Designing Hypothesis Tests for Digital Image Matching

Gregory Sean Cox

Thesis Presented for the Degree of
DOCTOR OF PHILOSOPHY
in the Department of Electrical Engineering
UNIVERSITY OF CAPE TOWN
August 2000

Designing Hypothesis Tests for Digital Image Matching

Gregory Sean Cox

August 2000

Abstract Image matching in its simplest form is a two class decision problem. Based on the evidence in two sensed images, a matching procedure must decide whether they represent two views of the same scene, or views of two different scenes. Previous solutions to this problem were either based on an intuitive notion of image similarity, or were modelled on solutions to the superficially similar problem of target detection in images. This research, in contrast, uses a decision theoretic formulation of the problem, with the image pair as unit of observation and probability of error in the match/mismatch decision as performance criterion. A stochastic model is proposed for the image pair, and the optimal test of match and mismatch hypotheses for samples of this random process is derived. The test is written conveniently in terms of a statistic of the two images and a scalar decision threshold. The analytical advantages of a solution derived from first principles are illustrated with the derivation of hypothesis conditional probability distributions, optimal decision thresholds, and expressions for the probability of error in the decision.

Two practical aspects of the optimal test are investigated. First, the error-rate performance is evaluated over a wide range of conditions using Monte Carlo methods in conjunction with a procedure for generating random image pairs. Even with deviations from the assumed model, the optimal test exhibits a significant performance advantage over tests based on traditional similarity measures. The optimal test performs poorly when the scene is occluded, however, and nonparametric similarity measures from the literature prove to be more effective in this case.

Second, the computational complexity of the optimal test is addressed. It is established that corresponding pairs of principal components from the two individual images provide an optimal compaction of the inter-image correlation structure under the proposed model. This knowledge is used to reduce the dimensionality of the problem, thereby reducing the computation required by the test statistic. Separable models provide alternative methods for reducing computation, and partitioned images are used to calculate the test statistic of a large image pair by summing the statistics for individual non-overlapping subimages.

The hypothesis testing approach is then extended to the sub-problem of image registration that is referred to as block matching. This formulation of the problem incorporates a rejection hypothesis and a prior for the position of correct register, where the latter is useful if there is *a priori* knowledge about the mechanism that originally put the images out of register. Techniques are developed for performing block matching efficiently for both the optimal test and for tests based on standard similarity measures. Experiments using purely synthetic random images and real images with artificial noise illustrate once again the error-rate performance benefits of an optimal test derived from first principles.

Acknowledgments

My experience of postgraduate study was significantly enhanced by fellow students and academic staff at the University of Cape Town, and by colleagues at the DebTech research laboratory. In particular, I thank Brendt Wohlberg and Fred Nicolls for their constructive input, and Professor Gerhard de Jager for his guidance.

I thank my wife Angélique, who made this endeavour possible through her support and understanding, and our parents for their unfailing encouragement.

The financial assistance of the National Research Foundation towards this research is hereby acknowledged. Opinions expressed in this work, or any conclusions arrived at, are those of the author and are not necessarily to be attributed to the National Research Foundation.

Contents

Glossary

ARE	Asymptotic Relative Efficiency
CBC	Coincident Bit Counting
CT	Computed Tomography
CCD	Charge Coupled Device
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DSC	Deterministic Sign Change
FFT	Fast Fourier Transform
GLRT	Generalized Likelihood Ratio Test
KLT	Karhunen Loève Transform
LRT	Likelihood Ratio Test
MAP	Maximum *A Posteriori*
MGF	Moment Generating Function
MI	Mutual Information
ML	Maximum Likelihood
MRF	Markov Random Field
MVN	Multivariate Normal
OSR	Occlusion-to-Signal Ratio
PCA	Principal Component Analysis
pdf	Probability Density Function
RMS	Root Mean Square
ROC	Receiver Operating Characteristic
SNR	Signal-to-Noise Ratio
SSC	Stochastic Sign Change
SSDA	Sequential Similarity Detection Algorithm

Notation

Scalars, Vectors and Matrices

a, b	scalars
$\mathbf{a}, \mathbf{b}$	vectors
$\mathbf{A}, \mathbf{B}$	matrices
$\mathbf{A}^T$	transpose of matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	inverse of matrix $\mathbf{A}$
$\det(\mathbf{A})$	determinant of matrix $\mathbf{A}$
$\mathbf{A}^*$	complex conjugate of the elements in $\mathbf{A}$
$\mathbf{I}$	identity matrix, where dimensions are determined by the context

Image Representation

$a(i, j)$	function image representation with discrete spatial coordinates
$a(x, y)$	function image representation with continuous spatial coordinates
$\mathbf{a}$	vector image representation
n_a	width or height of square image $\mathbf{a}$
$p_{\mathbf{a}}(\mathbf{a})$	pdf of random image $\mathbf{a}$
$p_{\mathbf{a}}(\mathbf{a}\|B)$	pdf of random image $\mathbf{a}$ conditioned on event B
$\mathbf{m}_{\mathbf{a}}$	mean vector for random image $\mathbf{a}$
σ_a^2	variance of random image $\mathbf{a}$
$\mathbf{K}_{\mathbf{a}}$	covariance matrix of random image $\mathbf{a}$
$\mathbf{R}_{\mathbf{a}}$	correlation coefficient matrix of random image $\mathbf{a}$ ($\mathbf{K}_{\mathbf{a}} = \sigma_a^2 \mathbf{R}_{\mathbf{a}}$)
$p_{\mathbf{a},\mathbf{b}}(\mathbf{a}, \mathbf{b})$	joint pdf of random images $\mathbf{a}$ and $\mathbf{b}$
$\mathbf{K}_{\mathbf{ab}}$	cross-covariance matrix of random images $\mathbf{a}$ and $\mathbf{b}$
$\mathbf{R}_{\mathbf{ab}}$	cross-correlation coefficient matrix of random images $\mathbf{a}$ and $\mathbf{b}$
ρ_{ab}	correlation coefficient between random images $\mathbf{a}$ and $\mathbf{b}$

Hypothesis Testing

H_1	match hypothesis
H_0	mismatch hypothesis
$P(A)$	probability of event A
P_I	probability of a type I detection error
P_{II}	probability of a type II detection error
$l(\mathbf{a}, \mathbf{b})$	match likelihood ratio for images $\mathbf{a}$ and $\mathbf{b}$
$l_G(\mathbf{a})$	generalized match likelihood ratio for images $\mathbf{a}$ and $\mathbf{b}$
$L(\mathbf{a})$	match log-likelihood ratio for images $\mathbf{a}$ and $\mathbf{b}$
P_1	*a priori* probability of match
P_0	*a priori* probability of mismatch

Operations

$\mathcal{X}(\mathbf{s}, \mathbf{v})$	correlation of search area $\mathbf{s}$ with image block $\mathbf{v}$
$\mathcal{S}_n(\mathbf{s})$	operation on image $\mathbf{s}$ that calculates local sums in overlapping $n \times n$ windows

Chapter 1

Introduction

Rapidly advancing fields of research frequently leave behind areas that lack rigorous theoretical attention. As researchers concentrate on topical and exciting new areas, the more mundane questions are neglected until they restrict further progress. One such question in the fields of computer vision and image processing asks what the optimal method is for determining whether two noisy images represent the same scene. Although the literature presents a range of solutions that vary greatly in approach and sophistication, a rigorous formulation of the problem in its simplest form is absent. This problem has new importance as applications such as low-dose medical imaging begin to test the limits of algorithm performance.

The research presented in this dissertation addresses the problem of direct image matching. Here a computer algorithm must decide whether the scenes represented by two distinct images have in common a component of interest. The adjective "direct" is used to indicate that no deterministic knowledge about the image content can be assumed and therefore a matching algorithm must process the pixel values in the images and not a higher level model-based representation of the scene.

1.1 Beyond Similarity and Detection

Methods in the literature often address the image matching problem by assessing the degree of similarity (or dissimilarity) between images. This seems to be a natural approach to the problem for reasons that perhaps go deeper than the obvious relationship between similarity and match. As Tversky puts it, similarity "serves as an organizing principle by which individuals classify objects, form concepts, and make generalizations. Indeed, the concept of similarity is ubiquitous in psychological theory" [1]. The emphasis on measures of similarity

in image matching may be the result of a natural human predilection for this approach.

Viewed independently of the intuitively appealing concept of similarity, image matching is nothing more than a decision problem where an observation of the physical world must be classified as belonging to one of two categories. The observation is a pair of digitized images and the categories are "match" and "mismatch". In principle this problem is distinct from the one of *target detection* in images, where the observation is a sensed image and the categories are match and mismatch between this image and an idealized template of the target in question. The superficial similarity between the problems of matching and detection, however, is potentially misleading. It leads to formulations of the image matching problem where the two images are treated as independent observations, whereas it is precisely the relationship between them — their interdependence — that is the determinant of match.

The approach taken in this work is to suspend intuitive notions of similarity and parallels with the classical detection problem, and to apply powerful results from decision theory to recognize match or mismatch in image pairs. It is argued that a measure of similarity should be a consequence of, rather than a starting point for, the design of an image matching algorithm, and that the unit of observation should be the image pair. It is shown that this perspective leads naturally to a useful formulation of the problem in that it supports more rigour than previous approaches and delivers solutions with superior performance.

1.2 A Scientific Approach

The level of scientific rigour in computer vision and related fields has been a topic of debate for some time. In the early 1990s a fascinating dialogue between several luminaries in computer vision revealed that opinions differed on the level of scientific rigour evident and the relative importance of theory and practice, but agreed on the importance of a scientific approach and of proper experimentation [2, 3, 4, 5, 6]. More recent meetings of computer vision academics and professionals still debate the issue of whether the field is a legitimate scientific discipline[1]. A scientific approach to the problem of image matching requires that the phenomenon of "match" is defined; that a model for the problem is developed; that a solution is derived on the basis of this model and accepted performance criteria; and that this solution is tested through rigorous experimentation. These elements are now explored further.

[1]For example, during panel discussions at the IEEE Computer Society Workshop: Vision Algorithms - Theory and Practice, which was held in conjunction with the Seventh International Conference on Computer Vision from 21 to 22 September 1999 on Corfu, Greece.

Definition For the purpose of this research, match is defined in terms of the relationship between the two scenes that the image pair captures. The nature of the relationship will depend on the particular application. For instance, stereo correspondence algorithms could consider matching scenes to be identical (barring sensor noise) and non-matching scenes to be statistically independent.

Modelling The assumption that deterministic information is not available leaves the option of modelling the image pair probabilistically. There is considerable precedent for the use of statistical image models in the areas of image compression, image restoration and object detection. Generally speaking, if the performance of an algorithm is measured over a large ensemble of images for which a statistical model captures the essence of the image data, then this sort of model is appropriate.

Derivation Given a definition of match and a statistical model for the image pair, the derivation of the optimal matching procedure is an optimization problem that is guided by performance criteria for the resulting algorithm. The fields of signal detection and pattern recognition have developed powerful tools for tackling this sort of problem.

Experimentation If an ensemble of image pairs that conforms to the model can be synthesized, then experimentation under ideal conditions is only limited by computational resources. This sort of numerical experiment using random numbers belongs to the class of Monte Carlo methods. In principle, the same results could be obtained analytically, but the mathematics involved is often intractable. Since the model is rarely a perfect representation of real images, further qualification of the algorithm involves experiments with real image data that is representative of the problem at hand.

1.3 Objectives of the Research

A decision theoretic formulation of the image matching problem allows the derivation of algorithms that are optimal with respect to the models chosen and the performance criteria applied. The primary hypothesis of this research is that the optimal approach based on parametric statistical models will outperform current techniques, which are based on ad-hoc notions of image similarity and target detection algorithms adapted for matching purposes. In order to test this hypothesis, the following objectives are set:

1. To review the current approaches to the problem of direct image matching.
2. To develop a statistical model for image pairs and a procedure for synthesizing an ensemble of images that conform to this model.
3. To derive the matching procedure that is optimal with respect to the aforementioned model in that it minimizes the probability of error when making the match/mismatch decision.
4. To compare the optimal procedure to existing sub-optimal procedures under the ideal conditions of the proposed model using Monte Carlo simulation experiments.
5. To apply this knowledge to a common image processing problem.
6. To test the effectiveness of the optimal algorithm using experiments with real images.
7. To draw conclusions on the basis of the results and to propose a programme of further research.

1.4 Outline of the Dissertation

A brief outline of the dissertation content is now given as an aid to the reader.

Chapter 2 reviews the literature on direct image matching. Early work in this area evolved from correlation-type filters used in one-dimensional signal processing. The matched filter and phase correlation are notable examples. Much of the work on direct image matching is based on measures of image similarity, which generally fall into the categories of correlation-based, difference-based, nonparametric and histogram-based measures. Some work has been done on quantifying the performance of direct matching methods, but these results are either experimental and specific to a set of test images, or analytical and restricted to the simplest measures. Although the different approaches are each based on a sound rationale, there is no evidence of measures that are derived from first principles using an image model.

Chapter 3 formulates the image matching problem in a decision theoretic framework. Tools available in the fields of signal detection theory and pattern recognition are investigated and the former are found to be appropriate for this problem. The high dimensionality of the observations, the assumed availability of an analytical image model and the impracticality of

unsupervised learning in this case support this decision. In order to illustrate the mechanics of formulating matching in this way, the optimal test is derived for the trivial problem of matching scalars. This has no practical significance, but shows some interesting characteristics of the solution that would be obfuscated in higher dimensions.

Chapter 4 derives a model for the image pair that includes a parameter to control the degree of match between the images. Multivariate normal models are almost unique in their analytical tractability, but do not appear to capture the essence of typical image ensembles. Simple techniques described in the literature, however, can transform typical images so that they do resemble samples of a multivariate normal random process. It is shown that reasonable assumptions about the image pair lead to a simple joint image model where the degree of match is controlled by the linear correlation coefficient between the ideal (noise-free) components of the images. Finally, the chapter derives an efficient method for synthesizing image pairs that are samples of the random process described by this model.

Chapter 5 derives the optimal likelihood ratio test for image matching, given the image-pair model of the previous chapter and probability of error as the performance criterion. It is shown that the test is conveniently written in terms of the normalized principal components of the images, suggesting a two stage procedure for calculation of the test statistic: a whitening transformation followed by calculation of the simplified likelihood ratio statistic for images with identically distributed, spatially independent pixels. The probability density function of the test statistic under match and mismatch hypotheses is derived, allowing optimal decision thresholds and the probability of error to be calculated analytically.

Chapter 6 documents Monte Carlo experiments that compare the error-rate performance of the optimal test and methods from the literature. These experiments reveal that under the assumed model the optimal test is far superior, suggesting that there is scope for significant improvement on the standard approaches to image matching. Experiments also investigate the error-rate when the image data deviates from the assumed model, and where there is occlusion. The optimal test is fairly robust under a wide range of conditions, but measures that are designed specifically to be robust under occlusion exhibit better performance than the optimal test when this type of distortion is present.

Chapter 7 addresses the high computational demands made by the likelihood ratio statistic. It is established that under the assumed model the principal components of the images correspond to the canonical variables of the image pair, making them an optimal compaction of the inter-image correlation structure. An efficient (but lossy) strategy for reducing the computational requirement only considers the subset of the canonical variables that have the highest correlation coefficients. Computation can also be reduced by assuming that row and column models are separable. If the image models can be approximated by separable Markov processes for the rows and columns with one-step correlation approaching unity, then the discrete cosine transform can approximate the whitening transforms for further efficiency. Finally, large images can be processed efficiently by using a block partitioned approximation of the test statistic.

Chapter 8 formulates the sub-problem of image registration known as block matching in a hypothesis testing framework. It is shown that the optimal test derived previously for matching is part of the resulting algorithm. Block matching is a highly computationally intensive operation and methods for performing the search for matching blocks efficiently are introduced. A method is also introduced for screening blocks that are unsuitable for matching, thereby reducing the number of control points where block matching has to be performed, and ensuring good registration performance from the control points that are selected. The error-rate performance of different block matching algorithms is then analyzed using experiments that range from Monte Carlo simulations with purely synthetic data to experiments that use real images with synthetic noise. The performance advantage of the optimal approach based on parametric statistical models is confirmed in the results of these experiments.

Chapter 9 concludes the dissertation. The key insights are summarized and the important results are highlighted in order to consolidate the contribution made by the previous chapters. This work has only scratched the surface of the image matching problem and directions for future research are suggested. The potential for using a similar approach to formulate other problems, such as target detection using a reference image, is also pointed out. Final remarks conclude the presentation.

Chapter 2

A Review of Direct Image Matching

The task of establishing whether two images match each other or not is ubiquitous in the field of image processing. Image alignment, indexing in image databases, the detection of changes over time: these are all tasks where a decision must be made as to whether two images represent the same scene. Given that equivalence in the scene itself is of importance, an obvious strategy would be to reconstruct and compare the scene that each image represents. If scene features can be identified in the images and used to establish the correspondence then a complete reconstruction may not be required. However, the sort of *a priori* information about scene content that is a prerequisite for this approach is often not available. In this situation the correspondence has to be determined directly from the image pixel values, or simple statistics of these values, by *direct* measures of image similarity. These measures make few assumptions about the scene content, using only vague and uncertain *a priori* information about the scene, sometimes embodied in a statistical model.

Advances in the measurement of image similarity have been reported in a diverse range of fields. The medical imaging, remote sensing, computer vision and pattern recognition literature are all represented in this review of image similarity measures. The review starts with early work in image matching, where image processing researchers were influenced by a more established body of signal processing research in the radar and communications fields. Early applications of image registration also only required translational alignment. This fact, together with the existing signal processing paradigm, lead to the classical filtering approach for image matching, which is reviewed briefly in Section 2.1. Section 2.2 then summarizes

the image similarity measures that have been reported in the literature. Evaluation and comparison is essential given the variety of both applications and available measures. Section 2.3 collects work in this area. Observations on the current state of direct image matching are made in Section 2.4.

Rather than reproduce each author's individual notational preferences, a consistent notation is used throughout the review to facilitate the comparison of different approaches. An image with m rows and n columns is either denoted as the mn-vector $\mathbf{u}$, where the elements are in row-column order, or the spatially discrete signal $u(j,k)$, where $j \in \{1,2,\ldots,m\}$ and $k \in \{1,2,\ldots,n\}$. Given image vector $\mathbf{u}$, the notation u_i refers to the i-th element of this vector, whereas $u(j,k)$ refers to the element corresponding to the pixel value at column j and row k in the image. Here u_i is equivalent to $u(j,k)$, where $i = j + n \cdot k$. Sometimes it will be convenient to think of the image as a continuous function in two dimensions, in which case it will be denoted as $u(x,y)$. By convention, coordinates in the spatial domain will be denoted (j,k) (or (x,y) in the continuous case) and coordinates in the spatial-frequency domain will be denoted (ω,ν).

2.1 Image Correlation Filters

No review of image matching would be complete without covering the classical filtering approach to aligning images. These techniques are typically only applicable to translational registration algorithms, but they do have general importance as the origin of many similarity measures with broader application.

2.1.1 Matched Filters and Cross Correlation

The convolution filter kernel that maximizes signal-to-noise ratio (SNR) where a finite length time-domain signal is embedded in additive white noise is the signal itself reflected about the amplitude axis [7]. The same result holds for a digital, spatially quantized two dimensional signal that is embedded in white noise [8]. Consider an $m_a \times n_a$ sensed image $a(j,k)$ and a known $m_v \times n_v$ template $v(j,k)$, where $m_v < m_a$ and $n_v < n_a$, and the pixel intensity values are real. The filter operation that maximizes SNR if a consists of instances of v embedded in white noise can be written as

$$f(j,k) = a(j,k) * \bar{v}(j,k), \tag{2.1}$$

where $*$ denotes linear convolution and $\bar{v}$ is equal to v rotated by π radians in the image plane. This operation is also referred to as the *correlation* of a and v and has been used to align two sensed images. In this context v is normally a subimage extracted from a second sensed image.

The matched filter of equation (2.1) can also be written as the *correlation function*

$$f(j,k) = \sum_s \sum_t a(j+s, k+t) \cdot v(s,t), \tag{2.2}$$

where (s,t) are coordinates in v with the origin at its center and f is defined for coordinates (j,k) where a and v overlap completely. This filter was derived for the detection of a (noise-free) template in white noise, however, and is not as useful a measure for registering two noisy images. For instance, the value of $f(j,k)$ is dependent on the signal strength in a and v. A more practical approach in this respect is the *normalized correlation function*

$$f_n(j,k) = \frac{\sum_s \sum_t a(j+s, k+t) \cdot v(s,t)}{\left[\sum_s \sum_t a(j+s, k+t)^2\right]^{1/2} \left[\sum_s \sum_t v(s,t)^2\right]^{1/2}}, \tag{2.3}$$

which has been used for change detection [9] and template matching [10].

Anuta pointed out that it is more efficient to calculate image correlation using Fast Fourier Transform (FFT) operations [11]. Since circular convolution in the spatial domain is equivalent to multiplication in the spatial-frequency domain [12],

$$f(j,k) = \mathcal{F}^{-1}\left[A(\omega,\nu) V^*(\omega,\nu)\right], \tag{2.4}$$

where $\mathcal{F}[\cdot]$ is the discrete Fourier transform (DFT), A and V are the DFTs of a and v respectively, and V^* denotes the complex conjugate of V. Note that if f is considered to be the same size as a then there are so-called *edge effects* in the result of both (2.1) and (2.4). For the former, the edge is undefined where v is translated to positions where there is only partial overlap with a, and for the latter a similarly sized edge area consists of aliased information because (2.4) is equivalent to *circular* convolution.

Given the importance of cross-correlation in communications and radar systems before image processing even existed, it is not surprising that this matching technique has received much theoretical and experimental attention. Several authors have investigated the statistical properties of cross-correlation when used to detect a known signal in additive noise. The probability density function (pdf) of the cross-correlator output for noise and ideal signal plus

noise is derived by Green [13]. The pdf for the case where both the received and reference signals are sampled and are corrupted by noise is derived by Roe and White [14]. Andrews extends this analysis to allow non-zero correlation between the input waveforms [15]. These results are of limited use in the analysis of the more general matching problem, because in this case the signal is not deterministically known.

The shape of the correlation function peak has also received attention. Dvornychenko establishes bounds for the normalized correlation of two signals that differ only by a relative shift under noise-free conditions [16].

2.1.2 Phase Correlation

A peak in f indicates the presence of v at the corresponding position in a. This peak is broad in practice, leading to inaccurate localization. Kuglin and Hines propose exploiting the Fourier shift property [17, p. 45] by using *phase correlation*, which produces a narrow peak and exhibits better performance in narrowband noise [18]. The phase correlation of a and v is given by the inverse Fourier transform of the phase difference between two images:

$$f_p(j,k) = \mathcal{F}^{-1}\left[\frac{A(\omega,\nu)\,V(\omega,\nu)^*}{|A(\omega,\nu)\,V(\omega,\nu)^*|}\right]. \tag{2.5}$$

This technique is not restricted to the case where v is contained within a, and can be used to find large relative displacements between two images of similar size. It is invariant to scaling or offset in image intensities and tolerates spurious low-frequency background intensity variations (e.g. illumination differences) very well, because these can be regarded as narrowband noise.

Variations on standard phase correlation have been proposed. Alliney and Morandi sacrifice performance for computational efficiency by using one dimensional phase correlation of projections [19]. De Castro and Morandi extend the technique by using an iterative search to cover rotational as well as translational differences between the images [20]. Pla and Bober show how linear deformation parameters that are separable in j and k can be estimated in a phase correlation framework [21].

2.1.3 Whitening Filters

Kuglin and Hines note that rewriting (2.5) as

$$f_p(j,k) = \mathcal{F}^{-1}\left[\left(\frac{A(\omega,\nu)}{|A(\omega,\nu)|}\right)\left(\frac{V(\omega,\nu)^*}{|V(\omega,\nu)^*|}\right)\right]$$

suggests an interesting interpretation of phase correlation: A and V are normalized by their magnitudes, which "effectively 'whitens' each image with respect to itself", before the standard correlation operation is carried out [18]. Pratt uses a similar idea, but proposes explicit whitening filters for a and v as preprocessing for the normalized correlation function [22]. Given two images, a and v, Pratt's statistical correlation measure is defined as

$$f_s(j,k) = \frac{\sum_s \sum_t \acute{a}(j+s,k+t) \cdot \acute{v}(s,t)}{\left[\sum_s \sum_t \acute{a}(j+s,k+t)^2\right]^{1/2} \left[\sum_s \sum_t \acute{v}(s,t)^2\right]^{1/2}}, \tag{2.6}$$

where $\acute{a}(j,k) = a(j,k) * D_a(j,k)$ and $\acute{v}(j,k) = v(j,k) * D_v(j,k)$. The spatial filter functions, $D_a(j,k)$ and $D_v(j,k)$, are chosen to decorrelate the image pixels (or "whiten" the images). The whitening filters are equivalent to gradient operators if the rows and columns can be modelled as separable Markov processes [22].

2.2 Image Similarity Measures

This section summarizes the literature on image similarity measures. Attention is restricted to the task of image matching — measures of similarity that are used to evaluate image fidelity after lossy image compression, for example, are not covered [23]. The specific definition of "match" for $\mathbf{u}$ and $\mathbf{v}$ will depend on the particular application of the measure. The definition of the term "similarity measure" is not restricted to measures that increase with the underlying degree of match between the two images. Measures that decrease with increasing match, often referred to in the literature as "distance measures", are also included in the definition.

2.2.1 Classical Image Similarity

Correlation- and difference-based measures are classical in the sense that they were used by the earliest image matching algorithms, but even recent literature compares new measures to examples from one or both of these categories [24, 25, 26].

Correlation-Based Measures The filters of Section 2.1 can be thought of as a windowed measurement of the local correlation between two images. The corresponding measures of similarity between two images of the same size are made explicit here. The cross correlation

of two $m \times n$ sensed images, $\mathbf{u}$ and $\mathbf{v}$, is their inner product

$$R(\mathbf{u}, \mathbf{v}) = \sum_i u_i v_i.$$

The normalized cross correlation is written as

$$\acute{R}(\mathbf{u}, \mathbf{v}) = \frac{\sum_i u_i v_i}{\sqrt{\sum_i u_i^2 \sum_i v_i^2}},$$

which is invariant to scaling of the pixel intensity values. A further modification aims for both scale and offset invariance by using the normalized cross-correlation of the zero-mean images, giving the *correlation coefficient* [27, p. 584]

$$r(\mathbf{u}, \mathbf{v}) = \frac{\sum_i (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum_i (u_i - \bar{u})^2 \sum_i (v_i - \bar{v})^2}} \tag{2.7}$$

where

$$\bar{u} = \frac{1}{N} \sum_i u_i \quad \text{and} \quad \bar{v} = \frac{1}{N} \sum_i v_i$$

and $N = mn$ is the number of pixels in each image. This measure has a statistical interpretation. If corresponding pixel pairs, (u_i, v_i), can be regarded as samples of two random processes, then r is their sample correlation coefficient [28, p. 328]. If these random processes have a bivariate normal distribution with covariance matrix,

$$K_{uv} = \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix},$$

then r is the maximum likelihood estimate of ρ [29, p. 144]. Like the statistical correlation coefficient, r is in the range $[-1, 1]$, where $r = 1$ represents positive correlation (identical images, except for scale and offset), $r = -1$ represents negative correlation (identical images with reverse "polarity") and $r = 0$ represents totally uncorrelated signals (images that don't match). If some of the pixel values are outliers due to noise, then r can be a very poor estimate of ρ. A potential solution can be found in the field of robust statistics, which addresses the problem of estimating statistical parameters using data that is corrupted by

outliers [30]. Brunelli and Messelodi compare several robust[1] estimates of ρ that outperform r in the presence of noise [31].

Other measures of linear correlation are also used to measure image similarity. Radcliffe, Rajapakshe and Shalev [32] calculate the χ^2 statistic, which is the basis of a classical test of independence between two data sets [28, p. 257], for a random subset of the corresponding pixel-pairs.

Difference-Based Measures The most common image similarity measures are based on differences between the intensity values of corresponding pixels. The simplest of these are the sum of absolute differences

$$d_1(\mathbf{u}, \mathbf{v}) = \sum_i |u_i - v_i| ,$$

which is popular for template matching [33, 34], and the sum of squared differences

$$d_2(\mathbf{u}, \mathbf{v}) = \sum_i (u_i - v_i)^2 .$$

The subscripts in the notation of d_1 and d_2 make reference to the fact that these measures are related to the metrics induced by L_1 and L_2 norms, respectively, on the space of images.

Svedlow, McGillem and Anuta claim that these measures have an advantage over cross-correlation in the case of additive noise in that the values of the latter have no absolute scale, whereas the values of d_1 and d_2 will depend on the statistical properties of the noise because the noise-free components are cancelled in the difference $(u_i - v_i)$ [35]. A statistical noise model can then be used to specify a confidence interval for d_1 or d_2 in the case of a match.

The sum of squared differences is related to the correlation-based measures of the previous section — Rosenfeld uses d_2 as the starting point for a derivation of cross correlation as a measure of match [36, p. 37]. Also, if $\acute{\mathbf{u}}$ and $\acute{\mathbf{v}}$ are zero-average images that have been normalized to have a sample standard deviation of one, then d_2 is related to the sample correlation coefficient by

$$d_2(\acute{\mathbf{u}}, \acute{\mathbf{v}}) = 2N \left[1 - r(\mathbf{u}, \mathbf{v})\right] . \tag{2.8}$$

[1] "Robust" here has a specific statistical interpretation, and not the general interpretation often used in image processing for a procedure that maintains performance under a wide range of conditions.

In this light r can be viewed as a variation on d_2 that compensates for different offsets and scaling in the pixel intensities of $\mathbf{u}$ and $\mathbf{v}$. Another measure related to d_2 is the *variance of the differences*,

$$d_v(\mathbf{u},\mathbf{v}) = \frac{1}{N}\sum_i \left[(u_i - v_i) - \frac{1}{N}\sum_i (u_i - v_i)\right]^2,$$

which was proposed by Cox and de Jager [37] for its insensitivity to pixel intensity offset differences between $\mathbf{u}$ and $\mathbf{v}$. It is equivalent to d_2 for two images where the sample mean has been subtracted beforehand.

Although d_1 is less sensitive to noise than d_2, neither tolerates pixel value outliers very well [31]. Speckle noise, or even a single dead pixel caused by a malfunction of the image formation device (common in charge coupled device (CCD) arrays), can cause otherwise matching images to seem dissimilar according to these measures. With appropriate scaling, d_1 can be thought of as an estimate of the mean absolute pixel difference. Considering that the median is a robust estimator of the mean [38], an obvious robust alternative to d_1 is the median absolute difference

$$d_m(\mathbf{u},\mathbf{v}) = \text{med}\left[|u_i - v_i|\right]. \tag{2.9}$$

Boninsegna and Rossi propose the *mixture distance* [39], which combines d_2 and d_m as follows:

$$\begin{aligned} d_k(\mathbf{u},\mathbf{v}) &= \sum_i e_k(u_i,v_i)^2 \\ e_k(u_i,v_i) &= \begin{cases} (u_i - v_i) & \text{if } |u_i - v_i| < k \\ m & \text{otherwise} \end{cases}, \end{aligned}$$

where m is the median of the differences and the parameter $0 \leq k < \infty$ controls the mixture.

Sequential Similarity Detection Barnea and Silverman propose the Sequential Similarity Detection Algorithms (SSDAs) for speeding up image registration [40]. An SSDA includes several elements that improve computational efficiency, but a simple example illustrates the SSDA notion of image similarity between two images. The error between two corresponding pixels at position (j,k) is evaluated with the absolute pixel difference

$$e(j,k) = |u(j,k) - v(j,k)|.$$

A random, non-repeating series of pixel positions, (j_n, k_n), is generated. The overall match error is then calculated by accumulating the individual errors in the order of this series. The number of pixels N, for which

$$\sum_{k=1}^{N} e(j,k) \geq T,$$

is used as the measure of similarity between the images, where T is a predetermined threshold. Similarity, then, is based on the rationale that matching images will require more pixel-pairs to exceed the threshold.

2.2.2 Nonparametric Similarity

A more recent trend in image similarity measurement is the use of *nonparametric* measures. These should not be confused with the *robust* measures that were mentioned previously[2]. Robust statistics is concerned with the effect of outliers on a procedure that was designed with a certain model in mind. Examples of robust statistics in image similarity measurement are Brunelli and Messelodi's correlation coefficient estimators [31] and the mixture distance proposed by Boninsegna and Rossi [39].

Nonparametric or distribution-free procedures, on the other hand, tolerate deviations from a classical model or the lack of model knowledge by making no (or very few) assumptions about the underlying statistical nature of the process that generates the input data. Where robust procedures use parameter estimators that are insensitive to outliers, nonparametric procedures ignore the parameters completely. Such procedures have largely had their origins in the field of detection theory, where optimal approaches were seen to have limitations when the model assumptions are violated [41].

Sign Change Measures Venot, Lebruchec and Roucayrol propose similarity measures called *sign change criteria* for situations where there is significant obscuration that can impede matching [42]. This approach is used predominantly for medical image registration [43, 44, 45].

Consider two $m \times n$ images $\mathbf{u}$ and $\mathbf{v}$ which differ only by additive noise. Their difference image, $\mathbf{e} = \mathbf{u} - \mathbf{v}$, will exhibit many sign changes between pixel pairs that are horizontally or vertically adjacent in the image plane. Images that have differences significantly greater

[2]Much of the literature treats these as equivalent concepts, but this discussion follows Huber [38] in the distinction made between robust and non-parametric procedures.

than the mean of the noise will not produce many sign changes between adjacent pixels. This motivates the use of sign change for measuring similarity between images. The nonparametric advantage of this approach lies in the fact that the values of intensity differences are not taken into account directly. Where an isolated outlier in either image would distort the value of a parametric measure, it will not have any significant effect on the number of sign changes in the difference image.

Formally then, define the function

$$\text{sgn}(x) = \begin{cases} 0 & x \geq 0 \\ 1 & x < 0 \end{cases}$$

and the stochastic sign change (SSC) measure of similarity is the number of sign changes found when scanning the image row-by-row or column-by-column. For the row-by-row case the SSC criterion is given by

$$s_s(\mathbf{u}, \mathbf{v}) = \sum_{j=1}^{n-1} \sum_{k=1}^{m} \text{sgn}\left[e\left(j,k\right) \cdot e\left(j+1,k\right)\right],$$

which increases with increasing similarity between $\mathbf{u}$ and $\mathbf{v}$.

The SSC relies on the presence of noise with a pdf that has zero median to supply sign changes in the difference image [46]. If the noise is low compared to the precision of digitization then this requirement is not satisfied. Venot, Lebruchec and Roucayrol deal with this case by adding a periodic pattern to one of the images as follows

$$\acute{v}\left(j,k\right) = \begin{cases} v\left(j,k\right) + q & \text{if } j+k \text{ is even} \\ v\left(j,k\right) - q & \text{if } j+k \text{ is odd} \end{cases}$$

and then calculating the SSC as before. The result, $s_d(\mathbf{u}, \mathbf{v}) = s_s(\mathbf{u}, \acute{\mathbf{v}})$, is called the deterministic sign change (DSC) criterion. In image registration experiments that compare it to the correlation function, correlation coefficient and sum of absolute differences, the SSC (or DSC) criterion purportedly provides a narrower match peak, better registration accuracy, and is more robust in the presence of obscuration [42].

The statistical properties of sign-change sequences can be used to specify the match threshold according to a confidence interval [42]. In the nonparametric statistical theory of run tests, it is known that if plus and minus signs are equiprobable, then the pdf of the number of sign changes in a sequence is normal. Hence, for images $\mathbf{u}$ and $\mathbf{v}$ that are perfectly matched except

for additive noise, the stochastic sign change has a 95% confidence interval of

$$\frac{N}{2} - 1 \pm 1.96\frac{\sqrt{N}}{2},$$

where N is the number of pixels in the sequence. The only restriction on the additive noise is that its pdf is shared by both images and has zero median [46].

Coincident Bit Counting Chiang and Sullivan propose a measure of similarity based on the number of coincident bits in the binary representation of corresponding pixels in two images [47]. Define the function, bits(x), as the number of bits set in the binary representation of the integer x. The coincident bit counting (CBC) measure of similarity can be formulated as

$$c(\mathbf{u}, \mathbf{v}) = \sum_i \text{bits}(u_i \bar{\oplus} v_i),$$

where $\bar{\oplus}$ is an exclusive NOR operator. Like the sign change criteria, the CBC measure is not proportionately affected by large pixel value differences and is therefore robust to pixel outliers.

According to Chiang and Sullivan the CBC measure can be made less sensitive to noise by excluding the lower order bits from its calculation [47]. The number of bits used can be dynamically adjusted according to the noise in the images. Chiang and Sullivan also suggest that in the case of template matching or image registration, a steeper peak in the match surface will be obtained at the position of correct match if the higher order bits are also omitted from the calculation of $c(\mathbf{u}, \mathbf{v})$, since these will probably be locally uniform [47].

Ordinal Measures An important class of nonparametric statistical tests is based on the rank, or ordering, of sample values. Corrupt data or outliers only affect a statistic that is based on rank if the incorrect data changes the relative ordering of the samples. Motivated by this observation, Bhat and Nayar propose a similarity measure where pixel intensity is viewed as an "ordinal variable" — that is, a variable drawn from a discrete ordered set [25]. Similarity is then based on rank permutations of the intensities rather than on absolute intensity values.

Given the pixel intensity pairs (u_i, v_i), π_u^i is defined as the rank of u_i among the pixels in image $\mathbf{u}$, and π_v^i is defined as the rank of v_i among the pixels in image $\mathbf{v}$. A composition

permutation s is defined as

$$s^i = \pi_v^k, \;\; k = \left(\pi_u^{-1}\right)^i$$

where π_u^{-1} is the inverse permutation of π_u, defined by

$$\pi_u^i = j \Rightarrow \left(\pi_u^{-1}\right)^j = i.$$

The permutation s is an ordering of $\mathbf{v}$ with respect to $\mathbf{u}$ and when $\mathbf{u}$ and $\mathbf{v}$ match, s is the identity permutation: $(1, 2, \ldots, N)$. A distance between π_u and π_v is then defined according to a distance measure between s and the identity permutation — the deviation

$$d_m^i = i - \sum_{j=1}^{i} J\left(s^j \leq i\right),$$

where $J(B)$ is the indicator function

$$J(B) = \begin{cases} 1 & B \text{ is true} \\ 0 & B \text{ is false.} \end{cases}$$

The similarity of $\mathbf{u}$ and $\mathbf{v}$ is then defined as

$$\kappa(\mathbf{u}, \mathbf{v}) = 1 - \frac{2 \max_{i \in \{1,2,\ldots,n\}} d_m^i}{\left\lfloor \frac{N}{2} \right\rfloor},$$

which has the range $[-1, 1]$. A perfect match is represented by $\kappa = 1$, and $\kappa = -1$ represents perfect negative correlation.

The nonparametric measure κ has some desirable properties. First, it is invariant to linear scaling and offsets in intensity. Second, it is not affected by monotonically increasing functions on $\mathbf{u}$ and $\mathbf{v}$ (i.e. $\kappa\left(f(\mathbf{u}), h(\mathbf{v})\right) = \kappa(\mathbf{u}, \mathbf{v})$, if $f(\cdot)$ and $h(\cdot)$ are monotonically increasing). Third, it is not affected by arbitrary ordinal relabelling of the intensity values.

2.2.3 Histogram-Based Similarity

As an estimator of the joint pdf of image-pair pixel values, the joint histogram seems to be a natural route to measuring similarity from an information theoretic point of view, since the structure of the bivariate pdf will reflect the dependence between the values of corresponding pixels. The correlation-based measures described previously implicitly model this as a linear

dependence and histogram-based measures have the potential to use a more general model for the relationship between two matching images.

Pairing Functions Garret, Reagh and Hibbs propose the pairing function $\mathcal{N}$ as the basis of an image correlation measure [48]. Given images $\mathbf{u}$ and $\mathbf{v}$ with pixels quantized to G levels, define the pairing function $\mathcal{N}$ as the $G \times G$ matrix where the entry $\mathcal{N}_{kl}$ represents the number of times the pixel value k from image $\mathbf{u}$ pairs with pixel value l in the corresponding pixel of $\mathbf{v}$. Note that exact pixel matches accumulate on the diagonal of $\mathcal{N}$. One correlation measure (or similarity measure) based on $\mathcal{N}$ is the simple sum

$$\phi_s(\mathbf{u}, \mathbf{v}) = \frac{1}{N} \sum_{k=0}^{G-1} \mathcal{N}_{kk}(\mathbf{u}, \mathbf{v}),$$

which is the total number of matches divided by the total number of possible matches. The normalized cross correlation can also be written in terms of pairing functions as

$$\phi_{cs} = \frac{\sum_{k=0}^{G-1} \sum_{l=0}^{G-1} kl\mathcal{N}_{kl}(\mathbf{u}, \mathbf{v})}{\left[\sum_{k=0}^{G-1} k^2 \mathcal{U}_k(\mathbf{u})\right]^{1/2} \left[\sum_{l=0}^{G-1} l^2 \mathcal{V}_l(\mathbf{v})\right]^{1/2}},$$

where

$$\begin{aligned} \mathcal{U}_k(\mathbf{u}) &= \sum_{l=0}^{G-1} \mathcal{N}_{kl}(\mathbf{u}, \mathbf{v}) \quad \text{the number of pixels with value } k \text{ in } \mathbf{u} \\ \mathcal{V}_l(\mathbf{v}) &= \sum_{k=0}^{G-1} \mathcal{N}_{kl}(\mathbf{u}, \mathbf{v}) \quad \text{the number of pixels with value of } l \text{ in } \mathbf{v}. \end{aligned}$$

Note that the pairing function itself is actually just the joint histogram of the pixels in $\mathbf{u}$ and $\mathbf{v}$. Figure 2-1 is a simple illustration of the pairing function concept.

One of the motivations for the pairing function concept is that the "easily calculated" expected values of the $\mathcal{N}_{kl}$ allow one to calculate appropriate match thresholds [48]. Garret, Reagh and Hibbs do this for $G = 4$, but following the same procedure for the large number of intensity levels that are common in modern imaging systems is impractical. Of course, the levels can be re-quantized to a manageable number, but this approach discards information from the original image.

(a) Subimage A of the 'peppers' image.

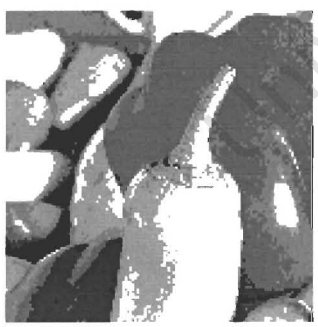
$$\mathcal{N} = \begin{bmatrix} 628 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1026 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2482 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4342 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1641 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3044 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2817 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 404 \end{bmatrix}$$

(b) $\mathcal{N}$ for two identical images: image (a). As expected, $\mathcal{N}$ is a diagonal matrix.

(c) Subimage B of the 'peppers' image.

$$\mathcal{N} = \begin{bmatrix} 8 & 6 & 41 & 127 & 81 & 128 & 125 & 112 \\ 1 & 2 & 96 & 201 & 464 & 185 & 44 & 33 \\ 267 & 105 & 328 & 537 & 383 & 438 & 258 & 166 \\ 80 & 62 & 533 & 1067 & 540 & 794 & 533 & 733 \\ 23 & 3 & 174 & 417 & 272 & 450 & 163 & 139 \\ 46 & 35 & 200 & 594 & 568 & 1034 & 394 & 173 \\ 484 & 290 & 434 & 479 & 155 & 605 & 322 & 48 \\ 57 & 22 & 61 & 155 & 48 & 46 & 9 & 6 \end{bmatrix}$$

(d) $\mathcal{N}$ for two dissimilar images: images (a) and (c). Off-diagonal elements are non-zero, indicating mismatch.



(e) Subimage A with intensity offset of 1 level.

$$\mathcal{N} = \begin{bmatrix} 0 & 628 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1026 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2482 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4342 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1641 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3044 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2817 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 404 \end{bmatrix}$$

(f) $\mathcal{N}$ for images differing only by an intensity offset: images (a) and (e). Non-zero elements are shifted one diagonal to the right with respect to (b).

Figure 2-1: Pairing function matrices for 64 × 64 pixel subimages of the 'peppers' image (quantized to 8 levels).

Mutual Information Viola and Wells [49] propose mutual information (MI) as a measure of fit between a model and an image in order to find the pose of an object. Maes *et al.* use the same principle — that the MI of the image pair will be maximized when **u** and **v** match — to register multimodal medical images [50, 51]. Consider the pixels in **u** and **v** to be samples of the random variables U and V, respectively. A state of mismatch between **u** and **v** can be modelled by statistical independence between U and V, where $p_{UV}(u,v) = p_U(u) \cdot p_V(v)$. A state of match can be modelled as maximal dependence, where $p_U(u) = p_V(v) = p_{UV}(u,v)$. The MI, then, measures the degree of dependence between **u** and **v** by calculating a distance between the estimated pdfs associated with the match and mismatch hypotheses. Maes *et al.* use the Kullback-Leibler distance

$$I(U,V) = \sum_{u,v} p_{UV}(u,v) \log \left[\frac{p_{UV}(u,v)}{p_U(u) \cdot p_V(v)} \right].$$

Another way of stating the rationale of using the MI in **u** and **v** as a measure of similarity is that the amount of information **u** contains about **v** (and vice versa) will be maximal if they match.

In practice $p_{UV}(u,v)$ is estimated by forming the joint histogram of the pixels in **u** and **v**, denoted here as $h(u,v)$. For convenience the pixel values in **u** and **v** are re-scaled to the range $[0, G_u - 1]$ and $[0, G_v - 1]$, where G_u and G_v are the number of histogram bins in each dimension. The required pdfs are then estimated as follows:

$$\begin{aligned} p_{UV}(u,v) &= \frac{h(u,v)}{\sum_{u,v} h(u,v)} \\ p_U(u) &= \sum_v p_{UV}(u,v) \\ p_V(v) &= \sum_u p_{UV}(u,v). \end{aligned}$$

This measure makes no assumptions about the form of the pdfs or the nature of the dependence between corresponding pixels, and its validity is therefore independent of the process generating the individual pixel values. Note that the images do have to be large enough to provide an adequate number of samples for the joint histogram. This rules out MI for tasks like the matching of small blocks for motion estimation and it is more suited to problems where large images from different modalities have to be registered.

Difference-Image Histograms Buzug and Weese measure similarity using the entropy associated with a histogram of the difference image $\mathbf{e} = \mathbf{u} - \mathbf{v}$ [52]. The G-bin histogram is calculated and then normalized to satisfy

$$\sum_{k=1}^{G} p_k(\mathbf{e}) = 1,$$

where $p_k(\mathbf{e})$ denotes the fraction of difference pixels that fall into bin k, where $k \in \{1, \ldots |, G\}$. The entropy is given by

$$M_{\text{entropy}}(\mathbf{u}, \mathbf{v}) = \sum_{k=1}^{G} f(p_k(\mathbf{e})),$$

where $f(x) = -p(x) \log p(x)$ is the entropy weighting function. The rationale behind the use of entropy is that images that don't match will produce a broad difference image histogram (high entropy), and matching images will produce a "peaky" histogram (low entropy). The measure is invariant to pixel value offsets.

The entropy measure is computationally expensive and Buzug *et al.* prove that a class of strictly convex, differentiable functions, which are faster to compute, retain its properties as far as similarity measurement is concerned [24]. Among these is the energy function, $f(x) = x^2$, which gives the similarity measure:

$$M_{\text{energy}}(\mathbf{u}, \mathbf{v}) = \sum_{k=1}^{G} p_k^2(\mathbf{e}).$$

2.2.4 Other Measures

Moghaddam, Nastar and Pentland propose a probabilistic measure for image matching in situations where (1) there are reasonably well-defined image classes that can be described by a Bayesian analysis of intra- and extra-class image differences and (2) training data are available [53]. Their application is face recognition, where the set of intra-class differences, Ω_I, models the variation in different images of the same person, and the set of extra-class differences, Ω_E, models the variation in images of different people. The similarity measure is then the *a posteriori* probability of the differences $d(\mathbf{u}, \mathbf{v})$ between two images belonging to the intra-class model, given by

$$S(\mathbf{u}, \mathbf{v}) = P(\Omega_I | d(\mathbf{u}, \mathbf{v})).$$

The differences are represented by a parameterized model of the deformation between $\mathbf{u}$ and $\mathbf{v}$, denoted here as $\tilde{\mathbf{U}}$.

Using the maximum *a posteriori* (MAP) rule, two images of the same person's face match if $P\left(\Omega_I|\tilde{\mathbf{U}}\right) > P\left(\Omega_E|\tilde{\mathbf{U}}\right)$, or equivalently, if $S(\mathbf{u},\mathbf{v}) > 1/2$. Bayes rule,

$$P\left(\Omega_I|\tilde{\mathbf{U}}\right) = \frac{p\left(\tilde{\mathbf{U}}|\Omega_I\right)P(\Omega_I)}{p\left(\tilde{\mathbf{U}}|\Omega_I\right)P(\Omega_I) + p\left(\tilde{\mathbf{U}}|\Omega_E\right)P(\Omega_E)},$$

represents the *a posteriori* pdfs in terms of the class conditional pdfs, $p\left(\tilde{\mathbf{U}}|\Omega_I\right)$ and $p\left(\tilde{\mathbf{U}}|\Omega_E\right)$, which can be estimated from training data. The high dimensionality of $\tilde{\mathbf{U}}$ makes the estimation impractical, but the authors overcome this problem by using the principal components of $\tilde{\mathbf{U}}$ and an efficient technique for directly obtaining the pdf of these components from training data.

Using this approach within the well-known Eigenfaces face recognition algorithm, Moghaddam, Jebara and Pentland were able to improve performance in tests using the FERET face database [54].

2.3 Evaluation and Comparison of Similarity Measures

Many algorithms, such as stereo matching and motion estimation, depend on an early stage of direct image matching for their success. The relative and absolute performance of different similarity measures should therefore be of interest to the designers of these algorithms. This section reviews work done in comparing similarity measures and analyzing their performance.

2.3.1 Similarity Measure Comparisons

Given the wide use of similarity measures in image processing, it is surprising that relatively few broad comparisons of the different approaches can be found. Different similarity measures have often been compared in the context of a very specific application, and the particular advantages of newly introduced similarity measures have often been demonstrated using limited experimentation. These investigations are too numerous to mention, and have limited value as the basis for a broad comparison of approaches to the problem of direct image matching.

There have been some attempts to assess similarity measures in a slightly broader context. Svedlow, McGillem and Anuta compare classical measures and image preprocessing operations

experimentally using multi-temporal images from the Landsat multispectral scanner over a series of test sites [35]. They compare the correlation function, correlation coefficient and the sum of absolute differences when used in translational image registration. Measuring the percentage of acceptable registrations in a given number of registration attempts, they find that the correlation coefficient is marginally better than the sum of absolute differences, but that the latter is desirable for its computational efficiency. The effect of different preprocessing methods — a gradient operator, a threshold operator and a combination of the two — is also analyzed. They show theoretically that if registration is viewed as a matched filter operation, then the gradient operator is the optimal pre-processor for image data that can be modelled as a first order Markov process. This is confirmed experimentally, with the magnitude of the gradient preprocessor producing the best overall results.

Aschwanden and Guggenbühl compare a broader range of similarity measures under a wider range of conditions [55]. Several variations of the cross-correlation, sum of squared differences and sum of absolute differences are selected as the subject of the experiment on the basis of their computational efficiency. Measures that involve preprocessing by gradient operators, high-pass filters and band-pass filters are also included. Three images serve as the source of the test data used in the experiments, representing various textures, various scenes with high edge content and a real-world laboratory scene. These images are subjected to varying degrees of illumination change, Gaussian noise, salt-and-pepper noise, image blur and magnification to provide sequences of test images. The experiments investigate translational registration in a region of interest around several pre-selected positions and use the deviation of the match-peak coordinates from the known correct coordinates to compare the performance of similarity measures. The authors conclude that the classical correlation-based measures are robust in the presence of distortion, that the normalized and zero-mean measures show the expected invariance to illumination changes, and that the performance of measures with high-pass filter preprocessing deteriorate catastrophically under high levels of distortion.

Matched filters and phase correlation have importance as realizable operations in optical processing systems. Horner and Gianino compare standard matched filters with amplitude-only and phase-only matched filters and conclude that phase correlation is superior in most situations [56].

Penny *et al.* compare the performance of six different similarity measures for registering 2D clinical fluoroscopy images to 2D radiographs that are reconstructed from 3D Computed Tomography (CT) data [26]. Scene changes in the time between imaging with the two modal-

ities, and differences in the image formation of the modalities complicate this matching task. The normalized cross correlation, difference image entropy, mutual information, normalized correlation in edge images, and two variations on difference image entropy called pattern intensity and gradient difference, are compared. The CT scan and fluoroscopy image of a spine phantom are used as test data and a "gold standard" correct registration is calculated using fiducial markers. These images are made more realistic by adding features segmented from clinical images such as soft tissue and interventional instruments. In experiments the mutual information has the worst performance, the correlation-based measures are found to be sensitive to the thin structure and large intensity differences created by the presence of a medical instrument in the fluoroscopy image, and the entropy-type measures are found to be sensitive to the slowly varying differences caused by soft-tissue. The pattern intensity and gradient difference measures register the images accurately in the presence of both medical instruments and soft tissue.

Meijering, Niessen and Vergiever provide a comprehensive review and qualitative comparison of the similarity measures that have been used in the registration of digital angiography images [57]. Aside from a general summary of the relative merits of different approaches, they question the competence of the CBC measure, noting that it suffers from inconsistent weighting of intensity differences and cannot live up to claims of noise insensitivity. They also conclude that the energy of the histogram differences by Buzug *et al.* [24] (M_{energy} above) is the best measure for the digital subtraction angiography application.

