. Hence the problem becomes parametrically posed, which detracts from a key attraction of using PCA.

There is another class of explorative data analysis derivative of PPA which attempts to solve the problem of mixed signal statistics in decorrelated datasets. For mention only, this is known as independent component analysis (ICA) [123] and it is readily interpreted in the "*cocktail party problem*" [124]. ICA has also been widely demonstrated in image processing problems [125]–[127]. Other extensions to PCA to name a few include non-linear PCA (NLPCA) [128], [129] and non-negative matrix factorization (NNF) [130], [131]. NLPCA is implemented usually through neural networks to optimise unknown parameters describing the data distributions nonlinearities, and NNF restricts the eigen analysis of the data to only positive characteristic values thus eliminating any unintentional cancellations between eigen values. To complete this section and worth more than a mention is the general class method PPA, the critical question that the general class asks is "what statistics does the analyst want to observe, or what are the statistics to be observed" [132]. Like PCA, the optimisation of the problem or minimisation of the error is characterised by a projection index; an underlying linear / nonlinear statistical measure describing the dataset. It is just that in PCA, variance optimises the problem from a linear basis. In PPA the projection index is set to describe a dejection from the normal distribution using higher-order moments [133], [134]. Since the technique is non-parametric, a method to compare interesting projections must be implemented. In the case of PPA this results in an iterative procedure that is computationally heavy [135] for $\Re^n$

problems. It also presents a real difficulty for translating the nonlinear departure from the normal distribution to the characteristics of the dataset in the original vector space. Explorative projection pursuit analysis has interesting applications within the unsupervised paradigm of machine learning for probability density estimation [136] and regression analysis [137], [138].

2.4.3 Structural Geometry, Applied Hough Transform

The Hough Transform (HT) [24] is a geometric feature extractor, subject of investigation in chapter 5 the algorithm is used to identify the structural shape of object parts. Prior to an appraisal of this signal processing technique, the mathematical development from line to plane parameterisation is presented following the methodology in [76]. To begin, consider two colinear points with coordinates $p(x,y)$. The two points are related by the line equation $y = mx + c$, where the line gradient is m and the y-intercept is c. The homogeneous line equation is $Ay + Bx + 1 = 0$, where $A = -1/c$ and $B = m/c$. This means that a line can be represented by the coefficient pair value (A,B). For a second point the equation is symmetric and the homogeneous from of the line equation simultaneously yields two points and a line. The HT operates by voting for points defining the same line along $Ay + Bx + 1 = 0$. This means that for two colinear points indicating the start and end of a line, all votes along the same line intercept at the same point in the Homogenous space (A,B). In practise, the presence of many lines in an image creates many solutions of the form $Ay + Bx + 1 = 0$, hence the residual errors are minimised by the voting procedure and the solution is resolved by the peak of the HT. A cartesian representation is bounded by the image size or parameter space and the number of lines, hence the computational time rises with the product of these two parameters. The practical parameter space of the HT is the polar form [139] of the line equation, where a point in the parameter space is fixed by an angle normal to the line $Ay + Bx + 1 = 0$ by (ρ, θ) .

For a circular plane described by $(x - x_0) + (y - y_0) = r^2$, a set of all points (x, y) centred at the origin (x_0, y_0) for a fixed radii generates a locus of points (x, y). In the circular HTs (CHT) parameter space, circles are defined by the value of the radii centred on the coordinate of the edges tangential point (x, y) in image space. Since the radius is not fixed each point in image space represents a circle centred at (x, y) over a radial range. Hence the accumulator space is a 3D cone projection at each point. The parametric form of cartesian coordinates (x, y) on the circular plane is $(x_0 + r\cos\theta, y_0 + r\sin\theta)$. The centre of the circle can be extracted easily from the parametric form of the accumulator space over a defined radial range. The parametric representation of an ellipse can also be implemented but for a 4D accumulator space corresponding to the centre, major and minor axis of the ellipse. A generalised HT [140] exists for arbitrary shapes, however the invariance of complex shapes to translation, rotation and scale results in complimentary accumulators that absorb more memory, and is not considered in this thesis. Modifications to the HT and CHT to reduce the parameter space can be made by separating the parameters using the gradient information of $p(x, y)$ and reforming the line and plane equations. There is a rigorous treatment in [76] of the HT and parameter space reduction.

The discussion of the HT from herein focuses on object recognition applications. The analysis of digital images relies upon shape detection, the simplicity of the HT makes the technique an ideal candidate for the following reasons: each $p(x,y)$ is treated independently from the next, this provides benefits in hardware/software implementations for real time processing; examples include field programmable gate array (FPGA) [141] and neural networks [142]. The voting mechanism in parameter space provides degrees of mitigation against additive noise and partial occlusions. By detecting multiple candidate lines, geometric shapes can be formulated. The reliability and accuracy of geometric configurations can be enhanced by considering pre-processing steps to obtain pixel intensity, location and gradient information. Using prior knowledge found from the specific application such as eye localisation [143] in biometrics or object tracking in autonomous vehicles [144], the parameter space of the HT presents opportunities to deploy in scenarios requiring complex scene analysis [145]. The features of the shapes are most critical, but this does not necessarily mean they are irregular when they can be linearly generalised in the HT.

2.4.4 Knowledge of Objects, Accumulating Evidence

In this thesis a regression model to correlate object part feature data is required. Linear correlations between combinations of features can be sought by inspection. Assuming the object parts can be appropriately described using linear geometry, projection viewing angles can also be discretised to estimate metrics between observations. Like biological vision this is only accomplished in artificial vision using learned knowledge from the experience of the visual world. This chapter concludes with a discussion of data driven object recognition [146].

Data driven object recognition is having a growing impact on the discovery of geometric, structural and semantic relationships between shapes [21], [147]. The data driven approach between geometry and structure is heavily related to the activity, hence functionality of the shape. Prevailing methods in object recognition study objects under isolation by transfer of information from exemplar to target through correlation. Limitations identified in the review of data driven 3D shape analysis in [146] highlight these issues: generalisation across different datasets; complexity and scalability of the datasets incurring computational training time penalty; size of available 3D datasets where 2D datasets are used to infer 3D data; generation of training data for geometric shape processing tasks; uncovering useful patterns within models to uncover useful information of the problem being solved. These issues, regardless of the geometry being either a plane or solid, are synonymous with object recognition and supervised learning. One problem that the authors in [148] discuss is the prediction of regions of prominent objects in images. The authors simplify the problem to a single class of object where there is at least one in the captured scene. This is commonly known as foreground detection but without manually drawing bounding boxes to allow isolated study of the object. In [149] evidence accumulation inference is also used to determine regions of interest to utilise shape-based recognition. Whilst data driven shape analysis is also applied in segmentation for applications such as moving object detection [150], the target application in this thesis is classification. Leading into chapter 5, classification of the complex object class cat is investigated by collection of reliable object geometry data to infer structural relationships between the geometry and precedence of critical features.

2.5 Software Simulation Tools

As part of this chapter's review on image processing techniques, the development of algorithms, parameter analysis, conditioning and manipulation of digital signals in the subsequent chapters of this work are fulfilled using the MATLAB programming environment and AUTOCAD mesh analysis.

MATLAB is host to a vast library of functions appropriate to the issues investigated in each of the three chapters that follow. However, scripts are developed unique to the configuration and application of the proposed signal processing techniques. This results in nonstandard MATLAB scripts, which are identified at the most appropriate point in each following chapter. To summarise, there is a unique algorithm applied to: correlation signal analysis (section 3.6-7), edge feature detection (section 4.4-6) and features measurement and analysis (section 5.3-6).

Due to the complexity of generating large amounts of reliable and consistent training images of similar or same non-rigid objects under motion or deformation, a key step demonstrated in chapter 5 is to track a cat's head under motion of rotation; which is based on prior knowledge of geometric form. As part of generating suitable training data to investigate this step AUTOCAD is used to identify and measure reference points onto a single mesh object exemplar of a cat.

Chapter 3 - Feature Alignment and Image Registration

The Fourier transform is a linear signal processing subject that is particularly suited for the efficient processing of the correlation function. The discussions in chapter 2 section 2.2 identify the methods and limitations of correlation techniques for measuring the similarity of signals across a spectrum of applications, in particularly image registration. The main critique of correlation in practice is its deficiency outside planer objects, closed form (ideal) signal representation and performance against distortions. Limiting distortion to linear and angular displacements, this chapter focuses on the sensitivity of correlation interpolation in linear and non-linear units for the accurate detection of displacement measurements. The analysis of the developed correlation method is extended and brought into context using simple background subtraction, whereby, the critical objective for the pre-processing stage of an object recognition system is object segmentation. Wider applications in image processing tasks are discussed and interpreted at the end of this chapter. Leading on from the process of correlation in the previous chapter, registration issues for achieving successful segmentation are first considered.

3.1 Chapter Synopsis

Readers can refer to this synopsis of the research contained in chapter 3 when reading the investigations within this chapter.

This chapter describes the procedure used to condition the input image, the aim is to prepare the image so that features of an object's edge information can be reliably detected and measured. This is done by minimising the undesirable effects caused by various types of transformations imposed onto the image. The chapter begins with a discussion and study of an object under isolation and the image processing requirements to resolve rigid body transformations under ideal and non-ideal conditions. Non-ideal conditions include sources of additive noise in imaging systems and distortion due to translations, rotations and scaling **(section 3.2)**. Appropriate correlation techniques are investigated to identify the most suitable arrangement for the correction in order to minimise the effects caused by distortion arising from various transformations. With an aim to establish the suitability of the correlation technique for a precision analysis in support of a platform to recover more complex angular translations in the image registration process **(section 3.3)**. Following the description of the Fourier-Mellin method to recover angular and scale displacements **(section 3.4)**, sampling issues associated with the sensitivity of measuring a correlation peak are discussed **(section 3.5)**. An adapted correlation signal analysis algorithm for the precision study of angular misalignments is presented. The method adapts non-uniform sampling and a relaxed interpolation strategy to enhance the signal bandwidth and the sensitivity of the correlation **(section 3.6)**. The correlation analysis algorithm is studied with images containing natural and rigid structures to demonstrate and validate the approach. This is achieved using autocorrelation to identify signal components contributing high SNR; input to an adaptive correlation filter to identify and track angular translation systematic to the image **(section 3.7)**. The chapter ends with a characterisation of the algorithms operating range by demonstrating the effect of noise upon the available signal bandwidth **(section 3.8)**.

To begin, a pictorial summary of this chapter’s investigation is presented,

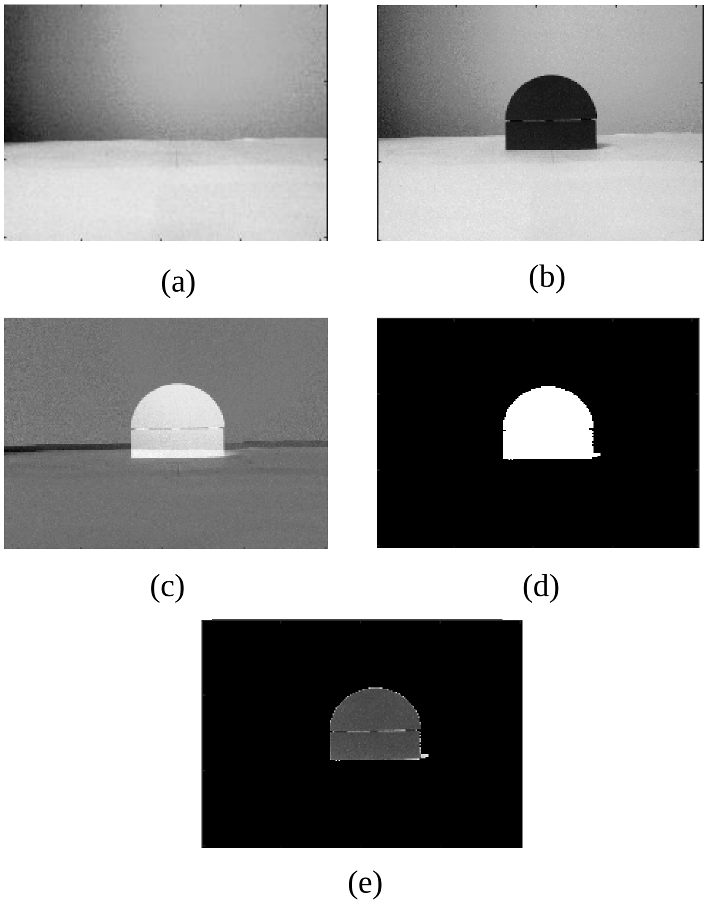


Figure 3.1: Example of image registration: a) background image, b) foreground image, c) subtracted image: b minus a, d) threshold mask image, e) detected foreground: product of d and b.

Under the ideal scenario presented in figure 3.1 the process to isolate the rigid object is simple due to homogeneous characteristics. In the case that either the background or foreground image has distorted by rigid body displacements, the alignment errors can perturb the image registration outcome. This chapter investigates the versatility of the correlation algorithm to mitigate such impacts for applications that follow this pre-processing routine to isolate an object of interest to inspect.

3.2 Object Isolation and Segmentation

The critical objective of the pre-processing stage in an object recognition system is to present the object suitably to the next processing step in the system; object feature identification and extraction. However, signal representation is not restricted to any one transformation. Most appropriately, the features of interest will dictate the selection of signal processing techniques that are able to acquire suitable signal representations for the most efficient extraction of its features. For example, this may include binary signatures; more on such topics in chapter 4. To illustrate the desired outcome of the segmentation routine, figure 3.2 presents an appropriate visual representation of a rigid object and a non-rigid object's data. There is no confusion from additional information that may have been surrounding each type of object that may impact the recognition process.

Figure 3.2: Representation of rigid and non-rigid object [151].

Importantly, it is easy to see that the buildings structure cannot change under normal situations except from the viewpoint and distance it is observed from, which for this object is generally restricted to scale, projective and affine geometry. The cats structure however can change under any given real situation. The first defining aspect to the object recognition system in task is the environment that it will be used in. Hence assumptions of background information, object types, object shapes and likely encounters of an objects form can be made through prior knowledge. The indoor world of visual experience is less complex and variable than its external counterpart. Therefore, by restricting the operating environment to household scenarios a database of information can be easily assimilated greatly simplifying the complexity of visual tasks. For example, a record of background images and knowledge of the objects of interest enables a system to have access to a reference and observation signal.

Using this framework, it is first possible to define a limited distortion invariant environment that enables suitable isolation and subsequent representation of an object. Presented in figure 3.3 is a process diagram showing the central discussion point for this chapter's investigations; image registration.

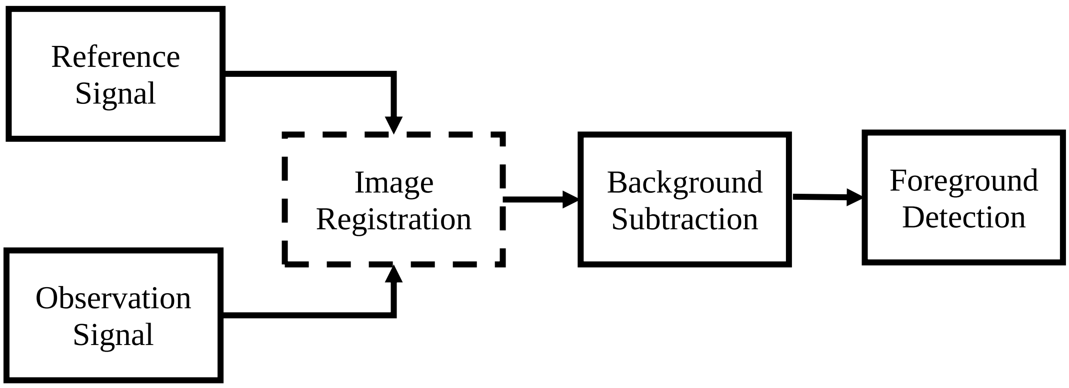


Figure 3.3: Adopted approach for achieving segmentation - highlighting the context of, and critical process of image registration.

3.3 Effects of Distortion in Image registration

For segmentation to succeed using simple background subtraction, the correlation method must be robust to the distortion that the environment or the acquisition system may present. Critically, to minimise any effects of distortions occurring in the processes of an object recognition system the signal of interest needs to be preserved. In this section, the sources of distortion identified in section 2.2 are given further grounding through example and discussion of their relative importance in the process of image registration and segmentation. These limitations include rigid body displacements (translation, rotation, scale and motion), intensity variation, occlusion and system level noise.

A noise contaminated image $I_{noise}(x, y)$ is composed of two parts: the true signal $I(x, y)$ and the noise signal $noise(x, y)$. In the case of non-additive noise, $noise(x, y)$ is any additional signal to the signal of interest that is component to the physical structure of the image content. This is occlusion and respective examples of blocking include: a cat in a tree, a cat curled up lying in foliage, or the tree branch or leaf, hiding the cat's body. Thus, increasing the complexity of the task to detect distinguishing features which may reduce the overall probability of successful recognition. This is a natural occurrence of human vision and one hurdle that computer vision systems must overcome. A sub-type of non-additive noise is the systematic elimination of object features or parts by occlusion caused by the segmentation algorithm itself. For the case of the background subtraction technique, intensity variation or similarity between foreground and background pixel intensity is one limiting factor and contributor of this type of error. Therefore, in the case of non-additive noise (I_{nan}) the noise contaminated image is

$$I_{nan}(x,y) = I_{fg}(x,y) + I_{bg}(x,y), \tag{3.1}$$

where I_{fg} and I_{bg} are the foreground and background components of the image. An additive noise source (I_{an}) is a time varying stochastic process R_{nd} , where R_{nd} denotes a random signal impacting the fundamental process of capturing the image. This could be of the environment or of the optical instrument receiving the signal information. For scientific imaging rigs, it would be usual to encounter a charge-coupled device (CCD) camera, whereas for everyday commercial use a complimentary metal-oxide-semiconductor (CMOS) camera system is typically employed. Each type of system has a unique formulation of R_{nd} . For now, consider the captured signal as the ideal noise free image and R_{nd} is statistically modelled as noise affecting the pixel nature of $I(x,y)$. Although a temporally stochastic signal I_{an} is an image component. Therefore, noise sources related to signal strength can be defined as

$$I_{an}(x,y) = I(x,y)R_{nd}(x,y). \tag{3.2}$$

Generally, sources of additive noise are considered as thermal processes (hardware), light particle interaction (photon shot noise) and natural image statistics independently affecting the capture of spatial distribution and light intensity. Therefore, the combined model of an image perturbed by potential noise sources is

$$I_{noise}(x,y) = I_{nan}(x,y) + I_{nan}(x,y)R_{nd}(x,y). \tag{3.3}$$

The noise level, hence, the amplitude of I_{an} and the similarity of I_{nan} are factors that influence the success of the correlation [51].

The average spatial intensity of foreground and background information of an image may be equal, similar or completely different. For example, the texture of animal fur in general appears to be random whereas a table is smooth. Such artefacts, that can belong to an object and its background, help to determine similarity because they are discriminate features. However, the colour of the worktop may also be similar to the animal's fur colour or even of a similar pattern. In this instance, the ability to discriminate between features is reduced because the variance of the pixel intensity across either the whole image (global) or local image region is closer to the observed mean value. For an object with a mean intensity value equal or close to the mean intensity of the background intensity no information exists for the correlation function to measure. Ideally information is required to be nonuniform and high variance. However, the environment must also be considered because visual information across one day causes any single object to appear differently. Illustrated in figure 3.4 are natural light scales with respect to human vision: day (photopic), night (scotopic) and twilight (mesopic).

Figure 3.4: Example of natural light boundaries, from left to right in grayscale: night, twilight, and day perspectives of an images content. Light scales are achieved using MATLAB functions to manipulate contrast.

The natural boundary is the amount of light interacting with the camera's sensor. This is called photon shot noise (PSN) and without high sensitivity instruments, discriminate features reflecting light are either poorly measured or not detected. This results in masked observation signals to perform correlation with.

Assuming the average pixel intensity between two images are equal except for some area of interest, the background subtraction method requires at the very least an observation signal to be aligned with a reference signal. Using a camera, images may be taken at varying positions (translations), planer angles (rotations) and distances (scales). Usually, these are global displacements. Since the problem is presented from static images, the processing methods considered in the following subsections are applied in 2D systems but can equally be extended to 3D. The transformation matrix that preserves lines and planes modified by translation (*T*), rotation (*R*) and scale (*S*) changes is called the similarity transform. Consider an image's global coordinate frame $I\left(X_1, Y_1, Z_1\right)$ that displaces to $I\left(X_2, Y_2, Z_2\right)$, the similarity transform, where θ is the amount of rotation in the line or plane, is

$$\begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \end{bmatrix} = \begin{bmatrix} S_X \cos\left(\theta\right) & -S_Y \sin\left(\theta\right) & T_X \\ S_X \sin\left(\theta\right) & S_Y \cos\left(\theta\right) & T_Y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix}. \tag{3.4}$$

A scale change across a line or plane is a linear process. For directional scaling $S_X \neq S_Y$, an object that has curve-linear features such as a circle would distort elliptically. Hence a line, plane and object, that are rotated and translated should be dealt with independently of directional scale changes. However, the matrix transformation in Eq. (3.4) can be used to model a cameras image that has undergone an isotropic scale change $S_X = S_Y > 1$. The Fourier-Mellin transformation [40] is a unique algorithm which treats rotation and scale as two independent 1D Fourier transforms, thus achieving lines, planes and objects invariant to both rotation and isotropic scaling. For both digital and optical implementations, scale changes are the main source of limitation for the application of the Fourier-Mellin transform. Whereby, small frequency scale distortions and uniformly sampled bandlimited signals of the polar grid contribute to a poor estimation of the original signals used to resolve correlation peak information. Of which, the imposed error simultaneously reduces the accuracy of the correlation to resolve rotational displacement.

3.4 Registering Translation

At the beginning of chapter 2 the essence of the correlation algorithm was presented. Here the technique is applied to register the translational displacement between two similar images: a background only image I_{bg} and foreground plus background image I_{nan}. The success of the correlation is based on the minimum Euclidean distance between a series of corrected observations and a reference signal. Whereby, the observation signal is divided into smaller image regions and a correlation is performed between each region and the reference signal to recover a series of translation measurements. Firstly, consider I_{bg} and I_{nan} as only differing by a translation. The resolution $M \times N$ of the observation signal can be subdivided into multiple image regions of resolution $M' \times N'$. For example, a square 1024×1024 pixel image can be divided into 64 sub-images each 128×128 pixels. Approaching the problem of aligning two similar images using sub-image regions generates two solutions; a refinement of the average or minimum correlation error and the identification of sub-images contributing useless information to the overall (global) correlation. Fundamental to this process is the method of correlation; for the purposes of this investigation the accuracy of the normalised and phase correlation method is restricted to the pixel level. To perform the correlation the observation signal is sub divided into image regions to obtain $g'(\cdot)$. The Fourier transform of each sub-image (Eq. (2.5)) is then taken to obtain $G'(\cdot)$. Whereby the correlation for each sub-image is obtained by correlating with the average reference image $\bar{h}(\cdot)$ using:

$$C'_n(u,v) = \frac{\bar{H}(u,v)\bar{G}'^*(u,v)}{\left\langle \left\|\bar{H}(u,v)\right\|, \left\|\bar{G}'(u,v)\right\| \right\rangle}, \tag{3.5}$$

and the phase correlation by

$$C'_\omega(u,v) = \frac{\bar{H}(u,v)\bar{G}'^*(u,v)}{\left|\bar{H}(u,v)\bar{G}'(u,v)\right|}. \tag{3.6}$$

The inverse Fourier transform of $C'_\omega(u,v)$ or $C'_n(u,v)$ reveals a correlation peak at the location of displacement $\delta(x \pm \Delta x, y \pm \Delta y)$. This provides a series of translation vectors $\vec{T}_X$ and $\vec{T}_Y$, where sub-images with the most similarity are either equal or near to a certain value. This will indicate a consensus for the correct displacement that has occurred between two images. In disregarding the outliers of the translation vector's an average displacement vector can be calculated using a directional translation threshold value, whereby values above or below a cut-off are also disregarded.

However, knowing what threshold value to apply requires a knowledge of the translation vector. To retain simplicity of this process, the original observation signal is corrected by the measured displacement from each image region's correlation. The Euclidean distance between each corrected observation signal $g_c(\cdot)$ and reference signal $h(\cdot)$ is obtained and the minimum error imposed by the translation is retained. Where i and j are the indices of the image coordinates, the error vector is obtained from

$$euclideanD = \frac{\sum_{i=0}^{M'-1}\sum_{j=0}^{N'-1}\left(g_c\left(x_i, y_j\right) - h\left(x_i, y_j\right)\right)^2}{M'N'}. \tag{3.7}$$

Issues of this approach that are considered for analysis are: if the two signals at each sub-image are not similar then the correlation will yield inconsistent error and translation vectors. For smaller sub-image regions, hence an increased number of sub-images, the number of correlations and evaluations to calculate will proportionally increase thus affecting speed of operation; if the total process time is still insignificant. In the following analysis of phase correlation (Eq. (3.6)) to recover translation displacements and a region of interest. The impact of an additional object to the correlation strength between a reference and test image is described The test image $g_c(\cdot)$ is the sub-image to perform correlation with a reference image $h(\cdot)$ that is the background only of the same image. The images are two separate images taken in quick succession in the day. The initial square region selected is 1024×1024 and the number of sub-image regions used in the analysis are 1, 4, 16, 64, 256, 1024 and 4096. The average correlation strength peaks are observed in table 3.1 for both the autocorrelation of $h(\cdot)$, $\bar{A}_{c(pk)}$, and the cross correlation $\bar{C}_{c(pk)}$ of each sub-image region $g_c(\cdot)$. The recovered translation is identified as the sub-image region with the lowest mean squared error to the shifted version of $h(\cdot)$, in the example used $T_X = -15$ and $T_Y = -1$, where the sign indicates the direction of translation. Importantly, the impact of the additional object is severe on the peak correlation strength and the point at which the sub-image regions become too small and yield an unreliable correlation signal is independent of this peak value.

Table 3.1: Sub-image region comparison: auto and cross-correlation strength for recovering regions of interest.

Number of Image Regions	$\bar{A}_{c(pk)}$	$\bar{C}_{c(pk)}$
1	1	0.0972
4	0.3846	0.0832
16	0.1665	0.0424
64	0.0780	0.0291
256	0.0418	0.0284
1024	0.0366	0.0329
4096	0.0385	0.0376

For the cross correlation the sub-image region returns poor estimation of the translational shift because $\bar{C}_{c(pk)}$ begins to increase for an 8×8 image region size. This is replicated in $\bar{A}_{c(pk)}$ where no translation shift occurs, except that the reversal point of the sub-region image size is 16×16. In figure 3.5, the regions of interest identified by the image regions listed (except the first) that are not shifted by the correct amount are presented. Critically, the similarity of background pixel intensity to object of interest pixel intensity places additional limitations to recovering the optimal sub-image region size. This procedure should be treated on a case by case basis.

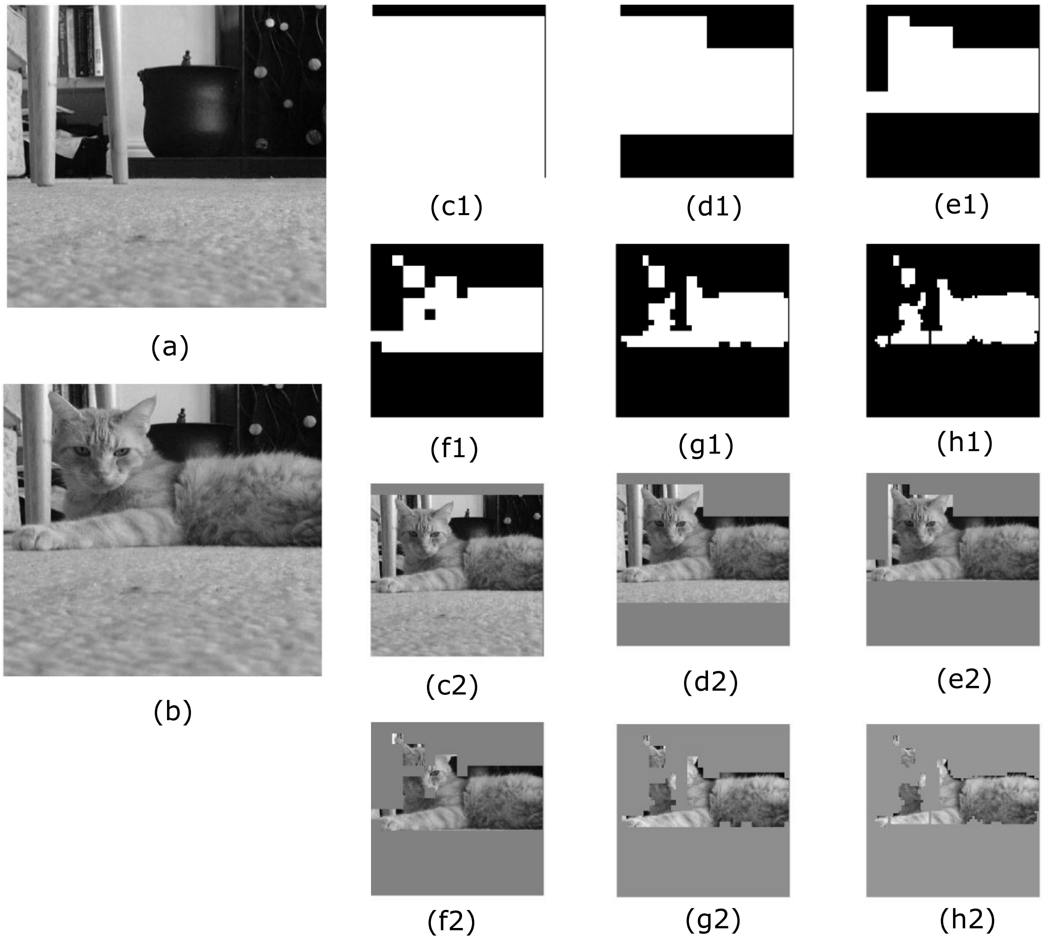
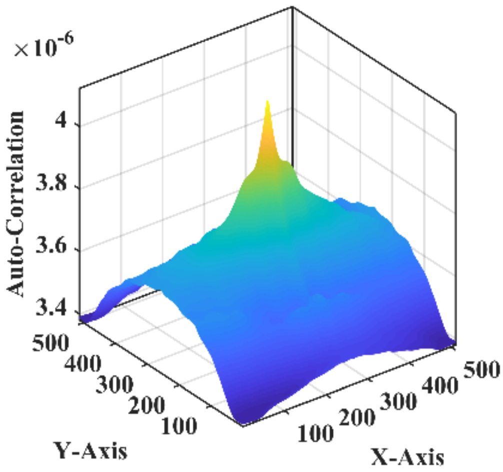


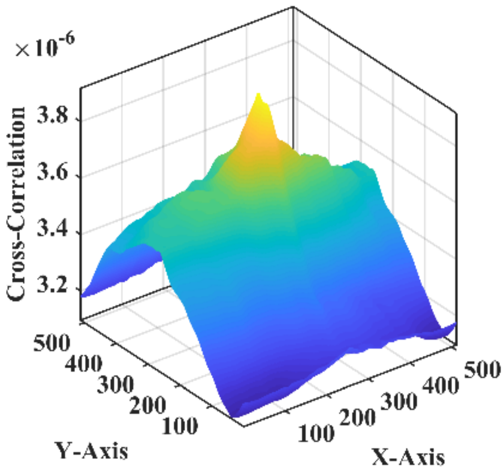
Figure 3.5: Regions of interest in sub-image region correlation. a) Reference image $h(\cdot)$, b) test image $g_c(\cdot)$. Image region mask (c1-h1) and detected region of interest (c2-h2): number of regions c) 4, d) 16, e) 64, f) 256, g) 1024 and h) 4096

The penultimate investigation of this chapter is an in-depth analysis of recovering angular displacement. The momentum for this is established from firstly, an analysis of the correlation accuracy to recover linear displacements (T) in the ideal noise free scenario. Secondly, an analysis of correlation precision against a simplified noise model of an imaging system.

3.4.1 Ideal, Noise Free Correlation

In the case of analytics, the phase correlation method is not an appropriate solution to investigate sensitivity of the correlation measurement. The main limitation to analysis using the phase correlation method is that the location of the correlation is restricted by the initial resolution of the shifted phase delta function. Secondarily, useful information that may be potentially related to the structural formation of the correlation is certainly lost. Hence, phase correlation is not in itself a discriminative feature outside of the phase displacement measurement. This is interesting because in the simplest of cases geometric shapes present unique autocorrelation surfaces, which distort due to linear and non-linear transformations. However, the complexity and dissimilarity of the functions being correlated can inhibit such an application. Of which, the accuracy of the correlation measurement must be equally considered because correctly sampled data should contain all the information of the original signal. The normalised correlation using Eq. (3.5) is demonstrated in figure 3.6 on two images under photopic conditions. Using two identical images except one with a feature removed two observations can be made, the correlation functions shape and the additional visual distortion of shape due to dissimilarity due to linear translation.





(a) (b)

(c) (d)

Figure 3.6: Normalised Correlation; a is resulted from the autocorrelation of c, and b resulted from the correlation of c and d.

Consider that the difference between the two images is any translation present from the original acquisition and presence of the cat occupying less than half of the image space. In the case of autocorrelation, the correlation peak shape is predominantly sharper (increased higher spatial frequency content) than when the cat is present. This illustrates the behaviour of spatial information and that a real signals composition is unique. Based on the images used in figure 3.6c,d, figure 3.7 demonstrates the process by which the accurate recovery of linear displacements between two signals is obtained after imposing additional translations.

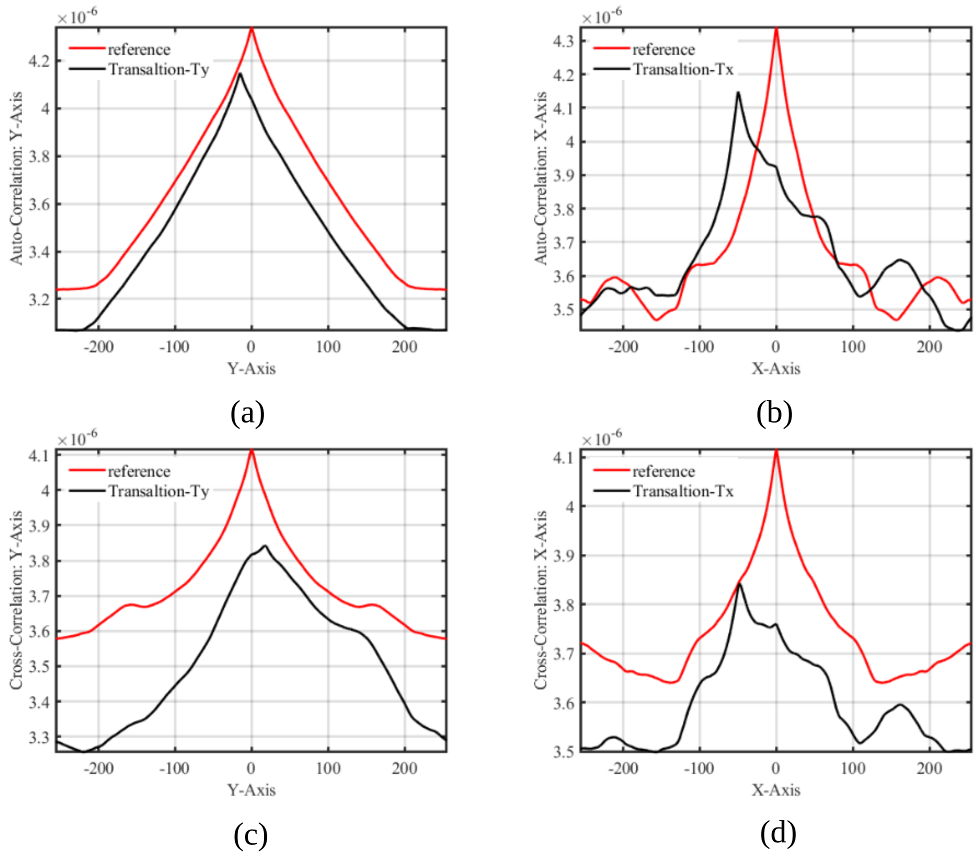


Figure 3.7: Linear displacements in correlation: $T_X = -50$, $T_Y = -15$. a) and b) envelopes of the x and y-axis for the autocorrelation of the reference image: figure 3.6c. c) and d) are the envelopes for the cross-correlation between the test and reference image: figure 3.6c-d.

When the content of functions $G(\cdot)$ and $H(\cdot)$ are not identical and undergo a linear translation the correlation function's shape distorts. A second consideration is the significance of wrap around in bandlimited signals when displacements are artificially imposed. Measuring the strength of correlation is normally addressed using the ratio in the correlation plane to the correlation peak or the fullwidth at half maximum. The correlation strength (peak value) in figure 3.7 reduces when both functions are dissimilar, but this is not necessarily the critical aspect to the measure of a correlation because regions between two images can be similar.

The critical component in correlation analysis is the peak location. Derived in the same way as Eq. (3.3) denoted as $C_{noise}(u,v)$, four terms are obtained that describe the whole correlation signal: the correlation of the true signal $G(\cdot)H^*(\cdot)$, the correlation of independent noise sources $N_G(\cdot)N_H^*(\cdot)$ and the correlation of the reference and observation signal $N_G(\cdot)H^*(\cdot)$ and $G(\cdot)N_H^*(\cdot)$.

3.4.2 Additive Noise in Correlation

The noise model used here is a simplification of a typical imaging system module, whereby the signal that is captured is indirectly degraded due to the sensitivity of the camera module and directly degraded due to the environment the image is taken from. Therefore, the sensitivity of correlation measurement is based on; white noise including photon shot noise and camera thermal noise; low frequency noise whose magnitude decreases as frequency increases. Detector flicker noise and environmental noise are typical examples of low frequency noise. For an overview of this noise model readers can turn to appendix A1. The correlation is performed over the segmented image, therefore the correlated noise patterns $N_G(\cdot)N_H^*(\cdot)$ have the least effect to the correlation signal. That is unless the background noise signal is close or greater than the strength of the signal, in which case the signal would be masked by noise. The components that are of critical interest are the noisy signal components $N_G(\cdot)H^*(\cdot)$ and $G(\cdot)N_H^*(\cdot)$ which directly affect the true correlation. To investigate the effect of random noise on the correlation, the model below is used. To begin with, the auto correlation of a 2D gaussian is perturbed with a uniformly distributed noise pattern. A gaussian signal is chosen due to symmetry and prior knowledge of the function. The

SNR increases from 1 through to 20000 photons per pixel by $\sqrt{N_{pht}}$. Secondly, a low frequency noise pattern attenuating at 3dB is normalised to the average signal level. At percentage levels of the signal's amplitude, noise levels of the correlation at 1 through to 150% are compared. For both phase and normalised correlation, figure 3.8 describes how each correlation method interacts with both noise patterns.

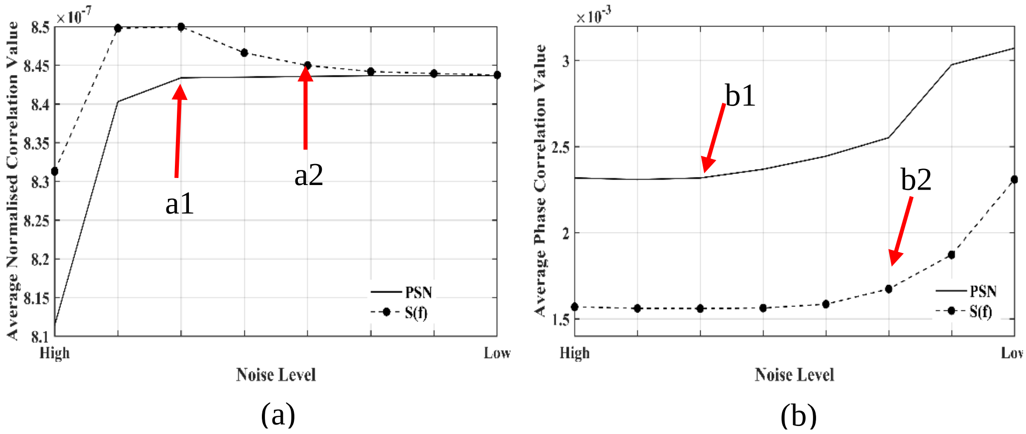


Figure 3.8: Average value of normalised correlation a), and phase correlation b). Photon shot noise (PSN) $\sqrt{N_{pht}} = \{1, 3.16, 10, 27.4, 31.6, 70.7, 100, 141.4\}$, and low frequency noise $S(f)_{\%} = \{150, 100, 50, 20, 10, 4, 2, 1\}$.

For normalised correlation the average correlation value includes the energy in the peak and whole plane. It is observed in figure 3.8 that low frequency noise has the greatest impact to normalised correlation and photon shot noise being the opposing constraint in phase correlation. This is visually plausible because there is low frequency shape in the correlation of normalised correlation and a single high frequency delta function for phase correlation. It is apparent that in phase correlation, the correlation signal only begins to increase beyond the amplitude of low frequency noise for amplitudes less than 10% of the signal amplitude (figure 3.8b-arrow b2). In the case of photon shot noise this limit is reached for a SNR of 10 (figure 3.8b-arrow b1). In normalised correlation these limits are paralleled: figure 3.8a-arrow a1-2.

To recover linear translations between known signals appropriate filtering schemes are applied to preserve true signal information and enhance the accuracy of the correlation measurement. Through knowledge of either the signal or noise characteristics perhaps the most important methods are matched and bandpass filtering. Angular (nonlinear) displacements can be resolved as a linear translation but the accuracy of the measurement ultimately also depends on the quality of the signal and the accuracy of the correlation method. To suitably apply filtering schemes a one size fits all approach may not be appropriate to the application, such as band pass filtering for the recovery of translations in temporal environments. Beginning with a traditional approach of implementing the discrete Fourier-Mellin transform this chapter's focus shifts onto the sensitivity and accuracy of the correlation process to recover linearly approximated nonlinear displacements.

3.5 Discrete Fourier-Mellin Transform

For the remainder of this chapter the discrete Fourier-Mellin transform is referred to as the DFMT. In the strictest sense the Fourier-Mellin transform is composed of a 1D polar Fourier transform to recover angular displacements and a 1D Mellin transform to recover uniform scale displacements. The issues associated with recovering scale simultaneously with rotation were outlined in section 2.2. This includes the distortion of the polar frequency domain due to scale changes in the spatial domain and the effects of sampling non-linear measurement units, θ. To reinforce that commentary the complete transform is first discussed to identify how the transform is implemented in practise.

First consider two functions where one is a rotated and translated replica of the other

$$h(x,y) = g\left(x\cos\theta_0 + y\sin\theta_0 - x_0, -x\sin\theta_0 + y\cos\theta_0 - y_0\right).$$

Through the Fourier translation and rotation property [32] this transforms to

$$H(u,v) = G(u\cos\theta + v\sin\theta, -u\sin\theta + v\cos\theta)e^{-j2\pi(ux_0+vy_0)},$$

where all points on the frequency grid are first rotated and then shifted in phase by the translation. The magnitude of $H(\cdot)$ is invariant to the phase displacement hence the function is only a rotated replica of the other. For small translations this is approximately the case because larger translations generate a greater wrap around, thus distorting the spectrum of one function. Where $\hat{u} = u\cos\theta + v\sin\theta$ and $\hat{v} = u\sin\theta + v\cos\theta$, the polar coordinate conversion of $\hat{u}$ and $\hat{v}$ is

$$H(r,\theta) = G(r,\theta \pm \theta_0). \tag{3.8}$$

For now, $|H(\cdot)|$ and $|G(\cdot)|$ are sampled uniformly on a polar grid. This means coordinates radiating from the centre are under sampled and oversampled at the centre. The opposite is also true: under sample toward the centre. Consider $\theta_{i(0:2\pi)}$ at r_j; these are 1D polar frequency coordinate values, recast back from a polar grid on to a cartesian grid. Hence, a rotation about the origin corresponds to a linear horizontal translation. Increasing the radial or the angular increment changes the grid space of each annular region. This adjusts the sampling rate and therefore a degree of freedom to manipulate nonlinear grid spacings. Applying a second Fourier transform to Eq. (3.8) generates $H(\omega_r,\omega_\theta)$ and $G(\omega_r,\omega_\theta)$, there exists a phase relationship that is deduced by applying the correlation function revealing its peak location at $\delta(r,\theta\pm\Delta\theta)$. This process is simplified by treating the correlation as a 1D transform in $F_\theta(r)$. The spatial frequency coordinates at where the signal amplitude is greater than the noise amplitude may not be well isolated to either the lower, mid or higher spatial frequencies. In the case of accurate and precise correlation this highlights the necessity for a matched filtering, akin to an adaptive filter, over the adaptation of a bandpass or a predefined class of filter to optimise the characteristics of the function.

The radius is not separable from θ in its current form. Consider the Fourier transform of two functions where one is a uniformly scaled (S_X, S_Y) replica of the other

$$H(u,v)=\frac{1}{S_X+S_Y}G\left(\frac{u}{S_X}+\frac{v}{S_Y}\right).$$

In polar coordinates θ is not affected by uniform scale changes because the orientation calculation cancels out the scaling factor. For the radius however, this is not the case;

$$r=\sqrt{\left(\frac{u^2}{S_X}+\frac{v^2}{S_Y}\right)},\theta=\arctan\left(\frac{vS_X}{uS_Y}\right).$$

Retaining the radial component yields a 1D transform in $F_\theta(r)$. For two functions, where one is a scaled replica of the other the shifting property identifies an equivalent transform,

$$|H(r)| = G\left(\frac{r}{S}\right). \tag{3.9}$$

A logarithmic conversion (usually $\log_e$) of $F_r(\theta)$ is equivalent to $F_\rho(\theta)$, where $\rho = \log_e(r)$. From this conversion Eq. (3.9) rearranges the inversely proportional scale constant to a logarithmic subtraction, $\rho - S$. As with the 1D treatment of recovering angular displacement the application of a second Fourier transform that is input to a correlation function reveals a vertical linear scale displacement at $\delta(\rho \pm \Delta\rho)$. Combining these two 1D transforms describes a function invariant to both rotation and scale.

This function is called the Fourier-Mellin transform [152] because of the unique property of the Mellin transform ($\dot{m}$), scale invariance. The definition of the Mellin transform or Mellin correlator [152] is

$$\dot{m}(s) = \int_0^\infty f(r) r^{s-1} dr, \tag{3.10}$$

where s is a complex variable. If s is restricted to the imaginary axis and $r = e^\rho$ Eq. (3.10) becomes

$$\dot{m}(j\omega) = \int_{0}^{\infty} f\left(e^{-\rho}\right) e^{-j\omega\rho} d\rho \,. \tag{3.11}$$

This is equivalent to a Fourier transform of a logarithmically scaled input. For the appropriate discrete implementation of this transform to recover rotations and scale changes these issues must be considered; nonlinear interpolation between coordinate transformation; sampling (1:1 - under-sampling at higher spatial frequencies); aliasing of low and high frequency components; scale displacements adding distortion; correlating both signal and noise.

3.6 Nonlinear Domain Transformation Error

Even though the discrete capture of a spatial signal is bandlimited (zero outside of the signal bandwidth - pixels/cycle), aliasing may still be present as compared to the original sample. Obeying Nyquist's theorem [65], an $M \times N$ image is sampled at 2 pixels/cycle. The spatial frequency coordinate map is thus $\pm M/2,\ \pm N/2$. For a signal that is under-sampled the exact discrete reconstruction of the signal is not possible due to aliasing. This being that the Fourier transform of a signal in time decomposes into a pulse train of signals that overlap and sum to compose the complete signature of the signal. Whereby, the degree of overlap is controlled by the sampling frequency. For an oversampled signal there is a direct waste in computational time as the sampling rate is satisfied. The interpretation of this process is that the image along either the x_i or y_j axis is a rect function bandlimited over the region *i=0:M-1 and j=0:N-1.* As presented in section 2.1.3, this is equivalent to an ideal 2D lowpass filter or a 2D sinc function. Therefore, any error within a signal obeying Nyquist's theorem is contained within the discrete representation of the sinc function. This concludes that errors exist due to the interpolation between the ideal reconstruction of the signal information at the sampling point locations in the image and the by-product that aliasing cannot be completely eradicated in a bandlimited signal.

For the recovery of angular translations, sources of error imposed by the polar Fourier correlation process are the nonlinear coordinate conversion and the casting of nonlinearly sampled units back to a linear grid. This leads the core investigation of this chapter into discussions of the initial signal quality and its effects on sampling nonlinear units suitable for a correlator.