2.3.2 The Effect of Distortion on Matching

Mostafavi and Smith conduct a theoretical investigation into the effect of affine geometric distortion on translational image registration with the correlation function, both with respect to the probabilities of false and correct registration [58] and with respect to the accuracy of registration [59]. The images are modelled as a reference image, $w_r(\mathbf{x}) = u(\mathbf{x})$, and a sensor image, $w_s(\mathbf{x}) = v(\mathbf{x}) + n(\mathbf{x})$, where the reference image is assumed to be a smaller part of the image plane and $\mathbf{x} = (j, k)$ represents image plane coordinates. Notice that for the sake of simplicity, only the sensor image is considered to contain noise, which, the authors claim, does not have a qualitative effect on the results. Since the distortion is assumed to be affine, the reference and sensor image are related by $v(\mathbf{x}) = u(\mathbf{A}\mathbf{x} + \mathbf{t}_0)$. For the affine distortion

the authors consider

$$\mathbf{A} = \alpha \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix},$$

which consists of rotation by θ radians and scaling by a factor of α. Image $v(\mathbf{x})$ and noise $n(\mathbf{x})$ are modelled as independent, stationary, multivariate normal random fields.

In order to assess the probabilities of false and correct registration, Mostafavi and Smith derive the peak-to-sidelobe ratio of the correlation function from the mean peak value and the variance of non-peak values. Modelling the correlation function output as a normal random variable, the probability of false acquisition P_{FA} is also analyzed as the probability of correct acquisition is held constant at 0.99. For a given signal-to-noise ratio, a maximum peak-to-sidelobe ratio and minimum P_{FA} suggest the same optimum window size. This image size is proportional to the image autocorrelation function, and is inversely proportional to the degree of distortion between $u(\mathbf{x})$ and $v(\mathbf{x})$.

Mostafavi and Smith also investigate the deviation of the registration peak position from the correct position that is caused by affine geometric distortion [59]. They show theoretically that a first order approximation to this local registration error is proportional to the gradient of the match surface at the position of correct registration, and inversely proportional to the curvature at the peak. Both results are intuitively appealing: the gradient at a peak in the match surface will be zero and will increase as one moves away from the peak, supporting the former. For the latter, high curvature implies a narrower peak and more accurate localization. One conclusion from this investigation is that, here too, there is an optimum window size proportional to the width of the autocorrelation function and inversely proportional to the amount of geometric distortion. In contrast with the result for probabilities of false and correct registration, however, the window size minimizing local registration error is smaller than the size that minimizes P_{FA} for a fixed geometric distortion.

2.3.3 Metrics for Matching Performance

Similarity measures have not been evaluated or compared using a universally accepted performance measure or set of measures. Most comparisons are based on a translational image registration experiment. Svedlow, McGillem and Anuta classify acceptable and unacceptable registrations manually and use the number of these as a relative performance measure in their experiments [35]. When evaluating the effect of geometric distortion on matching, Mostafavi

and Smith use the probability of correct and false registration as a performance measure in one study [58] and the spatial registration accuracy in another [59]. Aschwanden and Guggenbühl base their comparison on the spatial accuracy of registration alone [55]. Venot, Lebruchec and Roucayrol present results for a small number of specific cases, showing that their measure outperforms others [42]. Radcliffe, Rajapakshe and Shalev perform registration with translation, rotation and magnification and use deviations from all three of the known correct values in controlled experiments to evaluate their algorithm [32]. Chiang and Sullivan compare their measure to others using correct versus incorrect registration and a qualitative comparison of the match surfaces [47]. Buzug *et al.* also analyze the match surface, using the broadness of the peak and the extent of its "attractive basin" [24]. Bhat and Nayar compare measures using the percentage of mismatches that occur in a set of registration experiments [25]. Penny *et al.* use the RMS error between known correct registration parameters and the actual values produced by experiment [26].

When matching is discussed in the context of a filtering operation, the concept of SNR in the filtered image (match surface) is often used to evaluate matching performance. Kuglin and Hines ratio the power of the match peak and the power of background peaks in the match surface [18]. Horner and Gianino compare SNR for standard and phase-only matched filters [56].

The abovementioned performance measures are only useful in a relative sense since they are dependent on the experimental data. Sadjadi compares four separability measures that have independently meaningful values since they are estimates of the probability of error in the match/mismatch decision [60]. The first is the probability of error associated with the optimal Bayes decision rule for the correlation function. The second and third are the Chernoff and Battacharyya bounds on the Bayes probability of error, which are easier to compute. The fourth is Fisher's criterion

$$F = \frac{(m_1 - m_2)^2}{\sigma_1^2 + \sigma_2^2},$$

which is easiest to compute, but relies on the assumption that the similarity values are normally distributed and is unpredictable if this assumption is violated.

2.4 Discussion

The preceding review confirms that there are many different approaches to direct image matching. Similarity measures estimate linear correlation and differences between corresponding pixels, using classical, robust and nonparametric methods. More general measures evaluate the structure of joint and difference image histograms in order to identify dependence between corresponding pixels. The most common measures have been compared experimentally and some have been analyzed theoretically regarding their performance under various conditions of noise and image distortion.

Aside from very broad rules of thumb, however, the selection of a similarity measure in the design of an image matching algorithm is not guided by well understood principles. Although each of the available measures has a sound rationale, there is no guarantee that performance will be optimized by any one of them in a new application. The published comparisons of different measures are dependent on the data sets used in the experiments and there is no way to predict matching performance for a new application without repeating the analyses reported in the literature. There is also no systematic procedure for designing an application-specific similarity measure, nor are any of the available measures the outcome of such an approach. Finally, the upper bound on matching performance has only been investigated in idealized conditions and for the most tractable measures.

The remainder of this dissertation addresses these deficiencies through a rigorous formulation of the image matching problem.

Chapter 3

Formulating the Image Matching Problem

Before formulating the image matching problem, it is necessary to define what is meant by match and mismatch in the context of an imaging system. Consider the model of image formation shown in Figure 3-1. First some phenomenon creates a physical scene (e.g. man builds a production line, nature grows a lung). The scene is then irradiated, either by nature (e.g. the sun) or by the imaging system itself (e.g. artificial lighting, an X-ray generator). This radiation is then processed in some way by the scene and the result is transformed into a digital image by the image capture system. For the purpose of this investigation, the following

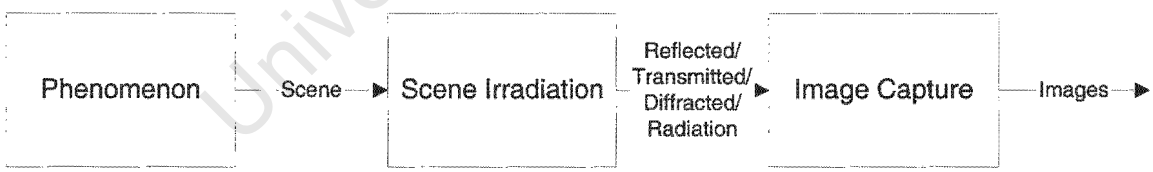


Figure 3-1: Image formation model.

important distinction is made: match or mismatch is a characteristic of the scene, rather than a characteristic of the captured images. The image-pair only contains *evidence* of match or mismatch in the scene. The objective of this research is to find ways to make the best possible use of this evidence and any available *a priori* information. To a large degree, the type of *a priori* information determines the approach taken by the designer. This research assumes that only vague information, typically described in terms of a stochastic model, is available. No deterministic assumptions are made about the image content.

The input data to the matching algorithm is a pair of images. Even if these images match, they usually differ in at least one image formation parameter — the images may have been formed at different times, in different orientations, by different image capture systems, or using different imaging modalities. Whatever the situation, the problem is one of using the evidence in the two images to decide whether they represent the same scene or not. Two fields tackle the problem of designing a decision rule for an observed quantity, where the *a priori* information about the source of the observation is expressed as a stochastic model: Detection Theory and Pattern Recognition. There is a large degree of overlap in the early theoretical foundations of these fields (both have their origins in statistical decision theory), but their paths have subsequently diverged. Detection theory has developed into a specialization of statistical hypothesis testing, whereas pattern recognition has become a branch of artificial intelligence. Sections 3.1 and 3.2 outline the primary concepts in these respective fields, and form the basis of a decision to formulate the matching problem in a hypothesis testing framework that is described in Section 3.3. Section 3.4 explores this formulation by applying it to the trivial problem of scalar matching. Section 3.5 concludes the chapter with a discussion on the material covered.

3.1 Detection Theory and Hypothesis Testing

The two major areas of statistical inference are parameter estimation and tests of hypotheses [61]. The former involves estimating a stochastic process parameter from an observation of the process, whereas the latter involves making an assertion, or conjecture, about the distribution of the stochastic process generating the data. This assertion is formulated as a *hypothesis*, denoted H, which the analyst either accepts or rejects on the basis of a *test*, denoted $\mathcal{T}$, on the observation. Many hypothesis testing problems involve two competing hypotheses. In this case one of them is referred to as the *null* hypothesis H_0, and the other as the *alternative* hypothesis H_1. In some cases one of the hypotheses, normally H_1, is in some sense more important than the other and is referred to as the *emphatic* hypothesis.

Detection theory is essentially an extension of statistical hypothesis testing that is specialized for the analysis of signals, particularly those that originate from communications and biological systems [62, 63]. Kazakos and Papantoni-Kazakos view detection and estimation as the search for the stochastic process that best describes a physical phenomenon, given observations of that phenomenon [64]. The distinction between detection and estimation is that the former searches a set of stochastic processes that has finite membership, whereas for

the latter the set is infinite.

3.1.1 The Hypothesis Test

The test is a procedure for deciding whether to accept or reject the hypothesis H_0 [61, p. 403]. Denote the sample space of observations as X and a single observation as $\mathbf{x}$. The nonrandomised, single hypothesis test rejects H_0 if and only if the observation $\mathbf{x} \in \mathbf{C}_{\mathcal{T}}$, where $\mathbf{C}_{\mathcal{T}}$ is called the *critical region* of $\mathcal{T}$ and is a subset of X. The test procedure is to observe $\mathbf{x}$ and check whether it falls inside the critical region $\mathbf{C}_{\mathcal{T}}$, accepting H_0 if it does. Design of the test can be viewed as an optimization problem. Given a stochastic model for the observation and a performance criterion, the optimal critical region must be found.

Models Models are generally described as well-known, parametrically known, or nonparametrically described [64, p. 2]. Well-known models have pdfs of known form with no unknown parameters. Parametric models have a pdf of known form, but include one or more unknown parameters. Nonparametric models may include some knowledge of the pdf, but its form is not completely specified. A hypothesis is described as *simple* or *non-composite* if the process associated with it is well-known. Hypotheses that have parametric models are referred to as *composite*.

Performance Criteria The performance criteria are related to the errors made by the test [61, p. 405]. Two types of error are possible: The *type I error*, or *false positive*, occurs if H_0 is rejected when true. The *type II error*, or *false negative*, occurs if H_0 is accepted when false. The probability of these errors occurring is referred to as the *size* of the errors in some texts. In some applications these error rates are part of a more general cost function that weights type I and type II errors differently.

Instead of the hypotheses representing two separate processes, it is sometimes useful to model them by disjoint subdivisions of the parameter space of a single parametrically known stochastic process [64, p. 41]. Denote the parameter space as Θ, Θ_1 as the subspace associated with the emphatic hypothesis H_1, and Θ_0 as the subspace associated with the null hypothesis H_0. The *power function* of the test $\mathcal{T}$, denoted $\pi_{\mathcal{T}}(\boldsymbol{\theta})$, is defined as the probability that H_1 is accepted as a function of the parameter $\boldsymbol{\theta} \in \Theta$. The power function is useful as a means of comparing alternative tests. Note that the type I and type II errors can be written in terms

of the power function as

$$P_I = P\left(H_1 \text{ accepted} | H_0 \text{ true}\right) = \int_{\Theta_0} \pi_{\mathcal{T}}\left(\boldsymbol{\theta}\right)$$

and

$$P_{II} = P\left(H_0 \text{ accepted} | H_1 \text{ true}\right) = 1 - \int_{\Theta_1} \pi_{\mathcal{T}}\left(\boldsymbol{\theta}\right),$$

respectively. The test that minimizes type II errors with the probability of type I errors fixed at α is called the *most powerful test* of size α. The *size of a test* is the maximum probability of accepting H_1 when the true hypothesis is H_0. This can be written in terms of the power function as $\sup_{\boldsymbol{\theta} \in \Theta_0} \pi_{\mathcal{T}}\left(\boldsymbol{\theta}\right)$.

3.1.2 Tests of Simple Hypotheses

The appropriate test for two simple hypotheses will be either a *Bayes test*, *minimax test* or *Neyman-Pearson test*, depending on the *a priori* information available [64, p. 46]. The Bayes test requires knowledge of the *a priori* probabilities of the hypotheses being true. If a cost function is available, the Bayes test is designed to minimize this cost. If not, the test, often referred to as the *ideal observer test*, minimizes the probability of error. If no *a priori* probabilities are available, but a cost function is given, then a minimax rule is used. This type of test essentially minimizes the cost function for the least favourable set of *a priori* probabilities. With no *a priori* probabilities or cost function, the problem is solved by setting an upper limit on the probability of an error in the non-emphatic hypothesis (type I error) and minimizing the probability of error in the emphatic hypothesis (type II error). The result is known as the Neyman-Pearson test.

It is often convenient to write the hypothesis test in terms of a statistic $s\left(\mathbf{x}\right)$ and a threshold λ, for example

$$\text{Accept} \begin{cases} H_1 \text{ if } s(\mathbf{x}) > \lambda \\ H_0 \text{ if } s(\mathbf{x}) \leq \lambda \end{cases}. \tag{3.1}$$

Let the observation $\mathbf{x}$ be a random sample from the pdf $p\left(\mathbf{x}|H_0\right)$ or $p\left(\mathbf{x}|H_1\right)$, conditioned on whether H_0 or H_1 is true, respectively. The likelihood ratio,

$$l\left(\mathbf{x}\right) = \frac{p\left(\mathbf{x}|H_1\right)}{p\left(\mathbf{x}|H_0\right)},$$

with an appropriate threshold is a common test. In fact, it has been proved that for two simple hypotheses the optimum Bayes test, minimax test and Neyman-Pearson test can all be written in terms of the likelihood ratio statistic and an appropriate threshold, which together constitute the *likelihood ratio test* (LRT) [61, p. 410-418].

3.1.3 Tests of Composite Hypotheses

Consider now the situation where the hypotheses are composite and share a parameterized pdf, $p(\mathbf{x}|\boldsymbol{\theta} \in \Theta)$, with a disjoint subdivision of the space of the unknown parameter: $\boldsymbol{\theta} \in \Theta_0$ for H_0 and $\boldsymbol{\theta} \in \Theta_1 = \Theta - \Theta_0$ for H_1. A common test for this scenario uses the generalized likelihood ratio

$$l(\mathbf{x}) = \frac{\sup_{\boldsymbol{\theta}\in\Theta_1} p(\mathbf{x}|\boldsymbol{\theta})}{\sup_{\boldsymbol{\theta}\in\Theta} p(\mathbf{x}|\boldsymbol{\theta})}$$

and an appropriate threshold, and is called the *generalized likelihood ratio test* (GLRT) [65]. Unlike the likelihood ratio test for simple hypotheses, the GLRT is not optimal in any sense, but has proved to be a good test in many (but not all) situations [61, p. 419]. A test is called the *uniformly most powerful test* of size α if it is the most powerful test among all tests of size α or less. This is a useful performance criterion for optimal tests of composite hypotheses.

3.1.4 Nonparametric and Robust Tests

The hypothesis testing schemes described in previous sections depend heavily on the accuracy of the stochastic model of the underlying process, since optimum performance is only guaranteed for observations of the process described by this model. However, the model is only an approximation of reality and where the approximation is inaccurate the performance might deteriorate. This problem is addressed by robust and non-parametric statistical methods.

Robust Tests The robust approach is to view the observations as consisting of two components — a well behaved component that can be described by a classical stochastic model as before, and a corrupt component that introduces deviations from this model. Corrupted observations are called outliers and robust tests are designed to tolerate these apparent anomalies. Robust statistics originated as ad-hoc procedures for eliminating outliers before doing statistical analysis, but has recently developed into a set of theoretical tools for designing and evaluating hypothesis tests and estimation schemes that tolerate various types of model contamination [38, 30]. Huber describes three desirable characteristics of a robust procedure:

near optimal performance under the classical model, a small drop in performance for small deviations from the model and non-catastrophic failure for larger deviations from the model [38, p. 5].

The use of robust methods in the design of hypothesis tests can vary from the use of robust estimators for unknown parameters, to the design of a test from first principles using a robust formulation of the problem. The field of robust statistics has mostly addressed univariate parameter estimation problems, making the former a more practical approach for high dimensional detection problems.

Robust statistical theory considers neighbourhoods of parametric models and is therefore a branch of parametric statistics [30, p. 9]. Another branch of statistical theory addresses the lack of accurate information by making weak assumptions and using nonparametric models.

Nonparametric Tests Nonparametric tests are used when a complete statistical model of the input data is not available or the underlying statistical process changes over time, or if the optimal test is too complex for practical implementation [41]. These methods differ from robust methods in that fewer modelling assumptions are made. Where robust methods operate in the neighbourhood of a parametric model, non-parametric models discard the parametric model completely. Tests are ad-hoc and based on empirical observations rather than the solution of a formal optimization problem [64, p. 199]. The oldest and most widely used class of nonparametric procedures are the sign tests, the simplest form of which identifies an observation as belonging to one of two processes that differ only in their means, which are symmetric around the origin [64, p. 200]. The test statistic in this case is

$$s(\mathbf{x}) = \sum_i \operatorname{sgn}(x_i) \quad \text{where} \quad \operatorname{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}.$$

Another common nonparametric approach uses the rank of observations, rather than their values, which results in approaches that are tolerant of large absolute deviations in isolated values and have useful invariance properties [64, p. 205].

Although ad-hoc, the nonparametric approach is the only way to arrive at practical tests in many situations. However, the lack of rigour leads to difficulty in comparing different nonparametric approaches analytically. The *asymptotic relative efficiency* (ARE), which measures the performance of one test relative to another, was designed to overcome this problem. Specific definitions differ, but the ARE generally measures the limiting ratio of the number of samples

per observation that two detectors require to achieve the same performance, as the distance between the hypotheses (or SNR) becomes infinitely small. Kazakos and Papantoni-Kazakos define ARE to compare a nonparametric detector with a Neyman-Pearson parametric detector on observations from the same parametric model [64, p. 199]. Helstrom compares two arbitrary detectors with fixed type I and type II error characteristics [62, p. 167]. Huber uses ARE with respect to a classical procedure as the performance criterion for a robust one [38, p. 5].

3.2 A Pattern Recognition Perspective

Pattern recognition also addresses the design of decision rules based on observations and models of physical phenomena [66, 67, 68, 69], but has strong ties with the research topics of computer vision and artificial intelligence. Many pattern recognition applications involve developing systems that mimic an aspect of human decision making ability. Weather forecasting, character recognition, speech recognition, and image interpretation are examples [67, p. 12].

Like detection theory, pattern recognition had its origins in statistical decision theory. Many of the concepts in this field are shared with detection theory and hypothesis testing, but the nomenclature and notation are, for the most part, completely different. There are also many differences on a practical level, which is not surprising if one considers the fundamental differences in their early application areas and the large differences in dimensionality and statistical complexity of the problems tackled by these respective fields.

Analogue versus Discrete Origins Many detection theory concepts originate from early work done in analog signal processing. Here signals were often sampled at fixed intervals to provide the observations for a hypothesis testing procedure. These sequential tests included both a stopping rule to recognize when enough data had been collected and a decision rule to test the actual hypothesis [64, p. 37]. Pattern recognition, on the other hand, originated in artificial intelligence research and early work attempted to classify a few heuristically selected features that were sensed from some phenomenon of interest. Hence, where the data vector in a detection theory problem is often a number of regularly spaced samples from the same process, the elements of a feature vector in a pattern recognition problem routinely consists of measurements of completely different physical phenomena.

Dimensionality Designers using detection theory in signal processing applications are accustomed to high dimensionality, and are restricted to analytical design and limited estimation of distribution parameters using training data. The pattern recognition community, on the other hand, typically deals with low dimensional problems where on-line training of discriminant functions is practical. Indeed, the "curse of dimensionality" is discussed in pattern recognition texts [68, p. 7] for dimensions that are commonplace in detection theory circles.

Statistical Complexity Historically, detection theory has been applied in areas where relatively simple fundamental models of the underlying physical phenomena can be developed and validated. Pattern recognition, on the other hand, has tackled more unruly data, often where explicit statistical modelling is not feasible. As a result, it can be argued that members of the pattern recognition community have had to be more resourceful and less rigorous than their detection theory counterparts.

More recently, with the increasing statistical complexity of the digital signals that are routinely tackled by detection theory and the increasing dimensionality of pattern recognition problems, a unification of these two fields seems appropriate. A brief overview of pattern recognition fundamentals is now given, with references to the previous sections on hypothesis testing where equivalent concepts or fundamental differences are identified.

3.2.1 Classes, Classifiers and Discriminant Functions

The pattern recognition equivalent to the detection theory hypothesis test is the *classifier*. Classifiers are derived from models in much the same way as hypothesis tests, or are learned from training data. They partition the observation space using *discriminant functions*, which Duda and Hart refer to as "something of a canonical form for classifiers" [66, p. 17]. In a similar fashion the hypothesis test statistic and decision threshold of Section 3.1 define a critical region for the null hypothesis.

The c distinct states of nature that the classifier must recognize, which are equivalent to detection theory hypotheses, are called *classes* and denoted ω_i, where $i \in \{1, 2, \ldots, c\}$. The observations are referred to as *feature vectors*, since they should encapsulate the salient features of the underlying physical phenomenon. A feature vector, denoted here as $\mathbf{x}$, is assigned to class ω_i if

$$g_i(\mathbf{x}) > g_j(\mathbf{x}) \quad \forall\, i \neq j,$$

where $g_i(\mathbf{x}) : i \in \{1, 2, \ldots, c\}$ are the discriminant functions. For the two class problem, a single discriminant function, $g(\mathbf{x}) = g_1(\mathbf{x}) - g_2(\mathbf{x})$, is used with the decision rule

$$\text{Decide} \begin{cases} \omega_1 \text{ if } g(\mathbf{x}) > 0 \\ \omega_2 \text{ if } g(\mathbf{x}) \leq 0 \end{cases}.$$

Notice the similarity to the hypothesis test of (3.1).

The surface defined by $g(\mathbf{x}) = 0$ partitions the feature space into two decision regions and is referred to as the *decision surface*. Pattern recognition places more emphasis on the decision surface than is the case in detection theory, which concentrates on characterizing the hypotheses (classes) rather than the boundaries between them.

3.2.2 Supervised Learning and Numerical Optimization

Like the design of tests in detection theory, the design of discriminant functions often requires that distribution parameters of the various classes are estimated from training data. This is usually referred to as *supervised learning* in the pattern recognition literature [66, p. 44]. However, pattern recognition takes supervised learning further than just estimating distribution parameters, by using data to *train* discriminant functions directly. The most common example of this approach is the neural network, a mathematical construct with many free parameters that are specified by supervised learning, or training [68]. The data models are developed within the neural network during the training phase and are limited only by the flexibility of the network architecture and the extent to which the training data faithfully represents the problem. More recent supervised learning techniques, such as the support-vector machine, use training data to build the decision surface directly [70].

In a sense, trainable discriminant functions like neural networks are just methods for solving the analytical optimization problem associated with deriving the optimal hypothesis test, but do so numerically with training data. Instead of specifying models for the different classes (hypotheses) and deriving the test using analytical optimization, a configurable discriminant function with many free parameters is specified and the best set of parameters sought using training data and a numerical optimization algorithm, such as back-propagation or a genetic algorithm. As is the case with designing hypothesis tests, the performance criteria that direct the optimization search are also based on the error-rate performance of the solution, or more generally, a risk or cost function that has different weightings for different types of error.

Supervised learning and numerical optimization share a problem that is absent from analytical optimization methods: unrepresentative training data or an inadequate search can lead

to a local, instead of global, maximum of the performance criteria. Although this problem is absent from the analytical optimization stage of the detection theorist, it can be compared to the problem of selecting a model for the data, since the supervised learning is, in a sense, both selecting the model and performing the optimization. For example, a nonparametric model may sacrifice performance in ideal simulations, relative to a parametric one, but the detection theorist may select it so that the system will tolerate deviations from the parametric model in real data. This designer will have effectively generalized better than the designer who chose the parametric model for its superior performance in simulations.

3.2.3 Unsupervised Learning and Clustering

Pattern recognition takes learning another step further than detection theory. In *unsupervised learning*, or *clustering*, the training set is unlabelled and the learning algorithm must search for natural groupings, or *clusters*, in the data. Essentially, the designer is still solving an optimization problem, but is doing so having specified different initial information, which could include the number of classes in the data and criteria that describe "good" clusters. Unsupervised learning procedures include estimation techniques (e.g. Gaussian mixture modelling [66, p. 190]), extensions of supervised techniques to unsupervised learning (e.g. unsupervised Bayesian learning [66, p. 203]), clustering algorithms and neural networks that identify classes during training (e.g. the Kohonen self organizing map [67, p. 162][71]).

3.2.4 Dimensionality Reduction

A common nonparametric procedure in pattern recognition is density estimation, where sample data are used to estimate the pdf of the feature vector. Nonparametric estimation methods, such as histograms and Parzen windows, have been developed, but a large number of samples are required and this number grows exponentially with the dimensionality of the feature space [66, p. 95]. This phenomena is referred to as the "curse of dimensionality" [68, p. 7] and also applies to supervised and unsupervised learning techniques. Detection theory has escaped this problem by rigorous analysis that provides parametric models, or where this fails by bypassing the density estimation problem and going directly to suboptimal nonparametric tests. The learning approach has been largely overlooked in detection theory, probably because high dimensionality and available computing technology have made it impractical.

In pattern recognition the curse of dimensionality has been addressed by using feature extraction or dimensionality reduction techniques, which preprocess the input data and extract

only the most salient features. The simplest approach is to use only a subset of the available features, chosen according to some set of criteria [68, p. 304]. A more general approach is to map the input data into a lower dimensional space. A common method for doing this is principal component analysis (PCA) [68, p. 310], which represents the n-dimensional data as the coefficients of a set of p principal components, where $p < n$. The principal components themselves are the eigenvectors associated with the p largest eigenvalues of the covariance matrix of the training data. A famous example of this approach in practice is the eigenfaces face recognition algorithm, where PCA reduces the large number of pixels in a face image to a more manageable feature vector [72, 73].

3.3 Problem Formulation

The previous sections in this chapter have given a brief overview of mathematical techniques developed for automating a decision making process that is based on observations of the real world. Attention now turns to formulating the decision making problem at hand — deciding whether a pair of images are in a state of match or mismatch.

Consider the space, Ψ, of all possible image pairs that the imaging system can generate, and a particular image-pair observation, $\bar{\psi} \in \Psi$. Based on the evidence in this observation, the algorithm must decide whether the two images represent equivalent scenes. If it is assumed that the algorithm makes a deterministic decision (i.e. the algorithm will always make the same decision for a given image pair), then the algorithm *must* create a partition of Ψ into image pairs that represent a state of match, $\bar{\psi} \in \Psi_1$, and image pairs that represent a state of mismatch, $\bar{\psi} \in \Psi_0 = \Psi - \Psi_1$. The ultimate solution, then, to the problem of designing a matching algorithm can be expressed very simply as the decision rule,

$$\hat{\delta}\left(\bar{\psi}\right) \equiv \text{Decide} \begin{cases} \text{match} & \text{if} \quad \bar{\psi} \in \Psi_1 \\ \text{mismatch} & \text{if} \quad \bar{\psi} \in \Psi_0 \end{cases}, \tag{3.2}$$

that optimizes the matching performance criteria over all potential decision rules, and satisfies any other requirements of a matching algorithm. Over and above being the optimal solution, the following general requirements are deemed important:

1. The algorithm should make full use of the *a priori* information available.

2. The algorithm should tolerate minor deviations from any modelling assumptions made.

3. The algorithm should be computationally feasible.

In practice, these requirements are in conflict, so any algorithm will be a compromise between them. However, the compromise should be made explicit and should not be the result of an ad-hoc approach to solving the problem.

Note that the decision rule of (3.2) suggests a formulation of the matching problem that has not been previously explored in any detail: model the situation in terms of two stochastic processes. Not one for each image, but one for each hypothesis:

1. A process that produces matching images, modelled with the multivariate pdf $p_{\bar{\psi}}\left(\bar{\psi}|H_1\right)$ and

2. a process that produces non-matching images, modelled with $p_{\bar{\psi}}\left(\bar{\psi}|H_0\right)$.

Since this model does not map directly onto a single physical imaging system it is rather counter-intuitive, but it is also the natural decision theoretic formulation of the problem given an observation (the image pair) and hypotheses that must be tested (match and mismatch).

3.3.1 Detection Theory versus Pattern Recognition

Superficially, detection theory and pattern classification both appear to be a natural framework for the image matching problem: A random process generates two images, which are the observation or feature vector. The hypotheses, or classes, are match and mismatch. The problem of designing an algorithm for image matching is now one of designing a good test of these hypotheses, or a good discriminant function to separate these classes. The preceding sections in this chapter have outlined the detection theory and pattern recognition approaches to this sort of problem. In summary, detection theory offers a more rigorous approach based on analytical models, has an emphasis on the optimal nature of a solution, routinely deals with highly dimensional observations and has been applied to problems where the observation dimensions are samples of a single process. In contrast, pattern recognition has an emphasis on unsupervised learning, rather than rigorous modelling of the underlying phenomena that are the source of the data, has been developed primarily for fewer, less well-known dimensions, and has been applied to a more diverse range of phenomena.

Given that the emphasis in this research is on the use of *a priori* stochastic models, the detection theory approach is chosen as the starting point for developing algorithms for image matching. This approach is also more suited to the high dimensionality of images and will

exploit the analytical image models that have been successfully applied in other areas of image processing.

3.3.2 Matching as Hypothesis Testing

In order to formulate the image matching problem as a hypothesis test, the following ingredients are required: an observation, a model of the physical process generating the observation under each hypothesis and the performance criteria that will be used to assess the test.

Observation The observation $\bar{\psi}$ is a pair of digitized images, which can be represented as the column vectors $\mathbf{u}$ and $\mathbf{v}$, where the pixels are stored in row-column order. In future this will be referred to as the image pair, denoted

$$\mathbf{w} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}.$$

Hypothesis Models The two hypotheses are match H_1, and mismatch H_0, and the image pair is modelled as two separate stochastic processes under these hypotheses. Note that unless stated otherwise, match is the emphatic hypothesis. Four scenarios are identified:

1. *Simple match*: The two simple hypotheses are modelled as well-known stochastic processes that share the pdf, $p_{\mathbf{w}}(\mathbf{w}|\boldsymbol{\theta})$, where $H_1 \iff \boldsymbol{\theta} = \boldsymbol{\theta}_1$ and $H_0 \iff \boldsymbol{\theta} = \boldsymbol{\theta}_0$.

2. *Simple match with nuisance parameters*: The two simple hypotheses are modelled as parametrically known stochastic processes that share the pdf, $p_{\mathbf{w}}(\mathbf{w}|\boldsymbol{\theta}, \boldsymbol{\varphi})$, where $H_1 \iff \boldsymbol{\theta} = \boldsymbol{\theta}_1$ and $H_0 \iff \boldsymbol{\theta} = \boldsymbol{\theta}_0$, and $\boldsymbol{\varphi}$ is a vector of unknown (nuisance) parameters.

3. *Composite match*: The two composite hypotheses are modelled as parametrically known stochastic processes that share the pdf, $p_{\mathbf{w}}(\mathbf{w}|\boldsymbol{\theta})$, where $H_1 \iff \boldsymbol{\theta} \in \Theta_1$ and $H_0 \iff \boldsymbol{\theta} \in \Theta_0$.

4. *Composite match with nuisance parameters*: The two composite hypotheses are modelled as parametrically known stochastic processes that share the pdf, $p_{\mathbf{w}}(\mathbf{w}|\boldsymbol{\theta}, \boldsymbol{\varphi})$, where $H_1 \iff \boldsymbol{\theta} \in \Theta_1$ and $H_0 \iff \boldsymbol{\theta} \in \Theta_0$, and $\boldsymbol{\varphi}$ is a vector of nuisance parameters.

Performance Criteria Type I and type II error rates will be used to compare the performance of different tests. This will be expressed in terms of the overall probability of error, or

as the size of a test with fixed type I error (false alarm) rate, where match is the emphatic hypothesis.

3.3.3 Defining Image Similarity

Without any loss in generality, the two-hypothesis problem can be expressed in terms of statistic of the image-pair, $s(\mathbf{w})$, and a decision threshold, λ, with the decision rule

$$\text{Accept} \begin{cases} H_1 \text{ if } s(\mathbf{w}) > \lambda \\ H_0 \text{ if } s(\mathbf{w}) \leq \lambda \end{cases}. \tag{3.3}$$

Note that $s(\mathbf{w})$ can be interpreted as a measure of similarity, or conversely, that the similarity measures reviewed in Chapter 2, together with their decision thresholds, can be viewed as tests of a match hypothesis. This suggests the following definition of an image similarity statistic:

Definition 1 *An* image similarity statistic *is defined to be any statistic that forms a test of the match or mismatch hypothesis on an image pair in conjunction with a scalar decision threshold.*

3.3.4 The Hypothesis Tests in Previous Work

Having formulated the matching decision rule in terms of hypothesis tests on the image pair, it is interesting to contrast this approach with the (sometimes implicit) decision rules in previous approaches to the problem of direct image matching. Previous work has either reformulated the matching problem as object detection and applied existing techniques, projected the image-pair into a subspace where modelling was simpler, or considered only the marginal pixel pdfs as the basis of comparison.

Matching as Object Detection Consider the image pair $\{\mathbf{u}, \mathbf{v}\}$. The design of many early matching algorithms proceeded by regarding one image, say $\mathbf{u}$, to be deterministically fixed and formulating matching as the detection problem: $H_1 \iff \mathbf{v} = \mathbf{u} + \boldsymbol{\eta}$ and $H_0 \iff \mathbf{v} = \boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is noise. Although intuitively appealing this formulation is theoretically problematic for a number of reasons. Among them are the facts that problem is not treated symmetrically in $\mathbf{u}$ and $\mathbf{v}$ and the pdf of $\mathbf{u}$ is not part of the formulation. Matching algorithms based on the matched filter and correlation function fall into this category [11, 59, 58, 10].

Matching in a Subspace A very common approach in matching is to project the image pair into a more manageable subspace and proceed from there. The most common example of such a subspace is the difference between the two images, $\mathbf{e} = \mathbf{u} - \mathbf{v}$. The hypotheses can be formulated in terms of the shared pdf $p_{\mathbf{e}}(\mathbf{e}|\boldsymbol{\theta})$, where $H_1 \iff \boldsymbol{\theta} \in \Theta_1$ and $H_0 \iff \boldsymbol{\theta} \in \Theta_0$. A potential problem with this approach is that the subspace sacrifices some information. For example, the joint image pdf, $p_{\mathbf{u},\mathbf{v}}(\mathbf{u}, \mathbf{v})$, is collapsed into the difference image pdf, $p_{\mathbf{e}}(\mathbf{e})$, sacrificing the marginal pdfs, $p_{\mathbf{u}}(\mathbf{u})$ and $p_{\mathbf{v}}(\mathbf{v})$. Examples of this approach include image difference measures [39] and the sign change criteria [42].

Matching by Probability Density Estimates Measures like pairing functions [48] and mutual information [51] base their match and mismatch hypotheses on the estimated joint pdf of corresponding pixels and therefore fully exploit all of the information in individual pixel pairs. However, they view the image as many observations of a univariate pdf rather than a single observation of a multivariate pdf and by doing so the intra-image inter-pixel interactions are ignored. Also, these are nonparametric approaches that do not exploit stochastic *a priori* information. As a result, they do not provide the most powerful tests.

3.4 Scalar Matching Example

It is instructive to reduce the image matching problem to the one of matching scalars (or single pixel images). Although there is very little practical use for a *scalar* matching procedure, the derivation of an optimal test illustrates concepts that are applicable to the more interesting problem of *image* matching, but without the intuitive obstacles inherent in results with high dimensionality.

To give the discussion context, an imaginary application in automatic vehicle navigation is considered: In order to recognize oncoming traffic at night, a system must distinguish between the two headlights of a single oncoming vehicle and two headlights that belong to separate vehicles, using the headlight intensity. The assumption is that variation in headlight intensity is much greater between vehicles. In addition, the sensing of headlight intensity is complicated by weather conditions, which can be modelled as additive noise.

3.4.1 The Optimal Test for Scalar Matching

An optimal test for deciding whether two scalars, which are corrupted by additive noise, are in a state of match or mismatch is derived here.

The Scalar-Pair Model

Consider the two scalars, $u = a + \mu$ and $v = b + \nu$, which have ideal components a and b, and are corrupted by additive noise components μ and ν. The ideal and noise components are all assumed to be normal. The noise components are assumed to be independent of the ideal components and of each other.

The joint pdf of the scalar pair, $p_{u,v}(u, v)$, is now derived for this simple additive noise model. Represent the scalar pair, ideal scalar pair and noise pair as the 2-vectors

$$\mathbf{w} = \begin{bmatrix} u \\ v \end{bmatrix}, \mathbf{c} = \begin{bmatrix} a \\ b \end{bmatrix} \text{ and } \boldsymbol{\eta} = \begin{bmatrix} \mu \\ \nu \end{bmatrix},$$

respectively. The bivariate pdf of the ideal pair can be written as

$$p_{\mathbf{c}}(\mathbf{c}) = \frac{1}{\sqrt{(2\pi)^{n^2} |\mathbf{K}_{\mathbf{c}}|}} \exp\left[-\frac{1}{2}(\mathbf{c} - \mathbf{m}_{\mathbf{c}})^T \mathbf{K}_{\mathbf{c}}^{-1} (\mathbf{c} - \mathbf{m}_{\mathbf{c}})\right], \tag{3.4}$$

which is characterized completely by the mean vector

$$\mathbf{m}_{\mathbf{c}} = \begin{bmatrix} m_a \\ m_b \end{bmatrix}$$

and the covariance matrix

$$\mathbf{K}_{\mathbf{c}} = \begin{bmatrix} \sigma_a^2 & \rho \cdot \sigma_a \sigma_b \\ \rho \cdot \sigma_a \sigma_b & \sigma_b^2 \end{bmatrix}.$$

Hence the popular abbreviated notation: $p_{\mathbf{c}}(\mathbf{c}) = N(\mathbf{c}; \mathbf{m}_{\mathbf{c}}, \mathbf{K}_{\mathbf{c}})$[1]. The quantity, $\rho \in [-1, 1]$, is the correlation coefficient between the ideal components. Assuming that a and b are generated by identical random processes, $m = m_a = m_b$ and $\sigma^2 = \sigma_a^2 = \sigma_b^2$, are defined. For the noise components,

$$p_{\boldsymbol{\eta}}(\boldsymbol{\eta}) = N(\boldsymbol{\eta}; \mathbf{0}, \mathbf{K}_{\boldsymbol{\eta}})$$

[1]Strictly speaking, the notation $N(\mathbf{m}, \mathbf{K})$ is commonly used to represent the multivariate normal *distribution* with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{K}$, but the presentation here will also use the notation $N(\mathbf{x}; \mathbf{m}, \mathbf{K})$ to represent the associated normal *probability density function*, where $\mathbf{x}$ is the independent variable.

where

$$\mathbf{K}_{\boldsymbol{\eta}} = \begin{bmatrix} \sigma_\mu^2 & 0 \\ 0 & \sigma_\nu^2 \end{bmatrix}.$$

The required pdf can now be written as

$$\begin{aligned} p_{u,v}(u,v|\boldsymbol{\theta}) &= p_{\mathbf{w}}(\mathbf{w}) \\ &= N(\mathbf{w};\mathbf{m_c},\mathbf{K_c}+\mathbf{K}_{\boldsymbol{\eta}}) \quad \text{[see Appendix B.1]} \\ &= k(\boldsymbol{\theta}) \cdot \exp\left[-\frac{1}{2}\cdot f(u,v,\boldsymbol{\theta})\right], \end{aligned} \tag{3.5}$$

where

$$\boldsymbol{\theta} = \{m,\sigma,\sigma_\mu,\sigma_\nu,\rho\},$$

$$k(\boldsymbol{\theta}) = \frac{1}{2\pi\sqrt{\sigma^4(1-\rho^2)+\sigma^2\sigma_\nu^2+\sigma_\mu^2\sigma^2+\sigma_\mu^2\sigma_\nu^2}}$$

and

$$f(u,v,\boldsymbol{\theta}) = \frac{(u-m)^2\left(\sigma^2+\sigma_\nu^2\right)-2(u-m)(v-m)\rho\sigma^2+(v-m)^2\left(\sigma^2+\sigma_\mu^2\right)}{\sigma^4(1-\rho^2)+\sigma^2\sigma_\nu^2+\sigma_\mu^2\sigma^2+\sigma_\mu^2\sigma_\nu^2}.$$

Hypothesis Models

In order to model the hypotheses it is necessary to model match and mismatch in the joint scalar pdf (3.5). This can be done by a process of elimination, considering the available parameters,

$$\boldsymbol{\theta} = \{m,\sigma,\sigma_\mu,\sigma_\nu,\rho\}$$

and the fact that any parameters determining match or mismatch must involve the relationship between the ideal components a and b. The noise is independent of a and b, so $\{\sigma_\mu,\sigma_\nu\}$ can be discarded. Since $\{m,\sigma\}$ are parameters in the marginal pdfs of a and b, they cannot affect any relationship between them and they too can be discarded. The remaining parameter is ρ, the correlation coefficient between a and b, which does indeed control the relationship between them.

Assuming that the other parameters are known, one possible model has the hypotheses sharing a pdf $p_{u,v}(u, v|\rho)$ that is parameterized on the correlation coefficient, where the hypotheses are defined by $H_1 \iff \rho = 1$ and $H_0 \iff \rho = 0$. For $\rho = 1$ the random vector $\mathbf{c}$ degenerates to a single random variable and $a = b$ in all scalar pairs. For $\rho = 0$, a and b are statistically independent. The probability of c having $a = b$ is

$$\begin{aligned} P(a = b|\rho = 0) &= \int_{a=b} p_{a,b}(a, b|\rho = 0)\, da\, db \\ &= 0, \end{aligned}$$

since the integration is over a set of measure zero. In practice the values are digitized and $P(a = b|\rho = 0)$ becomes a finite probability, but this is a consequence of the proposed model that is consistent with the matching problem. For the vehicle navigation scenario, this models the unlikely, but plausible situation of two oncoming vehicles having identical headlight intensity within the limitations of the measuring device. Notice that the mismatch hypothesis does not guarantee that $a \neq b$, but the match hypothesis does guarantee that $a = b$. This restriction is removed next.

For values of ρ, ρ_0 and ρ_1, where $0 \leq \rho_0 < \rho_1 < 1$, samples of the ideal components will be, on average, more similar for $\rho = \rho_1$ than for $\rho = \rho_0$. A more general hypothesis model, then, has $H_1 \iff \rho = \rho_1$ and $H_0 \iff \rho = \rho_0$. As before, the mismatch hypothesis does not guarantee that $a \neq b$, but now the match hypothesis does not guarantee that $a = b$ either. Returning to the vehicle navigation problem, this model allows some deviation between the intensities of a single vehicle's headlights. The model just ensures that on the whole, scalars that represent headlights of the same vehicle are closer together than scalars that represent headlights from different vehicles.

Note that in terms of the definition in Section 3.3.2, this is a *simple match* hypothesis model. Figure 3-2 shows an example of the match and mismatch hypothesis pdfs.

Deriving the Test

Assuming that the *a priori* probabilities of match, P_1, and mismatch, P_0, are known, the ideal observer test

$$\text{Accept} \begin{cases} H_1 \text{ if } l(u, v) > \lambda \\ H_0 \text{ if } l(u, v) \leq \lambda \end{cases}, \tag{3.6}$$

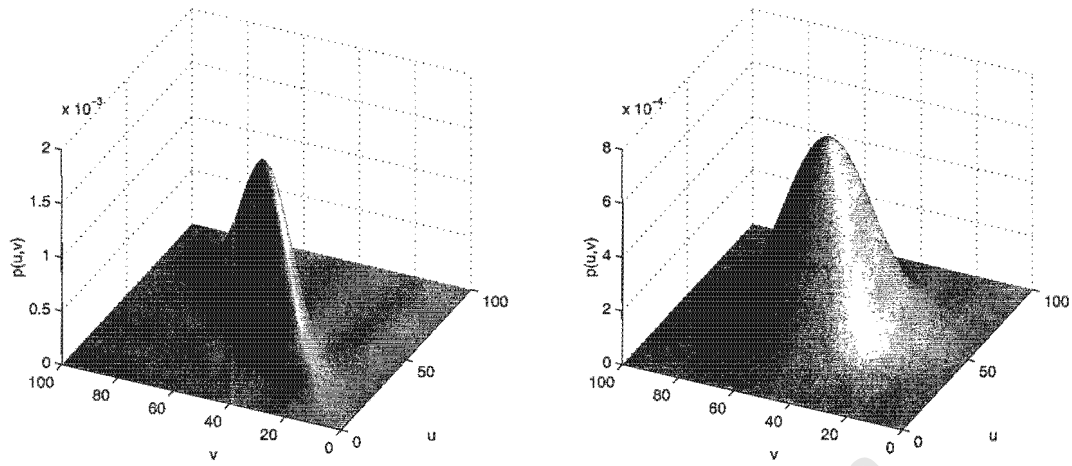


Figure 3-2: Hypothesis conditional pdf models ($m = 50$, $\sigma^2 = 200$, $\sigma_\mu^2 = \sigma_\nu^2 = 25$. $H_1 : \rho = 1$ (left) and $H_0 : \rho = 0$ (right)).

where

$$\lambda = \frac{P_0}{P_1}, \tag{3.7}$$

is optimal for testing two simple hypotheses. A more convenient notation for (3.6) is now introduced for the test (3.6):

$$l(u,v) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda.$$

Substituting the pdfs of the previous section into the likelihood ratio gives

$$\begin{aligned} l(u,v) &= \frac{p_{u,v}(u,v|\rho=\rho_1)}{p_{u,v}(u,v|\rho=\rho_0)} \\ &= \frac{k(\rho_1)}{k(\rho_0)} \exp\left[\frac{f(u,v,\rho_0) - f(u,v,\rho_1)}{2}\right] \end{aligned}$$

Taking the (strictly monotonic) logarithm of both sides and re-arranging the test,

$$f(u,v,\rho_0) - f(u,v,\rho_1) \underset{H_0}{\overset{H_1}{\gtrless}} 2\log\left[\frac{P_0}{P_1}\sqrt{\frac{\phi(\rho_0)}{\phi(\rho_1)}}\right],$$

where

$$\phi(\rho) = \left[\sigma^4\left(1-\rho^2\right) + \sigma^2\sigma_\nu^2 + \sigma_\mu^2\sigma^2 + \sigma_\mu^2\sigma_\nu^2\right]^{-1}$$

The test can now be written as

$$s(u,v) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$$

with the statistic

$$\begin{aligned} s(u,v) &= f(u,v,\rho_0) - f(u,v,\rho_1) \\ &= A(u-m)^2 + B(v-m)^2 + C(u-m)(v-m) \end{aligned}$$

and decision threshold

$$\lambda = 2\log\left[\frac{P_0}{P_1}\sqrt{\frac{\phi(\rho_0)}{\phi(\rho_1)}}\right],$$

where

$$\begin{aligned} A &= [\phi(\rho_0) - \phi(\rho_1)] \cdot \left(\sigma^2 + \sigma_\nu^2\right) \\ B &= [\phi(\rho_0) - \phi(\rho_1)] \cdot \left(\sigma^2 + \sigma_\mu^2\right) \\ C &= -[\rho_0\phi(\rho_0) - \rho_1\phi(\rho_1)] \cdot 2\sigma^2. \end{aligned}$$

Several special cases are now considered.

Special Case 1 Shared noise model ($\sigma_\mu^2 = \sigma_\nu^2$):

$$s(u,v) = A\left[(u-m)^2 + (v-m)^2\right] + C(u-m)(v-m)$$

$$\lambda = 2\log\left[\frac{P_0}{P_1}\sqrt{\frac{\phi(\rho_0)}{\phi(\rho_1)}}\right]$$

Special Case 2 Different noise models ($\sigma_\mu^2 \neq \sigma_\nu^2$), $\rho_1 = 1$ and $\rho_0 = 0$:

$$s(u,v) = \frac{\sigma_\mu^2 (u-m)^2}{(\sigma^2+\sigma_\mu^2)} + \frac{\sigma_\nu^2 (v-m)^2}{(\sigma^2+\sigma_\nu^2)} - (u-v)^2$$

$$\lambda = 2\frac{\sigma_\nu^2\sigma^2 + \sigma_\mu^2\sigma^2 + \sigma_\mu^2\sigma_\nu^2}{\sigma^2} \log\left[\frac{P_0}{P_1}\sqrt{\frac{\sigma_\nu^2\sigma^2 + \sigma_\mu^2\sigma^2 + \sigma_\mu^2\sigma_\nu^2}{(\sigma^2+\sigma_\mu^2)(\sigma^2+\sigma_\nu^2)}}\right]$$

Special Case 3 Shared noise model ($\sigma_\mu^2 = \sigma_\nu^2$), $\rho_1 = 1$ and $\rho_0 = 0$:

$$s(u,v) = \frac{\sigma_\nu^2}{(\sigma^2+\sigma_\nu^2)}\left[(u-m)^2 + (v-m)^2\right] - (u-v)^2 \tag{3.8}$$

$$\lambda = 2\frac{\sigma_\nu^2(2\sigma^2+\sigma_\nu^2)}{\sigma^2} \log\left[\frac{P_0}{P_1}\frac{\sqrt{\sigma_\nu^2(2\sigma^2+\sigma_\nu^2)}}{\sigma^2+\sigma_\nu^2}\right] \tag{3.9}$$

3.4.2 Analysis of the Optimal Test

Scalar matching offers the opportunity of analyzing the proposed hypothesis testing approach to matching without the high dimensionality of images. This may give some insight into the more complex tests for image matching that are investigated later.

The Test Statistic

Note that the statistic (3.8) in special case 3 of the previous section is written with two main terms. The second of these is the negated squared difference between the two measured scalars, $-(u-v)^2$. It is intuitively pleasing to have this term in the optimal test, since it corresponds to the use of the squared difference as a measure of dissimilarity. The sum of squared differences, or mean squared error, between corresponding pixels is a standard measure for comparing two images for matching or for measuring image fidelity.

The first term,

$$\frac{\sigma_\nu^2}{(\sigma^2+\sigma_\nu^2)}\left[(u-m)^2 + (v-m)^2\right],$$

effectively increases the statistic with decreasing likelihood of measuring the scalars u and v. The fact that two scalars are close together holds more information about whether they match or not if the individual scalars are unlikely. To illustrate this point with the night-time

navigation example, consider the situation where two headlights are detected with intensities near the average of modern vehicles and difference δ. Compare this with the situation where two headlights are detected with intensities close to the average of a Model T Ford, but also having difference δ. Since it is so unlikely that two classic motorcars are approaching it seems reasonable to decide that they belong to the same vehicle. In the first situation, on the other hand, it wouldn't be unlikely at all that two modern cars were approaching, and a decision that they belong to the same vehicle is more risky.

Signal to Noise Ratio

Here the SNR of the scalar is defined as the ratio of the ideal component standard deviation σ to the noise component standard deviation σ_ν, that is

$$\text{SNR} \equiv \frac{\sigma}{\sigma_\nu}.$$

The statistic (3.8) can then be rewritten in terms of the SNR as

$$\begin{aligned} s(u,v) &= \frac{\sigma_\nu^2}{(\sigma^2+\sigma_\nu^2)}\left[(u-m)^2+(v-m)^2\right]-(u-v)^2 \\ &= \frac{1}{\text{SNR}^2+1}\left[(u-m)^2+(v-m)^2\right]-(u-v)^2. \end{aligned}$$

The effect of the SNR can now be investigated. If the noise is negligible (i.e. SNR $\to \infty$) then

$$s(u,v) \to -(u-v)^2$$

and the optimal test reduces to the squared difference. In the extreme situation of the signal being drowned in noise (i.e. SNR ≈ 0)

$$s(u,v) \approx \left[(u-m)^2+(v-m)^2\right]-(u-v)^2,$$

suggesting that the likelihood term has increasing significance as noise levels increase. Notice, however, that the squared difference is significant for all SNR values. The squared difference, the likelihood term and the overall test statistic are plotted against different SNR values in Figure 3-3.

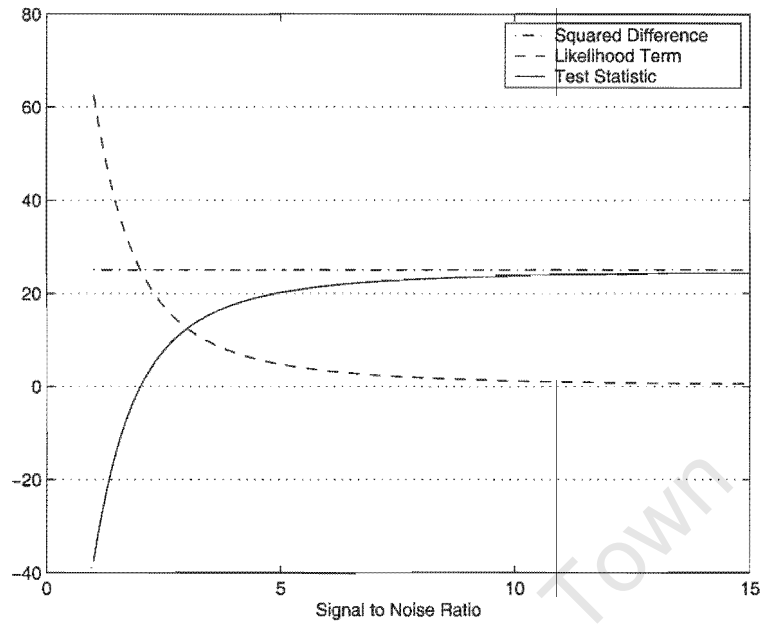


Figure 3-3: Relative importance of the squared difference and likelihood term as a function of SNR.

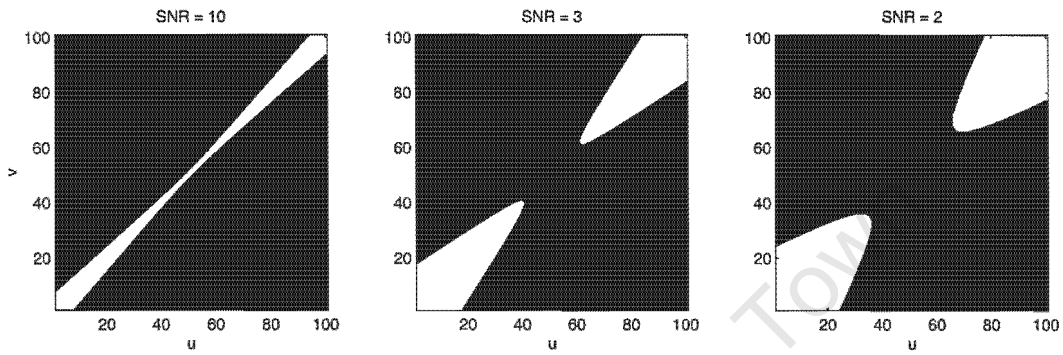
This relationship between SNR and the form of the test supports the frequent use of the squared difference test, or an equivalent measure, in many computer vision applications where the sensor noise can be neglected and it is the complexity of the scene that provides the challenge. However, applications such as low dose medical imaging that require the maximum attainable image processing performance under demanding SNR conditions might benefit from an optimal matching algorithm.

The Match/Mismatch Partition

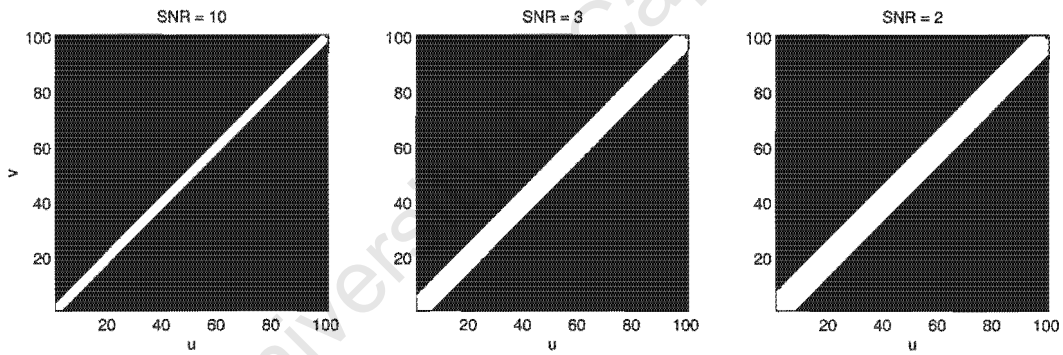
Figure 3-4(a) shows the match/mismatch partition induced on (u, v) by the ideal observer test of (3.8) and (3.9). The white area is the critical region, where the match hypothesis for scalar pairs is accepted. Notice that for some SNR levels there are values of u for which the scalar-pair will always be classified as mismatch, regardless of v, and vice versa. The partition induced by a test based on the squared difference term is shown in 3-4(b) for comparison. Note that the optimal ideal observer threshold, which is derived in Appendix B.2 for the squared difference statistic, is used here too.

The obvious implication of Figure 3-4(a) when extending the optimal test from scalars to images is that some images need not be tested for match with others. In an image registration

application using block matching, for example, one might exclude image blocks on this basis. The improbable image blocks, being the best candidates for matching, would be retained. This approach is similar to the use of *interest operators* [74] for block selection in image registration algorithms.



(a) Bayes test. $P_1 = 0.2$.



(b) Squared difference test. $P_1 = 0.5$.

Figure 3-4: The partition induced on (u, v) by scalar hypothesis tests.

The SNR has a lower bound for the situation where all scalars potentially have matching counterparts. Since the most likely scalars in the aforementioned models have maximum probability at their mean, this condition is satisfied if

$$s(m, m) > \lambda,$$

or, since $s(m, m) = 0$, an equivalent condition is $\lambda < 0$. Substituting (3.9), this condition

implies that

$$0 < \frac{P_0}{P_1} \frac{\sqrt{\sigma_\nu^2 \left(2\sigma^2 + \sigma_\nu^2\right)}}{\sigma^2 + \sigma_\nu^2} < 1$$

or, rearranging the inequality and writing it in terms of SNR, it implies that

$$0 < \frac{\sqrt{2 \cdot \mathrm{SNR}^2 + 1}}{\mathrm{SNR}^2 + 1} < \frac{P_1}{P_0}.$$

This inequality can be used to find the lower bound for SNR as a function of P_1 that guarantees potentially matching counterparts for all scalars. Figure 3-5 plots this lower bound and shows that if $P_1 > \frac{1}{2}$, then all scalars potentially have matching counterparts, regardless of SNR.

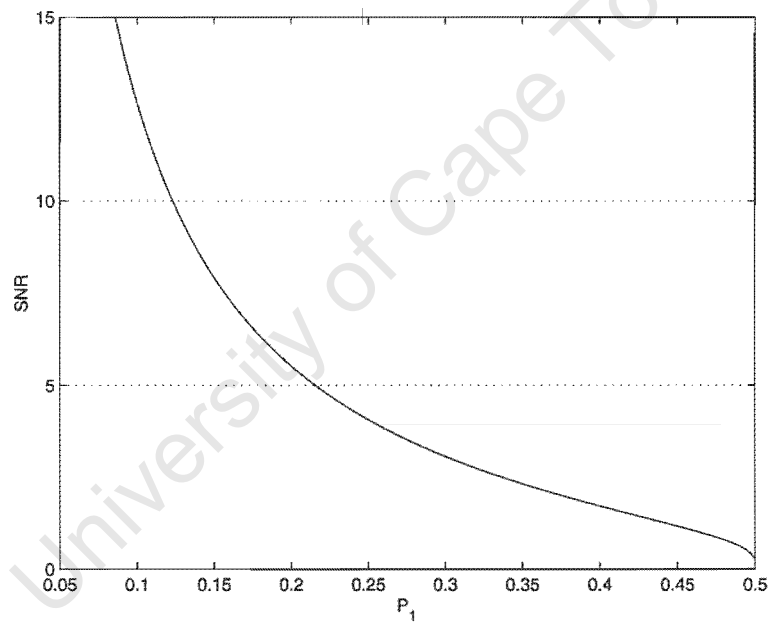


Figure 3-5: A lower bound for scalar SNR that guarantees potential matching counterparts.

3.5 Discussion

Having considered the theoretical machinery available for addressing decision-making problems, hypothesis testing has been selected as a suitable framework for the design of an image matching decision rule. With the image pair as unit of observation, the essential elements of the proposed formulation are separate stochastic models for the image pair under the

match and mismatch hypotheses. The approach has been illustrated with the simple example of scalar matching, which has very little practical significance, but reveals some interesting aspects of the proposed formulation via the simplicity of only two dimensions.

Application of the hypothesis testing approach to image matching is now required in order to establish whether it will satisfy the requirements set out for a matching algorithm in Section 3.3. First, however, attention is given to the aspect of image-pair modelling. Good models are crucial for deriving effective algorithms, and image synthesis based on the models will make it possible to do Monte Carlo experiments where an analytical approach is intractable.

Chapter 4

Modelling and Synthesis of Image Pairs

The proposed approach to the image matching problem has its subjectivity in the selection of the image-pair model. Aside from any compromises that are made for reasons of tractability of derivation, or practicality of implementation, the derivation of the optimal test is a mechanistic mathematical procedure. The model, on the other hand, is *selected* by the designer, where the degree of subjectivity in the selection is determined by the quantity and detail of prior knowledge about the physical phenomenon under consideration. The quality of this selection is important, because the test is derived from first principles and is optimal with respect to the model and appropriate performance criteria. The accuracy of the model, therefore, determines how close the test is to optimal for the actual problem.

The first subjective decision made here is the one to model the image pair as a random process. Stochastic models sometimes describe a process that is inherently random in nature, but more often they provide a way of developing a model when knowledge of the underlying phenomenon is incomplete or when the phenomenon is too complex to specify deterministically. For the matching applications considered here, specific information about the content of the image at any given time is assumed to be unknown and therefore the option of deterministic modelling is rejected. As Hunt and Cannon put it: "The world, as captured in an image, is so complex that complete *a priori* deterministic and mechanistic models seem out of reach" [75]. The stochastic modelling approach is common in the fields of image restoration [76, 77, 78], texture analysis [79, 80, 81], image compression [82], and target detection [83, 84], but has been applied less rigorously to the task of image matching.

This chapter develops a stochastic image-pair model that will be used as the basis for deriving the hypothesis test for image matching. Section 4.1 outlines the modelling assumptions that are made. Section 4.2 develops correlation-based and difference-based models for match and mismatch in the image pair. Synthetic images will be required for testing purposes, and Section 4.3 develops methods for generating random image pairs. Section 4.4 concludes the chapter with a discussion.

4.1 General Model Assumptions

A stochastic model expresses the characteristics of an image pair in terms of the probabilities associated with observing every possible image combination, but there are many different models to choose from. Kazakos and Papantoni-Kazakos describe the "best" model as the "simplest existing model that describes the phenomenon with satisfactory accuracy, with the emphasis on simplicity" [64, p. 1]. General assumptions that provide this simplicity and narrow the field of potential models for the image pair are provided here.

4.1.1 Stationary, Multivariate Normal Images

The requirement of simplicity for a stochastic model is normally embodied in the assumption that a multivariate normal (MVN) distribution (or equivalently, a Gaussian random field) describes observations of the phenomenon under consideration. As Muirhead concedes, analysis with other probability distributions is rarely tractable [29, p. 1]. The MVN pdf for $n \times n$ image $\mathbf{a}$ is given by

$$p_{\mathbf{a}}(\mathbf{a}) = \frac{1}{\sqrt{(2\pi)^{n^2} |\mathbf{K}_{\mathbf{a}}|}} \exp\left[-\frac{1}{2}(\mathbf{a}-\mathbf{m}_{\mathbf{a}})^T \mathbf{K}_{\mathbf{a}}^{-1}(\mathbf{a}-\mathbf{m}_{\mathbf{a}})\right],$$

or, using the abbreviated notation, by $p_{\mathbf{a}}(\mathbf{a}) = N(\mathbf{a}; \mathbf{m}_{\mathbf{a}}, \mathbf{K}_{\mathbf{a}})$.

In image processing models, the additional assumption of ergodicity (and therefore spatial stationarity) leads to easier analysis and more efficient algorithms. However, a cursory analysis of typical images with natural or man-made scene content reveals that these assumptions are unrealistic in general. Common violations are shown in Figure 4-1 — the pixel intensity values are positive, their histograms are asymmetrical and have multiple modes, and spatial averaging suggests that the stationarity assumption is questionable. Where the MVN and stationarity assumptions cannot be made, solutions can be found in ad-hoc (nonparametric)

(a) Medical radiograph - 'Skull'.

(b) Standard image - 'LAX' (subimage).

(c) Standard image - 'Lena' (subimage).

Figure 4-1: Local averages and histograms for three test images. The histograms are accompanied by the best fit normal pdf.

approaches or in more sophisticated models. As examples of the latter, multiresolution models [85, 86], mixture distributions [87], or generalized Gaussian models [78] have been reported. Alternatively, the images can be transformed so that they better resemble samples from a stationary MVN process. for example, Hunt and Cannon propose a model with additive nonstationary mean and stationary residual components [75], Hunt proposes normalization and spatial warping to enforce stationarity in the second order image statistics [88], and Chapple and Bertilone propose a pointwise transform to make image pixel statistics better resemble the normal distribution [89]. These methods can potentially overcome the non-normal and nonstationary characteristics of images, and Appendix A investigates them in more detail.

For the purposes of this research, then, the assumption is made that images can either be adequately modelled as an MVN process or they can be transformed to better resemble the samples of one. Normal marginal pdfs do not guarantee a normal joint pdf [29, p. 7], so the fact that MVN models are adequate for the individual images does not imply that the same is true for the image pair. Even so, for the tractability it offers, the additional assumption is made that a linear model adequately represents the match/mismatch relationship between the images. The resulting pdf for the image pair $\mathbf{w}^T = \left[\mathbf{a}^T, \mathbf{b}^T\right]$ is given by

$$p_{\mathbf{w}}(\mathbf{w}) = \frac{1}{\sqrt{(2\pi)^{2n^2}\,|\mathbf{K}_{\mathbf{w}}|}} \exp\left[-\frac{1}{2}(\mathbf{w}-\mathbf{m}_{\mathbf{w}})^T \mathbf{K}_{\mathbf{w}}^{-1}(\mathbf{w}-\mathbf{m}_{\mathbf{w}})\right], \tag{4.1}$$

where $\mathbf{m}_{\mathbf{w}}^T = \left[\mathbf{m}_{\mathbf{a}}^T, \mathbf{m}_{\mathbf{b}}^T\right]$ is simply a concatenation of the mean vectors for the individual images. $\mathbf{K}_{\mathbf{w}}$ is the joint image-pair covariance matrix, which can be written as

$$\mathbf{K}_{\mathbf{w}} = \begin{bmatrix} \sigma_a^2\mathbf{R}_{\mathbf{a}} & \sigma_a\sigma_b\mathbf{R}_{\mathbf{ab}} \\ \sigma_a\sigma_b\mathbf{R}_{\mathbf{ab}} & \sigma_b^2\mathbf{R}_{\mathbf{b}} \end{bmatrix},$$

where $\mathbf{K}_{\mathbf{a}} = \sigma_a^2\mathbf{R}_{\mathbf{a}}$ and $\mathbf{K}_{\mathbf{b}} = \sigma_b^2\mathbf{R}_{\mathbf{b}}$ are the covariance matrices of the individual images, and $\sigma_a\sigma_b\mathbf{R}_{\mathbf{ab}}$ is their cross-covariance matrix.

4.1.2 Shared Intra-Image Correlation Structure

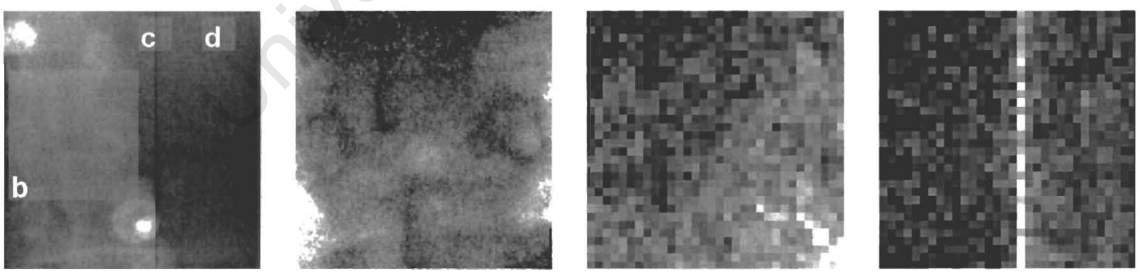
It is assumed that the images $\mathbf{a}$ and $\mathbf{b}$ share the same intra-image correlation structure, and therefore $\mathbf{R}_{\mathbf{a}} = \mathbf{R}_{\mathbf{b}} = \mathbf{R}$. In a matching application the two images will probably contain the same sort of subject matter, making this a reasonable assumption in most cases. Applications that require multi-modal matching are possible exceptions, although it should be noted that

the model still allows the images to differ by a systematic offset (mean vectors $\mathbf{m_a}$ and $\mathbf{m_b}$) and overall scale (variances σ_a^2 and σ_b^2). The image-pair covariance matrix can now be written as

$$\mathbf{K_w} = \begin{bmatrix} \sigma_a^2\mathbf{R} & \sigma_a\sigma_b\mathbf{R_{ab}} \\ \sigma_a\sigma_b\mathbf{R_{ab}} & \sigma_b^2\mathbf{R} \end{bmatrix}. \tag{4.2}$$

4.1.3 Additive Noise

The sensed image has two main components. First, there is information about the scene and second, there is superfluous information that was added during the generation of radiation, the irradiation of the scene and the image capture. This additional information is commonly referred to as noise. Figure 4-2 illustrates the distinction made between scene information and noise in a medical X-ray image: subfigure (a) is the original image, (b) highlights scene information in the form of the vertebrae, (c) highlights statistical noise in a quiet part of the image and (d) shows a structure noise artifact introduced by the line-scan operation of the imaging system. For now it is assumed that image formation artifacts are either absent, or that they can be removed by preprocessing that exploits their deterministic structure. Assumptions must now be made regarding the nature of the statistical noise.



(a) Medical X-ray image.

(b) Scene: vertebrae.

(c) Statistical noise.

(d) Structure artifact.

Figure 4-2: Image model components.

A common assumption is that the noise in an imaging system can be modelled in terms of additive (signal independent) and multiplicative (signal dependent) components [8, p. 268],

implying the model

$$\mathbf{u} = \boldsymbol{\mu}_{\times} \odot \mathbf{a} + \boldsymbol{\mu}_{+} \tag{4.3}$$

where $\odot$ denotes the Hadamard product[1], $\mathbf{a}$ represents the scene, $\boldsymbol{\mu}_{\times}$ represents multiplicative noise and $\boldsymbol{\mu}_{+}$ represents additive noise. In the event that one of the noise components dominates the other, the model can be further simplified by neglecting the smaller component. In some situations, purely multiplicative noise can be made additive by taking the logarithm of the image intensity values [17, p. 80], suggesting that the additive noise model, $\mathbf{u} = \mathbf{a} + \boldsymbol{\mu}$, can describe a wide range of imaging scenarios.

Further, it is assumed that the noise can be modelled as a zero-mean, stationary MVN random process. Since most imaging systems accumulate signal at various stages of the image formation process (e.g. scintillation, CCD camera integration and software summing of image frames), the normal approximation can be justified at each stage using the central limit theorem of statistics. Denoting the noise components of images $\mathbf{u}$ and $\mathbf{v}$ as $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ respectively, their pdfs are given by $p_{\boldsymbol{\mu}}(\boldsymbol{\mu}) = N(\boldsymbol{\mu}; \mathbf{0}, \mathbf{K}_{\boldsymbol{\mu}})$ and $p_{\boldsymbol{\nu}}(\boldsymbol{\nu}) = N(\boldsymbol{\nu}; \mathbf{0}, \mathbf{K}_{\boldsymbol{\nu}})$. If the noise is white, then $\mathbf{K}_{\boldsymbol{\mu}} = \sigma_{\mu}^2 \mathbf{I}$ and $\mathbf{K}_{\boldsymbol{\nu}} = \sigma_{\nu}^2 \mathbf{I}$.

Based on the rules governing sums of multivariate normal random vectors, the covariance matrix of $\mathbf{w}$ is the sum of the covariance matrices of the scene and noise components (see Mood, Graybill and Boes [61, p. 178] and Appendix B.1). Therefore, assuming that the noise in separate images is statistically independent, $\mathbf{K}_{\boldsymbol{\mu}\boldsymbol{\nu}} = \mathbf{0}$, and

$$\begin{aligned} \mathbf{K}_{\mathbf{w}} &= \begin{bmatrix} \sigma_a^2 \mathbf{R} & \sigma_a \sigma_b \mathbf{R}_{\mathbf{ab}} \\ \sigma_a \sigma_b \mathbf{R}_{\mathbf{ab}} & \sigma_b^2 \mathbf{R} \end{bmatrix} + \begin{bmatrix} \sigma_{\mu}^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma_{\nu}^2 \mathbf{I} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_a^2 \mathbf{R} + \sigma_{\mu}^2 \mathbf{I} & \sigma_a \sigma_b \mathbf{R}_{\mathbf{ab}} \\ \sigma_a \sigma_b \mathbf{R}_{\mathbf{ab}} & \sigma_b^2 \mathbf{R} + \sigma_{\nu}^2 \mathbf{I} \end{bmatrix}. \end{aligned} \tag{4.4}$$

4.2 Models for Match and Mismatch

The model for the image pair is characterized by the mean vector $\mathbf{m}_{\mathbf{w}}$ and the covariance matrix $\mathbf{K}_{\mathbf{w}}$ in equation (4.4). All of the quantities in the covariance matrix are characteristics of the individual images except for the normalized cross-covariance matrix $\mathbf{R}_{\mathbf{ab}}$. Since this matrix governs the relationship between images $\mathbf{u}$ and $\mathbf{v}$, it will be instrumental in defining

[1] $\mathbf{a} = \mathbf{b} \odot \mathbf{c} \Longrightarrow a_i = b_i \cdot c_i \ \forall i$

match and mismatch for the image pair. This section takes two approaches to the problem of defining meaningful structure for $\mathbf{R_{ab}}$ in the situations of match and mismatch.

4.2.1 Correlation-Based Model

This model bases match and mismatch on the correlation coefficient between the scene pixels of each image[2]. It makes the assumption that all corresponding pixel-pairs share a correlation coefficient ρ_{ab}, the value of which is specified differently in the separate models for match and mismatch. For guaranteed identical[3] images $\rho_{ab} = 1$, for statistically independent images $\rho_{ab} = 0$, and for values in-between the images exhibit varying degrees of correlation between corresponding pixels. The match and mismatch hypotheses can be defined as $H_1 \iff \rho_{ab} = \rho_1$ and $H_0 \iff \rho_{ab} = \rho_0$, respectively, where $0 \leq \rho_0 < \rho_1 \leq 1$. The normalized cross-covariance matrix is now given by

$$\mathbf{R_{ab}} = \begin{bmatrix} \rho_{ab} & & ? \\ & \rho_{ab} & \\ & & \ddots \\ ? & & \rho_{ab} \end{bmatrix},$$

where the off-diagonal elements are as yet unspecified. The following constraints limit the form of the matrix:

1. For $\rho_0 = 0$, mismatching images are statistically independent and $\mathbf{R_{ab}}\big|_{\rho_{ab}=0} = \mathbf{0}$.
2. For $\rho_1 = 1$, matching images are identical within a scaling factor and $\mathbf{R_{ab}}\big|_{\rho_{ab}=1} = \mathbf{R}$.

One valid structure for $\mathbf{R_{ab}}$ with these constraints is $\mathbf{R_{ab}} = \rho_{ab}\mathbf{R}$. This is not a unique solution[4], but its simplicity is appealing. Adopting it for the image-pair model, the joint covariance matrix for the correlation-based model becomes

$$\mathbf{K_w} = \begin{bmatrix} \sigma_a^2\mathbf{R}+\sigma_\mu^2\mathbf{I} & \sigma_a\sigma_b\rho_{ab}\mathbf{R} \\ \sigma_a\sigma_b\rho_{ab}\mathbf{R} & \sigma_b^2\mathbf{R}+\sigma_\nu^2\mathbf{I} \end{bmatrix}. \tag{4.5}$$

[2]Note that it is not the correlation coefficient between the pixels of the sensed images **u** and **v** (an estimate of which is often used as a measure of similarity between images) that is of interest here, but rather the correlation coefficient between the *scene* pixels of **a** and **b**.

[3]Identical, that is, to within the systematic offset and scaling factor determined by the respective image mean vectors and variances.

[4]Another has elements $\rho_k\mathbf{R}[i,j]^{\rho_k}$, where $\mathbf{R}[i,j]$ is the element of $\mathbf{R}$ at (i,j).

4.2.2 Difference-Based Model

The second approach uses a scene difference image to model the inter-image relationship. It is based on a hypothetical process that generates images with covariance matrices $\mathbf{K_a}$ and $\mathbf{K_b}$, mean vectors $\mathbf{m_a}$ and $\mathbf{m_b}$, and with control over an image difference parameter, δ_{ab}. The process generates two zero-mean, unit-variance images $\mathbf{a}_1$ and $\mathbf{b}_1$, that differ by a random image $\mathbf{d}_1$ with zero mean and standard deviation δ_{ab}. For guaranteed identical images, $\delta_{ab} = 0$, for statistically independent images, $\delta_{ab} = 1$, and for values in-between the images have differences of varying magnitudes. After the process has generated $\mathbf{a}_1$ and $\mathbf{b}_1$, it scales and translates their pixel intensities in order to produce images with the required mean vector and variance. For the matching problem there will be two processes: one that generates matching images, where $H_1 \iff \delta_{ab} = \delta_1$ and one that generates mismatching images, where $H_0 \iff \delta_{ab} = \delta_0$. Note that $0 \leq \delta_1 < \delta_0 \leq 1$. Figure 4-3 shows a schematic of the process. The cross-covariance matrix it implies will now be derived.

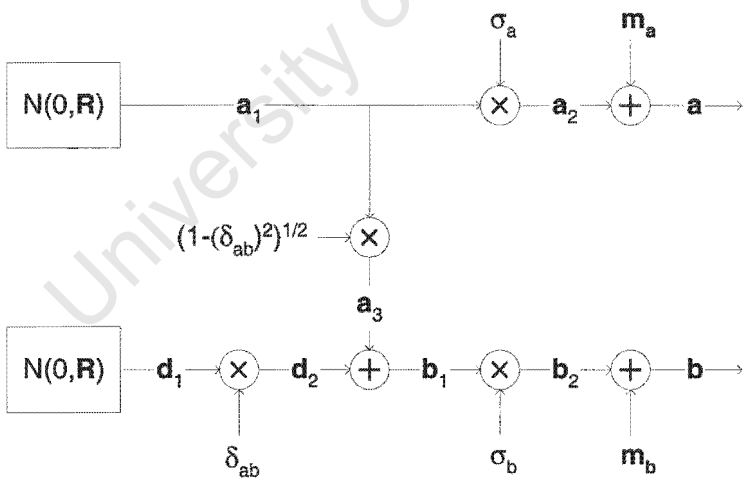


Figure 4-3: Process for generating an image pair with control over a difference parameter.

Independent root images $\mathbf{a}_1$ and $\mathbf{d}_1$ are generated using a MVN random process with zero mean and covariance matrix $\mathbf{R}$. The output images can be written in terms of the root images

as $\mathbf{a} = \sigma_a \mathbf{a}_1 + \mathbf{m_a}$ and

$$\begin{aligned}\mathbf{b} &= \sigma_b \mathbf{b}_1 + \mathbf{m_b} \\ &= \sigma_b (\mathbf{a}_3 + \mathbf{d}_2) + \mathbf{m_b} \\ &= \sigma_b \left(\delta_{ab} \mathbf{d}_1 + \sqrt{1 - \delta_{ab}^2} \mathbf{a}_1 \right) + \mathbf{m_b}.\end{aligned}$$

First, it is proved that the process in Figure 4-3 does indeed produce images with the required marginal covariance matrices. For the covariance matrix of $\mathbf{a}$:

$$\begin{aligned}\text{Var}[\mathbf{a}] &= E\left[(\mathbf{a} - \mathbf{m_a})(\mathbf{a} - \mathbf{m_a})^T \right] \\ &= \sigma_a^2 E\left[\mathbf{a}_1 \mathbf{a}_1^T \right] \\ &= \sigma_a^2 \mathbf{R},\end{aligned}$$

as required. For the covariance matrix of $\mathbf{b}$:

$$\begin{aligned}\text{Var}[\mathbf{b}] &= E\left[(\mathbf{b} - \mathbf{m_b})(\mathbf{b} - \mathbf{m_b})^T \right] \\ &= \sigma_b^2 E\left[\left(\delta_{ab} \mathbf{d}_1 + \sqrt{1 - \delta_{ab}^2} \mathbf{a}_1 \right) \left(\delta_{ab} \mathbf{d}_1 + \sqrt{1 - \delta_{ab}^2} \mathbf{a}_1 \right)^T \right] \\ &= \sigma_b^2 \delta_{ab}^2 E\left[\mathbf{d}_1 \mathbf{d}_1^T \right] + \sigma_b^2 \left(1 - \delta_{ab}^2 \right) E\left[\mathbf{a}_1 \mathbf{a}_1^T \right] + 2\sigma_b^2 \delta_{ab} \sqrt{1 - \delta_{ab}^2} E\left[\mathbf{a}_1 \mathbf{d}_1^T \right].\end{aligned}$$

Now $E\left[\mathbf{d}_1 \mathbf{d}_1^T \right] = E\left[\mathbf{a}_1 \mathbf{a}_1^T \right] = \mathbf{R}$, and $\mathbf{a}_1$ and $\mathbf{d}_1$ are independent by design, so $E\left[\mathbf{a}_1 \mathbf{d}_1^T \right] = 0$. Therefore $\text{Var}[\mathbf{b}] = \sigma_b^2 \mathbf{R}$ as required.

A similar approach can be used to express the cross covariance matrix in terms of $\mathbf{a}_1$ and $\mathbf{d}_1$ as follows:

$$\begin{aligned}\text{Cov}[\mathbf{a}, \mathbf{b}] &= E\left[(\mathbf{a} - \mathbf{m_a})(\mathbf{b} - \mathbf{m_b})^T \right] \\ &= E\left[\sigma_a \mathbf{a}_1 \left(\sigma_b \left(\delta_{ab} \mathbf{d}_1 + \sqrt{1 - \delta_{ab}^2} \mathbf{a}_1 \right) \right)^T \right] \\ &= \sigma_a \sigma_b \delta_{ab} E\left[\mathbf{a}_1 \mathbf{d}_1^T \right] + \sigma_a \sigma_b \sqrt{1 - \delta_{ab}^2} E\left[\mathbf{a}_1 \mathbf{a}_1^T \right] \\ &= \sigma_a \sigma_b \sqrt{1 - \delta_{ab}^2} \mathbf{R},\end{aligned}$$

since $E\left[\mathbf{a}_1 \mathbf{d}_1^T \right] = 0$ as before.

The joint covariance matrix for the difference-based model is now given by

$$\mathbf{K_w} = \begin{bmatrix} \sigma_a^2\mathbf{R}+\sigma_\mu^2\mathbf{I} & \sigma_a\sigma_b\sqrt{1-\delta_{ab}^2}\mathbf{R} \\ \sigma_a\sigma_b\sqrt{1-\delta_{ab}^2}\mathbf{R} & \sigma_b^2\mathbf{R}+\sigma_\nu^2\mathbf{I} \end{bmatrix}. \tag{4.6}$$

4.2.3 A Combined Model

Notice that covariance matrices of the correlation-based and difference-based models are actually the same, with the relationship

$$\delta_{ab}^2 = 1 - \rho_{ab}^2$$

between the difference parameter δ_{ab} and the cross-correlation coefficient ρ_{ab}. Figure 4-4 includes both parameters and illustrates how these essentially equivalent models differ in their controlling parameters.

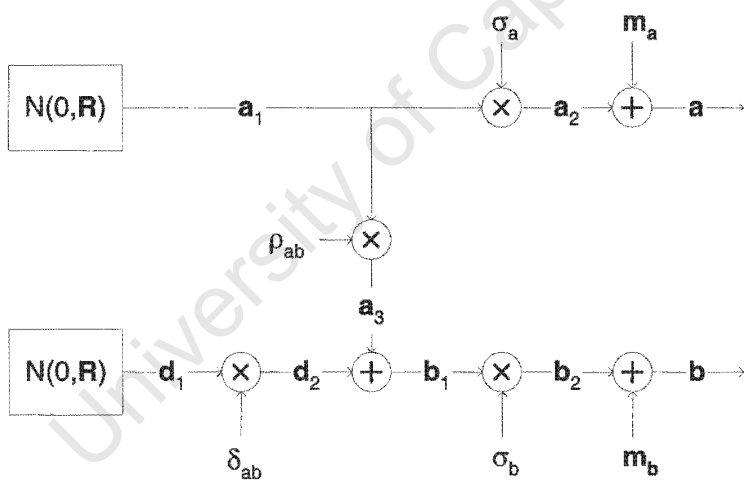


Figure 4-4: Process for generating an image pair with control over either a correlation or a difference parameter.

It is interesting to view the traditional correlation-based and difference-based similarity measures in the context of the combined process of Figure 4-4: the correlation-based measures (cross-correlation, correlation coefficient) estimate ρ_{ab}, and the difference-based measures (sum of squared differences, sum of absolute differences) estimate δ_{ab}. Seen in the light of the combined process these are essentially equivalent approaches, but differ with respect to the practical considerations associated with estimators: variance, bias, robustness and com-

putational economy. It should be noted that the proposed model suggests this interpretation of the traditional measures, but was not the rationale for their development. For the most part, authors considered the similarity measures to be a deterministic characteristic of the image pair, rather than the parameter of a stochastic model that had to be estimated in order to establish whether a match or mismatch model was in force.

To complete the image pair generation models under the match and mismatch hypotheses, two random processes that generate white noise fields are added to the process in Figure 4-5.

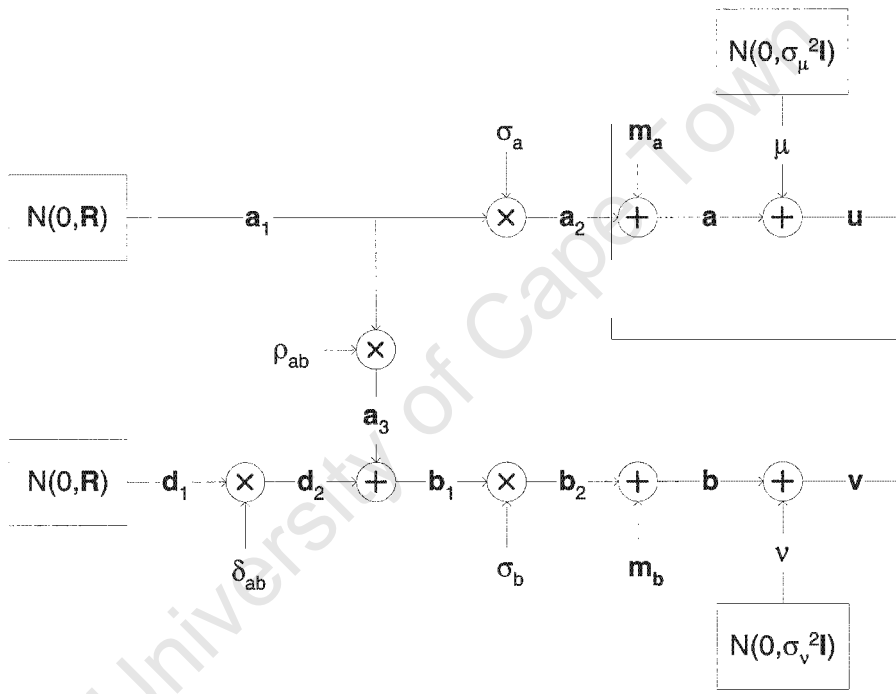


Figure 4-5: Process for generating an image pair with additive noise.

The hypothesis test for matching can now be based on a match and a mismatch model, where the two models differ only in the choice of value for δ_{ab} or ρ_{ab}. From this point onward the cross-correlation coefficient ρ_{ab} will be used as the match parameter[5]. Some observations are now made regarding the use of ρ_{ab} in the fields of image matching and multivariate statistics.

[5]The same procedure can be followed for the difference parameter δ_{ab} and an equivalent result will be obtained.

The Correlation Coefficient for Image Matching

If the elements of $\mathbf{a}$ are uncorrelated and the same is true of $\mathbf{b}$, then $\mathbf{R} = \mathbf{I}$ and the sample correlation coefficient

$$r(\mathbf{a}, \mathbf{b}) = \frac{\sum_i (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_i (a_i - \bar{a})^2}\sqrt{\sum_i (b_i - \bar{b})^2}} \quad \text{where } \bar{a} = \frac{1}{n}\sum_i a_i \quad \text{and } \bar{b} = \frac{1}{n}\sum_i b_i$$

is the maximum likelihood estimate of ρ_{ab}. Since images typically exhibit high degrees of spatial correlation, the form of (4.5) suggests that the sample correlation coefficient would be a better measure of image similarity if it were preceded by processing that whitened the image. Indeed, this sort of preprocessing has been motivated by several authors on both theoretical and experimental grounds [22, 35, 55].

Canonical Correlation Analysis

Principal component analysis (PCA) is a method for reducing the number of variables required to represent a correlated random vector while minimizing the loss of information incurred by doing so. Multivariate statistics provides an analogous method for reducing the correlation structure between two random vectors to its simplest possible form [29, p. 548]. This exploratory data analysis technique, *canonical correlation* analysis, provides an ordered set of linear transformations to extract the variables with maximum correlation from two random vectors.

Consider n^2-vectors $\mathbf{a}$ and $\mathbf{b}$. The first set of transformations $p_1 = \boldsymbol{\alpha}_1^T \mathbf{a}$ and $q_1 = \boldsymbol{\beta}_1^T \mathbf{b}$ give the first canonical variables, which have maximal correlation and unit variance. The second set of canonical variables, p_2 and q_2, have maximal correlation and unit variance, subject to the condition that they are uncorrelated with p_1 and q_1. This continues on to the n^2-th set of canonical variables, which have the lowest correlation and are uncorrelated with the $n^2 - 1$ previous sets of canonical variables. The i-th canonical correlation coefficient is the correlation coefficient between the i-th pair of canonical variables and lies in $[0, 1)$.

Given the joint covariance matrix of the ideal (noise-free) image pair

$$\begin{aligned} \mathbf{K}_c &= \begin{bmatrix} \mathbf{K_a} & \mathbf{K_{ab}} \\ \mathbf{K_{ab}} & \mathbf{K_b} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_a^2 \mathbf{R} & \sigma_a \sigma_b \rho_{ab} \mathbf{R} \\ \sigma_a \sigma_b \rho_{ab} \mathbf{R} & \sigma_b^2 \mathbf{R} \end{bmatrix} \end{aligned} \tag{4.7}$$

the squares of the canonical correlation coefficients are the eigenvalues of $\mathbf{K_a^{-1}K_{ab}K_b^{-1}K_{ab}}$ [29, p. 550]. Substituting (4.2),

$$\begin{aligned}\mathbf{K_a^{-1}K_{ab}K_b^{-1}K_{ab}} &= \left(\frac{\mathbf{R}^{-1}}{\sigma_a^2}\right)(\sigma_a\sigma_b\rho_{ab}\mathbf{R})\left(\frac{\mathbf{R}^{-1}}{\sigma_a^2}\right)(\sigma_a\sigma_b\rho_{ab}\mathbf{R}) \\ &= \rho_{ab}^2\mathbf{I}\end{aligned}$$

which has n^2 eigenvalues of ρ_{ab}^2. The n^2 canonical correlation coefficients of $\{\mathbf{a}, \mathbf{b}\}$ are therefore ρ_{ab}^2, confirming that all correlation between $\mathbf{a}$ and $\mathbf{b}$ is captured by the cross-correlation coefficient parameter ρ_{ab}.

4.3 Image-Pair Synthesis

A procedure for generating artificial images can be used to analyze matching algorithms numerically using Monte Carlo methods and to test them under controlled conditions with unlimited test data. Such a procedure for an MVN random field is described next, followed by the derivation of an efficient method for synthesizing correlated image pairs that are realizations of the joint model developed in the previous section.

4.3.1 Simulating Stationary MVN Fields

A procedure for synthesizing MVN random fields by generating a white noise image and transforming it to an image with the required covariance matrix is outlined by Johnson [90]. Conceptually, the procedure that generates $\mathbf{a}$, a sample of an $n \times n$ MVN field with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{K}$, is as follows:

1. Generate a zero mean, unit variance white noise image $\mathbf{z}$.

2. Derive the unitary transform matrix $\mathbf{G}$ that diagonalizes covariance matrix $\mathbf{K}$. Denote the diagonalized covariance matrix $\mathbf{K}_d$, where

$$\mathbf{K}_d = \mathbf{GKG}^T.$$

3. Scale the pixels in the white noise image by the square root of the elements on the diagonal of $\mathbf{K}_d$:

$$\acute{\mathbf{z}} : \acute{z}_i = \acute{w}_i\sqrt{\mathbf{K}_d(i,i)} \quad \forall\, i \in \{1, 2, \ldots, n\}$$

This step transforms $\mathbf{z}$ to a sample of a process with covariance matrix $\mathbf{K}_d$.

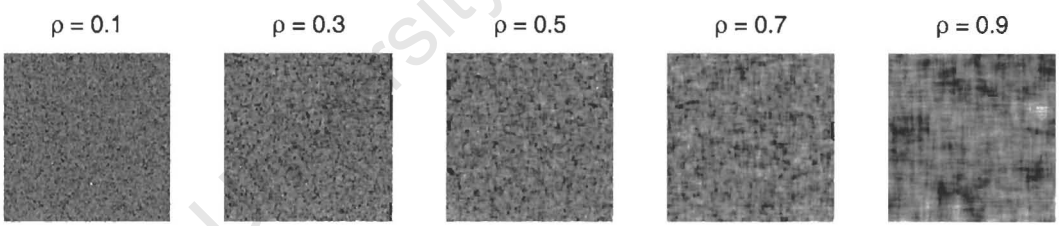
4. The unitary transform matrix $\mathbf{G}$ diagonalizes the covariance matrix $\mathbf{K}$. If $\acute{\mathbf{a}}$ is a sample of a $N(\mathbf{0}, \mathbf{K})$ process, then $\mathbf{G}\acute{\mathbf{a}}$ is a sample of a $N(\mathbf{0}, \mathbf{K}_d)$ process and vice versa, if $\acute{\mathbf{z}}$ is a sample of a $N(\mathbf{0}, \mathbf{K}_d)$ process, then $\mathbf{G}^{-1}\acute{\mathbf{z}}$ is a sample of a $N(\mathbf{0}, \mathbf{K})$ process [29, Theorem 1.2.6, p. 6]. The inverse of real unitary matrix $\mathbf{G}$ is its transpose, so

$$\acute{\mathbf{a}} = \mathbf{G}^T\acute{\mathbf{z}}$$

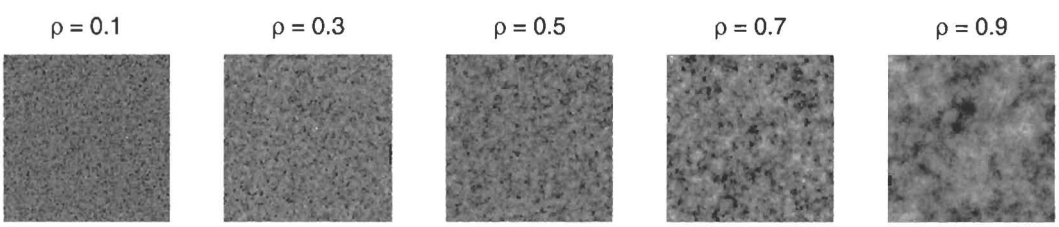
generates a sample with the required correlation structure.

5. Finally, the required mean is obtained by $\mathbf{a} = \acute{\mathbf{a}} + \mathbf{m}$.

In practice $\mathbf{K}$ can be embedded in a circulant matrix and diagonalized by the discrete Fourier transform (DFT) in step 2. Complications not discussed here include non-trivial embedding of some covariance matrices into a larger circulant matrix and the fact that the inverse DFT in step 4 provides a complex output, when a real sample is sought [91]. Figure 4-6 shows samples from separable and nonseparable Markov random fields [8, p. 33-37] that were generated using this procedure.



(a) Separable Model.



(b) Nonseparable Model.

Figure 4-6: Synthesized images based on Markov random fields.

4.3.2 Image-Pair Synthesis Equations

Image pairs that are samples of the model developed in the previous section can be synthesized for Monte Carlo simulation purposes. Generating independent white noise is trivial, but the generation of correlated image pairs is more difficult. One possibility is to use a method that generates image samples based on a known covariance matrix $\mathbf{K}$, like the one discussed in Section 4.3.1, but this approach suffers from the high dimensionality of $\mathbf{K}$ and does not exploit the Toeplitz-Block-Toeplitz structure that is commonly used in stationary image models with structured covariance.

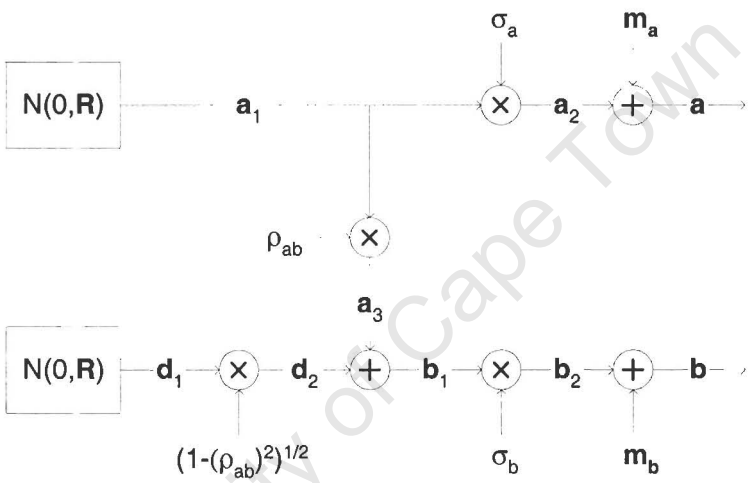


Figure 4-7: Process for synthesizing an image pair with specified cross-correlation coefficient.

A more realistic approach involves generating two images independently and transforming these to a correlated image-pair with the desired correlation coefficient. The process developed previously to illustrate the model (and repeated here in Figure 4-7) can be used for this purpose. Two random $N(0, \mathbf{R})$ images ($\mathbf{a}_1$ and $\mathbf{d}_1$ in Figure 4-7) are generated using the procedure in Section 4.3.1. The expressions

$$\mathbf{a} = \sigma_a \mathbf{a}_1 + \mathbf{m_a}$$

and

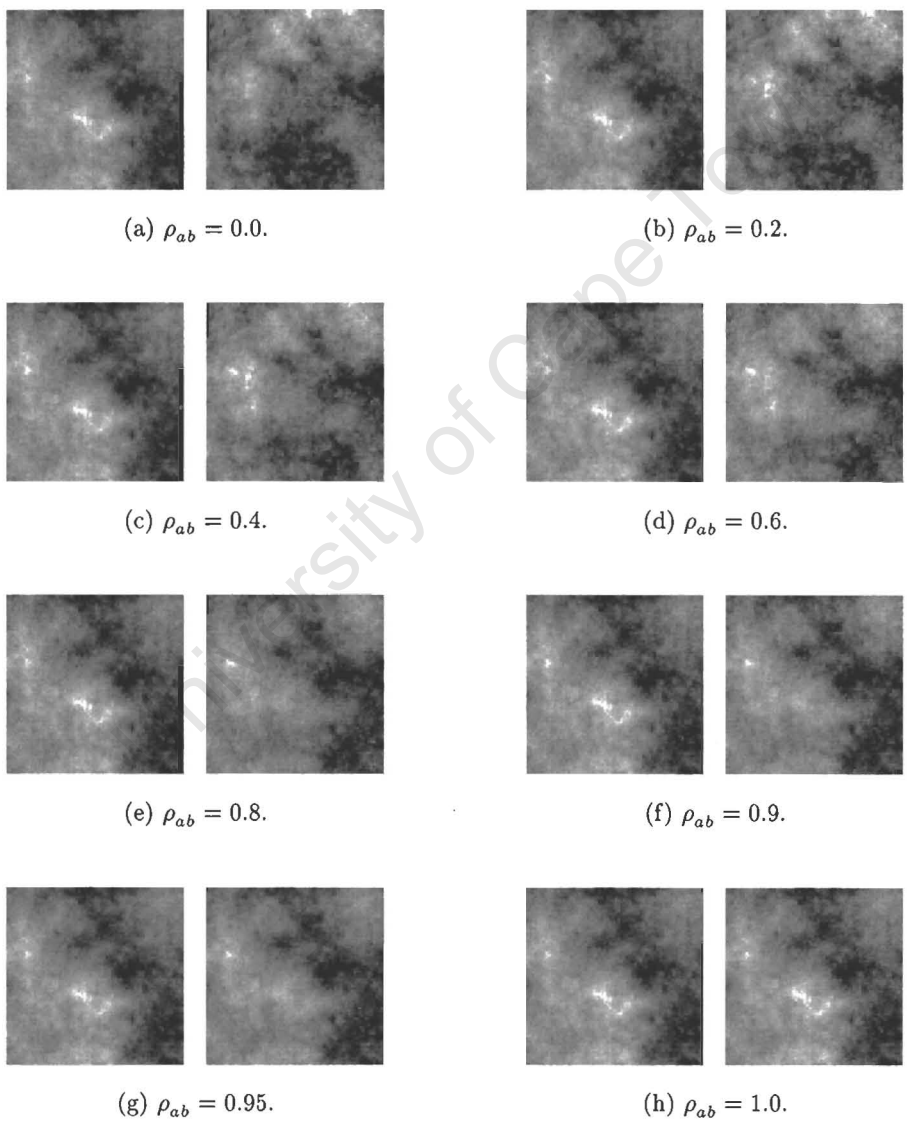
$$\mathbf{b} = \sigma_b \left(\rho_{ab} \mathbf{a}_1 + \sqrt{1 - \rho_{ab}^2} \mathbf{d}_1 \right) + \mathbf{m_b},$$

which are based directly on the process in Figure 4-7, can then be used to generate the

required images.

4.3.3 Example Image Pairs

Figure 4-8 shows examples of ideal image pairs synthesized with different match correlation coefficients. The individual images are first order Markov random fields (MRFs) [8, p. 36] with a one-step spatial correlation coefficient of $\rho = 0.8$.



(a) $\rho_{ab} = 0.0$. (b) $\rho_{ab} = 0.2$.

(c) $\rho_{ab} = 0.4$. (d) $\rho_{ab} = 0.6$.

(e) $\rho_{ab} = 0.8$. (f) $\rho_{ab} = 0.9$.

(g) $\rho_{ab} = 0.95$. (h) $\rho_{ab} = 1.0$.

Figure 4-8: Synthesized image pairs.

4.4 Discussion

The topic of image modelling is popular in certain parts of the image processing literature, and although many sophisticated models exist, a simple stationary multivariate normal model was selected here for its tractability. In general, real images violate these assumptions, but the use of an invertible transform to map the samples of nonstationary, non-normal random fields to realizations of stationary, normal ones makes this simple model more broadly applicable. Two other general assumptions were made: the (noise-free) scene components of the individual images share the same intra-image correlation structure, and the sensed images contain additive white noise.

The concepts of match and mismatch between images must be captured by the model in order to formulate the matching problem. This has been achieved by developing a joint probabilistic model for the image pair. It has been shown that an approach based on the correlation between corresponding pixel-pairs and an approach based on a scene difference image result in the same joint covariance matrix.

One of the benefits of having this model is that synthetic images can be generated for Monte Carlo analyses and algorithm testing. Procedures for generating synthetic image pairs efficiently have been developed for this purpose. The main reason for developing the model, however, was for the derivation of optimal hypothesis testing procedures for image matching, and this is the subject of the next chapter.

Chapter 5

Hypothesis Tests for Optimal Image Matching

Having proposed a linear model for the image pair in Chapter 4, attention now turns to the problem of designing an optimal hypothesis testing procedure for image matching. Optimal solutions realize the best possible performance if the model captures the nature of the observations accurately. Often, however, the model is an inadequate or incomplete representation of the observations and the test is suboptimal. Even in this case, the model-optimal solution can give insight into the important elements of a good test. The classic example of this situation is the matched filter, which was designed to maximize detectability of a deterministic signal in white noise, but is applied effectively on signals that deviate substantially from this model. Also, if all of the available *a priori* information is exploited, then the theoretical performance of the optimal solution on observations that conform to the model is the best performance attainable. Robust or non-parametric approaches can be evaluated in an absolute sense by comparing their model-theoretical performance to this maximum.

This chapter, then, aims to derive tests that optimize matching performance, to extract general insights on matching from these solutions and to provide a model-theoretical upper bound on matching performance that will be used to evaluate ad-hoc and intractable procedures later. Section 5.1 introduces the likelihood ratio and other tests for image matching that are based on the image-pair model of Chapter 4. The test is then simplified in Section 5.2, where a mathematically convenient representation for the test is developed on the basis of an eigendecomposition of the image covariance matrices. Section 5.3 investigates the statistical properties of the test. The test statistic is found to be asymptotically normal and this

fact is used to derive decision thresholds and analyze the probability of error. For the most part, tests with simple hypotheses are considered, but Section 5.4 formulates a generalized test with a composite match hypothesis. A final discussion concludes the chapter.

5.1 Hypothesis Tests Based on the Image-Pair Model

Chapter 3 formulated the image matching problem as several different hypothesis testing scenarios. The parametric model developed in Chapter 4 can now be used to derive specific tests. Given two random $n \times n$ images $\mathbf{u}$ and $\mathbf{v}$, the proposed model represents the image pair $\mathbf{w}^T = \begin{bmatrix} \mathbf{u}^T & \mathbf{v}^T \end{bmatrix}$ as an MVN random field, which has mean vector

$$\mathbf{m_w} = \begin{bmatrix} \mathbf{m_a} \\ \mathbf{m_b} \end{bmatrix}$$

and covariance matrix

$$\mathbf{K_w} = \begin{bmatrix} \sigma_a^2\mathbf{R}+\sigma_\mu^2\mathbf{I} & \sigma_a\sigma_b\rho_{ab}\mathbf{R} \\ \sigma_a\sigma_b\rho_{ab}\mathbf{R} & \sigma_b^2\mathbf{R}+\sigma_\nu^2\mathbf{I} \end{bmatrix}. \tag{5.1}$$

The parameter $\boldsymbol{\theta} = \{\rho_{ab}\}$, determines match or mismatch in the image pair. The parameters $\boldsymbol{\varphi} = \{\mathbf{R}, \sigma_a^2, \sigma_b^2, \sigma_\mu^2, \sigma_\nu^2\}$ are properties of the individual images. Different hypothesis testing scenarios are distinguished by the *a priori* information available about ρ_{ab} under the match and mismatch hypotheses, and by the extent of *a priori* knowledge about $\boldsymbol{\varphi}$.