3.6.1 Sampling Error

Uniform sampling is the most common approach in implementing the DFMT, of which figure 3.9 demonstrates two critical issues in sampling from a cartesian $F_c(x, y)$ to a polar domain $F_\rho(r, \theta)$. Firstly, at coordinate locations along the radial components at 0° and 90° in the polar grid the grid is exact except for any direction in-between $\pi/2$ intervals. Secondly, polar coordinates can be under-sampled towards the or away from the centre depending on the sampling arrangement. This has adverse effects in the overall DFMT algorithm as well as the accuracy in any correlation analysis because the image signal content is assumed to be within the lowest portion of the signal bandwidth whereas critical information to enhance the correlation may not be.

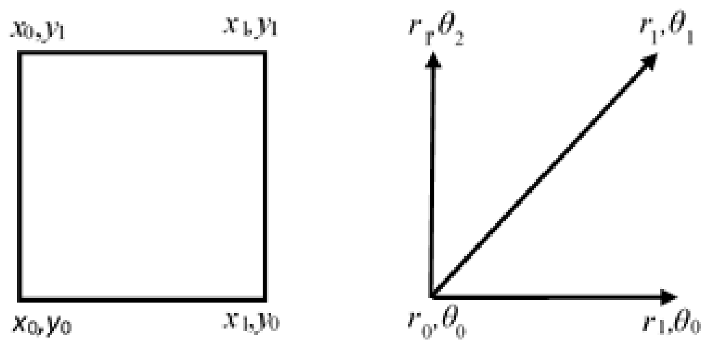


Figure 3.9: Uniform sampling of nonlinear units.

Assuming the sampling interval of a digital image is $\Delta x = \Delta y = 1$, to achieve a constant polar gradient sampling interval ($\Delta\theta$) along the radial direction a nonlinear sampling basis must be established. For an arc length $r \cdot \theta$ each radial component of the signal in Eq. (3.8) expressed as $F_\theta\left(r\right)$ must be $\le \Delta x$ if the sampling condition is met. The result of this is a linearly increasing sampling rate (s_{rate}) for increasing *r*, where the following condition is met

$$\Delta\theta_{r(1)} = \frac{2\pi r_1}{s_{rate}} \le 1, \Delta\theta_{r(\max)} = \frac{2\pi r_{\max}}{s_{rate} r_{\max}} \le 1 . \tag{3.12}$$

Consider an image 256×256 with a radial range of 1 to 128 in increments (Δr) of 1. The polar grid would sample the first radial ring over $\theta_{(0:2\pi)}$ using 8 samples and the last radial ring with 1024 samples with a constant $\Delta\theta$ equal to 0.785. For increasing image sizes this approach to re-sample nonlinear units becomes computationally expensive and reduces to a series of 1D correlations for the recovery of a rotational displacement. However, the full bandwidth of the signal is available to use or analyse with a confidence that the initial sampling error is mitigated. For grid locations not discretely identified i.e. those in-between $\pi/2$ intervals, the bandlimited signal remains estimated.

3.6.2 Interpolating Resampled Points

In this section, the ideal sinc interpolator and approximations to using polynomials is discussed. Typically, when evaluating non-discretely defined points on a polar grid for such algorithms as the DFMT either a nearest neighbour, bilinear or bicubic interpolator is used. Except for the nearest neighbour the accuracy increases using more samples within a neighbourhood to evaluate the central data point. For a bilinear interpolation the polynomial along one dimension is evaluated from one additional sample and for bicubic three additional samples. Hence, the interpolation accuracy in 2D is proportional to a four and eight-point neighbourhood as well as the nature of the surface intensity itself; where large differences between sample points may introduce unwanted bias into the weights of the polynomial coefficients.

To use a uniformly sampled algorithm to recover rotational displacement the correlation is a straightforward process obtained from matrices containing polar Fourier spatial frequency coefficients. There is an increase in accuracy through interpolation of the sampled data, however, the useful signal to perform correlation with must be assumed located in the lowest portion of the polar spatial frequency bandwidth. For oversampling from the centre, mid to high image spatial frequency information is distorted further from what its value is in an already under-sampled polar grid.

Usually the nearest neighbour interpolator is not of enough accuracy because no neighbourhood in the coordinate grid is considered, but in the case of nonuniformly sampled polar grid information its accuracy can be increased, and its ease of implementation exploited. To begin, consider that the more samples used to discretise a signal reduces the mean squared error between its true continuous representation. By zero padding a signal an alternative approach to implementing a sinc interpolator is accomplished. Here the sampling interval $1/\Delta x \Delta y$ in the spatial frequency domain requires reducing, hence the spatial interval must increase. Obviously for a nonuniformly sampled grid this may incur unrealistic processing requirements, but it is not necessary as this approach is used to generate just one resampled (finer) spatial frequency grid used as a look up table for replacing non-integer polar grid locations of the signals. Likewise, the accuracy using a bilinear or bicubic approximation can also be increased following this same approach but for the fact that computational expense also proportionally increases; along with the maximum radius of the input image and the complexity of the interpolator and neighbourhood size. As the ideal sinc interpolator is based on a signal of infinite extent, the maximum number of convolutions performed over a nonuniformly sampled grid is limited to the geometric sum of the resampling rate s_{rate} up to the maximum radial component (r_j) of the image boundary: $\sum_{j=0}^{r-1} r_j s_{rate}$.

A final consideration related to the rect form of an image boundary is spectral leakage. Whereby the rect window that is used to demonstrate and develop theory shows that a window function isolates a portion of a signal's frequency spectrum. Windows in practice are implemented to perform a short time Fourier transform, of which the signal captured is required to be periodic. For any real signal however, this condition may simply not be the case. Such artefacts are assumed in the FFT because a signal is assumed to repeat for all time before and all time after the bandlimited response, i.e. the sinc function exists on an infinite boundary. For aliasing, short time transitions can also occur at the beginning and end of each pulse response of the Fourier transformed signal if the signal is not periodic. Hence the frequency response of this artefact extends over a broad and high range. This means that the energy in the ideal signal loses amplitude and dissipates power into a range of broader frequencies decreasing from the ideal frequency. A window function reduces this leakage by smoothing sharp transitions at the beginning and end of the signal boundary, hence low pass filtering the signal. Ideally, windows are bell shaped and are required to start and end in zero and have a vertex amplitude of 1, such examples particularly suited to image processing include the Harris, Hann and Blackman window [153].

3.7 Correlation Signal Analysis Algorithm

Based on the discussions made on the sources of nonlinear domain transformation error, a correlation algorithm to measure the sensitivity and accuracy of angular displacement is presented; algorithm development is uniquely implemented in MATLAB. To isolate the rotation measurement the reference $h(\cdot)$ and test $g(\cdot)$ signals are exactly equal except for an imposed rotation. The centre of rotation is assumed to be at the centre of the image.

Firstly, a square region within a larger image is extracted from $h(x,y)$ and the signal region is then windowed using the Hanning function $w_h(x,y)$ [153]. The windowed signal $h_w(x, y)$ is then zero padded ensuring no spatial information is lost when an in-plane rotation is imposed. Since the image region is square ($M==N$) the pad size of the horizontal and vertical direction is $(M - diag)/2$. The padded signal $H_{wp}(u,v)$ is then zero padded again by an even sampling factor $s_f = 1,2,4,8...$thus, achieving a decrease in $\Delta u, \Delta v$. The Fourier transform of the up-sampled version of $h_{wp}(x, y)$ is replicated to form the test signal $G_{wp}\left(u,v;\theta_{(0:\pi)}\right)$. In the context of the correlation signal analysis algorithm an up-sampled signal is a resampled signal at a higher rate. Both $H_{wp}(u,v)$ and $G_{wp}\left(u,v;\theta_{(0:\pi)}\right)$ are transformed into the polar coordinate domain following the nonuniform sampling condition of Eq. (3.12) and the interpolation strategy as discussed in section 3.5.2; of which, each boundary of $F_\theta(r)$ is resampled by

$$r = \frac{0:(v/2)-1}{s_f}. \tag{3.13}$$

Hence, each location over the original signal boundary is replaced by the up-sampled coordinate value of $H_{wp}(u,v)$ and $G_{wp}\left(u,v;\theta_{(0:\pi)}\right)$. A series of 1D correlations that are calculated using Eq. (3.5) reveal the measured rotation θ_r at each $F_\theta\left(r_{0:(x/2)-1}\right)$. The relative error ($\varepsilon_r$) is then calculated from knowledge of the imposed range of θ on $g_{wp}\left(x, y;\theta_{(0:\pi)}\right)$.

The statistical figures of merit used to analyse this error are σ_θ, standard deviation, and $\bar{\sigma}_\phi$, average standard deviation. $\bar{\sigma}_\phi$ is obtained by imposing a second rotation to $h_{wp}\left(x, y; \phi_{(0:\pi)}\right)$ over the defined interval $\Delta\phi$. Where *n* is equal to the maximum rotation angle, the analysis of the measurements accuracy and precision is determined from these sets of equations:

$$\varepsilon_r = \theta_t - \theta_r , \tag{3.14}$$

$$\sigma_\theta = \sqrt{\frac{\sum_{i=0}^{n}\left(\varepsilon_{r(i)} - \bar{\varepsilon}_r\right)^2}{n+1}} , \tag{3.15}$$

$$\bar{\sigma}_\phi = \frac{\Delta\phi \sum_{i=0}^{n/\Delta\phi} \sigma_{\theta(i)}}{n+1} . \tag{3.16}$$

In figure 3.10 a function diagram of the correlation signal analysis algorithm is presented to accompany the observations of the algorithm's key steps and initial statistical analysis as presented in figure 3.11. Firstly, the type of input signal as used in figure 3.11a is a natural environment containing non-rigid structure. This forms a critical interpretation of the analysis because for a foreground detection algorithm, knowledge of the environment helps to build up a knowledge base. This can be used to either decide if the background subtraction method as described at the beginning of this chapter will have any success or potentially; identify a range that critical components of a signal's bandwidth occupy.

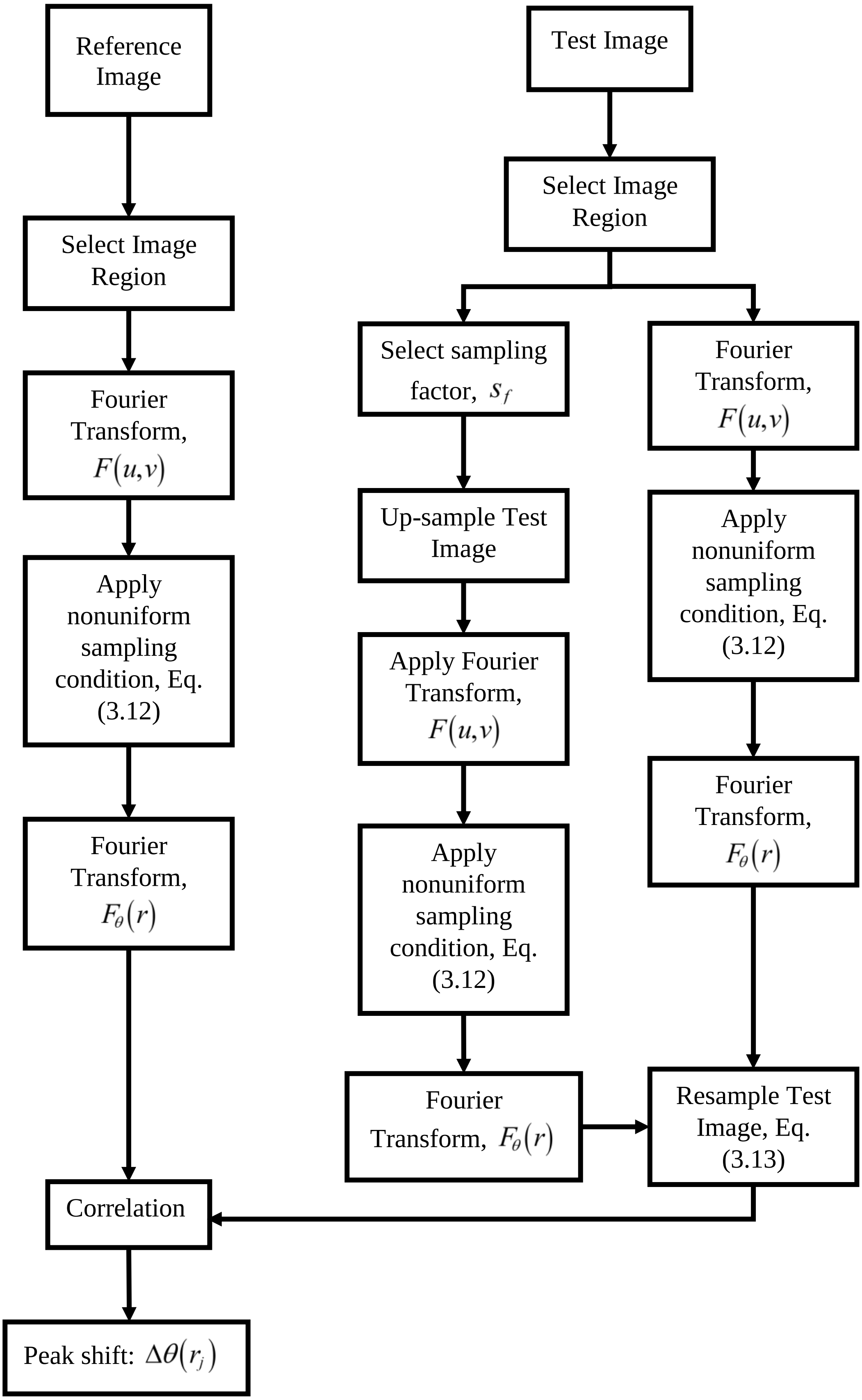


Figure 3.10: Correlation signal algorithm functional diagram.

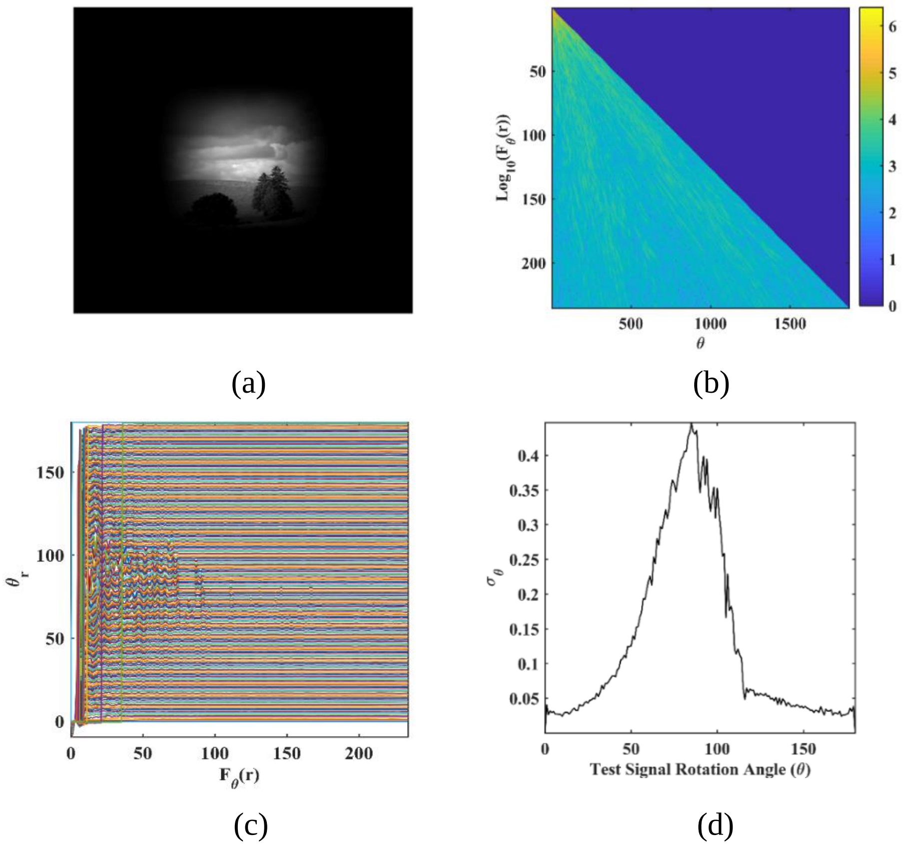


Figure 3.11: Key processes of correlation signal analysis. a) Identical input signal (470×470) - $h_{wp}\left(x, y; \phi_{(0:\pi)}\right)$ and $g_{wp}\left(x, y; \theta_{(0:\pi)}\right)$; set at: $\phi = 0^{\circ}$, $\theta = 0^{\circ}:180^{\circ}$, $s_f = 1$. b) Nonuniform polar Fourier amplitude spectrum of $H_{wp}\left(u, v; \phi_{(0:\pi)}\right)$; Log plot, constant sampling interval - $\Delta\theta = 0.785^{\circ}$, nearest neighbour interpolation. c) Recovered rotation angle measurements θ_r. d) Standard deviation of relative error measurement σ_θ, taken over the radial range $r_{(50:235)}$. Input signal taken from landscape database in ref [151]

The second observation is the general response of the recovered rotation angles over the radial range of the correlation between $H_{wp}\left(u, v; \phi_{(0:\pi)}\right)$ and $G_{wp}\left(u, v; \theta_{(0:\pi)}\right)$. Also demonstrated in figure 3.11c, is a congregation of error tending towards 90° and large errors occurring for approximately the first 10 radial components, which are assumed to be sampled correctly. This is initially confirmed by the standard deviation plot in figure 3.11d because the largest errors are tending to 90°. What is critical apart from the size of the standard deviation error measurement is whether the rotational errors are a pattern that is a function of the image content, or systematic to the algorithm itself. Such as the interpolation method used or the quality of the input signal that is appropriately conditioned. The radial range over which figure 3.11d is obtained can be adjusted to eliminate more radial components with larger errors but this only provides a reduced signal bandwidth with which to operate on. With application to matched filtering, the upper cut-off frequency of the useful signal in the bandwidth of $F_{\theta}\left(r_{0:(x/2)-1}\right)$ can be determined by the camera's bandwidth. This is fixed by its aperture.

3.8 Accuracy of the Recovered Rotation Signal

The operating environment of an object recognition algorithm are considered in this analysis of the correlations accuracy to recover in-plane rotations. To begin this investigation an image containing natural structure is first presented, of which, the ideal response to the approximated nonlinear correlation signals are analysed.

Firstly, by imposing a range for the lower cut off point of the polar spatial frequency components, the changes in the error response (Eq. (3.15)) of the natural image and the recovered rotation angle measurements in figure 3.11a-c can be identified.

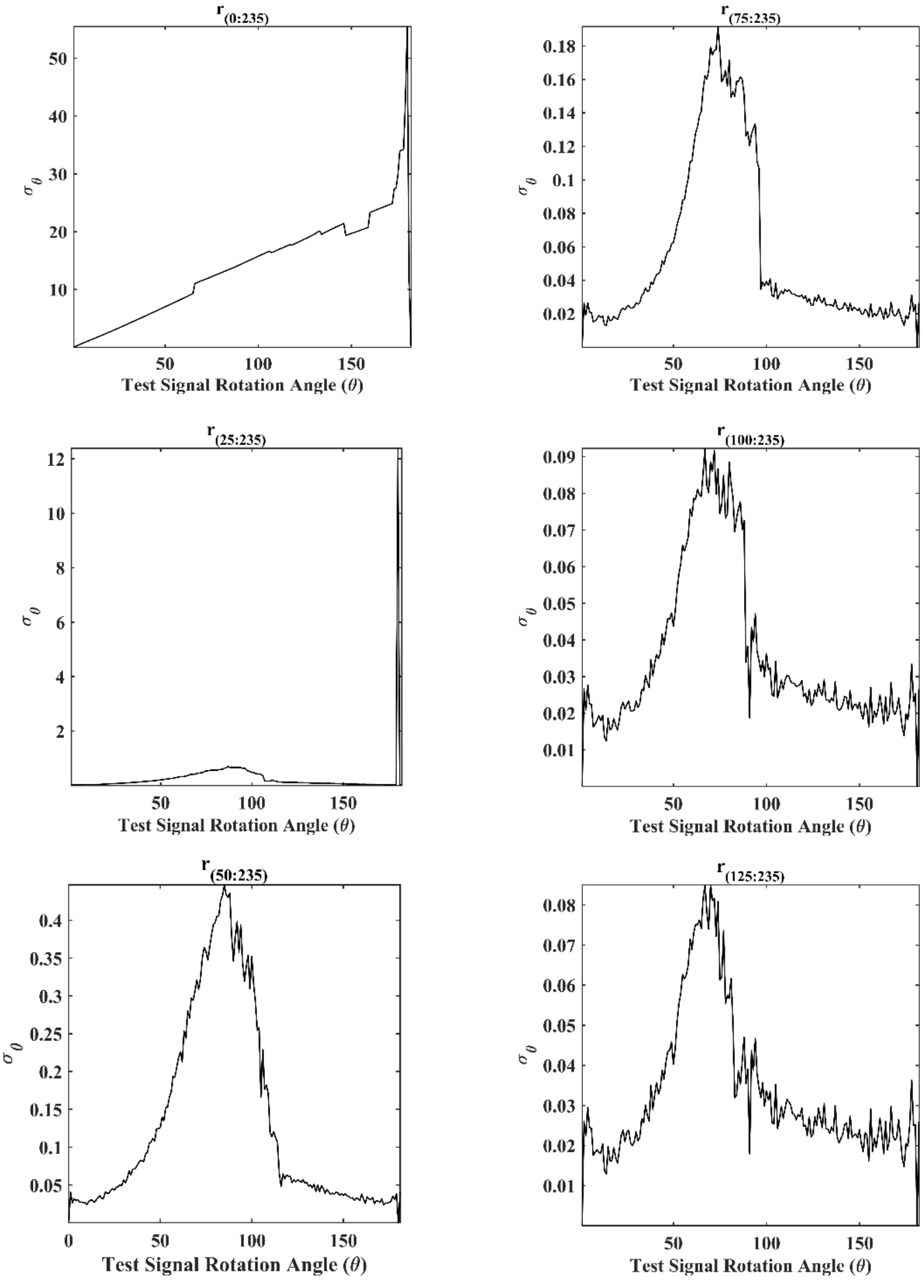
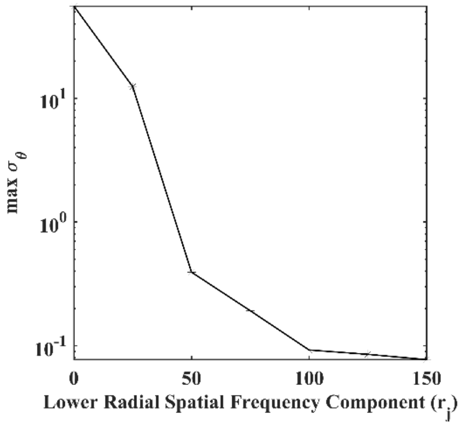
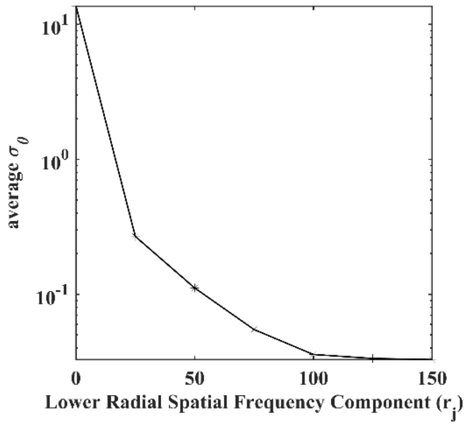


Figure 3.12: Standard deviation of error (σ_θ) for incremental polar spatial frequency lower cut off point $r_{(j:235)}$; $j = 0, 25, 50, 75, 100, 125$.

The radial bandwidth of this test image illustrated in figure 3.11c is $r_{(j:235)}$. For $j = 0, 25, 50, 75, 100, 125$, the error measurement shown in figure 3.12 develops the underlying structure of the error over the qualified range of radial lower cut off points. At $r_{(0:235)}$ the errors located around the first few polar spatial frequencies grossly perturb the range that the standard deviation measurement is taken from. Again, at $r_{(25:235)}$ the variation in the error at 180° still biases the overall error but some structure begins to present its self around the 90° test rotation angle in $g_{wp}\left(x, y; \theta_{(0:\pi)}\right)$. In both these cases the accuracy of the error is limiting the quality of the signal bandwidth analysis. For large errors that are characteristic of the signal content, they can occur at any polar frequency component not solely in the first few radial rings. When j increases further, two factors attributed to the signal become evident; the structure of the error and the natural variation within the signal; limited by its ideal reconstruction. As stated previously, it may be opportunely the case that the signal of interest to register has significant components that improve the correlation accuracy confined within a small bandwidth. However, it is much more likely that this is not the case and the increased polar spatial frequency lower cut off point is likely to limit the overall accuracy in the correlation. For $r_{(100:235)}$ and $r_{(125:235)}$ figures 3.13a-b demonstrate that both the maximum error value and the average variation beneath the error's structure have reduced variation. For wider bandwidths in this signal's polar spatial frequency representation, the error measurements accuracy deteriorates across a 180° field of view angle.



(a)



(b)

Figure 3.13: a) Maximum error, b) average error for incremental polar spatial frequency lower cut off point $r_{(j:235)}$; $j = 0, 25, 50, 75, 100, 125$.

Consider the error response at $r_{(50:235)}$ in figure 3.11c-d, the structure of the error begins to reveal itself without large bias from poorly reconstructed sampled points within the bandlimited signal; which potentially mask such structure for $\sigma_\theta > 1$. For a rotation interval $\Delta\phi = 5^\circ$ of the reference signal, each $h_{wp}\left(x, y; \phi_{(0:\pi)}\right)$ are correlated with $g_{wp}\left(x, y; \theta_{(0:\pi)}\right)$. For the imposed rotations on $h_{wp}\left(x, y; \phi_{(0:\pi)}\right)$ a continuous error is expected to be observed, however in figure 3.14.1 - 3.14.3 this is only the case between specific orientations as spurious errors occur at specific rotations. To interpret these orientation combinations figure 3.14.1-3 tracks the average standard deviation using Eq. (3.16) between each correlation measurement between $h_{wp}\left(x, y; \phi_{(0:\pi)}\right)$ and $g_{wp}\left(x, y; \theta_{(0:\pi)}\right)$.

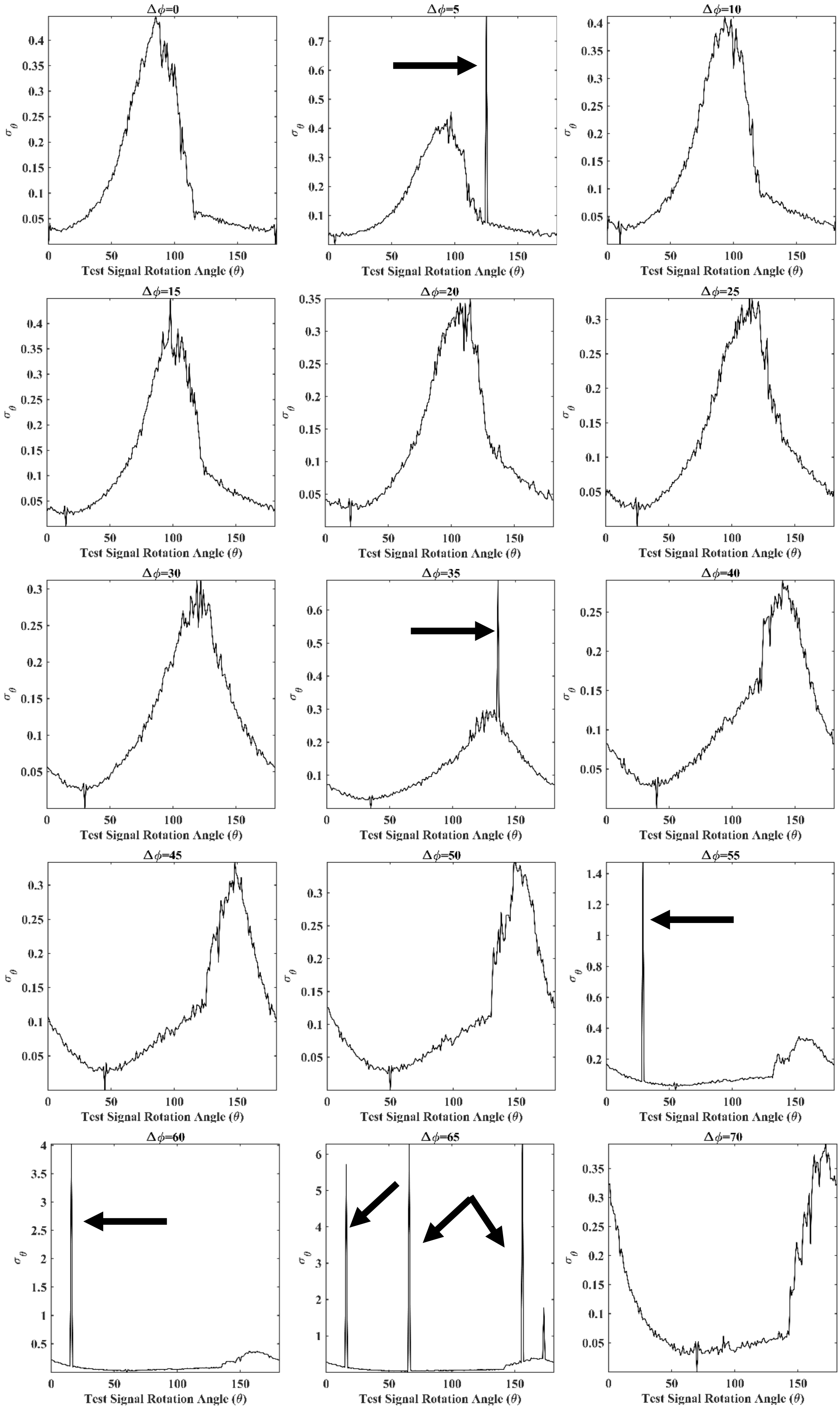


Figure. 3.14.1: Standard deviation of error (σ_θ); $\Delta\phi = 0^\circ : 70^\circ$. Arrows indicate spurious signals.

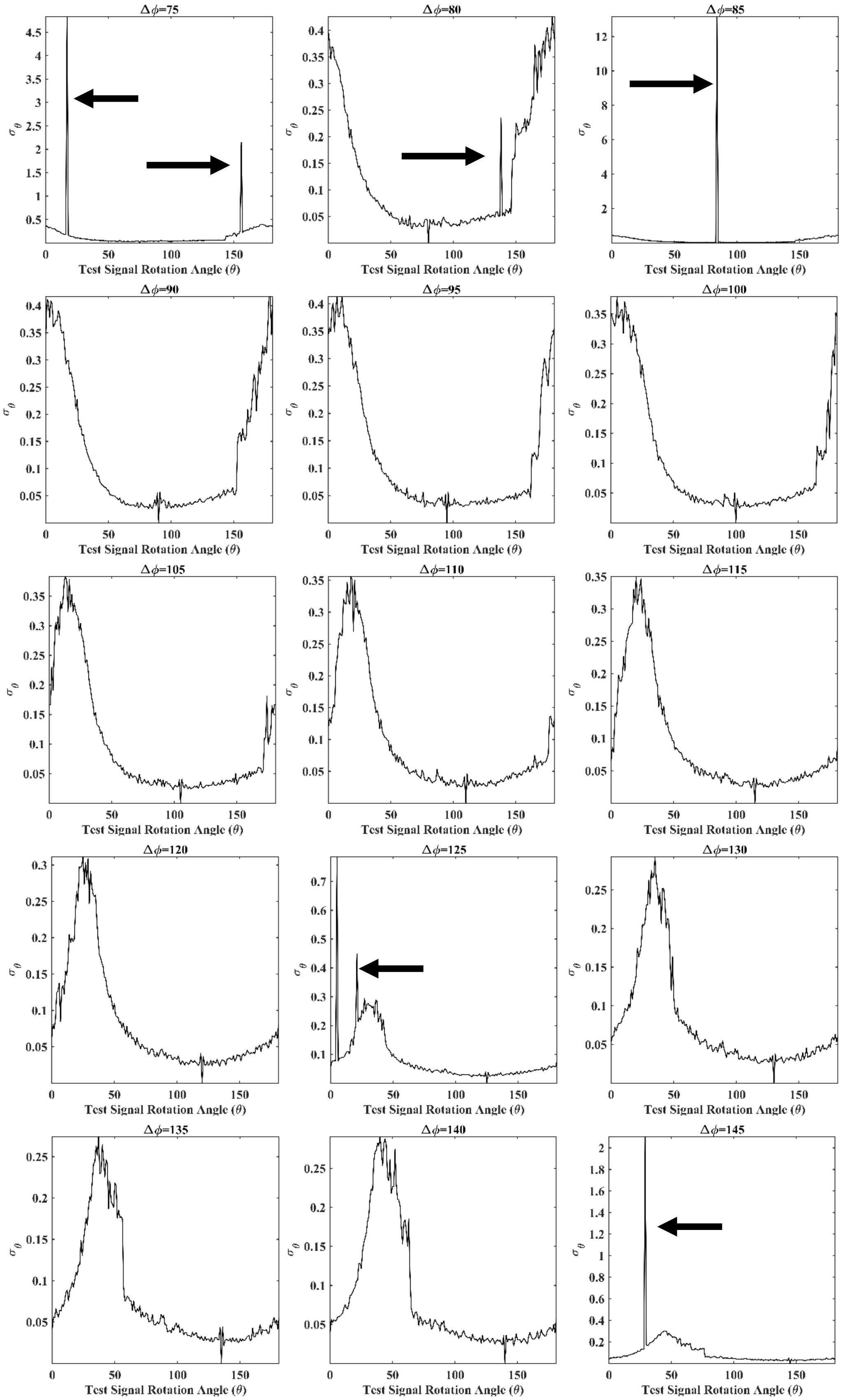


Figure. 3.14.2: Standard deviation of error (σ_θ); $\Delta\phi = 75^\circ : 145^\circ$. Arrows indicate spurious signals.

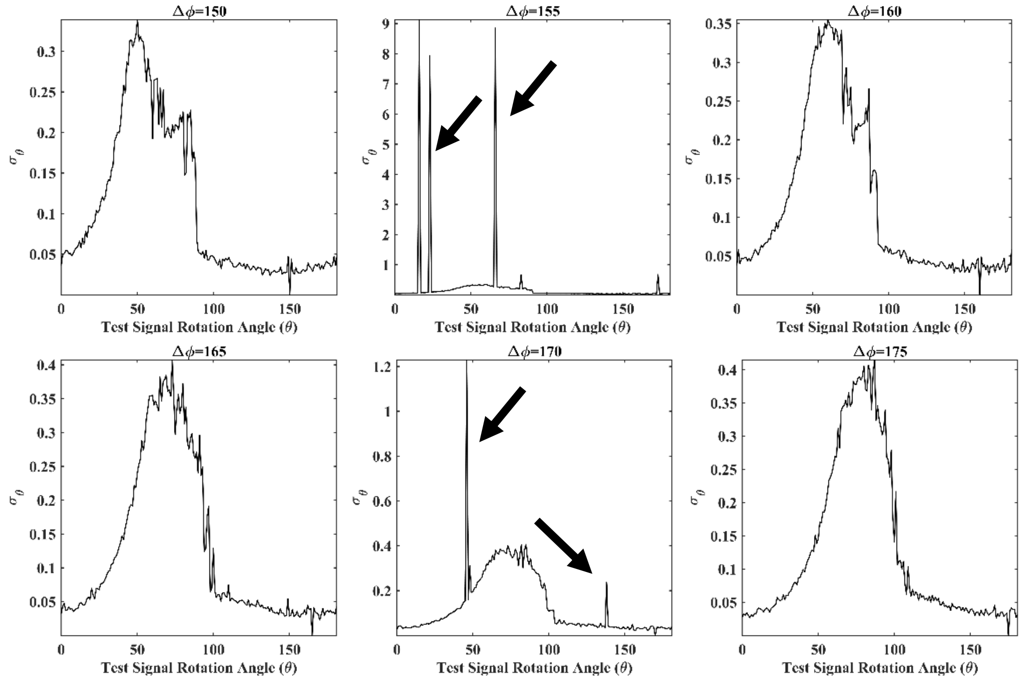


Figure. 3.14.3: Standard deviation of error (σ_θ); $\Delta\phi = 150^\circ : 175^\circ$. Arrows indicate spurious signals.

These apparent errors can be retraced back to frequency components of the test image through the location of the radial position using the equivalent rotational error measurement. Such as the tracked error over the radial bandwidth of the nonuniformly sampled signal presented in figure 3.11c.

Discounting the spurious locations of error, the structure of the error as discussed prior converges around 90°, although also displaying degrees of symmetry. The interpolation of the discretely sampled points using a nearest neighbour approach may be adding to these apparent spurious errors. This discussion leads onto the up-sampled method of correlation presented in section 3.6. For increased sampling using a nearest neighbour interpolation strategy, figure 3.15b-d demonstrates that the average standard deviation of error decreases accordingly whilst revealing further evidence of symmetry at the locations of error.

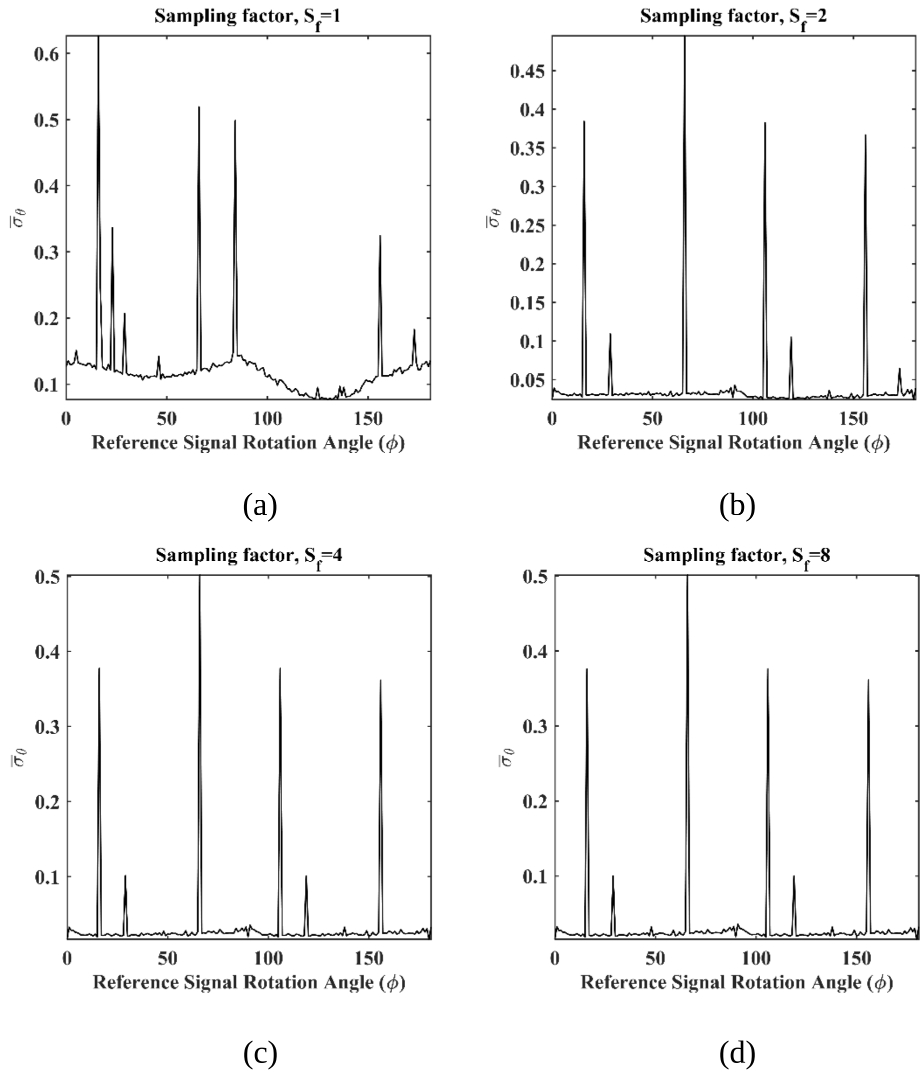


Figure 3.15: Average standard deviation ($\bar{\sigma}_\phi$) from nearest neighbour interpolation and up-sampling of the reference signal $h_{wp}\left(x, y; \phi_{(0:\pi)}\right)$; a) $s_f = 1$, b) $s_f = 2$, c) $s_f = 4$, d) $s_f = 8$.

In comparison of the nearest neighbour method linear interpolation of the same signal using the same up-sampling rates, as presented in figure 3.16, resulted in: marginal improvements to the baseline level of error ($\bar{\sigma}_b$) for sampling factors $s_f > 1$ and the same symmetries of error location and respective magnitudes. The corresponding analysis of accuracy against the interpolation approximation to the ideal reconstruction of the windowed natural signal of figure 3.11a is drawn to a conclusion in figure 3.17 through a characterisation of the baseline error $\bar{\sigma}_b$. Compared in table 3.2 are the relative differences in error between these two interpolation methods. The highlight of this comparison is that the relative error marginalises by an improvement of 6.2 for $s_f = 2$.

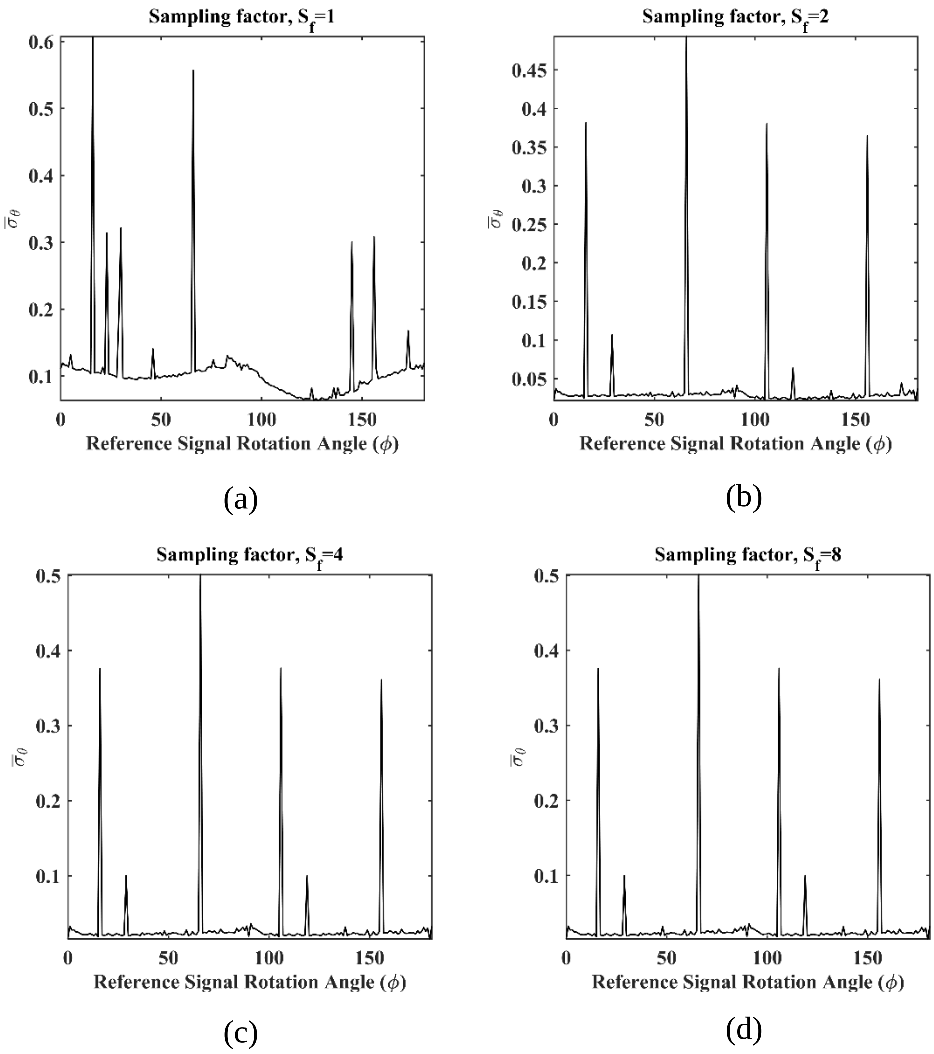


Figure 3.16: Average standard deviation ($\bar{\sigma}_\phi$) from linear interpolation and up-sampling of the reference signal $h_{wp}\left(x, y; \phi_{(0:\pi)}\right)$; a) $s_f = 1$, b) $s_f = 2$, c) $s_f = 4$, d) $s_f = 8$.

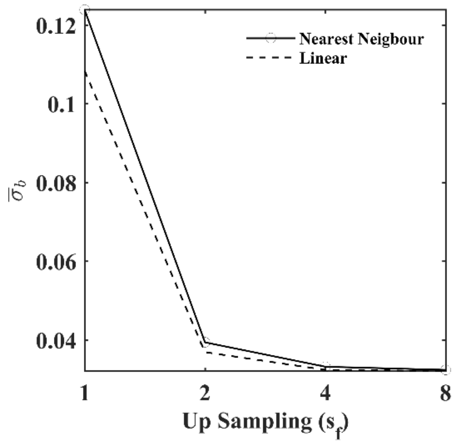


Figure 3.17: Baseline average standard deviation of error ($\bar{\sigma}_b$) of interpolation through nearest neighbour and linear approximation.

Table 3.2: Natural image, comparison of up-sampled interpolation methods.

	Nearest-Neighbour $\bar{\sigma}_b$	***Linear*** $\bar{\sigma}_b$	***Relative Error***	***Improvement-factor***
$s_f = 1$	0.1239	0.1084	0.0155	-
$s_f = 2$	0.0394	0.0369	0.0025	6.2
$s_f = 4$	0.0332	0.0324	0.0008	3.125
$s_f = 8$	0.0324	0.0321	0.0003	2.6

The relationship and trends of the error obtained from the approach to reconstruct an error free signal for performing nonlinear correlation are products of the analysis. Hence, any interpretations of the error signal will change along with the structure of the signal. To illustrate signal dependency two further naturally structured images are analysed in figure 3.18 and 3.19. As described in figure 3.15, the comparisons are drawn using nearest neighbour interpolation for increases in s_f .

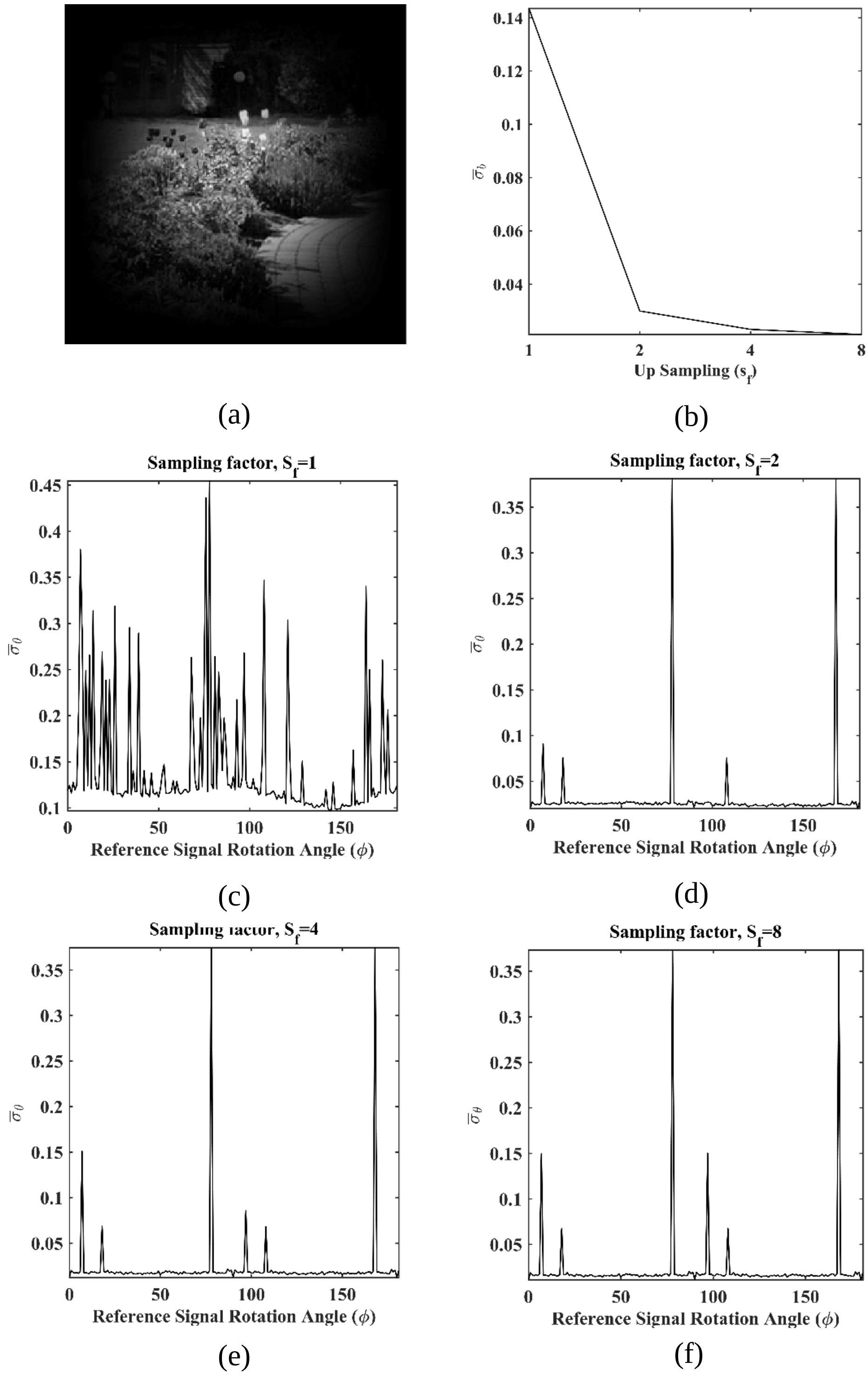


Figure 3.18: Landscape image [151]. Average standard deviation ($\bar{\sigma}_\phi$) from nearest neighbour interpolation and up-sampling of the reference signal. a) $h_{wp}\left(x, y; \phi_{(0:\pi)}\right)$, b) baseline average standard deviation of error ($\bar{\sigma}_b$), c) $s_f = 1$, d) $s_f = 2$, e) $s_f = 4$ f) $s_f = 8$.