5.1.1 The Likelihood Ratio Test

If $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ are well-known, then the hypotheses are simple and the optimal test is based on the likelihood ratio statistic. Here the simple match and mismatch hypotheses share the pdf $p_{\mathbf{w}}(\mathbf{w}|\rho_{ab})$, where $H_1 \iff \rho_{ab} = \rho_1$ and $H_0 \iff \rho_{ab} = \rho_0$, and the likelihood ratio test (LRT) is

$$l(\mathbf{w}) = \frac{p_{\mathbf{w}}(\mathbf{w}|\rho_{ab} = \rho_1)}{p_{\mathbf{w}}(\mathbf{w}|\rho_{ab} = \rho_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda, \tag{5.2}$$

where λ is an appropriate scalar decision threshold. Given the MVN image model, the likelihood ratio statistic can be written as

$$\begin{aligned} l(\mathbf{w}) &= \frac{N(\mathbf{w};\mathbf{m}_\mathbf{w},\mathbf{K}_{\mathbf{w}_1})}{N(\mathbf{w};\mathbf{m}_\mathbf{w},\mathbf{K}_{\mathbf{w}_0})} \\ &= \frac{|\mathbf{K}_{\mathbf{w}_0}|^{\frac{1}{2}}}{|\mathbf{K}_{\mathbf{w}_1}|^{\frac{1}{2}}} \exp\left[\frac{(\mathbf{w}-\mathbf{m}_\mathbf{w})^T \mathbf{K}_{\mathbf{w}_0}^{-1}(\mathbf{w}-\mathbf{m}_\mathbf{w})}{2} - \frac{(\mathbf{w}-\mathbf{m}_\mathbf{w})^T \mathbf{K}_{\mathbf{w}_1}^{-1}(\mathbf{w}-\mathbf{m}_\mathbf{w})}{2}\right], \end{aligned}$$

where $H_1 \iff \mathbf{K}_\mathbf{w} = \mathbf{K}_{\mathbf{w}_1}$ and $H_0 \iff \mathbf{K}_\mathbf{w} = \mathbf{K}_{\mathbf{w}_0}$. Noting that the logarithm is a monotonically increasing function, the test can be rearranged and expressed in terms of the statistic $s(\mathbf{w})$ and modified decision threshold $\acute{\lambda}$ as

$$s(\mathbf{w}) \underset{H_0}{\overset{H_1}{\gtrless}} \acute{\lambda},$$

where

$$\begin{aligned} s(\mathbf{w}) &= (\mathbf{w}-\mathbf{m}_\mathbf{w})^T \mathbf{K}_{\mathbf{w}_0}^{-1}(\mathbf{w}-\mathbf{m}_\mathbf{w}) - (\mathbf{w}-\mathbf{m}_\mathbf{w})^T \mathbf{K}_{\mathbf{w}_1}^{-1}(\mathbf{w}-\mathbf{m}_\mathbf{w}) \\ &= (\mathbf{w}-\mathbf{m}_\mathbf{w})^T \left(\mathbf{K}_{\mathbf{w}_0}^{-1} - \mathbf{K}_{\mathbf{w}_1}^{-1}\right)(\mathbf{w}-\mathbf{m}_\mathbf{w}) \end{aligned} \tag{5.3}$$

and

$$\acute{\lambda} = \log\left(\lambda^2 \frac{|\mathbf{K}_{\mathbf{w}_1}|}{|\mathbf{K}_{\mathbf{w}_0}|}\right). \tag{5.4}$$

The LRT, therefore, calculates the Mahalanobis distance between the sample image-pair and the mean image-pair under both hypotheses. The difference between these distances is then compared to a threshold. The decision threshold for the ideal observer LRT is given by

$$\lambda = \frac{1-P_1}{P_1},$$

where $P_1 = P(\rho_{ab} = \rho_1)$ is the *a priori* probability of a match. The ideal observer test, then, has statistic (5.3) and threshold

$$\acute{\lambda} = \log\left(\left(\frac{1-P_1}{P_1}\right)^2 \frac{|\mathbf{K}_{\mathbf{w}_1}|}{|\mathbf{K}_{\mathbf{w}_0}|}\right).$$

The minimax and Neyman-Pearson tests are also based on the likelihood ratio, but differ in the choice of threshold (see Section 3.1.2).

5.1.2 Generalized Tests

The LRT assumes that all of the model parameters are well-known. In practice, this is rarely the case, so implementation of the algorithm is preceded by an estimation stage (or training) that uses a representative set of data to establish what these parameters are. With good estimates, the parameters can be treated as well-known from this point onward. Often, however, the parameters will vary during the operation of the system. If a change in parameters is infrequent and can be detected, then the training can be repeated, but this requires system down-time and there is the possibility that the incorrect parameters will cause suboptimal operation for some time before the change is detected. An alternative is to use a nonparametric detection scheme that is not affected by parameter changes, but this approach sacrifices performance in favour of weaker assumptions. An approach that makes better use of the available *a priori* information might estimate the parameters on-line from the current observation. One such approach is the generalized likelihood ratio test (GLRT).

Where the model is parametrically known, one or both of the hypotheses must be composite. A distinction is now made between two types of parameters. The parameter ρ_{ab} determines match or mismatch and is therefore referred to as the match parameter. The parameters in $\boldsymbol{\varphi}$ are characteristics of the individual images and are referred to here as nuisance parameters. It is now shown that an asymptotically optimal test where match or nuisance parameters are unknown can be formulated as the LRT with maximum likelihood estimates of the unknown parameters.

Match Parameter

The GLRT can be used when the behaviour of ρ_{ab} is well-known under only one of the hypotheses. Consider the scenario where ρ_{ab} is only well-known under the mismatch hypothesis. Here the composite match hypothesis H_1 and simple mismatch hypothesis H_0 share the pdf $p_{\mathbf{W}}(\mathbf{w}|\rho_{ab})$, where $H_0 \iff \rho_{ab} = \rho_0$ and $H_1 \iff \rho_{ab} > \rho_0$. One intuitively appealing special case of this scenario is where images are assumed to be statistically independent under the mismatch hypothesis and $\rho_0 = 0$.

The GLRT with a simple mismatch and composite match hypothesis is

$$l_G(\mathbf{w}) = \frac{p_{\mathbf{W}}(\mathbf{w}|\rho_{ab} = \rho_0)}{\max_{\rho_{ab} \in (\rho_0, 1]} p_{\mathbf{W}}(\mathbf{w}|\rho_{ab})} \underset{H_1}{\overset{H_0}{\gtrless}} \lambda,$$

where the denominator maximizes the joint probability over all possible values of ρ_{ab}. The

GLRT can also be written in terms of an estimate of ρ_{ab} [92, p. 240] and then

$$l_G(\mathbf{w},\bar{\rho}_{ab}) = \frac{p_{\mathbf{W}}(\mathbf{w}|\rho_{ab}=\rho_0)}{p_{\mathbf{W}}(\mathbf{w}|\rho_{ab}=\bar{\rho}_{ab})} \underset{H_1}{\overset{H_0}{\gtrless}} \lambda, \tag{5.5}$$

where

$$\bar{\rho}_{ab} = \underset{\rho_{ab}\in(\rho_0,1]}{\text{argmax}} \left[p_{\mathbf{W}}(\mathbf{w}|\rho_{ab})\right]$$

is the maximum likelihood (ML) estimate of ρ_{ab}. Note that $\bar{\rho}_{ab}$ is an estimate based on the *observation* $\mathbf{w}$, and not on a set of *training data*. For the purposes of this discussion, parameters that can be estimated beforehand using training data are treated as well-known.

The presentation will now depart from statistical tradition in two respects. First, H_0 will always refer to the mismatch hypothesis, regardless of the formulation. For instance, the hypotheses for the scenario where the match hypothesis is simple and the mismatch hypothesis is composite are $H_0 \iff \rho_{ab} < \rho_1$ and $H_1 \iff \rho_{ab} = \rho_1$. In contrast, the statistical literature normally treats the well-known hypothesis as the null hypothesis H_0. Second, the generalized likelihood ratio will always be written so that the threshold is exceeded for the match hypothesis. With this convention, both forms of the GLRT are simply the LRT where ρ_{ab} is replaced by its ML estimate in the composite hypothesis.

The GLRT with simple mismatch hypothesis in (5.5) is rewritten as

$$l_G(\mathbf{w},\bar{\rho}_{ab}) = \frac{p_{\mathbf{W}}(\mathbf{w}|\rho_{ab}=\bar{\rho}_{ab})}{p_{\mathbf{W}}(\mathbf{w}|\rho_{ab}=\rho_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda. \tag{5.6}$$

The GLRT for the scenario with a simple match hypothesis and composite mismatch hypothesis is now

$$l_G(\mathbf{w},\bar{\rho}_{ab}) = \frac{p_{\mathbf{W}}(\mathbf{w}|\rho_{ab}=\rho_1)}{p_{\mathbf{W}}(\mathbf{w}|\rho_{ab}=\bar{\rho}_{ab})} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda,$$

where

$$\bar{\rho}_{ab} = \underset{\rho_{ab}\in[0,\rho_1)}{\text{argmax}} \left[p_{\mathbf{W}}(\mathbf{w}|\rho_{ab})\right].$$

Nuisance Parameters

The GLRT principle can also be applied if parameters other than ρ_{ab} are unknown, or if they cannot be reliably estimated from training data. Given the unknown parameter vector $\boldsymbol{\varphi}$, the

GLRT is

$$l_G(\mathbf{w}, \bar{\varphi}) = \frac{p_{\mathbf{w}}(\mathbf{w}|\rho_{ab} = \rho_1, \varphi = \bar{\varphi}_1)}{p_{\mathbf{w}}(\mathbf{w}|\rho_{ab} = \rho_0, \varphi = \bar{\varphi})} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda,$$

where $\bar{\varphi}$ is the ML estimate of φ, and $\bar{\varphi}_1$ is the ML estimate of φ under H_1.

Note that the scenarios of unknown match parameters and nuisance parameters can be combined. For example, the GLRT with well-known mismatch hypothesis, composite match hypothesis, and nuisance parameters φ is

$$l_G(\mathbf{w}, \bar{\rho}_{ab}, \bar{\varphi}) = \frac{p_{\mathbf{w}}(\mathbf{w}|\rho_{ab} = \bar{\rho}_{ab}, \varphi = \bar{\varphi})}{p_{\mathbf{w}}(\mathbf{w}|\rho_{ab} = \rho_0, \varphi = \bar{\varphi}_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda, \tag{5.7}$$

where $\bar{\varphi}_1$ is the ML estimate of φ under H_1.

The GLRT is *asymptotically* optimal with respect to the number of pixels in an image under certain regularity conditions on the ML estimates [92, p. 262], which essentially require that in the limit the estimates are as good as knowing the parameters. In practice there is no guarantee of optimum properties, but the GLRT often provides a good test [61, p. 419].

5.2 A Convenient Representation for the Test

The LRT can be implemented directly using the test statistic and decision threshold given in (5.3) and (5.4) respectively. However, these expressions are difficult to compute because of the dimensionality of $\mathbf{K}_{\mathbf{w}}$, and give little insight into the nature of the optimal test. A more convenient representation that uses an eigendecomposition of the individual image covariance matrices is now introduced.

5.2.1 The LRT as a Function of Whitened Images

A simplified LRT for $n \times n$ images $\mathbf{u}$ and $\mathbf{v}$ is derived here. It is assumed that the simple match and mismatch hypotheses share the pdf, $p_{\mathbf{w}}(\mathbf{w}|\rho_{ab}) = N(\mathbf{w}; \mathbf{m}_{\mathbf{w}}, \mathbf{K}_{\mathbf{w}})$, where $H_1 \iff \rho_{ab} = \rho_1$ and $H_0 \iff \rho_{ab} = \rho_0$. Recall that the composite hypotheses can be dealt with by replacing the unknown parameter with the appropriate maximum likelihood estimate of that parameter.

It is convenient to work with random images that have independent, identically distributed (iid) pixels. Consider an image $\mathbf{u}$ with covariance matrix $\mathbf{K}_{\mathbf{u}}$. The Karhunen-Loève (KL)

transform[1] is defined as [8, p. 163]

$$\acute{\mathbf{u}} = \mathbf{V}^T\mathbf{u},$$

where $\mathbf{V}$ is an orthogonal matrix with columns that are the eigenvectors of $\mathbf{K_u}$. The resulting vector has uncorrelated elements $\acute{u}_i$ with variance $\lambda_i^{\mathbf{u}}$, where $\lambda_i^{\mathbf{u}}$ is the eigenvalue corresponding to the i-th column-eigenvector in $\mathbf{V}$. If $\Lambda_{\mathbf{u}}$ is a matrix with these eigenvalues on the diagonal and zero elsewhere, then the transformation,

$$\acute{\mathbf{u}} = \mathbf{T_u}\mathbf{u} = \Lambda_{\mathbf{u}}^{-\frac{1}{2}}\mathbf{V}^T\mathbf{u}, \tag{5.8}$$

whitens the image in that the result has independent, identically distributed (iid) pixels with unit variance (see Appendix B.3). The whitened image has the mean vector

$$\mathbf{m}_{\acute{\mathbf{u}}} = \Lambda_{\mathbf{u}}^{-\frac{1}{2}}\mathbf{V}^T\mathbf{m}_{\mathbf{u}}.$$

Since $\mathbf{K_u}$ and $\mathbf{K_u}$ share the eigenvectors of $\mathbf{R}$ (see Appendix B.5), the corresponding transformation for $\mathbf{v}$ is

$$\acute{\mathbf{v}} = \mathbf{T_v}\mathbf{v} = \Lambda_{\mathbf{v}}^{-\frac{1}{2}}\mathbf{V}^T\mathbf{v}.$$

Appendix B.6 shows that if the random images $\mathbf{u}$ and $\mathbf{v}$ with joint covariance matrix

$$\mathbf{K_w} = \begin{bmatrix} \sigma_a^2\mathbf{R}+\sigma_\mu^2\mathbf{I} & \sigma_a\sigma_b\rho_{ab}\mathbf{R} \\ \sigma_a\sigma_b\rho_{ab}\mathbf{R} & \sigma_b^2\mathbf{R}+\sigma_\nu^2\mathbf{I} \end{bmatrix}$$

are independently transformed to unit variance iid images, then the resulting image pair has covariance matrix

$$\begin{aligned} \mathbf{K}_{\acute{\mathbf{w}}} &= \begin{bmatrix} \mathbf{T_u} & \mathbf{0} \\ \mathbf{0} & \mathbf{T_u} \end{bmatrix} \begin{bmatrix} \sigma_a^2\mathbf{R}+\sigma_\mu^2\mathbf{I} & \sigma_a\sigma_b\rho_{ab}\mathbf{R} \\ \sigma_a\sigma_b\rho_{ab}\mathbf{R} & \sigma_b^2\mathbf{R}+\sigma_\nu^2\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{T_u} & \mathbf{0} \\ \mathbf{0} & \mathbf{T_u} \end{bmatrix}^T \\ &= \begin{bmatrix} \mathbf{I} & \mathbf{D} \\ \mathbf{D} & \mathbf{I} \end{bmatrix}, \end{aligned} \tag{5.9}$$

[1] The literature also refers to this as the Hotelling transform and the method of principal components.

where $\mathbf{D}$ is a diagonal matrix with elements

$$\mathbf{D}[i,i] = k_i\rho_{ab} = \frac{\sigma_a\sigma_b\rho_{ab}\omega_i}{\sqrt{\left(\sigma_a^2\omega_i + \sigma_\mu^2\right)\left(\sigma_b^2\omega_i + \sigma_\nu^2\right)}} \tag{5.10}$$

and ω_i is the eigenvalue of $\mathbf{R}$ associated with the eigenvector in the i-th column of $\mathbf{V}$. This result implies that the correlation coefficient between the i-th pair of corresponding pixels in the whitened images is $k_i\rho_{ab}$.

The LRT statistic of equation (5.3) is now simplified for the case of unit variance, iid images with joint covariance according to (5.9). If the pixels in an image are independent of each other, then the state-conditional image-pair pdfs can be written as the product of the n^2 state-conditional pixel-pair pdfs. Consequently, the likelihood ratio is written in terms of the whitened image pair as

$$l\left(\acute{\mathbf{w}}\right) = \prod_{i=1}^{n^2} \frac{p_{\acute{\mathbf{w}}_i}\left(\acute{\mathbf{w}}_i|\rho_{ab}=\rho_1\right)}{p_{\acute{\mathbf{w}}_i}\left(\acute{\mathbf{w}}_i|\rho_{ab}=\rho_0\right)}, \tag{5.11}$$

where $\mathbf{w}_i = [u_i, v_i]^T$ is the i-th pixel-pair. The pixel-pairs are unit variance with correlation coefficient $k_i\rho_{ab}$ and therefore have the bivariate normal pdf

$$p_{\acute{\mathbf{w}}_i}\left(\acute{\mathbf{w}}_i\right) = \frac{\left(1-k_i^2\rho_{ab}^2\right)^{-\frac{1}{2}}}{2\pi} \exp\left[\frac{2k_i\rho_{ab}\left(\acute{u}_i - m_{\acute{u}_i}\right)\left(\acute{v}_i - m_{\acute{v}_i}\right) - \left(\acute{u}_i - m_{\acute{u}_i}\right)^2 - \left(\acute{v}_i - m_{\acute{v}_i}\right)^2}{2\left(1-k_i^2\rho_{ab}^2\right)}\right].$$

Substituting this pdf into (5.11), the likelihood ratio becomes

$$l\left(\acute{\mathbf{w}}_i\right) = \prod_{i=1}^{n^2}\left(\frac{1-k_i^2\rho_0^2}{1-k_i^2\rho_1^2}\right)^{\frac{1}{2}} \exp\left[\frac{\beta_i\left(\acute{u}_i - m_{\acute{u}_i}\right)\left(\acute{v}_i - m_{\acute{v}_i}\right) - \alpha_i\left(\left(\acute{u}_i - m_{\acute{u}_i}\right)^2 + \left(\acute{v}_i - m_{\acute{v}_i}\right)^2\right)}{2}\right],$$

where

$$\alpha_i = \frac{k_i^2\left(\rho_1^2 - \rho_0^2\right)}{\left(1-k_i^2\rho_0^2\right)\left(1-k_i^2\rho_1^2\right)} \tag{5.12}$$

and

$$\beta_i = 2\frac{k_i\left(\rho_1 - \rho_0\right)\left(1+k_i^2\rho_0\rho_1\right)}{\left(1-k_i^2\rho_0^2\right)\left(1-k_i^2\rho_1^2\right)}. \tag{5.13}$$

Noting that the logarithm is a monotonically increasing function, the LRT, $l(\acute{\mathbf{w}}) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$, can be rearranged and expressed in terms of the sufficient statistic

$$s(\acute{\mathbf{u}}, \acute{\mathbf{v}}) = \sum_{i=1}^{n^2} \beta_i (\acute{u}_i - m_{\acute{u}_i})(\acute{v}_i - m_{\acute{v}_i}) - \alpha_i \left((\acute{u}_i - m_{\acute{u}_i})^2 + (\acute{v}_i - m_{\acute{v}_i})^2 \right) \tag{5.14}$$

and modified decision threshold

$$\acute{\lambda} = \log \lambda^2 + \sum_{i=1}^{n^2} \log \left(\frac{1 - k_i^2 \rho_1^2}{1 - k_i^2 \rho_0^2} \right). \tag{5.15}$$

5.2.2 Performing the Test

The procedure for performing the LRT for images $\mathbf{u}$ and $\mathbf{v}$ now has three steps:

1. Whiten $\mathbf{u}$ and $\mathbf{v}$ using transformations, $\acute{\mathbf{u}} = \mathbf{T}_\mathbf{u}\mathbf{u}$ and $\acute{\mathbf{v}} = \mathbf{T}_\mathbf{v}\mathbf{v}$, respectively (see (5.8)).

2. Calculate the LRT statistic for whitened images (see (5.14)).

3. Compare the LRT statistic to the decision threshold (see (5.15)).

Note that this formulation of the test does not require the inversion of the joint image covariance matrix, but rather requires computation of the eigenvalues and eigenvectors associated with $\mathbf{K}_\mathbf{u}$ and $\mathbf{K}_\mathbf{v}$, the covariance matrices of the individual images. The eigendecomposition is required in order to specify the whitening transforms $\mathbf{T}_\mathbf{u}$ and $\mathbf{T}_\mathbf{v}$, and to calculate the weighting factors k_i. The latter are a function of the scene correlation coefficient matrix eigenvalues ω_i, which can be calculated from the eigenvalues of either individual image. Rearranging the result of Appendix B.4,

$$\omega_i = \frac{\lambda_i^\mathbf{u} - \sigma_\mu^2}{\sigma_a^2} \quad \text{or} \quad \omega_i = \frac{\lambda_i^\mathbf{v} - \sigma_\nu^2}{\sigma_b^2}.$$

5.2.3 Special Cases

Three special cases of the statistic are now considered.

Special Case 1: Mismatch Implies Independence

Here it is assumed that $H_0 \iff \rho_{ab} = 0$. Substituting $\rho_0 = 0$ into (5.14) the statistic

$$s\left(\acute{\mathbf{u}}, \acute{\mathbf{v}}\right) = \sum_{i=1}^{n^2} \frac{k_i \rho_1}{1 - k_i^2 \rho_1^2} \left[2\left(\acute{u}_i - m_{\acute{u}_i}\right)\left(\acute{v}_i - m_{\acute{v}_i}\right) - k_i \rho_1 \left(\left(\acute{u}_i - m_{\acute{u}_i}\right)^2 + \left(\acute{v}_i - m_{\acute{v}_i}\right)^2\right)\right]$$

is obtained. This is likely to be the most useful version of the LRT statistic, because statistical independence is a natural way to represent mismatch between images.

Special Case 2: Negligible Noise

If noise is negligible, then $k_i = 1$ and

$$s_{\mathrm{NF}}\left(\acute{\mathbf{u}}, \acute{\mathbf{v}}\right) = \sum_{i=1}^{n^2} \left[\beta\left(\acute{u}_i - m_{\acute{u}_i}\right)\left(\acute{v}_i - m_{\acute{v}_i}\right) - \alpha\left(\left(\acute{u}_i - m_{\acute{u}_i}\right)^2 + \left(\acute{v}_i - m_{\acute{v}_i}\right)^2\right)\right],$$

where

$$\alpha = \frac{\rho_1^2 - \rho_0^2}{\left(1 - \rho_0^2\right)\left(1 - \rho_1^2\right)}$$

and

$$\beta = 2\frac{\left(\rho_1 - \rho_0\right)\left(1 + \rho_0 \rho_1\right)}{\left(1 - \rho_0^2\right)\left(1 - \rho_1^2\right)}.$$

Neglecting the noise does not trivialize the matching problem, because if ρ_1 is less than unity, then the scene components of the images are in general not identical under the match hypothesis. This is a useful approximation of the test if some phenomenon other than additive noise, and that can be modelled by the match parameter ρ_1, is the dominant source of distortion in the matching problem (e.g. minor geometric transformations).

Special Case 3: Images with Independent Pixels

If the pixels in the individual images are spatially independent, then the eigendecomposition of the images, and hence the whitening transform, is trivial. In this case $T_{\mathbf{u}}\left(\mathbf{u}\right) = \sigma_a^{-1}\mathbf{u}$ and $T_{\mathbf{v}}\left(\mathbf{v}\right) = \sigma_b^{-1}\mathbf{v}$. The weighting factors are

$$k_i = k = \frac{\sigma_a \sigma_b}{\sqrt{\left(\sigma_a^2 + \sigma_\mu^2\right)\left(\sigma_b^2 + \sigma_\nu^2\right)}}$$

and the LRT statistic simplifies to

$$s_{\text{IP}}\left(\hat{\mathbf{u}}, \hat{\mathbf{v}}\right) = \sum_{i=1}^{n^2} \left[\beta\left(\acute{u}_i - m_{\acute{u}_i}\right)\left(\acute{v}_i - m_{\acute{v}_i}\right) - \alpha\left(\left(\acute{u}_i - m_{\acute{u}_i}\right)^2 + \left(\acute{v}_i - m_{\acute{v}_i}\right)^2\right)\right], \tag{5.16}$$

where

$$\alpha = \frac{k^2\left(\rho_1^2 - \rho_0^2\right)}{\left(1 - k^2\rho_0^2\right)\left(1 - k^2\rho_1^2\right)}$$

and

$$\beta = 2\frac{k\left(\rho_1 - \rho_0\right)\left(1 + k^2\rho_0\rho_1\right)}{\left(1 - k^2\rho_0^2\right)\left(1 - k^2\rho_1^2\right)}.$$

This version of the statistic is equivalent to an extension of the scalar matching test in Chapter 3 from scalars to images with independent pixels .

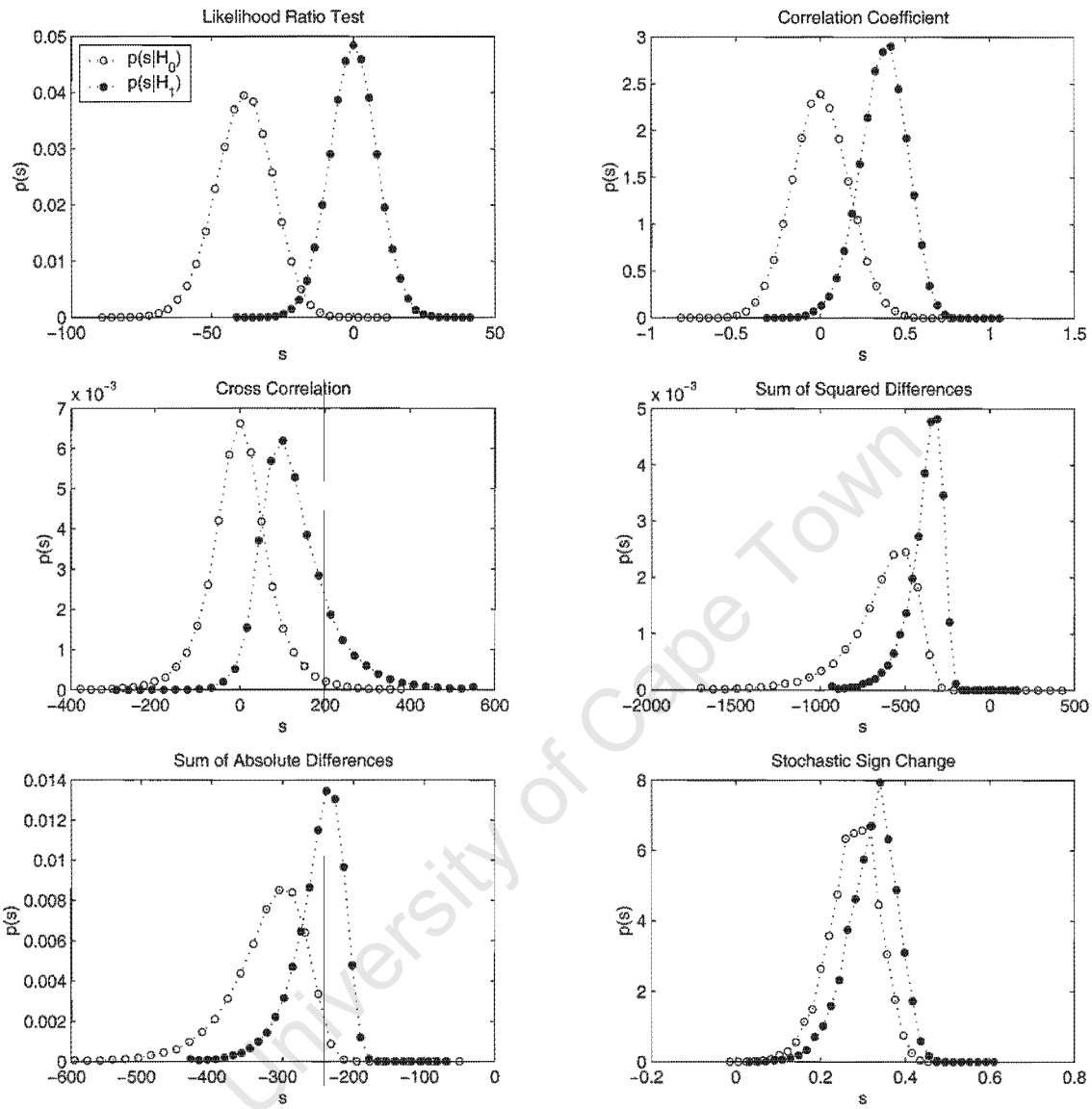
5.3 Properties of the Test

Figure 5-1 shows the results of a Monte Carlo simulation experiment that compares the pdfs under match and mismatch hypotheses for the LRT and five other image similarity statistics. The minimum error rate for equal *a priori* probability of match and mismatch in each case is represented by the area of overlap between the pdfs. In this particular experiment the LRT statistic is clearly superior to the others. A solution derived from first principles has another, perhaps even more important advantage than performance however: since it is based on a mathematical model, the LRT is more amenable to further analysis than might be the case for an ad-hoc solution.

In this section an expression for the pdf of the LRT statistic is derived. This allows optimal ideal observer and Neyman-Pearson decision thresholds to be found. The error rates can then be investigated and the relationships established between performance and important system parameters, such as image size and SNR.

5.3.1 PDF of the LRT Statistic

Since the LRT statistic is a function of two random images, it too is a random variable. The pdfs of the statistic under the match and mismatch hypotheses are important because they determine the error rates of the decision rule and can therefore be used to establish optimal



(a) Monte Carlo histograms.

Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.5
Scene one-step correlation (ρ)	0.8
SNR	2

(b) Simulation parameters.

Figure 5-1: Monte Carlo histograms under match and mismatch hypotheses for the LRT and five other similarity statistics. The experiment used an ensemble of 10000 image pairs generated using the procedure given in Chapter 4.

decision thresholds. The approach taken here is to first derive the expectation and variance of the statistic and then show that the pdf is asymptotically normal.

Expectation and Variance

Since the statistic of (5.14) has been written as a function of images with iid pixels, the pdf can be written as

$$p_s(s) = \prod_{i=1}^{n^2} p_{s_i}(s_i)$$

where

$$s_i(\acute{u}_i, \acute{v}_i) = \beta_i(\acute{u}_i - m_{\acute{u}_i})(\acute{v}_i - m_{\acute{v}_i}) - \alpha_i\left((\acute{u}_i - m_{\acute{u}_i})^2 + (\acute{v}_i - m_{\acute{v}_i})^2\right).$$

To enhance the clarity of the presentation this is rewritten in terms of the zero-mean, unit variance random variables $x_i = \acute{u}_i - m_{\acute{u}_i}$ and $y_i = \acute{v}_i - m_{\acute{v}_i}$. Note that from (5.9) and (5.10), x_i and y_i have correlation coefficient $\rho_i = \rho_{ab}k_i$, where k_i is defined as before. Consider a single random pixel $s_i = \alpha_i x_i^2 + \alpha_i y_i^2 - 2\beta_i x_i y_i$. Noting that the random variables x_i and y_i have the bivariate normal pdf

$$p_{x_i y_i}(x, y) = \frac{1}{2\pi\sqrt{1-\rho_i^2}} \exp\left[-\frac{1}{2(1-\rho_i^2)}\left(x^2 - 2\rho_i xy + y^2\right)\right],$$

the expectation and variance of s_i can be found using the method of moment generating functions (MGFs). The MGF of s_i is given by

$$\begin{aligned} m_{s_i}(t) &= E[\exp st] \\ &= E\left[\exp\left[\left(\beta_i xy - \alpha_i x^2 - \alpha_i y^2\right)t\right]\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left[\left(\beta_i xy - \alpha_i x^2 - \alpha_i y^2\right)t\right] \cdot p_{x_i y_i}(x, y)\ dx\ dy \\ &= \left(1 - 4\alpha_i^2 t^2 \rho_i^2 + \beta_i^2 t^2 \rho_i^2 + 4\alpha_i^2 t^2 - \beta_i^2 t^2 + 4\alpha_i t - 2\rho_i \beta_i t\right)^{-\frac{1}{2}} \end{aligned} \tag{5.17}$$

The r-th moment of s_i, $E[s_i^r]$, can be found by differentiating this MGF r times with respect to t and taking the limit of the result as $t \to 0$ [61, p. 78]. Following this procedure yields

$$E[s_i] = \rho_i \beta_i - 2\alpha_i$$

for the first and

$$E\left[s_i^2\right] = 3\left(\rho_i\beta_i - 2\alpha_i\right)^2 + \left(1-\rho_i^2\right)\left(\beta_i^2 - 4\alpha_i^2\right) \tag{5.18}$$

for the second raw moment. The variance is then

$$\begin{aligned}\mathrm{Var}\left[s_i\right] &= E\left[s_i^2\right] - \left(E\left[s_i\right]\right)^2 \\ &= 2\left(\rho_i\beta_i - 2\alpha_i\right)^2 + \left(1-\rho_i^2\right)\left(\beta_i^2 - 4\alpha_i^2\right). \end{aligned} \tag{5.19}$$

The overall statistic is now

$$s\left(\hat{\mathbf{u}}, \hat{\mathbf{v}}\right) = \sum_{i=1}^{n^2} s_i\left(\hat{u}_i, \hat{v}_i\right), \tag{5.20}$$

which is the summation of n^2 independent random variables, s_i, that have expectation and variance according to (5.18) and (5.19) respectively. With the linearity of expectation

$$E\left[s\right] = m_s = \sum_{i=1}^{n^2}\left(\rho_{ab}k_i\beta_i - 2\alpha_i\right) \tag{5.21}$$

and using the rule governing the variance of a summation of independent random variables[2] the variance of the statistic can be written as

$$\mathrm{Var}\left[s\right] = \sigma_s^2 = \sum_{i=1}^{n^2}\left[2\left(\rho_{ab}k_i\beta_i - 2\alpha_i\right)^2 + \left(1-\rho_{ab}^2k_i^2\right)\left(\beta_i^2 - 4\alpha_i^2\right)\right]. \tag{5.22}$$

Asymptotically Normal Distribution

The central limit theorem states that under certain conditions the distribution of a sum of random variables $\sum s_i$ is asymptotically normal [93, p. 214]. Sufficient conditions are that [93, p. 219]

1. $\lim_{n\to\infty}\sum \mathrm{Var}[s_i] = \infty$
2. There exists a number $p > 2$ and finite constant c such that $\int_{-\infty}^{\infty} s^p p_{s_i}(s)\ ds < c < \infty$, for all i.

[2] $\mathrm{Var}\left[\sum a_i x_i\right] = \sum a_i^2 \mathrm{Var}\left[x_i\right]$ [61, p. 178].

The sum in (5.20) satisfies the first condition since the variance of each term is positive. Using the moment generating function of (5.17) again, the third moment of each term is found to be

$$E\left[s_i^3\right] = 3\left(\rho_{ab}k_i\beta_i - 2\alpha_i\right)\left[5\left(\rho_{ab}k_i\beta_i - 2\alpha_i\right)^2 + 3\left(1-\rho_{ab}^2k_i^2\right)\left(\beta_i^2 - 4\alpha_i^2\right)\right],$$

satisfying the second condition, since ρ_{ab} is bounded. The pdf of the LRT statistic is therefore asymptotically normal with respect to the number of pixels in each image.

The pdfs of the statistic under the match and mismatch hypotheses are obtained by setting $\rho_{ab} = \rho_1$ and $\rho_{ab} = \rho_0$ respectively. Figure 5-2 compares the pdfs obtained by Monte Carlo simulation experiment with the theoretical normal approximation. Even for relatively small images the normal approximation is reasonable under both hypotheses. In Figure 5-3 the histogram reaches the normal approximation for larger image sizes than was the case in Figure 5-2, illustrating that convergence to the normal approximation is dependent on the image parameters. In particular, a high degree of spatial correlation in the images reduces the amount of independent information that is conveyed by a single pixel. As a consequence, larger images are required before the normal approximation becomes reasonable. If results that depend on normality of the LRT test are used (such as the Neyman Pearson threshold below), then the normality assumption under the expected image parameters can be tested by Monte Carlo experiment beforehand.

5.3.2 Optimal Decision Thresholds

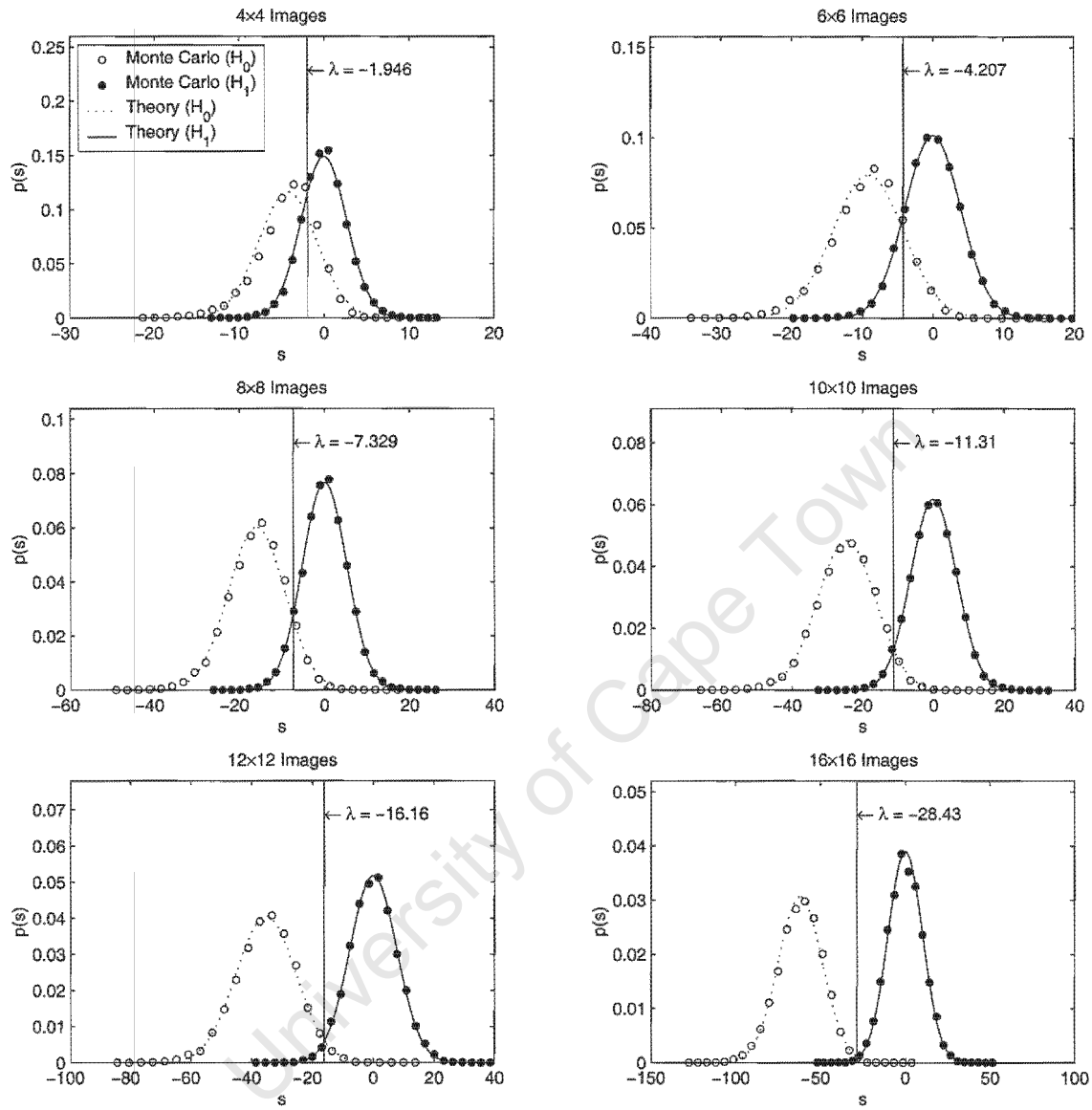
The decision thresholds for the ideal observer and Neyman-Pearson test are now specified.

Ideal Observer Test

The ideal observer test assumes that the *a priori* probability of a match P_1, is known and minimizes the overall probability of error. In this case the LRT decision threshold is

$$\acute{\lambda} = 2\log\left(\frac{1-P_1}{P_1}\right) + \sum_{i=1}^{n^2}\log\left(\frac{1-k_i^2\rho_1^2}{1-k_i^2\rho_0^2}\right),$$

where k_i is defined as before (see equation (5.10)). For $P_1 = 0.5$, the values of $\acute{\lambda}$ should correspond to the intersection of the match and mismatch pdfs, which is indeed the case in Figure 5-2. Note that the ideal observer threshold does not rely on a normal approximation of

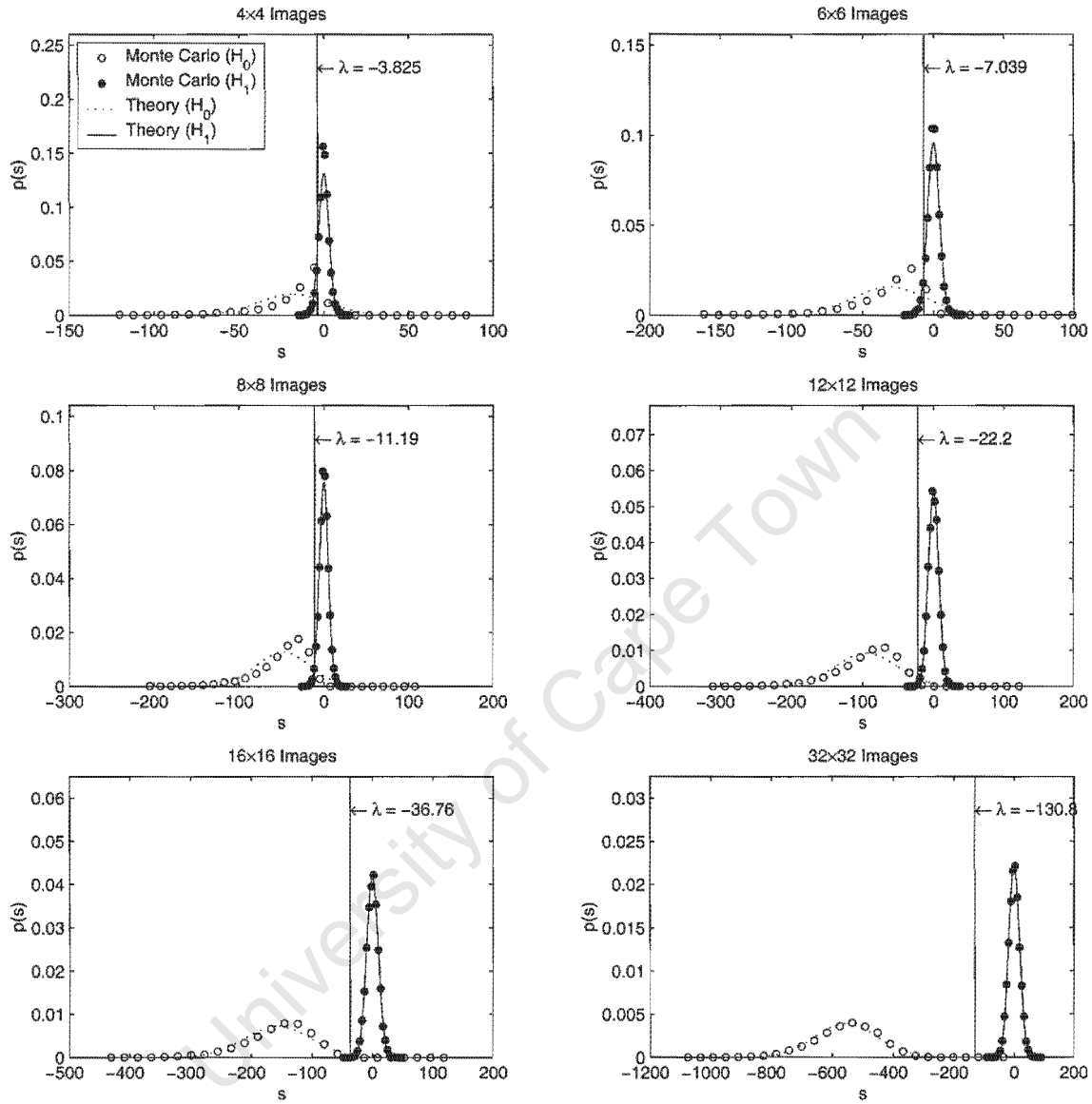


(a) Monte Carlo histograms and normal approximations.

Parameter	*Value*
Image size (n)	Range [4,16]
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.5
Scene one-step correlation (ρ)	0.6
SNR	2

(b) Simulation parameters.

Figure 5-2: Monte Carlo experiment illustrating asymptotic normality of the LRT statistic. The experiment used an ensemble of 10000 image pairs generated using the procedure given in Chapter 4.



(a) Monte Carlo histograms and normal approximations.

Parameter	*Value*
Image size (n)	Range [4,32]
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.95
Scene one-step correlation (ρ)	0.95
SNR	2

(b) Simulation parameters.

Figure 5-3: Monte Carlo experiment showing slower convergence to the normal distribution where spatial correlation is high. The experiment used an ensemble of 10000 image pairs generated using the procedure given in Chapter 4.

the LRT statistic. Where the normal approximation is inadequate for 4×4 images in Figure 5-2, the threshold corresponds to the intersection of the Monte Carlo histograms under the match and mismatch hypotheses.

Neyman-Pearson Test

The Neyman-Pearson test does not assume knowledge of P_1 and instead minimizes type II errors with a fixed upper bound on the type I error rate[3]. Since the likelihood ratio and an appropriately chosen threshold constitute an optimal test of simple binary hypotheses, the Neyman-Pearson test can be specified by simply choosing the threshold for a given upper bound α on the type I error rate. If the statistic under the mismatch hypothesis has the normal pdf, $p_s(s|H_0) = N(s; m_0, \sigma_0)$, and the decision threshold is $\acute{\lambda}$, then the probability of a type I error is given by (see Appendix B.7)

$$P_I = \frac{P_0}{2}\left(1 - \operatorname{erf}\left[\frac{\acute{\lambda} - m_0}{\sqrt{2}\sigma_0}\right]\right). \tag{5.23}$$

The *a priori* probability of a mismatch, P_0, is unknown so the threshold is chosen to minimize the probability of a type II error, P_{II}, with the constraint that $P_I \leq \alpha$, or

$$\frac{P_0}{2}\left(1 - \operatorname{erf}\left[\frac{\acute{\lambda} - m_0}{\sqrt{2}\sigma_0}\right]\right) \leq \alpha.$$

Since $P_0 \leq 1$,

$$\frac{1}{2}\left(1 - \operatorname{erf}\left[\frac{\acute{\lambda} - m_0}{\sqrt{2}\sigma_0}\right]\right) \leq \alpha \tag{5.24}$$

guarantees that $P_I \leq \alpha$. Noting that the error function, $\operatorname{erf}[\cdot]$, is monotonically increasing, (5.24) can be rewritten as a constraint on the Neyman-Pearson threshold

$$\acute{\lambda} \geq \sqrt{2}\sigma_0 \operatorname{erf}^{-1}[1 - 2\alpha] + \mu_0.$$

[3]Recall that type I errors occur when the test accepts a match hypothesis when the image-pair is in a state of mismatch, and vice versa for type II errors.

The probability of a type II error is (see Appendix B.7)

$$P_{II} = \frac{P_1}{2}\left(1 + \operatorname{erf}\left[\frac{\acute{\lambda} - m_1}{\sqrt{2}\sigma_1}\right]\right), \tag{5.25}$$

which is monotonically increasing in $\acute{\lambda}$. Therefore the threshold,

$$\acute{\lambda} = \sqrt{2}\sigma_0 \operatorname{erf}^{-1}[1 - 2\alpha] + \mu_0,$$

minimizes P_{II} with the constraint $P_I \leq \alpha$. Substituting the mean (5.21) and variance (5.22) of the asymptotically normal LRT statistic

$$\begin{aligned}\acute{\lambda} &= \sqrt{2}\sigma_0 \operatorname{erf}^{-1}[1 - 2\alpha] + \mu_0 \\ &= \sqrt{2\sum_{i=1}^{n^2}\left[2\left(\rho_{ab}k_i\beta_i - 2\alpha_i\right)^2 + \left(1 - \rho_{ab}^2 k_i^2\right)\left(\beta_i^2 - 4\alpha_i^2\right)\right]}\operatorname{erf}^{-1}[1 - 2\alpha] \\ &\quad + \sum_{i=1}^{n^2}\left(\rho_{ab}k_i\beta_i - 2\alpha_i\right).\end{aligned}$$

5.3.3 Probability of Error

Knowledge of the error rate as a function of image parameters is important for setting the best decision threshold, and for comparing different matching techniques. It could also be important for specifying design parameters such as minimum image size and SNR when designing an imaging system. The error rate for the LRT can be calculated analytically because the statistic has an approximately normal distribution for a useful range of image sizes. In general, however, it will be difficult to find these pdfs analytically and numerical Monte Carlo methods will be used later to make comparisons with other matching techniques.

Expressing the mean (5.21) and variance (5.22) of the LRT statistic as a function of the match correlation ρ_{ab},

$$m_s\left(\rho_{ab}\right) = \sum_{i=1}^{n^2}\left(\rho_{ab}k_i\beta_i - 2\alpha_i\right)$$

and

$$\sigma_s^2\left(\rho_{ab}\right) = \sum_{i=1}^{n^2}\left[2\left(\rho_{ab}k_i\beta_i - 2\alpha_i\right)^2 + \left(1 - \rho_{ab}^2 k_i^2\right)\left(\beta_i^2 - 4\alpha_i^2\right)\right].$$

Assuming the normal approximation is valid, $p_s(s|H_0) = N\left(s; m_s(\rho_0), \sigma_s^2(\rho_0)\right)$ and $p_s(s|H_1) = N\left(s; m_s(\rho_1), \sigma_s^2(\rho_1)\right)$, and the probabilities of type I and type II errors are (see Appendix B.7)

$$P_I(\rho_0) = \frac{P_0}{2}\left(1 - \operatorname{erf}\left[\frac{\acute{\lambda} - m_s(\rho_0)}{\sqrt{2\sigma_s^2(\rho_0)}}\right]\right) \tag{5.26}$$

and

$$P_{II}(\rho_1) = \frac{P_1}{2}\left(1 + \operatorname{erf}\left[\frac{\acute{\lambda} - m_s(\rho_1)}{\sqrt{2\sigma_s^2(\rho_1)}}\right]\right). \tag{5.27}$$

In (5.26) and (5.27) the error probabilities are written explicitly as a function of the inter-image correlation coefficient under the match and mismatch hypotheses, ρ_0 and ρ_1, but they are also dependent on other factors, such as image size (n), the extent of the spatial correlation in the individual images (ρ) and the SNR. Since the effects of these parameters are interrelated, the analysis of error rate is difficult, but Figures 5-4, 5-5, 5-6 and 5-7 investigate a range of parameter combinations in order to establish general trends. For these experiments, the *a priori* probability of match is assumed to be $P_1 = 0.5$.

Figure 5-4 shows that, as would be expected, the error rate decreases as the image size increases. Similarly, error rate decreases with increasing SNR and match correlation in Figures 5-5 and 5-7, respectively. Figure 5-6 shows that greater spatial correlation in the images leads to more errors, since an image with greater spatial correlation conveys less information. In a sense, increased spatial correlation is analogous to reduced image size.

5.4 The Composite Match Hypothesis

Suppose now that the value of ρ_{ab} for matching images is an unknown parameter. In this case the GLRT,

$$l_G(\mathbf{w}) = \frac{p_{\mathbf{W}}(\mathbf{w}|\rho_{ab} = \bar{\rho}_{ab})}{p_{\mathbf{W}}(\mathbf{w}|\rho_{ab} = \rho_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda,$$

must be formed with a composite match hypothesis. The test must be slightly reformulated and an ML estimator for ρ_{ab} derived.

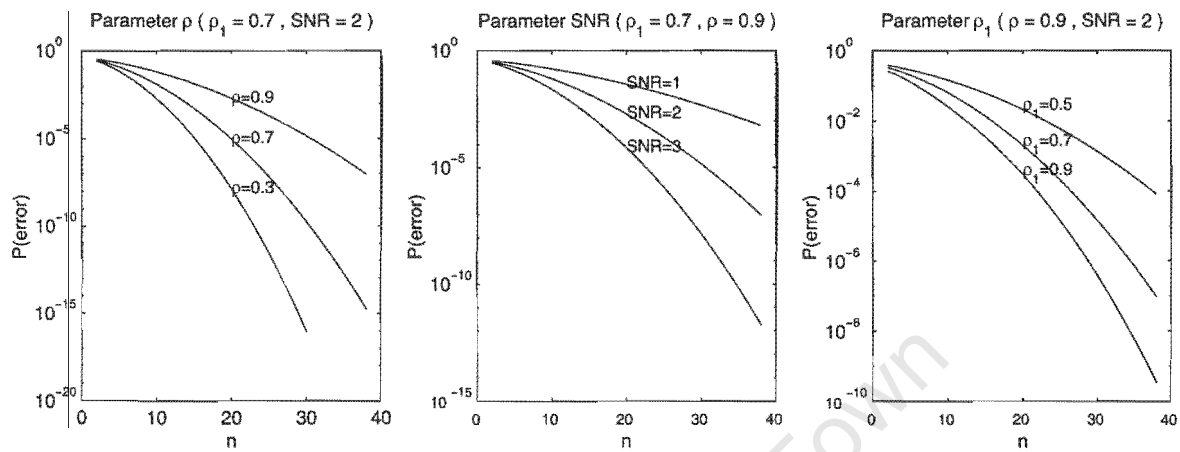


Figure 5-4: Probability of error versus image size.

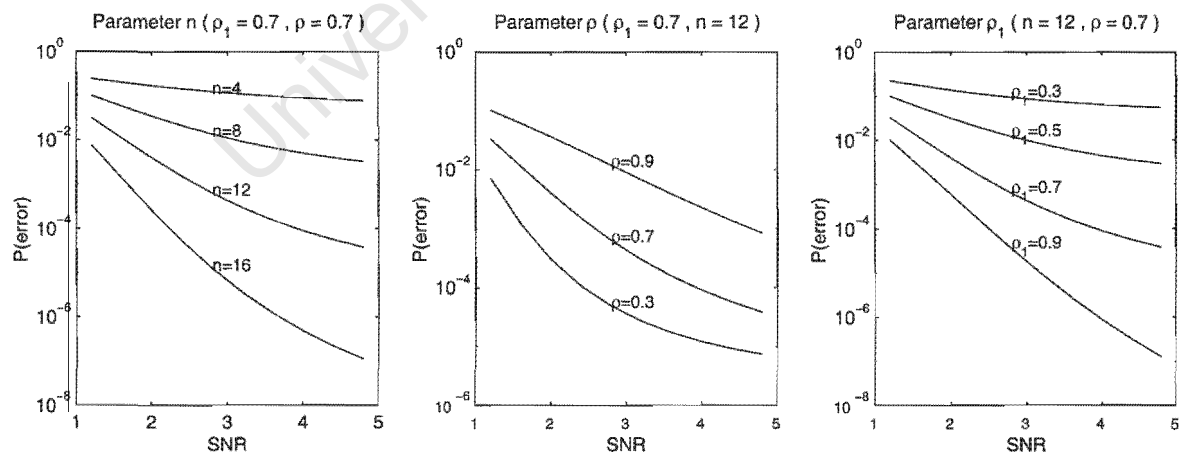


Figure 5-5: Probability of error versus SNR.

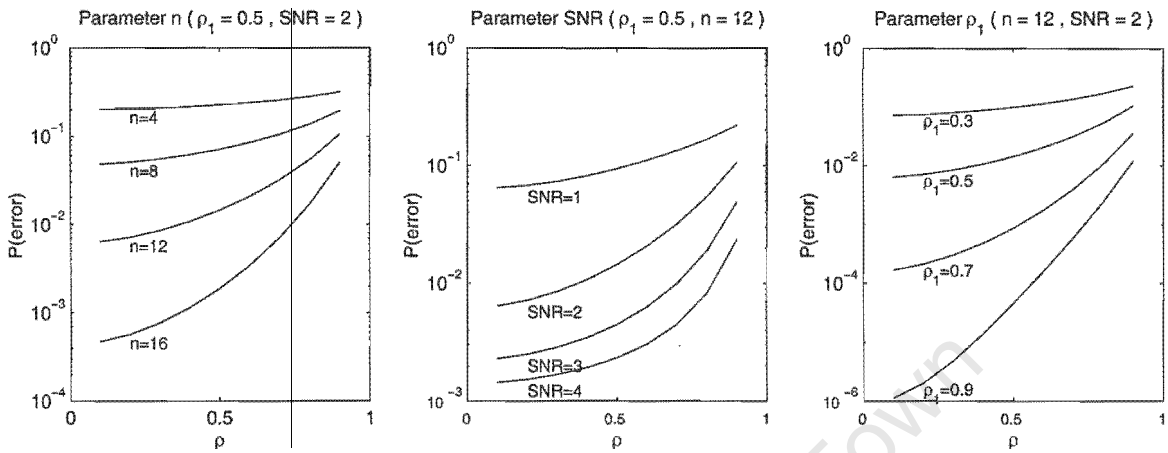


Figure 5-6: Probability of error versus one-step spatial correlation coefficient.

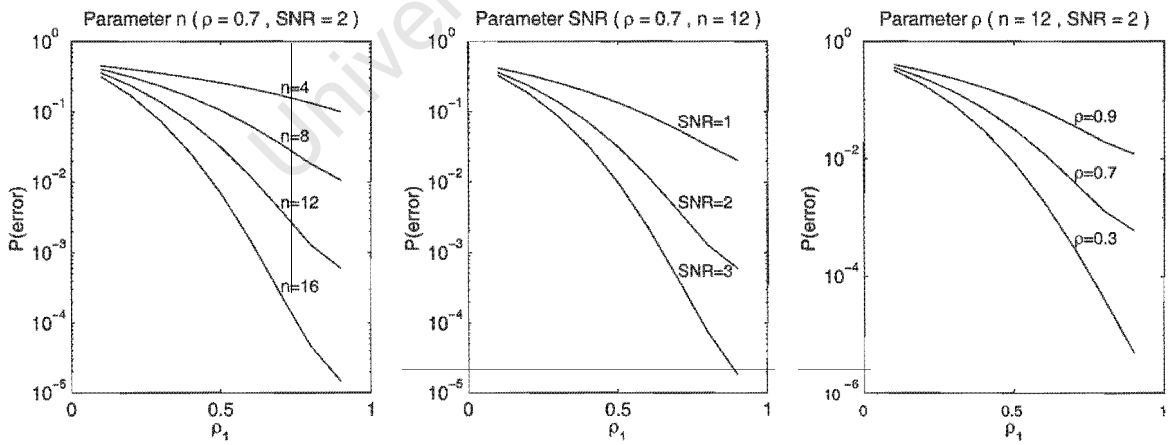


Figure 5-7: Probability of error versus match correlation coefficient.

5.4.1 Reformulating the Test

With one exception, the test statistic and decision threshold can be simplified in the same manner as the test based on simple hypotheses in the previous section. There is a term in the decision threshold that contains ρ_{ab} and since this quantity is no longer constant, it must be brought into the test statistic. Doing so, the GLRT statistic is

$$\begin{aligned} s_G\left(\acute{\mathbf{u}}, \acute{\mathbf{v}}, \bar{\rho}_{ab}\right) &= \sum_{i=1}^{n^2}\left[\beta_i\left(\bar{\rho}_{ab}\right)\left(\acute{u}_i - m_{\acute{u}_i}\right)\left(\acute{v}_i - m_{\acute{v}_i}\right) - \alpha_i\left(\bar{\rho}_{ab}\right)\left(\left(\acute{u}_i - m_{\acute{u}_i}\right)^2 + \left(\acute{v}_i - m_{\acute{v}_i}\right)^2\right)\right] \\ &\quad - \sum_{i=1}^{n^2}\gamma_i\left(\bar{\rho}_{ab}\right), \end{aligned}$$

where

$$\alpha_i\left(\bar{\rho}_{ab}\right) = \frac{k_i^2\left(\bar{\rho}_{ab}^2 - \rho_0^2\right)}{\left(1 - k_i^2\rho_0^2\right)\left(1 - k_i^2\bar{\rho}_{ab}^2\right)},$$

$$\beta_i\left(\bar{\rho}_{ab}\right) = 2\frac{k_i\left(\bar{\rho}_{ab} - \rho_0\right)\left(1 + k_i^2\rho_0\rho_1\right)}{\left(1 - k_i^2\rho_0^2\right)\left(1 - k_i^2\bar{\rho}_{ab}^2\right)}$$

and

$$\gamma_i\left(\bar{\rho}_{ab}\right) = \frac{\left(\sigma_a^2\omega_i + \sigma_\mu^2\right)\left(\sigma_b^2\omega_i + \sigma_\nu^2\right) - \left(\sigma_a\sigma_b\bar{\rho}_{ab}\omega_i\right)^2}{\left(\sigma_a^2\omega_i + \sigma_\mu^2\right)\left(\sigma_b^2\omega_i + \sigma_\nu^2\right) - \left(\sigma_a\sigma_b\rho_0\omega_i\right)^2}.$$

The corresponding decision threshold is

$$\acute{\lambda} = 2\log\lambda.$$

5.4.2 Estimating the Match Correlation Coefficient

The GLRT for a composite match hypothesis requires a ML estimate of the match correlation coefficient ρ_{ab}. For image matching the sample correlation coefficient between the pixels in $\mathbf{u}$ and $\mathbf{v}$, also referred to as Pearson's r and defined as

$$r_{\mathbf{uv}} = \frac{\sum_{i=1}^{n^2}\left(u_i - m\left(\mathbf{u}\right)\right)\left(v_i - m\left(\mathbf{v}\right)\right)}{\left[\sum_{i=1}^{n^2}\left(u_i - m\left(\mathbf{u}\right)\right)^2\sum_{i=1}^{n^2}\left(v_i - m\left(\mathbf{v}\right)\right)^2\right]^{\frac{1}{2}}}, \quad m\left(\mathbf{u}\right) = \frac{1}{n^2}\sum_{i=1}^{n^2}u_i \tag{5.28}$$

is routinely used as an estimate of ρ_{ab}. The presence of additive noise, however, biases the mean of this estimate away from the true value. Consider the joint covariance matrix of the image pair

$$\mathbf{K_w} = \begin{bmatrix} \sigma_a^2\mathbf{R}+\sigma_\mu^2\mathbf{I} & \sigma_a\sigma_b\rho_{ab}\mathbf{R} \\ \sigma_a\sigma_b\rho_{ab}\mathbf{R} & \sigma_a^2\mathbf{R}+\sigma_\mu^2\mathbf{I} \end{bmatrix}. \tag{5.29}$$

It is clear that (5.28) is estimating

$$\rho_{uv} = \frac{\sigma_a\sigma_b\rho_{ab}}{\sqrt{\left(\sigma_a^2+\sigma_\mu^2\right)\left(\sigma_b^2+\sigma_\nu^2\right)}}$$

and not the desired quantity, ρ_{ab}. As the additive noise increases, ρ_{uv} diverges away from ρ_{ab}. If the parameters are assumed known at this stage, then the bias can be removed by calculating

$$\bar{\rho}_{ab} = \frac{\sqrt{\left(\sigma_a^2+\sigma_\mu^2\right)\left(\sigma_b^2+\sigma_\nu^2\right)}}{\sigma_a\sigma_b}\rho_{uv},$$

but this estimator still treats the pixel pairs as independent. An ML estimate based on the model represented by (5.29) will make better use of the *a priori* information about spatial correlation that is captured in the image-pair covariance matrix.

Maximum Likelihood Estimate

A procedure for obtaining the match correlation coefficient ML estimate,

$$\bar{\rho}_{ab} = \underset{\rho_{ab}\in(\rho_0,1]}{\text{argmax}} \left[p_{\mathbf{w}}\left(\mathbf{w}|\rho_{ab}\right)\right],$$

given an image-pair sample $\mathbf{w}$ is introduced here. With known mean and covariance matrix for $\mathbf{w}$, an optimization procedure can be used to find the $\bar{\rho}_{ab} \in (\rho_0, 1]$ that maximizes the likelihood. However, the dimensionality of $\mathbf{w}$ makes this impractical for all but the smallest of images and a simplified procedure is now derived.

The likelihood is given by

$$p_{\mathbf{w}}\left(\mathbf{w}|\rho_{ab}\right) = \frac{1}{\sqrt{(2\pi)^{2n^2}\left|\mathbf{K_w}\right|}} \exp\left[-\frac{1}{2}\left(\mathbf{w}-\mathbf{m_w}\right)^T\mathbf{K_w^{-1}}\left(\mathbf{w}-\mathbf{m_w}\right)\right],$$

where $\mathbf{K_w}$ is given by (5.29). Now maximizing the likelihood of $\mathbf{w}$ over ρ_{ab} is equivalent to maximizing the likelihood of its whitened equivalent $\acute{\mathbf{w}}$. The latter has likelihood

$$p_{\acute{\mathbf{w}}}\left(\mathbf{w}|\rho_{ab}\right) = \frac{1}{\sqrt{\left(2\pi\right)^{2n^2}\left|\mathbf{K}_{\acute{\mathbf{w}}}\right|}} \exp\left[-\frac{1}{2}\left(\mathbf{w}-\mathbf{m}_{\acute{\mathbf{w}}}\right)^T \mathbf{K}_{\acute{\mathbf{w}}}^{-1}\left(\mathbf{w}-\mathbf{m}_{\acute{\mathbf{w}}}\right)\right], \tag{5.30}$$

where according to (5.10),

$$\mathbf{K}_{\acute{\mathbf{w}}} = \begin{bmatrix} \mathbf{I} & \mathbf{D} \\ \mathbf{D} & \mathbf{I} \end{bmatrix}, \text{ for } \mathbf{D}\left[i,i\right] = k_i\rho_{ab} = \frac{\sigma_a\sigma_b\rho_{ab}\omega_i}{\sqrt{\left(\sigma_a^2\omega_i+\sigma_\mu^2\right)\left(\sigma_b^2\omega_i+\sigma_\nu^2\right)}}.$$

In order to calculate (5.30) the inverse and determinant of $\mathbf{K}_{\acute{\mathbf{w}}}$ are required.

Inverse First, a special case of a property of partitioned matrices (see Muirhead's text, for example [29, Theorem A5.2, p. 580]) is introduced. If $\mathbf{K}$ is a nonsingular $n \times n$ matrix that is partitioned into four $\frac{n}{2} \times \frac{n}{2}$ submatrices as follows:

$$\mathbf{K} = \begin{bmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{A} & \mathbf{I} \end{bmatrix},$$

then the inverse, $\mathbf{B} = \mathbf{K}^{-1}$, can be partitioned into four $\frac{n}{2} \times \frac{n}{2}$ submatrices,

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix},$$

that can be expressed in terms of the submatrices of $\mathbf{K}$ as

$$\mathbf{B}_{11} = \mathbf{B}_{22} = \left(\mathbf{I}-\mathbf{A}^2\right)^{-1} \text{ and } \mathbf{B}_{12} = \mathbf{B}_{21} = -\mathbf{A}\left(\mathbf{I}-\mathbf{A}^2\right)^{-1}.$$

Using this property, the inverse of $\mathbf{K}_{\acute{\mathbf{w}}}$ can be written as

$$\mathbf{K}_{\acute{\mathbf{w}}}^{-1} = \begin{bmatrix} \left(\mathbf{I}-\mathbf{D}^2\right)^{-1} & -\mathbf{D}\left(\mathbf{I}-\mathbf{D}^2\right)^{-1} \\ -\mathbf{D}\left(\mathbf{I}-\mathbf{D}^2\right)^{-1} & \left(\mathbf{I}-\mathbf{D}^2\right)^{-1} \end{bmatrix}. \tag{5.31}$$

Determinant Once again, consider the partition of a matrix $\mathbf{K}$ into four equally sized submatrices:

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{12} & \mathbf{K}_{22} \end{bmatrix}.$$

If $\mathbf{K}_{22}$ is nonsingular, then [29, p. 581]

$$\det(\mathbf{K}) = \det(\mathbf{K}_{22}) \det\left(\mathbf{K}_{11} - \mathbf{K}_{12}\mathbf{K}_{22}^{-1}\mathbf{K}_{12}\right).$$

Using this property and (5.10)

$$\begin{aligned} \det(\mathbf{K}_{\acute{\mathbf{w}}}) &= \det(\mathbf{I}) \det\left(\mathbf{I} - \mathbf{D}^2\right) \\ &= \prod_{i=1}^{n^2} \left(1 - k_i^2 \rho_{ab}^2\right). \end{aligned} \tag{5.32}$$

Substituting (5.31) and (5.32) into the likelihood of (5.30), and rewriting in terms of the individual whitened images $\acute{\mathbf{u}}$ and $\acute{\mathbf{v}}$,

$$p_{\acute{\mathbf{u}},\acute{\mathbf{v}}}(\acute{\mathbf{u}}, \acute{\mathbf{v}} | \rho_{ab}) = \frac{1}{(2\pi)^{n^2} \sqrt{g(\rho_{ab})}} \exp\left[-\frac{1}{2} f(\acute{\mathbf{u}}, \acute{\mathbf{v}}, \rho_{ab})\right],$$

where

$$f(\acute{\mathbf{u}}, \acute{\mathbf{v}}, \rho_{ab}) = \sum_{i=1}^{n^2} \frac{1}{1 - k_i^2 \rho_{ab}^2} \left[(\acute{u}_i - m_{\acute{u}_i})^2 + (\acute{v}_i - m_{\acute{v}_i})^2 - 2k_i \rho_{ab} (\acute{u}_i - m_{\acute{u}_i})(\acute{v}_i - m_{\acute{v}_i})\right]$$

and

$$g(\rho_{ab}) = \prod_{i=1}^{n^2} \left(1 - k_i^2 \rho_{ab}^2\right).$$

The value of $\rho_{ab} \in (\rho_0, 1]$ that maximizes this expression for any given image pair is the ML estimate $\bar{\rho}_{ab}$. Note that it is equivalent to maximize the log-likelihood

$$L(\acute{\mathbf{u}}, \acute{\mathbf{v}} | \rho_{ab}) = -\frac{1}{2} \log g(\rho_{ab}) - \frac{1}{2} f(\acute{\mathbf{u}}, \acute{\mathbf{v}}, \rho_{ab}) - n^2 \log 2\pi,$$

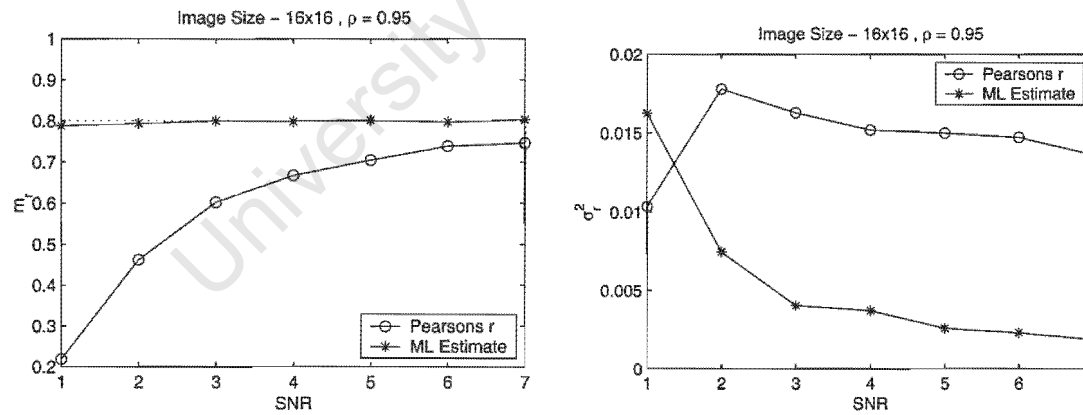
or, dropping constant terms and dividing by common coefficients, the modified log-likelihood

$$\acute{L}\left(\acute{\mathbf{u}}, \acute{\mathbf{v}} | \rho_{ab}\right) = -\log g\left(\rho_{ab}\right) - f\left(\acute{\mathbf{u}}, \acute{\mathbf{v}}, \rho_{ab}\right).$$

Calculus can be used to find an expression for $\bar{\rho}_{ab}$ in terms of $\acute{\mathbf{u}}$ and $\acute{\mathbf{v}}$. Alternatively, $\acute{L}\left(\acute{\mathbf{u}}, \acute{\mathbf{v}} | \rho_{ab}\right)$ can be maximized directly for given $\acute{\mathbf{u}}$ and $\acute{\mathbf{v}}$ using a numerical optimization procedure.

Comparing Estimators

Figure 5-8 shows the result of a Monte Carlo experiment that evaluates the performance of Pearson's r and the ML estimator based on direct maximization. The result supports the hypothesis that the ML estimator is unbiased. Pearson's r, on the other hand, is a biased estimator of ρ_{ab} as long as the noise is not negligible, but the bias can be removed with knowledge of the scene and noise variance. The ML estimate is asymptotically the minimum variance unbiased estimator, and Figure 5-8 confirms that for increasing n, the ML estimate of ρ_{ab} does indeed have the lower variance. Although the variance of the Pearson's r disqualifies



(a) Bias of estimator.

(b) Variance of estimator.

Figure 5-8: Performance of correlation coefficient estimators in the presence of additive noise (true $\rho_{ab} = 0.8$).

it as an estimator of ρ_{ab} for the GLRT, this does not imply that it cannot be used effectively as a measure of similarity in many applications. The reason for this is that as a measure of similarity, it is discrimination between match and mismatch that is important, rather than

performance as an estimator of some joint parameter of the image pair. In the next chapter Monte Carlo experiments will compare the discriminating power of Pearson's r with that of the optimal LRT and those of other similarity statistics.

5.5 Discussion

The hypothesis test for matching that is optimal with respect to a joint MVN model of the image pair has been derived. The test statistic and decision threshold have been simplified and are conveniently expressed in terms of the eigenvectors and eigenvalues of the correlation coefficient matrix that is shared by the two images. Performing the test is a three stage procedure. First the individual images are whitened. Then a likelihood ratio test statistic for iid images is calculated. Finally, the result is compared to a decision threshold that is chosen to optimize some performance criteria.

The fact that this statistic and decision threshold have been derived from first principles makes it possible to do analyses that would otherwise be difficult, and possibly intractable. Under the assumed model the test statistic is asymptotically normal and the mean and variance have been derived. Together with the expression for the optimal decision threshold, this facilitates the derivation of expressions for the error rates of the test. Knowledge of theoretical error rates and their relationships with image parameters can then be used to predict and optimize the performance of an image processing system.

However, assessing the performance of the LRT outside the assumed model, and deriving the error rates for the traditional approaches to direct image matching in order to draw comparisons, are tasks that are not as easy to perform analytically. The next chapter employs Monte Carlo simulation techniques in conjunction with the image-pair synthesis equations of Chapter 4 to analyze and compare the LRT error rate over a wide range of conditions.

Chapter 6

Error-Rate Performance of the Optimal Test

The likelihood ratio test (LRT) for image matching is now compared with other direct image matching methods over a range of imaging conditions. Both variation within the assumed model and deviation away from it are considered. Pearson's r (the sample correlation coefficient)[1], cross-correlation, the sum of squared differences, the sum of absolute differences and the stochastic sign change criterion are used as a basis for comparison (see Chapter 2 for definitions of these measures). The criteria for this selection were (1) that these statistics are widely used and (2) that they are useful over a wide range of image sizes. Table 6.1 lists the pertinent test and similarity statistics along with their symbols.

[1]In this chapter the nomenclature "Pearson's r" will be used instead of the more common "correlation coefficient" when referring to the sample correlation coefficient similarity statistic r. This will reduce the potential for confusion of this quantity with the match correlation coefficient ρ_1 or the spatial one-step correlation coefficient ρ.

Name	*Symbol*
Likelihood ratio test statistic	s
LRT statistic (noise-free approximation)	s_{NF}
LRT (independent pixel approximation)	s_{IP}
GLRT with unknown parameters θ	$s_G(\theta)$
Pearson's r	r
Cross correlation	R
Sum of squared differences	d_2
Sum of absolute differences	d_1
Stochastic sign change	s_s

Table 6.1: Test and similarity statistics investigated by Monte Carlo experiment.

For the purpose of the experiments it is assumed that the scene component of the individual images can be modelled as 2D, nonseparable, first-order Markov random fields. The correlation coefficient matrix $\mathbf{R}$ is therefore parameterized on the one-step spatial correlation coefficient ρ. It is also assumed that mismatch is conditioned on independent images, that is, $H_0 \iff \rho_{ab} = 0$.

Section 6.1 outlines experimental procedures. Section 6.2 presents the results of experiments that analyze error rate as a function of parameters in the joint image model. The image ensembles used here conform to the assumed model, but Section 6.3 investigates the effect of deviations from the model on error rate. Section 6.4 closes the chapter with a discussion on the material covered.

6.1 Monte Carlo Simulation

Monte Carlo methods, which comprise the branch of experimental mathematics that is concerned with random numbers, are used extensively in areas such as nuclear physics [94]. They can be categorized as either probabilistic, where random numbers simulate the random processes of the original problem, or deterministic, where they are used to solve a problem that is not random in nature. Methods in the former category, sometimes referred to as *direct simulation* methods [94, p. 43], are appropriate for the analysis of matching error rate. Image formation has been modelled as a random process and equations that transform individual random numbers into random image pairs were derived in Chapter 4. The Monte Carlo procedure involves using a random number generator in conjunction with these equations to generate a random ensemble of image pairs that conforms to the model. Matching techniques can then be applied to the image pairs in the ensemble and the resulting error rates observed.

A crucial element of this procedure is the generation of random numbers, which should conform to the desired probability distribution. Typically, a computer-generated pseudorandom sequence is used, and the Gaussian random numbers used in this chapter's experiments were generated by the pseudorandom number generator provided in the MATLAB numerical mathematics software library.

6.1.1 Experimental Procedure

The matching test consists of observing an image pair, calculating a scalar statistic of the image pair data and comparing the statistic to a decision threshold. A simple Monte Carlo

procedure for estimating the error rate associated with a particular matching test and *a priori* probability of match P_1 (and therefore *a priori* probability of mismatch $P_0 = 1 - P_1$) is as follows:

1. Denoting as $T = T_0 + T_1$ the overall number of Monte Carlo trials, generate an ensemble of $T_1 = P_1 T$ matching images and an ensemble of $T_0 = P_0 T$ non-matching images.
2. Use the test to make a match/mismatch decision for each image pair.
3. Compare the decision with the true class for each image pair and count the number of type I and type II errors, denoting them N_I and N_{II} respectively.
4. Calculate estimates of the type I and type II probabilities using $P_I = N_I \cdot T^{-1}$ and $P_{II} = N_{II} \cdot T^{-1}$.

This procedure is impractical for two reasons. First, if $P_1 \gg P_0$ then a very large number of trials may be required in order to obtain good estimate of P_I and likewise for $P_0 \gg P_1$ and P_{II}. This problem can be overcome by performing the simple procedure with equally sized ensembles for match and mismatch and scaling the resulting error rates by P_1 and P_0. Second, the procedure cannot be used as it stands if the decision threshold is as yet unspecified, which rules out comparisons with the traditional direct image matching techniques since these typically do not have a specified optimal decision threshold. Two methods for comparing error rate performance without a specified threshold are now considered.

The Receiver Operating Characteristic

The receiver operating characteristic (ROC) [95] characterizes the error rate performance of a test by the hypothesis conditional probabilities of match detection $P_d = P_{II} \cdot T \cdot T_1^{-1} = N_{II} \cdot T_1^{-1}$, and match false alarm, $P_f = P_I \cdot T \cdot T_0^{-1} = N_I \cdot T_0^{-1}$. The ROC plot is the locus of $\{P_d, P_f\}$ pairs for all possible decision thresholds. Figure 6-1 shows an example of an ROC plot for three different tests. The superior test is the one closest to the top left corner of the graph, where P_d is maximized and P_f is minimized. Note that the ROC plot does not require knowledge of the *a priori* probability P_1.

Minimum Error Rate and the Ideal Observer Test

In order to compare different test statistics and similarity measures it is convenient to represent the error-rate performance with a single value. One possibility is the overall probability

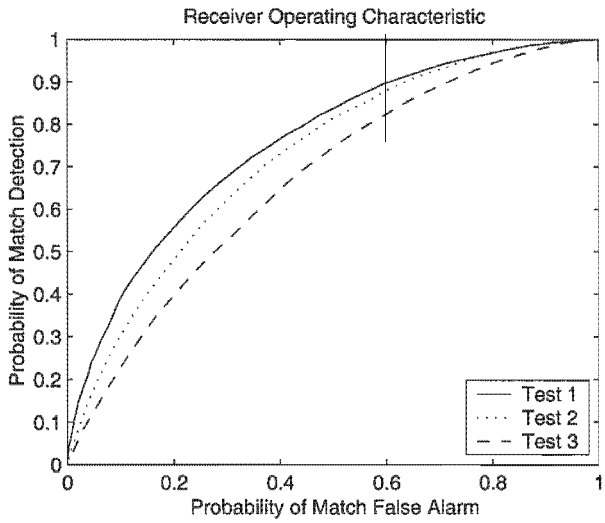


Figure 6-1: Receiver operating characteristic plot.

of error, or error rate, $P(\text{error}) = P_I + P_{II}$, which can be calculated using the simple procedure outlined above. If there is no specified threshold, then the threshold that minimizes the error rate can be used. Together with the test statistic, this threshold constitutes the ideal observer test (see Section 5.3.2) and is the value for which normal hypothesis conditional pdfs intersect.

6.1.2 Simulation Parameters and Default Values

The parameters in the assumed model are given in Table 6.2, together with the default values that will be used in experiments that follow. Some motivation for the selection of these particular values is in order. Jain suggests that for many classes of images a value of $\rho = 0.95$ is appropriate for the Markov one-step spatial correlation coefficient [8, p. 37]. Most experiments will be performed for two match correlation coefficients: $\rho_1 = 0.6$ models the situation where matching scenes can have significant differences under the match hypothesis, and $\rho_1 = 0.99$ models scene images that are nearly identical under the match hypothesis. The mismatch correlation coefficient is assumed to be $\rho_0 = 0$, implying that non-matching images are statistically independent.

In many applications the images in the pair will be generated by similar or identical imaging systems. For this reason the scene and noise variance are assumed to be shared by the two images and denoted as $\sigma^2 = \sigma_a^2 = \sigma_b^2$ and $\sigma_\eta^2 = \sigma_\mu^2 = \sigma_\nu^2$ respectively. Without loss of

Name	*Symbol*	*Default Value*
Image mean vector	$\mathbf{m}$	$\mathbf{0}$
Mismatch correlation coefficient	ρ_0	0
Match correlation coefficient (low)	ρ_1	0.60
Match correlation coefficient (high)	ρ_1	0.99
Markov spatial one-step correlation coefficient	ρ	0.95
Signal-to-Noise Ratio (low)	SNR	1
Signal-to-Noise Ratio (medium)	SNR	2
Signal-to-Noise Ratio (high)	SNR	3

Table 6.2: Default values for model parameters in Monte Carlo experiments.

generality, scene variance can be set to unity and the noise variance specified via signal-to-noise ratio, where SNR $= \sigma/\sigma_\eta$. Different imaging applications will exhibit widely different SNR performance, so any choice of a particular default value will be somewhat arbitrary. Depending on the experiment, values of SNR $= 1$, SNR $= 2$ or SNR $= 3$ are used in this chapter. By default the image pixels have zero mean, that is $\mathbf{m_u} = \mathbf{m_v} = \mathbf{0}$.

In each experiment the number of Monte Carlo trials (T_0 and T_1) is selected in order to obtain an error in the result that is negligible in comparison with any trends reported regarding differences between the error rates of the respective tests[2].

6.1.3 Selection of the Competitors

The results of experiments reported in this chapter do not represent an exhaustive comparison of the LRT with all other similarity statistics. Error-rate performance is compared with standard correlation- and difference-based similarity measures and a nonparametric measure, which is used for comparison where reality deviates from the assumed model. The chosen measures, which are shown in Table 6.1, are widely used and are useful over a range of image sizes. Measures that have a narrow range of applicability are not included. For example, although histogram based measures like mutual information are popular, particularly for multimodal image matching, they are ineffective where there are insufficient pixels to form an adequate histogram.

[2]The satisfaction of this requirement was verified by assessing the repeatability of the result in multiple experiments.

6.2 Error Rate and Model Parameters

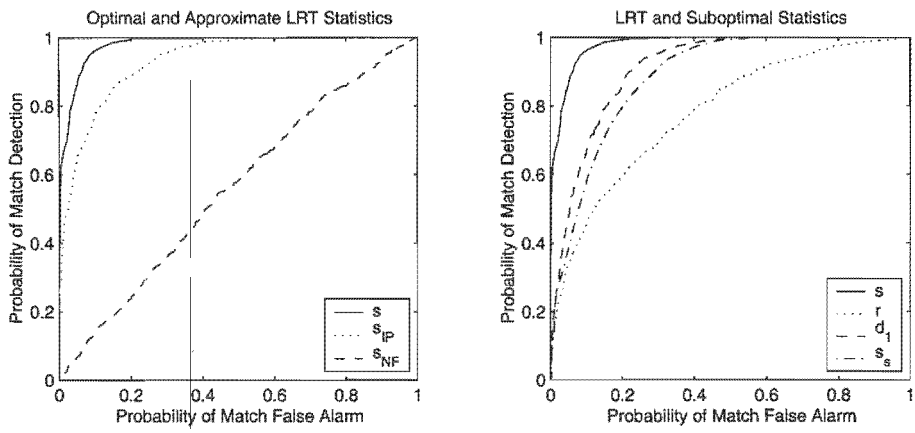
This section investigates the error-rate performance of direct matching techniques under changes within the assumed image-pair model. Figure 6-2 shows ROC plots for three sets of match correlation coefficient and SNR values, with default values used for the other model parameters. Some deductions can be made from these graphs. First, as would be expected for images generated under the assumed model, the LRT exhibits the best performance in all six graphs. Second, the independent-pixel approximate LRT exhibits close-to-optimal performance for low SNR and low match correlation (Figure 6-2(b)). Third, the best suboptimal measure for low match correlation and high SNR is the correlation coefficient. Fourth, the best suboptimal measure for high match correlation and low SNR is the sum of absolute differences. Finally, the stochastic sign change criterion exhibits poor performance for low match correlation.

In order to analyze performance over a wider range of parameter values, results are now presented in terms of the overall probability of error associated with the ideal observer test. It is assumed that the *a priori* probability of match is $P_1 = 0.5$.

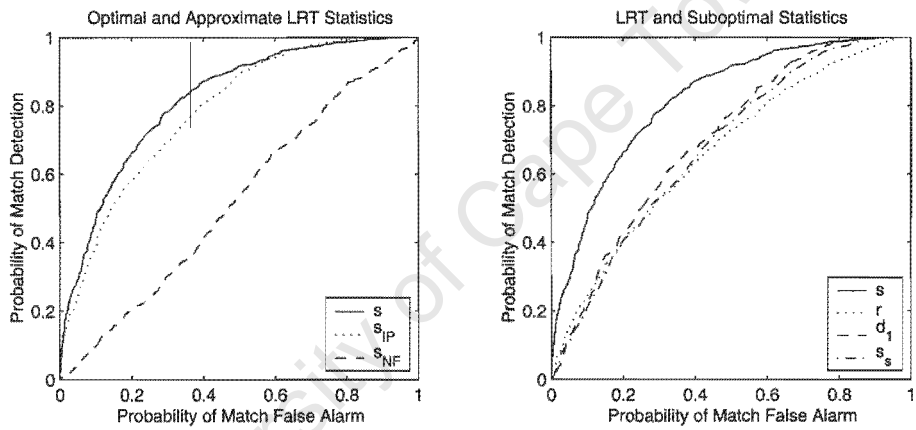
6.2.1 Image Parameters

Here error rate is compared for parameters of the individual images. Image size, SNR, and spatial correlation are investigated. Before discussing each experiment in more detail, some general observations are made. As expected, the optimal LRT produces the best performance. This is no surprise, since the test was derived from first principles to minimize error rate. What is of more interest is the extent to which the optimal measure outperforms the suboptimal measures, and indeed, the LRT does exhibit a large performance advantage in the results. For example, in the range $n \in [15, 25]$ in Figure 6-3(a) and the range SNR $\in [2, 3]$ in Figure 6-4(a), the LRT error is almost an order of magnitude lower than that of the next best measure.

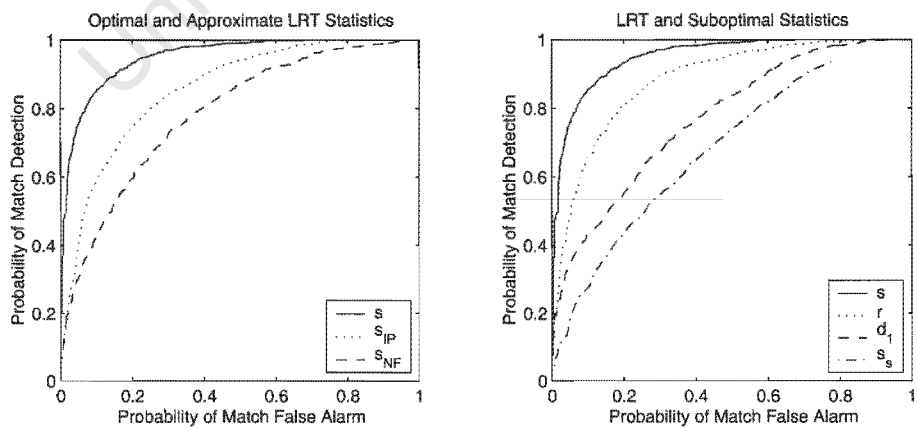
Among the suboptimal measures, Pearson's r consistently produces better performance than the sum of absolute differences when the match correlation coefficient is low. The opposite is true when the match correlation coefficient is high. This can be attributed to the fact that as the sample correlation coefficient, Pearson's r is more sensitive to changes in the match correlation coefficient ρ_1 even though it is not directly estimating this parameter. The better suboptimal measure always outperforms the stochastic sign change (SSC) criterion, which is to be expected, since a test based on the SSC criterion is nonparametric and makes few *a priori* assumptions about the problem at hand. It should be less powerful than parametric



(a) $\rho_1 = 0.99$. SNR $= 1$.



(b) $\rho_1 = 0.60$. SNR $= 1$.



(c) $\rho_1 = 0.60$. SNR $= 3$.

Figure 6-2: ROC comparison of the LRT statistic and common similarity measures. Graphs on the left compare the LRT to its noise-free and pixel-independence approximations. Graphs on the right compare the LRT statistic to other suboptimal similarity measures ($T_0 = T_1 = 1000$).

measures in experiments based on parametric models. In this respect, a fairer comparison is made in Section 6.3, where deviations from the parametric model are considered.

It is also clear from the majority of the results that the independence and noise-free assumptions severely compromise the performance of the LRT when the images do not support them. This is an indication of the potential for lost performance in other image processing algorithms, where these common assumptions are often used to simplify the mathematics of derivation or to reduce the computational complexity of the derived solution. This result would not surprise researchers in the field of robust statistics — Hampel, Ronchetti, Rousseeuw and Stahel comment, for example, that the independence assumption is "apart from systematic errors the most dangerous violation of usual statistical assumptions" [30, p. 8, sec. 1.1b].

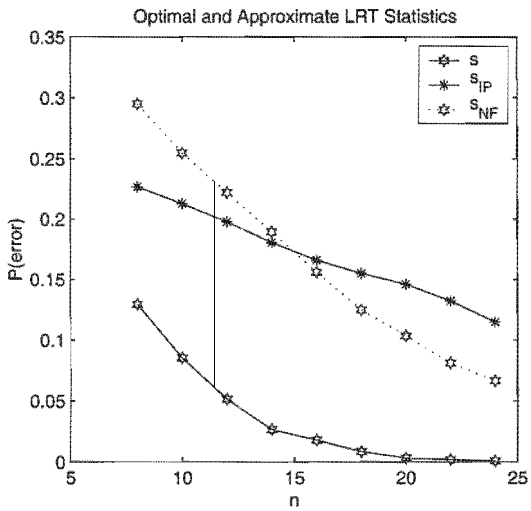
Image Size

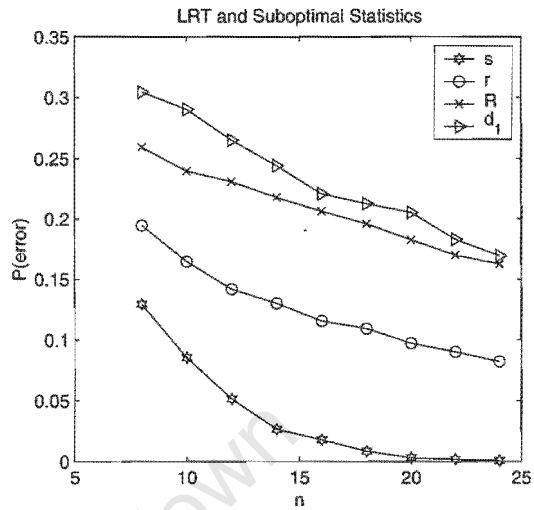
All of the statistics show better performance for larger images in Figure 6-3. This is expected, since for larger images more information is available to discriminate between matching and non-matching image pairs. One way of comparing the statistics is to find the image size that is required to keep the error rate below a certain limit. For example, if the upper bound $P(\text{error}) \leq 0.1$ is set in Figure 6-3(a), then the optimal LRT requires a 9×9 image, whereas the next best statistic, Pearson's r, requires a 21×21 image to match this performance.

Signal-to-Noise Ratio

Figure 6-4 is a good illustration of the difference between the optimal and approximate LRT statistics. The penalty of the independence assumption is different for low and high match correlation coefficient. For the former (Figure 6-4(a)), the penalty is greater for increasing SNR, and vice versa for the latter (Figure 6-4(b)). In both cases the error rate of the LRT with the noise-free assumption is high for low SNR, but approaches the performance of the optimal LRT as the noise becomes less significant.

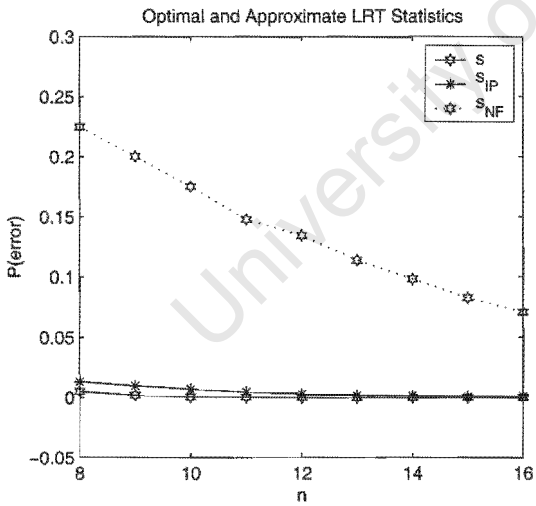
As with the image size experiment, the LRT outperforms the suboptimal measures by a wide margin. One interesting point is that the SSC error rate reaches a minimum and then begins to increase with increasing SNR. The SSC criterion relies on the presence of noise to produce sign changes in the difference image and when the noise has lower amplitude than other low frequency deviations between the images (due to the fact that $\rho_1 < 1$ in this experiment), the performance deteriorates. This is where the deterministic sign change

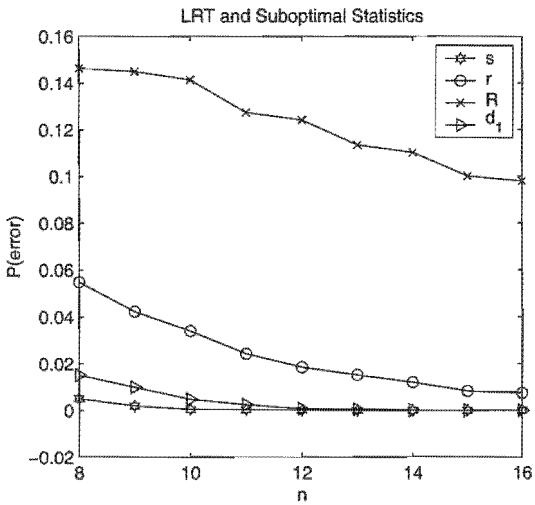




Parameter	*Value*
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
Scene one-step correlation (ρ)	0.9
SNR	2

(a) Low match correlation coefficient.





Parameter	*Value*
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.99
Scene one-step correlation (ρ)	0.9
SNR	2

(b) High match correlation coefficient.

Figure 6-3: Monte Carlo investigation of error rate versus image size ($T_0 = T_1 = 5000$).

(DSC, see Section 2.2.2) would be a better nonparametric measure. The exact point that DSC should be introduced is at the SSC error rate minimum on Figure 6-4, where SNR ≈ 1.2.

Spatial Image Correlation

The effect of spatial correlation in the scene (non-noise) component of the images, represented here by the one-step spatial correlation coefficient ρ, is investigated in Figure 6-5. All of the measures, except the SSC, have increasing error with increasing spatial correlation. As discussed in Section 5.3.3, this is due to the fact that high spatial correlation reduces the information content in the scene component of the image and leaves the similarity statistics little to work with. The performance of the SSC, on the other hand, is not dependent on the information content in the scene in the same way: the more spatial correlation there is, the more likely it is that the scene component will be subtracted perfectly, allowing the SSC to better analyze the sign changes in the difference image and thereby reducing the error rate.

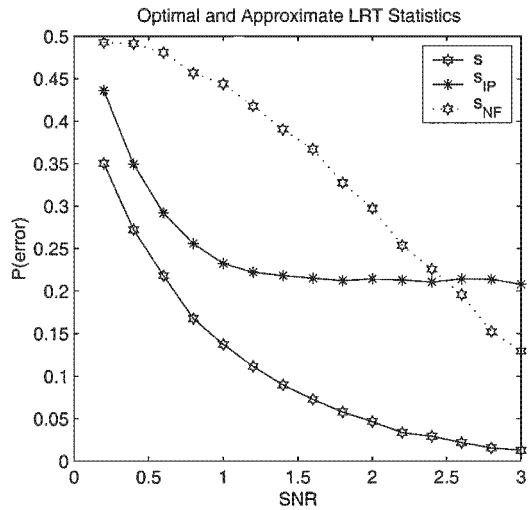
6.2.2 Match Correlation Coefficient

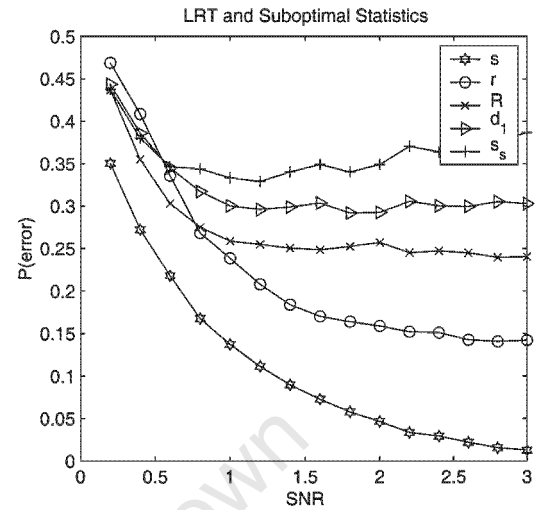
The effect of the match correlation coefficient ρ_1, which represents the correlation between corresponding pixels in the scene (non-noise) component of a matching image pair, is investigated in Figure 6-6. All measures have an increasing error rate as ρ_1 decreases. The optimal LRT has a significant performance advantage over the range: $\rho_1 \in [0.1, 0.9]$.

The experiment of Figure 6-6 assumes prior knowledge of ρ_1. The advantage of the LRT is that it offers an opportunity to incorporate this knowledge, but in practice the value of ρ_1 used to calculate the statistic is likely to be inaccurate. Figure 6-7 shows the results from an experiment that investigates incorrect knowledge of ρ_1. The LRT was used with a fixed value for ρ_1 that did not change with the value used to synthesize image pairs over the range of the experiment. In the graph on the left the LRT used $\rho_1 = 0.2$ and on the right it used $\rho_1 = 0.8$. It is evident from the results that the LRT is relatively insensitive to inaccuracies in ρ_1.

6.2.3 Unknown Parameters

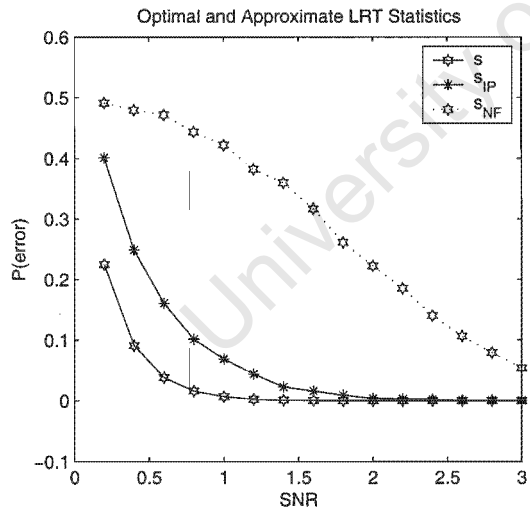
This section investigates the situation where the *a priori* knowledge required by the LRT is incomplete. This knowledge consists of the parameters in the joint image model, that is, the

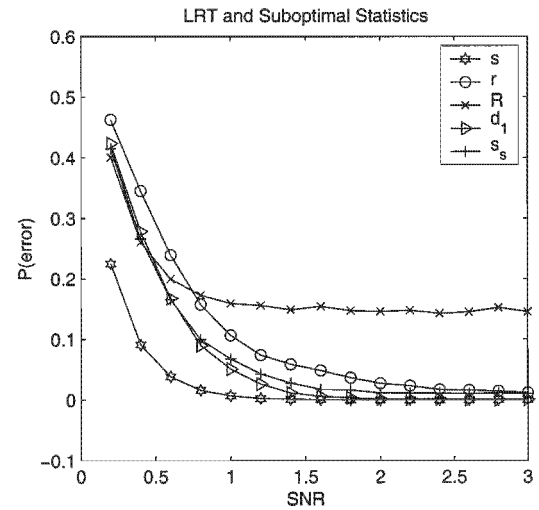




Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
Scene one-step correlation (ρ)	0.95

(a) Low match correlation coefficient.

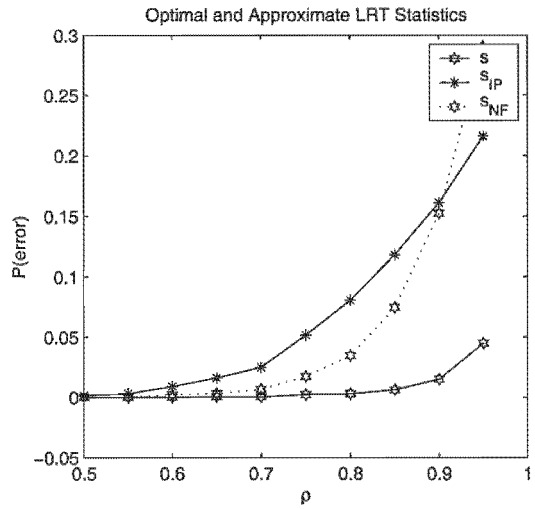


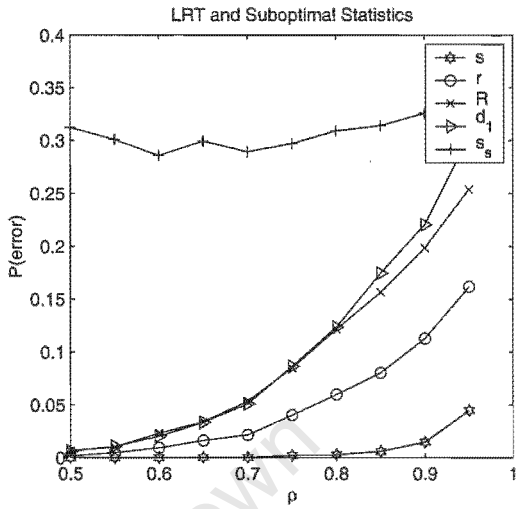


Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.99
Scene one-step correlation (ρ)	0.95

(b) High match correlation coefficient.

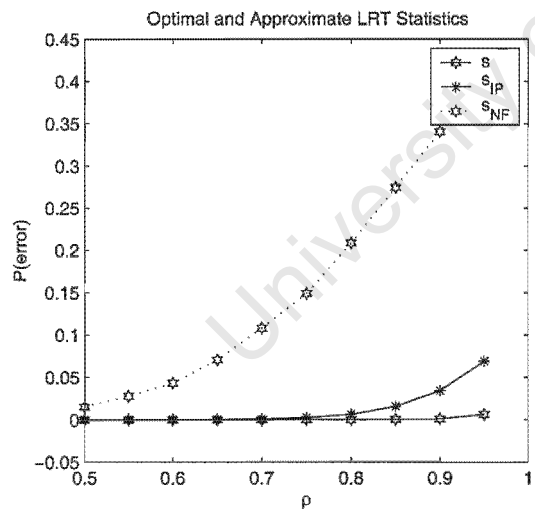
Figure 6-4: Monte Carlo investigation of error rate versus SNR ($T_0 = T_1 = 5000$).

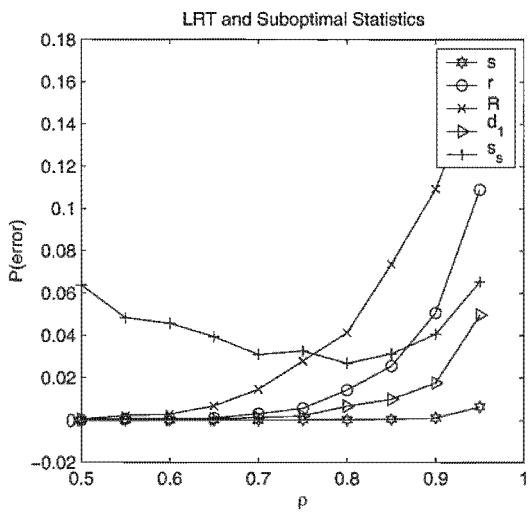




Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
SNR	2

(a) Low match correlation coefficient.

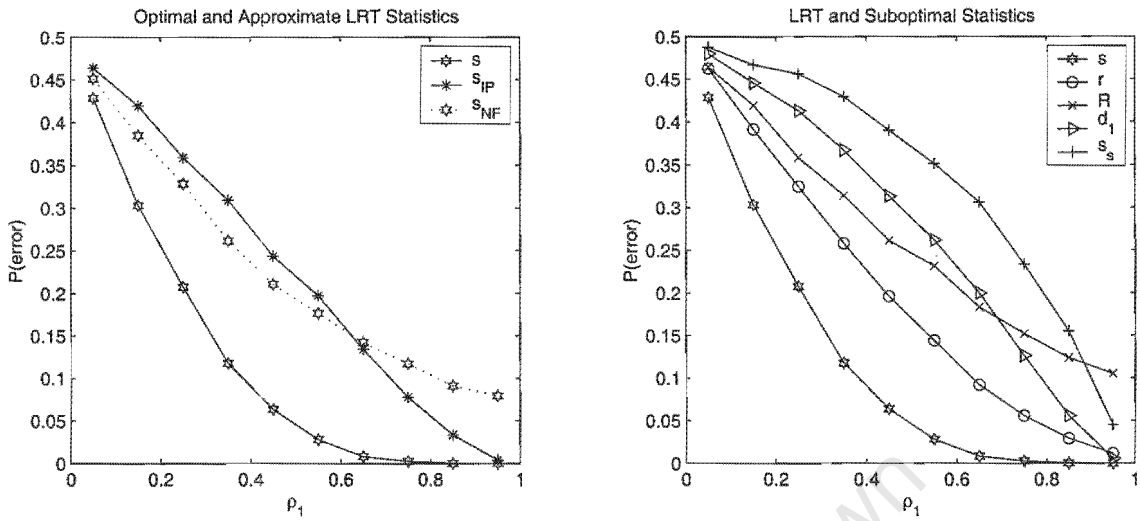




Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.99
SNR	1

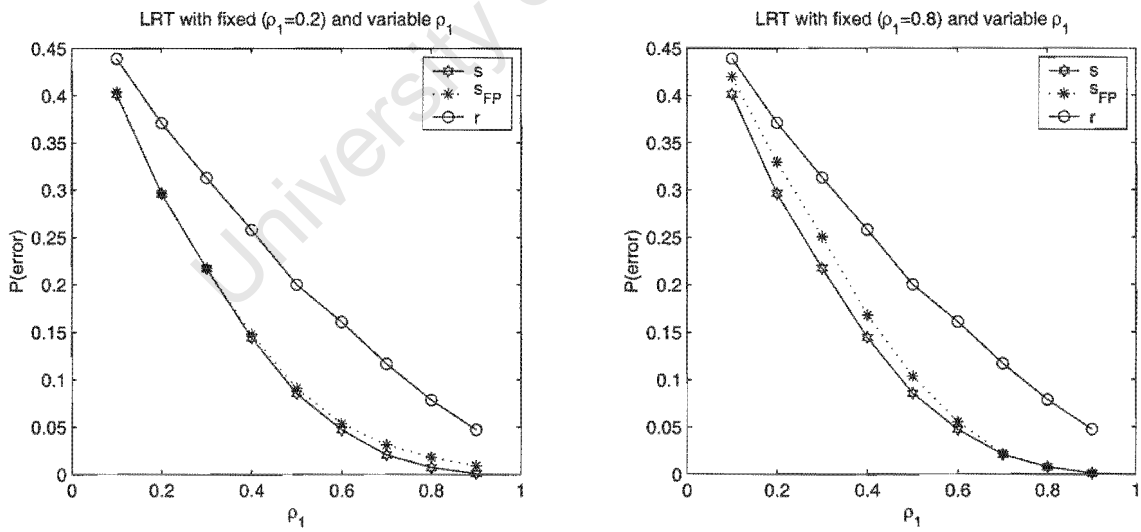
(b) High match correlation coefficient.