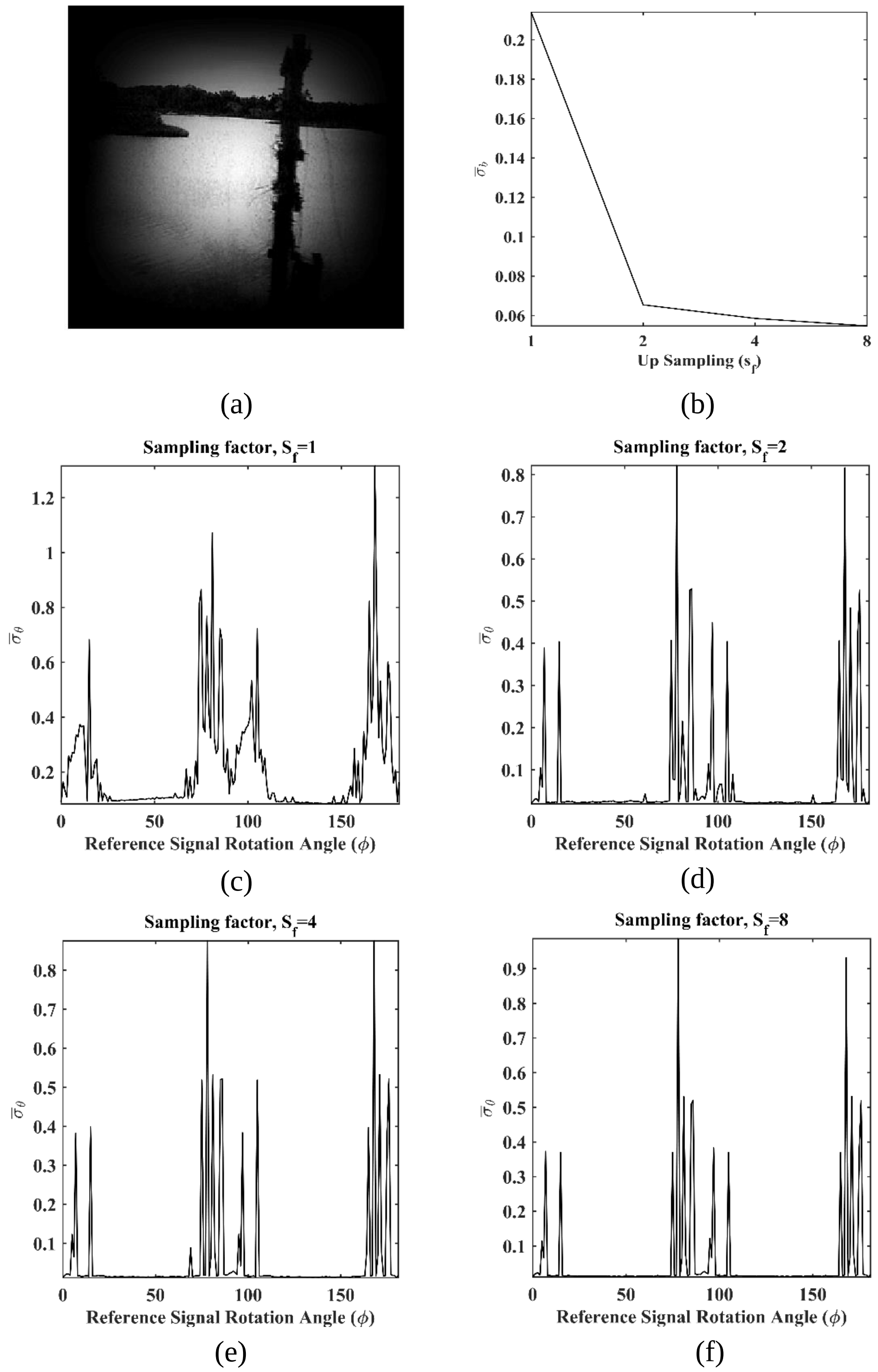


Figure 3.19: Landscape image [151]. Average standard deviation ($\bar{\sigma}_\phi$) from nearest neighbour interpolation and up-sampling of the reference signal. a) $h_{wp}\left(x, y; \phi_{(0:\pi)}\right)$, b) baseline average standard deviation of error ($\bar{\sigma}_b$), c) $s_f = 1$, d) $s_f = 2$, e) $s_f = 4$ f) $s_f = 8$.

A common theme across the three examples of natural images in figure 3.11a, and 3.18 - 3.19 is a rapid increase in the rate of change in accuracy when $s_f = 2$. The measured values in table 3.3 show that for increases in sampling of the test image, the rate of improvement in accuracy consistently decreases by, $\bar{\sigma}_{b(s_f)} / \bar{\sigma}_{b(s_{f+1})}$.

Table 3.3: Natural images, rate of improvement: up-sampled NN interpolation.

Natural Image: Figure 3.11		
	Nearest-Neighbour $\bar{\sigma}_b$	***Improvement-factor***
$s_f = 1$	0.1239	-
$s_f = 2$	0.0394	3.14
$s_f = 4$	0.0332	1.19
$s_f = 8$	0.0324	1.02
Natural Image: Figure 3.18		
$s_f = 1$	0.1435	-
$s_f = 2$	0.0299	4.8
$s_f = 4$	0.023	1.3
$s_f = 8$	0.0221	1.04
Natural Image: Figure 3.19		
$s_f = 1$	0.214	-
$s_f = 2$	0.0654	3.27
$s_f = 4$	0.0585	1.12
$s_f = 8$	0.0547	1.07

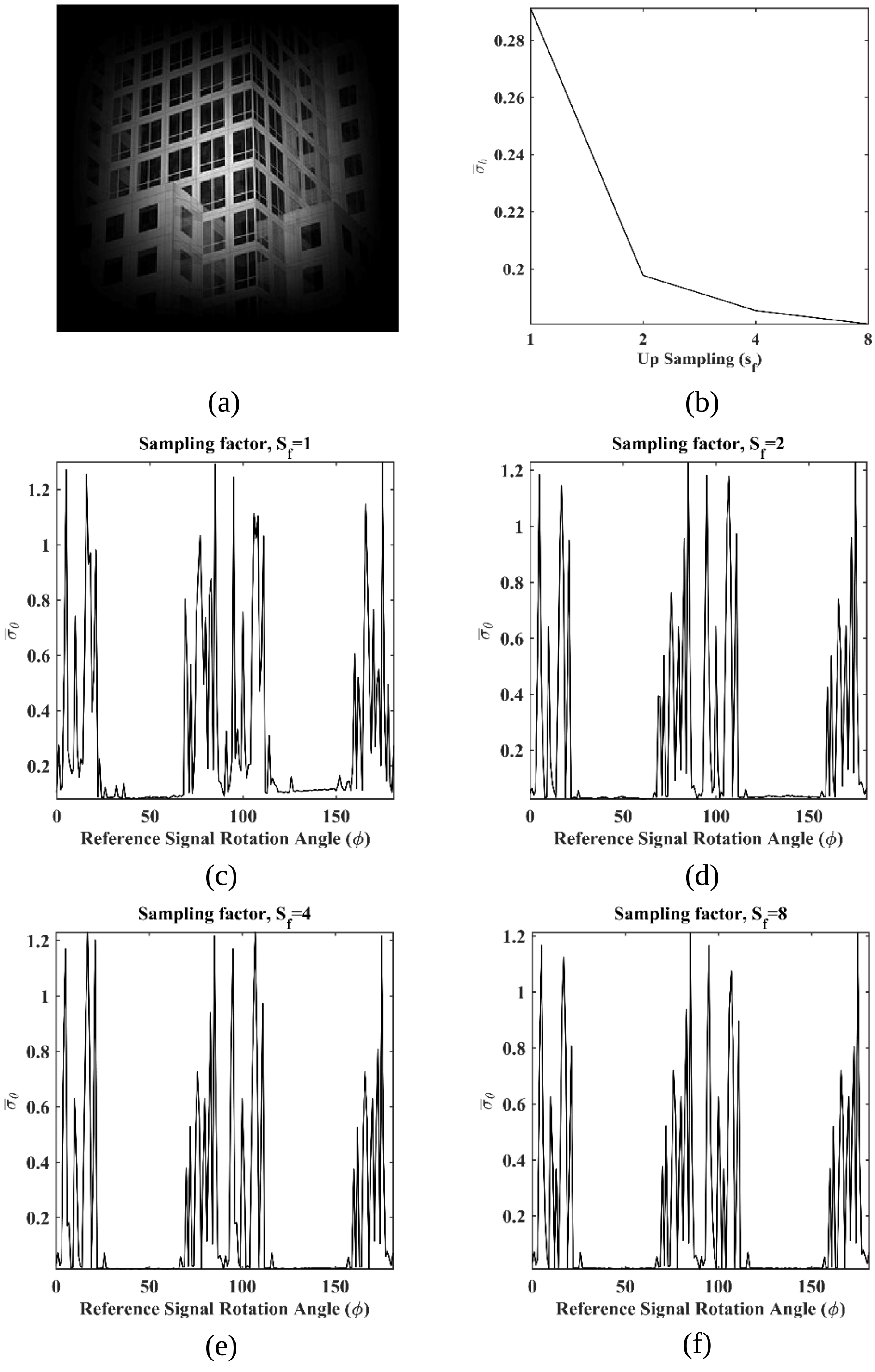


Figure 3.20: Building image [151]. Average standard deviation ($\bar{\sigma}_\phi$) from nearest neighbour interpolation and up-sampling of the reference signal. a) $h_{wp}\left(x, y; \phi_{(0:\pi)}\right)$, b) baseline average standard deviation of error ($\bar{\sigma}_b$), c) $s_f = 1$, d) $s_f = 2$, e) $s_f = 4$, f) $s_f = 8$.

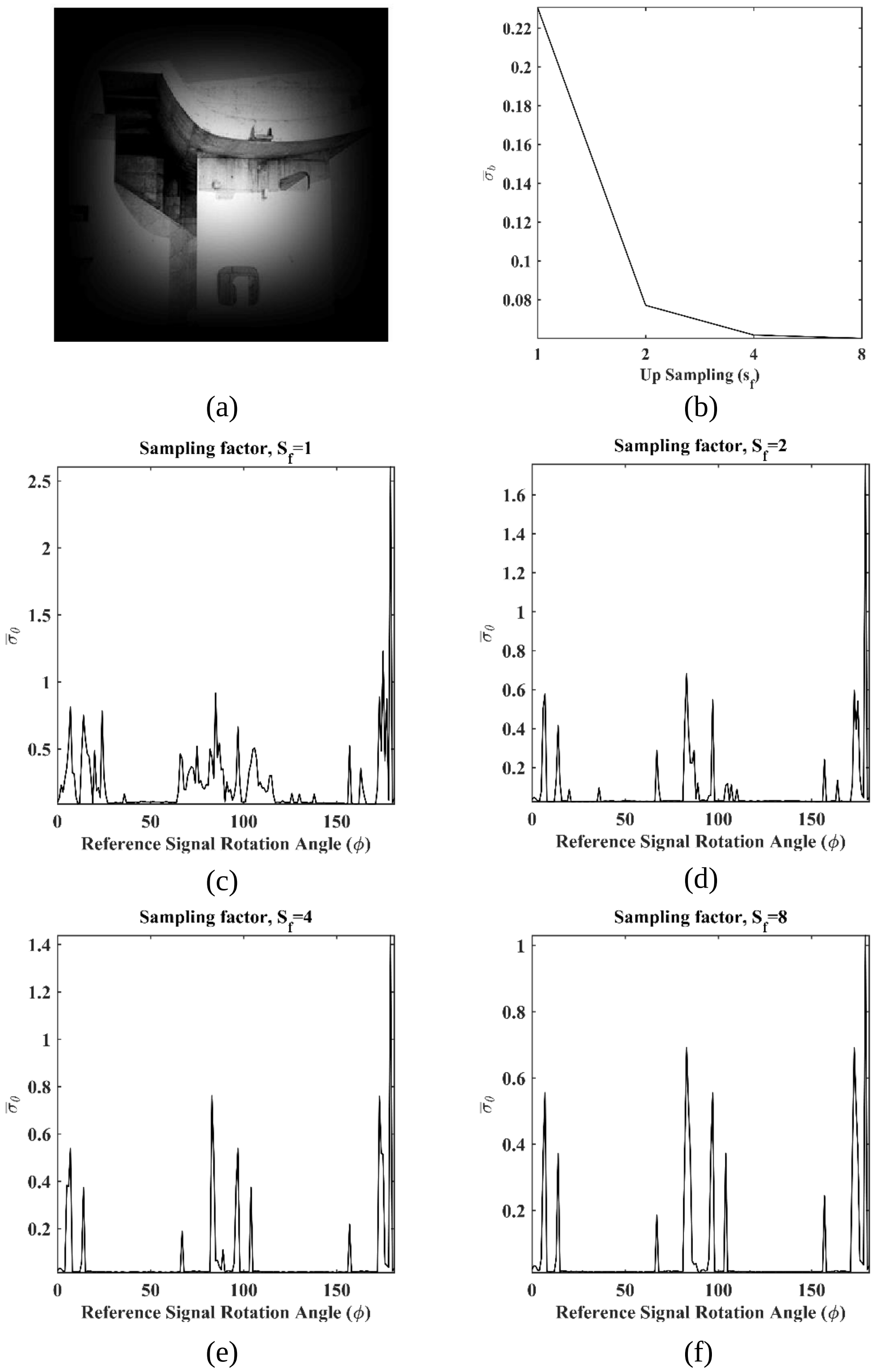


Figure 3.21: Building image [151]. Average standard deviation ($\bar{\sigma}_{\phi}$) from nearest neighbour interpolation and up-sampling of the reference signal. a) $h_{wp}\left(x, y; \phi_{(0:\pi)}\right)$, b) baseline average standard deviation of error ($\bar{\sigma}_b$), c) $s_f = 1$, d) $s_f = 2$, e) $s_f = 4$, f) $s_f = 8$.

Table 3.4: Rigid images, rate of improvement: up-sampled NN interpolation.

Rigid Image: Figure 3.20		
	Nearest-Neighbour $\bar{\sigma}_b$	***Improvement-factor***
$s_f = 1$	0.2912	-
$s_f = 2$	0.1977	1.47
$s_f = 4$	0.1854	1.07
$s_f = 8$	0.1807	1.03
Rigid Image: Figure 3.21		
$s_f = 1$	0.2307	-
$s_f = 2$	0.0771	3
$s_f = 4$	0.0618	1.25
$s_f = 8$	0.0601	1.03

The two rigidly structured images in figures 3.20 - 3.21 indicate that the rate of improvement in accuracy may also be reliant on additional parameters. One key difference between these two images is that in figure 3.21 the scene contains pockets of high variance and non-uniform structure within localised regions of the image. This is similar to either a sky, water or landscape scene, such as those presented in figure 3.11a and figures 3.18-3.19. Whereas, the image in figure 3.20 contains uniform structure with four-fold symmetries and equidistant regions of high variance, thus behaving quite differently to any of the other images presented. In this case, the peak errors at locations of $\phi_{(0:\pi)}$ which are included in the measurement of $\bar{\sigma}_b$ cause the rate of improvement in accuracy to not change rapidly or significantly. A second likewise feature of the images used for both natural and rigidly structured scenes is that figures 3.19 and 3.20 both have features lying with symmetries. In figure 3.19, the tree line and post due to the image composition give a false impression of rigid structure such as that found in the windows and building corners in figure 3.20. In the data plots for increases in s_f, for both images, there are prominent errors about 0° and 90°. Hence, biasing the peak errors over the average error. This is a further indication that the lower radial frequency cut off frequency varies with each signal.

3.9 Precision of the Recovered Rotation Signal

The precision of nonlinear correlation measurements is completely defined by the amount of noise in the image signal; see appendix A1 for the noise model. In this section, the precision of the correlation measurement highlights the available bandwidths to formulate matched filter responses. For consistency in the analysis, the image in figure 3.11a is used in this investigation. For $s_f = 1, 2, 4, 8$ the up-sampled signal is approximated using nearest neighbour interpolation, 10 samples are obtained to average each signal rotation; the reference image $h_{wp}\left(x, y; \phi\right)$ is fixed at 0° and the test image $g_{wp}\left(x, y; \theta_{(0:\pi)}\right)$ is scanned over π with 1° increments.

The precision analysis begins with figure 3.22 confirming the relationship between accuracy and signal bandwidth under low noise conditions. The available signal bandwidth of the test image (figure 3.11) in relation to the recovered rotations of a correlation between $h_{wp}\left(x, y; \phi\right)$ and $g_{wp}\left(x, y; \theta_{(0:\pi)}\right)$ changes according to the noise content in the signal. The bandwidth range is defined by lower and upper polar spatial frequency cut off points, $r_{(lower)}$ and $r_{(upper)}$. The relationship between accuracy and up-sampling, discussed in section 3.7, is only true when components with high signal to noise are chosen to process the correlation.

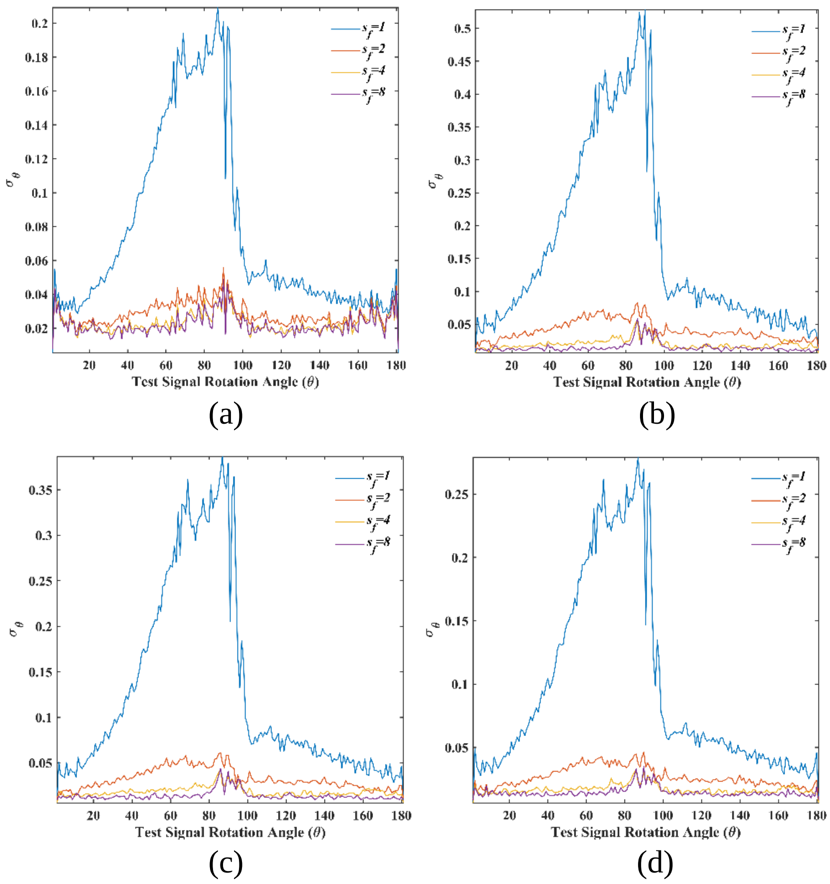


Figure 3.22: Up-sampling and low noise trade off: $s_f = 1, 2, 4, 8$, fixed photon shot noise - $\sqrt{N_{pht}} = 100$ and fixed low frequency noise - 1% of the signal's amplitude.

a) $r_{(50:75)}$, b) $r_{(50:100)}$, c) $r_{(50:150)}$, d) $r_{(50:235)}$.

For a moderate gain in noise, figure 3.23 depicts the interdependency of up-sampling and the noise signal. Within this signal there may be larger errors biasing the overall error across the correlation. Therefore, the correct selection of frequency components across rotations of $g_{wp}\left(x, y; \theta_{(0:\pi)}\right)$ becomes quite critical.

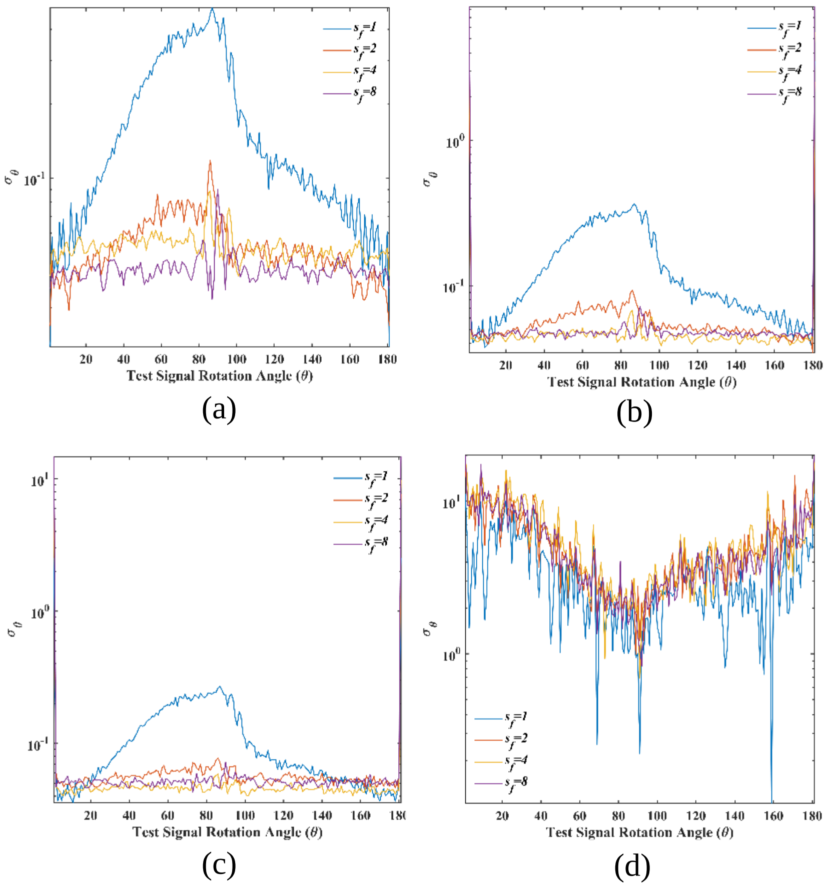


Figure 3.23: Up-sampling and increased noise trade off: $s_f = 1, 2, 4, 8$, fixed photon shot noise - $\sqrt{N_{pht}} = 31.6$ and fixed low frequency noise - 5% of the signal's amplitude. a) $r_{(50:75)}$, b) $r_{(50:100)}$, c) $r_{(50:150)}$, d) $r_{(50:235)}$.

At bandpass intervals for $r_{(upper)} = 75, 100, 150$ figure 3.23a-c shows that σ_θ may not be affected by noise to significant extents. When the whole bandwidth is considered for the measurement of σ_θ, as shown in figure 3.23d, errors are completely dominated by the overall noise response. This section now investigates the response of the noise model components: photon shot noise and low frequency noise. Figure 3.24a-d demonstrates the noise response of the chosen test image for fixed photon shot noise and increases of low frequency noise, and visa-versa, over a small and full signal bandwidth: $r_{(50:75)}$ and $r_{(50:235)}$. In figure 3.24a,c $\sqrt{N_{pht}} = 141.4$ and the range of $S(f)$ is 1%, 5%,10%, 20%. In figure 3.24b,d $S(f) = 1\%$ and the range of $\sqrt{N_{pht}}$ is 100, 70.7, 31.6, 10. For both bandwidth ranges, the up-sampling rate s_f is set to 2.

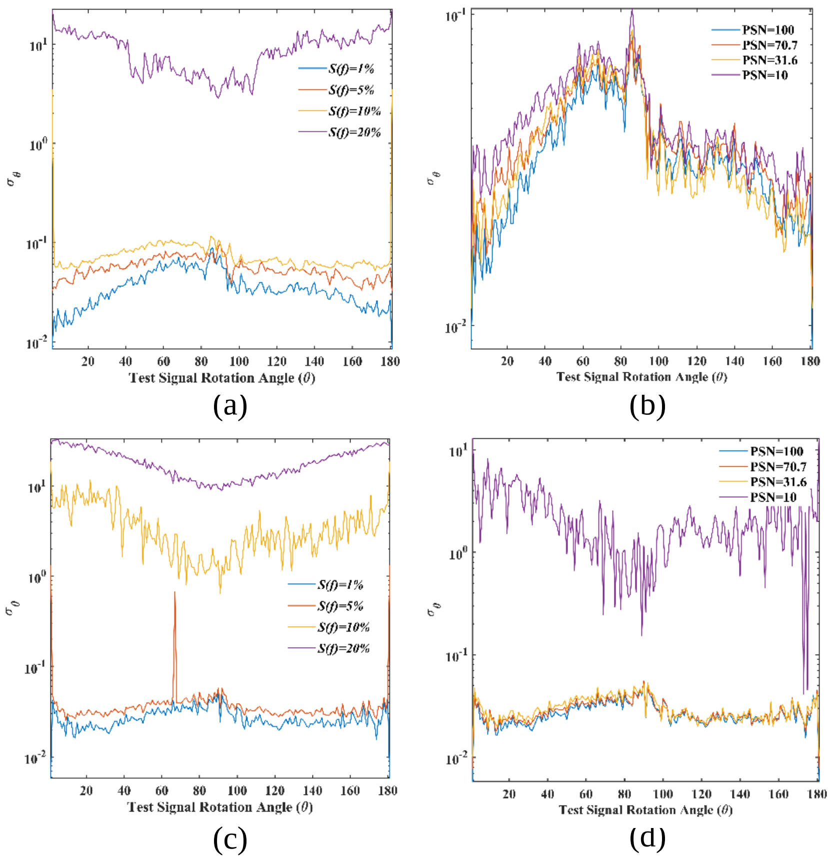


Figure 3.24: Bandwidth and noise trade off: $s_f = 2$, a) fixed photon shot noise (PSN), $r_{(50:75)}$, b) fixed low frequency noise $S(f)$, $r_{(50:75)}$, c) fixed PSN, $r_{(50:235)}$, d) fixed $S(f)$, $r_{(50:235)}$.

The effect of low frequency noise on a low to mid-range reduced correlation bandwidth has greater impact on the error signal across θ than photon shot noise. As described in section 3.4.2 typical filters used in uniformly sampled correlation measurements, assume true signals lie over a reduced range of frequency components. If this range is over a higher portion of the spectrum or over a wider range of the spectrum, photon shot noise will dominate and perturb the higher frequency components used in a correlation. Hence, for matched filtering low frequency noise must be equally considered when selecting a spread of components to form the filter's key. In the case of the test image used over the radial range $r_{(50:235)}$, and for $S(f) < 10\%$ and $\sqrt{N_{pht}} = 31.6$ the average value of σ_θ over $\theta_{(0:\pi)}$ is of the order of 0.05°.

The correlation analysis algorithm presented in section 3.7 develops an approach to optimise a matched filter to recover angular displacements through analysis of the correlation error. One drawback of this method is that a screening of an image over a range of θ is required, as such this is not an online method of implementing a correlation filter. A second limitation to consider is that for cross correlation, the dissimilar information between the known signal and observation will decrease the correlation accuracy if taken over a global frame. The actual available bandwidth to select the key of the filter from is dictated by the under-sampling of non-linear units at the lowest portion of the signal bandwidth and the cameras aperture at its highest. Like binary masks, activated components are chosen based on a threshold condition. Hence, it is reasonable to impose a cut off condition to select frequency locations of the key by only retaining polar spatial frequency components that have a SNR >1dB. Measuring or estimating the amount of noise in a signal without knowledge of noise is a separate issue. Since natural structures in images can impose variabilities into binary decisions and miss information, it is difficult to characterise generalised filter keys. However, the spectrum of an image can indicate the type of underlying noise signal present as well as the makeup of the signal.

3.10 Summary

The aim of this chapters has been to analyse the sensitivity of registering angular displacements via a combination of nonuniform polar spatial frequency sampling and spatial re-sampling of the input function. The primary purpose of the analysis is to enable the recovery of invariant features in images from the translational and rotational elements of a rigid body transformation. Secondary, is an autocorrelation algorithm to identify potential structure of an image's composition. Up-sampling of the input signal replaces the spatial amplitudes and phase information of a coarser polar coordinate grid with a finer estimate of the interpolated grid points. Based on simple nearest-neighbour interpolation the accuracy approaches the accuracy of a linear interpolant for just two times the original image resolution. Beyond four times the resolution of the original sampled image, the average errors between the two interpolants, dependent on the amount of rigid structure, can be as low as 10^{-4}. The analysis is constrained to the example images. The motivation for the investigations came from the application of correlation filters. Where the reliability of the displacement measurement can be constructed from signals with low SNR, even within the passband of a filter. The matched filter analysis identifies the signal's signal strength for the entire bandwidth of the signal and as such, the optimum filter can be applied. Although there is a signal dependency of this application, generality can be reduced using the spread of image frequency spectra.

Chapter 4 - Identifying and Extracting Critical Features

The boundary of an object is the first critical feature for an object recognition system to detect after the reduction of random noise and unreliable image components. In section 2.3 it is identified that for digital images a boundary exists for a rate of change in pixel intensity. Forming the fundamental building block in edge detection, the rate of change is measured using numerical differentiation and implemented through convolution. Rather than convolution affecting the robustness of detecting true edge components the reliability of an edge detectors measurement is limited for two reasons: the formation of discrete kernels and the nature and order of continuous differentiation… An example of a low-level feature detection is an object's edge. Localising low-level features within an image crucially provides a means of building up to a higher-level abstraction of an object and its parts. Even in simplified environments through segmentation routines, the complex nature of non-rigid objects may impose additional requirements upon an edge detector's operating range. In this chapter, a first order differentiator kernel to encompass a wider range of detection accuracy and precision is presented. Application of the developed edge detection algorithm is demonstrated on computer generated binary images of edges. Through the process of convolution, object localisation issues for achieving successful critical feature extractions are first considered.

4.1 Chapter Synopsis

Readers can refer to this synopsis of the research contained in chapter 4 when reading the investigations within this chapter.

This chapter describes the procedure used to detect edge information within the input image, the aim is to simplify and isolate potential structures of interest so that object features can be localised. This is done by emphasising those image structures whose intensities change rapidly as compared to neighbouring pixels. The chapter begins with a general discussion on the image processing steps leading to feature extraction to isolate an object or image area (chapter 3). In describing how features can be represented by their boundary, issues in localising such features are discussed **(section 4.2)**. An important aspect of a non-rigid object boundary is that the gradient direction may not be necessarily occurring at discrete angles defined on a linearly sampled grid. The suitability of kernel coefficients for the purpose of sub-pixel edge detection and arbitrary gradient direction kernel generation are discussed in **(section 4.3)**. Based on a defined response within an image region to an orientated filter, an edge detection technique is detailed. The detection range is extended by convolving a set of orientated filters with an input image and revaluating the edge response using a second-order polynomial. The method is called region maxima (RM) edge detection **(section 4.4).** An analysis of this technique is performed based on the number of sampling points used in the polynomial reconstruction of the edge and the gradient interval of the orientated filters. The purpose of the analysis characterises the sensitivity of reconstructed edge symmetry properties in the presence of imaging system noise **(section 4.5)**. Owing to the recovered operating range and accuracy of the RM method an iterative version of the edge detection algorithm is described. To test the refined estimate of the edge gradient directions special test functions are implemented to accurately determine the precision of the iterative method **(section 4.6)**.

To begin, a pictorial summary of this chapter's investigation is presented,

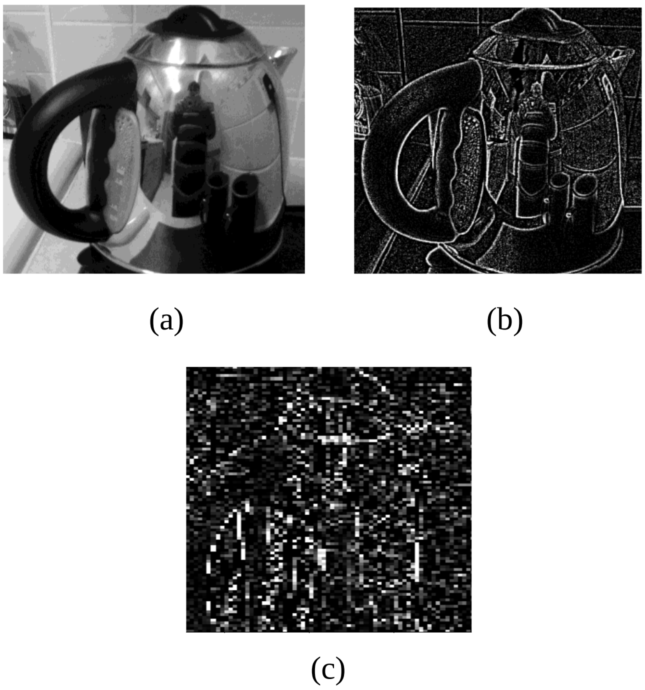


Figure 4.1: Example of edge detection: a) grayscale image, b) Canny edge image - standard MATLAB function, c) down sampled Canny edge image.

The illustrative image in figure 4.1a is a cropped area of a 1056 $\times$2048 capture, the edges extracted using the standard canny method in figure 4.1b detect the abrupt transitions in pixel intensities. However, there is a lot of background information being detected which may also be edges themselves, not just the edges contributing to the makeup of the object kettle. This emphasises the benefits of studying an object under isolation. Furthermore, for low resolution images the saliency of the detected edges in figure 4.1c diminishes significantly in this example when it is down sampled to a 100 $\times$100 capture. This chapter investigates the process of edge detection and the accuracy of gradient features to recover the most salient features of an object's boundary.

4.2 Object Feature Extraction Process

Generally, the removal of as much visual information that potentially masks the true observation of the object relieves the potential stresses upon additional processing steps. In terms of feature extraction algorithms, performance criteria such as accuracy and precision and operating sensitivity are predefined at the initial image processing stage. To achieve an extraction of a feature, which for example could be a geometric, probabilistic, statistical or visual account of an object and its parts, it is imperative to first detect possible features and then localise the object or object part containing those features. The flow of visual information and the relative processing stages to achieve a feature extraction is presented in figure 4.2.

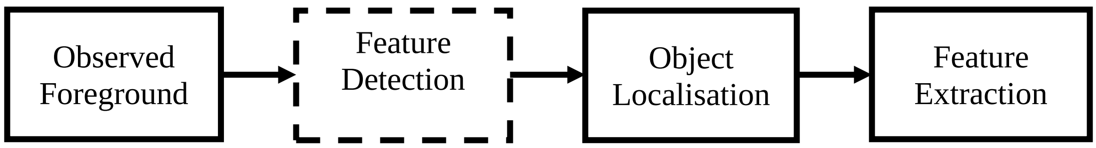
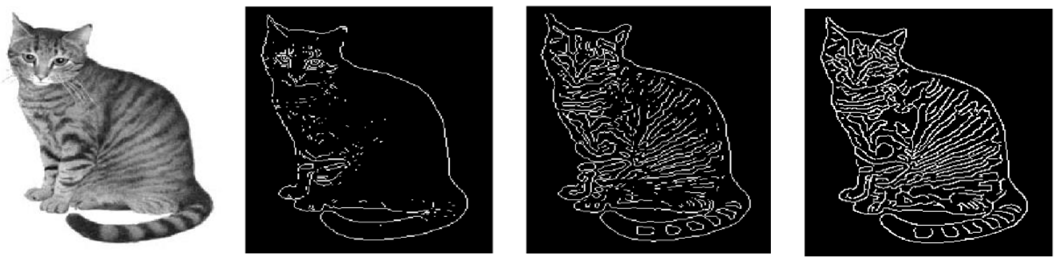


Figure 4.2: Adopted approach for extracting features - highlighting the context of, and critical process, feature detection.

The pertinent point that begins this chapter's investigations is the process of object feature extraction. However, this chapter first describes and develops the importance of the prevalent processes as highlighted in figure 4.2; feature detection leading to object localisation. After which, the concept of feature levels and abstractions (types) of features are presented to bring the focus of discussion in this chapter towards detecting edge features in non-rigid objects.

4.2.1 Feature Representation and Measurement

An object's feature can be defined as a tangible piece of information that is discriminative of the overall presentation of information. For a similar or identical object, the feature may be exact or similar, thus an accumulation of the most discriminative (critical) features has an aim of achieving a unique representation of an object. What these features are has impact upon the representation and the measurement of them. In line with the discussion in chapter 1's introduction of Hubel and Wiesel [5] and the sensitivity to orientated edges in human vision and the representation of shape recognition, (simple and complex) in Gestalt theory [12], the first discriminative feature that is identified are the edges of an object. In section 2.3 the principal methods for detecting edges were presented: first order, second order and hybrid edge algorithms., The characteristics of these algorithms are illustrated in figure 4.3b-d using standard MATLAB functions. Whereby, the strengths and weaknesses in finding a suitable implementation to maximise the detection of edges are highlighted.



(a) (b) (c) (d)

Figure 4.3: Examples of edge detection techniques: a) input image [151], b) Sobel (first-order), c) Laplacian of Gaussian (second-order), c) Canny (hybrid).

In figure 4.3b the edges are defined by the gradient maximum of the differentiated edge, for figure 4.3c the edges are defined by the zeros crossing in the second differentiation of the edge, and in figure 4.3d edge liking and double thresholding is applied to enhance the strongest edges.

The cat head in figure 4.3 is an object part which consists of features describing the eye, ears, nose and mouth. Hence, it is the detection of hierarchical patterns of-and-within features that is most important for the overall process of recognition; such as abrupt transitions in image intensity depicting possible boundaries. This in turn allows specific feature information such as geometrical and textural components of the cat's ears to be recovered. The abstraction of feature layers to higher-levels and the accumulation of evidence for the existence of an object, is a topic of investigation in chapter 5. Therefore, the depiction of low and high-level feature abstractions in this chapter is kept to a definition within the context of feature extraction. For example, the detection of intersecting points using the edge image is an increased level of feature abstraction albeit still a low-level feature. A higher-level abstraction of a feature specifies something more unique to the object such as the texture across a plane and the spatiotemporal relationship to a second located feature; that is also initially described by a boundary. What is most critical post edge detection and perhaps the most challenging, and in support of recognition by parts, is the isolation of an objects part.

4.2.2 Localising Features

Parts of a cat are identified as the head, eyes, nose, ears, tail, leg, paw, body etc…In comparison to another generic class of animal such as a dog or rabbit some object parts or relationship between parts may not be so discriminative, hence redundant. As correlation is a method of shape matching, which is discussed in section 2.2, the representation of the object part must be made as simple as it needs to be whilst retaining its discriminative features. Obviously if certain parts are occluded, not present or undetectable, additional restrictions present themselves for achieving higher probabilities of successful recognition. A localisation process based on prior knowledge built from evidence accumulation is explored in chapter 5, central to this process and forming the main investigation in this chapter is the detection accuracy of an object's boundary and its characteristic features: edge magnitude and direction. The reasons behind this investigation follow on from three points of view: non-rigid objects are irregular, accumulated evidence is generated from the lowest level of feature and that industrial and bioscience imaging and vision applications, particularly analysis and monitoring processes, may benefit from improved detection accuracy.

The motivations behind low-level feature detection and the key requirements of edge detection for pattern recognition are now established. The next section conveys and develops the necessary theory for generating legitimate gradient kernels beyond the pixel constraint of a discrete 2D cartesian grid.

4.3 Arbitrary Gradient Kernel Direction Generation

The proceeding analysis of the gradient kernels looks at the symmetry of the first-order differential operator across sub-pixel gradient directions. The necessity of implementing kernels, using integer coefficient weights, is discussed alongside Sobel's kernel formulation and the coefficient weights normal to the line of symmetry at arbitrary gradient directions.

The general issue of using a discrete grid is the pixel resolution it is bounded by, whereby a continuous line of symmetry is completely defined at discrete locations. Since a circle is an asymmetrical function, a repeated pattern exists for a range of 0° and 90°. Most importantly, a linear coordinate grid does not line up with a nonlinear coordinate grid when discrete integer grid points are considered except for $\pm 90^{\circ}$ and $\pm 45^{\circ}$. This is easily recognisable for an angle of 22.5° on the unit circle using a 3×3 kernel, whereby the linear coordinates $\left(x_i, y_j\right)$ equal to $\left(1, 0.5\right)$ generate a line of symmetry at 26.57°. Increasing the size of a gradient kernel creates additional discrete locations at which extended gradient directions lie. The line of symmetry in adopting this approach becomes noncontinuous and only approximates the gradient to what is already perturbed to a systematic discrete error. A second issue of using an increased kernel size is that the minimum translational geometry of a detectable edge increases with it. This chapter's investigations first show how a standard 3×3 kernel is established using a vector gradient. Then a method of upscaling to larger kernels via convolution is presented before the approach to generate arbitrary kernel gradient directions is derived.

4.3.1 Kernel Mask Generation, Vector gradient

The central difference method stated in Eq. (2.11) is implemented as a vector sum of the gradient. Whereby, the dot product between central difference coordinate pairs approximate the projected gradient ***G***. In figure 4.4, a 3×3 kernel with nine locations (*a* to *i*) depicts how a 3×3 region of an image is overlaid with the radial position of kernel gradient.

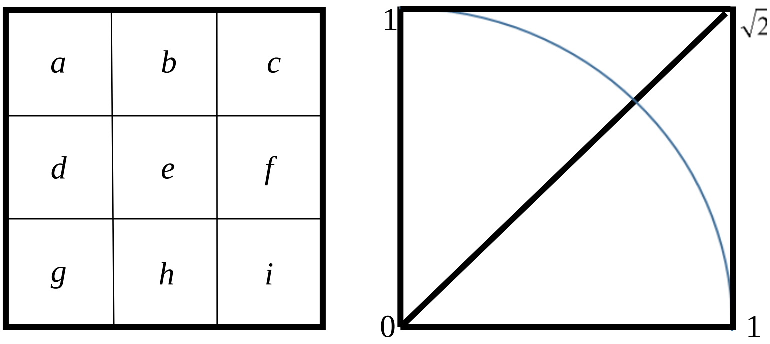


Figure 4.4: Overlay of kernel locations (*a* to *i*) and a 3×3 image regions corresponding radial positions.

The coordinate pairs as depicted in figure 4.3 are (*a*,*i*), (*c*,*g*), (*b*,*h*) and (*d*,*f*). For the radial positions $\boldsymbol{r}1=0$, $\boldsymbol{r}2=1$ and $\boldsymbol{r}3=\sqrt{2}$ and collecting likewise terms in the kernel locations where $\langle \cdot,\cdot \rangle$ is the distance from the centre of the kernel grid *e*, the projected gradient is

$$\begin{aligned}\mathbf{G} = \ & \frac{a-i}{\mathbf{r}3}\cdot\frac{\langle -1,1\rangle}{\mathbf{r}3}+\frac{c-g}{\mathbf{r}3}\cdot\frac{\langle 1,1\rangle}{\mathbf{r}3} \\ & +\frac{b-h}{\mathbf{r}2}\cdot\frac{\langle 0,1\rangle}{\mathbf{r}2}+\frac{d-f}{\mathbf{r}2}\cdot\frac{\langle 1,0\rangle}{\mathbf{r}2} \\ & +\frac{e}{\mathbf{r}1}\cdot\frac{\langle 0,0\rangle}{\mathbf{r}1}.\end{aligned} \tag{4.1}$$

The centre component (*e*) is a zero vector and the common factors are the radial positions ***r***, hence the orthogonal gradient components are

$$\begin{aligned}\mathbf{G_X} &= \frac{1}{(\mathbf{r}3)^2}(-a-i+c-g)+\frac{1}{(\mathbf{r}2)^2}(d-f), \\ \mathbf{G_Y} &= \frac{1}{(\mathbf{r}3)^2}(a-i+c-g)+\frac{1}{(\mathbf{r}2)^2}(b-h).\end{aligned} \tag{4.2}$$

The constant multiplicative for each central difference pairing in Eq. (4.2) are the kernel coefficient weights. By factoring $\mathbf{G}$ to $2\mathbf{G}$ the coefficients become integer values, hence the radial constants of the kernel matrix for the gradient vector directions become

$$2\left(\mathbf{G}_{\mathbf{X}}\right) = D_{\theta=0^\circ} = \begin{bmatrix} -1 & 0 & 1 \\ 2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad 2\left(\mathbf{G}_{\mathbf{Y}}\right) = D_{\theta=90^\circ} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}. \qquad (4.3)$$

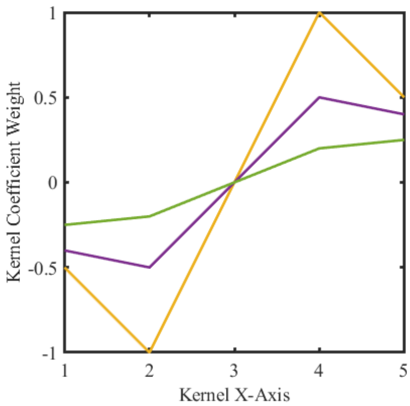
For larger kernels coordinate pairs are collected along each equidistant radius, for example, a 5×5 kernel would add three more radial positions $\boldsymbol{r}$ to evaluate eight further coordinate pairs.

4.3.2 Larger Kernel Masks, Convolution

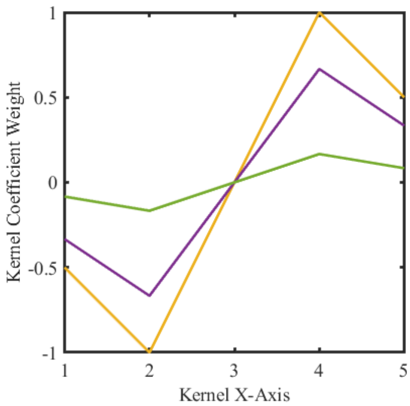
As described for the standard 3×3 kernel in the last section, the same procedure using the vector summed gradient for arbitrary kernel sizes can be easily implemented. Albeit that due to potentially large differential values, normalisation of the kernel values is a general practice before performing convolution with an image. As presented in section 2.3, the Sobel kernel is the outer product of a 1D gradient vector normalised to the unit circle with a triangle filter (also 1D). The filter aides in smoothing high frequency oscillations generated in the convolution between the kernel and the image. In the same vein larger kernels can be digitally implemented equivalent to the vector gradient method that has just been derived. For a 5×5 kernel along the *x*-direction this approach is demonstrated by performing a 2D convolution between $D_{\theta=0^\circ}$ and the outer product of two triangular filter vectors; one filter is transposed.

$$\mathbf{G_X}(5\times5) = \left(\left[121\right]^T \otimes \left[121\right]\right) * D_{\theta=0^\circ}\ . \tag{4.4}$$

The generation of $\mathbf{G_X}(5\times5)$ along the *x*-direction using the vector sum gradient and the convolution method in Eq. (4.4) is compared in figure 4.5. The ratios between opposing coefficients of the symmetry line in the two methods are not equal and the height of the non-normalised filter in the convolution method changes the rate of the slope along the perimeter of the kernel. So that the convolution between the kernel and the image is relative to the pixel intensities of the image, the slope of the kernel is normalised to either the size of the kernel or the maximum weight value.