Figure 6-5: Monte Carlo investigation of error rate versus spatial correlation ($T_0 = T_1 = 5000$).



Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Scene one-step correlation (ρ)	0.9
SNR	2

Figure 6-6: Monte Carlo investigation of error rate versus match correlation coefficient ($T_0 = T_1 = 5000$).



Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Scene one-step correlation (ρ)	0.95
SNR	2

Figure 6-7: Monte Carlo investigation of error rate versus variable match correlation coefficient with fixed LRT parameters ($T_0 = T_1 = 5000$).

mean vector $\mathbf{m_w} = [\mathbf{m_a}, \mathbf{m_b}]^T$ and the covariance matrix

$$\mathbf{K_w} = \begin{bmatrix} \sigma_a^2\mathbf{R}+\sigma_\mu^2\mathbf{I} & \sigma_a\sigma_b\rho_{ab}\mathbf{R} \\ \sigma_a\sigma_b\rho_{ab}\mathbf{R} & \sigma_b^2\mathbf{R}+\sigma_\nu^2\mathbf{I} \end{bmatrix}. \tag{6.1}$$

Unknown parameters call for the use of the generalized LRT (GLRT) and experiments here contrast the performance of the GLRT with the performance of the LRT that knows the parameters. The unknown parameters are either characteristics of the individual images or are the match parameter itself. The former are referred to as nuisance parameters, and the latter leads to a composite match hypothesis.

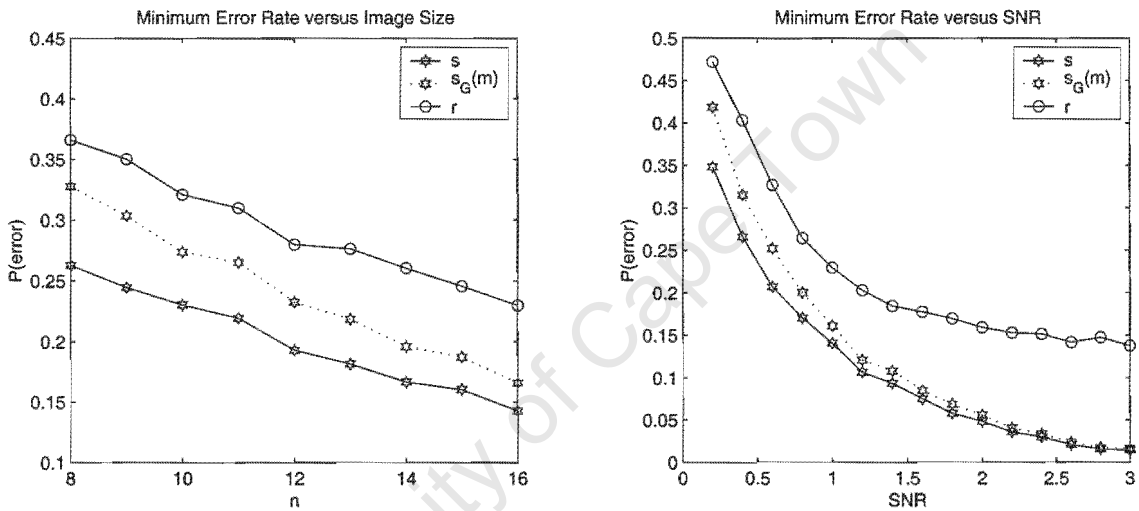
Image Offset Nuisance Parameter

The parameters $\{\mathbf{m_a}, \mathbf{m_b}, \sigma_a^2, \sigma_b^2, \rho, \sigma_\mu^2, \sigma_\nu^2\}$ are characteristics of the individual images and can often be estimated with training data. If not, they can be estimated from individual images, but only when the images are sufficiently large. Often it will be the case that illumination levels vary from image to image, resulting in an unpredictable scalar offset to the mean vectors $\mathbf{m_a}$ and $\mathbf{m_b}$. This offset, which may be different for each image, can be treated as a nuisance parameter. Figure 6-8 compares the error rate for the LRT where the offset is known to the GLRT that uses an estimate. Results for Pearson's r are also given, since this suboptimal statistic removes an estimate of the mean in its calculation, thereby providing some degree of offset invariance. The result shows that the GLRT sacrifices little performance in this case.

Composite Match Hypothesis

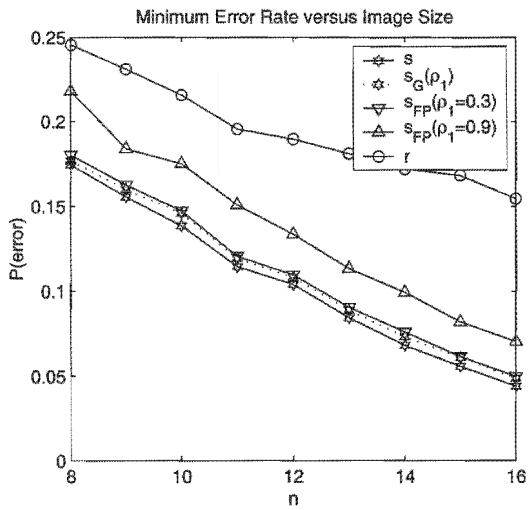
In Section 6.2.2 an experiment showed that the LRT was relatively tolerant of inaccuracies in the value of ρ_1 specified for the model. If this parameter is unknown, however, it may be better to use a generalized test than to make an educated guess about its true value. In Figure 6-9 the optimal LRT, the GLRT with an estimate of ρ_1, two LRTs with fixed ρ_1 (0.3 and 0.9), and suboptimal statistics are compared for a range of image size and SNR.

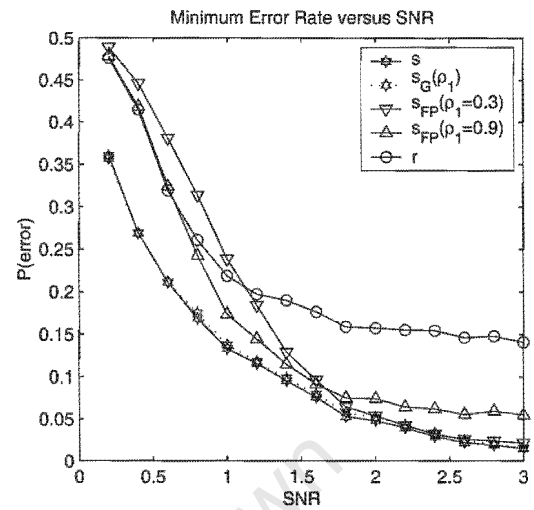
In Figure 6-9(a) the true match correlation coefficient is $\rho_1 = 0.6$. Here the GLRT is superior to the inaccurate guesses of 0.3 and 0.9 in both cases. The LRT with match correlation coefficient fixed at a value lower than the true value has fewer errors than the LRT with match correlation coefficient fixed at a value higher than the true value. For the most part, all of the LRT variations outperform Pearson's r.



Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
Scene one-step correlation (ρ)	0.95
SNR	1

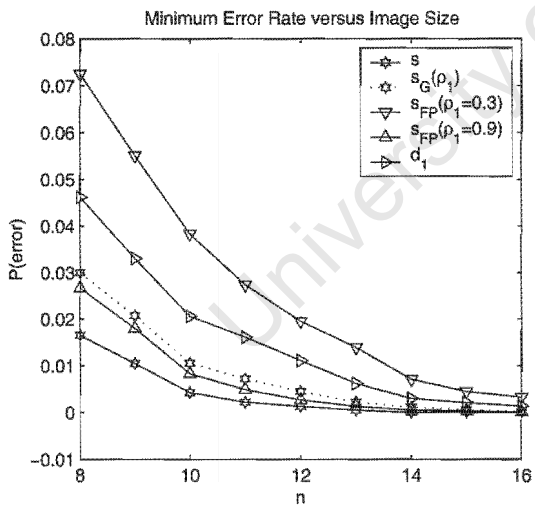
Figure 6-8: Monte Carlo investigation of the GLRT with image mean nuisance parameter ($T_0 = T_1 = 5000$).

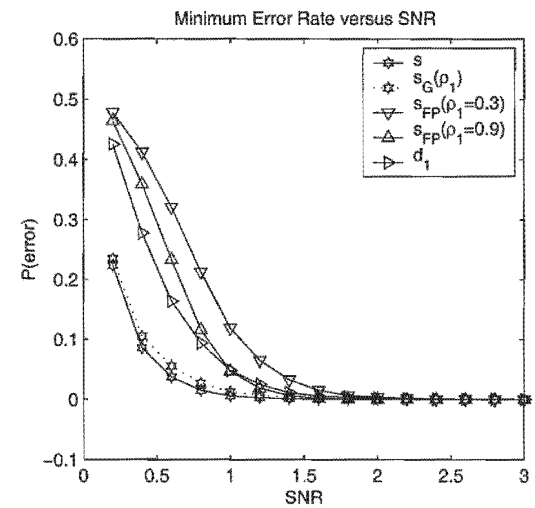




Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
Scene one-step correlation (ρ)	0.95
SNR	2

(a) Low match correlation coefficient.





Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.99
Scene one-step correlation (ρ)	0.95
SNR	2

(b) High match correlation coefficient.

Figure 6-9: Monte Carlo investigation of the GLRT with composite match hypothesis ($T_0 = T_1 = 5000$).

In Figure 6-9(b) the true match correlation coefficient is $\rho_1 = 0.99$. The sum of absolute differences is a better suboptimal measure in this regime and is therefore used for comparison instead of Pearson's r. On the whole, the GLRT is once again superior, although for high SNR, the LRT with ρ_1 fixed at 0.9 is marginally better than the GLRT. This is not entirely unexpected, since 0.9 is relatively close to the true value of 0.99. Despite the proximity to the true value, however, the guess of 0.9 has very poor relative performance for SNR $\lessapprox 1$.

In all results the GLRT is significantly better than the suboptimal statistic used for comparison. In summary, these results suggest that if there is uncertainty about the true value of the match correlation coefficient, then the GLRT should be employed.

6.3 Error Rate and Deviations from the Model

The previous experiments were performed on an ensemble of image pairs that conformed to the assumed model. Real images rarely conform exactly to such a model and the effect of deviations from the model are investigated in this section.

6.3.1 Sensitivity to Model Parameters

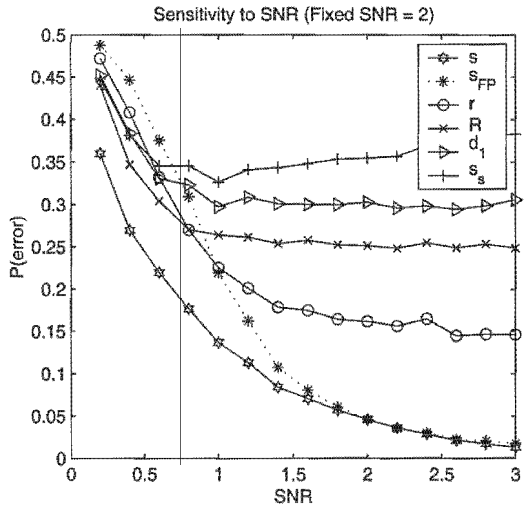
The performance of the LRT in the vicinity of the parameter values chosen for the image pair model is an indication of the robustness of the test. The LRT statistic (5.14) and decision threshold (5.15) summarize the parameters of the individual image models in the quantities

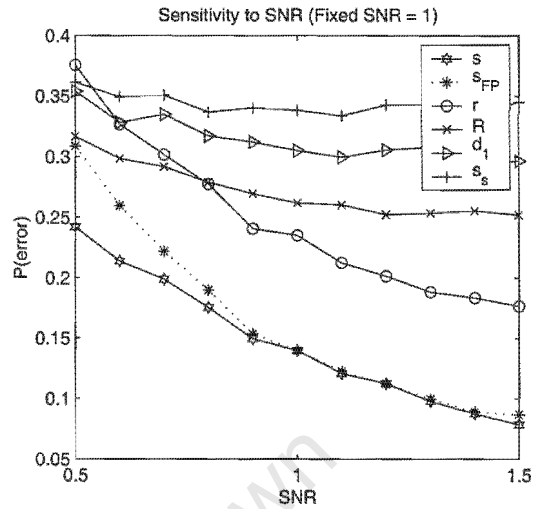
$$k_i = \frac{\omega_i}{\omega_i + \mathrm{SNR}^{-1}}$$

where $\mathrm{SNR} = \sigma^2/\sigma_\eta^2$. The quantities ω_i are eigenvalues of $\mathbf{R}$ and are determined by the one-step correlation coefficient of the Markov scene, denoted ρ. The model is therefore parameterized on SNR and ρ.

Figure 6-10(a) compares the error rate of the optimal LRT and several suboptimal statistics with that of the LRT where SNR is fixed. The result is an indication of the LRT sensitivity to inaccuracy in the SNR value used to calculate the test statistic and suggests that small inaccuracies will not have a seriously detrimental effect on the error rate. Figure 6-10(b) suggests that the same is true for small inaccuracies in the value of ρ used to calculate the LRT statistic. These particular graphs also suggest that overestimating the SNR is more detrimental than underestimating it, whereas it is better to overestimate ρ.

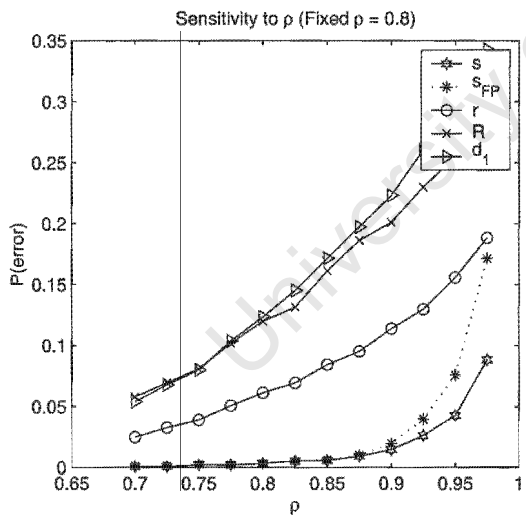
The experiment of Figure 6-10 explores the discriminating power of the test, but does

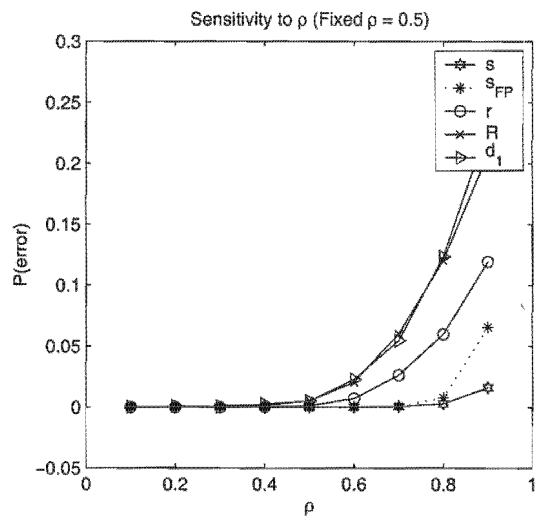




Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
Scene one-step correlation (ρ)	0.95

(a) LRT with SNR fixed.



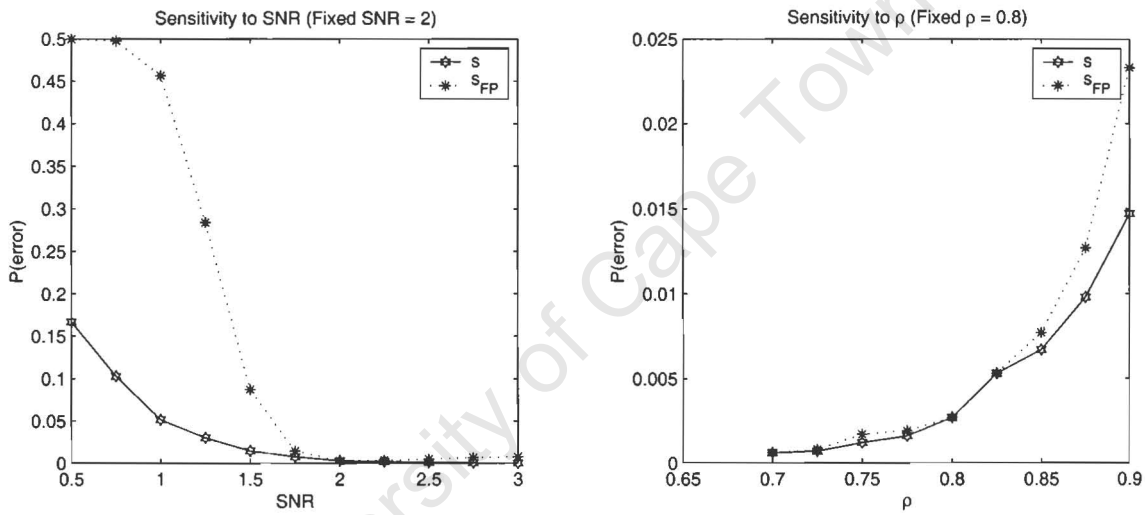


Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
SNR	2

(b) LRT with ρ fixed.

Figure 6-10: Monte Carlo investigation of error rate for the LRT with a fixed parameter ($T_0 = T_1 = 5000$).

not say anything abut the sensitivity of the decision threshold to inaccuracy in the model parameters. This is the case because the minimum error rate is estimated from the hypothesis conditional pdfs, and does not require that the threshold be specified. In practice, however, the decision threshold must be specified beforehand and an incorrect threshold will increase the error rate. Figure 6-11 shows the result of an experiment that used the theoretical ideal observer threshold given in Chapter 5. The error rate of an LRT that knows the correct model parameters is compared to an LRT with fixed model parameters. Once again, the test with fixed parameters is reasonably insensitive to parameter inaccuracies. Overestimating SNR by more than 0.5, however, might have a severe effect on the error rate.

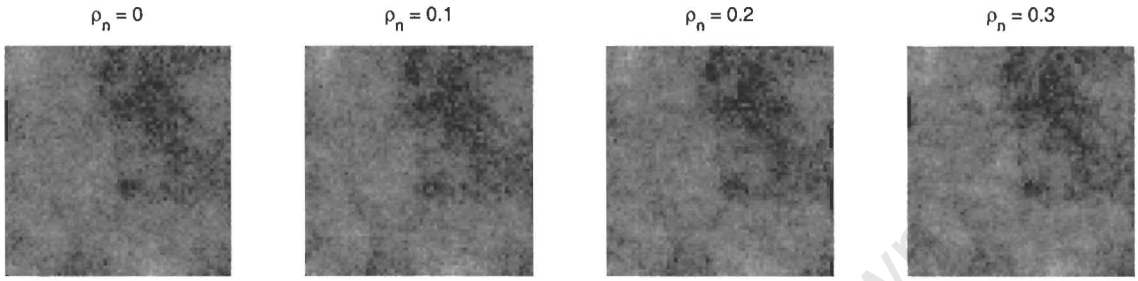


Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
Scene one-step correlation (ρ)	0.8
SNR	2

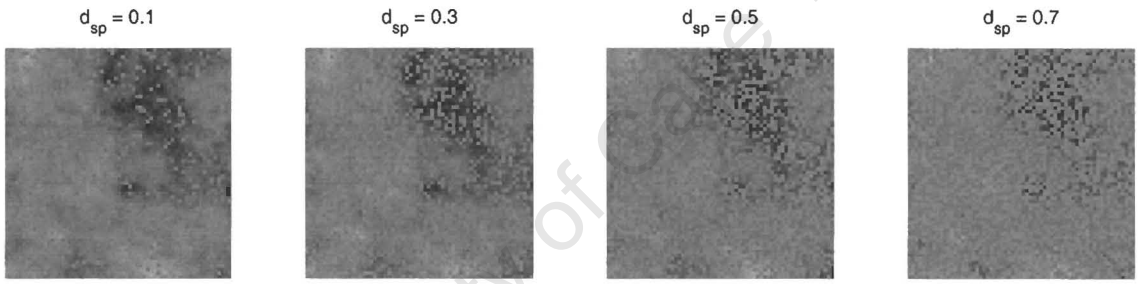
Figure 6-11: Monte Carlo investigation of LRT decision threshold sensitivity to model parameter inaccuracies ($T_0 = T_1 = 5000$).

6.3.2 Noise Deviations

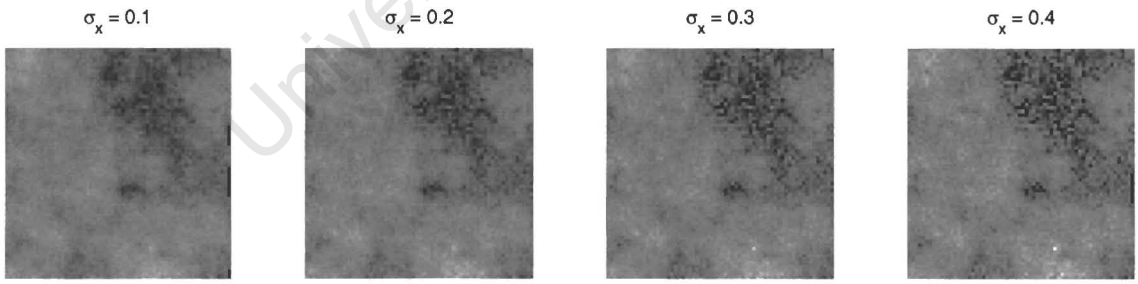
The assumed model caters for additive white noise. Real imaging systems also generate other types of noise and the effect of three common examples of these on matching performance are investigated here. They are correlated noise (or coloured noise), salt-and-pepper noise and multiplicative noise (or speckle).



(a) Additive noise of varying one-step correlation (SNR = 2).



(b) Salt and pepper noise for a range of noise density.



(c) Multiplicative noise for a range of standard deviation.

Figure 6-12: Synthesized images with varying degrees of noise.

Correlated Noise

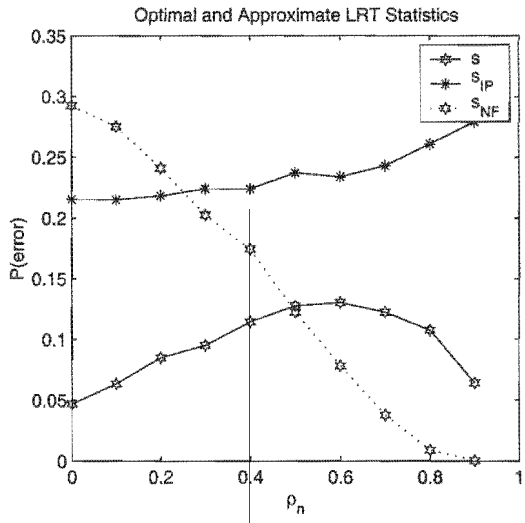
The proposed image-pair model assumes that the noise in distinct pixels is statistically independent, but often imaging systems introduce noise with slight correlation between adjacent pixels. Figure 6-12(a) shows examples of correlated noise, where it was modelled as a nonseparable first-order Markov field with one-step correlation coefficient ρ_n. Figure 6-13(a) graphs the error rate over a range of values for ρ_n. The experiment shows that the error rate performance of the LRT is not seriously affected by spatial correlation in the noise component. For a noise correlation coefficient of $\rho_n \gtrapprox 0.5$, the noise-free LRT special case outperforms the optimal LRT, but this degree of noise correlation is very uncommon in imaging systems.

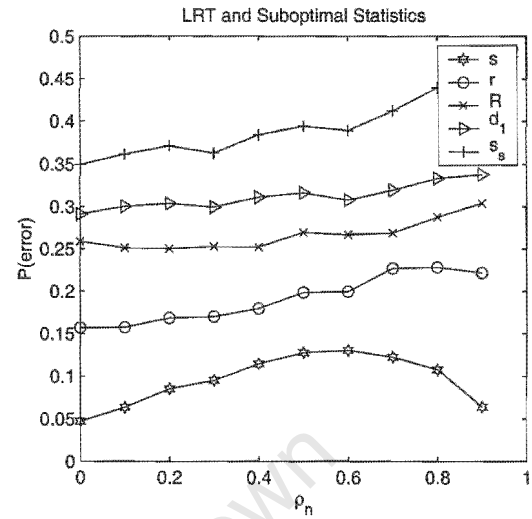
Salt-and-Pepper Noise

Another common variety of image corruption is so-called salt-and-pepper noise, examples of which are shown in Figure 6-12(b) for a range of noise density, d_{sp}. Figure 6-14 shows the error rate obtained from Monte Carlo experiments that were performed for a range of noise density values. It is evident that the error rate advantage of the LRT disappears for $d_{sp} \gtrapprox 0.1$. As Figure 6-12(b) shows, however, values this high correspond to very extreme conditions of salt-and-pepper noise.

Multiplicative Noise

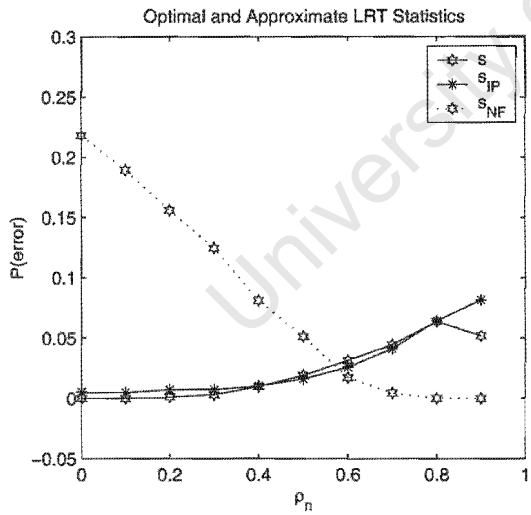
Each individual pixel in the assumed model can be written as $u = a + \mu$ where a is the scene component and μ is the noise component of the pixel intensity value. A more general model that also incorporates multiplicative noise can be written as $u = \mu_\times a + \mu_+$, where $\mu_\times$ is the multiplicative noise component. A normal random variable with unity mean and standard deviation $\sigma_\times$ is a reasonable model for $\mu_\times$. Figure 6-12(c) shows multiplicative noise for several values of $\sigma_\times$. Figure 6-15 graphs error rate performance for $\sigma_\times \in [0, 1]$. The LRT maintains superiority for a reasonable degree of multiplicative noise — if the multiplicative SNR is defined as $\text{SNR}_\times = \frac{\sigma}{\sigma_\times}$, then superiority is maintained for $\text{SNR}_\times \gtrapprox 2$. The results also suggest that the SSC criterion is fairly robust in the presence of multiplicative noise. This is particularly true for the high match correlation case, where it is almost impervious to increasing noise variance.

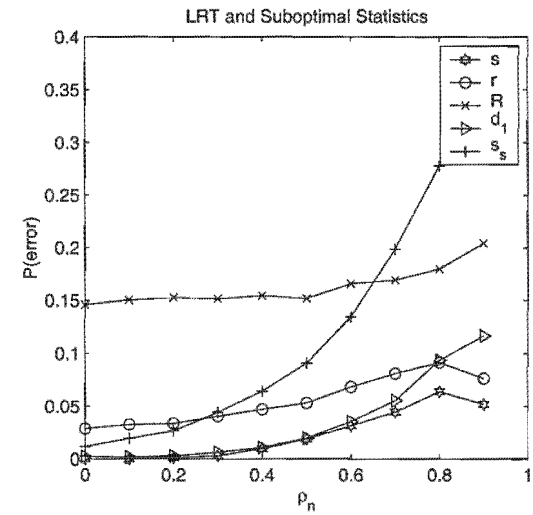




Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
Scene one-step correlation (ρ)	0.95
SNR	2

(a) Low match correlation coefficient.

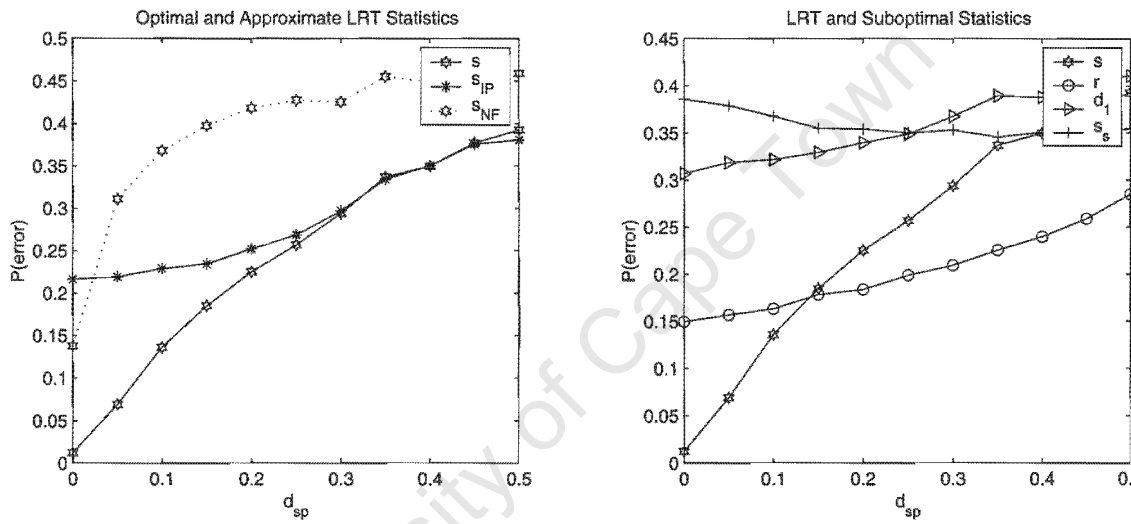




Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.99
Scene one-step correlation (ρ)	0.95
SNR	2

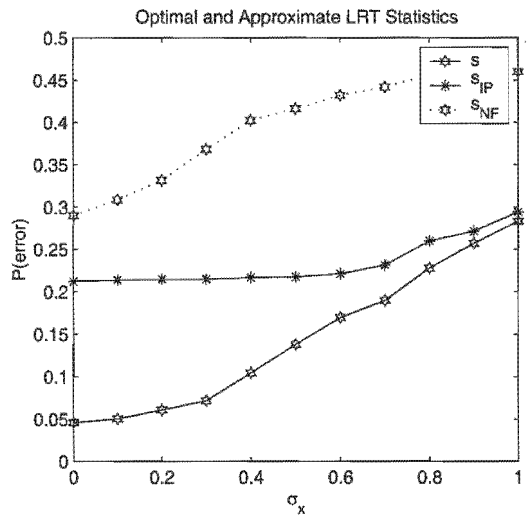
(b) High match correlation coefficient.

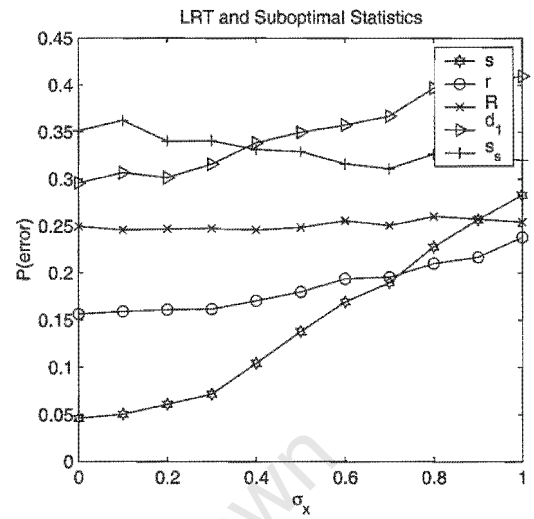
Figure 6-13: Monte Carlo investigation of error rate versus spatial noise correlation ($T_0 = T_1 = 5000$).



Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
Scene one-step correlation (ρ)	0.95
SNR	3

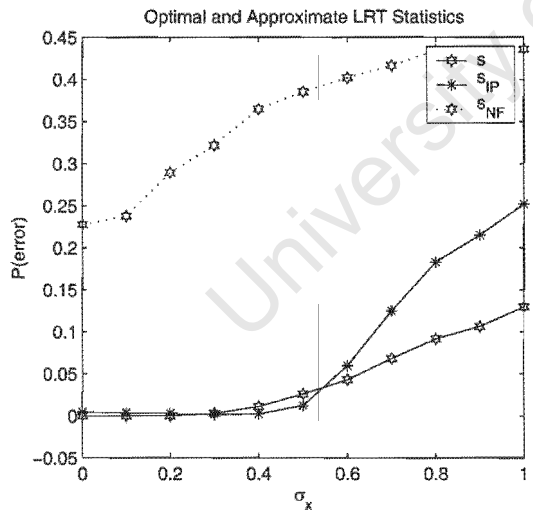
Figure 6-14: Monte Carlo investigation of error rate versus salt-and-pepper noise density ($T_0 = T_1 = 5000$).

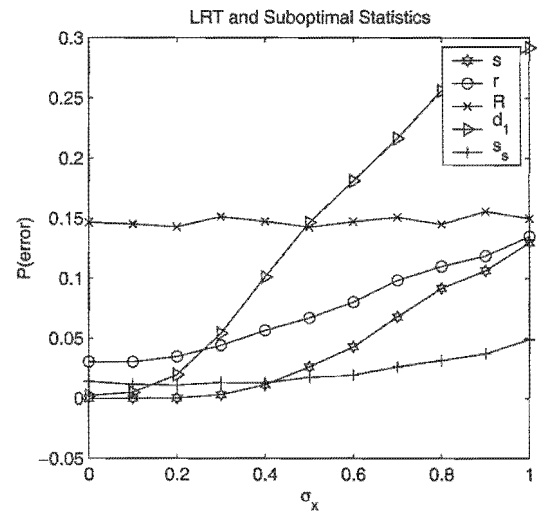




Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
Scene one-step correlation (ρ)	0.95
SNR	2

(a) Low match correlation coefficient.





Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.99
Scene one-step correlation (ρ)	0.95
SNR	2

(b) High match correlation coefficient.

Figure 6-15: Monte Carlo investigation of error rate versus the standard deviation of a multiplicative noise component ($T_0 = T_1 = 5000$).

6.3.3 Occlusion

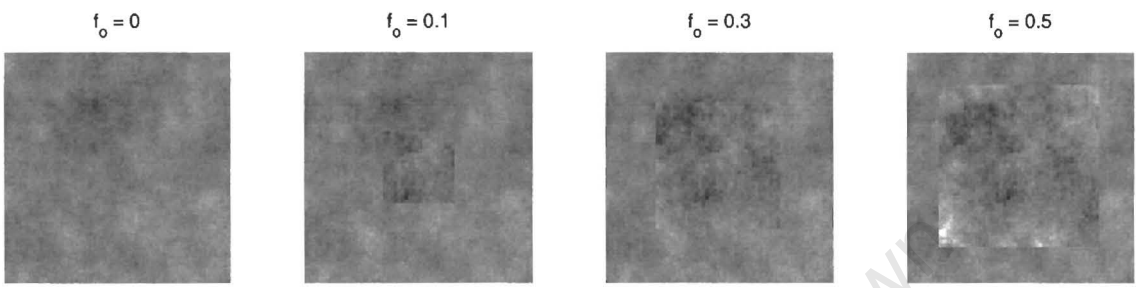
A more serious deviation from the assumed model occurs when the scene is occluded by another object in either or both of the images in the pair. Some similarity statistics, like the SSC criterion, have been motivated by the need for robust operation in the presence of this sort of deviation [42]. The experiments here characterize occlusion in terms of (1) the fraction of image area occluded f_o, (2) the occlusion-to-signal ratio (OSR), which is defined as the ratio of the pixel intensity standard deviation in the occluded portion to the standard deviation in the original image, and (3) whether the occlusion is opaque (reflected radiation imaging) or additive (attenuated radiation imaging). Figure 6-16 shows examples of images with different degrees of occlusion.

The results of experiments show that the LRT is the least robust method in the presence of occlusion. Figure 6-17(a) graphs error rate for additive and opaque occlusion over the occlusion fraction f_o. The suboptimal measures outperform the LRT, with the nonparametric SSC criterion proving to be the most robust statistic as should be expected. Over a range of OSR in Figure 6-17(b) the same trend is observed. Here the true advantage of the SSC criterion is observed — it is invariant to the increasing energy in the occluded part of the image (i.e. the increasing OSR) because it does not take the pixel intensity values directly into account.

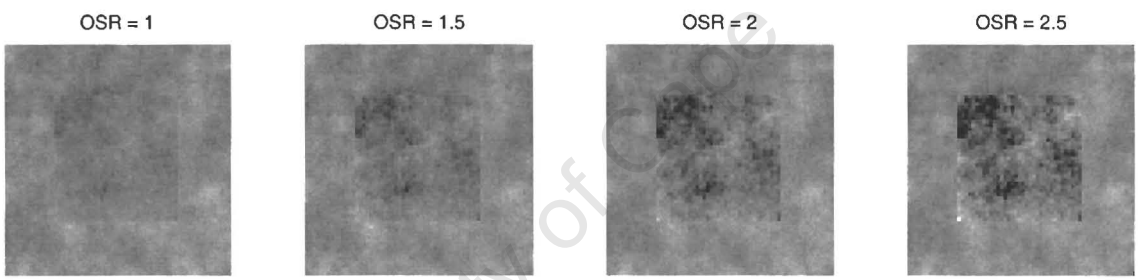
Huber describes near optimal performance under the classical model as one of the desirable characteristics of a robust procedure [38, p. 5]. Strictly speaking, the SSC criterion is nonparametric rather than robust, but the same principle can be applied. So even if the optimal LRT is unsuitable in a particular situation because of outliers, it can be used as a benchmark for performance under the classical model. For example, the experiments in this chapter suggest that although the SSC criterion is effective in the presence of occlusion, it sacrifices significant performance in comparison to the optimal error rate without occlusion. There would therefore appear to be scope for developing a test that exhibits close-to-optimal performance under the classical model, but is robust in the presence of occlusion.

6.4 Discussion

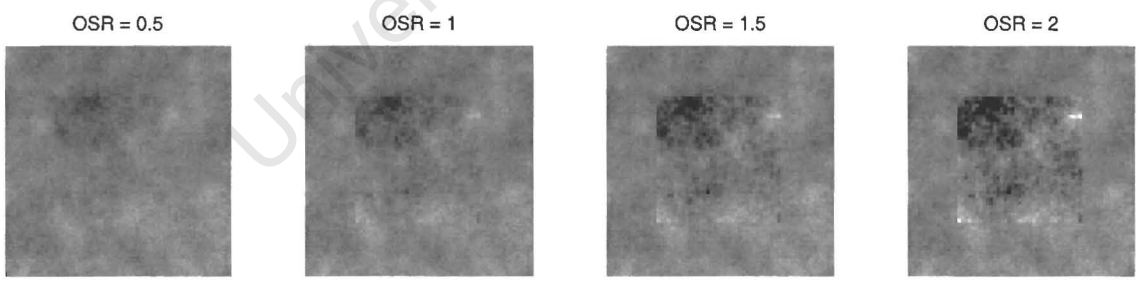
This chapter has demonstrated the utility of Monte Carlo simulation methods for evaluating matching techniques. Image-pair synthesis equations that were derived in Chapter 4 were used to generate ensembles of matching and non-matching image pairs. The performance



(a) Opaque occlusion for a range of occlusion fraction.

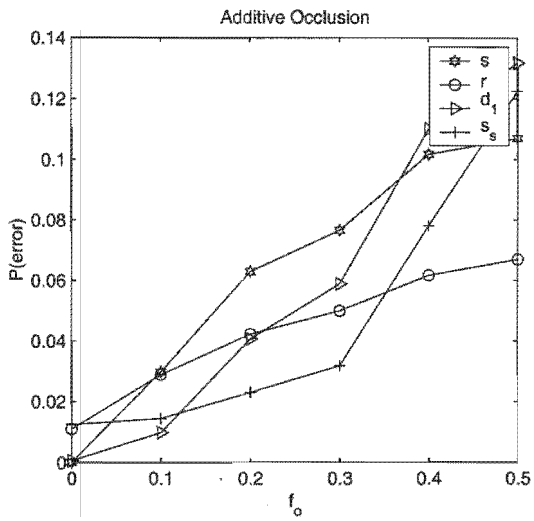


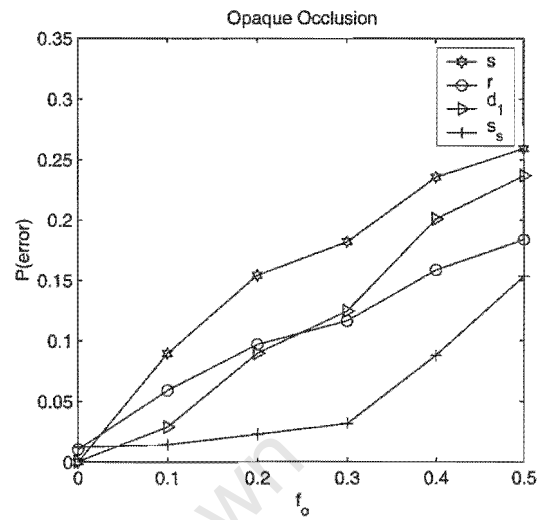
(b) Opaque occlusion for a range of occlusion-to-signal ratio.



(c) Additive occlusion for a range of occlusion-to-signal ratio.

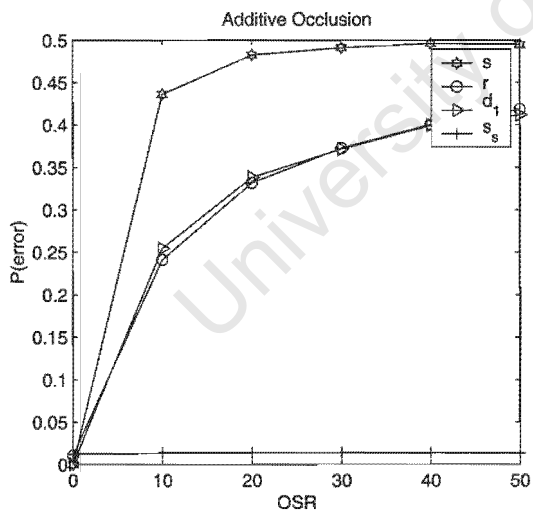
Figure 6-16: Synthesized images with varying degrees of occlusion.

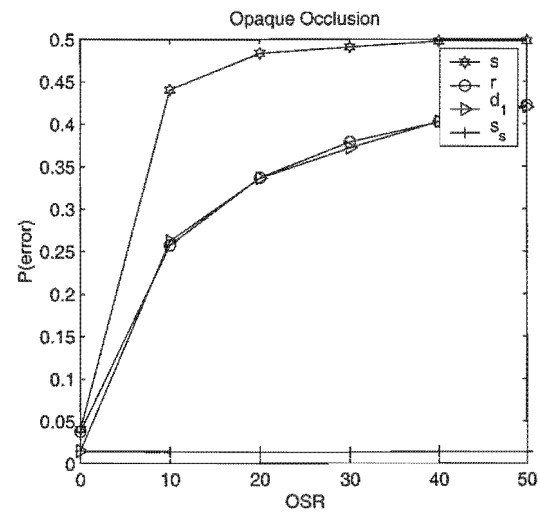




Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.99
Scene one-step correlation (ρ)	0.95
SNR	3

(a) Error rate versus occlusion fraction with OSR = 1.





Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.99
Scene one-step correlation (ρ)	0.95
SNR	3

(b) Error rate versus occlusion-to-signal ratio with $f_o = 0.1$.

Figure 6-17: Monte Carlo investigation of the effect of occlusion on the matching error rate ($T_0 = T_1 = 50000$).

of the standard similarity statistics could then be compared with each other and with the model-optimal LRT that was derived in Chapter 5. These comparisons were made on the basis of the ROC plot and the minimum error rate of an ideal observer test.

The results for image pairs that conform to the model of Chapter 4 demonstrate the superiority of the LRT, but since this test was derived to be optimal under the assumed model, it is not surprising that it outperforms suboptimal approaches. The more significant result is the extent of superiority over other tests that the LRT exhibits, which indicates that there is much potential for improvement on current methods. Among the suboptimal statistics, a general rule emerges: correlation-based statistics are better for image pairs with low match correlation coefficient and difference-based statistics are better when the match correlation coefficient is close to unity.

Results also suggest that the LRT is relatively insensitive to inaccurate knowledge of the model parameters. Where there is significant uncertainty about the true value of a parameter, however, the GLRT provides better performance than the LRT that uses an educated guess of the parameter value. Other deviations from the proposed model were also investigated. For correlated, salt-and-pepper and multiplicative noise, the LRT retains its performance advantage under normal imaging conditions. The LRT is very sensitive to occlusion, however, and it is here that the advantage of a nonparametric measure, such as the stochastic sign change, becomes evident. Finally, it appears that the noise-free and independent-pixel assumptions, which are widely used in analyses of image processing algorithms, seriously degrade performance if they are not warranted.

On the basis of the results presented in this chapter, it appears that the LRT will be effective under a wide variety of imaging conditions. The emphasis now shifts to the practical aspects of implementing the test, and the next chapter introduces efficient methods for computing the LRT statistic.

Chapter 7

Efficient Implementation of the Optimal Test

Chapter 5 developed an optimal hypothesis test for direct image matching that exhibits markedly improved matching performance in terms of error rate, but has significantly higher computational complexity than the standard approaches. This chapter introduces methods that can reduce the computational requirements so that they are comparable to those of the standard similarity statistics.

The LRT statistic is calculated in two stages: an $O\left(n^4\right)$ whitening transform on the images and an $O\left(n^2\right)$ calculation of the LRT statistic for the whitened images. One strategy for economizing on computation is to preprocess the images with the whitening transform. In the case of an image database, the transformed image could be calculated for each new addition to the database and stored along with, or, since the transform is invertible, instead of the new image. Where the application is image registration, the subimages around all control points and positions in the corresponding search area could be whitened before performing local block matching[1]. This approach rearranges the calculations to enhance efficiency, but still computes the exact LRT statistic. An alternative is to sacrifice optimal performance and use an approximate statistic that significantly reduces computation. The "lossy" approach is considered in this chapter.

The relationship between the optimal LRT statistic and the canonical inter-image correlation structure inherent in the image-pair model is an important source of inspiration when

[1]This approach may require excessive storage if an exhaustive search is used, since the whitened subimage around each position in the search area must be stored.

finding efficient implementation strategies. Section 7.1 makes this relationship explicit and discusses its consequences. The canonical variables are then used in Section 7.2 to reduce the dimensionality required for testing match in the image pair. Section 7.3 reduces computation in the whitening transform by using simplified image models, and Section 7.4 proposes a practical method for calculating the LRT with large images. A general discussion closes the chapter.

7.1 The LRT in Terms of Canonical Variables

Canonical correlation analysis was introduced in Chapter 4, where it was shown that if the scene component of the image pair model, denoted here as $\mathbf{c} = [\mathbf{a}^T, \mathbf{b}^T]^T$, has the covariance matrix

$$\mathbf{K}_\mathbf{c} = \begin{bmatrix} \sigma_a^2\mathbf{R} & \sigma_a\sigma_b\rho_{ab}\mathbf{R} \\ \sigma_a\sigma_b\rho_{ab}\mathbf{R} & \sigma_b^2\mathbf{R} \end{bmatrix},$$

then the correlation structure of the scene-pair can be reduced to n^2 canonical correlation coefficients with values ρ_{ab}^2. Recall that the canonical correlation coefficients of the n^2-vectors $\mathbf{a}$ and $\mathbf{b}$ are the correlation coefficients between a series of n^2 random variable pairs that are obtained by orthogonal transformations of $\mathbf{a}$ and $\mathbf{b}$. The first pair of transformations, say $p_1 = \boldsymbol{\alpha}_1^T\mathbf{a}$ and $q_1 = \boldsymbol{\beta}_1^T\mathbf{b}$, give the first canonical variables, which have maximal correlation and unit variance. The second set of canonical variables, $p_2 = \boldsymbol{\alpha}_2^T\mathbf{a}$ and $q_2 = \boldsymbol{\beta}_2^T\mathbf{b}$, have maximal correlation and unit variance, subject to the condition that they are uncorrelated with p_1 and q_1, and so on.

It is now instructive to derive the canonical variables and corresponding canonical correlation coefficients of the image pair with additive noise. Their relationship with the principal components of the individual images is explained, and it is shown that the LRT of Chapter 5 is already conveniently represented in terms of the canonical variables.

7.1.1 Image-Pair Canonical Variables

Expressions for the canonical variables and the canonical correlation coefficients of an image pair under the model of Chapter 4 are derived here. In Chapter 5 it was shown that a

whitening transform on the individual images leads to the joint image-pair covariance matrix

$$\mathbf{K}_{\acute{\mathbf{w}}} = \begin{bmatrix} \mathbf{I} & \mathbf{D} \\ \mathbf{D} & \mathbf{I} \end{bmatrix}. \tag{7.1}$$

Now this happens to be a canonical form for $\mathbf{K}_{\mathbf{w}}$ under the group of transformations $\mathbf{K}_{\mathbf{w}} \longrightarrow \mathbf{T}\mathbf{K}_{\mathbf{w}}\mathbf{T}$ [29, p. 550], where corresponding pixels in the whitened images are the canonical variables and the squares of the diagonal elements of $\mathbf{D}$ are the canonical correlation coefficients. Stated differently, it is evident that applying a whitening transform to each image independently, transforms the image pair into a canonical form for the correlation structure between the images.

Using the notation introduced in Section 5.2[2], the canonical variables are the corresponding pixels in the whitened images $\acute{\mathbf{u}} = \Lambda_{\mathbf{u}}^{-\frac{1}{2}}\mathbf{V}^T\mathbf{u}$ and $\acute{\mathbf{v}} = \Lambda_{\mathbf{v}}^{-\frac{1}{2}}\mathbf{V}^T\mathbf{v}$. The i-th pair of canonical variables is therefore $\{\acute{u}_i, \acute{v}_i\}$, where

$$\acute{u}_i = (\lambda_i^{\mathbf{u}})^{-\frac{1}{2}} [\mathbf{V}]_i^T \mathbf{u} \quad \text{and} \quad \acute{v}_i = (\lambda_i^{\mathbf{v}})^{-\frac{1}{2}} [\mathbf{V}]_i^T \mathbf{v}, \tag{7.2}$$

and $[\mathbf{V}]_i$ is the i-th column vector of $\mathbf{V}$. The i-th canonical correlation coefficient is $\mathbf{D}^2[i,i] = \rho_i^2$, where

$$\rho_i^2 = k_i^2\rho_{ab}^2 = \frac{\sigma_a^2\sigma_b^2\omega_i^2\rho_{ab}^2}{\left(\sigma_a^2\omega_i + \sigma_\mu^2\right)\left(\sigma_b^2\omega_i + \sigma_\nu^2\right)}. \tag{7.3}$$

One can also derive the canonical correlation coefficients of the image pair from the joint covariance matrix. This is now done for the sake of completeness. If the image pair has the partitioned joint covariance matrix

$$\mathbf{K}_{\mathbf{w}} = \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{bmatrix},$$

then the canonical correlation coefficients are the eigenvalues of $\mathbf{K}_{11}^{-1}\mathbf{K}_{12}\mathbf{K}_{22}^{-1}\mathbf{K}_{21}$ [29, p. 550].

[2]Recall from Chapter 5, Section 5.2, that $\mathbf{V}$ is an orthogonal matrix with columns that are the eigenvectors of $\mathbf{R}$, the correlation coefficient matrix shared by images $\mathbf{u}$ and $\mathbf{v}$. The diagonal matrices $\Lambda_{\mathbf{u}}$ and $\Lambda_{\mathbf{v}}$ have the eigenvalues of $\mathbf{u}$, denoted $\lambda_i^{\mathbf{u}}$, and the eigenvalues of $\mathbf{v}$, denoted $\lambda_i^{\mathbf{v}}$, as their i-th diagonal elements respectively. The i-th column vector of $\mathbf{V}$ is the eigenvector associated with $\lambda_i^{\mathbf{u}}$ and $\lambda_i^{\mathbf{v}}$.

Noting that $\mathbf{K_w} = \acute{\mathbf{T}}_\mathbf{w}\mathbf{K}_{\acute{\mathbf{w}}}\acute{\mathbf{T}}_\mathbf{w}^T$ where

$$\acute{\mathbf{T}}_\mathbf{w} = \begin{bmatrix} \mathbf{T}_\mathbf{u}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_\mathbf{v}^{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{V}\Lambda_\mathbf{u}^{\frac{1}{2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}\Lambda_\mathbf{v}^{\frac{1}{2}} \end{bmatrix} \text{ and } \acute{\mathbf{T}}_\mathbf{w}^T = \begin{bmatrix} \Lambda_\mathbf{u}^{\frac{1}{2}}\mathbf{V}^T & \mathbf{0} \\ \mathbf{0} & \Lambda_\mathbf{v}^{\frac{1}{2}}\mathbf{V}^T \end{bmatrix}, \tag{7.4}$$

the joint image-pair covariance matrix can be written as

$$\begin{aligned} \mathbf{K_w} &= \acute{\mathbf{T}}_\mathbf{w}\mathbf{K}_{\acute{\mathbf{w}}}\acute{\mathbf{T}}_\mathbf{w}^T \\ &= \begin{bmatrix} \mathbf{V}\Lambda_\mathbf{u}^{\frac{1}{2}}\Lambda_\mathbf{u}^{\frac{1}{2}}\mathbf{V}^T & \mathbf{V}\Lambda_\mathbf{u}^{\frac{1}{2}}\mathbf{D}\Lambda_\mathbf{v}^{\frac{1}{2}}\mathbf{V}^T \\ \mathbf{V}\Lambda_\mathbf{v}^{\frac{1}{2}}\mathbf{D}\Lambda_\mathbf{u}^{\frac{1}{2}}\mathbf{V}^T & \mathbf{V}\Lambda_\mathbf{v}^{\frac{1}{2}}\Lambda_\mathbf{v}^{\frac{1}{2}}\mathbf{V}^T \end{bmatrix} \end{aligned}$$

The canonical correlation coefficients are then the eigenvalues of

$$\begin{aligned} \mathbf{K}_{11}^{-1}\mathbf{K}_{12}\mathbf{K}_{22}^{-1}\mathbf{K}_{21} &= \left(\mathbf{V}\Lambda_\mathbf{u}\mathbf{V}^T\right)^{-1}\left(\mathbf{V}\Lambda_\mathbf{v}^{\frac{1}{2}}\mathbf{D}\Lambda_\mathbf{u}^{\frac{1}{2}}\mathbf{V}^T\right)\left(\mathbf{V}\Lambda_\mathbf{v}\mathbf{V}^T\right)^{-1}\left(\mathbf{V}\Lambda_\mathbf{v}^{\frac{1}{2}}\mathbf{D}\Lambda_\mathbf{u}^{\frac{1}{2}}\mathbf{V}^T\right) \\ &= \mathbf{V}\Lambda_\mathbf{u}^{-1}\Lambda_\mathbf{v}^{\frac{1}{2}}\mathbf{D}\Lambda_\mathbf{u}^{\frac{1}{2}}\Lambda_\mathbf{v}^{-1}\Lambda_\mathbf{v}^{\frac{1}{2}}\mathbf{D}\Lambda_\mathbf{u}^{\frac{1}{2}}\mathbf{V}^T \\ &= \mathbf{V}\mathbf{D}^2\mathbf{V}^T. \end{aligned}$$

If $\mathbf{A}$ is a nonsingular square matrix and $\mathbf{B} = \mathbf{C}^{-1}\mathbf{A}\mathbf{C}$, then $\mathbf{A}$ and $\mathbf{B}$ have the same eigenvalues [29, p. 583]. Therefore, $\mathbf{V}\mathbf{D}^2\mathbf{V}^T$ and $\mathbf{D}^2$ have the same eigenvalues. The eigenvalues of a diagonal matrix are the elements on the diagonal [29, p. 550], and so the canonical correlation coefficients, denoted ρ_i^2, are the squares of the elements on the diagonal of $\mathbf{D}$ as required.

7.1.2 Principal Components and Canonical Variables

Principal component analysis (PCA) is an important technique in statistical data analysis and has also found many applications in image processing. The principal components of an image are the result of a coordinate transformation of the image vector that is based on the image covariance matrix. This transform has optimal variance properties: the first principal component is the linear combination of the image pixel values that has maximal variance. The second principal component is the linear combination of the pixel values that has maximal variance and satisfies the condition of being uncorrelated with the first principal component, and so on for $i \in \{3, 4, \ldots, n^2\}$. Statistical treatments of PCA are given by Anderson [96, ch. 11] and Muirhead [29, ch. 9]. The optimal compaction of variance into a subset of the principal components has made PCA useful in signal and image processing applications, where it is often referred to as the Karhunen-Loève (KL) transform [8, p. 163].

Using the notation of the previous section, the KL transform of $\mathbf{u}$ is $\mathbf{V}^T\mathbf{u}$. The principal components are therefore given by $\dot{u}_i = [\mathbf{V}]_i^T\,\mathbf{u}$ for $i \in \{1, 2, \ldots, n^2\}$. Similarly, the principal components of $\mathbf{v}$ are $\dot{v}_i = [\mathbf{V}]_i^T\,\mathbf{v}$ for $i \in \{1, 2, \ldots, n^2\}$. Assume that $\mathbf{V}$, $\Lambda_\mathbf{u}$, and $\Lambda_\mathbf{v}$ are constructed in such a way that $\dot{u}_i$ and $\dot{v}_i$ are the i-th principal components of $\mathbf{u}$ and $\mathbf{v}$ (i.e. they have the i-th highest variance among all the components). Then the eigenvalues of $\mathbf{u}$ and $\mathbf{v}$, $\lambda_i^\mathbf{u}$ and $\lambda_i^\mathbf{v}$, which are the variance of $\dot{u}_i$ and $\dot{v}_i$ respectively, are ordered with decreasing value for increasing i. Writing the eigenvalues of $\mathbf{R}$ in terms of these quantities (see B.4),

$$\omega_i = \frac{\lambda_i^\mathbf{u} - \sigma_\mu^2}{\sigma_a^2} \quad \text{or} \quad \omega_i = \frac{\lambda_i^\mathbf{v} - \sigma_\nu^2}{\sigma_b^2}. \tag{7.5}$$

Note that both of the expressions in (7.5) enforce an ordering of ω_i with decreasing magnitude.

Referring to (7.2), it is seen that the canonical variable pair $\{\acute{u}_i, \acute{v}_i\}$ can be written in terms of the principal components as follows:

$$\begin{aligned} \acute{u}_i &= (\lambda_i^\mathbf{u})^{-\frac{1}{2}}\,[\mathbf{V}]_i^T\,\mathbf{u} \\ &= (\lambda_i^\mathbf{u})^{-\frac{1}{2}}\,\dot{u}_i \end{aligned} \tag{7.6}$$

and

$$\begin{aligned} \acute{v}_i &= (\lambda_i^\mathbf{v})^{-\frac{1}{2}}\,[\mathbf{V}]_i^T\,\mathbf{v} \\ &= (\lambda_i^\mathbf{v})^{-\frac{1}{2}}\,\dot{v}_i. \end{aligned} \tag{7.7}$$

Inspecting equation (7.3), it is evident that ρ_i^2 decreases with ω_i, implying that i orders the canonical variable pairs by decreasing canonical correlation coefficient in (7.6) and (7.7).

This result reveals an important property of the image-pair model that was introduced in Chapter 4: the i-th canonical variable pair consists of the i-th principal components from each image, normalized to have unit variance. Furthermore, it follows from (7.1) that the i-th canonical correlation coefficient in the image pair is the square of the correlation coefficient between the i-th principal components of each image.

7.1.3 Canonical Correlation Coefficients and the LRT

Repeating it here for convenient access, the optimal test derived in Chapter 5 has statistic

$$s\left(\acute{\mathbf{u}},\acute{\mathbf{v}}\right)=\sum_{i=1}^{n^2}\beta_i\left(\acute{u}_i-m_{\acute{u}_i}\right)\left(\acute{v}_i-m_{\acute{v}_i}\right)-\alpha_i\left(\left(\acute{u}_i-m_{\acute{u}_i}\right)^2+\left(\acute{v}_i-m_{\acute{v}_i}\right)^2\right) \tag{7.8}$$

and decision threshold

$$\acute{\lambda}=\log\lambda^2+\sum_{i=1}^{n^2}\log\left(\frac{1-k_i^2\rho_1^2}{1-k_i^2\rho_0^2}\right). \tag{7.9}$$

where

$$\alpha_i=\frac{k_i^2\left(\rho_1^2-\rho_0^2\right)}{\left(1-k_i^2\rho_0^2\right)\left(1-k_i^2\rho_1^2\right)}, \tag{7.10}$$

$$\beta_i=2\frac{k_i\left(\rho_1-\rho_0\right)\left(1+k_i^2\rho_0\rho_1\right)}{\left(1-k_i^2\rho_0^2\right)\left(1-k_i^2\rho_1^2\right)} \tag{7.11}$$

and

$$k_i=\frac{\rho_i}{\rho_{ab}}.$$

The statistic is written as a function of whitened images and therefore assumes a preprocessing stage involving the KL transform and normalization of the pixels to unit variance.

Notice that the statistic can be written as a sum of independent terms that operate on the canonical variable pairs $\{\acute{u}_i,\acute{v}_i\}$, where each term is weighted by a function of the associated canonical correlation coefficient ρ_i^2. It is possible for coefficients α_i and β_i to be uniform, as is the case for noise-free images, where $\rho_i^2=\rho_{ab}^2$, but in general this will not be the case and therefore some of the canonical variables will be emphasized when the LRT statistic is calculated. The relationship between the LRT statistic and the canonical variables will be investigated further in Section 7.2, where the weighting of terms in (7.8) will be exploited to reduce the dimensionality of the calculation.

7.2 Economy by Reduced Dimensionality

The motivation for using canonical correlations in exploratory data analysis is to identify the variables that represent the most correlation between two random vectors and then reduce the dimensionality of the problem by only considering these variables during further analysis of the data. The previous section showed that each term in (7.8) is a function of a canonical variable pair, $\{\acute{u}_i, \acute{v}_i\}$, and is weighted by a function of the associated canonical correlation coefficient between $\acute{u}_i$ and $\acute{v}_i$, denoted as ρ_i^2. If there are large differences between the magnitudes of the canonical correlation coefficients, then it is possible that some of the terms in (7.8) and (7.9) can be neglected, thereby reducing the computation required by the LRT. This possibility is investigated here.

7.2.1 Significance of Terms in the LRT Statistic

Without loss of generality, assume that the eigenvalues ω_i (and therefore the quantities k_i) are ordered with decreasing magnitude. Inspecting (7.10) and (7.11) reveals that α_i and β_i, coefficients of the summed terms in the test statistic of (7.8), are now also ordered with decreasing magnitude. Figure 7-1 illustrates this point for the situation where $\mathbf{R}$ is nonseparable Markov and the SNR is 2.

Since the pixels in the whitened images have unit variance, the effect of non-uniform ρ_i (and therefore the effect of additive noise, since for images free of noise, $\rho_i = \rho_{ab}\ \forall i \in \{1, 2, \ldots, n^2\}$) is to emphasize the contribution of certain pixels in the whitened images when calculating the test statistic. Figure 7-2 orders the basis vectors of the whitening transform by decreasing eigenvalue for 16×16 images with a nonseparable Markov covariance matrix. Notice that the basis vector associated with the lowest canonical correlation coefficient (corresponding to $i = 256$ in Figure 7-2) represents the image component that has least spatial correlation, which in the case of the proposed model and the Markov covariance matrix used in Figure 7-2 is predominantly additive white noise. The optimal test, therefore, appears to emphasize the contribution of image components with better SNR.

To summarize: for images with additive noise the significance of the terms in s decreases with increasing i for two reasons. First, the whitened pixel-pairs are ordered with decreasing correlation coefficient and second, the coefficients α_i and β_i are ordered with decreasing magnitude. Chapter 5 showed that the expectation and variance of the terms in s are given

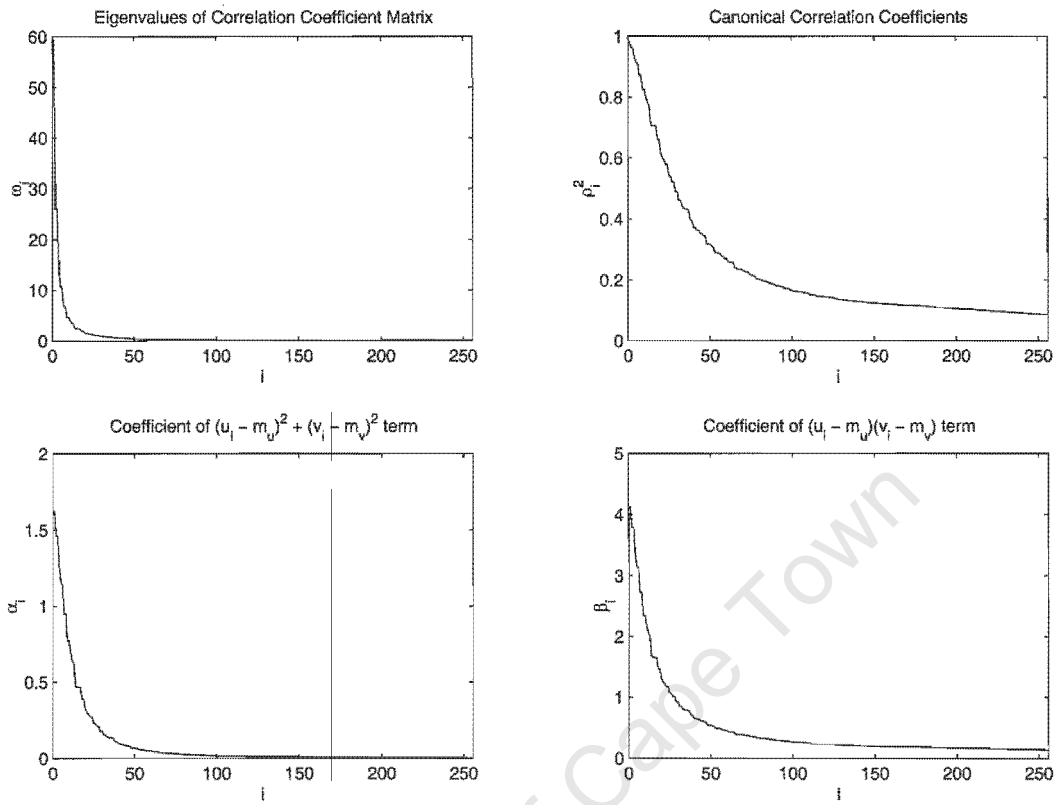


Figure 7-1: Coefficients of the LRT statistic for 16×16 images ($\rho = 0.8$, $\rho_1 = 0.8$, $SNR = 1.0$).

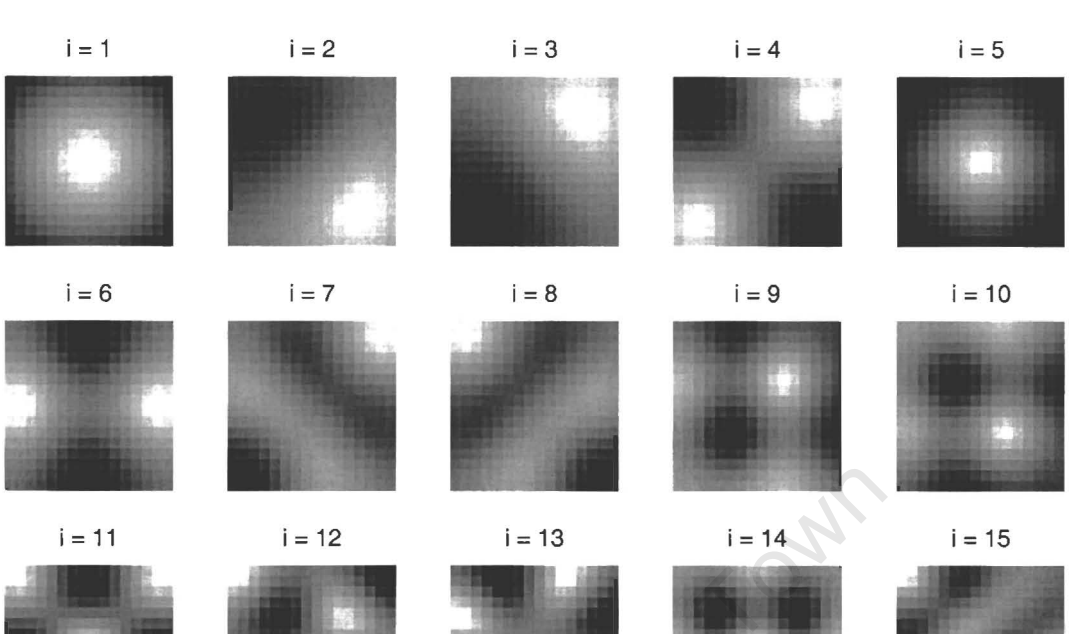
by

$$E[s_i] = \rho_{ab} k_i \beta_i - 2\alpha_i$$

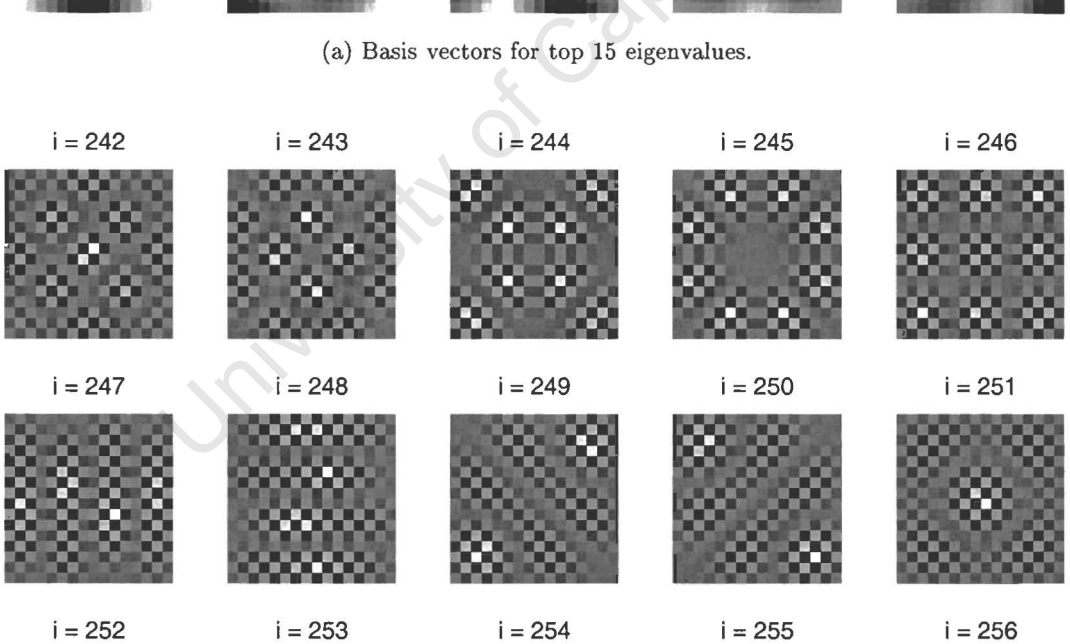
and

$$\mathrm{Var}[s_i] = 2(\rho_{ab} k_i \beta_i - 2\alpha_i)^2 + (1 - \rho_{ab}^2 k_i^2)(\beta_i^2 - 4\alpha_i^2),$$

respectively. Figure 7-3 plots the former to illustrate the decreasing significance of the terms with increasing i. The variance of each term also determines its significance, and a dashed line shows the one standard deviation confidence interval.



(a) Basis vectors for top 15 eigenvalues.



(b) Basis vectors for bottom 15 eigenvalues.

Figure 7-2: Basis vectors associated with covariance matrix eigenvalues of 16×16 images.

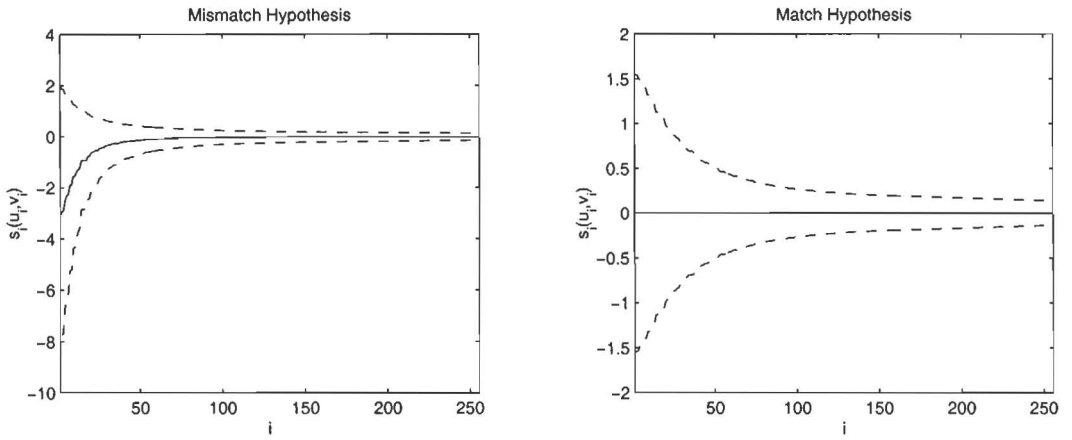


Figure 7-3: Significance of the pixel-pair terms in the LRT statistic when ordered with descending canonical correlation coefficient ($n = 16$, $\rho = 0.8$, $\rho_1 = 0.8$, and $SNR = 0.5$). The solid line depicts the mean of the terms. The dashed line delineates a one sigma confidence interval around the mean.

7.2.2 An Approximate Test Based on the Canonical Subset

An approximate version of the LRT can be written with statistic

$$s_c(\acute{\mathbf{u}}, \acute{\mathbf{v}}) = \sum_{i=1}^{c} \hat{\beta}_i (\acute{u}_i - m_{\acute{u}_i})(\acute{v}_i - m_{\acute{v}_i}) - \hat{\alpha}_i \left((\acute{u}_i - m_{\acute{u}_i})^2 + (\acute{v}_i - m_{\acute{v}_i})^2 \right) \tag{7.12}$$

and decision threshold

$$\acute{\lambda}_c = \log \lambda^2 + \sum_{i=1}^{c} \log \left(\frac{1 - \hat{k}_i^2 \rho_1^2}{1 - \hat{k}_i^2 \rho_0^2} \right), \tag{7.13}$$

where $c < n^2$. Note that the new notation, $\hat{\alpha}_i$, $\hat{\beta}_i$ and $\hat{k}_i$ indicates that these coefficients have been ordered with decreasing magnitude for increasing i.

The c components of $\acute{\mathbf{u}}$ and $\acute{\mathbf{v}}$ used by (7.8) capture more of the correlation structure between random images $\mathbf{u}$ and $\mathbf{v}$ than any other c-component reduction of the joint image-pair dimensionality. In other words, the property of the individual KL transforms on $\mathbf{u}$ and $\mathbf{v}$ that ensures the optimal compaction of energy into c components [8, p. 168] translates into an optimal compaction of the correlation structure into c component-pairs, $\acute{w}_i = \{\acute{u}_i, \acute{v}_i\}$, under the proposed image-pair model[3]. Therefore (7.8) represents the most powerful test for match

[3]Note that the words "optimal compaction" are used here in a mean squared sense. In other words, the compaction is optimal over the population of possible image pairs, rather than for a single image pair.

between $\mathbf{u}$ and $\mathbf{v}$ given the restriction of using only c components of the original images. The c pixel-pairs used to calculate (7.12) and (7.13) are now referred to as the c-component *canonical subset* of the image pair.

Figure 7-4 shows the pdf obtained under the match and mismatch hypotheses in a Monte Carlo experiment with 16×16 images and unity SNR. In this case the pdf of the approximate LRT statistic that uses a 50% canonical subset of the 256 whitened pixel-pairs is almost indistinguishable from the pdf of the full LRT statistic. Looking at the overlap of the hypothesis conditional pdfs, it is evident that the approximate LRT using a 10% canonical subset outperforms Pearson's r for the ideal observer test.

7.2.3 Probability of Error

The expression for the expectation and variance of the approximate statistic is identical to that of the full statistic, except that the sum is only taken over the terms that are in the canonical subset. As a function of ρ_{ab} and the number of terms c, the mean and variance are

$$m_s(\rho_{ab}, c) = \sum_{i=1}^{c} \left(\rho_{ab} \hat{k}_i \hat{\beta}_i - 2\hat{\alpha}_i \right)$$

and

$$\sigma_s^2(\rho_{ab}, c) = \sum_{i=1}^{c} \left[2\left(\rho_{ab} \hat{k}_i \hat{\beta}_i - 2\hat{\alpha}_i \right)^2 + \left(1 - \rho_{ab}^2 \hat{k}_i^2 \right) \left(\hat{\beta}_i^2 - 4\hat{\alpha}_i^2 \right) \right]$$

respectively. The pdf of the statistic is asymptotically normal with respect to c. Therefore for large enough c the probability of type I and type II errors in an ideal observer test can be calculated analytically using expressions similar to (5.26) and (5.27) of the previous chapter. These expressions, which are now a function of c, are

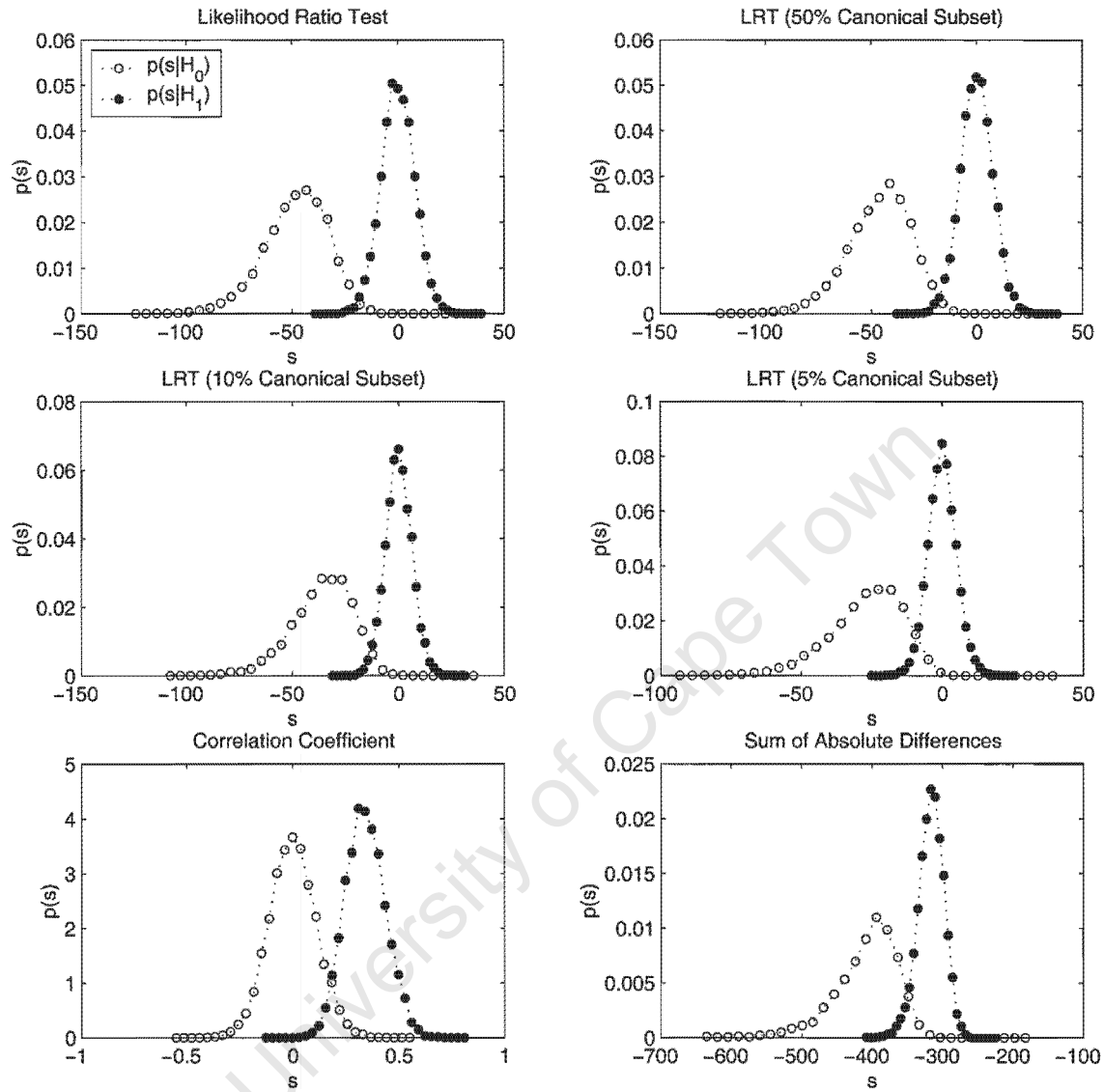
$$P_I(\rho_0, c) = \frac{P_0}{2} \left(1 - \operatorname{erf} \left[\frac{\acute{\lambda} - m_s(\rho_0, c)}{\sqrt{2\sigma_s^2(\rho_0, c)}} \right] \right)$$

and

$$P_{II}(\rho_1, c) = \frac{P_1}{2} \left(1 + \operatorname{erf} \left[\frac{\acute{\lambda} - m_s(\rho_1, c)}{\sqrt{2\sigma_s^2(\rho_1, c)}} \right] \right),$$

respectively.

Figure 7-5 graphs these theoretical error rates for 16×16 images over SNR and the



(a) Monte Carlo histograms.

Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.8
Scene one-step correlation (ρ)	0.8
SNR	1

(b) Simulation parameters.

Figure 7-4: Monte Carlo histograms under match and mismatch hypotheses for the canonical subset LRT statistic, the full LRT statistic, and two suboptimal statistics. The experiment used an ensemble of 10000 image pairs.

degree of spatial correlation in the images. A nonseparable Markov model was assumed for the shared correlation coefficient matrix $\mathbf{R}$. This matrix is parameterized on the one-step correlation coefficient ρ. Each graph plots the error rate for approximate tests that use a 90% ($c = 231$), 30% ($c = 77$) and 10% ($c = 26$) canonical subset.

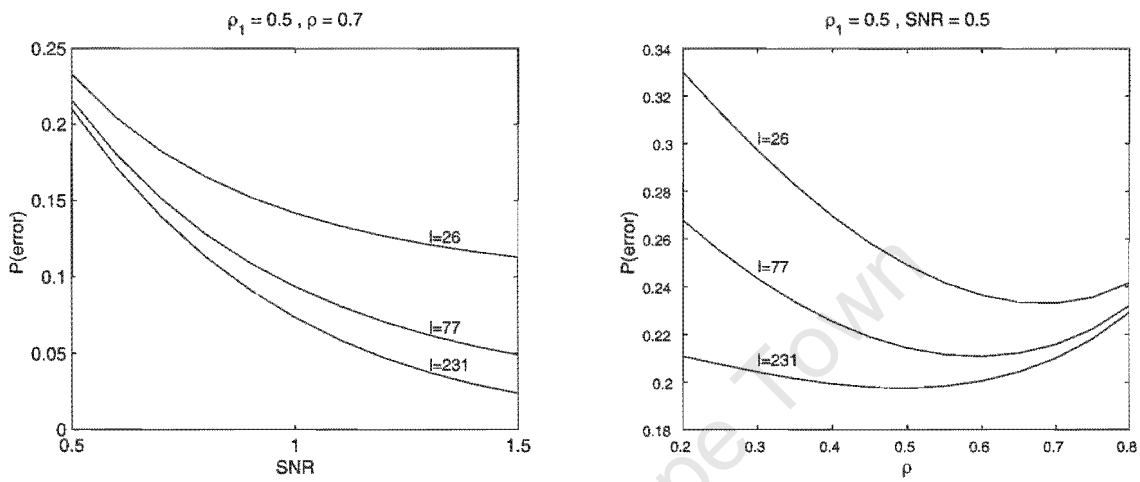


Figure 7-5: Probability of error versus SNR and spatial correlation for different size canonical subsets.

Increasing SNR has the predictable effect of lowering the error rate. The effect of different levels of spatial correlation is more interesting, with the error rate decreasing to a minimum and then increasing. This is the result of two competing effects relating to the effective decrease in scene information that an increase in ρ represents. A decrease in information makes the approximate representation of the correlation structure less "lossy" and the approximate test is closer to the full LRT. This effect dominates for lower ρ and causes the initial decreasing trend in error rate. The second effect was mentioned in the previous chapter with regard to the error rate of the full LRT: less information in the scene component of the image leads to lower discriminating power in the test. For the higher values of ρ this effect is dominant.

7.2.4 Principal Components and Image Matching

At this point a comparison of the canonical subset and the role of principal components in other image matching algorithms is appropriate. Face recognition is one area in particular where PCA has been applied with some success. Since many systems store one face image as a reference and then compare a later face image to determine whether they do indeed represent the same person, it can be argued that this is a matching problem as it was defined

in Chapter 1 — a decision must be made as to whether two observations of the real world (the face images) represent the same scene (a particular person's face). The use of principal components for this problem originated in work by Sirovich and Kirby that sought to find an ideal representation for face images [97, 98]. Turk and Pentland subsequently used this representation in a pattern recognition framework for face recognition in a scheme called *eigenfaces* [73].

There are several variations on the theme [72, 53, 54], but in essence the eigenfaces algorithm can be summarized as follows: use a training set of faces to estimate a linear model for the face population that is characterized by a mean vector $\mathbf{m}$ and covariance matrix $\mathbf{K}$. Define the eigenfaces as the eigenvectors (denoted $\mathbf{h}_i$) that are associated with the N largest eigenvalues (denoted λ_i) of $\mathbf{K}$. Represent the training face associated with each person in the original data-set as the set of weights obtained when the mean, $\mathbf{m}$, is subtracted from the face image and it is projected onto the eigenfaces. In order to classify a new face image, subtract the mean face $\mathbf{m}$ and project the result onto the eigenfaces. Then calculate the Mahalanobis distance between the eigenface representation of the new face and that of each of the known faces. If any of the distances fall below a predefined threshold, ϵ, then this indicates a potential match. The new face is recognized as belonging to the same person as the training face with the smallest distance in the list of potential matches.

For matching new face $\mathbf{u}$ to training face $\mathbf{v}$, this procedure effectively uses the image distance measure

$$d_{\text{EIG}}(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^{N} \left[\frac{1}{\sqrt{\lambda_i}} \mathbf{h}_i \cdot (\mathbf{u} - \mathbf{v}) \right]^2 . \tag{7.14}$$

As a *representation* the eigenfaces approach is optimal in that, given a simple linear model for the images, it minimizes error in a mean square sense if the dimensionality is constrained. This, however, does not guarantee that the eigenfaces recognition algorithm is optimal with respect to the probability of recognition error. In contrast, the optimal LRT does guarantee optimal matching given the joint image model proposed in Chapter 4.

Figure 7-6 serves to underscore this last point. Under the proposed model the LRT statistic outperforms the eigenfaces difference measure in an ideal observer test, even when the dimensionality is constrained to 20% of the full number of pixels. The advantage is particularly evident for low SNR. Error rates for Pearson's r are given to provide additional context and it is evident that r is more tolerant of high noise levels than the eigenfaces

distance. It is interesting to note that for SNR $\lessapprox 2$, the eigenfaces distance using 20% of the eigenfaces (Figure 7-6(b) and 7-6(d)) has a lower error rate than the eigenfaces distance based on the full set of eigenfaces (Figure 7-6(a) and 7-6(c)). This is because noise predominates in the eigenfaces associated with the lower 80% of the eigenvalues, which are discarded in the eigenfaces distance of Figure 7-6(b) and 7-6(d). Using 20% of the eigenfaces is effectively a crude approximation of the weighting of important canonical variables that is inherent in the LRT statistic.

7.3 Economy by Model Simplification

Certain simplifications of the image model result in whitening transforms with significantly better computational economy. Two of these are mentioned here. The first uses the simple efficiency inherent in a separable image model and the second exploits an approximation of the KL transform for images with high spatial correlation.

7.3.1 Matching with a Separable Model

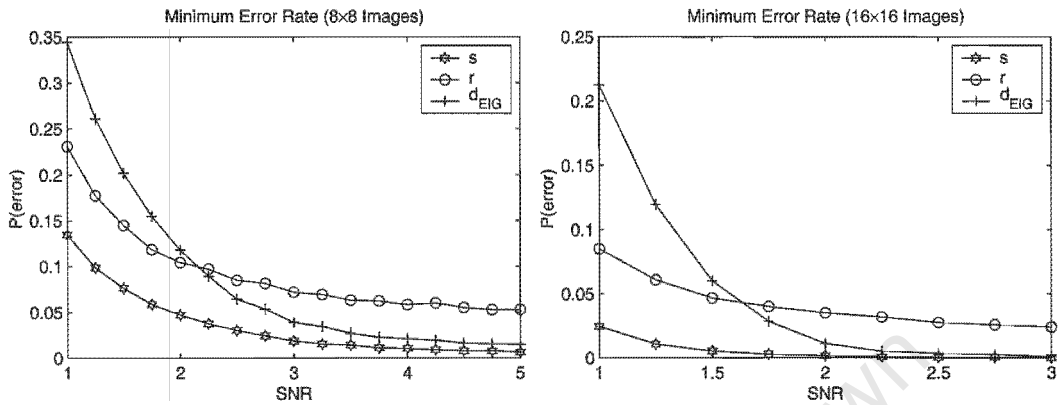
In the optimal test the images are whitened with a transformation of the form $\acute{\mathbf{u}} = \Lambda^{-\frac{1}{2}}\mathbf{V}^T\mathbf{u}$, where the columns of $\mathbf{V}$ are the eigenvectors of the ideal scene correlation coefficient matrix $\mathbf{R}$, and the elements on the diagonal of Λ are the associated eigenvalues. If the scene model is separable, then $\mathbf{R}$ can be written as the Kronecker product of the row and column correlation coefficient matrices: $\mathbf{R} = \mathbf{R}_c \otimes \mathbf{R}_r$. The eigenvector matrix associated with $\mathbf{R}$ can then be written in terms of those associated with $\mathbf{R}_c$ and $\mathbf{R}_r$ as $\mathbf{V} = \mathbf{V}_c \otimes \mathbf{V}_r$ [8, p. 30], and

$$\acute{\mathbf{u}} = \Lambda^{-\frac{1}{2}} \left(\mathbf{V}_c \otimes \mathbf{V}_r\right)^T \mathbf{u}. \tag{7.15}$$

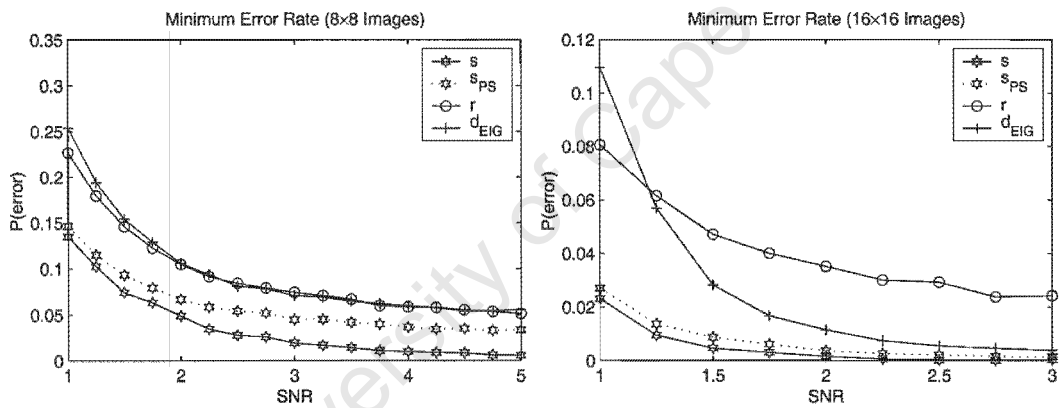
If $\mathbf{U}$ denotes the image $\mathbf{u}$ written in matrix form, then (using the properties of Kronecker products [8, p. 30]), the operation in (7.15) can be written as

$$\acute{\mathbf{u}} = \Lambda^{-\frac{1}{2}} \mathrm{vec}\left(\mathbf{V}_c^T \mathbf{U} \mathbf{V}_r\right),$$

where vec $(\cdot)$ denotes the operation that takes a $n \times n$ matrix to a n^2 vector with the elements in row-column order. Here $\mathbf{V}_c^T\mathbf{U}\mathbf{V}_r$ is an $O\left(n^3\right)$ operation, in contrast with the $O\left(n^4\right)$ operation required for $\left(\mathbf{V}_c \otimes \mathbf{V}_r\right)^T \mathbf{u}$ in (7.15). Usually the row and column models are the



(a) Experiment using all all eigenfaces.



(b) Experiment using the top 20% of the eigenfaces (d_{EIG}) and a 20% canonical subset in the partial LRT (s_{PS}).

Parameter	*Value*
Image size (n)	8
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.8
Scene one-step correlation (ρ)	0.9
SNR	[1,5]

(c) Simulation parameters.

Figure 7-6: Monte Carlo error-rate comparison of the LRT statistic and the eigenfaces distance (10000 trials).

same and the whitening transformation can be written as

$$\acute{\mathbf{u}} = \Lambda^{-\frac{1}{2}} \text{vec}\left(\mathbf{V}_s^T \mathbf{U} \mathbf{V}_s\right), \tag{7.16}$$

where $\mathbf{V}_s = \mathbf{V}_c = \mathbf{V}_r$.

Now consider the overall whitening transformation. Assume that $\tilde{\mathbf{u}} = \text{vec}\left(\mathbf{V}_s^T \mathbf{U} \mathbf{V}_s\right)$ is calculated first. Note that although the remaining operation, $\Lambda^{-\frac{1}{2}}\tilde{\mathbf{u}}$, is written as an $O\left(n^4\right)$ matrix-vector multiplication, $\Lambda^{-\frac{1}{2}}$ is diagonal and therefore $\Lambda^{-\frac{1}{2}}\tilde{\mathbf{u}}$ can be computed with an $O\left(n^2\right)$ operation. The overall separable computation of (7.16) is then $O\left(n^3\right)$, whereas the overall computation of (7.15) is $O\left(n^4\right)$. This represents a substantial reduction in computation.

This section has assumed that a separable model is adequate for the problem at hand. Another situation is the one where a nonseparable model is known to be the better option, but a separable solution must suffice because of limited computational resources. In this situation the separable approximation to the KL transform can be used to reduce computation and the whitening transformation can be written as

$$\acute{\mathbf{u}} \approx \Lambda_{\text{SEP}}^{-\frac{1}{2}} \text{vec}\left(\mathbf{V}_s^T \mathbf{U} \mathbf{V}_s\right), \tag{7.17}$$

which uses diagonal matrix Λ_{SEP} instead of the matrix of eigenvalues. In order to best approximate the nonseparable case

$$\Lambda_{\text{SEP}} = \text{diag}\left[\left(\mathbf{V}_s \otimes \mathbf{V}_s\right)^T \mathbf{K}_{\mathbf{u}} \left(\mathbf{V}_s \otimes \mathbf{V}_s\right)\right], \tag{7.18}$$

where $\mathbf{K}_{\mathbf{u}}$ is the nonseparable covariance matrix of the original image $\mathbf{u}$ and $\text{diag}\left[\mathbf{A}\right]$ is $\mathbf{A}$ with non-diagonal elements set to zero.

7.3.2 Matching with the Discrete Cosine Transform

If the row and column models are separable, stationary, first order Markov sequences with one-step correlation coefficients approaching unity, then the discrete cosine transform (DCT) is a good approximation of the KL transform [8, p. 153]. Assuming the approximation is adequate, then (7.16) can be rewritten as

$$\acute{\mathbf{u}} \approx \Lambda_{\text{DCT}}^{-\frac{1}{2}} \text{vec}\left(\text{DCT}\left\{\mathbf{U}\right\}^T\right),$$

where Λ_{DCT} replaces the eigenvalue matrix, is diagonal, and ensures that the elements of $\acute{\mathbf{u}}$ are unit variance. If the DCT is written as the matrix transformation:

$$\mathrm{DCT}\{\mathbf{U}\} = \mathbf{C}\mathbf{U}\mathbf{C}^T,$$

then

$$\Lambda_{\mathrm{DCT}} = \mathrm{diag}\left[(\mathbf{C}\otimes\mathbf{C})\,\mathbf{K}_{\mathbf{u}}\,(\mathbf{C}\otimes\mathbf{C})^T\right], \tag{7.19}$$

where $\mathbf{K}_{\mathbf{u}}$ is the covariance matrix of the original image $\mathbf{u}$. Computation is therefore reduced by replacing the $O\left(n^3\right)$ separable KL transform with the $O\left(n^2\log_2 n\right)$ DCT. Figure 7-7 shows a comparison of the time taken to compute $\mathbf{V}_s\mathbf{U}\mathbf{V}_s^T$ and $\mathrm{DCT}\{\mathbf{U}\}$ in the MATLAB programming environment.

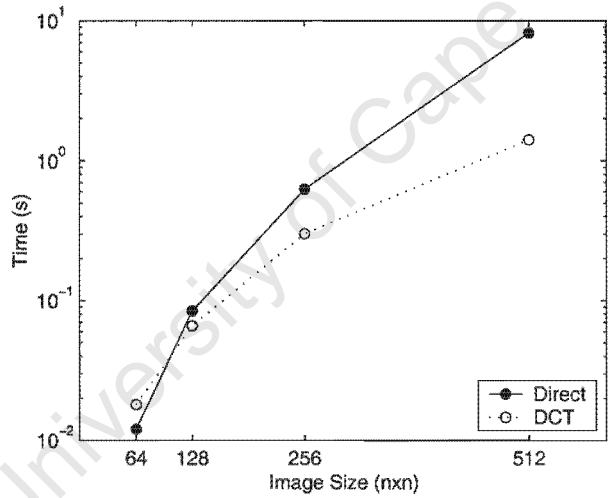


Figure 7-7: Comparison of computation times for direct and DCT-based calculation of the separable whitening transform.

The approach based on the DCT has the additional advantages that the whitening transformation is independent of the exact image model (except for calculating the weightings in (7.19)) and that no eigendecomposition of the covariance matrix is required beforehand.

7.3.3 Monte Carlo Experiments

The error-rate performance of the separable and DCT-based tests is now compared to that of the optimal LRT using Monte Carlo experiments. The results of two experiments are

presented (with Pearson's r for context) in Figure 7-8(a). In order to show small differences the results are also presented as an error rate relative to that of the LRT statistic. It is apparent that the separable and DCT approximations sacrifice very little error-rate performance. As expected, the former is the better of the two, but only by a very small margin.

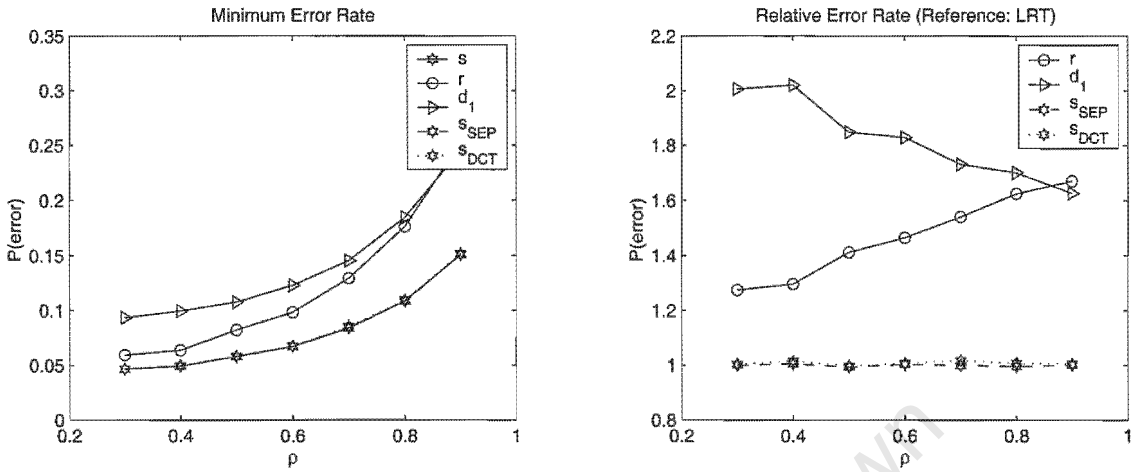
Figure 7-8(b) shows a result where the model that generated the images was nonseparable, whereas the covariance matrix used in (7.18) and (7.19) to calculate the weighting of the whitened components was assumed to be separable. This inaccuracy increases the error rate, illustrating the utility of calculating (7.18) and (7.19) using the known nonseparable covariance matrix, even when the separability assumption is being made in order to use the KL or DCT transform for their better efficiency. This last result suggests that it is not the exact KL transform, but rather the correct weighting of the transformed components that is important in retaining the error-rate properties of the LRT.

7.4 A Practical Test for Large Images

The LRT statistic is based on the covariance matrix of the images in the image pair. For an $n \times n$ image, the covariance matrix is $n^2 \times n^2$, which, even for relatively small images and today's ever improving computing resources, is sometimes impractical to manipulate. Consider a 32×32 image: the full covariance matrix is 1024×1024 elements and the eigenvector/eigenvalue decomposition is a significant computational task. The simplification of a separable model reduces the problem to two 32×32 covariance matrices, which is manageable. But if large images (say 512×512 pixels and larger) must be processed, then even the separable model is impractical. One could argue that in time computing resources will advance to a point where 512×512 images will not be problematic, but when this is true there will undoubtedly be applications that deal with still larger images. A simple, but suboptimal solution to this problem is to break the images into smaller blocks, process these blocks individually and then combine the results to form the overall LRT statistic of the image pair. This approach is investigated here.

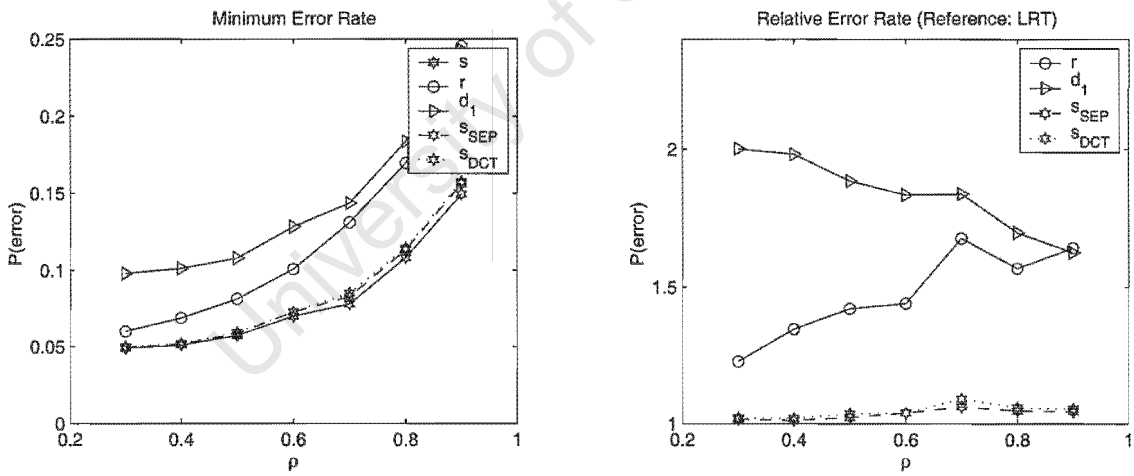
7.4.1 The Blockwise Partitioned LRT

Figure 7-9 shows a larger image partitioned into smaller, non-overlapping blocks. For full $n \times n$ images $\mathbf{u}$ and $\mathbf{v}$ the $n_b \times n_b$ subimage blocks are denoted $\mathbf{u}^i$ and $\mathbf{v}^i$ for $i \in [1, 2, \ldots, n_b^2]$. If the simplifying assumption is made that the blocks are statistically independent, then the



Parameter	*Value*
Image size (n)	8
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.8
SNR	1

(a) Image model and weighting both nonseparable.



Parameter	*Value*
Image size (n)	8
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.8
SNR	1

(b) Nonseparable image model and separable weighting.

Figure 7-8: Monte Carlo investigation of error rates for the separable and DCT LRT approximations (10000 trials).

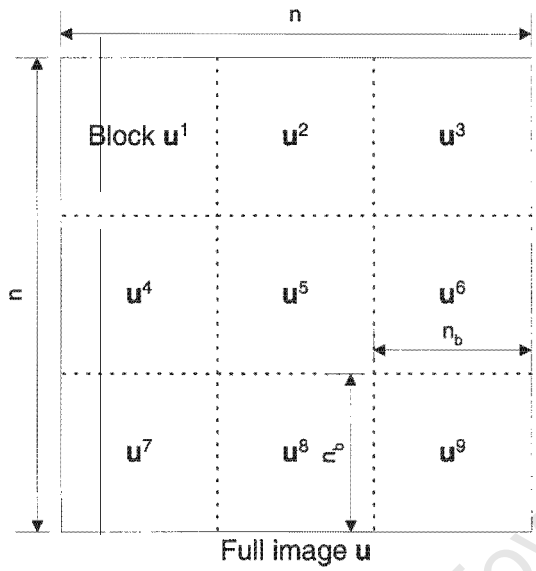


Figure 7-9: Block partitioned image.

pdf of the full image-pair can be written in terms of the block-pair pdfs as

$$p_{\mathbf{u},\mathbf{v}}(\mathbf{u},\mathbf{v}) = \prod_{i=1}^{n_b^2} p_{\mathbf{u}^i,\mathbf{v}^i}\left(\mathbf{u}^i,\mathbf{v}^i\right)$$

and the match log-likelihood ratios are related by

$$L(\mathbf{u},\mathbf{v}) = \sum_{i=1}^{n_b^2} L_{n_b}\left(\mathbf{u}^i,\mathbf{v}^i\right), \tag{7.20}$$

where $L(\cdot)$ and $L_{n_b}(\cdot)$ are the log-likelihood ratios of the full image-pair and the block pair, respectively. Using (7.20) and the expression for the LRT statistic derived in Chapter 5, it can be shown that the statistic for the full image can be written as the sum of the statistics s_{n_b} for the individual block pairs as follows

$$s(\acute{\mathbf{u}},\acute{\mathbf{v}}) = \sum_{i=1}^{n_b^2} s_{n_b}\left(\acute{\mathbf{u}}^i,\acute{\mathbf{v}}^i\right).$$

Recall that $\acute{\mathbf{u}}^i$ and $\acute{\mathbf{v}}^i$ are the whitened versions of $\mathbf{u}^i$ and $\mathbf{v}^i$ respectively. Similarly, the corresponding decision threshold is

$$\acute{\lambda} = n_b^2 \cdot \acute{\lambda}_{n_b} = n_b^2 \left[2 \log \left(\frac{1 - P_1}{P_1} \right) + \sum_{i=1}^{n_b^2} \log \left(\frac{1 - k_i^2 \rho_1^2}{1 - k_i^2 \rho_0^2} \right) \right],$$

where $\acute{\lambda}_{n_b}$ is the decision threshold for an individual block.