(a)



(b)

Figure 4.5: Comparison of methods to generate arbitrary kernel sizes: 5×5, a) *x*-profile of vector gradient sum method, b) *x*-profile of convolution method.

The purpose of comparing the two edge kernels in this section has been to demonstrate that the kernel weights of an edge detectors are subjective to the application. In figure 4.5a the slope of the inner equidistant radii changes less rapidly than the slopes of the kernel in figure 4.5b. This highlights the inclusion of the triangle filter into the kernel.

4.3.3 Sub-Pixel Gradient Directions

Standard edge kernels operate over an interval $\Delta\theta = 45^{\circ}$, using Eq. (4.2) kernel masks at $\pm\Delta\theta$ can be determined by a circular shift of the coefficient weights by 1 position within the equidistant radii. However, for $\Delta\theta = \arctan(1/2)$ the coefficient weights must shift by a 1/2 to remove the systematic error of the cartesian grid at the 1/2 position. This source of error is described at the beginning of section 4.2. Therefore, to yield an arbitrary kernel gradient direction, it is sensible to restate Eq. (4.2) in polar coordinates. Before this, however, integer coefficient sub-pixel kernels for $\Delta\theta = \arctan(1/2)$ are formulated to illustrate how a 1/2-pixel shift can occur.

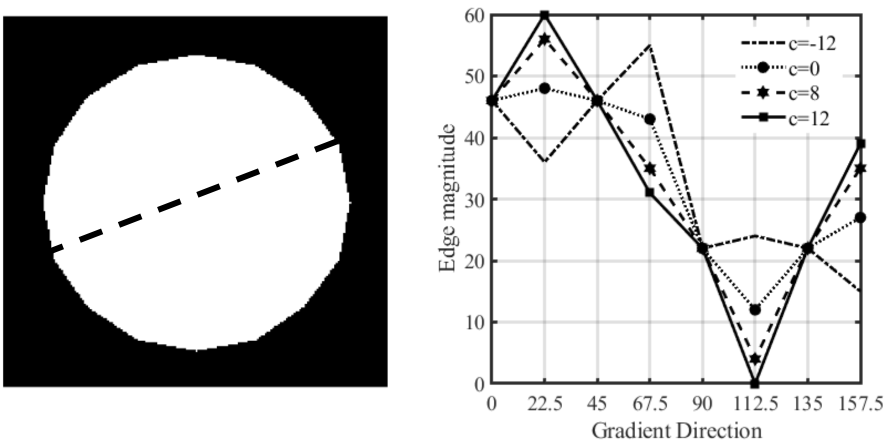
Consider the integer coefficients, using the convolution method, for a 5×5 kernel with a line of symmetry at 45° (the kernel is not normalised as presented in figure 4.4b)

$$D_{\theta=45^{\circ}} = \begin{bmatrix} 0 & 2 & 1 & 4 & 6 \\ -2 & 0 & 8 & 12 & 4 \\ -1 & -8 & 0 & 8 & 1 \\ -4 & -12 & -8 & 0 & 2 \\ -6 & -4 & -1 & -2 & 0 \end{bmatrix}.$$

For a circular shift right by 1-pixel of the outer equidistant radii, radial symmetry is completely defined on the cartesian grid weighted by a vector magnitude away from the centre. The vector magnitude is equal to $\sqrt{5}$. For an equal circular shift of the inner equidistant radius, the point on the discretely sampled line of symmetry is not uniquely defined,

$$\mathrm{D}_{\theta=22.5^{\circ}} = \begin{bmatrix} 2 & 0 & -2 & -1 & -4 \\ 1 & c_{\theta}^{+}(i-1, j-1) & -8 & -12 & -6 \\ 4 & 8 & 0 & -8 & -4 \\ 6 & 12 & 8 & c_{\theta}^{-}(i+1, j+1) & -1 \\ 4 & 1 & 2 & 0 & -2 \end{bmatrix}.$$

The kernel weights $c_{\theta}^{+}(i-1, j-1)$ and $c_{\theta}^{-}(i+1, j+1)$ indicate the position of the required kernel weight. Lying within the range of the known kernel weight values, figure 4.6b reveal the effects on symmetry in the differentiator for a 1/2-pixel shift in the kernels inner equidistant radii by tracking the value of c_{θ} . The dotted line in figure 4.6a indicates the angle at which the edge detector is applied to the test function.



(a) (b)

Figure 4.6: a) test image, b) profile of edge magnitude for $c_{\theta} = -12, 0, 8, 12$ and an edge gradient at 22.5°.

When $c_{\theta} = -12$, the gradient direction with the maximum edge magnitude is shifted to 67.5° from 22.5°, whereas for $c_{\theta} \geq 0$ the gradient direction for the maximum edge magnitude is maintained at 22.5°. This is a factor of the sign either side of symmetry in the differentiator. Since a second-order polynomial fits the profile of the maximum gradients curve, the distortion to the parabola shape of the function is critical. For $c_{\theta}=0$ the function is smoother at the peak location, hence low in sensitivity to changes in neighboring points (P). At $c_{\theta}=12$ the peak is prominent but not all measured points are orthogonally symmetric. This is true for this value of c_{θ} at $P\left(67.5^{\circ}\right)$, $P\left(157.5^{\circ}\right)$. Orthogonal symmetry is maintained throughout for $c_{\theta}=sign\left(8\right)$, this is confirmed from the phase response in figure 4.7 for the values of $c_{\theta}^{\pm}$.

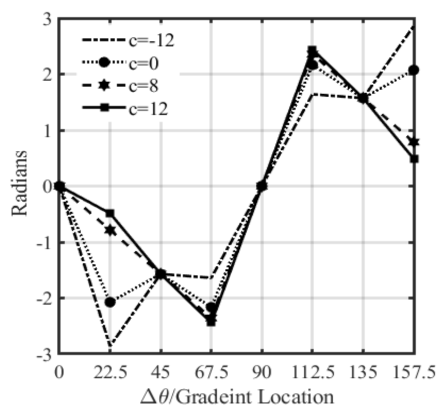


Figure 4.7: Phase gradient response at a 22.5° gradient for $c_{\theta}=-12,0,8,12$.

The gradient direction in figure 4.6-7 is specified and calculated at 22.5° and the kernel is configured to detect a 26.57° gradient. Hence, the cartesian grid systematically enforces a symmetry error for $\Delta\theta=22.5^{\circ}$.

Restating the vector sum gradient in polar coordinates allows any coefficient weight for any angle to be defined. Thus, the fixed error of the cartesian grid can be easily sidestepped. The gradient $\mathbf{G}$ defined in Eq. (4.1) in polar coordinates is

$$\mathbf{G}_{\boldsymbol{\theta}} = \frac{\left(\cos(\theta), \sin(\theta)\right)}{\mathbf{r}} \cdot \frac{\left\langle kx_i, ky_j \right\rangle}{\mathbf{r}}, \tag{4.5}$$

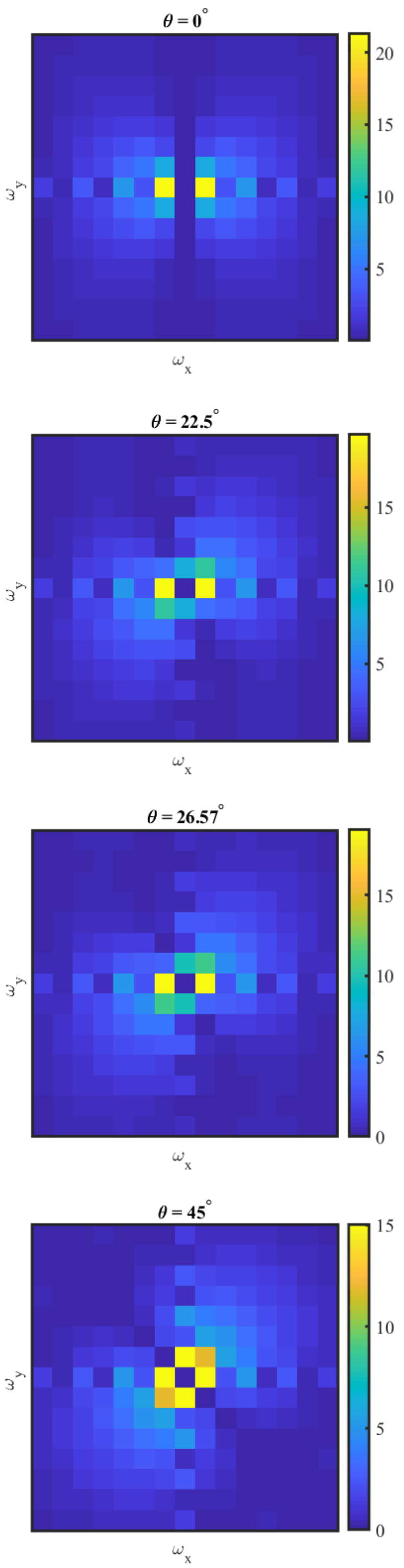
where kx_i and ky_j are the cartesian distances from the center of the kernel. For $\theta = 0^{\circ}, 90^{\circ}$ the cartesian vector components $\mathbf{G_X}$ and $\mathbf{G_Y}$ become

$$\begin{aligned} \mathbf{G}_{\boldsymbol{\theta}=0^{\circ}} &= \frac{\cos(\theta)}{(\mathbf{r})^2} \cdot \left\langle kx_i \right\rangle, \\ \mathbf{G}_{\boldsymbol{\theta}=90^{\circ}} &= \frac{\sin(\theta)}{(\mathbf{r})^2} \cdot \left\langle ky_j \right\rangle. \end{aligned} \tag{4.6}$$

Since the radius is a factor of the cartesian distances $kx_i / (\mathbf{r})^2$ and $ky_j / (\mathbf{r})^2$, the inner products can simply be expressed as $kx_i / kx_i^{\,2} + ky_i^{\,2}$ and $ky_i / kx_i^{\,2} + ky_i^{\,2}$. This results in an arbitrary kernel gradient direction algorithm:

$$\mathbf{G}_{\boldsymbol{\theta}} = \frac{\left(kx_i \cos(\theta) + ky_j \sin(\theta)\right)}{kx_i^2 + ky_i^2}. \tag{4.7}$$

To demonstrate Eq. (4.7), the magnitudes of the filter kernels frequency spectrum for $\theta = 0^{\circ}, 22.5^{\circ}, 26.57^{\circ}, 45^{\circ}$ are shown in figure 4.8. The kernel size is enlarged to 15×15 so that there is greater visual clarity in the kernel's line of symmetry between non-integer gradient directions.



$$\begin{matrix} -0.2500 & -0.2000 & 0 & 0.2000 & 0.2500 \\ -0.4000 & -0.5000 & 0 & 0.5000 & 0.4000 \\ -0.5000 & -1.0000 & 0 & 1.0000 & 0.5000 \\ -0.4000 & -0.5000 & 0 & 0.5000 & 0.4000 \\ -0.2500 & -0.2000 & 0 & 0.2000 & 0.2500 \end{matrix}$$

$$\begin{matrix} -0.1353 & -0.0317 & 0.1913 & 0.3378 & 0.3266 \\ -0.2930 & -0.2706 & 0.3827 & 0.6533 & 0.4461 \\ -0.4619 & -0.9239 & 0 & 0.9239 & 0.4619 \\ -0.4461 & -0.6533 & -0.3827 & 0.2706 & 0.2930 \\ -0.3266 & -0.3378 & -0.1913 & 0.0317 & 0.1353 \end{matrix}$$

$$\begin{matrix} -0.1118 & 0.0000 & 0.2236 & 0.3578 & 0.3354 \\ -0.2683 & -0.2235 & 0.4473 & 0.6708 & 0.4472 \\ -0.4472 & -0.8944 & 0 & 0.8944 & 0.4472 \\ -0.4472 & -0.6708 & -0.4473 & 0.2235 & 0.2683 \\ -0.3354 & -0.3578 & -0.2236 & -0.0000 & 0.1118 \end{matrix}$$

$$\begin{matrix} 0 & 0.1414 & 0.3536 & 0.4243 & 0.3536 \\ -0.1414 & 0 & 0.7071 & 0.7071 & 0.4243 \\ -0.3536 & -0.7071 & 0 & 0.7071 & 0.3536 \\ -0.4243 & -0.7071 & -0.7071 & 0 & 0.1414 \\ -0.3536 & -0.4243 & -0.3536 & -0.1414 & 0 \end{matrix}$$

Figure 4.8: Magnitude spectrum of filter kernels using Eq. (4.7) for $\theta = 0^\circ, 22.5^\circ, 26.57^\circ, 45^\circ$ and corresponding 5×5 coefficient weights.

4.4 Region Maxima (RM) Edge detection Algorithm

Obtained from a single pixel or region of pixels from an image, the magnitude of an edge is extracted using gradient kernels such as those presented in the last section. Its orientation is refined by fitting the sampled neighbouring points with a second order polynomial. Implementing the edge detection kernels using the maxima of a pixel intensity or distribution of intensities in a region is explored in this section. This will determine an optimum edge algorithm using an extended gradient interval.

The algorithm process is as follows: an input image $I(x, y)$ is convolved with a set of gradient filters, $D_\theta(i, j)$, spanning π with an angular interval $\Delta\theta$; thereby concatenated to obtain a 3D matrix of gradient images $Lmap(x_i, y_j, \theta_k)$. As discussed in section 2.2.2 standard implementations of edge detection algorithms set $\Delta\theta = 45^\circ$, here following Eq. (4.7) $\Delta\theta = 22.5^\circ$. As shown in figure 4. 9, the maximum value within a local region $Ir(n_i, n_j)$ along the dimension of orientated filters θ_k is extracted and thus stored to form an output edge image.

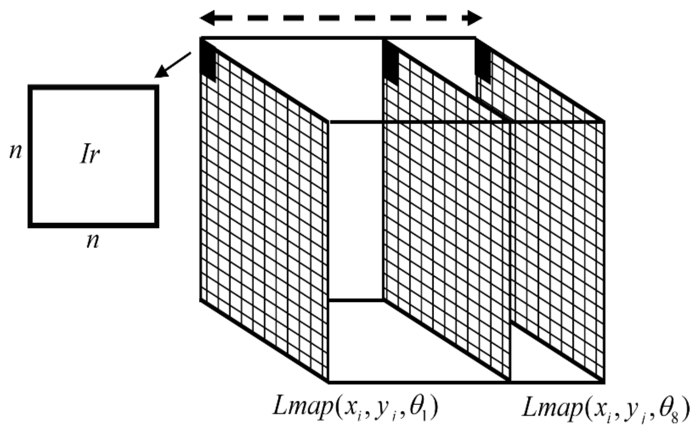


Figure 4.9: Concatenated set of gradient filters $Lmap(x_i, y_j, \theta_k)$ and local maxima extraction.

Once the edge image is formed, a refined edge image is constructed by forming a parabolic curve for every pixel of interest, using the adjacent pixels of the interest point. The curve is evaluated using a second order polynomial to identify the parabola's vertex, thus obtaining the true edge magnitude and location. Hence, the refined gradient direction is simply a transformation of the location proportional to the predefined gradient interval $\Delta\theta$. How many sampling points are required is discussed in the next section? Using a computer-generated circle ($r = 40$) as an input function and an edge kernel width of five pixels, the algorithms processes and outcomes are demonstrated in figures 4.10 – 11.

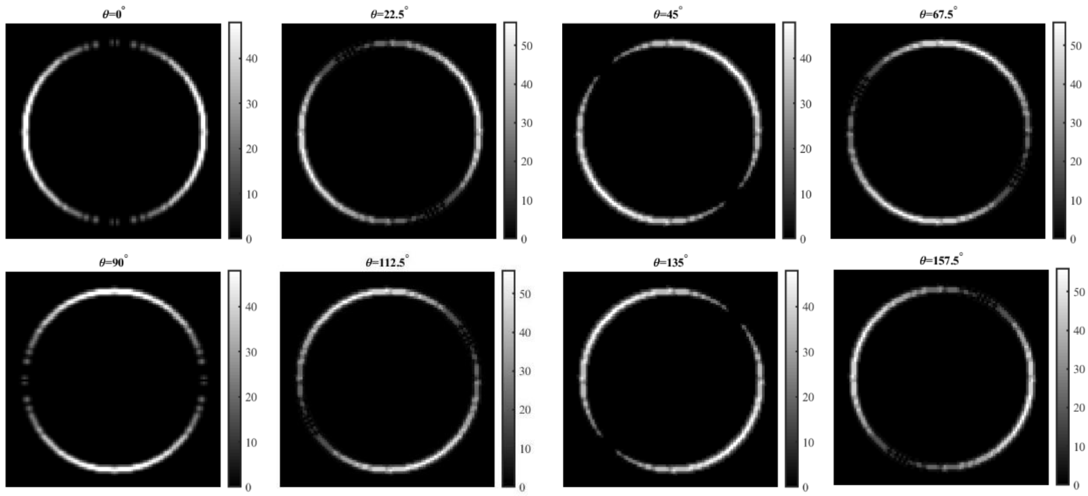


Figure 4.10: Gradient magnitude images forming $Lmap(x_i, y_j, \theta_k)$.

For an image region of one pixel only, the algorithm scans θ_k pixel-by-pixel. Thus, in this example fixing the minimum detectable edge-to the width of the edge kernel. The resultant gradient magnitude and direction in figure 4.11 shows that by selecting the image pixel $Ir(n_i, n_j)$ along θ_k, corresponding to the maximum magnitude, the optimum representation of the circular edge is obtained.

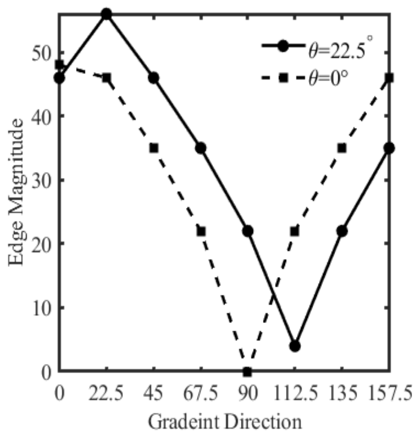


Figure 4.11: Location and value of $P_{pk-1} : P_{pk+1}$ and $P_{pk-2} : P_{pk+2}$ of $Lmap(x_i, y_j, \theta_k)$ for $I(r,\theta): r = 40$ and $\theta = 0^\circ, 22.5^\circ$.

Consider a line distorted by noise, by including more sampling points the accuracy of the line's reconstruction improves. For a curve which is symmetric, 3-points would reconstruct the curve perfectly, hence reveal the location and value of the true vertex (P_{pk}). Increasing the number of points would theoretically increase the accuracy of the location and value of the true vertex. These observations suggest that the vertex location and amount of distortion on the parabolic curve is due to the amount of noise at each sub-pixel gradient point. For the gradient interval $\Delta\theta = 45^\circ$ the mean value across all points is fixed. The configured sub-pixel kernel such as that shown in figure 4.8, has a P_{pk} polynomial point value fixed by non-integer coefficient values $c_\theta^\pm(i,j)$. For a kernel configuration with a gradient interval $\Delta\theta = 22.5^\circ$ across $Lmap(x_i, y_j, \theta_k)$, all adjacent points used to construct the second-order polynomial curve are symmetric and do not change unless either: $c_\theta^\pm(i,j)$ varies or system noise levels increase. Hence, if $P_{pk-1,-2}$ and $P_{pk+1,+2}$ are equal then there is no effect or apparent improvement to the accuracy of the gradient location for the extracted value of P_{pk}. This situation based on an iterative refinement of P_{pk} to maintain symmetry, develops the RM edge detection algorithm to enable a more accurate and precise estimate of the detected edges gradient.

4.5 RM Edge Detection Analysis

This section investigates the accuracy and precision of the RM edge detection algorithm for one coarse iteration perturbed by the noise model in appendix A1. In line with the approach towards sub-pixel edge detection outlined in this chapter, a coarse estimate of the edge gradient direction is analysed with two fixed sampling intervals $\theta = 22.5^{\circ}, 45^{\circ}$ over the angular range $\theta = 0^{\circ} : \Delta\theta : 45^{\circ}$. To analyse the RM edge detector, three model variants according to gradual theoretical improvements in sub-pixel measurement precision are scripted, compared and analysed in MATLAB. The variants are: *1*) $\Delta\theta = 45^{\circ}$ with 3 polynomial sampling points, *2*) $\Delta\theta = 22.5^{\circ}$ with 3 polynomial sampling points and, *3*) $\Delta\theta = 22.5^{\circ}$ with 5 sampling points. The test image that this analysis is performed on is that displayed in figure 4.5; a 16-sided polygon. The radius of the polygon at the intersect points is 40 pixels and each intersect point over the angular range is sampled 100 times to obtain an average measurement.

4.5.1 Sensitivity of Polynomial Sampling Points

The gradient error response for each model variant is presented in tables 4.1 - 4.6 over the specified angular range distorted by photon shot noise (PSN) and low frequency noise $S(f)$.

Table 4.1: Gradient error distorted by PSN for $\Delta\theta = 45^{\circ}$ and 3-sampling points

$\sqrt{N_{pht}}$	***Gradient Error*** $\sigma(\theta)$		
	$\theta = 0^{\circ}$	$\theta = 22.5^{\circ}$	$\theta = 45^{\circ}$
3.16	4.501	6.254	4.952
10	1.050	2.582	1.334
31.6	0.344	0.733	0.489
70.7	0.155	0.315	0.214
100	0.121	0.208	0.141
141.4	0.081	0.154	0.092

Table 4.2: Gradient error distorted by $S(f)$ for $\Delta\theta = 45^{\circ}$ and 3-sampling points

$S(f)_{\%}$	***Gradient Error*** $\sigma(\theta)$		
	$\theta = 0^{\circ}$	$\theta = 22.5^{\circ}$	$\theta = 45^{\circ}$
100	6.789	8.248	7.370
75	4.518	5.992	5.033
50	3.080	3.947	3.463
25	1.779	2.278	1.793
10	0.682	0.794	0.606
1	0.064	0.086	0.067

Table 4.3: Gradient error distorted by PSN for $\Delta\theta = 22.5^{\circ}$ and 3-sampling points

$\sqrt{N_{pht}}$	Gradient Error $\sigma(\theta)$		
	$\theta = 0^{\circ}$	$\theta = 22.5^{\circ}$	$\theta = 45^{\circ}$
3.16	7.596	2.137	8.252
10	2.260	0.640	2.977
31.6	0.711	0.200	0.963
70.7	0.266	0.094	0.330
100	0.225	0.063	0.250
141.4	0.160	0.040	0.188

Table 4.4: Gradient error distorted by $S(f)$ for $\Delta\theta = 22.5^{\circ}$ and 3-sampling points

$S(f)_{\%}$	Gradient Error $\sigma(\theta)$		
	$\theta = 0^{\circ}$	$\theta = 22.5^{\circ}$	$\theta = 45^{\circ}$
100	11.074	3.056	10.95
75	9.952	1.837	8.722
50	6.908	1.323	6.567
25	4.290	0.626	3.982
10	1.682	0.257	1.483
1	0.153	0.025	0.145

Table 4.5: Gradient error distorted by PSN for $\Delta\theta = 22.5^{\circ}$ and 5-sampling points

$\sqrt{N_{pht}}$	***Gradient Error*** $\sigma(\theta)$		
	$\theta = 0^{\circ}$	$\theta = 22.5^{\circ}$	$\theta = 45^{\circ}$
3.16	4.266	4.174	3.629
10	1.270	1.351	1.126
31.6	0.422	0.442	0.352
70.7	0.187	0.185	0.151
100	0.114	0.123	0.116
141.4	0.097	0.096	0.079

Table 4.6: Gradient error distorted by $S(f)$ for $\Delta\theta = 22.5^{\circ}$ and 5-sampling points

$S(f)_{\%}$	***Gradient Error*** $\sigma(\theta)$		
	$\theta = 0^{\circ}$	$\theta = 22.5^{\circ}$	$\theta = 45^{\circ}$
100	7.727	5.518	7.143
75	5.665	3.313	4.932
50	3.351	2.641	3.229
25	1.630	1.199	1.896
10	0.738	0.517	0.642
1	0.069	0.054	0.067

The tabulated error measurements of both noise distributions refer to two patterns of behaviour, firstly, a gradual decrease in measurement error of the intersect point of the polygon at 22.5° when the gradient interval increases from $\pi/4$ to $\pi/8$. For $\Delta\theta = 22.5^{\circ}$ there is an increase in error for the intersect points 0° and 45° when 3-points were used to sample the polynomial curve. Secondly, when 5-points were used to sample the polynomial curve, the errors are random and stable across all intersect points of the polygon. These error patterns are highlighted in figure 4.12 - 4.13 for photon shot noise and low frequency noise, respectively.

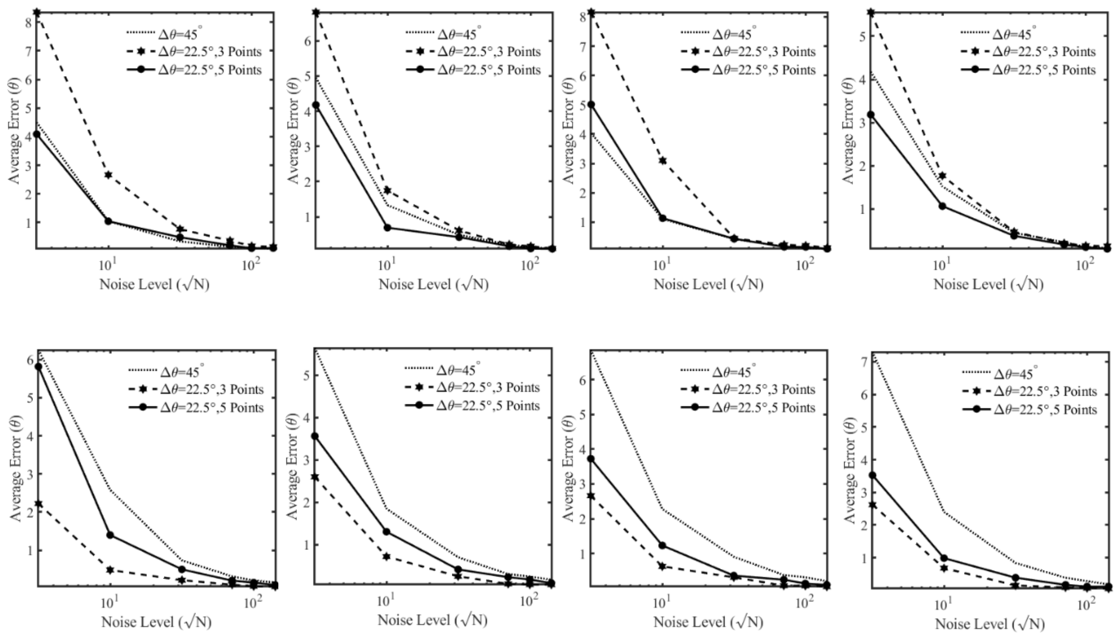


Figure 4.12: Sensitivity of polynomial edge reconstruction to photon shot noise, $\sqrt{N_{pht}} = \{3.16, 10, 31.6, 70.7, 100, 141.4\}$; $45^{\circ}, 22.5^{\circ}$ interval using 3 and 5-points. Top row: pixel level; gradients $0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}$. Bottom row: sub-pixel level gradients $22.5^{\circ}, 67.5^{\circ}, 112.5^{\circ}, 157.5^{\circ}$.

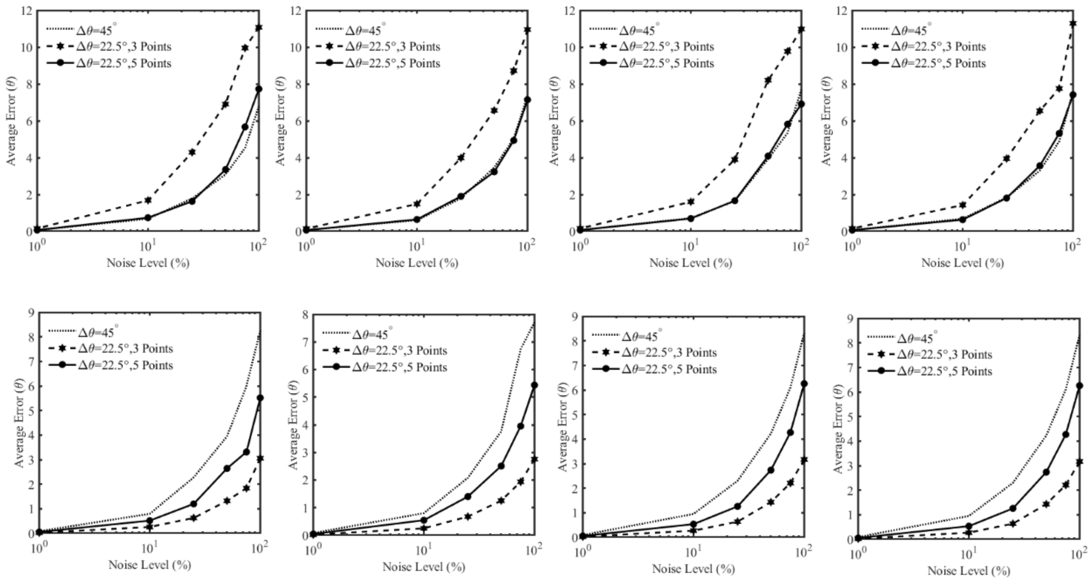


Figure 4.13: Sensitivity of polynomial edge reconstruction to low frequency noise, $1\%,10\%,25\%,50\%,75\%,100\%$ of signal amplitude at 3dB; $45^{\circ}, 22.5^{\circ}$ interval using 3 and 5-points. Top row: pixel level; gradients $0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}$. Bottom row: sub-pixel level gradients $22.5^{\circ}, 67.5^{\circ}, 112.5^{\circ}, 157.5^{\circ}$.

The analysis of a second-order polynomial used to refine the gradient estimate at a specified point has been performed to identify an operating range. Most useful, perhaps, is the use of the RM edge detection algorithm to detect boundaries of curve linear object features for applications requiring a precision of $<1^{\circ}$ error: SNR >31.6 and $S(f)<10\%$. For $\Delta\theta = 22.5^{\circ}$ using 3-sampling points the operating range increases to: $SNR>10$ and $S(f)<25\%$ for multiples of edge gradient orientations of 22.5°. However, the analysis has revealed and supported from earlier discussions, that the number of points used to sample the parabola only enhances the measurement of the vertex's position when the points are appropriately spaced and above the noise level. Hence, equal sampling separation and symmetry are the critical components of this analysis. The application of the RM algorithm in this analytical example is done so with little regard for the accuracy of the actual test function. A circle was not used in this analysis because points along the perimeter of the function may be sampled more within the test function, creating flat spots, such as 0° and 45°, which are definitely defined on a cartesian grid as opposed to 22.5° which is not.

A standard edge detector interval is 45° and the RM algorithm, using Eq. (4.7), has demonstrated a (coarse) gradient estimate using a 22.5° kernel interval. For an arbitrary generation of a gradient kernel, a logical progression of this algorithm is to iterate the same procedure to a fine estimate of a gradient's direction. Using an iterative procedure, in this manner, requires two considerations. Firstly, for realistic application, prior knowledge of potential gradient orientations of the feature of interest is required. Secondly, an accurate test function confined on a Cartesian grid is required to measure the precision of the iteration. The next section investigates this extension to the RM edge detection algorithm.

4.6 Iterative RM Edge Detection Algorithm

It may not be necessary to have a more precise estimate of an edge's gradient, particularly for the detection of macro sized object boundaries in images. Such as animals, vehicles or household items. However, for niche applications where the feature of interest is at a smaller scale, i.e. micro and nanoscale, the measurements sensitivity to random and systematic errors is critical. Examples include semiconductor wafer layer alignment via fiducials, in photolithography fabrication techniques and strain deformation tracking and analysis in semiconductor wafer layers.

Justification for the development of the RM algorithm's accuracy and precision is founded on applications such as these.

Firstly, consider the coarse estimation of an edge's gradient direction obtained by equal separation ($\Delta\theta = 22.5^{\circ}$) of the polynomials data points. Ignoring the number of data points used to sample the parabolic curve for the moment, re-evaluating the magnitude of the vertex of P_{pk} but with data points closer to the coarse estimate would theoretically provide a more accurate estimate of the gradient direction. Likewise, additional refinement of this new peak value and location and so on…would increase the accuracy further. However, for the case of a gradient direction, that is uniquely described by the coarse kernel interval the separation and symmetry of the parabola is exact and has limited justification for the necessity of an iterative procedure. Whereas for gradient directions not defined uniquely by $\Delta\theta$ such as 23° or 27° etc…, the parabola would be skewed. By redefining $\Delta\theta$ in the next iteration so that a curve's symmetry is more accurately modelled, the accuracy in this case is relevantly justified. Distortion, due to random noise and the number of sampling points, forms part of the analysis that follows. However, first an accurate test function for this procedure is presented, whereby, the test function is equally important because a circle can only be approximated through tangential lines confined on a cartesian grid.

4.6.1 Generating Accurate Test Functions

To obtain an accurate measurement of a gradient direction using the RM method, a set of discrete accurate binary edges described by $\Delta\theta = \arctan(1/2)$ are generated (figure 4.14): note that the edge gradients are $90^{\circ} - \theta$. A cartesian grid can accurately contain sub-pixel edges at 14.04°, 18.4° and 26.57° for coordinates (x, y) equal to $(2,1)$, $(3,1)$ and $(4,1)$. To maintain an equal number of convolutions along the binary edge's step, spatial sampling along the image dimension is proportional to coordinate $(\cdot, y)$ that is initially set to 10 for (x, y) equal to $(1,1)$.

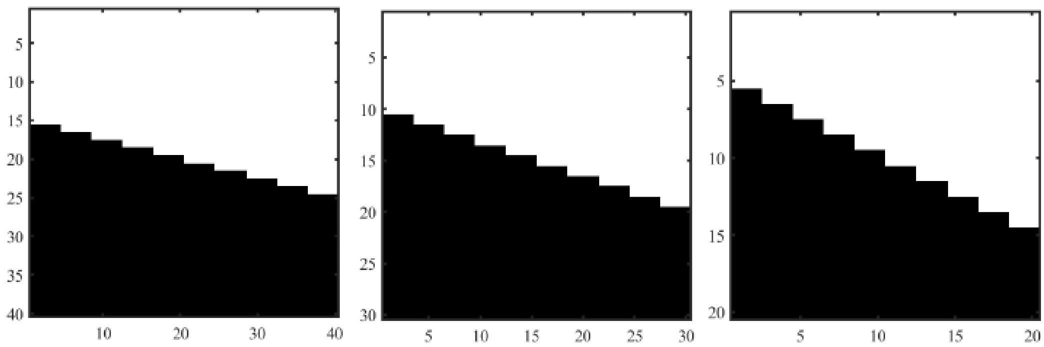


Figure 4.14: Test edge functions: from left to right $\theta = 75.96^{\circ}, 71.6^{\circ}, 63.4^{\circ}$; for coordinates (y, x) equal to $(4,1)$, $(3,1), (2,1)$.

4.6.2 Accuracy of Fine gradient Measurement

Since the step edges of the test functions increase for acute gradient orientations, the width of the feature causes the number of convolutions across the edge to increase. Thus, to appropriately sample the edges in the convolution for increases in *y* the kernel size increases. The accuracy of the iterative method is determined by measuring the error across the coarse and second iteration for $k^{n \times n}, n = 3, 5, 7, 9, 11, 13, 15$.

For the analysis that follows using the test functions in figure 4.14 the gradient interval of the coarse iteration is confined to $\Delta\theta = 22.5^{\circ}$ using 3-points to sample the polynomial. The second iteration is also evaluated with the interval $\Delta\theta = 22.5^{\circ}$ using 3-points. The measurement is taken at a single point located at the centre of the convolution and edge image. It is obvious that the width of the edge would increase along with *k;* which has effect in noise reduction at a cost of feature resolution. However, to attain an accurate measurement across gradient directions the size of *k* becomes important for sensitive measurements. The accuracy of measuring the edge gradient image directions $\theta = 75.96^{\circ}, 71.6^{\circ}, 63.4^{\circ}$ are presented in table 4.7 - 4.9. These measurements are the coarse recovered gradient $\theta_{r(c)}$, the absolute error of the coarse recovered gradient $\left|\varepsilon\left(\theta_{r(c)}\right)\right|$ and reciprocation of these for the second iteration: $\theta_{r(f)}$ and $\left|\varepsilon\left(\theta_{r(f)}\right)\right|$. A third measurement recorded is k_s which tracks the sampling ratio across *k* using the ratio of image width to kernel width k^n.

Table 4.7: Edge gradient direction: 63.4°.

k^n	$\theta_{r(c)}$	$\left\|\varepsilon\left(\theta_{r(c)}\right)\right\|$	$\theta_{r(f)}$	$\left\|\varepsilon\left(\theta_{r(f)}\right)\right\|$	k_s
3	63.48	0.08	63.44	0.04	6.67
5	64.63	1.23	64.59	1.19	4
7	64.89	1.49	64.85	1.45	2.86
9	65.07	1.67	65.04	1.64	2.22
11	65.16	1.76	65.14	1.74	1.82
13	65.24	1.84	65.21	1.81	1.54
15	65.28	1.88	65.26	1.86	1.33

Table 4.8: Edge gradient direction: 71.6°.

k^n	$\theta_{r(c)}$	$\left\|\varepsilon\left(\theta_{r(c)}\right)\right\|$	$\theta_{r(f)}$	$\left\|\varepsilon\left(\theta_{r(f)}\right)\right\|$	k_s
3	63.48	8.12	63.44	8.16	10
5	70.31	1.29	70.35	1.25	6
7	71.35	0.25	71.39	0.21	4.29
9	71.33	0.27	71.37	0.23	3.33
11	72.05	0.45	72.1	0.5	2.73
13	72.27	0.67	72.32	0.72	2.307
15	72.25	0.65	72.97	1.37	2

Table 4.9: Edge gradient direction: 75.96°.

k^n	$\theta_{r(c)}$	$\left\|\varepsilon\left(\theta_{r(c)}\right)\right\|$	$\theta_{r(f)}$	$\left\|\varepsilon\left(\theta_{r(f)}\right)\right\|$	k_s
3	63.48	12.48	63.44	12.52	13.33
5	70.31	5.65	70.35	5.61	8
7	74.04	1.92	74.1	1.86	5.7
9	74.9	1.06	74.95	1.01	4.44
11	74.9	1.06	74.96	1.0	3.63
13	75.32	0.64	75.37	0.59	3.07
15	75.88	0.08	75.92	0.04	2.67

For a square odd kernel there are equal distances away from the inspected pixel in all directions. For the edge test images, the optimum square odd kernel sizes for $\theta = 75.96^{\circ}, 71.6^{\circ}, 63.4^{\circ}$ are $15\times15, 7\times7$ and 3×3. The gradient direction accuracy of the detected edge after the second iteration is $\pm0.04^{\circ}$ when $(\cdot, y)$ is even and equal to 2 and 4, whereas for the value of $(\cdot, y)$ equal to 3 the accuracy decreases to $\pm0.21^{\circ}$. The decrease in accuracy is a result of the optimum kernel size for (x, y) equal to $(3,1)$, where the kernel width would have to impossibly become a non-integer value; approximately 5.3 . The proceeding analysis of precision for these edge directions is measured using the optimum k^{N} that have been identified in tables 4.7 - 4.9.

4.6.3 Sensitivity of Polynomial Sampling Points

The gradient error of the iterative RM edge detection algorithm is measured after the second iteration which is distorted by photon shot noise (PSN) and low frequency noise $S(f)$. In this analysis of the measurement's precision the gradient interval of the coarse iteration is confined to $\Delta\theta = 22.5^{\circ}$ with 3 and 5-points to sample the polynomial. The second iteration is evaluated using both 3 and 5-points but with an interval range $\Delta\theta = 1^{\circ} : 45^{\circ}$.

Consider the test function in figure 4.14 for 63.4°, the gradient kernels used to evaluate the gradient of the test function are $D_{\theta} = 45^{\circ}, 67.5^{\circ}, 90^{\circ}$ when the number of sampling points is 3, and $D_{\theta} = 22.5^{\circ}, 45^{\circ}, 67.5^{\circ}, 90^{\circ}, 112.5^{\circ}$ when they are 5. The main contributor to parabolic symmetry distortion is due to a gradient direction not specified by any of the coarse iterations' gradient kernels. The amount of noise (due to photon shot noise as shown in figure 4.15) on the curve changes the magnitude of the sampled points and the sensitivity of the measurement to the level of curvature about the peak above and below the noise threshold.

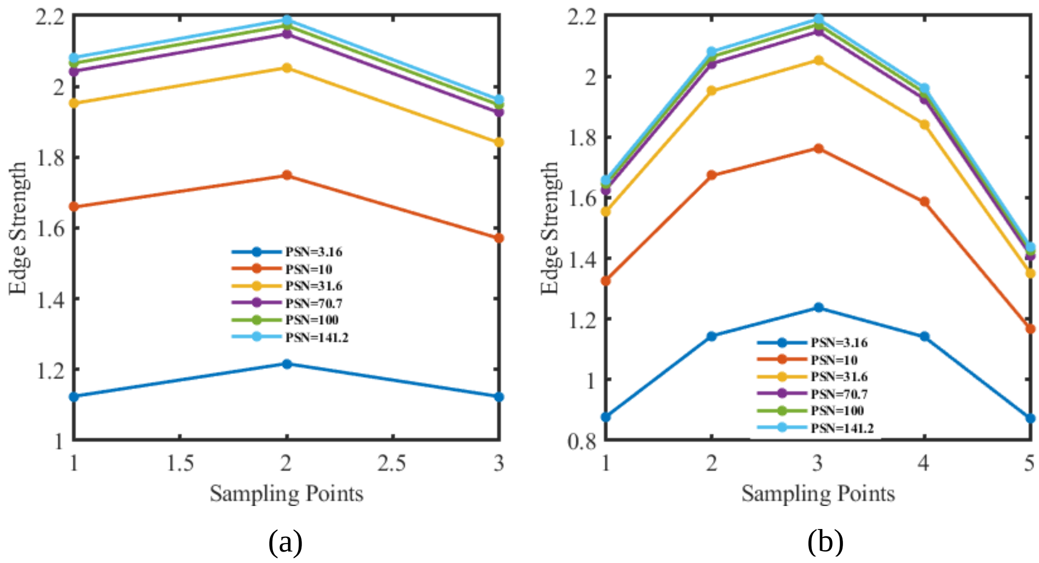
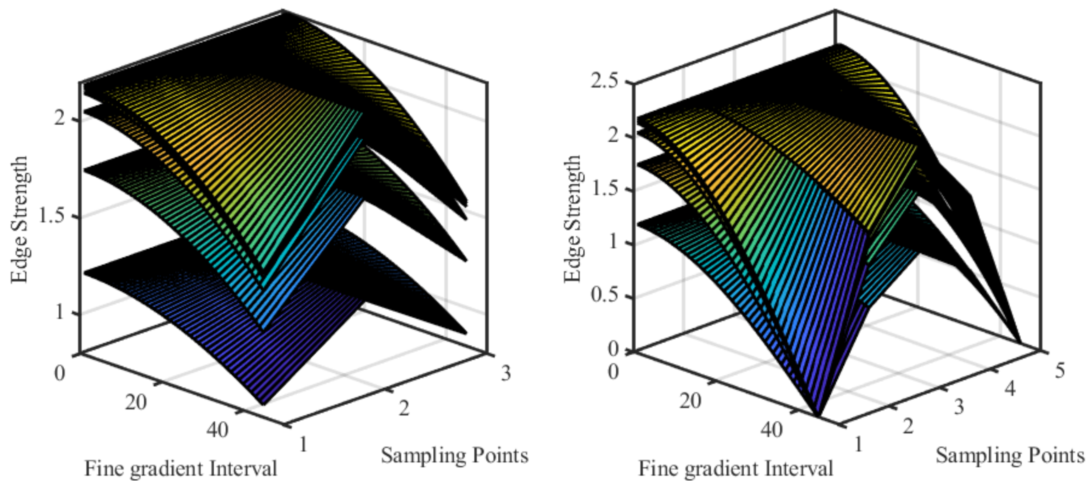


Figure 4.15: Coarse iteration curve distortion due to edge gradient direction 63.4° and photon shot noise (PSN) for: a) 3-points, b 5-points.

For both sampling intervals of the coarse iteration the error is systematic to the symmetry of the curve, hence including more points results in a more precise but less accurate measurement. In contrast to this, figure 4.16 shows the function that the second iteration measurement is extracted from for 3 and 5-points. A similar function exists for a second iteration extracted from a coarse estimate using 5-points; this is not illustrated as its purpose closely follows 3-points.



(a) (b)

Figure 4.16: Fine iteration curve distortion for edge gradient direction 63.4° and photon shot noise (PSN) for: a) 3-points, $\Delta\theta = 1^{\circ} : 45^{\circ}$, b) 5-points, $\Delta\theta = 1^{\circ} : 45^{\circ}$.

The fine gradient interval (illustrated in figure 4.16) demonstrates that for instances where the curve is relatively flat there is less sensitivity to measurement variation. This results in a second iteration that provides a method of optimising the attainable measurement based upon knowledge of the noise within the system. For a coarse measurement taken from a distorted polynomial curve, which is due to an edge direction in-between kernel gradient direction, the function is thus revaluated by a second-order polynomial that is more accurately symmetric.

For the precision analysis of the edge gradient 63.4° there are four model variants of the RM iterative edge detection algorithm : 1) coarse iteration $\Delta\theta = 22.5^{\circ}$ 3-point polynomial evaluation and second iteration $\Delta\theta = 1^{\circ} : 45^{\circ}$ using 3-points, 2) coarse iteration $\Delta\theta = 22.5^{\circ}$ 3-point polynomial evaluation and second iteration $\Delta\theta = 1^{\circ} : 45^{\circ}$ using 5-points, 3) coarse iteration $\Delta\theta = 22.5^{\circ}$ 5-point polynomial evaluation and second iteration $\Delta\theta = 1^{\circ} : 45^{\circ}$ using 3-points, 4) coarse iteration

$\Delta\theta = 22.5^{\circ}$ 5-point polynomial evaluation and second iteration $\Delta\theta = 1^{\circ} : 45^{\circ}$ using 5-points. This analysis characterizes the symmetry of the parabola for each model variant by measuring the *L1-norm* (Eq. (4.8)) of the curves' points ($P_{pk} \pm P_n, n = 1, 2$) $\Delta\theta = 1^{\circ} : 45^{\circ}$:

$$\chi\left(P\left(\Delta\theta\right)\right) = \left\|\left(P_{pk}\left(\Delta\theta\right) - P_{pk-n}\left(\Delta\theta\right)\right) - \left(P_{pk}\left(\Delta\theta\right) - P_{pk+n}\left(\Delta\theta\right)\right)\right\|. \quad (4.8)$$

The relationship between symmetry and the expected error of the precise measurement for each configuration of the RM edge detection algorithm will be determined. There is no precedent for a model variant that is optimum, because the signal dependencies, gradient angle and the level of noise, provide no basis of a consistent analysis for a comparison between a refined estimate of the coarse measurement. So that the accuracy of the measurement remains $< 1^{\circ}$ the systems' noise level operating range is maintained at: SNR > 31.6 and $S\left(f\right) \leq 10\%$.

The following figures present: figures 4.17.1-3 compares the noise symmetry analysis of the error range for model 1,2,3,4, for $\theta = 63.4^{\circ}$, $\theta = 71.6^{\circ}$, $\theta = 75.96^{\circ}$ distorted by photon shot noise; figure 4.18 replicate the comparison for the four model variants and gradient directions for the same order but distorted by low frequency noise. In figure 4.17.3 the effect of distortion and gradient angle to the polynomial model of the edge is also illustrated. Whereby, it is observed that the data trend line becomes cubic as gradients gradually become acuter. A full presentation of these symmetry curves are plotted in appendix 2.

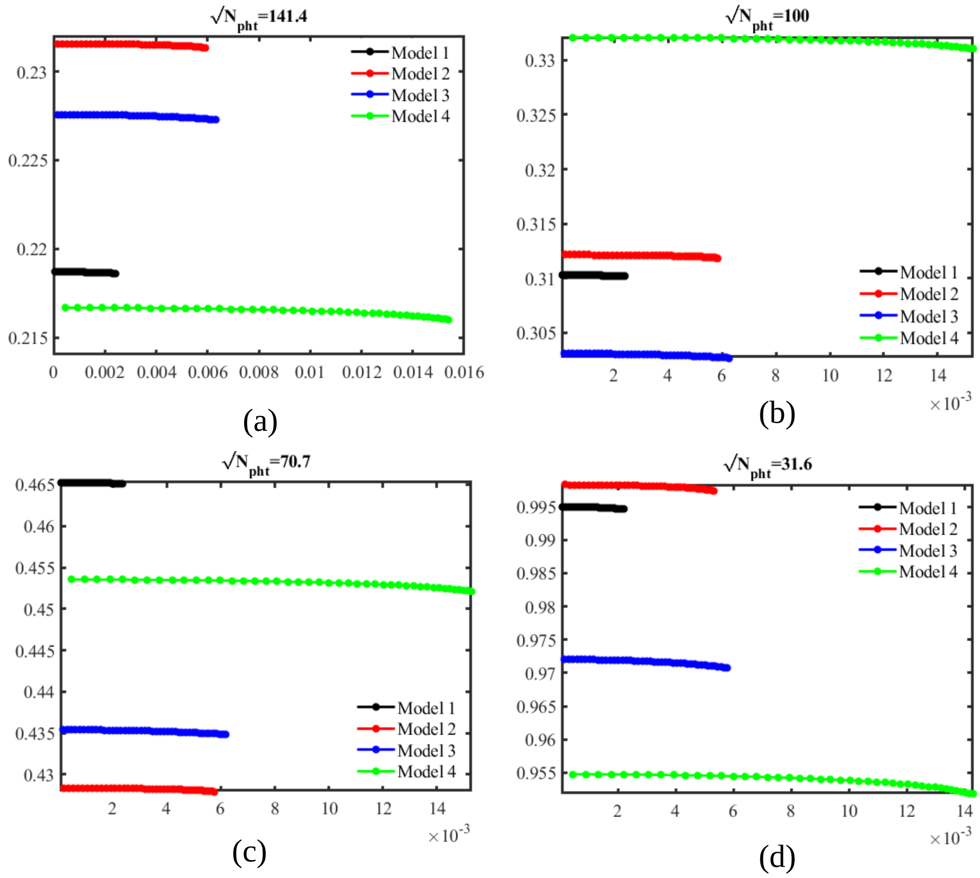


Figure 4.17.1: RM edge detection noise symmetry analysis for distortion to polynomial sampling points of iterative RM model 1 - 4; edge direction $\theta = 63.4^{\circ}$; distortion - PSN. a) $\sqrt{N_{pht}} = 141.4$, b) $\sqrt{N_{pht}} = 100$, c) $\sqrt{N_{pht}} = 70.7$, d) $\sqrt{N_{pht}} = 31.6$. Y-axis: Error $\sigma(\Delta\theta)$ vs X-axis: Symmetry $\chi(P(\Delta\theta))$.

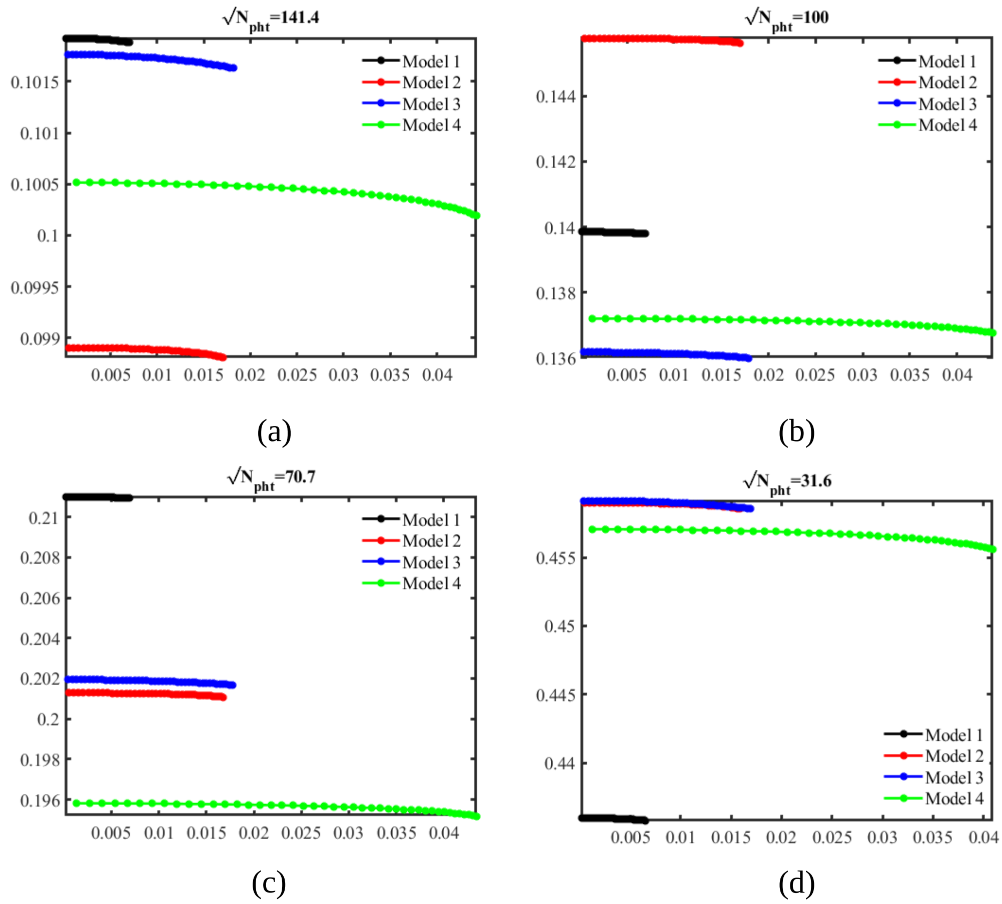


Figure 4.17.2: RM edge detection noise symmetry analysis for distortion to polynomial sampling points of iterative RM model 1 - 4; edge direction $\theta = 71.6^{\circ}$; distortion - PSN. a) $\sqrt{N_{pht}} = 141.4$, b) $\sqrt{N_{pht}} = 100$, c) $\sqrt{N_{pht}} = 70.7$, d) $\sqrt{N_{pht}} = 31.6$. Y-axis: Error $\sigma(\Delta\theta)$ vs X-axis: Symmetry $\chi(P(\Delta\theta))$.

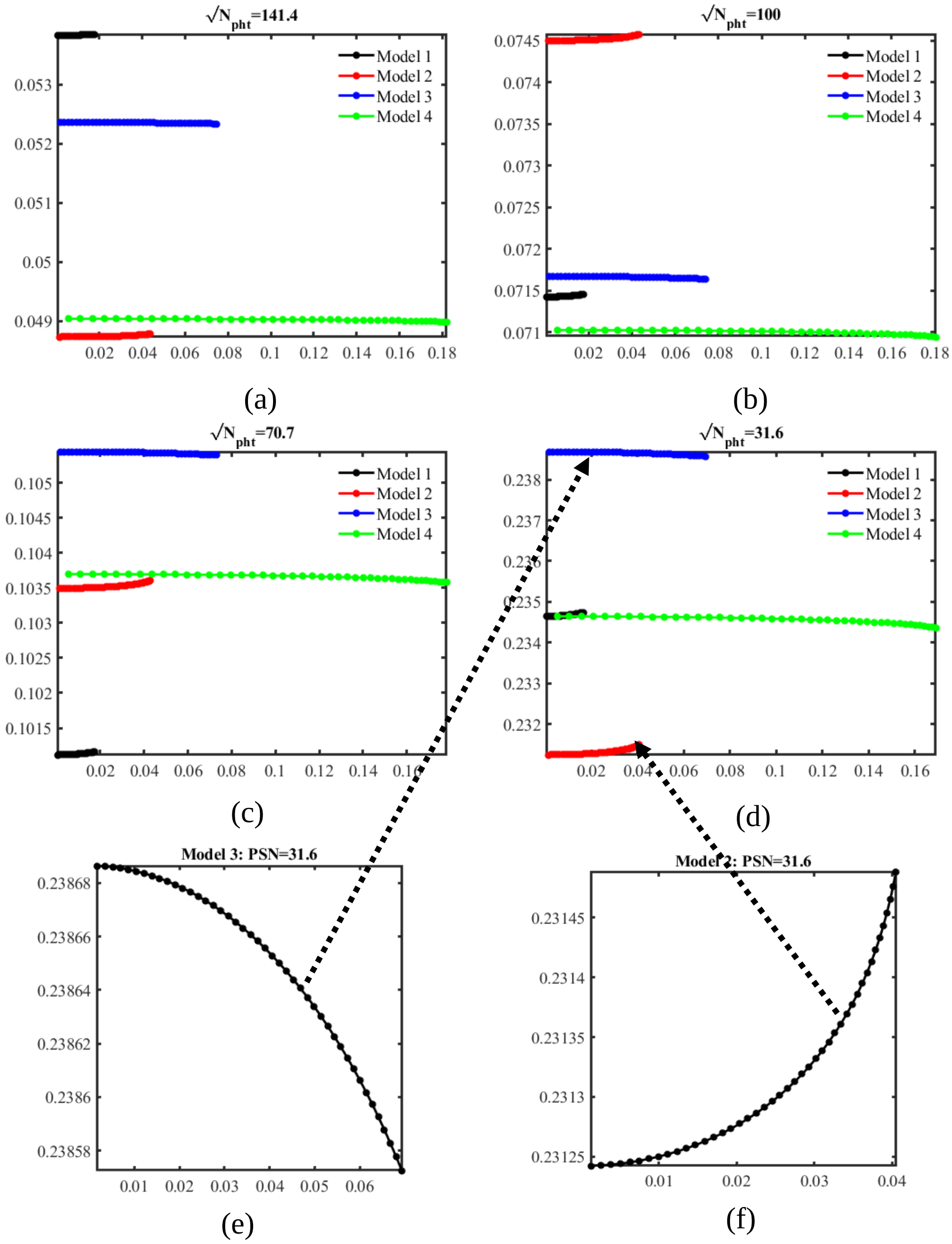


Figure 4.17.3: RM edge detection noise symmetry analysis for distortion to polynomial sampling points of iterative RM model 1 - 4; edge direction $\theta = 75.96^{\circ}$; distortion - PSN. a) $\sqrt{N_{pht}} = 141.4$, b) $\sqrt{N_{pht}} = 100$ c) $\sqrt{N_{pht}} = 70.7$, d) $\sqrt{N_{pht}} = 31.6$, e) second order error trend line, f) cubic error trend line . Y-axis: Error $\sigma(\Delta\theta)$ vs X-axis: Symmetry $\chi(P(\Delta\theta))$.

In the preceding figures the effect of $\Delta\theta$ on the symmetry and standard error is characterised against the gradient edge directions of the test functions and the configuration models of the iterative RM algorithm. The gradient directions of the test function become acuter as coordinate $(\cdot, y)$ increases. One effect of this increase is that the symmetry of the parabola formed by the kernels at the coarse estimate distorts to a function that can no longer be characterised by a second-order polynomial. In the case of evaluating $\theta = 75.96^{\circ}$ using 3-points, $P_{pk+1} > P_{pk}$ and the function begins to be characterised by a cubic term. This is evident in figure 4.17.3 for models' 1 and 2 when the trend is compared to figure 4.17.3d-e; which are evaluated at the coarse iteration using 3-points (model 1) and 5-points (model 2). The result of change to the polynomial order causes the standard error of the measurement to be proportional to the symmetry of polynomial curve. This is not evident for all remaining results in figures 4.17.1-3. The reason behind this, is that for $\Delta\theta = 1^{\circ}$ the distance between $P_{pk} \pm P_n,\, n = 1,2$ is smaller than it is for increases in $\Delta\theta$, thus the precision measurement of parabolic symmetry reduces as $\Delta\theta$ increases. Hence, it is sensible to state that the standard $\sigma(\Delta\theta)$ increases as symmetry $\chi(P(\Delta\theta))$ becomes poorer. The relationship between $\sigma(\Delta\theta)$ and $\chi(P(\Delta\theta))$ is therefore inversely proportional:

$$\sigma(\Delta\theta) \propto \frac{1}{\chi(P(\Delta\theta))} + error\,. \tag{4.9}$$

In the case of $\theta = 75.96^{\circ}$, this relationship yields that there must be more than 3-sampling points in the coarse iteration to appropriately evaluate such an acute angle. This response is echoed in figures 4.18 for low frequency noise distortion.

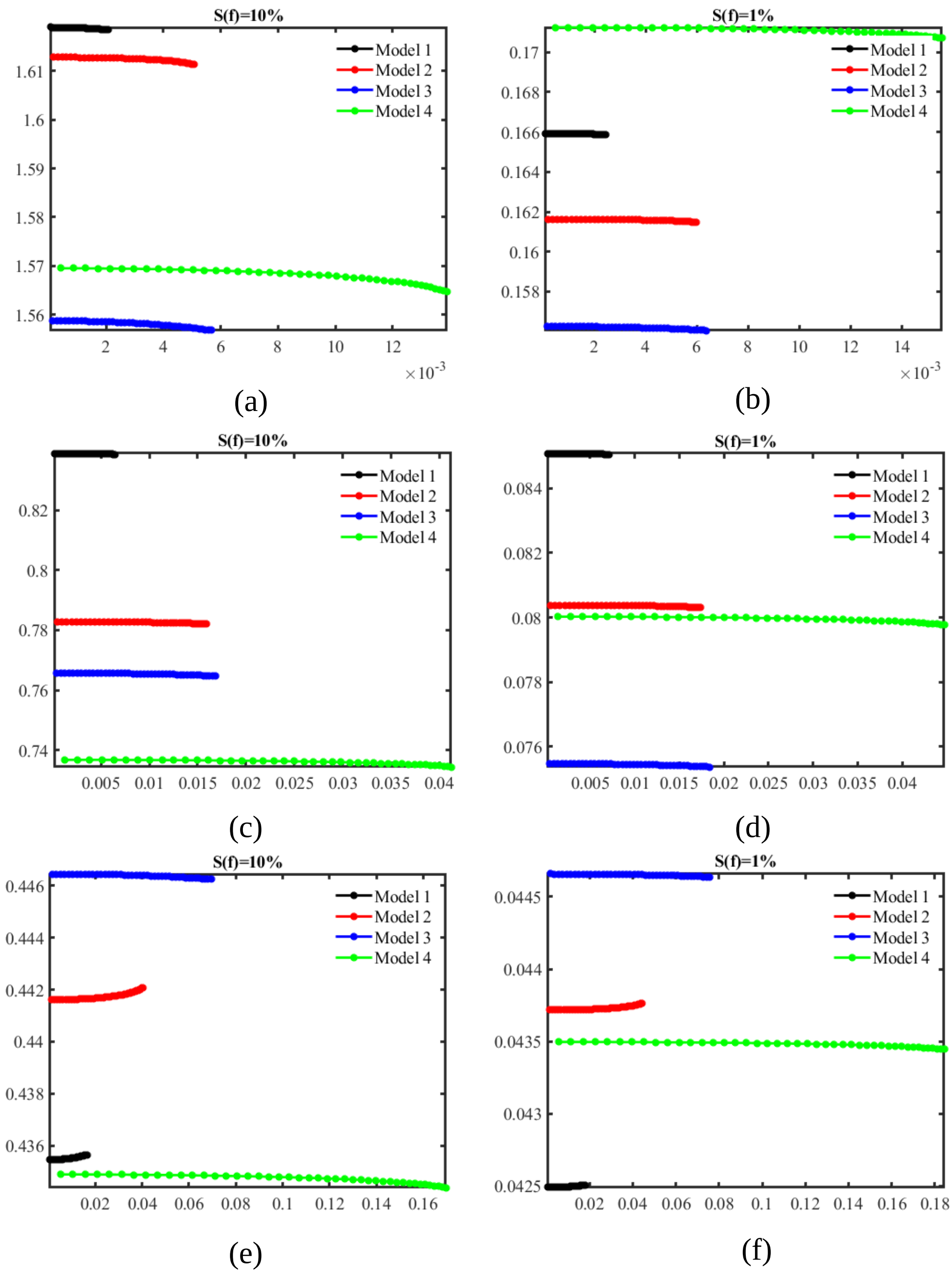


Figure 4.18: RM edge detection noise symmetry analysis for distortion to polynomial sampling points of iterative RM model 1 – 4; distortion - $S(f)$. a) $\theta = 63.4^{\circ}$: $S(f) = 10\%$, b) $\theta = 63.4^{\circ}$: $S(f) = 1\%$, c) $\theta = 71.6^{\circ}$: $S(f) = 10\%$, d) $\theta = 71.6^{\circ}$: $S(f) = 1\%$, e) $\theta = 75.96^{\circ}$: $S(f) = 10\%$, f) $\theta = 75.96^{\circ}$: $S(f) = 1\%$. Y-axis: Error $\sigma(\Delta\theta)$ vs X-axis: Symmetry $\chi(P(\Delta\theta))$.

The accuracy of the measurement for these test functions is dependent of on k^n, where large k^n increases the neighbourhood that the pixel of interest is evaluated from. However, for the optimum k^n that is identified in section 4.5.2 it is also shown that accuracy degrades for changes in k^n along with changes to the number of convolutions sampling the pixels of interest.

As stated previously there is not a single optimum configuration of coarse-to-fine polynomial evaluation via the iterative RM algorithm. One example of this is the case of acute gradients distorting the parabola beyond a second-order term. System and signal noise, photon shot noise and low frequency noise, presents the amount of error you can expect based on the degree of symmetry in the polynomial and the gradient interval spacing of the fine iteration. Based on the identified operating range for the noise sources, table 4.10 to 11 demonstrates the consistency of the error ranges for each model configuration and edge direction. Across all combinations of iterative RM model configuration, edge gradient direction and imposed noise levels the calibration between gradient error and symmetry extends in range as the noise level decreases. In exceptional conditions where the test function is well illuminated the precision measurement is of the order 10^{-4}. For the generated edge gradient directions this precision figure of merit extends to higher noise levels when the configuration of coarse to fine sampling of the polynomial is considered

Table 4.10: Photon shot noise error range: model configuration 1 – 4.

$\theta = 63.4^{\circ}$	$\sqrt{N_{pht}}$			
Model	31.6	70.7	100	141.4
1	3.7106 e^{-4}	1.8153 e^{-4}	1.2155 e^{-4}	8.5619 e^{-5}
2	9.2004 e^{-4}	4.1529 e^{-4}	2.9955 e^{-4}	2.2399 e^{-4}
3	0.0012	5.5781 e^{-4}	3.8897 e^{-4}	2.9312 e^{-4}
4	0.003	0.0014	0.001	6.8740 e^{-4}
$\theta = 71.6^{\circ}$				
1	1.8715 e^{-4}	9.09751 e^{-5}	6.0282 e^{-5}	4.3883 e^{-5}
2	0.0029	0.0013	9.3779 e^{-4}	6.0282 e^{-5}
3	5.9804 e^{-4}	2.6407 e^{-4}	1.7795 e^{-4}	1.3293 e^{-4}
4	0.0015	6.2942 e^{-4}	4.4215 e^{-4}	3.2395 e^{-4}
$\theta = 75.96^{\circ}$				
1	-	-	-	-
2	-	-	-	-
3	1.1379 e^{-4}	5.0163 e^{-5}	3.3985 e^{-5}	2.4877 e^{-5}
4	2.7881 e^{-4}	1.2150 e^{-4}	8.2940 e^{-5}	5.7409 e^{-5}

Table 4.11: Low frequency noise $S(f)$ error range: model configuration 1 – 4.

Model	$\theta = 63.4^{\circ}$		$\theta = 71.6^{\circ}$		$\theta = 75.96^{\circ}$	
	$S(f)$10%	$S(f)$1%	$S(f)$10%	$S(f)$1%	$S(f)$10%	$S(f)$1%
1	5.6095 e^{-4}	6.5231 e^{-4}	3.4798 e^{-4}	3.6768 e^{-5}	-	-
2	0.0015	1.5614 e^{-4}	0.0049	5.1591 e^{-4}	-	-
3	0.002	2.0087 e^{-4}	9.9651 e^{-4}	9.8699 e^{-5}	2.1099 e^{-4}	5.0357 e^{-4}
4	0.0048	5.4231 e^{-4}	0.0023	2.5772 e^{-4}	2.1188 e^{-5}	5.0987 e^{-5}

4.7 Summary

The aim of this chapter has been to analyse the sensitivity of arbitrary sub-pixel edge feature detection proposed in Eq. (4.7). The purpose of the developed algorithm (region maxima (RM)) is to isolate salient information with regard to an object boundary so that discriminative features lying in between integer sampled grid points can be measured. Observations from the analysis of the algorithm's sensitivity demonstrated that the extent of capability goes beyond the requirements of applications involving macro sized objects. The course of the investigation led to the characterisation of symmetry of the differentiator's coefficients. The expression in Eq. (4.8) concludes that symmetry is inversely proportional to standard error. The analysis showed that evaluating the edge using 3 and 5-sampled points between coarse and fine iterations of RM resolved the sensitivity of the technique down to 10^{-5}. Object and misalignment error tracking in manufacturing processes, particularly semiconductor layer fabrication stages, and medical image analysis operating at the micron scale are perhaps the most suitable applications at this degree of sensitivity.

Chapter 5 – Geometric Recognition of Non-Rigid Objects

The core investigation in chapter 4 analyses the method and accuracy of detecting an object's boundary information. In biological vision this feature is the initial analysis within the recognition task and as such, it is the common signal processing task to a machine/computer vision system. Based on this pre-processing step, the motivations for this chapter are founded from the utilisation of appropriate prior knowledge accumulated through the recognition process, and the discussions in section 2.4; the spatial relationship of object information and the linear geometric characterisation of object features using edges. This chapter's discourse coalesces towards opportunities in complex object recognition by focusing on a process of recognition that is initiated by edge feature extraction and measurement using the Hough Transform [23], [24], and linear analysis. Investigations are developed based on appropriate features and their stability for discriminative statistics in correlation. The object class cat is used throughout the proceeding investigations because of view invariance and object familiarity. This chapter begins by presenting the recognition method with a central discussion of systematic and random issues affecting recognition success.

5.1 Chapter Synopsis

Readers can refer to this synopsis of the research contained in chapter 5 when reading the investigations within this chapter.

This chapter describes the procedure used to recognise a non-rigid object; the aim is to build up a confidence recognition level through the object parts that are detected. This is achieved by fitting line and plane geometry to the simplified structure in the image. The beginning of this chapter brings together the system level process of an object recognition system and the research in chapter 2 and chapter 3: image registration and feature detection. A method of recognition using the Hough transform to extract feature geometries is discussed. For the example object (cat), critical features of the object parts are identified to measure and track in the outline method **(section 5.2)**. The functional routines for the training and test procedures show how prior knowledge is built into the recognition process. The focus areas of investigation are: the formation of image data suitable for the selected signal processing algorithms; a resolve of object view to select appropriate training templates; to track changes in the object through geometric inference. Whereby, inference is used to estimate a set of relationships between features and their linear geometries **(section 5.3)**. The recognition method is demonstrated for the simplest case, a front profile of a cat's head. Using anatomical measurements of the object features, a statistical analysis determines the reliability of those features **(section 5.4)**. A 3D model of a cat's head is used to ascertain further features to track the object horizontally and vertically. Using a series of ratio-of-ratio measures, a rotational map of the identified features is created. These measurements enable scale invariance of the linear distance measurements and an operating range for the discriminate feature ratios **(section 5.5)**. The purpose of tracking a cat's head is to establish the appropriate training profile to use in the feature detection process. The method is tested by selecting views that are contained within the training data and in-between the discretely sampled profiles. A further analysis inspects the contributions of individual feature ratios to the overall error.

To begin, a pictorial summary of this chapter's investigation is presented,

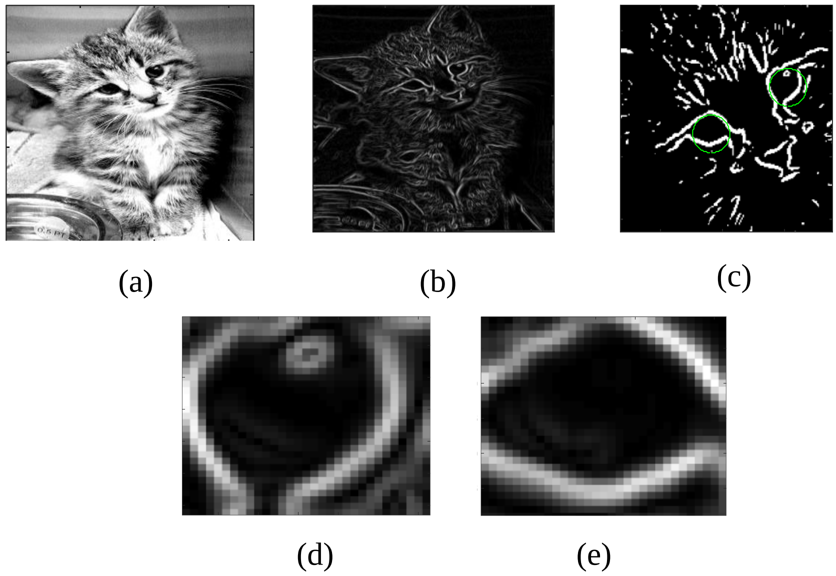


Figure 5.1: Application of Hough transform to extract circular features: a) grayscale image [151], b) RM edge image, c) threshold edge image: green circles indicate detection of circular object, d) image region – right, e) image region - left.

The image of a cat in figure 5.1a is pre-processed with the RM edge detection algorithm (figure 5.1b). After applying a threshold to convert the edge image to a binary image (figure 5.1c) the circular Hough transform is applied to the conditioned signal. Upon extraction of circular features, the image regions (figure 5.1d-e) are then probed to determine if there is an eye present. Following successful identification of this object part, the process is repeated utilising prior knowledge of the object's parts positional relationship to other critical parts and their specific planar geometric properties. This chapter develops a data-driven recognition routine based on identifying and tracking geometric properties of non-rigid objects. The application of the Hough transforms to challenging image scenes is investigated.

5.2 Recognition, System Level Processes

The object recognition system in figure 5.2 isolates the key processes that chapters 3 and 4 have addressed through their central aims. These aims are image registration and feature detection. The requirements of these processes for the prescribed recognition system are to identify regions of interest, detect the potential features and to suitably formulate obtained feature measurements input into a comparator to identify the recognition confidence level. The central focus of this chapter is the method of recognition, whereby signal processing issues are identified, presented and discussed.

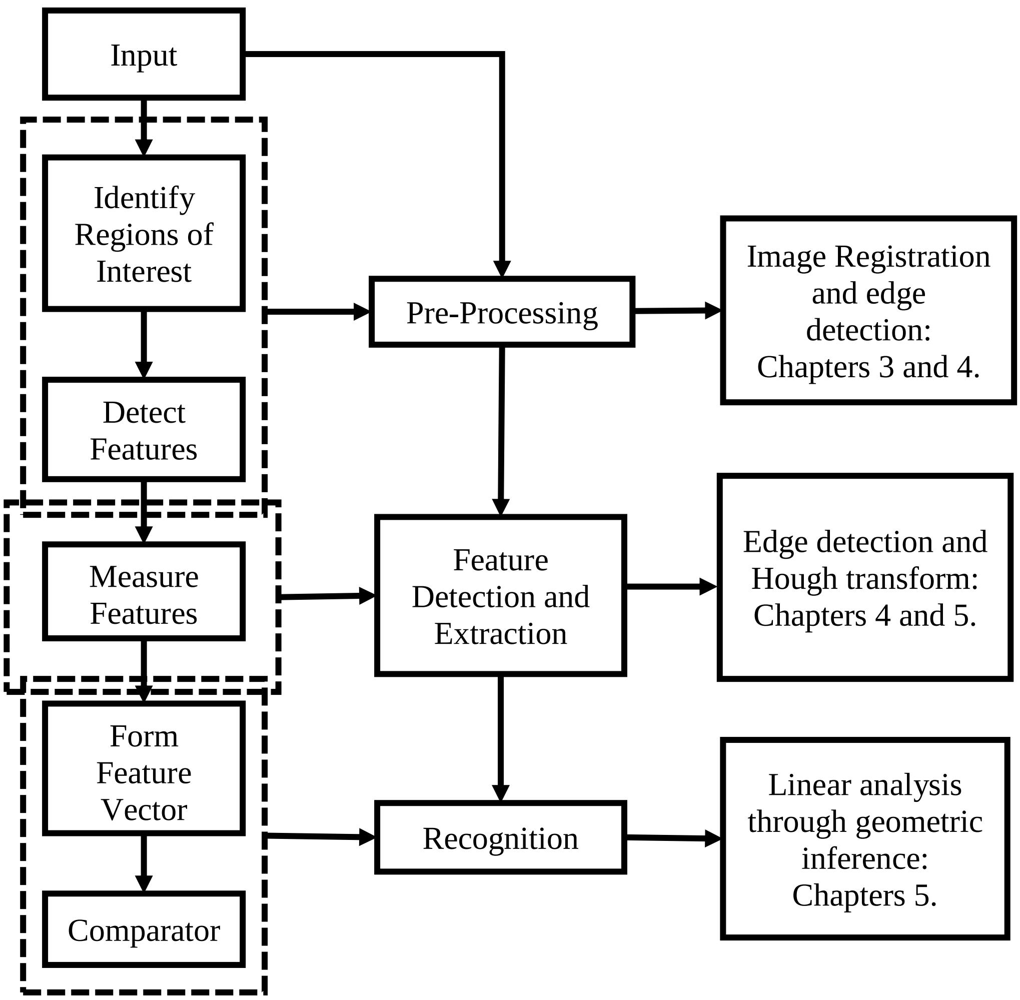


Figure 5.2: Object recognition Process, relating system level and processing tasks to the thesis chapters.

In the introduction in section 2.4.4 to data driven object recognition, there are identified processing tasks that benefit from the use of prior knowledge; such as object localisation within images and bounding box positioning for regions of interest. In the scheme of the proposed method of recognition, a data driven methodology is applied to all three stages in the recognition system level processes: pre-processing, feature detection and extraction, and recognition. The aim of utilising prior knowledge in this manner is that likelihoods of object instances are discriminated along the processing path as opposed to solely the end stage of recognition. Examples of prior knowledge for the object class cat may include: head position and pose; existence of features such as a second eye or ear; the distances to search for additional features in the input image; the appropriate profile to corelate detected features with.

5.2.1 Method of Recognition: Geometric Inference

For a cat the most discriminative body part is the head, which contains the features: ear, eye, nose, mouth. There are textural features, such as pattern frequency associated with fur. However, the reasoning behind the proposed method of recognition is based on which facial features and the spatial relationship of the facial features. The example cat head in figure 5.3 illustrates key features and a triangulated observation of the feature space.

Figure 5.3: Cat head [151] illustrating the triangular nature of feature space relationships and the key feature positions: ears, eyes, nose.

The key measurement of the features are the separation distances and the specific geometries of the feature. For example, the approximate circular eye radius or ear edge length and width, and incident angle at the tip of the ear. Within images the feature distances are projective, when a cat's head turns these projections are preserved in 2D. Since scale invariance becomes a critical issue (for reasons explained in section 5.2.2), preserving projective measurements becomes important when selecting the appropriate template profile to base the recognition decisions on. Hence, feature ratios which are dimensionless and therefore invariant to scale displacement, are used to normalise the feature data.

5.2.2 Processing Issues

To recognise a cat using linear geometry the first step is to detect the edges within the image. Using the RM edge detection method presented in chapter 4, the 3×3 edge kernel's sampling interval is 22.5° and the accuracy of the measurement is obtained by evaluating the peak of the edge reconstructions polynomial using 3-points and one coarse iteration. Additive noise is not considered. Using the edge image as input into the Hough transform (described in section 2.4.3), the threshold is obtained by selecting a range of the maximum peaks in the parametric Hough accumulator: rho and theta. Thresholding of the accumulator is important due to weakly detected edges in the image.

Once the Hough transform is applied and candidate feature locations are identified, the obtained initial geometric measurement will be in pixels. To have likewise comparisons at scale differences to training data measurements, pixel distances are required to be converted back to length (for example, mm or cm) and then normalised with respect to the feature in question. This process is complex in that the feature in question for the training data will have its standard deviation removed from the average value of the feature. Hence, collecting the ratio between critical features is a more considered approach to forming a feature vector.

Take the eye as the feature in question; the initial process will signal that the feature in question has a high probability to be an eye (based on the shape and size). The final determination must rely on the geometric relationships between the feature and the other features located subsequently. The application and the operating range required by the environment the system operates in will also dictate a range as to what scale and viewpoint it expects the object to present itself in. The overall approach is as follows; a visual device may have an operating range of 5 meters with a viewing angle of $\pm 50^{\circ}$. Whereby for a defined interval of scale, rotation and image resolution, a search range of the radius in the Hough transform can be applied. Once a candidate location has been found, closer inspection using a second Hough transform may present a smaller circular object (pupil perhaps). Hence information is accumulated that discriminates the candidate as an eye. For additional verification, eigenspace analysis of the candidate eye image region(s) to a constrained database of eyes, ears and noses can be applied. The only viable outcome is a probability that the feature is the feature of interest. However, cat eyes are elliptical and although the major axis does not change, the minor axis does. This is dependent on the closure of the cat's eye lid. Therefore, the starting point in the recognition process is an important consideration.

Due to the eye being deformable, the apparent stable object feature of the cat's face is the ear. Unlike dogs, cats' ears are almost always perpendicular to the boundary of the head. Therefore, the starting point sequences in the recognition process to consider are: *1) Ear, second ear, eye, second eye, nose, 2) Eye, second eye, ear, second ear, nose, 3) Nose, eye, second eye, ear, second ear.* Within these sequences there is the possibility that a second feature (eye or ear) may not be detectable or even exist. This is because the system cannot detect it, or it is occluded from the input. Therefore, missing features play a role in determining pose as well as decreasing the recognition methods probability of success. A final step to make use of the number of features located, and their geometric relationships, a number indicates the level of confidence that the object is a cat.

5.3 Recognition Method Routine

In this section, the overall approach to recognition and how prior knowledge is generated and implemented is presented. To begin, a functional diagram in figure 5.4 highlights the processes for both training and testing the method of recognition. Following this, each processing element is outlined separately. The first clarification of this recognition approach is that a detectable and identifiable process consists of geometric anatomical measurements and ratios of said object part features. Secondly, due to the constraints in the number of cat pictures available and the computing time required, the investigations shown should be treated as a demonstration of the method.

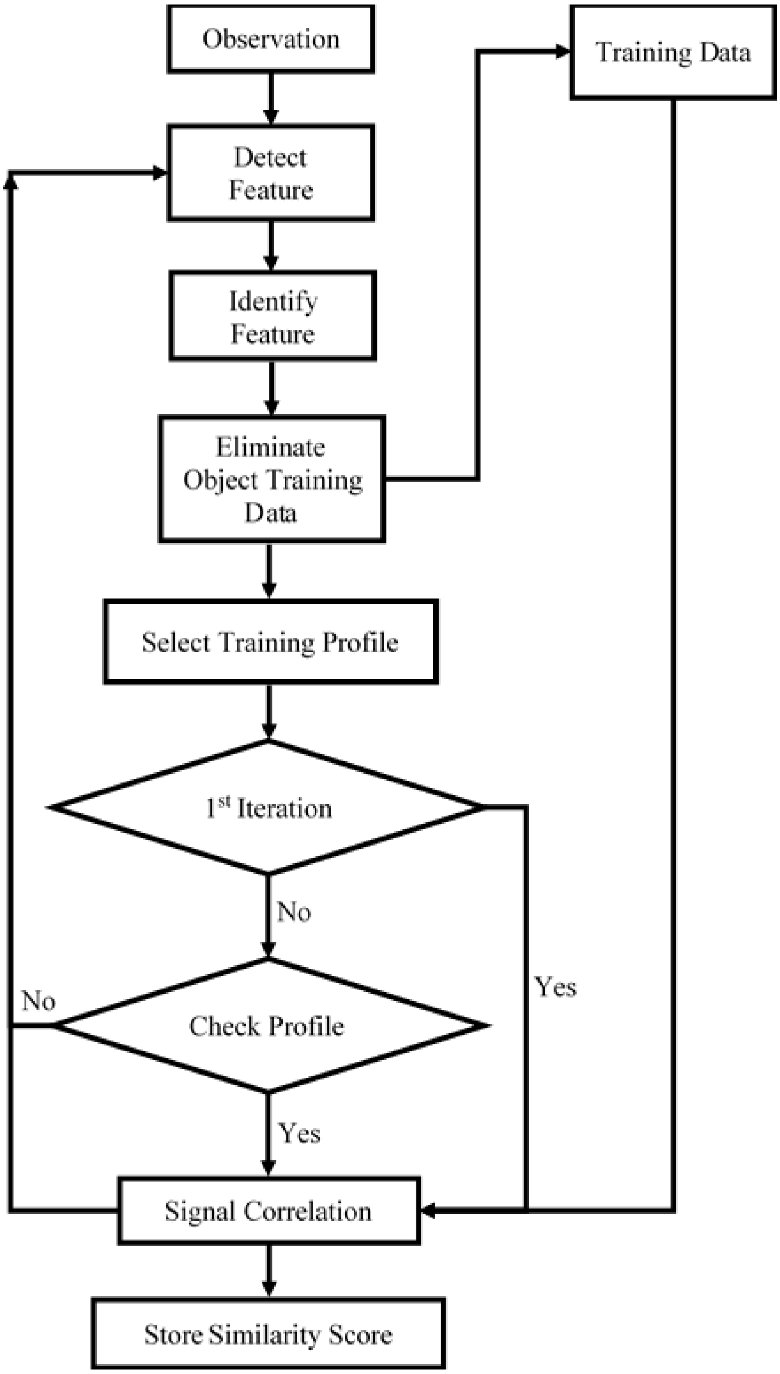


Figure 5.4: Functional diagram of the recognition method.

Observation

- Find area of interest: application of correlation, Eq. (3.7)

Detect Object Part

- Preserve shapes in image: application of edge detection, Eq. (4.7)

- Identify lines/curves to maximise the criteria of plane geometry input to the Hough transform

Identify Object Part

- Extract geometric shape information from Hough transform
- Identify geometric range of object part: scale and position/view
- Verify geometric distances with training examples
    - Euclidean distance metrics of feature geometry
    - Holistic shape similarity: eigenspaces

Eliminate Object Training Data

- Remove outcome of feature identification from training data

Signal Correlation

- Compare geometric distances with training profile feature
- Populate feature table with similarity score of object feature

Within this method three focus areas of investigation are identified. These are: the formulation of image data suitable for use in the Hough transform and correlation to find object parts, in order to resolve the pose/view of the object class cat; track changes in the object through the geometric inference of features; demonstrate occlusion to build up complexities into the recognition approach.

5.4 Demonstration of the Recognition Method

In this section the training data is presented to demonstrate the recognition method for two scenarios: firstly, a cat looking straight on at the camera and secondly, a cat whose pose is rotated away from the front profile view. Taken from a sub-category of the

ImageNet database [21], there are 25 examples of a cat's front profile (Ω_F) in figure 5.5. These are used to obtain an average profile of the cat; these profiles are depicted in figure 5.5. Whereby, the approximated feature measurements (see appendix A3) using Microsoft Word line measurement are: *1) eye major axis, 2) eye minor axis, 3) head width, 4) head height, 5) nose width, 6) nose height, 7) ear length, 8) ear width, 9) eye separation, 10) ear separation, 11) eye to nose separation, 12) eye to ear separation*. Any non-measurable feature is given a zero-measurement value. However, the features chosen are preliminary at this stage. The stability and discriminant ability across view invariance of these features is revealed most appropriately in section 4.4.2.

Figure 5.5: Cat profile $\Omega_{F(i=1:25)}$ training set [151]: 25 examples of a cat's front profile.

Each $\Omega_{F(i)}$ of the training set are presented in figure 5.6, information about their feature's measurements reveal those with the largest distance values. Whereby, the head width and length, ear separation and eye separation are the features with the largest linear distance. In these training images the ear length and width (feature 7 and feature 8) seems to exhibit the greatest variability of the obtained approximated feature measurements.

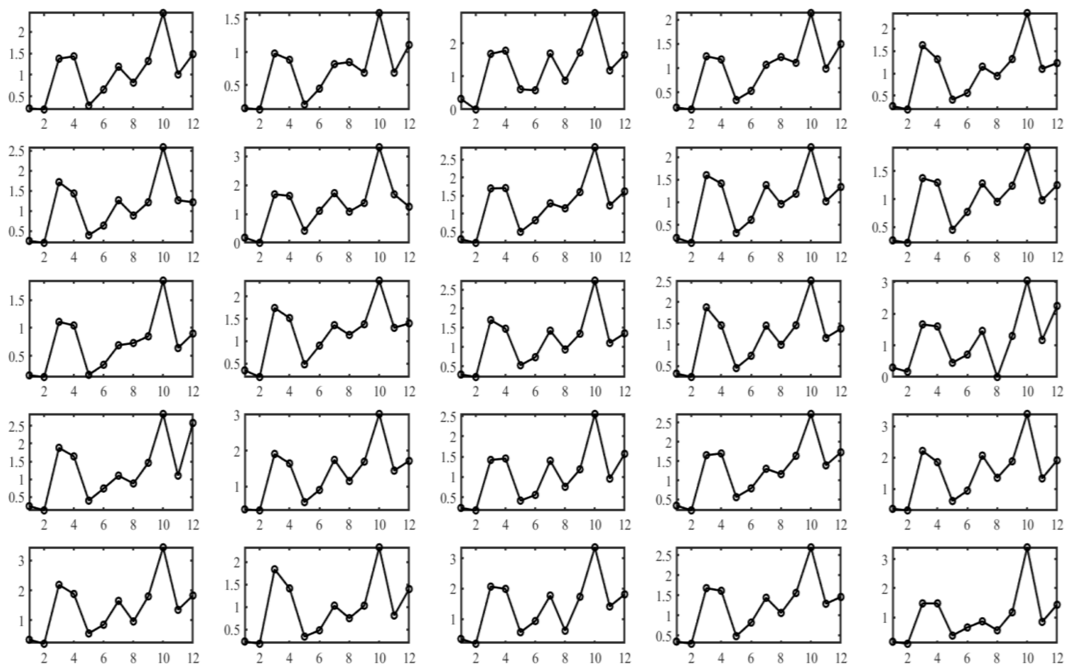
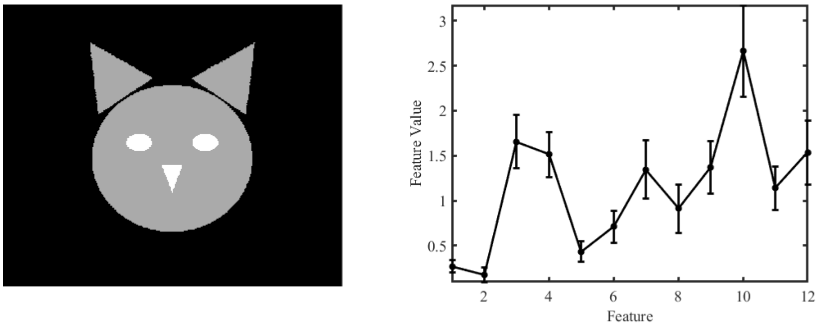


Figure 5.6: Individual plots of cat front profile feature vector $\Omega_{F(i)}$, each chart is one cat in figure 5.4. From left to right: number of feature measurements (X-axis) vs feature value (Y-axis), cm.

The average cat face of $\bar{\Omega}_F$ in figure 5.7a illustrates feature location, feature relationship distances (cm), and geometric approximations of the features; eyes are elliptical, head is elliptical, nose and ears are triangular. The average cats face appears regular because the object part shape is simplified by a regular geometric shape. The $\bar{\Omega}_F$ and standard deviation $\sigma\left(\bar{\Omega}_F\right)$ in figure 5.7b is important because it potentially indicates which are the most stable features: those features with a low variance and large mean. Comparing the statistics of figure 5.7b to the coefficient of variation $cv = \sigma\left(\bar{\Omega}_F\right) / \bar{\Omega}_F$ in table 5.1, the result show that the feature with the lowest dispersion is the ear separation distance, and the highest dispersion being the minor axis of the ellipse eye.



(a) (b)

Figure 5.7: a) Average cat face $\bar{\Omega}_F$ and b) feature vector: $\Omega_{F(i)} \pm \sigma\left(\Omega_{F(i)}\right)$.

Table 5.1: Front profile statistics: coefficient of Variation (*cv*), average $\bar{\Omega}_F$, standard deviation $\sigma\left(\Omega_F\right)$.

	Feature	***cv (%)***	$\bar{\Omega}_F$ ***(cm)***	$\sigma\left(\Omega_F\right)$ ***(cm)***
1	***Eye major axis***	25.68	0.269	0.069
2	***Eye minor axis***	44.7	0.178	0.079
3	***Head width***	18.4	1.656	0.298
4	***Head height***	16.7	1.513	0.252
5	***Nose width***	26.86	0.435	0.117
6	***Nose height***	25.01	0.713	0.178
7	***Ear length***	24.0	1.334	0.322
8	***Ear width***	29.53	0.911	0.269
9	***Eye separation***	21.31	1.371	0.292
10	***Ear separation***	19.1	2.662	0.508
11	***Eye to nose separation***	21.53	1.138	0.245
12	***Eye to ear separation***	23.36	1.535	0.3568

The head measurements can be considered redundant measurements because the boundary of the head, even simplified to circular shape, is difficult to define as a boundary of fur which is spectrally a high frequency component. The high dispersion of the eye closing seems to resemble the variation in the training set in figure 5.5. Where some eyes are partially open and some closed. The nose has a *cv* that is relatively equal to the remaining features (21.53% to 29.53%). By a visual inspection of the cat's nose, it is a small lesser defined object relative to an eye and an ear. Hence, the sequences of recognition stated in section 5.2.2 place the least importance to nose recognition in the recognition process.

5.4.1 Front Profile

In this section the recognition process is demonstrated for a cat front profile using two of the sample images in the training set, Ω_T . As each feature is detected and identified, its similarity to the average cats face feature range $\bar{\Omega}_{F(i)} \pm \sigma\left(\Omega_{F(i)}\right)$ is scored through its relative error in the range of $-\sigma\left(\Omega_{F(i)}\right) \leq error\left(\Omega\right) \leq \sigma\left(\Omega_{F(i)}\right)$:

$$error\left(\Omega_{(i)}\right) = \frac{\left|\Omega_{T(i)} - \bar{\Omega}_{F(i)}\right|}{\bar{\Omega}_{F(i)}} . \qquad \text{Eq. (5.1)}$$

The starting point of the process assumes that an area of interest has been obtained using Eq. (3.6). Therefore, the demonstration begins by looking for an eye, this is achieved by applying a circular Hough transform (CHT) to build up a range of accumulator spaces over a radial range. A second assumption here is that the eyes are open.

The test image [151] $\Omega_{T(1)}$ and its edge image using Eq. (4.7) are depicted in figure 5.8. In this demonstration the radial range to apply the CHT is obtained by manually extracting the approximate radius of the eye.

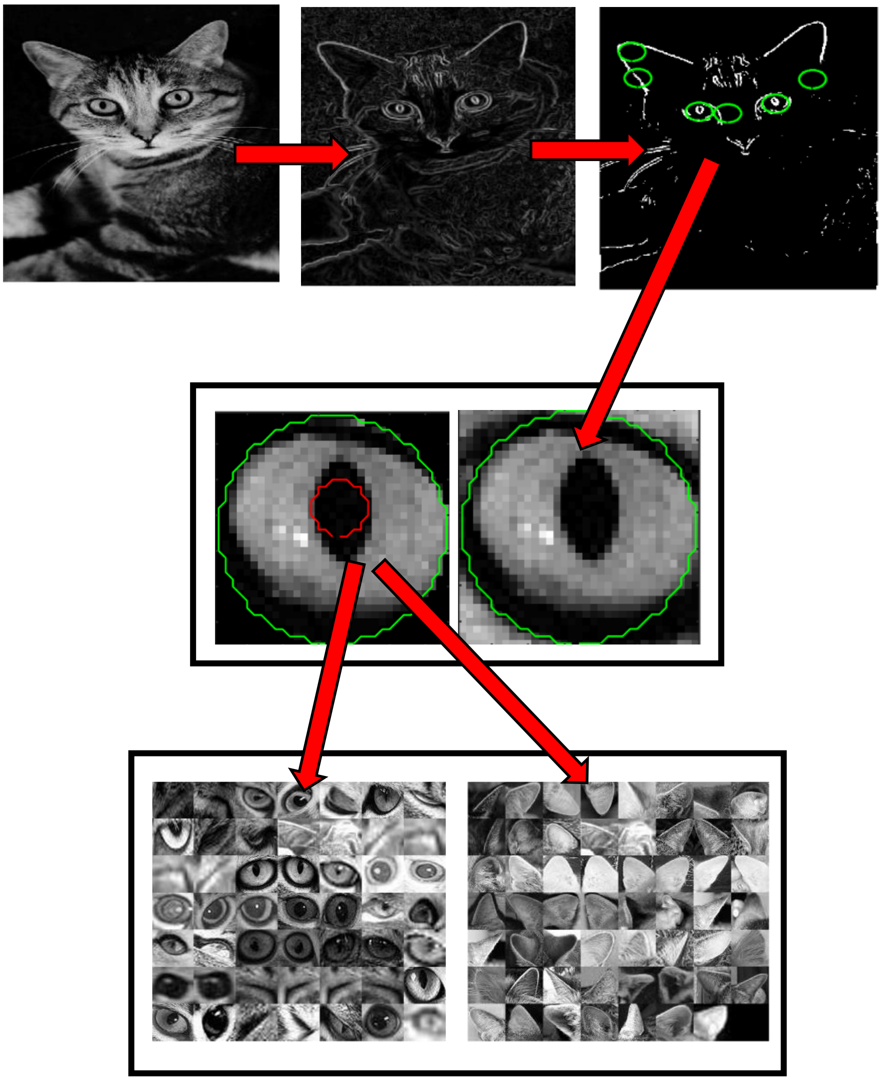
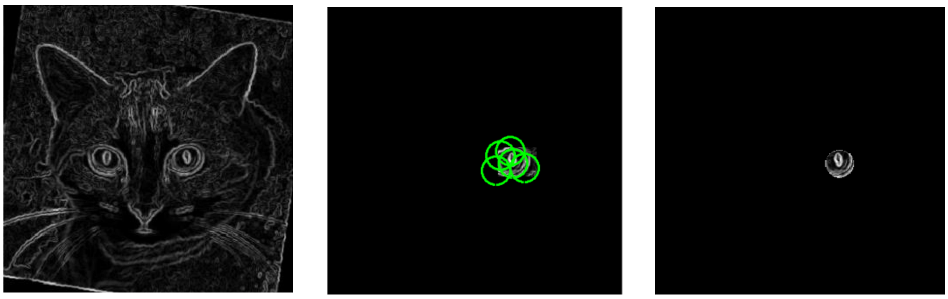


Figure 5.8: Example of process to identify the object part eye; the threshold to convert to a binary image is set at 0.5, the threshold of the CHT accumulator space is 90%, the CHT radial range is between 14 and 18-pixels (the potential locations of the eye are shown for 16-pixel radii only). A second CHT is applied to identify a smaller circle with radii of 4-pixels. The image region data is verified by matching it to a data base of cat object parts using the eigenspace methodology.

Once the eye has been detected and verification is satisfactory, a process of elimination can begin. Based on the average eye separation distance and its standard deviation (figure 5.7b) a search region can be calculated. The scale of $\Omega_{T(1)}$ is 0.014cm/pixel, therefore the average eye separation distance from table 5.1 at this scale is 98 pixels and its standard deviation 20 pixels. The search radius remains at 16 pixels because the eye is assumed to be open. The radial range of the searchable region is reduced by implying that the region does not exist outside of the image frame and that any in-plane rotation of the cat face can be recovered using autocorrelation in polar coordinates (as discussed in chapter 3) to scan for the closest matching rotation. This also implies that the cat is presenting itself at an in-plane angle. The search for the second eye is illustrated in figure 5.9. It is important to also consider that in the case that two eyes are initially detected, correlation between regions identifies the translation between the eyes. Hence their distance and in-plane rotation of the face is recoverable.



(a) (b) (c)

Figure 5.9: Second eye location: a) input image, b) search region and candidate locations c) Similar location based on correlation of spatial intensity map of first eye.

From the centre locations of the eye and the average separation distances of $\bar{\Omega}_F$, the search area for an image region can be determined to locate a nose. The geometry of the eye to nose feature space is illustrated as thus:

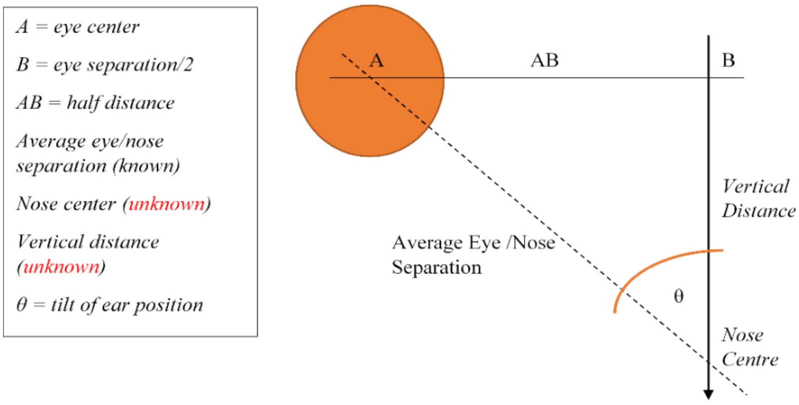


Figure 5.9: Geometric feature space of the eye and nose separation.

This geometric relationship is based on left eye to nose separation, since the cat face profile is forward facing there is an expected degree of symmetry if the opposite eye is used instead. The simplified shape characteristic of a cat's nose is triangular. As demonstrated in figure 5.11a-c, the Hough transform (HT) applied to a plane extracts the line parameters ρ and θ, whereby triangulated line relationships can be formed.

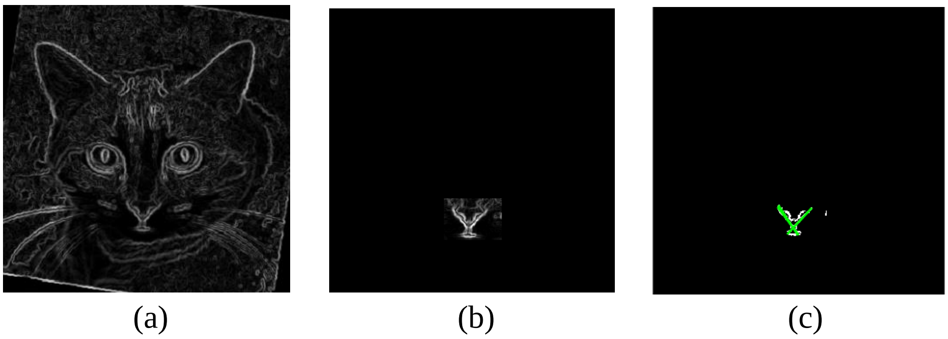


Figure 5.11: Nose location: a) input image, b) search region, c) candidate lines.

For the object part nose, the average eye to nose separation search distance is 81 ± 18 pixels. From this initial assumption the two strongest lines in the accumulator space of the HT reveal two lines with a length of 49 pixels and 47.5 pixels separated by an angle subtending -40° and 41°. A line's start and end point are isolated using an iterative count of points intercepting a line. Based upon the start and end points of each detected line the width and centre of the nose is deduced: nose width is 39 pixels and nose height (distance to eye centre along the y-direction) is 66 pixels. The centre of the nose becomes the second possible starting point for locating the ear; the first is the eye centre and the eye to ear separation distance. The recovered nose centre and eye centre coordinates allow the measured eye to nose separation distance to be averaged by repeating the process on the opposite eye. This distance is 74 pixels.

The geometry of the nose to ear feature space is formed from the subtended angle of the line describing the nose edge. In figure 5.12 These angles are based on the standard deviation of the nose height and eye separation half distance (AB).

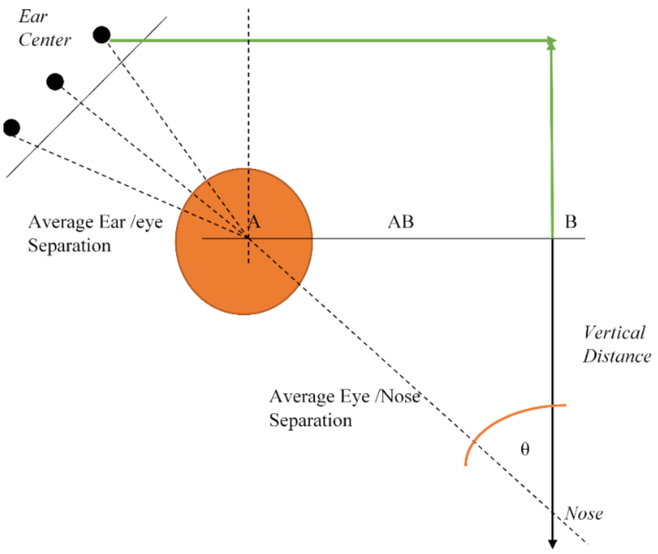
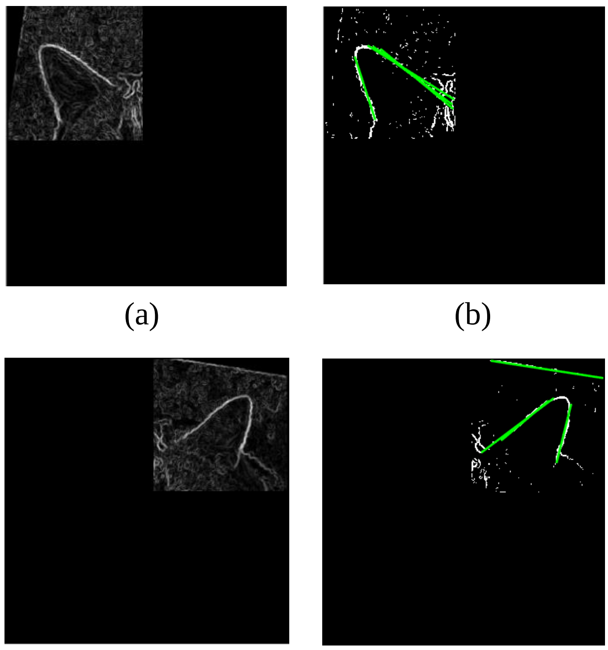


Figure 5.12: Geometric feature space of the nose, eye and nose separation.

For three projected angles of $\Omega_{T(1)}$ and three vertical distances, there are nine locations for the ear centre. Using the extremities of the ear centre coordinate locations an image region is generated to search for the ear. Within this region the HT is applied to inspect for straight lines, whereby peaks in the accumulator space are isolated using a threshold to retain a percentage of the strongest peaks. The width and height of the feature are extracted using the line segments and angles extracted in the HT. The extracted line segments for the ear are shown in figure 5.13a-b. The measured ear width and height are 83 pixels and 73 pixels respectively. The ears centre location is the coordinate at half the ear height, this information presents the distances to project an image region to detect the opposite ear, figure 5.13c-d.



(a) (b)

(c) (d)

Figure 5.13: Ear location: a) left search region, b) left ear candidate lines, c) right search region, d) right ear candidate lines.

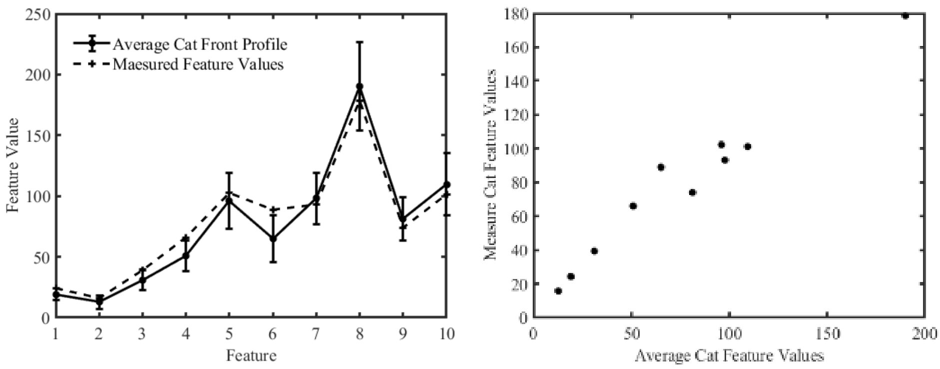
This distance is initially based on the average ear separation distance in $\bar{\Omega}_F$, the image frame and the exact location of the first ear within the frame boundary limit the possibility of which the second ear can be detected. In this case, the opposite (right) ear width and height are 95 pixels and 132 pixels. The left and right ear feature measurements can be averaged. Furthermore, the centre coordinate of the averaged right ear height denotes the stop point to measure the ear separation distance. This value is 179 pixels. For the left ear, the average subtended angles of the most congruent detected lines are -54° and -22°. For the right ear the average subtended angles are 52° and 14°. The average absolute angle formed by the approximate apex of the ear's geometric simplification is 35°. This angle is close to the angle of incidence of the nose line segments.

Based on the measured feature distances in table 5.2, the relative error obtained using Eq. (5.1) demonstrates the recognition success of the cat for this test by comparing the relative error between the average cat feature data and the measured cat feature data. A second point to consider, is the consistency of test data measurements used to obtain an average profile, whereby the same points in each test example is approximated. Whereas, the measured distances of the location coordinates are in error to the process of recognitions; these are limited to a pixel level accuracy. The absolute error between the measured values of $\Omega_{T(1)}$ and the observed values used in generating the average profile $\Omega_{T(0)}$, quantitively highlights the degree of systematic error in manually extracting the feature data in the training set for the sample used.

Table 5.2: Recognition measurements of $\Omega_{T(1)}$: measurements are in pixels except where % is denoted, cm/pixel scale value is 0.014.

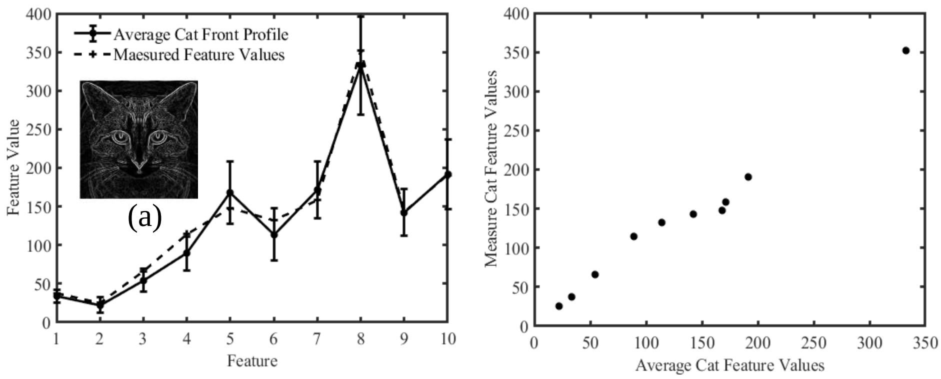
Feature	$\bar{\Omega}_{F(i)}$	$\pm\sigma\left(\Omega_{F(i)}\right)$	$\Omega_{T_1(i)}$	$error\left(\Omega_{(i)}\right)_{\%}$	$\Omega_{T_o(i)}$	$\left\vert\Omega_{T_o(i)}-\Omega_{T_1(i)}\right\vert$
1	19.2	4.9	24.2	26	18.9	5.3
2	12.7	5.7	16	26	13.6	2.4
3	-	-	-	-	-	-
4	-	-	-	-	-	-
5	31.01	8.4	39	26	29.3	9.7
6	50.9	12.7	66	29.7	40	26
7	96	23	102.6	6.8	82.1	20.45
8	65.1	19.2	88.75	36	67.1	21.65
9	97.9	20.9	93	5	94.3	1.3
10	190.2	36.3	178.6	6	167.6	11
11	81.3	17.5	74	9	78.6	4.6
12	109.6	25.6	101.4	7.5	87.9	13.45

Since the eye's radius is approximated to be a circle, the minor axis of the ellipse approximation to the eye in the test data is used to deduce a relative measurement of this feature. The head feature measurements are omitted due to the saliency of the feature. The feature vector of the measured values compared to the average profile is depicted in figure 5.14, the average error of $error\left(\bar{\Omega}_{T(1)}\right)_{\%}$ is 17.8%. The features with the highest degree of error are the width and length of the ear and nose and eye radius. However, the features with the lowest degree of error and lowest measurement error are the distances between object parts based on approximated start and end point locations. To ascertain confidence in the method of recognition for the front profile, the process is repeated for a second test sample $\Omega_{T(2)}$ (figure 5.15a). Whereby the observed and recorded measurements are documented in figure 5.15b-c and table 5.3.



(a) (b)

Figure 5.14: Comparison of average and measured feature values for $\Omega_{T(1)}$: a) average cat profile and measured feature vector, b) correlation between average and measured feature values.



(b) (c)

Figure 5.15: Comparison of average and measured feature values for $\Omega_{T(2)}$: a) edge image of $\Omega_{T(2)}$ b) average cat profile and measured feature vector, c) correlation between average and measured feature values.

Table 5.3: Recognition measurements of $\Omega_{T(2)}$: measurements are in pixels except where % is denoted, cm/pixel scale value is 0.008.

Feature	$\bar{\Omega}_{F(i)}$	$\pm\sigma\left(\Omega_{F(i)}\right)$	$\Omega_{T_1(i)}$	$error\left(\Omega_{(i)}\right)_{\%}$	$\Omega_{T_o(i)}$	$\left\|\Omega_{T_0(i)-}\Omega_{T_1(i)-}\right\|$
1	33.6	8.6	37.8	13	32.3	5.5
2	22.3	10	25	12	26.3	1.3
5	54	14.6	66	22	50	7.4
6	89	22.3	114	28	80	34
7	168	40.3	148	12	158.8	10.8
8	114	33.6	132	16	111.3	20.7
9	171	36.5	158	8	152.5	5.5
10	333	63.6	352	6	323.7	28.3
11	142	30.6	143	0.7	158.8	15.75
12	192	44.8	191	0.5	152.5	38.5

The $error\left(\bar{\Omega}_{T(2)}\right)_{\%}$ of the measured feature vector values against the average profile is 8.2%. Attempting to measure the length of an edge via a thresholding process of pre-processing may not always provide a reliable account of the feature. This is because the distances of the object parts are relatively small, therefore any fluctuations above the measurement will distort the value. Whereas, larger separation distances seem to present more stable features that are less susceptible to the measurement error of an edge. This is investigated in the next section to select better accurate profiles to compare feature distances against. The ratios of such lengths for resolving a cats pose/view is the primary mechanism to account for nonlinear feature interactions undergoing elevated and out-of-plane rotations.

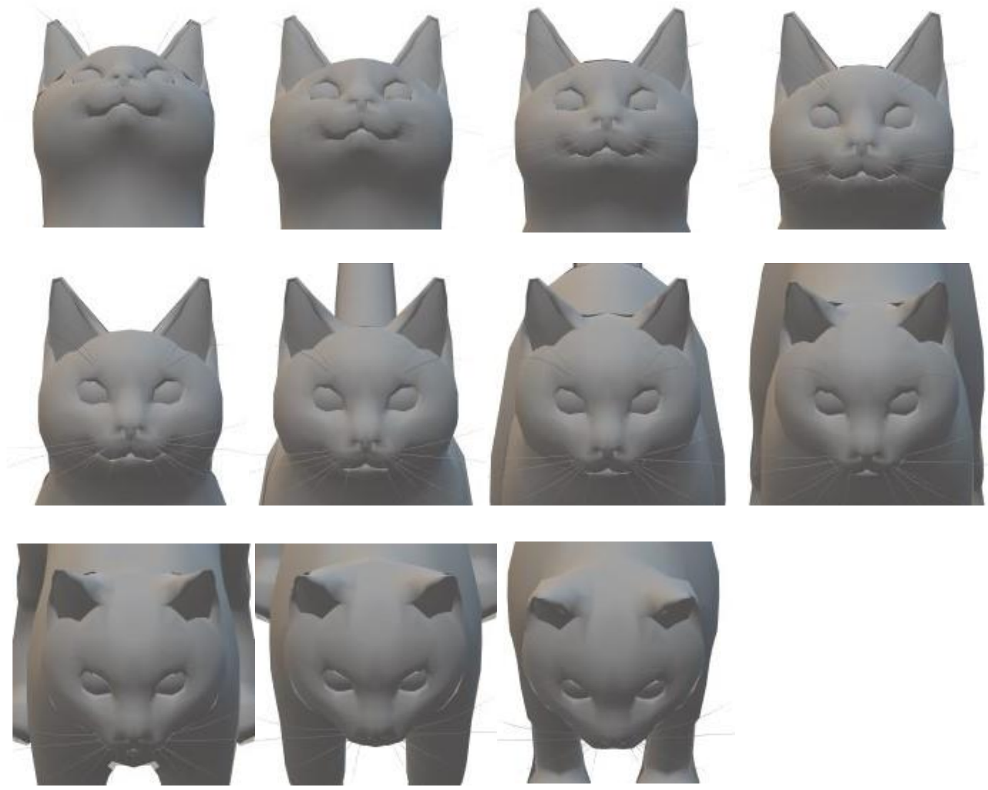
5.5 Accounting for Object View

For the identification of rotation in a cats head, a 3D model obtained from [155] is used to extract key feature distances across out-of-plane rotations θ and elevation angles ϕ. This feature space is denoted as $\Omega_F(\theta,\phi)$. For the training sets in figure 5.4 it is difficult to obtain examples of a cats head undergoing a series of rotations at set intervals, therefore the model used in the training data bases for Ω_F and $\Omega_F(\theta,\phi)$ are different. However, for demonstration of these two critical tasks: front profile and rotated profile, it is simple to see how one dataset can replace the other.

5.5.1 Rotated Profile

To begin the analysis of a rotated profile, the training object in figure 5.16a-b displays what is considered as the average cat head. This is a clear assumption that these results and analysis are based on. As will be revealed the appropriate cat head features to track through a 2D rotation map $\Omega_{F(i)}(\theta,\phi)$, are: ear separation, eye separation, nose to ear separation, nose to ear separation half distance and eye to ear cross separation. Of these features, there are (more than) two length tractable features through θ and ϕ rotation planes. This process is a factor of feature distortion identified in lesser measurable distances in the analysis of the front profile feature space. The cat head undergoing rotations in θ will always be symmetrical because the change in view is purely horizontal; except under occluded views. Under elevation angles ϕ this is not the case as the change in view occurs vertically, whereby the top of the cat head is not symmetric to the bottom view.

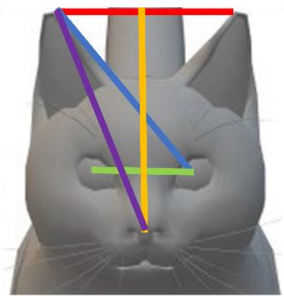
(a)



(b)

Figure 5.16: Test rotated cat head profiles [155]: a) out-of-plane rotation θ, rotation range left to right -50° to 0°, $\Delta\theta = 10^{\circ}$. b) elevation angle ϕ, left to right 50° to -50°, $\Delta\phi = 10^{\circ}$.

For $\Omega_F\left(\theta_{-50^\circ:0^\circ},\phi\right)$ in figure 5.16a, the features used to track the rotation displacement are the ear separation, eye separation and cross eye separation (left eye to right ear separation or vice-versa). These are chosen because of their length and availability to measure up to $\pm 50^\circ$. Whereas for $\Omega_F\left(\theta,\phi_{-50^\circ:50^\circ}\right)$ in figure 5.16b, the suitable features to track are the nose to ear separation and the nose to half ear separation distance. The features that are used for tracking combinations of 3D rotational displacement in a 2D images are highlighted in figure 5.17. The primary purpose of this step is to identify the most appropriate profile to use, to identify the class object cat. Described in figure 5.4, this is seen as an iterative process until variability in the profile selection process reduces to a stable and ideally singular response.



Feature 1: ear separation

Feature 2: eye separation

Feature 3: nose ear center separation

Feature 4: nose ear separation

Feature 5: eye ear cross separation (L2R)

Figure 5.17: Cat head [155] features used to track the rotated profile $\Omega_F\left(\theta,\phi\right)$.

Consistent and reliable reference points to measure distances is an issue identified in the cat's front profile (section 5.4.1). Propagating such variability into the rotated profile must be eliminated to enable isolation of correct template profiles in the recognition process. To remedy the systematic error in the location of the measurement points in the rotation profiles, the 3D cat head model is processed through AUTOCAD to fix the measurement point locations in the object's triangular mesh.

5.5.2 Spatial Relationship Analysis of Features

In this section the features illustrated in figure 5.17 (*feature 1: ear separation, feature 2: eye separation, feature 3: nose ear centre separation, feature 4: nose ear separation, feature 5: eye ear cross separation*) are tracked through combinations of $\Omega_F(\theta,\phi)$. Normalisation of the feature is delivered using the ratio of separation distances. The rationale for this normalisation process in the recognition method is discussed in section 5.2.2, whereby its critical point is that normalising on a per feature basis increases the number of comparisons to make. For each of the feature ratio plot in figures 5.18.1-5, the quotient's denominator equal to the numerator is omitted because the quotient value is 1 for all rotations. Feature 2 becomes unmeasurable and as such its value is fixed to zero. The obtained measured values are presented in appendix A3.

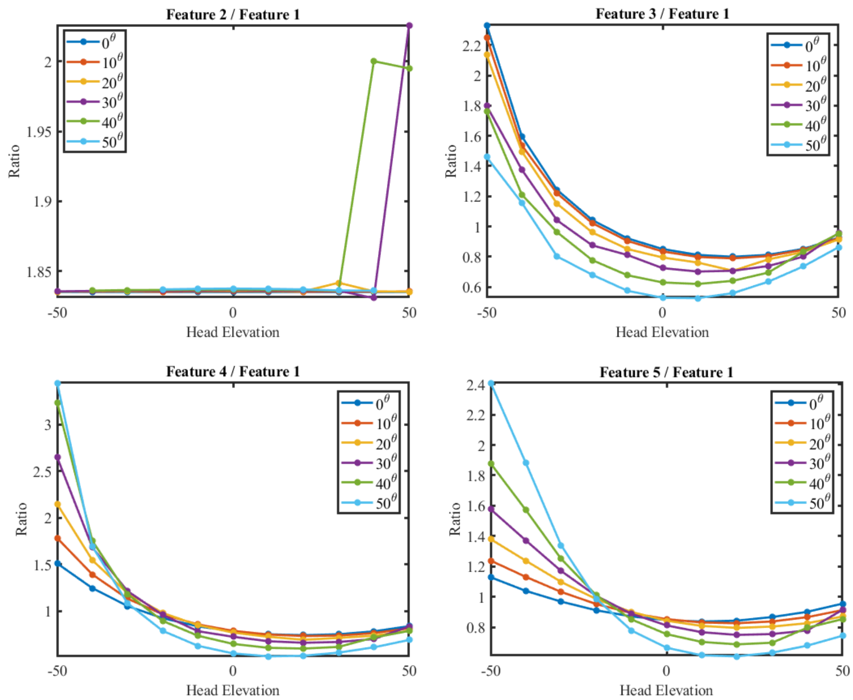


Figure 5.18.1: Tracking feature distance ratio 2,3,4,5: normalisation by feature 1.

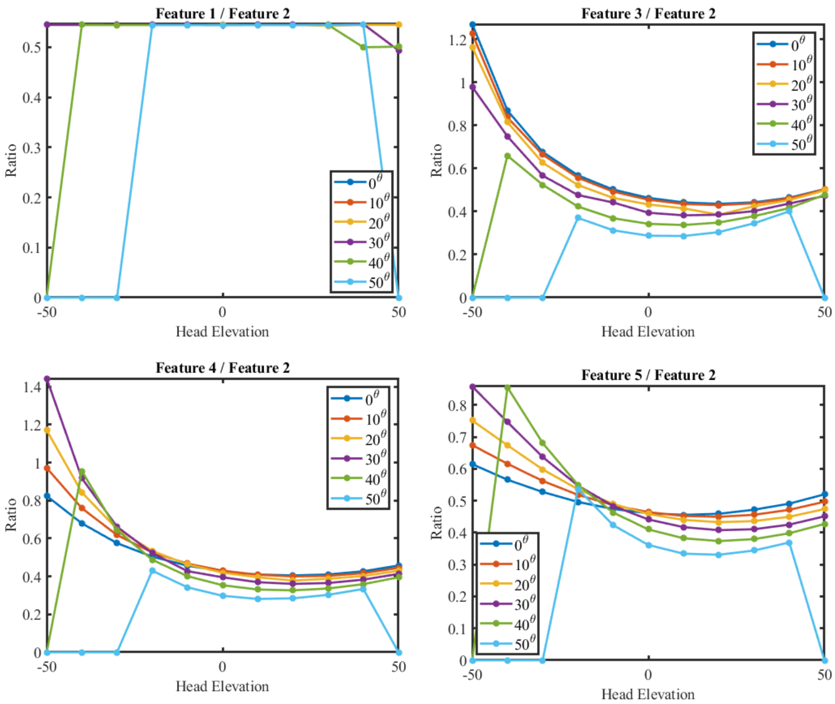


Figure 5.18.2: Tracking feature distance ratio 1,3,4,5: normalisation by feature 2.

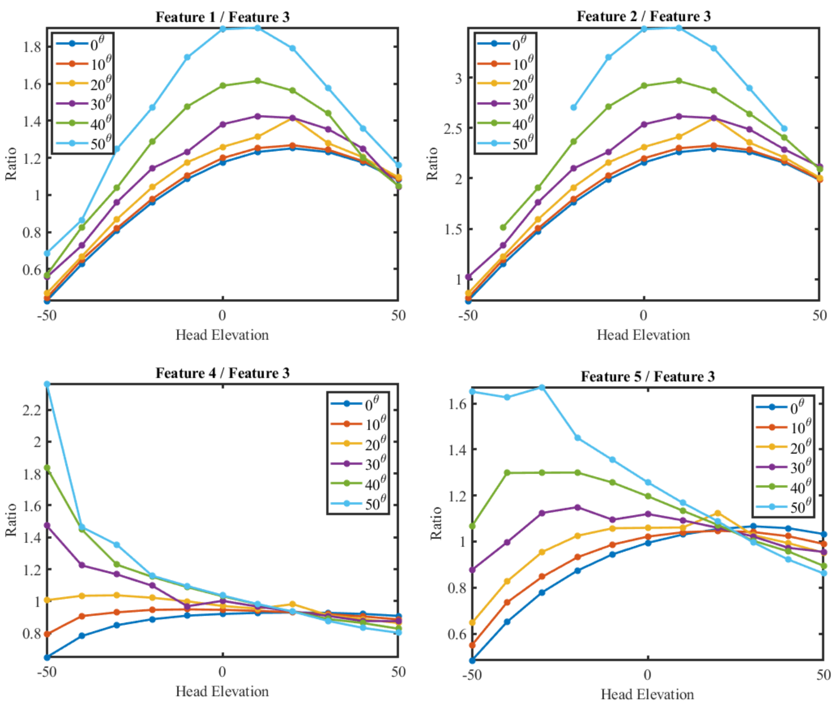


Figure 5.18.3: Tracking feature distance ratio 1,2,4,5: normalisation by feature 3.

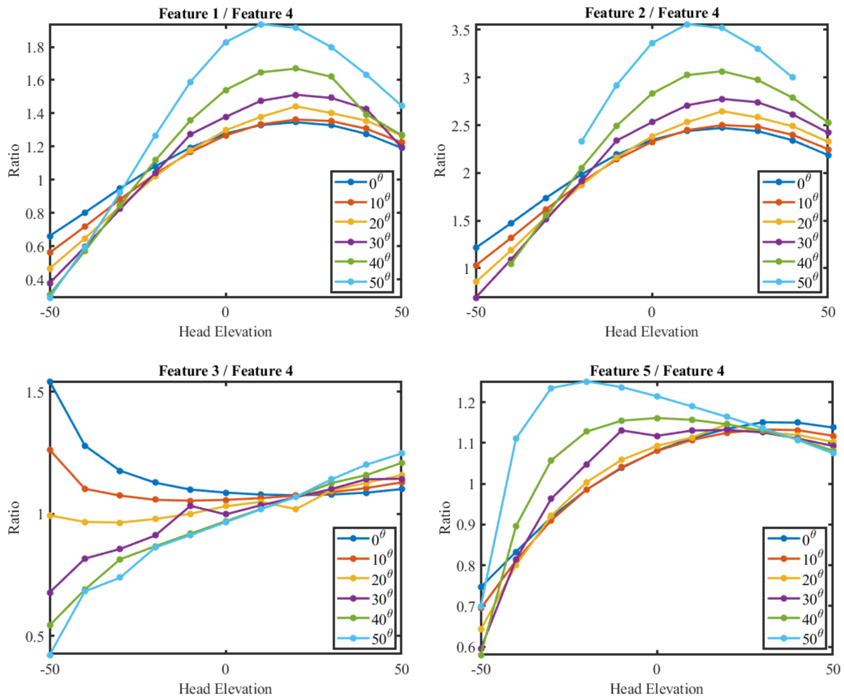


Figure 5.18.4: Tracking feature distance ratio 1,2,3,5: normalisation by feature 4.

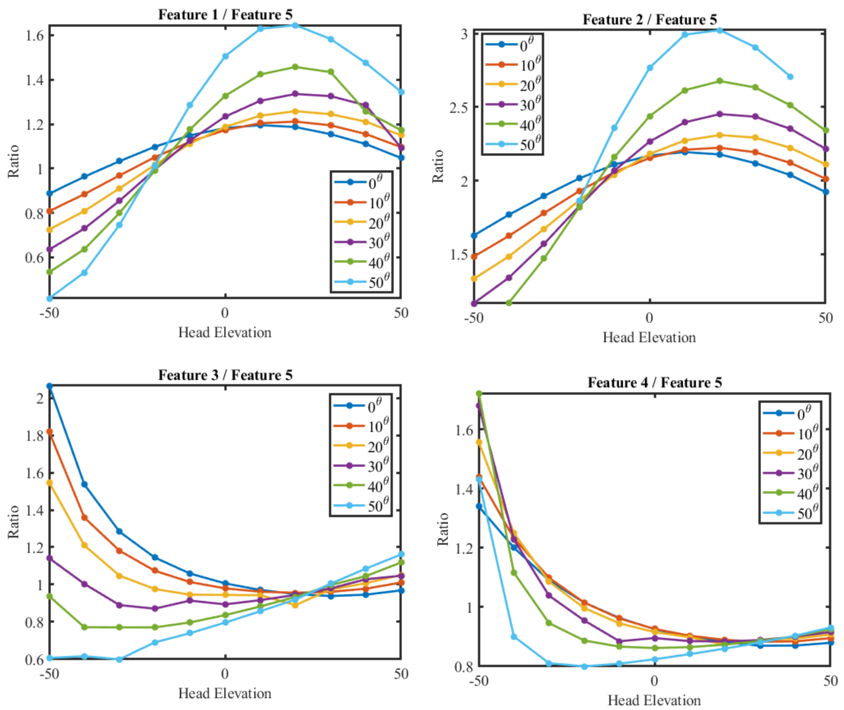


Figure 5.18.5: Tracking feature distance ratio 1,2,3,4: normalisation by feature 5.

For features that are perpendicular to each other (ear separation/nose ear centre separation, and eye separation/nose ear centre separation) there are no cross over points for the inversion of the ratio dependent on $\Omega_F(\theta,\phi)$. All ratios involving features 3,4 and 5 have a crossover inversion point, the ratio at this point is not symmetric. For features 4 and 5 there are two crossover inversion points, furthermore these points are not symmetric and change as θ increases. The disproportionality of the ratio is accounted by the vertical transitions of the features in the elevation angle of the cat's face. Features 1 and 2 do not change by large amounts vertically, therefore these ratios can be considered uniform across horizontal transitions. For the purpose of identifying an unknown feature sets view to select a recognition profile, the error in the similarity between the training data set must be minimised. To use these features, an ambiguity can exist for two identified points along one trajectory of θ, therefore cross checking between the inversion points is necessary. In the next section, the tracking of a cat head is demonstrated using the observations in the training set.

5.6 Object Pose Tracking and Profile Identification

The profile $\Omega_F(\theta,\phi)$ is a 2D rotational map for each feature separation distance ratio. Since each feature is normalised by each feature, there are $25\times\theta\times\phi$ feature ratio values to compare with an input feature ratio vector. Of size 5×5. The comparison is calculated using the Euclidean distance between the feature point values. To demonstrate this tracking method the cat head is rotated by firstly, a horizontal transition and secondly a combination of $\Omega_F(\theta,\phi)$ in between the discretised training set.

5.6.1 Error Calculation Method

To determine the location of the cat's head the Euclidean distance between the input feature ratios obtained during the feature measurement process (as described in section 5.3) and the known values obtained through the training examples (see figure 5.16a-b) is calculated. The process begins by determining the ratio between the feature measurements. As illustrated in figure 5.19 each ratio is an entry into the feature space $\Omega_F\left(\theta,\phi,R_n\right)$, where R_n is the normalised ratio of the features for n being the normalising feature in the ratio. For the purpose of the following examples it is assumed that all features have been obtained. Each entry of the feature space $\Omega_F\left(\theta,\phi\right)$ along R_n is used to compare to the measured values in the recognition process. The measured values parameters in this demonstration are denoted by $\Omega_T\left(\theta,\phi,R_n\right)$. Whereby, the equality of ratio proportionality, known as the ratio of ratios, are the discriminant affine features to track similarity of object pose.

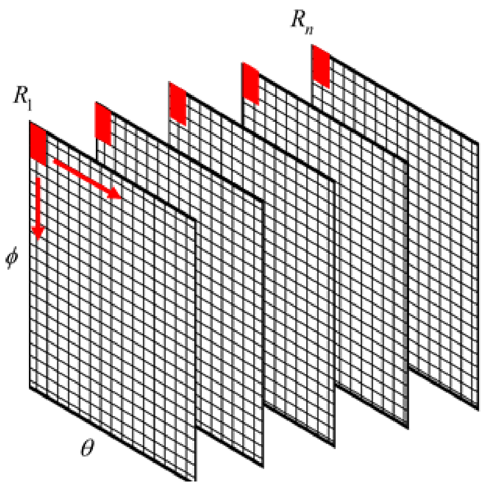


Figure 5.19: Formation of feature ratio matrix $\Omega_F\left(\theta,\phi,R_n\right)$. Red squares indicate column data input into the error calculation using the Euclidean distance to compare with $\Omega_T\left(\theta,\phi,R_n\right)$.

5.6.2 Object Pose Tracking Tests

There are two input signals in the tracking test: table 5.4.1-2. One for a transition taken from the training set $\Omega_F(\theta,\phi)$ for $\Omega_F(10^\circ,0^\circ)$, and one for a transition in-between the discretised training samples $\Omega_F(35^\circ,-25^\circ)$.

Table 5.4.1: Input signal, ratio of ratios for $\Omega_T(10^\circ,0^\circ,R_{1:5})$.

Ratio of Ratios					
Feature	***1***	***2***	***3***	***4***	***5***
1	1	1.836	0.835	0.789	0.853
2	0.545	1	0.455	0.430	0.464
3	1.198	2.20	1	0.946	1.021
4	1.267	2.326	1.058	1	1.080
5	1.173	2.1533	0.979	0.926	1

Table 5.4.2: Input signal, ratio of ratios for $\Omega_T(35^\circ,-25^\circ,R_{1:5})$.

Ratio of Ratios					
Feature	***1***	***2***	***3***	***4***	***5***
1	1	1.860	0.668	0.615	0.721
2	0.538	1	0.359	0.331	0.387
3	1.498	2.786	1	0.921	1.079
4	1.626	3.024	1.086	1	1.172
5	1.388	2.581	0.927	0.854	1

Table 5.5: Ratio errors for $\Omega_T\left(10^{\circ},0^{\circ},R_1\right)$.

θ,ϕ	***0°***	***10°***	***20°***	***30°***	***40°***	***50°***
-50°	1.686	1.774	1.953	2.217	3.356	3.64
-40°	0.907	0.967	1.077	1.166	1.26	2.313
-30°	0.5	0.547	0.568	0.57	0.574	1.923
-20°	0.255	0.276	0.265	0.233	0.199	0.205
-10°	0.101	0.106	0.078	0.043	0.166	0.315
0°	0.019	0	0.045	0.134	0.268	0.434
10°	0.046	0.058	0.107	0.194	0.319	0.476
20°	0.059	0.077	0.17	0.208	0.319	0.455
30°	0.045	0.061	0.105	0.182	0.271	0.379
40°	0.052	0.03	0.058	0.121	0.188	0.268
50°	0.145	0.109	0.081	0.239	0.198	1.842

In this first test for $\Omega_F\left(10^{\circ},0^{\circ}\right)$, the input feature ratio values $\Omega_T\left(10^{\circ},0^{\circ},R_1\right)$ are exactly equal to the values in the training data set. Hence the red square in table 5.5 indicates the location of rotation in the cat's head that has the least error (exactly zero) compared to the input signal feature ratios. This demonstrates, for the ideal scenario, the computational method to track the cat head for known rotational combinations in the training set.

In the next test, the cats head view is in-between the discretised training samples, the rotational combination is $\Omega_F\left(35^{\circ},-25^{\circ}\right)$. In this example the location of minimum error will not be equal to zero, this part of the investigation provides an account of the error's contribution for each feature ratio combination of $\Omega_T\left(35^{\circ},-25^{\circ},R_{1:5}\right)$. The ratio of errors for $\Omega_T\left(35^{\circ},-25^{\circ},R_{1:5}\right)$ are presented in appendix A4, whereby the location of $\Omega_F\left(\theta,\phi\right)$ for each feature ratio combination R_n are rounded to the nearest interval Δ in the training set.

The average ratio error $\Omega_T\left(35^{\circ},-25^{\circ},R_{1:5}\right)$ is 0.0606. This figure of merit, using all features, places a value as to the amount of systematic error due to approximating a profile at exactly half the interval rotational distance of $\Omega_F\left(\theta,\phi\right)$. For rotational angles in-between, for example a quarter or three-quarter distance from the training sample the errors approach zero. In the position of maximum error, each R_n is decomposed into its individual feature ratios that contribute to the total error per R_n. Firstly, the feature labels are restated: *feature 1: ear separation, feature 2: eye separation, feature 3: nose ear centre separation, feature 4: nose ear separation, feature 5: eye ear cross separation*. In table 5.6 the error contribution of each feature and their percentage of the ratio error are presented for $\phi = 25^{\circ}$. Observing the cat head at $\Omega_T\left(35^{\circ},-25^{\circ}\right)$, the feature combinations in ascending order contributing least error to each ratio error of R_n (ignoring feature parallelism) are: feature 4/feature 1, feature 5/feature 3, feature 4/feature 5, feature 4/feature 2 and feature 3/feature 4; R_1, R_3, R_5, R_2, R_4.

Table 5.6: Error contribution of each feature ratio to the ratio error for

$\Omega_T\left(35^{\circ}, -25^{\circ}, R_{1:5}\right)$.

	Feature 1	***Feature 2***	***Feature 3***	***Feature 4***	***Feature 5***	***Ratio Error***
R_1	0	-0.0240	0.0271	0.0021	-0.0232	0.043
% (Ratio Error)	-	56	63	5	54	-
R_2	0.0069	0	-0.0103	-0.0045	-0.0137	0.0190
% (Ratio Error)	36	-	54	24	72	-
R_3	0.0640	0.0825	0	0.0144	-0.0072	0.106
% (Ratio Error)	60	78	-	14	7	-
R_4	0.0434	0.0416	-0.0167	0	-0.0258	0.068
% (Ratio Error)	64	61	25	-	38	-
R_5	0.0356	0.0331	-0.0448	0.0109	0	0.067
% (Ratio Error)	53	49	67	16	-	-

To compare the percentage of ratio errors, table 5.7 presents the same data but for $\phi = 25^{\circ}$. The difference between the two elevation angles is that the observation of the cat's head is from the bottom, whereby features present themselves differently in the view. This is exhibited in the disproportionate symmetry in figure 5.18.1-5.

Table 5.7: Error contribution of each feature ratio to the ratio error for $\Omega_T\left(35^{\circ},25^{\circ},R_{1:5}\right)$.

	Feature 1	***Feature 2***	***Feature 3***	***Feature 4***	***Feature 5***	***Ratio Error***
R_1	0	0.0002	0.005	-0.0198	-0.096	0.098
% (Ratio Error)	-	0.2	5.1	20.2	97.9	-
R_2	0.0001	0	-0.0027	-0.0107	-0.0522	0.053
% (Ratio Error)	0.2	-	5.09	20.2	98.5	-
R_3	0.0065	0.0117	0	-0.0163	-0.1025	0.105
% (Ratio Error)	6.2	11.1	-	15.5	97.6	-
R_4	0.0211	0.0385	0.0133	0	-0.0768	0.089
% (Ratio Error)	23.7	43.3	14.9	-	86.3	-
R_5	0.0823	0.1516	-0.0294	-0.0025	0	0.175
% (Ratio Error)	47	86.6	16.8	1.4	-	-

The average ratio error for $\Omega_T\left(35^{\circ},25^{\circ},R_{1:5}\right)$ is 0.104. Observing the cat head at $\Omega_T\left(35^{\circ},25^{\circ}\right)$, the feature combinations in ascending order contributing least error to each ratio error of R_n (ignoring feature parallelism) are: feature 4/feature 5, feature 3/feature 2, feature 3/feature1, feature1/ feature 3 and feature 3/feature 4; R_5, R_2, R_1, R_3, R_4.

Comparing the two elevation angles, the results of the analysis show that the average ratio error is increased for a positive elevation, and that the precedence of the error contribution from each feature ratio is dependent on the observation of the cat's head. One source of error for the increased average ratio error is the availability of the cat's facial features as compared to looking toward the top of a cat's head. A second source is that feature ratio symmetry is disproportionate at a zero-elevation angle. This symmetry also carries a small dependency on the out-of-plane head rotation (see figure 5.18.1-5). Hence, the average ratio errors for in-between angles may follow the disproportionate symmetry. Observing a cat's head from underneath at $\phi = 25^{\circ}$, the results demonstrate a greater reliability on features perpendicular to each other. This is in opposition to diagonal distances carrying both a vertical and horizontal component of transition. Such as ear and eye separation and nose to ear half separation distance. In comparison, observing a cat's head from above at $\phi = -25^{\circ}$, the most reliable feature ratios are those with either a horizontal or vertical component and a feature carrying both transitional components. Such as, nose to ear separation and ear separation distance or two angled features containing both transitional contents.

5.7 Summary

The aim of this chapter has been to present a method of object recognition by acquiring prior knowledge suitable for identifying non-rigid objects. Two key processes are demonstrated: data driven recognition and object view tracking. This is achieved through a unique algorithm using the edge detection process presented in chapter 4 and the Hough transform across a line and plane. For the object class cat, its salient features are analysed to isolate those features appropriate for the recognition method. The linear analysis is based on the measurement and tracking of separation distances between features and ratios of them. The most discriminative and stable features are those accountable over lengthier distances, such as ear separation and nose to ear separation. A second type of features that can be relied upon are those oriented at an angle relative to the elevation or rotation of the head, so that a change in the head position will alter the length of the orthogonal component of the feature. Which feature is suitable will depend on the actual movement of the head, it is therefore essential that a minimum of two features, with their directions $\sim 90^{\circ}$ with each other, should be used in the identification process. To track a 3D object through a 2D basis, the displacements between features occurs both horizontally and vertically and unequally. The complexity of feature tracking is exemplified by interactions between the feature ratio combinations of rotational displacements exhibiting nonlinear object movements. For recognition of a cat with a frontal view the recognition error, obtained using a training sample, is as small as 8.2%. The consistency of the measurement location contributed to this error, and as such introduces a bias into the error measurement. The second training example error is 17.8%. The contributions of these errors are measured using the difference between the recorded and recovered feature distances. This revealed the susceptibility of shorter feature distances to measurement point location distortion. An additional influence of this error is the small variations to the front profile approximation in the training images. For a rotation of the cat's head $\Omega_F\left(35^{\circ},\pm 25^{\circ}\right)$, the average error peaks at the half distance of the rotational intervals are 0.104 and 0.0606 respectively.

Chapter 6 - Conclusions

The chapters of this thesis have investigated the bottle necks of an object recognition system's core processes and their application to identify geometrical forms of a non-rigid object. In this final chapter, the research is discussed against the aims and objectives communicated in the introductory chapter. Building on from the work in this thesis on applied correlation, edge feature detection and data driven recognition further research is identified. This includes possible novel application development of the explored signal processing techniques. Lastly, a summary of outcomes is provided.

6.1 Thesis Discussion

In chapter 1 the topic of object recognition is introduced through the perspective of the biological visual system. The central theme of the discourse provides a commentary on the physiological and cognitive components of vision whilst placing a thesis narrative towards applied signal processing in computational vision. The vital aim here, for the reader, is to be provided with an overview of object recognition systems and the necessary background information to understand the main issues and concepts in non-rigid object recognition (section 1.2): signal sensing, processing and analysis. Along with a survey of these issues, the reader's attention is focused upon view invariance in the real world and appropriate utilisation of prior knowledge and object information. To further develop subject specific knowledge, chapter 2 identifies three linear techniques in visual signal computation: correlation, edge detection and Hough transform. The systems statistical application of the signal processing techniques in image analysis, object detection and recognition are displacement and similarity measurement, the detection of boundary information, and the extraction and measurement of linear geometrical form. Each of the analytical functions have been mathematically presented and context is drawn, perhaps more importantly, towards limitations in measurement sensitivity and range, and spatial reasoning through inference in between object features (sections 2.2.3, 2.3.4 and 2.4.4). In non-rigid object detection and recognition processes the digital implementation of algorithms has also been considered and outlined. Each of the three central discussions forms part of the critical precis for the investigative components in this thesis; chapters 3 - 5.

The invariant nature of objects and features are confined to images and limited by rigid body displacements. In chapter 3, correlation is applied to the task of image registration to recover angular displacements. The examination of the process builds on from translation displacements using the digital implementation of the Fourier-Mellin transform. Critically, this technique is sensitive to scale distortion which is also propagated into angular displacements; even though rotation and scale elements are treated separately using a log transformation of the radial axis. A second limitation is the filtering mechanisms employed into the correlation. It is assumed that all object information occupies a mid-band range of spatial frequencies and because displacement is the feature of interest, phase correlation is usually applied. For broad band information, noise occurs everywhere in the signal spectrum and the phase response of the correlation maybe beholden to signal components with a low SNR. This results in a poor estimation of any displacement measurement. In response to these concerns (section 3.6), a matched filter analysis is employed to identify components with a high SNR. Particular attention focuses on the sampling of a polar transformation, hence, the available bandwidth in an analysis. In conjunction with the sampling condition of the signal, the intersection and interpolation of grid points in-between signal domains is of concern also. The adapted correlation method employs normalised correlation to encapsulate broad band signals. The SNR of the whole bandwidth is increased by adopting nonuniform sampling and a higher sampled input, with the aim of removing systematic errors and to respond to random processes. The up sampling of the input signal increases the accuracy in the polar conversion because it replaces grid points calculated from a lower sampling rate. Achievable average errors are 0.1% for the test images used to test the process. Also, via the simplicity of nearest-neighbour interpolation, reasonable processing times for larger images can be retained. In the context of image registration, autocorrelation is used to identify regions of interest in images where dissimilarity occurs. Rotational displacement, like transitional displacement, can be measured globally or locally and averaged over smaller defined regions for two purposes. Firstly, improved estimations of the displacements between a test and reference image sample (section 3.8 and 3.9). Secondly, identification and inference of underlying geometric structure in images to base further processing decisions on (sections 3.2 and 3.4).

To address the issue of detecting critical object information, the process of edge detection is investigated. Since non-rigid objects undergo articulated and elastic motion, chapter 4 looks at the boundary information lying beneath the pixel resolution. Instead of increasing the size of the edge kernel, which limits the detectable size of the feature, the region maxima (RM) method, adapted from the Sobel operator (section 4.4) estimates an arbitrary edge gradient direction and is then reconstructed to a higher degree of accuracy. The edge is modelled on a parabola and the process is iterated to refine the accuracy level of the measurement. After two iterations this accuracy fluctuates around a 10^{-5} level. For macro sized objects, such as a cat, knowing that the edge direction resolution is 1°, is normally appropriate for the process of feature extraction. However, smaller scale applications where the precision of the edge gradient direction is vital, the incentive of implementing iterative RM comes into its own. The validity of the analysis in section 4.5 and 4.6, is reliant on the test function input into the edge detector. Since all visual information is discretised, the test functions could only be estimated based on the arctangent coordinates of a discretised space along a continuous line. Therefore, for an infinite boundary any angle can be created. Hence, the analysis is restricted to edges at $14.04^{\circ}, 18.4^{\circ}, 26.56^{\circ}$. For acuter edge directions (14.04°), the symmetry of the parabola distorts to a function that can no longer be described by a second-order polynomial. Here, the adjacent sampled point of the edge kernel has a higher intensity than the centre point. Importantly this formulates a relationship between symmetry and error, whereby in the unusual case described, a proportional relationship exists. However, the relationship is inversely proportional when the RM method is operating correctly. This is because the sampling interval distance of the kernel in-between points along the parabola increase as the sampling interval increases.

The approach to non-rigid object recognition capitalises on the eigenfaces routine to identify key facial features. The process builds on from the RM edge detection process in chapter 4 to detect boundaries that are inputs for the Hough transform to validate. The recognition procedure described in section 5.3 bases decisions of the current feature (position, length and orientation) that are obtained from the Hough transform parameters, to identify what to detect and measure next. The starting point of the process is to look for a stable feature that does not vary non-rigidly relative to other features. The analysis of the cat's facial features identifies the ear as the most stable object part and features with longer measurable distances belonging to object parts. By acquiring evidence along each process, the recognition method is non-parametric. The approach is data driven and it acknowledges that confidence in the probability of recognition is based on the cumulated evidence. Whilst the comprehensive aim is to develop, test and validate a fully integrated algorithm, chapter 5 demonstrates two of the key elements in the recognition procedure. Firstly, recognition of the cat's head from a front observation (section 5.4) and secondly, tracking of the cat's head through out-of-plane rotations and elevated views (section 5.5 and 5.6). The purpose of these investigations is to demonstrate the procedure and employ a method to select appropriate profiles to compare acquired feature distance measurements. Limitations of the investigation are as follows: the twenty-five examples of a cat's head in the front observation training set is not a definite size; the method is tested using two examples from the training set; acquisition of a real cats head at defined rotational intervals is a complicated task, hence the rotated profiles are based on a computer generated 3D models of a cats head; variance is required to be built into the rotated training profiles, one solution to this, other than acquiring many more 3D models, is to jitter the template feature geometries. Even so, the analysis of the recognition method revealed an 8.2% error recognition rate of the cat from a front observation. An estimate of the error impact at half angle intervals of the selected training profile is given and is shown to vary for positive and negative elevation angles, 0.104 and 0.0606 respectively. The source of this variance is seen to be dependent on the asymmetry of the cat's head across elevated views. Such patterns are depicted in the feature spatial relationship analysis in section 5.5.2.

The developed algorithms in this thesis are unique to the pre-processing stages of the proposed data driven recognition routine. The figure of merits for the adapted algorithms are now compared to similar work.

The correlation analysis algorithm in chapter 3 establishes an average rotation error of 0.1% of a degree for an interval of 180° in 1° increments. In [156] the authors compare their algorithm to recover rotations based on the Radon transform [157]. They demonstrate accuracies on popular test images, for example *''Lena''*, against phase correlation (PC) and SIFT/SURF. The average accuracy of their method is as low as 10% of a degree over a range of 45°, which outperforms both PC and SURF [26]. In [158] a registration accuracy of 10^{-5} is demonstrated, but for a shift in translation only. Depending on the sampling combination selected, the detection accuracy of the RM method for angular units reduces to 10^{-5}. In [159] the authors demonstrate a third-order process in differential geometry to measure orientation, curvature and position. Considering no scale distortion, their achieved accuracy, based on simple binary shapes (200×200 pixels) is typically of order 10^{-2}; image scale and edge blur was shown to reduce this accuracy. The accuracy of the RM method, based on sampling points along a curve, outperforms higher-order differentiation to recover sub-pixel gradient information. Further investigation of the RM method is required; particularly distortion due to blur and scale.

The authors in [160] investigate active skeletons as a binary shape descriptor for animals and reduced training sets. For a training set size of 15, they tested 313 binary representations with ~80% recognition success for changes in non-rigid object pose. A recognition confidence level is not discussed in [160] and the measured recognition errors in the data driven recognition routine proposed in this work, are based on isolated examples. Initial reports for the proposed method are promising and substantive evidence to recognise and track key non-rigid object features has been retrieved. With further investigation, comparable metrics can be established.

6.2 Further Work

In this research study, the limitations of the adapted signal processing tool's measurement capabilities and, the approach in non-rigid object recognition have indicated these areas as recommendations for further work.

Tracking image structure through correlation analysis

- Using simple shaped nonbinary uniform and nonuniform test functions, verify symmetry errors in the angular correlation scan.
- Impose simple object shapes into general image scenes to assimilate real object shapes; identify changes in the 1D correlation shape and symmetry.

Misalignment in medical/manufacturing process: millimetre/micron application

- Identify fiducial patterns to track 2D translation and orientation displacements between sample and test images.
- Utilise accurate edge gradient information input into the image registration process.
- Model changes to edge direction and intensity in temporal environments to indirectly measure and track changes in features, particularly nonuniform circular objects.

Recognition through geometric incidence structure

- Develop a more rigorous approach to eliminate non-interesting information in the localisation process via image registration. Identify feature characteristics earlier on in the recognition process
- Explore training set sizes, establish the amount and limit of error in the average profile across rotational views and intervals.
- Develop the feature disambiguation, through the extraction process, of the object cat to generate finer detailed features of cat breeds and “my cat”.
- Employ PCA as the comparator to investigate orientation of the features, and distortion, through the position and distance of the characteristic values.

6.3 Summary

The aim of this thesis examines the incidence and geometry of object features for applications in non-rigid object recognition. The study begins by inspecting object recognition system processes to determine appropriate signal processing techniques in addressing the critical issues of view invariance, feature detection and appropriation of features in recognition. The research in this thesis describes and presents these central outcomes. Firstly, a measurement sensitivity analysis of an adopted correlation tool to detect angular displacements. Secondly, an edge feature extraction tool whose measurement sensitivity is characterised toward the micron scale. Lastly, demonstration of a recognition approach, suitable for characterising non-rigid objects which is based on a data driven acquirement of features and a linear assessment of feature geometry. It is hoped that useful insights have been gained for identifying complex object patterns that may be hidden, and by which, can be simplified within linear subspaces.

References

[1] "Wiring diagram Eye pattern Drawing free commercial clipart - Mavic Pro,Anatomy,Biology free clip art." [Online]. Available: https://www.kisscc0.com/clipart/wiring-diagram-eye-pattern-drawing-eye-diagram-3mp7i4/. [Accessed: 01-Nov-2018].

[2] P. Artal, "Optics of The Eye and Its Impact In Vision : A Tutorial," no. 34, pp. 340–367, 2014.

[3] M. Carandini *et al.*, "Do We Know What the Early Visual System Does ?," *J. Neurosci.*, vol. 25, no. 46, pp. 10577–10597, 2005.

[4] T. Gollisch and M. Meister, "Review Eye Smarter than Scientists Believed : Neural Computations in Circuits of the Retina," *Neuron*, vol. 65, no. 2, pp. 150–164, 2010.

[5] T. N. Wiesel and D. H. Hubel, "Receptive Fields and Functional Architecture of Monkey Striate Cortex," *J. physio.,* Vol. 195, no 04 pp. 215–243, 1968.

[6] M. J. Farah, "Is an Object an Object an Object? Cognitive and Neuropsychological Investigations of Domain Specificity in Visual Object Recognition," *Curr. Dir. Psychol. Sci.*, vol. 1, no. 5, pp. 164–169, 1992.

[7] N. K. Logothetis and D. L. Sheinberg, "Visual Object Recognition.," *Annu. Rev. Neurosci.*, vol. 19, pp. 577–621, 1996.

[8] I. Biederman, "Recognition-by-Components: A Theory of Human Image Understanding.," *Psychol. Rev.*, vol. 94, no. 2, pp. 115–147, 1987.

[9] M. Juttner, T. Caelli, and I. Rentschler, "Recognition-by-Parts: A Computational Approach To Human Learning and Generalization of Shapes.," *Biol. Cybern.*, vol. 74, no. 6, pp. 521–535, Jun. 1996.

[10] K. Koffka, "Principles of Gestalt Psychology Chapter I Why Psychology? An Introductory Question," *Princ. Gestalt Psychol.*, pp. 1–14, 1935.

[11] J. B. Watson, "Psychology As The Behaviorist Views It.," *Psychol. Rev.*, vol. 20, no. 2, pp. 158–177, 1913.

[12] M. A. Johan Wagemans, James H. Elder, Michael Kubovy, Stephen E. Palmer and and R. von der H. Peterson, Manish Singh, "A Century of Gestalt Psychology in Visual Perception I. Perceptual Grouping and Figure-Ground Organization," *Psychol. Bull.*, vol. 138, no. 6, pp. 1172–1217, 2012.

[13] D. Marr and E. Hildreth, "Theory of Edge Detection," *Proc. R. Soc. B Biol. Sci.*, vol. 207, no. 1167, pp. 187–217, 1980.

[14] T. J. Palmeri and I. Gauthier, "Visual Object Understanding," *Nat. Rev. Neurosci.*, vol. 5, p. 291, Apr. 2004.

[15] W. S. Mcculloch and W. Pitts, "A Logical Calculus Nervous Activity," *Bull. Math. Biol.*, vol. 52, no. l, pp. 99–115, 1943.

[16] W. Pitts and W. S. McCulloch, "How We Know Universals The Perception of Auditory and Visual Forms," *Bull. Math. Biophys.*, vol. 9, no. 3, pp. 127–147, Sep. 1947.

[17] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, pp. 193-202, 1980.

[18] K. Fukushima, "Neocognitron : A Hierarchical Neural Network Capable of Visual Pattern Recognition," *Neural Networks.*, vol. 1, pp. 119–130, 1988.

[19] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, “Object Recognition with Gradient-Based Learning,” *Proc. IEEE.*, Vol. 86, no. 12, pp. 2278–2324, 1998.
[20] Y. Lecun, Y. Bengio, and G. Hinton, “Deep Learning,” 2015.
[21] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” *2009 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 248–255, 2009.
[22] D. Gabor, “Theory of Communication. Part 1: The Analysis of Information,” *Electr. Eng. - Part III Radio Commun. Eng. J. Inst.*, vol. 93, no. 26, pp. 429–441, 1946.
[23] R. O. Duda and P. E. Hart, “Use of The Hough Transformation to Detect Lines and Curves in Pictures,” *Commun. ACM*, vol. 15, no. 1, pp. 11–15, 1972.
[24] P. Hart, “How the Hough transform was Invented,” *IEEE Signal Process. Mag.*, vol. 26, no. 6, pp. 18–22, 2009.
[25] D. G. Lowe, “Object Recognition From Local Scale-Invariant Features,” *Proc. Seventh IEEE Int. Conf. Comput. Vis.*, pp. 1150–1157 vol.2, 1999.
[26] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *Int. J. Computer Vision.*, vol. 20, no. 01, pp. 91–110, 2003.
[27] J. Canny, “A Computational Approach to Edge Detection.,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, Jun. 1986.
[28] R. K. McConnell, “Method of and Apparatus for Pattern Recognition.” Patent, 28-Jan-1986.
[29] M. Tarr and H. Bülthoff, “Is Human Object Recognition Better Described by Geon Structural Descriptions or by Multiple Views?,” *J. Exp Psych: Human Perception and Performance*., vol. 21,l no. 6, 1995.
[30] R. Giryes and M. Elad, “Reinforcement Learning: A Survey,” *Eur. Signal Proc. Conf.*, pp. 1475–1479, 2011.
[31] L. B. Merabet, J. F. Rizzo, A. Amedi, D. C. Somers, and A. Pascual-leone, “What Blindness Can Tell us About Seeing Again: Merging Neuroplasticity and Neuroprostheses,” *Nat Rev Neurosci.*, vol. 6, no. 01, January, pp. 1–7, 2005.
[32] E. O. Brigham, The Fast Fourier Transform and its Applications, 2nd ed. Englewood Cliffs, N.J: Prentice-Hall, 1974.
[33] J. W. Cooley and J. W. Tukey, “An Algorithm for the Machine Calculation of Complex Fourier Series,” *Math. Comput.*, vol. 19, no. 90, p. 297, 1965.
[34] R. N. Bracewell, The Fourier Transform and its Applications, 2nd ed. New York: McGraw-Hill, 1978.
[35] N. S. Hashemi, R. B. Aghdam, A. S. B. Ghiasi, and P. Fatemi, “Template Matching Advances and Applications in Image Analysis,” *article.*, no. 10, 2016.
[36] T. R. Crimmins, “A Complete Set of Fourier Descriptors for Two-Dimensional Shapes,” no. 6, pp. 848–855, 1982.
[37] T. P. Wallace and O. R. Mitchell, “Abstracts of the Workshop Comput. Anal, of Time-Varying Imagery,” no. 6, pp. 583–588, 1980.
[38] A. El Oirrak and D. Aboutajdine, “Affine Invariant Descriptors using Fourier Series,” *Pattern. Recognition Letters.*, vol. 23, no. 8, pp. 1109–1118, 2002.
[39] D. Casasent and D. Psaltis, “Scale Invariant Optical Correlation using Mellin Transforms,” *Optics Communications.*, vol. 17, no. 1, pp. 59–63, 1976.
[40] D. Casasent and D. Psaltis, “Position , Rotation , and Scale Invariant Optical Correlation,” *Applied Optics.*, vol. 15, no. 7, pp. 1795–1799, 1976.
[41] H. S. Stone, B. Tao, and M. Mcguire, “Analysis of Image Registration Noise due to Rotationally Dependent Aliasing,” *J. Vis. Comm. Image Representation.*, Vol. 14, pp. 114 - 135, 2003.

[42] J. Altmann and H. J. Reitbock, “A Fast Correlation Method for Scale-and Translation-Invariant Pattern Recognition.,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 1, pp. 46–57, 1984.
[43] T. Caelli and G. Moraglia, “On the Detection of Signals Embedded in Natural Scenes,” *Percept Psychophys.*, vol. 39, no. 2, pp. 87–95, 1986.
[44] M. G. A. Thomson and D. H. Foster, “Role of Second-and Third-Order Statistics in the Discriminability of Natural Images,” *J. Opt. Soc. Am. A*, vol. 14, no. 9, p. 2081, 1997.
[45] M. G. A. Thomson, “Higher-Order Structure in Natural Scenes,” *J. Opt. Soc. Am. A*, vol. 16, no. July, pp. 1549–1553, 1999.
[46] D. Zoran, “Natural Image Statistics for Human and Computer Vision,” 2013.
[47] B. Srinivasa Reddy and B. N. Chatterji, “An FFT-Based Technique for Translation, Rotation, and Scale-Invariant Image Registration,” *IEEE Trans. Image Process.*, vol. 5, no. 8, pp. 1266–1271, 1996.
[48] F. E. A. El-gamal, M. Elmogy, and A. Atwan, “Current Trends in Medical Image Registration and Fusion,” *Egypt. Informatics J.*, vol. 17, no. 1, pp. 99–124, 2016.
[49] I. Despotović, B. Goossens, and W. Philips, “MRI Segmentation of the Human Brain: Challenges, Methods, and Applications,” *Comput. Math. Methods Med.*, vol. 2015, 2015.
[50] Q. Chen, M. Defrise, and F. Deconinck, “Symmetric Phase-Only Matched Filtering Image Registration and Recognition,” *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 16, no. 12, pp. 1156–1168, 1994.
[51] H. Yan and J. G. Liu, “Robust Phase Correlation Based Feature Matching for Image Co-Registration and DEM Generation,” *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 37, no. 07, pp. 1751–1756, 2008.
[52] Á. Ordóñez, F. Argüello, and D. B. Heras, “Fourier–Mellin Registration of Two Hyperspectral Images,” *Int. J. Remote Sens.*, vol. 38, no. 11, pp. 3253–3273, 2017.
[53] R. Dahyot, “Unsupervised Camera Motion Estimation and Moving Object Detection in Videos,” *Image Process.*, no. 1, pp. 102–109, 2006.
[54] D. Stricker, “Tracking with Reference Images : A Real-Time and Markerless Tracking Solution for Out-Door Augmented Reality Applications,” *Proc. 2001 Conf. Virtual reality, Archeol. Cult. Herit. - VAST ’01*, no. May, p. 77, 2001.
[55] L. Tai Sing, “Image Representation using 2D Gabor Wavelets,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 959–971, 1996.
[56] Q. Tian and M. N. Huhns, “Algorithms for Subpixel Registration,” *Comput. Vis. Graph. Image Process.*, vol. 35, no. 2, pp. 220–233, Aug. 1986.
[57] G. Ifrah, The Universal History of Computing: From the Abacus to Quantum Computing, New York, NY, USA: John Wiley & Sons, Inc., 2000.
[58] Pierre Ambs, “Optical Computing: A 60-Year Adventure,” *Advances in Optical Technologies*, vol. 2010, 15 pages, 2010.
[59] D. J. Heeger, “Model for The Extraction of Image Flow,” *J. Opt. Soc. Am. A*, vol. 4, no. 8, pp. 1455–1471, 1987.
[60] H. Foroosh, J. B. Zerubia, and M. Berthod, “Extension of Phase Correlation to Subpixel Registration,” *IEEE Trans. Image Process.*, vol. 11, no. 3, pp. 188–199, 2002.
[61] A. M. R. Sousa, J. Xavier, M. Vaz, J. J. L. Morais, and V. M. J. Filipe, “Cross-Correlation and Differential Technique Combination to Determine Displacement Fields,” *Strain*, vol. 47, no. 2, pp. 87–98, 2011.

[62] Y. Douini, J. Riffi, A. M. Mahraz, and H. Tairi, "An Image Registration Algorithm Based on Phase Correlation and the Classical Lucas–Kanade Technique," *Signal, Image Video Process.*, vol. 11, no. 7, pp. 1321–1328, 2017.

[63] F. Ghorbel, "Robust and Efficient Fourier – Mellin Transform Approximations for Gray-Level Image Reconstruction and Complete Invariant Description," *J. Comp. Vis. Image Understanding.*, vol. 83, no. 7, pp. 57–78, 2001.

[64] R. Gonzalez, "Improving Phase Correlation for Image Registration," *Image Vis. Comput. New Zeal. 2011*, 2011.

[65] C. E. Shannon, "Communication in the Presence of Noise," *Proc. IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[66] E. H. W. Meijering, W. J. Niessen, and M. A. Viergever, "Piecewise Polynomial Kernels For Image Interpolation : A Generalization Of Cubic Convolution," *Proceed. 1999 Int. Conf. Image Process.*, vol. 3, pp. 647–651, 1999.

[67] A. P. Mahmoudzadeh and N. H. Kashou, "Evaluation of Interpolation Effects on Upsampling and Accuracy of Cost Functions-Based Optimized Automatic Image Registration," *Int. J. Biomed. Imaging*, vol. 2013, 2013.

[68] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient Subpixel Image Registration Algorithms," *Opt. Lett.*, vol. 33, no. 2, p. 156, 2008.

[69] A. Yousef, J. Li, and M. Karim, "High-Speed Image Registration Algorithm with Subpixel Accuracy," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1796–1800, 2015.

[70] C. L. Wang et al., "Local Upsampling Fourier Transform for High Accuracy Image Rotation Estimation", *Advanced Materials Research.*, Vols. 268-270, pp. 1488-1493, 2011

[71] V. N. Dvornychenko, "Bounds on (Deterministic) Correlation Functions with Application to Registration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-5, no. 2, pp. 206–213, 1983.

[72] edited by G. M. Shepherd, The Synaptic Organization of the Brain. 5th ed. Oxford ; New York : Oxford University Press, 2004., 2004.

[73] P. Dhankhar and N. Sahu, "A Review and Research of Edge Detection Techniques for Image Segmentation," *Int. J. Comput. Sci. Mob. Comput.*, vol. 2, no. July, pp. 86–92, 2013.

[74] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. Syst. Man. Cybern.*, vol. 9, no. 1, pp. 62–66, 1979.

[75] M. Sezgin and B. Sankur, "Survey Over Image Thresholding Techniques and Quantitative Performance Evaluation," *J. Electronic Imaging.*, vol. 13, no., pp. 146–168, 2004.

[76] M. Nixon and A. S. Aguado, Feature Extraction & Image Processing for Computer Vision, 3rd ed. Academic Press, 2012.

[77] I. Sobel and G. Feldman, "A 3x3 Isotropic Gradient Operator for Image Processing," *Pattern Classification Scene Analysis*, pp. 271-272, 1973.

[78] T. Pouli, E. Reinhard, and D. W. Cunningham, Image Statistics in Visual Computing, 1st ed. Natick, MA, USA: A. K. Peters, Ltd., 2013.

[79] G. T. Shrivakshan and C. Chandrasekar, "A Comparison of Various Edge Detection Techniques used in Image Processing," *Int. J. Comput. Sci. Issues*, vol. 9, no. 5, pp. 269–276, 2012.

[80] S. Saluja, A. K. Singh, and S. Agrawal, "A Study of Edge-Detection Methods," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 2, no. 1, pp. 994–999, 2013.

[81] F. A. Pellegrino, W. Vanzella, and V. Torre, "Edge Detection Revisited," I*EEE Trans. Syst Man Cybern B Cybern.*, vol. 34, no. 3, pp. 1500–1518, 2004.

[82] R. Maini and H. Aggarwal, “Study and Comparison of Various Image Edge Detection Techniques,” *Int. J. Image Process.*, vol. 3, no. 1, pp. 1–11, 2009.
[83] M. Pitter, C. W. See, and M. Somekh, “Subpixel Microscopic Deformation Analysis using Correlation and Artificial Neural Networks.,” *Opt. Express*, vol. 8, no. 6, pp. 322–327, 2001.
[84] M. Pitter, C. See, J. Goh, and M. Somekh, “Focus Errors and Their Correction in Microscopic Deformation Analysis using Correlation.,” *Opt. Express*, vol. 10, no. 23, pp. 1361–7, 2002.
[85] Y. Lecun and K. Kavukcuoglu, “Convolutional Networks and Applications in Vision,” *ISCAS 2010 - 2010 IEEE Int Symp. Circuits and Systems: Nano-Bio Circuit Fabrics and System.*, pp. 253–256, 2010.
[86] S. Srinivas, R. K. Sarvadevabhatla, K. R. Mopuri, N. Prabhu, S. S. S. Kruthiventi, and R. V. Babu, “A Taxonomy of Deep Convolutional Neural Nets for Computer Vision,” *J. Frontiers in Robotics and AI.*, Vol. 2, no. 01, pp. 1–18, 2016.
[87] J. R. Movellan, “Tutorial on Gabor Filters,” vol. 49, pp. 1–23, 2002.
[88] E. Bengtsson and P. Malm, “Screening for Cervical Cancer Using Automated Analysis of PAP-Smears,” *Comput. Math. Methods Med.*, vol. 2014, p. 12, 2014.
[89] J. C. Histol, R. U. Deepak, R. R. Kumar, N. B. Byju, and P. N. Sharathkumar, “Cytology & Histology Computer Assisted Pap Smear Analyser for Cervical Cancer Screening using Quantitative Microscopy,” *J. Cytol. Histol.*, Vol. S:3, no. 04, 2015.
[90] H. Komagata, T. Ichimura, Y. Matsuta, M. Ishikawa, K. Shinoda, N. Kobayashi, A. Sasaki “Feature Analysis Of Cell Nuclear Chromatin Distribution In Support Of Cervical Cytology,” *J. Medical Imaging .*, vol. 4, no. 10, 2017.
[91] E. F. Badran, E. G. Mahmoud, and N. Hamdy, “An Algorithm for Detecting Brain Tumors in MRI Images,” *2010 Int. Conf. Comput. Eng. Syst.*, pp. 368–373, 2010.
[92] Tomasz Węgliński and A. Fabijańska, “Brain Tumor Segmentation From Mri Data Sets,” *Memstech*, no. 4, pp. 11–14, 2011.
[93] A. Fabijańska, “A Survey Of Subpixel Edge Detection Methods for Images of Heat-Emitting Metal Specimens,” *Int. J. Appl. Math. Comput. Sci.*, vol. 22, no. 3, pp. 695–710, 2012.
[94] G.-S. Xu, “Sub-Pixel Edge Detection Based on Curve Fitting,” *2009 Second Int. Conf. Inf. Comput. Sci.*, vol. 1, pp. 373–375, 2009.
[95] Q. Sun, Y. Hou, Q. Tan, C. Li, and M. Liu, “A Robust Edge Detection Method With Sub-Pixel Accuracy,” *Optik (Stuttg).*, vol. 125, no. 14, pp. 3449–3453, 2014.
[96] Y. You, “The Sub-Pixel Edge Detection Based on Improved Zernike Moment for Brain CT Image,” *Int. Conf on Electronic Information Technology and Intellectualization (ICEITI 2016).*, pp. 123–128, 2016.
[97] Q. Ying-Dong, C. Cheng-Song, C. San-Ben, and L. Jin-Quan, “A Fast Subpixel Edge Detection Method using Sobel-Zernike Moments Operator,” *Image Vis. Comput.*, vol. 23, no. 1, pp. 11–17, 2005.
[98] W. C. Lee *et al*, “Subpixel Edge Location using Orthogonal Fourier-Mellin Moments Based Edge Location Error Compensation Model,” *2008 Eighth Int. Conf. Intell. Systems Design and Applications*, 2008, vol. 3, pp. 346–351.
[99] D. G. Bailey, “Sub-Pixel Profiling,” *ICICS*, pp. 1311–1315, 2005.

[100] P. Rockett, “The Accuracy of Sub-Pixel Localisation in The Canny Edge Detector,” *Bmvc*, pp. 392–401, 1999.
[101] A. Fabijańska, “Gaussian-Based Approach to Subpixel Detection of Blurred and Unsharp Edges,” *2014 Fed. Conf. Comput. Sci. Inf. Syst*, 2014, pp. 641–650.
[102] P. V. C. Hough, “Machine Analysis of Bubble Chamber Pictures,” *Int. Conf. on High Energy Accelerators and Instrumentation*, 1959.
[103] T. L. Heath and Euclid, The Thirteen Books of Euclid’s Elements, Books 1 and 2. New York, NY, USA: Dover Publications, Inc., 1956.
[104] J. N. Cederberg, “Geometric Transformations of the Euclidean Plane,” *A Course in Modern Geometries*, New York, NY: Springer New York, 1989, pp. 74–126.
[105] M. Audin, “Affine Geometry,” in *Geometry*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 7–42, 2003.
[106] J. W. Cannon, “The Recognition Problem: What is a topological manifold?,” *Bull. Amer. Math. Soc.*, vol. 84, no. 5, pp. 832–866, 1978.
[107] J. Lee, Introduction to Riemannian Manifolds, 2nd ed. New York: Springer International Publishing, 2018.
[108] J. T. and J. Friedman, “A Projection Pursuit Algorithm for Exploratory Data Analysis,” *IEEE Trans. Comput.*, vol. 23, pp. 881–890, 1974.
[109] J. H. Friedman, “Exploratory Projection Pursuit,” *J. Am. Stat. Assoc.*, vol. 82, no. 397, pp. 249–266, 1987.
[110] I. T. Jolliffe, “Principal Component Analysis,” *Appl. Opt.*, vol. 44, no. May, p. 6486, 2005.
[111] I. T. Jolliffe, J. Cadima, and J. Cadima, “Principal Component Analysis : A Review and Recent Developments Subject Areas,” J. *Philosophical Trans. R. Soc. A: Mathem, Phys. Eng Sci.*, Vol. 374, no. 4, 2016.
[112] L. I. Smith, “A tutorial on Principal Components Analysis Introduction,” *Statistics (Ber).*, vol. 51, p. 52, 2002.
[113] J. Shlens, M, “A Tutorial on Principal Component Analysis,” 2014.
[114] W. F. Massy, “Principal Components Regression in Exploratory Statistical Research,” *J. Am. Stat. Assoc.*, vol. 60, no. 309, pp. 234–256, Mar. 1965.
[115] H. Shum, K. Ikeuchi, and R. Reddy, “Principal Component Analysis with Missing Data and its Application to Polyhedral Object Modeling,” *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. 17, no. 10, pp. 854–867, 1995.
[116] J. Karhunen and J. Joutsensalo, “Representation and Separation of Signals using Nonlinear PCA Type Learning,” *Neural Networks*, vol. 7, no. 1, pp. 113–127, 1994.
[117] C. K. Wikle, “Modern Perspectives on Statistics for Spatio-Temporal Data,” *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 7, no. 1, pp. 86–98, 2015.
[118] M. Turk and A. Pentland, “Eigenfaces for Recognition,” *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86, Jan. 1991.
[119] S. Lee, S. Jung, J. Kwon, and S. Hong, “Face Detection and Recognition using PCA,” *Proc. IEEE. IEEE Reg. 10 Conf. TENCON 99. “Multimedia Technol. Asia-Pacific Inf.” (Cat. No.99CH37030)*, vol. 1, pp. 84–87, 1999.
[120] P. Viola and M. Jones, “Robust Real-Time Face Detection,” *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.
[121] A. S. Tolba, A. H. El-baz, and A. A. El-Harby, “Face Recognition : A Literature Review,” *Int. J. Signal Process.*, vol. 2, no. 2, pp. 88–103, 2006.

[122] S. Aly and M. Hassaballah, “Face Recognition: Challenges, Achievements and Future Directions,” *IET Comput. Vis.*, vol. 9, no. 4, pp. 614–626, 2015.
[123] E. Oja and A. Hyva, “Independent Component Analysis : Algorithms and Applications,” vol. 13, pp. 411–430, 2000.
[124] J. Shlens, “A Tutorial on Independent Component Analysis,” vol. 94043, pp. 1–13, 2014.
[125] B. A. Draper, K. Baek, M. S. Bartlett, and J. R. Beveridge, “Recognizing Faces with PCA and ICA,” *Comput. Vis. Image Underst.*, vol. 91, no. 1–2, pp. 115–137, 2003.
[126] J. B. E. L. L. Anthony and T. J. Sejnowski, “The " Independent Components " of Natural Scenes are Edge Filters,” *J. Vision research.,* vol. 37, no. 01, pp. 3327–3338, 1997.
[127] J. H. van Hateren and D. L. Ruderman, “Independent Component Analysis of Natural Image Sequences Yields Spatio-Temporal Filters Similar To Simple Cells In Primary Visual Cortex,” *Proc. R. Soc. Lond. B*, vol. 265, no. 1412, pp. 2315–2320, 1998.
[128] O. Unit, O. Sciences, C. P. Group, and B. Columbia, “Nonlinear Principal Component Analysis by Neural Networks: Theory and Application to the Lorenz System,” *J. Climate.,* Vol. 13, no. 02, pp. 821–835, 2000.
[129] M. Linting, J. J. Meulman, P. J. F. Groenen, and A. J. van der Koojj, “Nonlinear Principal Components Analysis: Introduction and Application.,” *Psychol. Methods*, vol. 12, no. 3, pp. 336–358, 2007.
[130] L. I. U. Weixiang, Z. Nanning, and Y. O. U. Qubo, “Nonnegative Matrix Factorization and its Applications in Pattern Recognition,” *Chinese Sci. Bull.,* vol. 51, no. pp.7-18, 2006.
[131] D. D. Lee and H. S. Seung, “Learning the Parts of Objects by Non-Negative Matrix Factorization,” *Nature.,* vol. 401, no. 11, pp. 788–791, 2000.
[132] P. J. Huber, “Projection Pursuit,” *Ann. Stat.*, vol. 13, no. 2, pp. 435–475, 1985.
[133] C. Posse, “Projection Pursuit Exploratory Data Analysis,” *j. Computat. Stat. Data Anal.,* vol. 20, no. 02, 1995.
[134] D. Cook, A. Bujat, and J. Cabrerat, “Projection Pursuit Indexes Based on Orthonormal Function Expansions Projection Pursuit Indexes Based On Orthonormal Function Expansions,” *J. Comput. Graph. Stat.,* vol. 2, no. 3, pp. 225–250, 2015.
[135] E. K. Lee and D. Cook, “A Projection Pursuit Index for Large p Small n Data,” *Stat. Comput.*, vol. 20, no. 3, pp. 381–392, 2010.
[136] J. H. Friedman, W. Stuetzle, and A. Schroeder, “Projection Pursuit Density Estimation,” *J. Am. Stat. Assoc.*, vol. 79, no. 387, pp. 599–608, 1984.
[137] J. H. Friedman and W. Stuetzle, “Projection Pursuit Regression,” *J. Am. Stat. Assoc.*, vol. 76, no. 376, pp. 817–823, Dec. 1981.
[138] N. Intrator, “Combining Exploratory Projection Pursuit and Projection Pursuit Regression with Application to Neural Networks,” *Neural Computation* .,vol. 5, no. 07, pp. 443–455, 1993.
[139] R. O. Duda and P. E. Hart, “Use of the Hough Transformation To Detect Lines and Curves in Pictures,” *Commun. ACM.,* vol. 15, no. 1, 1972.
[140] D. H. Ballard, “Readings in Computer Vision: Issues, Problems, Principles, and Paradigms,” M. A. Fischler and O. Firschein, Eds. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1987, pp. 714–725.
[141] S. Tagzout, K. Achour, and O. Djekoune, “Hough Transform Algorithm for FPGA Implementation,” *J. Sig Process.,* vol. 81. no. 06, pp. 1295–1301, 2001.

[142] J. Basak, “Learning Hough Transform: A Neural Network Model,” *Neural computation.*, vol. 13. no. 04, pp 651–676, 2001.
[143] D. Borza, A. S. Darabant, and R. Danescu, “Real-Time Detection and Measurement of Eye Features from Color Images,” *Sensors (Basel).*, vol. 16, no. 7, p. 1105. 2016.
[144] N. Yokoya and A. Iwasaki, “Object Detection Based on Sparse Representation and Hough Voting for Optical Remote Sensing Imagery,” *IEEE J. Sel. Topics in Applied Earth Obser. Rem Sens.*, vol. 8, no. 5, pp. 2053–2062, 2015.
[145] D. Prasad, “Survey of the Problem of Object Detection In Real Images,” *Int. J. Image Process.*, no. 6, pp. 441–466, 2012.
[146] K. Xu, V. G. Kim, Q. Huang, N. Mitra, and E. Kalogerakis, “Data-Driven Shape Analysis and Processing,” *Comput. Graph. Forum.*, Vol. 36, no. 01, pp. 101 - 132, 2017.
[147] R. Girshick, J. Donahue, T. Darrell, J. Malik, and U. C. Berkeley, “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation,” *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, no. 11, pp. 580–587, 2013.
[148] J. A. Rodriguez-Serrano, D. Larlus, and Z. Dai, “Data-Driven Detection of Prominent Objects,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 10, pp. 1969–1982, 2016.
[149] H. Wei, Q. Yu, and C. Yang, “Shape-Based Object Recognition via Evidence Accumulation Inference,” *Pattern Recognit. Lett.*, vol. 77, no. 825, pp. 42-49, 2016.
[150] Z. Li, Y. Xia, and Z. Qu, “Data-Driven Background Representation Method to Video Surveillance,” *J. Opt. Soc. Am. A Opt. Image Sci. Vis.*, vol. 34, no. 2, pp. 193–202, 2017.
[151] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
[152] J. W. Goodman, “Introduction to Fourier Optics McGraw-Hill Series in Electrical and Computer Engineering,” *Quantum Semiclassical Opt. J. Eur. Opt. Soc. Part B*, vol. 8, no. 5, pp. 491, 1996.
[153] F. J. Harris , “On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transfrom,” *Proc. IEE*, vol. 66, no. 1, pp. 51–83, 1978.
[154] R. Goecke, “Optical Flow Estimation using Fourier Mellin Transform,” 2008.
[155] “cat.3D CAD Model Library GrabCAD.” [Online]. Available: https://grabcad.com/library/cat-41. [Accessed: 05-Feb-2019].
[156] T. Fujisawa and M. Ikehara, “High-Accuracy Image Rotation and Scale Estimation Using Radon Transform and Sub-Pixel Shift Estimation,” *in IEEE Access*, vol. 7, pp. 22719–22728, 2019.
[157] B. H, Harrison, “III The Radon Transform and Its Applications,” *In Prog. Opt*, vol. 21, No. C, pp. 217-286, 1984.
[158] A. HajiRassouliha et al., “Subpixel Phase Based Image Registration using S aitzky-Golay differentiators in Gradient-Correlation,” *J. Comput. Vis. Image Understanding*, vol. 170, no. 11, pp. 1077–3142, 2018.
[159] B.B, Kimia, “Differential Geometry in Edge Detection: Accurate Estimation of Position, Orientation and Curvature,” IEEE *Trans. Pat. Analysis. Mach. Intel*, vol. 41, no.7, pp. 1573–1586, 2019.
[160] X. Bai *et al.*, “Active Skeleton for Nonrigid object Detection,” I*EEE 12th Int. Conf. Comp. Vis (ICCV)*, pp. 575-582, 2009.

Appendix

Appendix A1 – Noise Model

The simple imaging module in figure A1.1 illustrates how an object recognition algorithm could receive distorted signals.

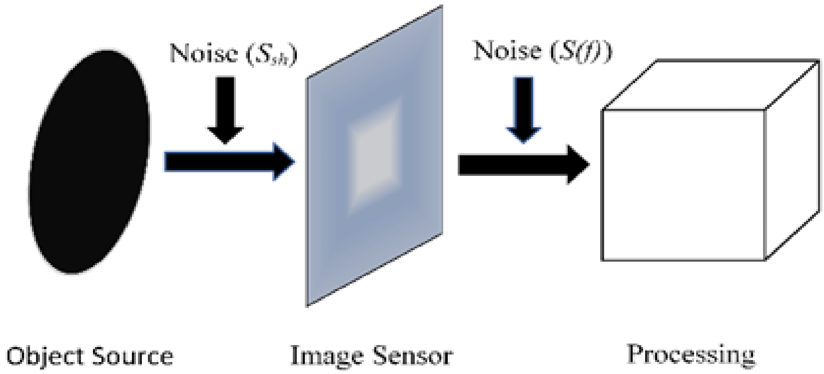


Figure A1.1: Noise sources within an imaging module: white noise s_{sh} and low frequency noise $S(f)$.

Firstly, light reflected from an object is captured by the image sensor of the camera; CMOS or CCD. The random arrival of photons from the object do not interact with each pixel of the sensor at the same time. The efficiency in converting photons to electrons, which varies with the incident wavelength, is controlled by the quantum efficiency of the sensor. When the camera sensor, the imperfect electronics of the camera and the 1/*f* pattern in the power spectrum of images are considered, the dominating response in the bandlimited output of the camera is governed by white and low frequency distributed noise.

White Noise Model

The probability (P) of the number of photons (N) interacting with a pixel of the sensor can be accurately modelled following Poisson statistics. For greater numbers of photons interacting with a pixel ($N > 20$), the probability of detecting interacting photons (q) increases and the Poisson model can be approximated using Gaussian statistics, Eq. (A2a). The variance (σ^2) of the distribution relates to the average number of incident photons.

$$P(q) \approx \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{\frac{-(q-N)^2}{2\sigma^2}}. \qquad \text{(A2a)}$$

Low Frequency Noise Model

Random signal fluctuations S are dominated by the lowest frequency portion of the noise-modulated spectrum. The spectral density of low frequency noise is

$$S(f) \approx \frac{1}{f^{\beta_{dB}}}, \qquad \text{(A2b)}$$

which increases as frequency (f) decreases. The term β_{dB} is a definition of the power spectral density of the noise over the bandwidth of the signal. The amplitude of $S(f)$ is determined from the inputs normalised signal strength. To vary the influences of $S(f)$, the matched signal is scaled as a percentage between 1% and 100%.

Appendix A2 – RM Edge Detection Symmetry Curves

The symmetry curves presented here parameterise the range of error in the symmetry of the second order model calculated from Eq. (4.8) of a detected as edge using the RM method presented in chapter 4. These figures accompany figures 4.17.1-3 and figure 4.18…

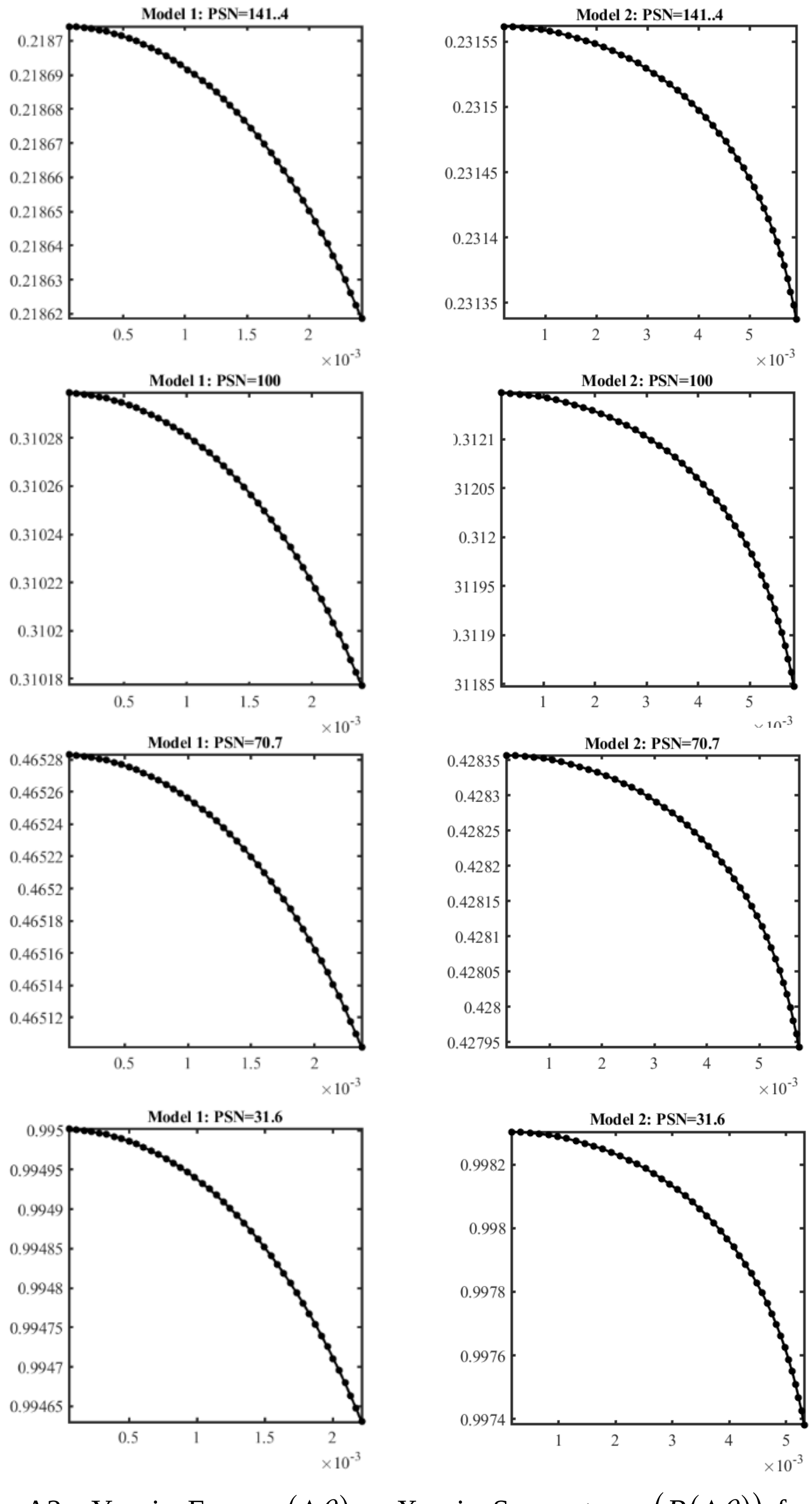


Figure A2a: Y-axis: Error $\sigma(\Delta\theta)$ vs X-axis: Symmetry $\chi(P(\Delta\theta))$ for iterative RM model 1 - 2; edge direction 63.4° distortion (PSN) $\sqrt{N_{pht}} = 141.4, 100, 70.7, 31.6$.

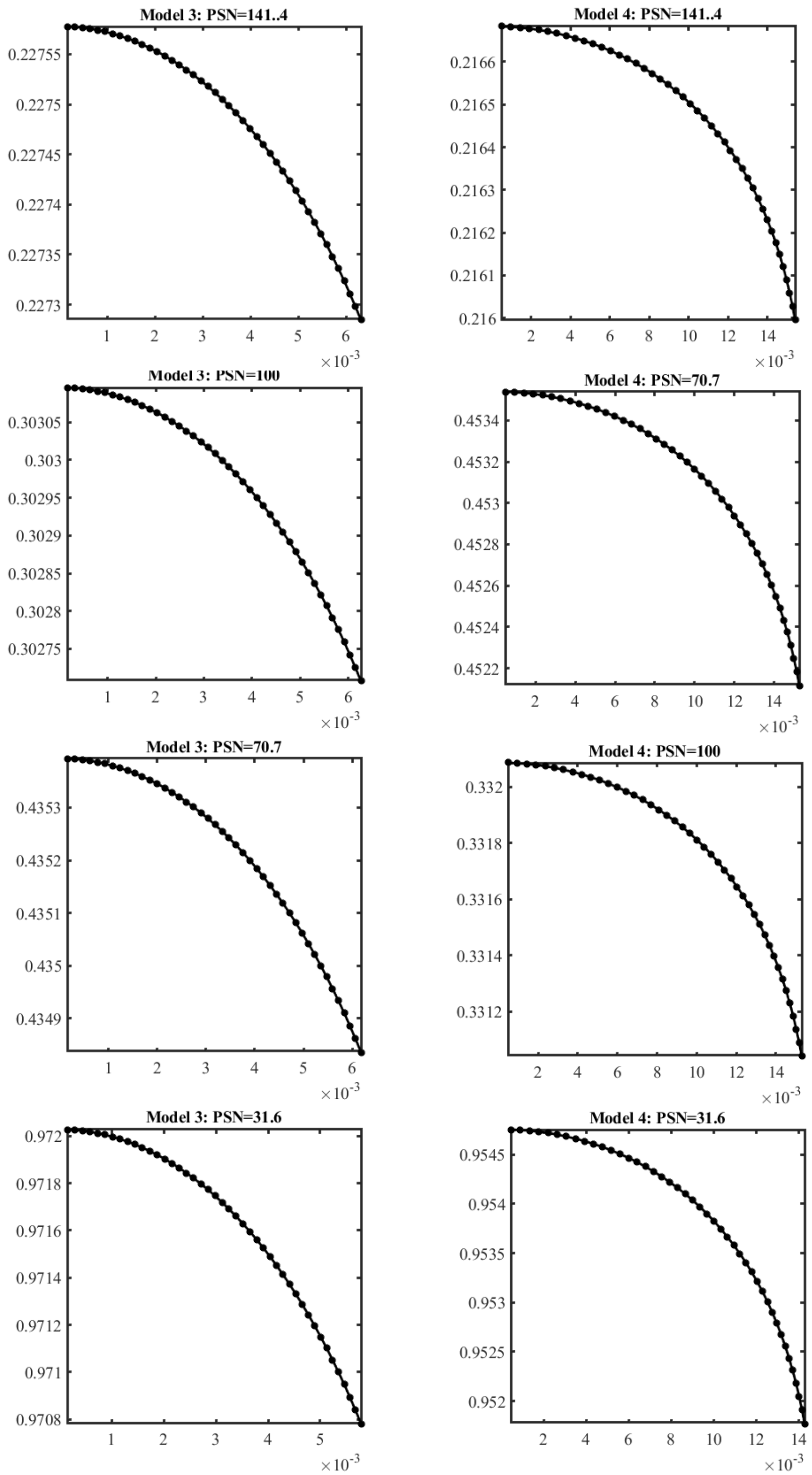


Figure A2b: Y-axis: Error $\sigma(\Delta\theta)$ vs X-axis: Symmetry $\chi(P(\Delta\theta))$ for iterative RM model 3 - 4; edge direction 63.4° distortion (PSN) $\sqrt{N_{pht}} = 141.4, 100, 70.7, 31.6$.

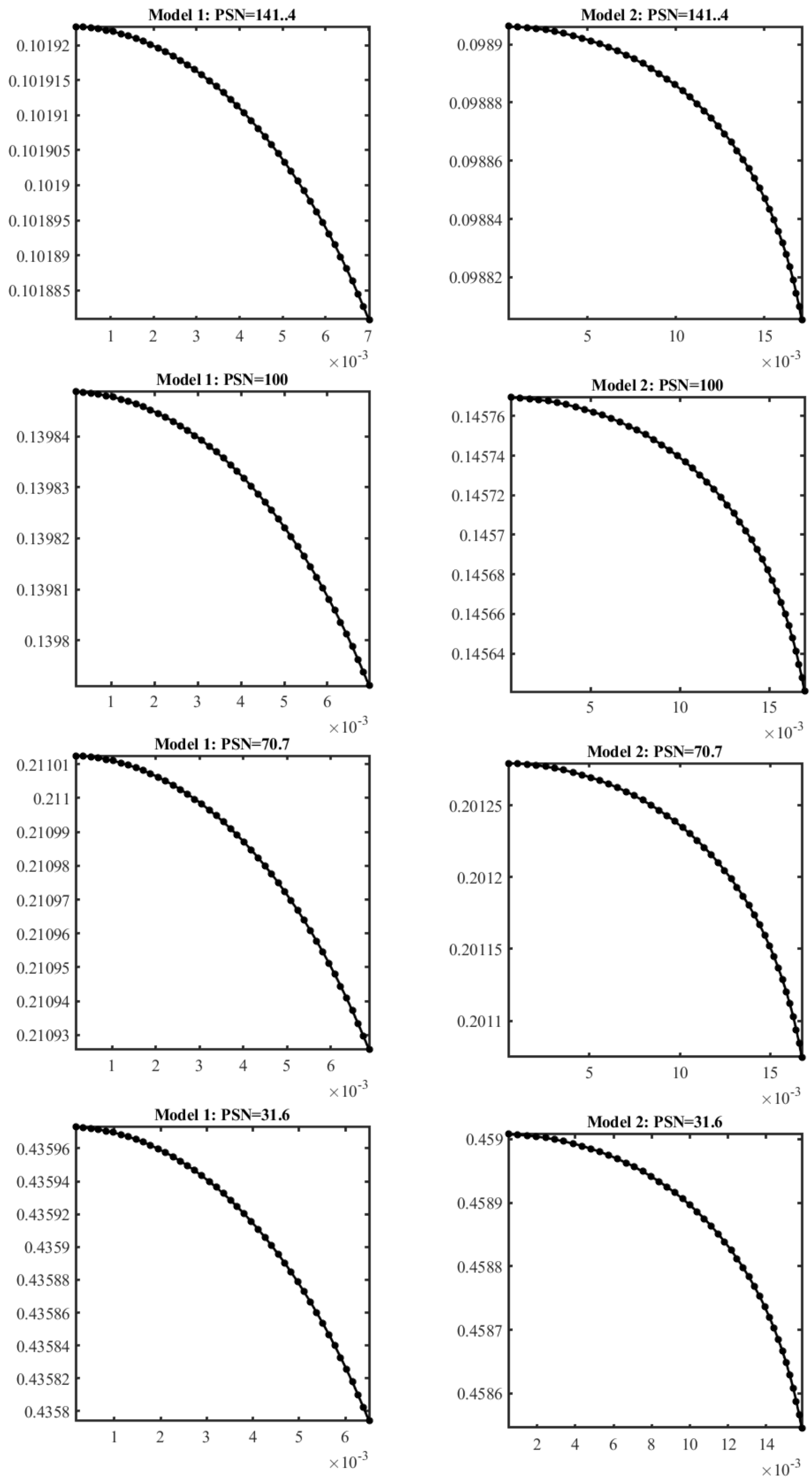


Figure A2c: Y-axis: $\sigma(\Delta\theta)$ vs X-axis: Symmetry $\chi(P(\Delta\theta))$ for iterative

RM model 1 - 2; edge direction 71.6° distortion (PSN)

$\sqrt{N_{pht}} = 141.4, 100, 70.7, 31.6$.

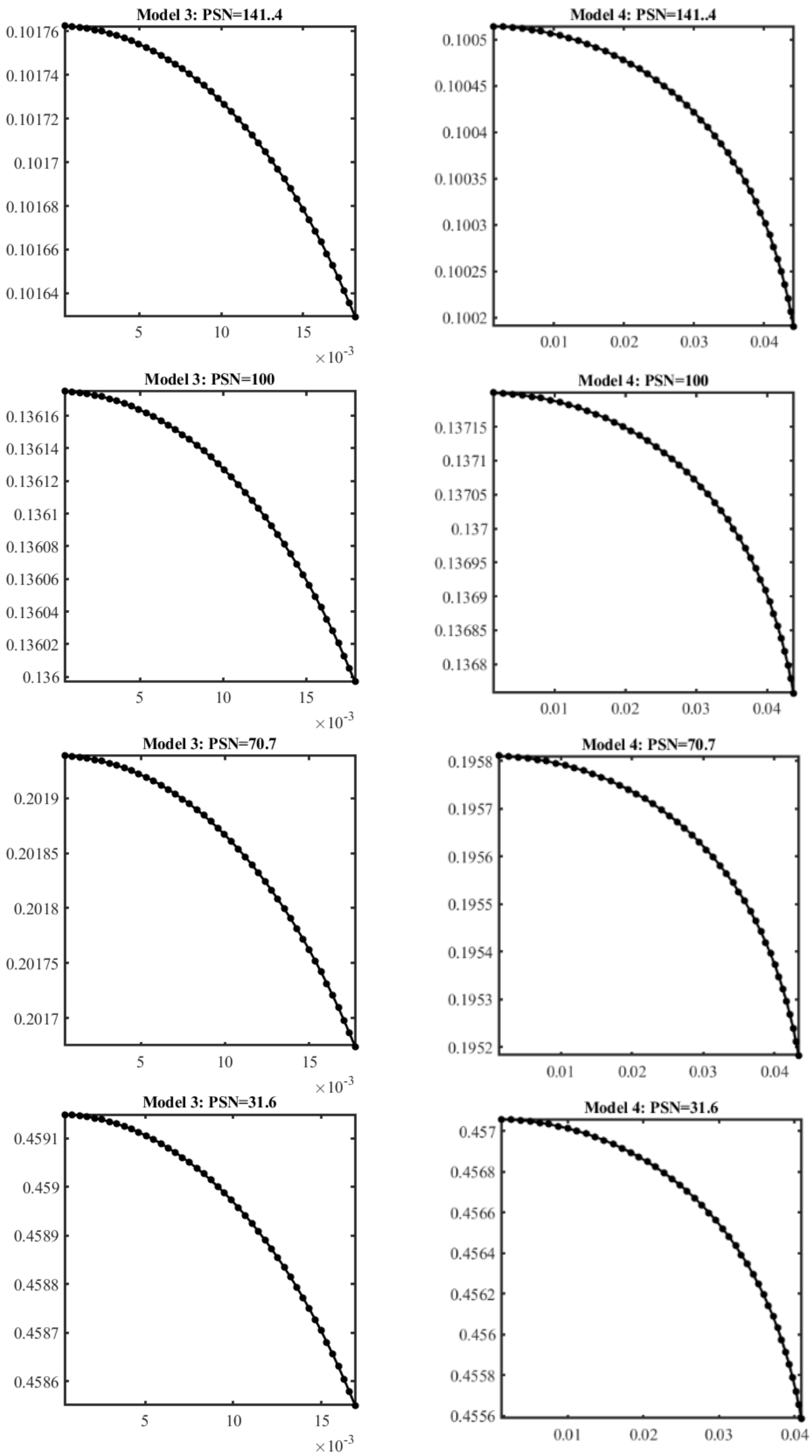


Figure A2d: Y-axis: Error $\sigma(\Delta\theta)$ vs X-axis: Symmetry $\chi(P(\Delta\theta))$ for iterative RM model 3 - 4; edge direction 71.6° distortion (PSN) $\sqrt{N_{pht}} = 141.4, 100, 70.7, 31.6$.

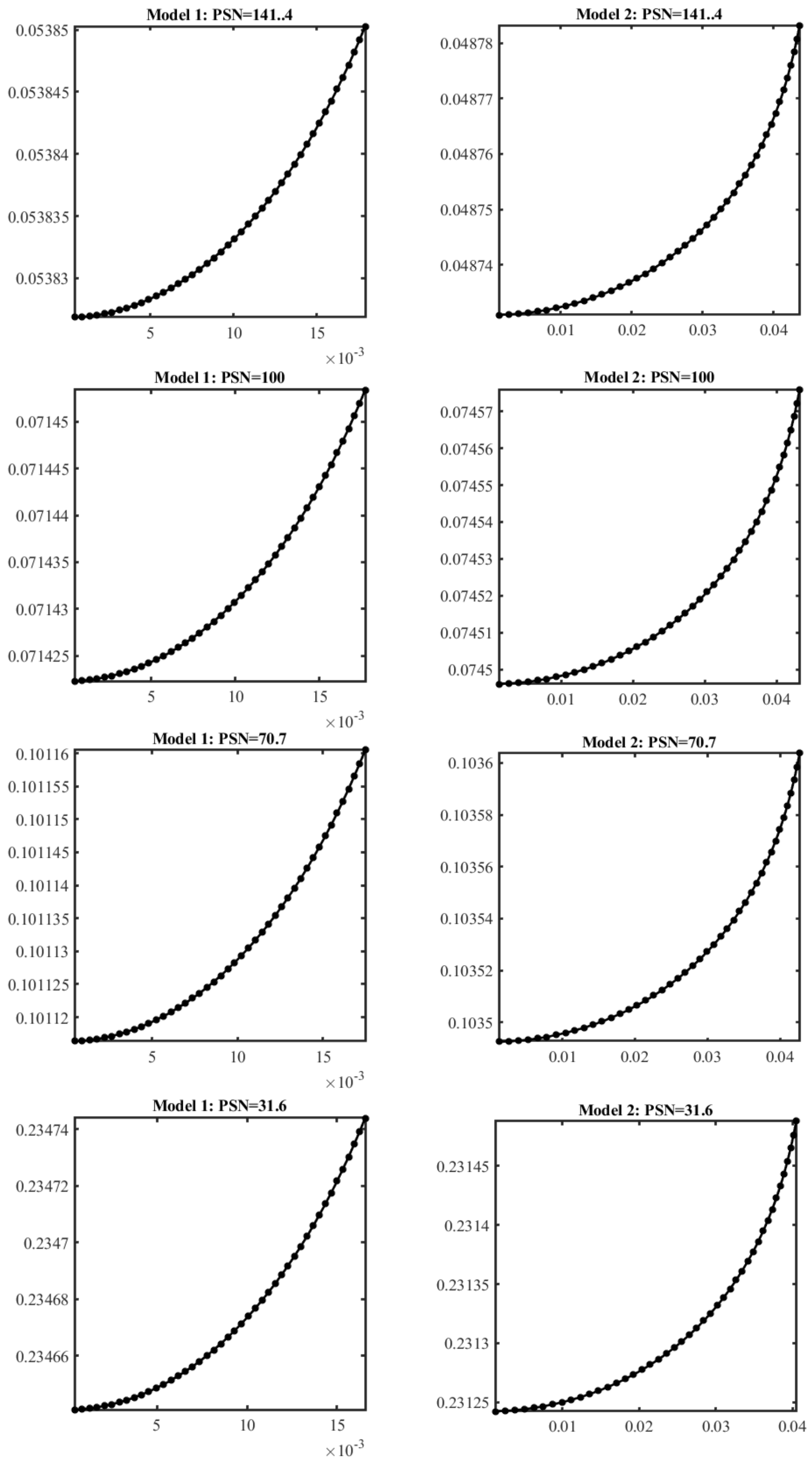


Figure A2e: Y-axis: Error $\sigma(\Delta\theta)$ vs X-axis: Symmetry $\chi(P(\Delta\theta))$ for iterative RM model 1 - 2; edge direction 75.96° distortion (PSN) $\sqrt{N_{pht}} = 141.4, 100, 70.7, 31.6$.

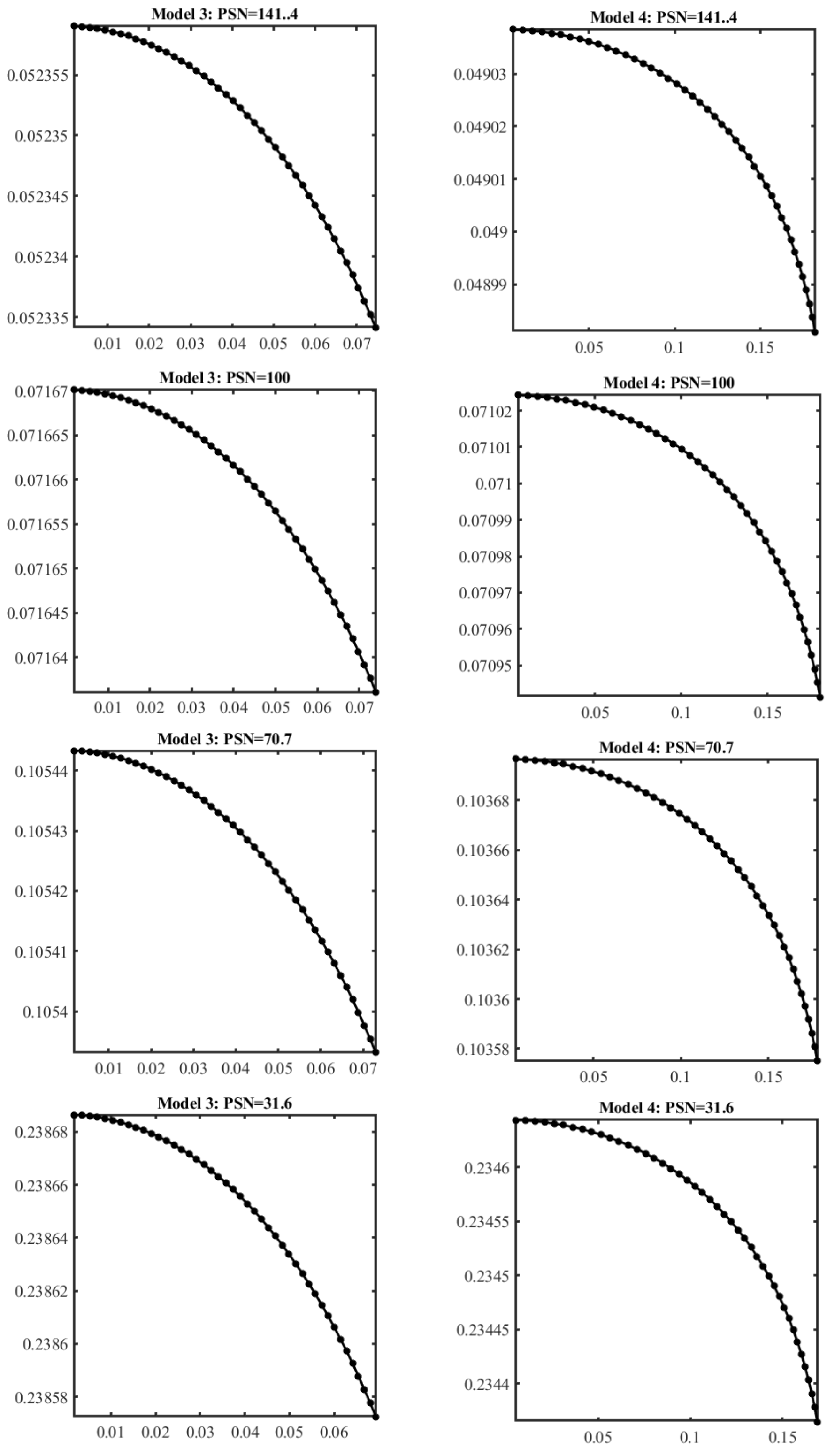


Figure A2f: Y-axis: $\sigma(\Delta\theta)$ vs X-axis: Symmetry $\chi(P(\Delta\theta))$ for iterative RM model 3 - 4; edge direction 75.96° distortion (PSN) $\sqrt{N_{pht}} = 141.4, 100, 70.7, 31.6$.

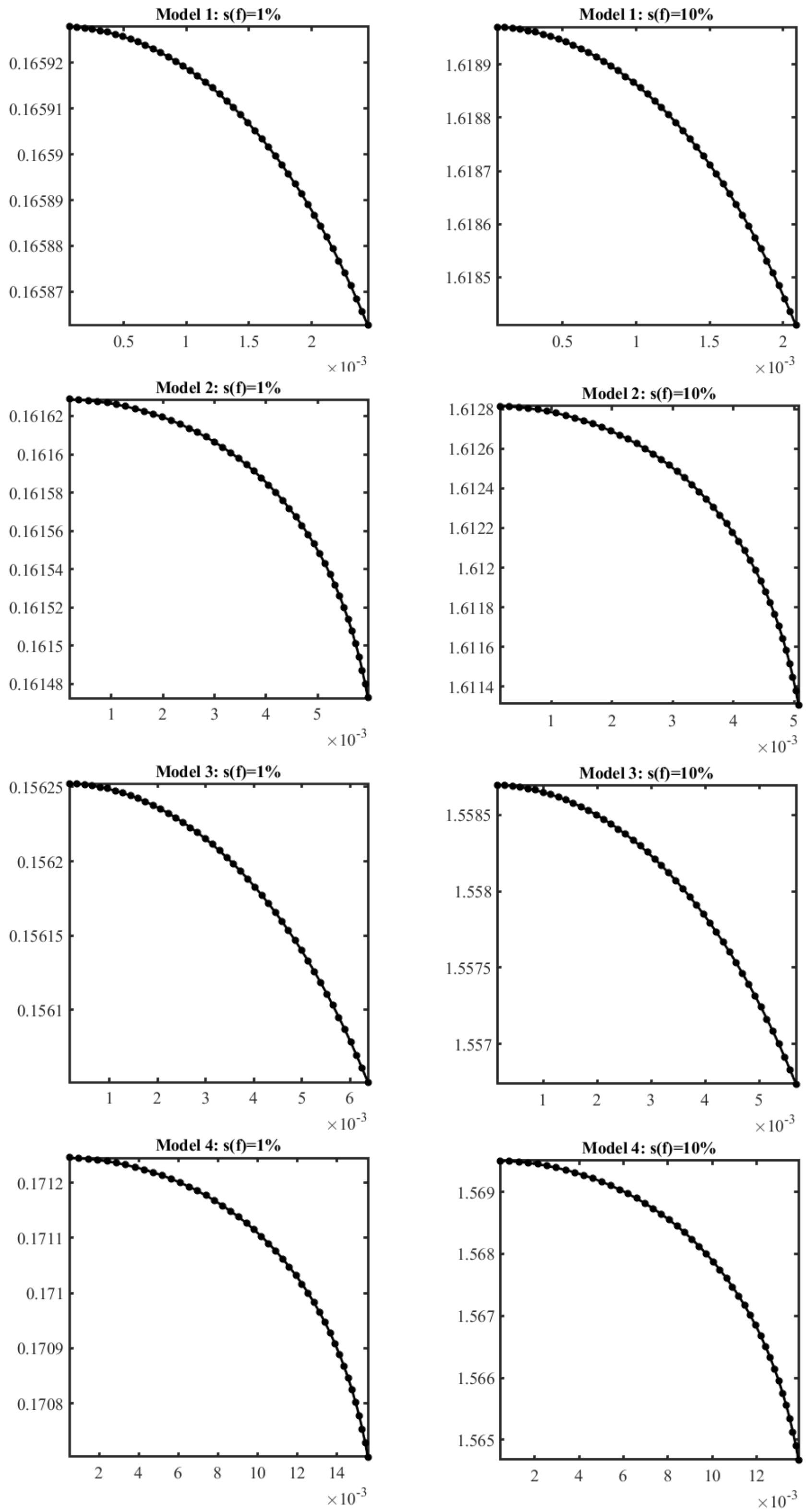


Figure A2g: Y-axis: Error $\sigma\left(\Delta\theta\right)$ vs X-axis: Symmetry $\chi\left(P\left(\Delta\theta\right)\right)$ for iterative RM model 1 - 4; edge 63.4° distortion $S\left(f\right)=1\%,10\%$.

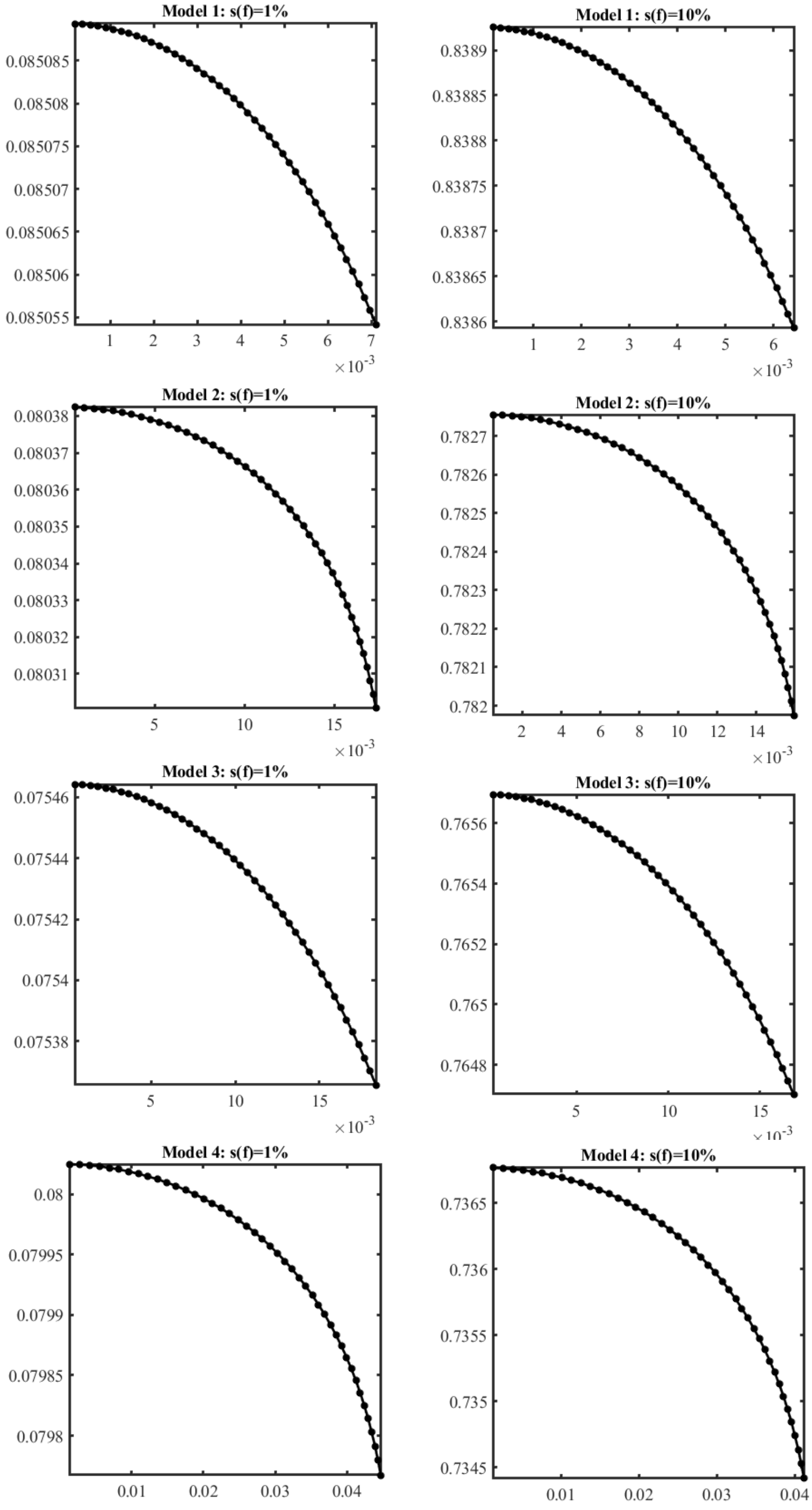


Figure A2h: Y-axis: Error $\sigma(\Delta\theta)$ vs X-axis: Symmetry $\chi(P(\Delta\theta))$ for iterative RM model 1 - 4; edge direction 71.6° distortion $S(f)=1\%,10\%$.

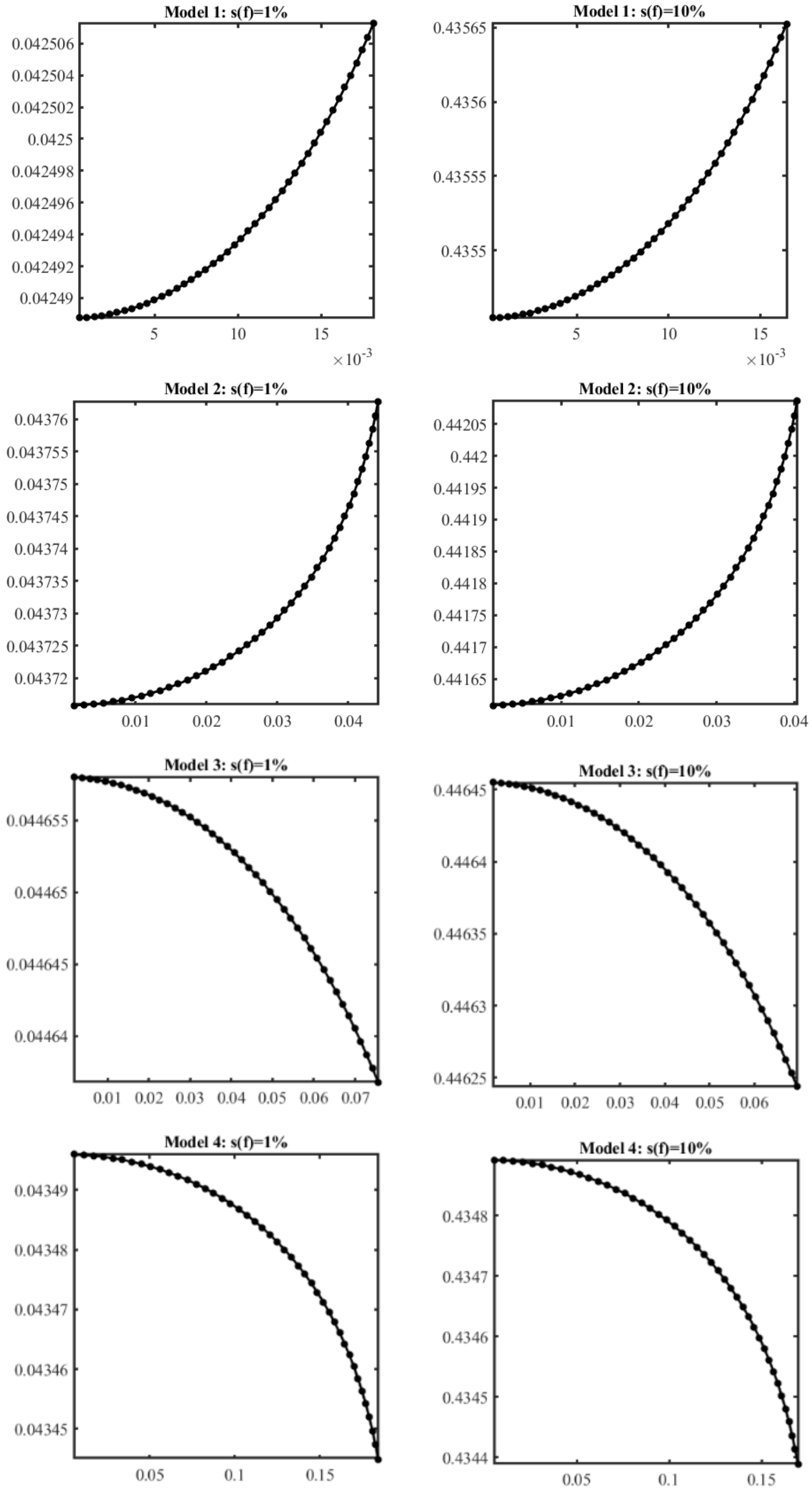


Figure A2i: Y-axis: Error $\sigma(\Delta\theta)$ vs X-axis: Symmetry $\chi(P(\Delta\theta))$ for iterative RM model 1 - 4; edge direction 75.96° distortion $S(f)=1\%,10\%$.

Appendix A3 – Measured Feature Distances

Front profile (cm): $\Omega_{F(1:25)}$

Eye Width
(0.45 0.31 0.64 0.37 0.53 0.517 0.36 0.59 0.42 0.52 0.3 0.69 0.58 0.64 0.6 0.49 0.72 0.48 0.68 0.72 0.7 0.46 0.68 0.69 0.32)/2
Eye Height
(0.4 0.27 0 0.29 0.38 0.42 0 0.39 0.2 0.43 0.24 0.39 0.45 0.48 0.33 0.25 0.66 0.36 0.43 0.62 0.43 0.36 0.36 0.58 0.18)/2
Head Width
(2.75 1.96 3.36 2.49 3.26 3.44 3.36 3.4 3.2 2.75 2.22 3.49 3.41 3.76 3.33 3.76 3.81 2.84 3.31 4.42 4.37 3.68 4.13 3.36 2.95)/2
Head Height
(2.86 1.77 3.55 2.36 2.63 2.88 3.25 3.42 2.83 2.59 2.09 3.04 2.96 2.91 3.2 3.28 3.28 2.91 3.4 3.7 3.76 2.83 3.98 3.23 2.95)/2
Nose Width
0.29 0.21 0.61 0.34 0.41 0.4 0.42 0.5 0.32 0.45 0.16 0.48 0.53 0.45 0.45 0.4 0.56 0.42 0.56 0.61 0.56 0.34 0.56 0.48 0.37
Nose Height
0.66 0.45 0.58 0.53 0.56 0.64 1.11 0.82 0.61 0.77 0.34 0.9 0.74 0.74 0.71 0.74 0.9 0.56 0.79 0.94 0.85 0.48 0.93 0.82 0.65
Ear Length
1.19 0.82 1.69 1.07 1.15 1.27 1.72 1.29 1.38 1.28 0.69 1.36 1.43 1.45 1.46 1.1 1.74 1.4 1.3 2.06 1.65 1.03 1.77 1.44 0.86
Ear Width
0.82 0.85 0.87 1.23 0.94 0.89 1.08 1.15 0.96 0.95 0.73 1.14 0.94 1 0 0.88 1.15 0.76 1.16 1.35 0.96 0.75 0.61 1.06 0.55
Eye Separation
1.32 0.69 1.72 1.11 1.32 1.22 1.38 1.6 1.19 1.24 0.85 1.38 1.35 1.46 1.3 1.46 1.69 1.19 1.64 1.88 1.8 1.03 1.73 1.56 1.17
Ear Separation
2.43 1.59 2.91 2.14 2.35 2.59 3.31 2.84 2.22 1.93 1.85 2.36 2.73 2.51 3.05 2.83 3.02 2.55 2.73 3.39 3.44 2.33 3.36 2.7 3.4
Eye - Nose separation
1.01 0.69 1.18 0.99 1.1 1.27 1.68 1.23 1.02 0.98 0.64 1.3 1.11 1.16 1.17 1.1 1.44 0.96 1.39 1.33 1.35 0.81 1.41 1.29 0.84
Eye - Ear separation
1.48 1.11 1.65 1.5 1.23 1.22 1.25 1.62 1.34 1.25 0.9 1.4 1.36 1.38 2.25 2.58 1.71 1.57 1.73 1.91 1.83 1.4 1.81 1.46 1.43

Rotated profile (base measurement unit unknown): $\Omega_{F(i)}(\theta,\phi)$

$\Omega_F\left(0^\circ,\phi_{-50^\circ:50^\circ}\right)$

Ear Separation
0.8580 0.8580 0.8580 0.8580 0.8580 0.8580 0.8580 0.8580 0.8580 0.8580 0.8580
Eye Separation
0.4675 0.4675 0.4675 0.4675 0.4675 0.4675 0.4675 0.4675 0.4675 0.4675 0.4675
Nose Centre to Ear Separation Mid-point
0.3678 0.5377 0.6907 0.8227 0.9310 1.0084 1.0566 1.0732 1.0560 1.0074 0.9280
Nose Centre to Ear Centre
0.5674 0.6881 0.8130 0.9278 1.0239 1.0965 1.1403 1.1552 1.1398 1.0949 1.0225
Cross Separation: Eye Centre to Ear Left to Right
0.7603 0.8261 0.8860 0.9418 0.9857 1.0142 1.0254 1.0180 0.9904 0.9522 0.8986

$\Omega_F\left(10^\circ,\phi_{-50^\circ:50^\circ}\right)$

Ear Separation
0.8527 0.8504 0.8483 0.8465 0.8454 0.8450 0.8454 0.8465 0.8483 0.8504 0.8527
Eye Separation
0.4645 0.4633 0.4621 0.4612 0.4605 0.4603 0.4605 0.4612 0.4621 0.4633 0.4645
Nose Centre to Ear Separation Mid-point
0.3786 0.5533 0.6958 0.8282 0.9347 1.0123 1.0588 1.0730 1.0548 1.0049 0.9251
Nose Centre to Ear Centre
0.4780 0.6102 0.7480 0.8760 0.9854 1.0705 1.1273 1.1535 1.1481 1.1113 1.0446]
Cross Separation: Eye Centre to Ear Left to Right
0.6886 0.7520 0.8218 0.8891 0.9471 0.9910 1.0177 1.0252 1.0131 0.9822 0.9346

$\Omega_F\left(20^\circ,\phi_{-50^\circ:50^\circ}\right)$

Ear Separation
0.8370 0.8281 0.8195 0.8125 0.8079 0.8063 0.8079 0.8125 0.8222 0.8281 0.8370
Eye Separation
0.4560 0.4511 0.4464 0.4425 0.4400 0.4391 0.4400 0.4426 0.4464 0.4511 0.4560
Nose Centre to Ear Separation Mid-point
0.3920 0.5535 0.7123 0.8458 0.9497 1.0145 1.0610 1.1494 1.0516 0.9957 0.9156
Nose Centre to Ear Centre
0.3896 0.5354 0.6871 0.8281 0.9505 1.0461 1.1135 1.1716 1.1524 1.1225 1.0608
Cross Separation: Eye Centre to Ear Left to Right
0.6067 0.6693 0.7460 0.8252 0.8978 0.9574 0.9998 1.0221 1.0231 1.0026 0.9620

$\Omega_F\left(30^{\circ},\phi_{-50^{\circ}:50^{\circ}}\right)$

Ear Separation
0.8125 0.7926 0.7734 0.7574 0.7468 0.7431 0.7468 0.7574 0.7734 0.7904 0.8966
Eye Separation
0.4426 0.4317 0.4212 0.4125 0.4067 0.4046 0.4067 0.4125 0.4212 0.4317 0.4426
Nose Centre to Ear Separation Mid-point
0.4518 0.5763 0.7432 0.8655 0.9205 1.0260 1.0640 1.0715 1.0477 0.9876 0.9368
Nose Centre to Ear Centre
0.3066 0.4708 0.6365 0.7890 0.9509 1.0240 1.1016 1.1449 1.1544 1.1279 1.0717
Cross Separation: Eye Centre to Ear Left to Right
0.5155 0.5781 0.6609 0.7530 0.8408 0.9164 0.9744 1.0114 1.0252 1.0152 0.9808

$\Omega_F\left(40^{\circ},\phi_{-50^{\circ}:50^{\circ}}\right)$

Ear Separation
0.7814 0.7468 0.7128 0.6838 0.6642 0.6573 0.6642 0.6838 0.7128 0.8136 0.8490
Eye Separation
0 0.4067 0.3881 0.3723 0.3616 0.3578 0.3616 0.3723 0.3882 0.4068 0.4256
Nose Centre to Ear Separation Mid-point
0.4439 0.6169 0.7405 0.8798 0.9807 1.0440 1.0720 1.0678 1.0258 0.9780 0.8908
Nose Centre to Ear Centre
0.2415 0.4258 0.6028 0.7634 0.9015 1.0128 1.0937 1.1415 1.1550 1.1336 1.0768
Cross Separation: Eye Centre to Ear Left to Right
0.4161 0.4753 0.5700 0.6769 0.7807 0.8722 0.9454 0.9962 1.0221 1.0220 0.9958

$\Omega_F\left(50^{\circ},\phi_{-50^{\circ}:50^{\circ}}\right)$

Ear Separation
0.7468 0.6948 0.6420 0.5956 0.5632 0.5515 0.5632 0.5956 0.6420 0.6948 0.7468
Eye Separation
0 0 0 0.3242 0.3065 0.3001 0.3065 0.3242 0.3496 0.3784 0
Nose Centre to Ear Separation Mid-point
0.5119 0.6008 0.8006 0.8754 0.9803 1.0440 1.0710 1.0658 1.0123 0.9443 0.8656
Nose Centre to Ear Centre
0.2166 0.4103 0.5916 0.7549 0.8953 1.0085 1.0910 1.1404 1.1551 1.1347 1.0799
Cross Separation: Eye Centre to Ear Left to Right
0.3100 0.3694 0.4791 0.6034 0.7237 0.8303 0.9171 0.9800 1.0163 1.0247 1.0048

Appendix A4 – Feature Ratio Errors

$\Omega_T\left(35^{\circ},-25^{\circ};R_{1:5}\right)$

Out-of-plane rotation $\theta=\left(0^{\circ}:50^{\circ};\Delta\theta=10^{\circ}\right)$ vs elevation angle

$\phi=\left(-50^{\circ}:50^{\circ};\Delta\phi=10^{\circ}\right)$: red square indicates rotation error minimum.

R_1

	θ					
	1.935	2.036	2.222	2.480	3.586	3.868
	1.167	1.237	1.350	1.437	1.522	2.492
	0.765	0.819	0.843	0.838	0.831	2.021
	0.523	0.551	0.539	0.494	0.419	0.319
	0.370	0.381	0.348	0.280	0.180	0.113
ϕ	0.279	0.276	0.236	0.155	0.065	0.167
	0.232	0.219	0.171	0.088	0.057	0.204
	0.222	0.201	0.118	0.071	0.052	0.184
	0.249	0.220	0.173	0.098	0.043	0.114
	0.309	0.276	0.232	0.171	0.251	0.084
	0.414	0.377	0.338	0.445	0.383	1.872

R_2

	θ					
	1.062	1.117	1.219	1.359	1.295	1.295
	0.644	0.682	0.743	0.791	0.836	1.295
	0.425	0.454	0.467	0.464	0.460	1.295
	0.293	0.308	0.301	0.277	0.235	0.180
	0.210	0.215	0.197	0.160	0.104	0.060
ϕ	0.160	0.158	0.136	0.092	0.037	0.083
	0.134	0.127	0.1	0.055	0.023	0.103
	0.129	0.117	0.071	0.045	0.019	0.092
	0.143	0.127	0.1	0.060	0.023	0.054
	0.176	0.158	0.134	0.102	0.075	0.04
	0.234	0.213	0.192	0.160	0.146	1.295

R_3

	θ					
ϕ	2.359	2.30	2.228	2.085	3.076	3.290
	1.907	1.835	1.788	1.670	1.545	2.959
	1.512	1.467	1.356	1.180	1.060	2.890
	1.177	1.128	0.993	0.796	0.570	0.451
	0.905	0.857	0.710	0.588	0.256	0.579
	0.711	0.661	0.535	0.290	0.227	0.826
	0.591	0.546	0.419	0.190	0.228	0.823
	0.549	0.515	0.219	0.207	0.106	0.581
	0.591	0.565	0.485	0.336	0.175	0.165
	0.709	0.695	0.656	0.568	0.501	0.369
	0.903	0.90	0.887	0.819	0.85	2.817

R_4

	θ					
ϕ	2.145	2.319	2.519	2.738	3.395	3.405
	1.801	1.967	2.119	2.237	2.293	3.223
	1.478	1.612	1.704	1.741	1.691	3.125
	1.189	1.285	1.318	1.274	1.122	0.818
	0.949	1.01	0.984	0.775	0.619	0.215
	0.76	0.791	0.728	0.562	0.241	0.413
	0.659	0.650	0.557	0.356	0.070	0.623
	0.621	0.588	0.426	0.277	0.068	0.572
	0.658	0.606	0.499	0.317	0.076	0.335
	0.767	0.704	0.604	0.465	0.346	0.135
	0.944	0.876	0.791	0.748	0.629	3.0360

R_5

	θ					
ϕ	1.644	1.639	1.699	1.818	2.855	2.837
	1.156	1.226	1.333	1.457	1.629	2.738
	0.883	0.972	1.061	1.160	1.272	2.680
	0.692	0.767	0.821	0.860	0.876	0.848
	0.557	0.605	0.615	0.578	0.491	0.311
	0.473	0.488	0.453	0.356	0.181	0.258
	0.438	0.419	0.347	0.206	0.067	0.481
	0.453	0.402	0.304	0.144	0.1190	0.511
	0.519	0.437	0.328	0.172	0.103	0.388
	0.612	0.520	0.410	0.275	0.196	0.226
	0.744	0.646	0.546	0.488	0.382	2.593

$\Omega_T\left(35^\circ, 25^\circ; R_{1:5}\right)$

Out-of-plane rotation $\theta=\left(0^\circ:50^\circ;\Delta\theta=10^\circ\right)$ vs elevation angle

$\phi=\left(-50^\circ:50^\circ;\Delta\phi=10^\circ\right)$: red square indicates rotation error minimum.

R_1

	θ					
	1.547	1.596	1.737	1.964	3.137	3.392
	0.766	0.777	0.848	0.902	0.964	2.137
	0.393	0.379	0.344	0.293	0.265	1.856
	0.257	0.207	0.142	0.098	0.161	0.299
	0.275	0.243	0.242	0.297	0.404	0.567
ϕ	0.324	0.317	0.344	0.417	0.542	0.709
	0.356	0.365	0.406	0.485	0.605	0.760
	0.360	0.381	0.454	0.506	0.612	0.746
	0.334	0.366	0.412	0.487	0.574	0.679
	0.282	0.321	0.370	0.434	0.437	0.579
	0.208	0.254	0.302	0.312	0.359	1.893

R_2

	θ					
	0.843	0.870	0.947	1.070	1.473	1.473
	0.418	0.424	0.462	0.491	0.525	1.473
	0.214	0.207	0.188	0.160	0.145	1.473
	0.140	0.113	0.077	0.053	0.088	0.163
	0.149	0.132	0.132	0.162	0.220	0.309
ϕ	0.176	0.172	0.187	0.227	0.295	0.387
	0.193	0.198	0.221	0.264	0.330	0.414
	0.196	0.207	0.247	0.275	0.334	0.407
	0.182	0.199	0.226	0.265	0.312	0.370
	0.153	0.174	0.201	0.235	0.278	0.315
	0.113	0.138	0.164	0.198	0.225	1.470

R_3

	θ					
	1.731	1.641	1.526	1.320	2.288	2.506
	1.268	1.158	1.070	0.90	0.732	2.166
	0.879	0.797	0.637	0.392	0.238	2.144
	0.577	0.490	0.316	0.105	0.321	0.728
	0.384	0.319	0.240	0.294	0.713	1.270
ϕ	0.332	0.313	0.350	0.540	0.952	1.587
	0.351	0.366	0.446	0.641	1.015	1.610
	0.360	0.388	0.610	0.638	0.925	1.389
	0.329	0.364	0.428	0.551	0.716	0.985
	0.285	0.324	0.370	0.433	0.505	0.637
	0.322	0.364	0.401	0.393	0.466	2.146

R_4

	θ					
	1.060	1.115	1.261	1.462	2.107	2.110
	0.663	0.736	0.850	0.948	0.991	1.935
	0.380	0.402	0.43	0.446	0.381	1.887
	0.297	0.214	0.146	0.089	0.203	0.533
	0.421	0.351	0.348	0.545	0.705	1.197
ϕ	0.571	0.540	0.595	0.754	1.092	1.696
	0.669	0.675	0.763	0.960	1.316	1.923
	0.703	0.736	0.889	1.040	1.368	1.879
	0.668	0.721	0.827	1.008	1.274	1.646
	0.566	0.633	0.736	0.878	1.020	1.316
	0.411	0.485	0.578	0.626	0.768	1.954

R_5

	θ					
	1.347	1.178	1.074	1.035	1.906	1.831
	0.809	0.659	0.585	0.544	0.611	1.718
	0.585	0.454	0.317	0.207	0.234	1.688
	0.541	0.424	0.306	0.205	0.175	0.263
	0.571	0.499	0.454	0.471	0.561	0.797
ϕ	0.610	0.584	0.603	0.688	0.879	1.254
	0.624	0.640	0.705	0.839	1.083	1.511
	0.601	0.653	0.738	0.906	1.157	1.549
	0.534	0.620	0.731	0.892	1.118	1.427
	0.447	0.547	0.667	0.815	0.947	1.220
	0.337	0.447	0.564	0.632	0.792	1.761

Appendix A5 – Publications

Submitted to Elsevier Signal Processing: Image Communication (currently under review)

Nicholas Wells, Chung. W. See, “Image Signature Analysis: Nonuniformly Sampled Adaptive Correlation Filtering,” 2019.

Submitted to Taylor and Francis Journal of Imaging Science (currently under review)

Nicholas Wells, Chung. W. See, “Polynomial Edge Reconstruction Sensitivity, Sub-Pixel Sobel Gradient Kernel Analysis,” 2019.