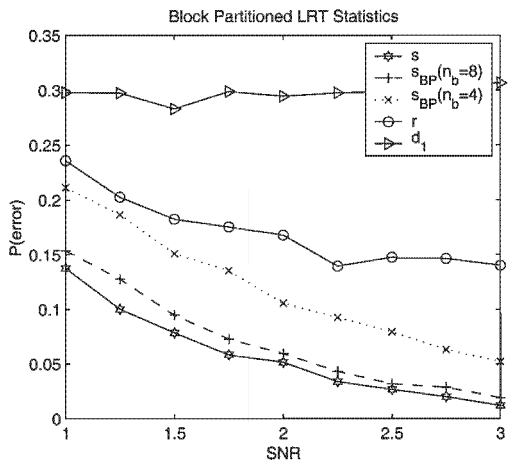
7.4.2 Monte Carlo Experiments

Figure 7-10 shows the result of an experiment that compares the block partitioned approximation with the full LRT for a range of SNR values. The image size is 16×16 and both 8×8 and 4×4 approximations are compared. In this experiment the performance of the 8×8 block partition is very close to that of the full LRT, while the 4×4 approximation fares significantly worse. In both cases the block-independence assumption is violated at the block edges, but for larger blocks there are fewer edges overall and the effect is lessened. But even the 4×4 approximation, where the block independence assumption is grossly violated, exhibits better performance than the suboptimal statistics. This is to be expected, since any partitioned LRT with blocks larger than a single pixel in size incorporates information about the spatial correlation in the image, which is arguably what the suboptimal statistics are lacking. The error-rate for matching larger 64×64 images using the block partitioned statistic is shown in Figure 7-11 and similar trends are observed for a range of block sizes.

7.5 Discussion

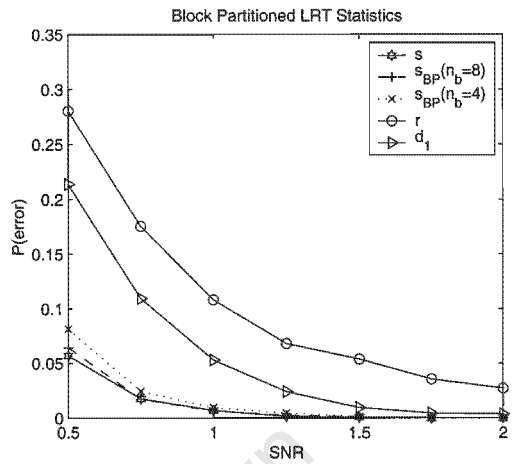
The optimal test expressed as an operation on whitened images has significance beyond that of being a mathematically convenient representation. The pixel-pairs in the whitened images, which correspond to the pairs of image principal components, are actually the canonical variables of the image pair. What is more, the ordering of the principal components by variance is equivalent to the ordering of the canonical variable pairs by canonical correlation coefficient. There is therefore an optimal compaction (in a mean square sense) of the correlation between the two images in pairs of corresponding principal components from the individual images.

This optimal compaction leads to a natural method for reducing the dimensionality of the matching problem by forming an approximate test that uses only a subset of the canonical variable pairs. Two mechanisms ensure that the computation required of optimal LRT statis-



Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
Scene one-step correlation (ρ)	0.95

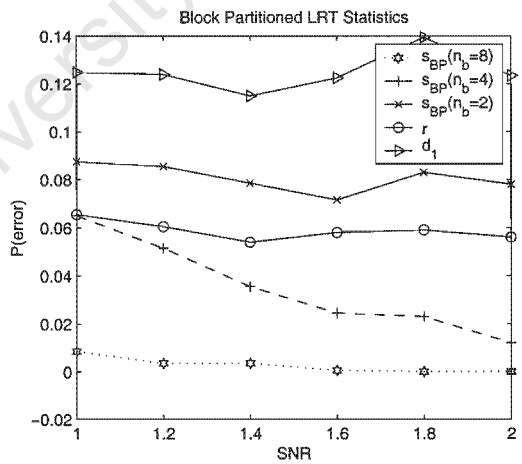
(a) Low match correlation coefficient.



Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.99
Scene one-step correlation (ρ)	0.95

(b) High match correlation coefficient.

Figure 7-10: Monte Carlo results that compare the full LRT with the block partitioned LRT approximation (500 trials).



Parameter	*Value*
Image size (n)	64
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
Scene one-step correlation (ρ)	0.95

Figure 7-11: Monte Carlo results for large images and the block partitioned LRT approximation (500 trials).

Statistic	*Order*
Pearson's r	$O\left(n^2\right)$
Sum of squared differences	$O\left(n^2\right)$
LRT statistic	$O\left(n^4\right)$
LRT statistic (canonical subset with c basis images)	$O\left(cn^2\right)$
LRT statistic (separable)	$O\left(n^3\right)$
LRT statistic (DCT)	$O\left(n^2 \log n\right)$
LRT statistic ($n_b \times n_b$ partitions)	$O\left(n^2 n_b^2\right)$

Table 7.1: Algorithm order for similarity statistic calculation with $n \times n$ images.

tic can be significantly reduced with very little compromise in error rate: (1) the whitening transform compresses the essential correlation structure into fewer pixel-pairs, and (2) the LRT statistic weights the contribution of these pairs more heavily than the other pixel-pairs.

Another method for improving the economy of the LRT computation uses simplified models. If the row and column models of the image are separable, then the whitening transform is reduced from an $O\left(n^4\right)$ to an $O\left(n^3\right)$ operation. If the separable model has one-step spatial correlation coefficient approaching one, then the KL transform aspect of the whitening transform is approximated well by the DCT, making the further improvement to an $O\left(n^2 \log n\right)$ operation. The error introduced by the separability assumption when the images are actually nonseparable is reduced by calculating the weightings of the principal components using the original nonseparable covariance matrix.

A third method is targeted at large images, for which even the separable covariance matrices are difficult to manipulate. The strategy is to simply partition the image into separate non-overlapping blocks and to sum the LRT statistic from each block to obtain the overall statistic. The underlying assumption is that the blocks are statistically independent and although this is not strictly true, experiments show that the method is effective and does not sacrifice significant error-rate performance if the blocks are not made too small. The order of the algorithms that calculate two standard statistics and five variations on the LRT statistic are given in Table 7.1.

The optimal test for image matching has now been derived, compared to other methods and implemented efficiently. The next chapter formulates a classic image processing problem — translational image registration — using a hypothesis testing procedure. This problem is arguably one of the most important applications of the similarity statistic and it is therefore important that the principles of the LRT are applied here too.

Chapter 8

Hypothesis Tests for Image Registration

Image registration is the task of finding a mapping between two images that represents scene equivalence, where this problem is non-trivial due to spatial misalignment or geometric distortion of scene information in the images, image intensity differences caused by inconsistency in the irradiation of the scene or the image acquisition system, or changes in the scene itself [99]. Typical applications include scene reconstruction from stereo images, comparative analysis of medical images and change detection.

This chapter considers the registration sub-problem that is commonly referred to as *block matching*. Block matching in two images proceeds by extracting the subimage, or block, around a control point in the first image and comparing this subimage to a region, or search area, in the second image to find the corresponding point. This point is chosen to be the one that maximizes some similarity measure between the block and the search area. A mapping from one image to the other is established by following this procedure for a number of control points. Early work in image registration was based on block matching alone [11, 40], but modern approaches are far more sophisticated, operating on multi-modal image sets [100, 50] and catering for complex spatial distortion [101, 102].

Although translational block matching is not the state of the art in image registration algorithms, many sophisticated algorithms depend on it in an early stage of their processing (e.g. [103]) and can therefore benefit from improved accuracy and efficiency in the search for matching blocks. Section 8.1 formulates this procedure as a hypothesis testing problem that exploits the LRT matching statistic developed in Chapter 5. Efficient implementation of the

search algorithm is considered in Section 8.2. In order to register two complete images to one another, block matching must be performed across the image to find a set of local mappings, and Section 8.3 proposes methods for selecting the set of control points that will provide the best result. Section 8.4 evaluates the block matching performance of different techniques using Monte Carlo simulation and experiments with real images. Section 8.5 concludes the chapter.

8.1 Formulating the Block Matching Problem

The following registration sub-problem is now considered: given the $n \times n$[1] block $\mathbf{v}$ and the $n_s \times n_s$ search area $\mathbf{s}$, where $n < n_s$, find the position $\mathbf{p} = (i, j)$, if there is one, where the scene content of the $n \times n$ subimage of $\mathbf{s}$ centered on $\mathbf{p}$ corresponds to the scene content in $\mathbf{v}$. The coordinates of $\mathbf{p}$ are discrete, have their origin at the center of the search area, and take on values that represent an integer number of increments in the translational exhaustive search. In most cases this increment is the pixel size, but for subpixel accuracy it will be smaller. Call $\mathbf{p}$ the position of *correct register* between $\mathbf{v}$ and $\mathbf{s}$. Now $\mathbf{p} \in \{\mathbf{p}_k : k \in \{1, \ldots, N_p\}\}$ where N_p is the number of possible positions in $\mathbf{s}$.

Denote the $n \times n$ subimage of $\mathbf{s}$ centered on an arbitrary position $\mathbf{p}_k$ as $\mathbf{u}_k$. The problem is then one of finding out which subimage $\mathbf{u}_k \in \mathbf{s}$, if any, corresponds to $\mathbf{v}$, where $k \in \{1, 2, \ldots, N_p\}$. It is not assumed that there is one unique match between $\mathbf{v}$ and one subimage $\mathbf{u}_k \in \mathbf{s}$, and therefore the registration problem is not equivalent to finding out which subimage $\mathbf{u}_k$ *matches* $\mathbf{v}$. There are potentially up to N_p subimages in $\mathbf{s}$ that match $\mathbf{v}$, but there is at most one position of correct register. This distinction between match (the two scenes look the same) and registration (they are views of the same scene) is an important one, because it introduces the possibility of scene ambiguity into the image registration model.

The search is depicted in Figure 8-1. If only the positions where $\mathbf{v}$ and $\mathbf{s}$ completely overlap are considered to be candidates for registration, then $N_p = (n_s - n + 1)(n_s - n + 1)$. If the reference point of the $n \times n$ image $\mathbf{v}$ is defined to be its center at $\left(\left\lfloor \frac{n+1}{2} \right\rfloor, \left\lfloor \frac{n+1}{2} \right\rfloor\right)$, then the candidate positions in the search area are given by

$$\mathbf{p}_k = (i, j) : i \in \left\{ \left\lfloor \frac{n+1}{2} \right\rfloor, n_s - \left\lfloor \frac{n}{2} \right\rfloor \right\} \ , \ j \in \left\{ \left\lfloor \frac{n+1}{2} \right\rfloor, n_s - \left\lfloor \frac{n}{2} \right\rfloor \right\}.$$

[1] The extension to non-square images is trivial.

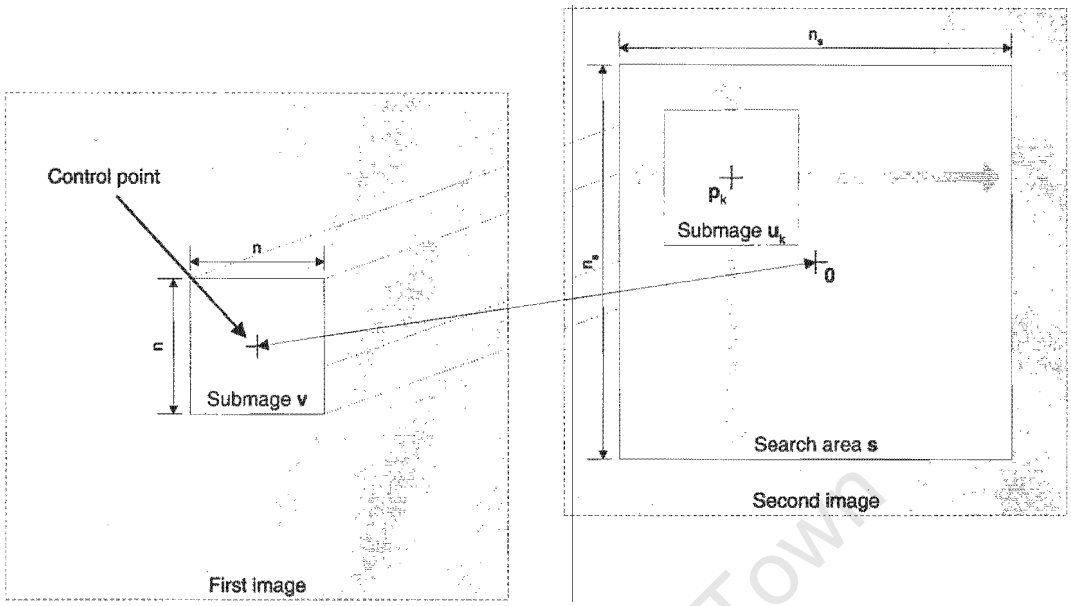


Figure 8-1: Block matching schematic.

8.1.1 The Search as a MAP Test

The decision of which position constitutes registration of $\mathbf{v}$ with the search area $\mathbf{s}$ is now formulated as a hypothesis test. The test that chooses the hypothesis with maximum *a posteriori* probability, the so-called MAP test, minimizes the overall probability of error. Denoting H_k as the hypothesis that $\mathbf{p}_k$ is the position of correct register and as R_k the event that this is correct, this test can be written as

$$\text{Accept } H_c \iff P(R_c|\mathbf{u}_c, \mathbf{v}) > P(R_k|\mathbf{u}_k, \mathbf{v}) \,\forall k. \tag{8.1}$$

This test, however, does not address the situation where the position of correct register is not in the search area. This can be a consequence of the fact that the images have finite extent and each image will have some unique scene content where they do not overlap. It may also happen when the correct position is outside a search area that has been truncated for practical reasons (see Section 8.1.3.) In either case, the decision rule should have a reject option and this can be provided in the form of a *no-registration*, or *rejection hypothesis*, H_0. A lower limit, denoted λ_R, is specified for the *a posteriori* probability of correct registration,

giving the decision rule with rejection hypothesis:

$$\begin{aligned} &\text{Accept } H_c \iff \exists c \text{ s.t. } P(R_c|\mathbf{u}_c,\mathbf{v}) > P(R_k|\mathbf{u}_k,\mathbf{v})\,\forall k \text{ and } P(R_c|\mathbf{u}_c,\mathbf{v}) > \lambda_R \\ &\text{Accept } H_0 \quad \text{Otherwise.} \end{aligned} \tag{8.2}$$

8.1.2 The Registration Statistic and Rejection Threshold

In order to specify the test it is necessary to derive the *a posteriori* probability $P(R_k|\mathbf{u}_k,\mathbf{v})$. The following information is available to the designer:

1. The image data $\mathbf{s}$ and $\mathbf{v}$ and their associated statistical models.
2. Defined states of match and mismatch for the scene (noise-free) components of the image data, where M_k denotes the event that the pair $\{\mathbf{u}_k,\mathbf{v}\}$ match.
3. The fact that registration implies matching images: $R_k \Longrightarrow M_k$. Note that the converse is not true in general because of potential scene ambiguity..
4. The *a priori* probability of correct register for each position $\mathbf{p}_k$, denoted $P(R_k)$.

A Posteriori Probability of Registration as a Function of Match Likelihood

By Bayes' theorem,

$$P(R_k|M_k,\mathbf{u}_k,\mathbf{v}) = \frac{P(R_k|\mathbf{u}_k,\mathbf{v})\cdot P(M_k|R_k,\mathbf{u}_k,\mathbf{v})}{P(M_k|\mathbf{u}_k,\mathbf{v})}.$$

Since $R_k \Longrightarrow M_k$, $P(M_k|R_k,\mathbf{u}_k,\mathbf{v}) = 1$ and

$$P(R_k|M_k,\mathbf{u}_k,\mathbf{v}) = \frac{P(R_k|\mathbf{u}_k,\mathbf{v})}{P(M_k|\mathbf{u}_k,\mathbf{v})}. \tag{8.3}$$

Also, $P(R_k|M_k,\mathbf{u}_k,\mathbf{v}) = P(R_k|M_k)$, since the only information that $\mathbf{u}_k$ and $\mathbf{v}$ have to offer in this formulation of the problem is their state of match or mismatch. Making this substitution and rearranging equation (8.3)

$$P(R_k|\mathbf{u}_k,\mathbf{v}) = P(R_k|M_k)\,P(M_k|\mathbf{u}_k,\mathbf{v}).$$

Now, using Bayes' theorem once again,

$$P(R_k|M_k) = \frac{P(R_k)\cdot P(M_k|R_k)}{P(M_k)} = \frac{P(R_k)}{P(M_k)},$$

since $P(M_k|R_k) = 1$ (registration implies match), and therefore

$$\begin{aligned} P(R_k|\mathbf{u}_k, \mathbf{v}) &= P(R_k) \frac{P(M_k|\mathbf{u}_k, \mathbf{v})}{P(M_k)} \\ &= P(R_k) \frac{p(\mathbf{u}_k, \mathbf{v}|M_k)}{p(\mathbf{u}_k, \mathbf{v})}. \end{aligned}$$

Now, using the laws of conditional probability,

$$p(\mathbf{u}_k, \mathbf{v}) = P(M_k)\, p(\mathbf{u}_k, \mathbf{v}|M_k) + P(\bar{M}_k)\, p(\mathbf{u}_k, \mathbf{v}|\bar{M}_k)$$

and

$$P(R_k|\mathbf{u}_k, \mathbf{v}) = P(R_k) \frac{p(\mathbf{u}_k, \mathbf{v}|M_k)}{P(M_k)\, p(\mathbf{u}_k, \mathbf{v}|M_k) + P(\bar{M}_k)\, p(\mathbf{u}_k, \mathbf{v}|\bar{M}_k)}.$$

Dividing the numerator and denominator by $p(\mathbf{u}_k, \mathbf{v}|M_k)$ the result is

$$P(R_k|\mathbf{u}_k, \mathbf{v}) = P(R_k) \frac{l(\mathbf{u}_k, \mathbf{v})}{P(M_k)\, l(\mathbf{u}_k, \mathbf{v}) + P(\bar{M}_k)}, \tag{8.4}$$

where

$$l(\mathbf{u}_k, \mathbf{v}) = \frac{p(\mathbf{u}_k, \mathbf{v}|M_k)}{p(\mathbf{u}_k, \mathbf{v}|\bar{M}_k)}$$

is the likelihood ratio with respect to a match or mismatch event at position $\mathbf{p}_k$. If the models for match and mismatch in the image pair $\{\mathbf{u}_k, \mathbf{v}\}$ are taken from Chapter 4, then $l(\mathbf{u}_k, \mathbf{v})$ is equivalent to the match/mismatch likelihood ratio introduced in Chapter 5.

A Priori Probability of Match

The *a priori* probability of a match event at position $\mathbf{p}_k$, denoted $P(M_k)$, is as yet unspecified. Using the laws of conditional probability, this probability can be written in terms of $P(R_k)$ as

$$P(M_k) = P(R_k)\, P(M_k|R_k) + P(\bar{R}_k)\, P(M_k|\bar{R}_k),$$

where $\bar{R}_k$ is the event that $\mathbf{p}_k$ is *not* the position of correct register and $P(M_k|\bar{R}_k)$ is the probability that $\mathbf{u}_k$ and $\mathbf{v}$ match anyway. $P(M_k|\bar{R}_k)$ is essentially the probability of a match occurring by chance (the scenes look the same, but are not the same scene) and is denoted

P_1. Now, since $P(M_k|R_k) = 1$ (registration implies match) and $P(\bar{R}_k) = 1 - P(R_k)$,

$$P(M_k) = P(R_k)(1 - P_1) + P_1.$$

In most applications the probability of a chance match between the ideal (noise-free) scene components is very low ($P_1 \ll 1$) and

$$P(M_k) \approx P(R_k) + P_1 \tag{8.5}$$

will be a good approximation.

Simplified Rejection Threshold

The registration test chooses the position that has maximum *a posteriori* probability of correct registration *if* this probability exceeds a rejection threshold, denoted λ_R. Using the expression for $P(R_k|\mathbf{u}_k, \mathbf{v})$ in equation (8.4), the condition that must be satisfied is

$$P(R_k)\frac{l(\mathbf{u}_k, \mathbf{v})}{P(M_k)\,l(\mathbf{u}_k, \mathbf{v}) + P(\bar{M}_k)} > \lambda_R.$$

This condition can be rearranged and written as

$$l(\mathbf{u}_k, \mathbf{v}) > \frac{\lambda_R P(\bar{M}_k)}{P(R_k) - \lambda_R P(M_k)}$$

Substituting (8.5), the condition is $l(\mathbf{u}_k, \mathbf{v}) > \lambda$, where

$$\begin{aligned}\lambda &= \frac{\lambda_R(1 - P(R_k) - P_1)}{P(R_k) - \lambda_R(P(R_k) + P_1)} \\ &\approx \frac{\lambda_R(1 - P(R_k))}{P(R_k) - \lambda_R(P(R_k) + P_1)} \quad [\text{since } P_1 \ll 1].\end{aligned}$$

Since the logarithm is monotonically increasing, the condition can be written as $L(\mathbf{u}_k, \mathbf{v}) > \log\lambda$, where $L(\mathbf{u}_k, \mathbf{v})$ is the log-likelihood function for match between $\mathbf{u}_k$ and $\mathbf{v}$. From Chapter 5,

$$L(\mathbf{u}_k, \mathbf{v}) = \frac{1}{2}s(\mathbf{u}_k, \mathbf{v}) - \frac{1}{2}\log\left(\frac{|\mathbf{K}_{\mathbf{w}_1}|}{|\mathbf{K}_{\mathbf{w}_0}|}\right),$$

where $s(\mathbf{u}_k, \mathbf{v})$ is the LRT statistic for image matching, and $\mathbf{K}_{\mathbf{w}_1}$ and $\mathbf{K}_{\mathbf{w}_0}$ are the image-pair covariance matrices under the match and mismatch hypotheses, respectively. The condition

can now be written as $s(\mathbf{u}_k, \mathbf{v}) > \acute{\lambda}_R$, where

$$\acute{\lambda}_R = \log\left(\frac{|\mathbf{K}_{\mathbf{w}_1}|}{|\mathbf{K}_{\mathbf{w}_0}|}\right) + 2\log\lambda, \tag{8.6}$$

is the rejection threshold for the LRT statistic of the image pair. Making the further assumption that the *a priori* probability of correct register at any point $\mathbf{p}_k$ in the search area is far greater than the general *a priori* probability of a chance match, then $P_1 \ll P(R_k)$ and it can be shown that equation (8.6) simplifies to

$$\acute{\lambda}_R = \log\left(\frac{|\mathbf{K}_{\mathbf{w}_1}|}{|\mathbf{K}_{\mathbf{w}_0}|}\right) + 2\log\frac{\lambda_R}{1-\lambda_R} + 2\log\frac{1-P(R_k)}{P(R_k)}.$$

Simplified Registration Statistic

Combining (8.4) and (8.5), the *a posteriori* registration probability for position $\mathbf{p}_k$ can be written as

$$\begin{aligned} P(R_k|\mathbf{u}_k, \mathbf{v}) &= P(R_k)\frac{l(\mathbf{u}_k, \mathbf{v})}{P(M_k)\,l(\mathbf{u}_k, \mathbf{v}) + P(\bar{M}_k)} \\ &= P(R_k)\frac{l(\mathbf{u}_k, \mathbf{v})}{(P(R_k) + P_1)\,l(\mathbf{u}_k, \mathbf{v}) + 1}. \end{aligned} \tag{8.7}$$

Instead of maximizing (8.7) directly, a simplified, but equivalent statistic can be derived. Making the assumption that probability of registration should be far greater than the probability of a chance match in the search area, $P_1 \ll P(R_k)$, and

$$P(R_k|\mathbf{u}_k, \mathbf{v}) \approx \frac{P(R_k)\,l(\mathbf{u}_k, \mathbf{v})}{P(R_k)\,l(\mathbf{u}_k, \mathbf{v}) + 1}.$$

In this expression, the k that maximizes $P(R_k)\,l(\mathbf{u}_k, \mathbf{v})$ will also maximize $P(R_k|\mathbf{u}_k, \mathbf{v})$ and therefore the former is equivalent to the latter as a registration statistic. The logarithm is monotonically increasing, so

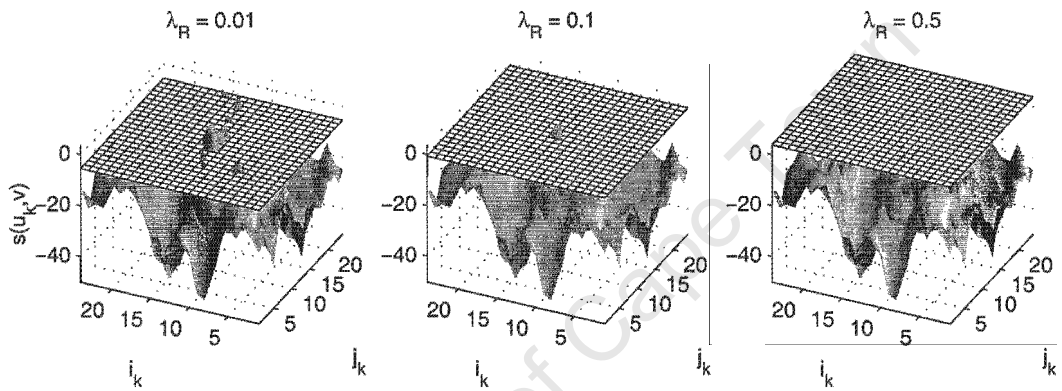
$$\log P(R_k)\,L(\mathbf{u}_k, \mathbf{v}) = \log P(R_k) + \frac{1}{2}s(\mathbf{u}_k, \mathbf{v}) - \frac{1}{2}\log\left(\frac{|\mathbf{K}_{\mathbf{w}_1}|}{|\mathbf{K}_{\mathbf{w}_0}|}\right) \tag{8.8}$$

is also equivalent in this respect. Assuming spatial stationarity in the search area, the final term in (8.8) can be dropped because it is common to all positions, and therefore

$$t(\mathbf{u}_k, \mathbf{v}) = s(\mathbf{u}_k, \mathbf{v}) + 2\log P(R_k)$$

is a simplified registration statistic that is equivalent to the *a posteriori* registration probability for position $\mathbf{p}_k$.

The result of calculating a statistic for each position in the search area is now referred to as the *match surface*. Figure 8-2 shows an example of the match surface calculated for random Markov images with the matching block at the center of the search area. Different values of the rejection threshold are shown as a horizontal plane above the surface. Notice in this case that for $\lambda_R = 0.01$ there are several false hits, whereas for $\lambda_R = 0.5$ even the correct position is rejected.



(a) Rejection threshold shown above the likelihood ratio surface.

Parameter	*Value*
Image size (n)	8
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.9
Scene one-step correlation (ρ)	0.7
SNR	1

(b) Simulation parameters.

Figure 8-2: Rejection thresholds and likelihood ratio match surfaces.

8.1.3 Prior for the Position of Correct Register

$P(R_k)$, the *a priori* probability of $\mathbf{p}_k$ being the position of correct register, must still be specified. This prior can be expressed in terms of the expected translational distortion, $\mathbf{d} = (x, y)$, where the coordinates x and y are distances from the center of the search area with units of the horizontal and vertical search increment. Given a pdf for $\mathbf{d}$, $P(R_k)$ is the probability that distortion $\mathbf{d}$ corresponds to a translational offset to position $\mathbf{p}_k$, and is given

by

$$P(R_k) = \int_{A_k} p_{\mathbf{d}}(\mathbf{d})\ d\mathbf{d},$$

where A_k is the area around position $\mathbf{p}_k$ on the search grid. For a one-pixel search increment, $A_k = 1$. A first order approximation is given by

$$P(R_k) \approx A_k \cdot p_{\mathbf{d}}(\mathbf{p}_k).$$

The pdf of $\mathbf{d}$ may follow from knowledge of the mechanisms at work in a particular application. Where detailed knowledge is not available, the prior can be based on the Bayesian definition of probability as a "subjective degree of belief" [104]. The information that is available can then be used in conjunction with a maximum entropy argument to specify the pdf [104]. For example, if the only information available are limits on $\mathbf{d}$, then a uniform pdf is the maximum entropy prior. In the most general case the size of the images could provide the limits of distortion in x and y. However, this could result in an impractical algorithm, and it may be necessary to arbitrarily truncate the extent of the search area.

Alternatively, if the covariance matrix of $\mathbf{d}$ is known (or can be estimated), entropy is maximized by a multivariate normal prior [104]. Assuming that this model is appropriate, and that the components x and y are statistically independent, the marginal pdfs $p_x(x) = N(x; m_x, \sigma_x^2)$ and $p_y(x) = N(x; m_y, \sigma_y^2)$ completely describe the random translation. For a search increment of one pixel,

$$P(R_k) = \frac{1}{2\pi\sqrt{\sigma_x^2\sigma_y^2}} \exp\left[-\frac{(i_k - m_x)^2}{2\sigma_x^2} - \frac{(j_k - m_y)^2}{2\sigma_y^2}\right],$$

where $\mathbf{p}_k = (i_k, j_k)$. It is interesting to note that previous authors have proposed a similar approach on the basis of a rather more ad-hoc argument by applying a Gaussian weighting function to the response of the similarity statistic in the search area (e.g. Mori, Kidode and Asada [105]).

8.1.4 Search Algorithm Summary

The search algorithm can now be summarized for search area $\mathbf{s}$ and $n \times n$ block $\mathbf{v}$ as follows.

1. Extract the $n \times n$ subimages of $\mathbf{s}$, $\mathbf{u}_k$ for $k \in \{1, 2, \ldots, N_p\}$, where N_p is the number of search positions in $\mathbf{s}$.

2. Find the position $\mathbf{p}_m$ in the search area that maximizes the registration statistic, where

$$m = \arg\max_k \left[t\left(\mathbf{u}_k, \mathbf{v}\right) \right].$$

3. Accept this as the position of correct register if $s\left(\mathbf{u}_m, \mathbf{v}\right) > \acute{\lambda}_R$. Then $\mathbf{p} = \mathbf{p}_m$.

Having developed a search algorithm based on hypothesis tests, attention now turns to its efficient implementation.

8.2 Efficient Block Matching Implementation

Methods for the efficient calculation of the match surface are now considered. Core operations are identified and strategies for their efficient implementation are proposed. Block matching with the LRT statistic and standard similarity measures are then formulated as a computationally efficient combination of these operations.

8.2.1 Core Operations

Calculating the match surface for search area $\mathbf{s}$ and block $\mathbf{v}$ can be viewed as an (often nonlinear) filtering operation on $\mathbf{s}$. The most computationally expensive part of this operation typically involves a calculation in one of two forms. Reverting to non-vector notation for images, the first is the cross-correlation function of $n_s \times n_s$ image s with a smaller $n \times n$ image v:

$$f\left(i + \frac{n}{2} - 1, j + \frac{n}{2} - 1\right) = \sum_{k=1}^{n} \sum_{l=1}^{n} s\left(i + k - 1, j + l - 1\right) v\left(k, l\right). \tag{8.9}$$

The second is the summation of image intensities in a $n \times n$ sliding window:

$$f\left(i + \frac{n}{2} - 1, j + \frac{n}{2} - 1\right) = \sum_{k=1}^{n} \sum_{l=1}^{n} s\left(i + k - 1, j + l - 1\right). \tag{8.10}$$

Note that in both cases n is assumed to be even and $f\left(i, j\right)$ is only defined for

$$i \in \left\{\frac{n}{2}, \ldots, n_s - \frac{n}{2}\right\} \text{ and } j \in \left\{\frac{n}{2}, \ldots, n_s - \frac{n}{2}\right\}.$$

Methods for performing these simple calculations efficiently using the Fourier transform convolution theorem and by eliminating redundant calculations are investigated here. The

results are then applied to the LRT and other similarity statistics. Alternative methods, such as the three-step search [106] and coarse-fine matching [107, 34, 10], can speed up the search, but these approaches only approximate the exhaustive search and are not considered further.

Cyclic Convolution using FFTs

The Fourier transform convolution theorem is a well-known result that is routinely used for efficient calculation of the cross-correlation function [11]. If (8.9) is viewed as a discrete cyclic convolution operation of s with v rotated by 180° as the convolution kernel, then f is the non-cyclic part of the result:

$$f = \mathcal{F}^{-1}\left[S \cdot V^*\right]$$

where $\mathcal{F}^{-1}$ denotes the inverse Fourier transform, S is the Fourier transform of s, V is the Fourier transform of v padded to the size of s with zeros, and * denotes complex conjugation. If the fast Fourier transform (FFT) algorithm is used, this calculation is $O\left(n_s^2 \log n_s^2\right)$ compared to $O\left(n^2 \left(n_s - n\right)^2\right)$ for the direct calculation.

Depending on where the origin is defined to be in the Fourier transform result, f must be rearranged to obtain the cross-correlation function. Typically the quadrants must be swapped diagonally and the non-cyclic part of the result extracted, as demonstrated in Figure 8-3. The valid cross-correlation function, denoted f', is an $n'_s \times n'_s$ image, where $n'_s = n_s - n$.

Barnea and Silverman claim that this technique can only be used to speed up calculation of the numerator in the normalized correlation function

$$f\left(i + \frac{n}{2} - 1, j + \frac{n}{2} - 1\right) = \frac{\sum_{k=1}^{n}\sum_{l=1}^{n} s\left(i+k-1, j+l-1\right) v\left(k,l\right)}{\sqrt{\sum_{k=1}^{n}\sum_{l=1}^{n} s\left(i+k-1, j+l-1\right)^2}\sqrt{\sum_{k=1}^{n}\sum_{l=1}^{n} v\left(k,l\right)^2}}$$

and that calculation of $\sum_{k=1}^{n} s\left(i+k, j+l\right)^2$ is still a time consuming windowed operation [40]. This is not true, since by letting $q\left(i,j\right) = s\left(i,j\right)^2$ and defining a as the $n \times n$ kernel of ones, this operation can be written as the cross-correlation of q with the kernel a, which *can* be computed using FFTs. In fact, since the FFT of s need only be calculated once, calculating the denominator in this way requires only one additional inverse FFT operation if the FFT of a is computed beforehand. As the next section shows, however, it is possible to do this sort of windowed summation even more efficiently using cumulative sums.

The vector notation $\mathcal{X}\left(\mathbf{s}, \mathbf{v}\right)$ is now defined for the cross-correlation of image $\mathbf{s}$ with kernel

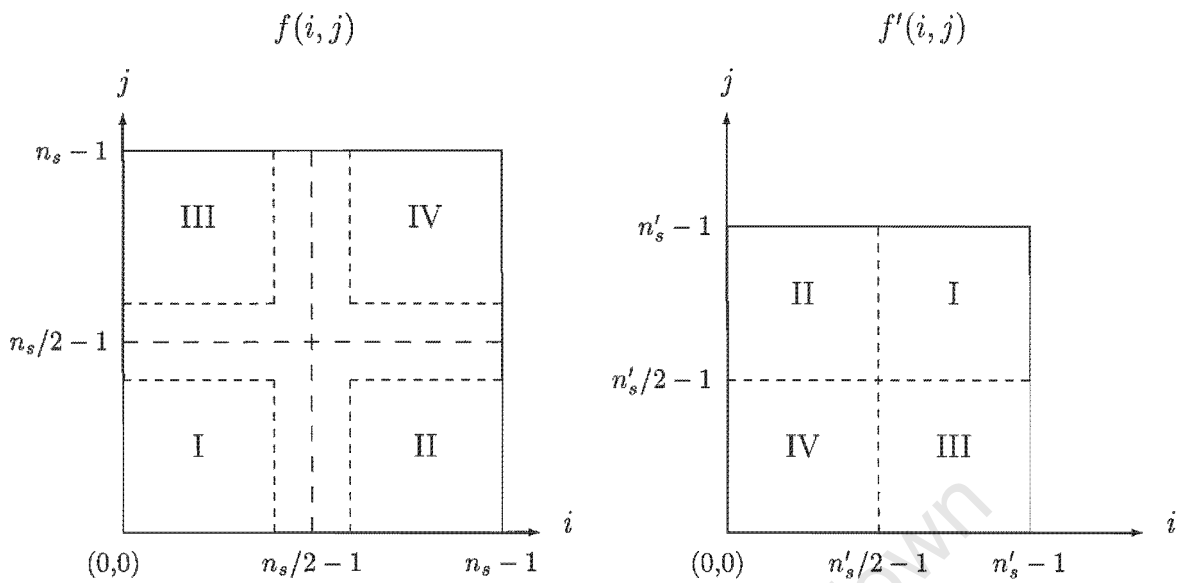


Figure 8-3: Extracting the valid cross-correlation function from the FFT convolution result.

v, where the result is a n_s^2 row-column image vector that has valid cross-correlation values stored at positions corresponding to the center of the sliding kernel.

Fast Cumulative Sums

A method is now introduced for eliminating redundant calculations when calculating the summation of image intensities in a $n \times n$ sliding window. This is the calculation

$$f(i,j) = \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} s(i+k, j+l), \forall\, i,j \in \{1,2,\ldots,n_s - n + 1\}, \tag{8.11}$$

where each result is recorded at the top left of the window (rather than its center) for notational convenience. Calculated directly, an $O\left(n^2 (n_s - n)^2\right)$ operation is required to compute f. A more efficient algorithm uses cumulative sums to reduce the number of operations required. For the sake of simplicity, consider the one dimensional signal vector **s** with n_s elements. In this case the vector of windowed sums **f** is given by

$$f_i = \sum_{j=i}^{i+n-1} s_j, \forall\, i \in \{1,2,\ldots,n_s - n + 1\}, \tag{8.12}$$

which requires $n(n_s - n)$ calculations. Define the cumulative sum vector $\mathbf{a}$ as

$$a_i = \sum_{j=1}^{i} s_j, \forall\, i \in \{1, 2, \ldots, n_s\}.$$

This vector can be calculated efficiently with the initial condition $a_1 = s_1$ and the iteration

$$a_i = a_{i-1} + s_i \quad \forall\, i \in \{2, 3, \ldots, n_s\}. \tag{8.13}$$

The windowed sums can then be calculated by setting $f_1 = a_n$ and iterating

$$f_i = a_{i+n-1} - a_{i-1} \quad \forall\, i \in \{2, 3, \ldots, n_s - n + 1\}. \tag{8.14}$$

The alternative method now involves calculating $\mathbf{a}$ using (8.13) and then using (8.14) to calculate the window sums. The algorithm requires only $(2n_s - n - 1)$ calculations, but requires additional storage for the cumulative sums. Note that storage is only required for $(n + 1)$ cumulative sums if a circular buffer is used.

Returning to two dimensional images, the situation is more complicated. Two buffers are now required for the cumulative sums. The first of these, denoted here as $\mathbf{a}$, stores cumulative sums for all rows in the image:

$$a_{i,j} = \sum_{k=1}^{j} s_{i,k}, \forall\, i, j \in \{1, 2, \ldots, n_s\},$$

which can be calculated efficiently using

$$\begin{aligned} a_{i,1} &= s_{i,1} && \forall\, i \in \{1, 2, \ldots, n_s\} \\ a_{i,j} &= a_{i,j-1} + s_{i,j} && \forall\, i \in \{1, 2, \ldots, n_s\} \text{ and } j \in \{2, 3, \ldots, n_s\}. \end{aligned} \tag{8.15}$$

The second buffer, $\mathbf{b}$, accumulates column sums of the n-pixel cumulative row sums in $\mathbf{a}$:

$$\begin{aligned} b_{i,1} &= \textstyle\sum_{k=1}^{i} a_{k,n} && \forall\, i \in \{1, 2, \ldots, n_s\} \\ b_{i,j} &= \textstyle\sum_{k=1}^{i} (a_{k,j+n-1} - a_{k,j-1}) && \forall\, i \in \{1, 2, \ldots, n_s\} \text{ and } j \in \{2, 3, \ldots, n_s - n + 1\}, \end{aligned}$$

which can be calculated efficiently using

$$\begin{aligned} b_{1,1} &= a_{1,n} \\ b_{1,j} &= a_{1,j+n-1} - a_{1,j-1} \qquad \forall\, j \in \{2,3,\ldots,n_s-n+1\} \\ b_{i,j} &= b_{i-1,j} + (a_{i,j+n-1} - a_{i,j-1}) \quad \forall\, i \in \{2,3,\ldots,n_s\} \text{ and } j \in \{2,3,\ldots,n_s-n+1\}. \end{aligned} \tag{8.16}$$

The $n \times n$ pixel window sums are then calculated using

$$\begin{aligned} f_{1,j} &= b_{n,j} \qquad \forall\, j \in \{1,2,\ldots,n_s-n+1\} \\ f_{i,j} &= b_{i+n-1,j} - b_{i-1,j} \quad \forall\, i \in \{2,3,\ldots,n_s-n+1\} \text{ and } j \in \{1,2,\ldots,n_s-n+1\}. \end{aligned} \tag{8.17}$$

The efficient algorithm calculates the buffer (8.15), then the buffer (8.16) and then calculates the $n \times n$ window sums using (8.17). The algorithm requires a total of

$$\begin{aligned} & \qquad\qquad \text{equation (8.15)} \qquad\qquad \text{equation (8.16)} \qquad\qquad \text{equation (8.17)} \\ & \qquad\qquad\quad \downarrow \qquad\qquad\qquad\quad \downarrow \qquad\qquad\qquad\quad \downarrow \\ N_{\text{cumulative}} &= n_s\,(n_s-1) + (2n_s-1)\,(n_s-n) + (n_s-n+1)\,(n_s-n) \\ &= 4n_s^2 - 4n_s n - n_s + n^2 \end{aligned}$$

operations (excluding assignments), as opposed to the

$$N_{\text{direct}} = n^2\,(n_s-n+1)^2$$

operations required of the direct calculation of equation (8.11). For example, if $n_s = 1024$ and $n = 32$, then the efficient algorithm is approximately 300 times faster than the direct calculation. If storage is limited, full image buffers for $\mathbf{a}$ and $\mathbf{b}$ are not required — the algorithm can use a buffer of size n_s for $\mathbf{a}$ and a $(n_s\,(n+1))$ size circular buffer for $\mathbf{b}$.

For clarity of presentation, the vector notation $\mathcal{S}_n\,(\mathbf{s})$ is defined for the operation that calculates sums in a sliding $n \times n$ window of image $\mathbf{s}$, where the result is a n_s^2 row-column image vector that has the sums stored at positions corresponding to the window centers and zeros elsewhere.

8.2.2 Efficient Filtering with the LRT Statistic

The block matching algorithm requires that the LRT statistic $s(\mathbf{u}_k, \mathbf{v})$ be calculated for each position in the search area $\mathbf{s}$. Define the *match surface* as the image of similarity statistics

$$\mathbf{f} : f_k = s(\mathbf{u}_k, \mathbf{v}) \ \forall\, k \in \{1, 2, \ldots, N_p\}.$$

where the subimages $\mathbf{u}_k$ are extracted from $\mathbf{s}$ in row-column order. The calculation of $s(\mathbf{u}_k, \mathbf{v})$ consists of two stages. First the images are whitened using

$$\acute{\mathbf{u}}_k = \mathbf{T}_\mathbf{u}\mathbf{u}_k \quad \text{and} \quad \acute{\mathbf{v}} = \mathbf{T}_\mathbf{v}\mathbf{v}$$

and second the statistic of the whitened images

$$s(\acute{\mathbf{u}}, \acute{\mathbf{v}}) = \sum_{i=1}^{n^2} \beta_i \left(\acute{u}_{k,i} - m_{\acute{u}_i}\right)\left(\acute{v}_i - m_{\acute{v}_i}\right) - \alpha_i \left(\left(\acute{u}_{k,i} - m_{\acute{u}_i}\right)^2 + \left(\acute{v}_i - m_{\acute{v}_i}\right)^2\right)$$

is calculated. For the purposes of this discussion the coefficients α_i and β_i are constants determined by the image-pair model.

Since $\acute{\mathbf{v}}$ is only calculated once, the whitening transform on the N_p subimages $\mathbf{u}_k$ is the computationally demanding part of the calculation. Each pixel in the whitened image $\acute{\mathbf{u}}_k$ is the projection of $\mathbf{u}_k$ onto a basis image of the transformation $\mathbf{T}_\mathbf{u}$. Pixel $\acute{u}_{k,i}$ corresponds to the basis image that is the i-th column of $\mathbf{T}_\mathbf{u}^T$. Denoting this basis image as the column vector $\mathbf{t}_\mathbf{u}^i$, $\acute{u}_{k,i} = \mathbf{t}_\mathbf{u}^i \cdot \mathbf{u}_k$. Now the i-th whitened pixel can be calculated for all subimages in the search area by computing the cross-correlation of $\mathbf{s}$ and $\mathbf{t}_\mathbf{u}^i$. Using the vector notation introduced for cross-correlation in Section 8.2.1, this can be written as

$$\acute{\mathbf{s}}^i = \mathcal{X}\left(\mathbf{s}, \mathbf{t}_\mathbf{u}^i\right).$$

The k-th pixel in the match surface for $\mathbf{s}$ and $\mathbf{v}$ can now be written as

$$f_k = \sum_{i=1}^{n^2} \beta_i \left(\acute{s}_k^i - m_{\acute{u}_i}\right)\left(\acute{v}_i - m_{\acute{v}_i}\right) - \alpha_i \left(\left(\acute{s}_k^i - m_{\acute{u}_i}\right)^2 + \left(\acute{v}_i - m_{\acute{v}_i}\right)^2\right),$$

where $\acute{s}_k^i$ is the corresponding pixel in $\acute{\mathbf{s}}^i$, the i-th basis expansion of the search area.

An efficient method for the search algorithm can now be summarized as follows:

1. Compute the whitened transform $\acute{\mathbf{v}} = \mathbf{T}_\mathbf{v}\mathbf{v}$.

2. Perform the cross-correlation of search area image $\mathbf{s}$ and each basis image $\mathbf{t}_{\mathbf{u}}^{i}$, giving $\acute{\mathbf{s}}^{i} = \mathcal{X}\left(\mathbf{s}, \mathbf{t}_{\mathbf{u}}^{i}\right)$.

3. Calculate the match surface contribution associated with each basis image:

$$\mathbf{f}^{i} : f_{k}^{i} = \beta_{i}\left(\acute{s}_{k}^{i} - m_{\acute{u}_{i}}\right)\left(\acute{v}_{i} - m_{\acute{v}_{i}}\right) - \alpha_{i}\left(\left(\acute{s}_{k}^{i} - m_{\acute{u}_{i}}\right)^{2} + \left(\acute{v}_{i} - m_{\acute{v}_{i}}\right)^{2}\right).$$

4. Sum these contributions to obtain the match surface: $\mathbf{f} = \sum_{i=1}^{n^2} \mathbf{f}^{i}$.

Most of the computation in this procedure is found in the $\left(1 + 2n^2\right)$ FFT operations on $n_s \times n_s$ images that are required to compute the cross-correlations of the search area with the basis images. The algorithm is therefore $O\left(n^2 n_s^2 \log_2 n_s\right)$ as opposed to $O\left(n^2\left(n^2+1\right)\left(n_s - n\right)^2\right)$ for the direct calculation.

Note that computation can be further reduced by using the lossy canonical subset of Chapter 7. If the whitened pixel pairs are ordered with decreasing canonical correlation coefficient, then $\mathbf{f} = \sum_{i=1}^{c} \mathbf{f}^{i}$ is the optimal approximation of the match surface that uses only $c < n^2$ basis images. Another lossy strategy from Chapter 7 uses the DCT as an approximation of the whitening transform, reducing block matching with the LRT statistic to an $O\left(\left(n_s - n\right)^2 n^2 \log_2 n\right)$ algorithm.

8.2.3 Efficient Filtering with other Similarity Statistics

Efficient algorithms are now proposed for computing the match surface for other similarity statistics. In each case the computation will be expressed in terms of cross-correlation and windowed summation operations. Calculating the inner product of every subimage $\mathbf{u}_k$ with $\mathbf{v}$ is equivalent to performing cross-correlation of the search area $\mathbf{s}$ with $\mathbf{v}$. Likewise, the sum of the pixels in every subimage $\mathbf{u}_k$ corresponds to windowed summation in the search area $\mathbf{s}$. Formally

$$\mathbf{f} = \mathcal{X}\left(\mathbf{s}, \mathbf{v}\right) \iff f_k = \sum_{i=1}^{n^2} u_{k,i} v_i \;\forall\; k \in \{1, 2, \ldots, N_p\} \tag{8.18}$$

and

$$\mathbf{f} = \mathcal{S}_n\left(\mathbf{s}\right) \iff f_k = \sum_{i=1}^{n^2} u_{k,i} \;\forall\; k \in \{1, 2, \ldots, N_p\}, \tag{8.19}$$

respectively. These two operations are the basis of the efficient calculations presented here.

The Sum of Squared Differences

This statistic can be expanded as follows:

$$\begin{aligned} d_2(\mathbf{u}_k, \mathbf{v}) &= \sum_{i=1}^{n^2} (u_{k,i} - v_i)^2 \\ &= \sum_{i=1}^{n^2} u_{k,i}^2 - 2\sum_{i=1}^{n^2} u_{k,i} v_i + E_v, \end{aligned}$$

where

$$E_v = \sum_{i=1}^{n^2} v_i^2$$

and need only be calculated once. Using (8.18) and (8.19) the match surface can now be written as a function of the search area as follows,

$$\mathbf{f} = \mathcal{S}_n\left([\mathbf{s}]^2\right) - 2\mathcal{X}(\mathbf{s}, \mathbf{v}) + E_v,$$

where $[\mathbf{s}]^2$ denotes a vector that has elements s_i^2. The FFT operation in the cross-correlation term dominates the computation and therefore this method is $O\left(n_s^2 \log_2 n_s\right)$ as opposed to $O\left(n^2 (n_s - n)^2\right)$ for the direct calculation.

The Correlation Coefficient

The sample correlation coefficient (Pearson's r) is given by

$$r(\mathbf{u}_k, \mathbf{v}) = \frac{\sum_{i=1}^{n^2} (u_{k,i} - m(\mathbf{u}_k))(v_i - m(\mathbf{v}))}{\sqrt{\sum_{i=1}^{n^2} (u_{k,i} - m(\mathbf{u}_k))^2 \sum_{i=1}^{n^2} (v_i - m(\mathbf{v}))^2}}, \text{ where } m(\mathbf{u}) = \frac{1}{n^2}\sum_{i=1}^{n^2} u_i.$$

The match surface is now expressed in terms of a numerator $\mathbf{x}$ and denominator $\mathbf{y}$ as

$$\mathbf{f} : f_k = \frac{x_k}{y_k} \; \forall \; k \in \{1, 2, \ldots, N_p\}.$$

Denoting $\hat{\mathbf{v}} = \mathbf{v} - m(\mathbf{v})$, the numerator of r can be expanded to

$$\begin{aligned}\sum_{i=1}^{n^2} (u_{k,i} - m(\mathbf{u}_k))\,\hat{v}_i &= \sum_{i=1}^{n^2} u_{k,i}\hat{v}_i - m(\mathbf{u}_k) \sum_{i=1}^{n^2} \hat{v}_i \\ &= \sum_{i=1}^{n^2} u_{k,i}\hat{v}_i \text{ [since } \sum_{i=1}^{n^2} \hat{v}_i = m(\hat{\mathbf{v}}) = 0],\end{aligned}$$

which corresponds to the search area operation

$$\mathbf{x} = \mathcal{X}(\mathbf{s}, \hat{\mathbf{v}}). \tag{8.20}$$

The square of the denominator of r can be written as

$$\sum_{i=1}^{n^2} (u_{k,i} - m(\mathbf{u}_k))^2 \sum_{i=1}^{n^2} \hat{v}_i^2 = E_{\hat{v}} \cdot \left[\sum_{i=1}^{n^2} u_{k,i}^2 - \frac{1}{n^2} \left(\sum_{i=1}^{n^2} u_{k,i} \right)^2 \right] \tag{8.21}$$

where $E_{\hat{v}}$ is the once-off calculation

$$E_{\hat{v}} = \sum_{i=1}^{n^2} \hat{v}_i^2.$$

Using (8.21) the denominator of the sample correlation coefficient corresponds to the search area operation

$$\mathbf{y} = \sqrt{E_{\hat{v}} \cdot \left(\mathcal{S}_n\left([\mathbf{s}]^2\right) - \frac{1}{n^2} \left[\mathcal{S}_n(\mathbf{s})\right]^2 \right)}, \tag{8.22}$$

where the square root is taken element by element.

The element-by-element ratio of (8.20) and (8.22) now provides the match surface for the sample correlation coefficient. The FFT operation in the numerator dominates the computation and therefore this method is $O\left(n_s^2 \log_2 n_s\right)$ as opposed to $O\left(n^2 (n_s - n)^2\right)$ for the direct calculation.

8.3 Selection of Control Points

The previous sections of this chapter developed an efficient hypothesis testing procedure for matching the block $\mathbf{v}$ from a single position in the first image to a search area $\mathbf{s}$ in the second

image in order to find the subimage $\mathbf{u}_k$ that represents registration of the two images at that position. The full registration procedure must perform this task for all positions in the first image. This is normally impractical from a computational point of view, however, and a subset of the image positions are selected as control points. Aside from computational economy, this approach can also improve the error-rate by not using blocks that typically provide an unreliable registration result. For images with isolated edges separated by large flat areas, for example, it is blocks containing the former that will support reliable registration. Researchers have proposed the use of interest operators for the task of finding areas of the image with information that supports registration (for example, the approach of Barnard and Thompson [74]). This section seeks to formalize this practice within the framework of the hypothesis testing procedure derived earlier.

8.3.1 Control Point Screening with an Absolute Condition

Chapter 3 derived optimal tests for matching scalars to give insight into the more complex problem of image matching. One observation made there was that the partition induced by the test on the space of scalars (u, v) sometimes leaves certain values of u and v with no potentially matching counterparts. The critical regions in Figure 8-4 for SNR = 2 and SNR = 3 are examples of this scenario. This observation is also relevant for images and can

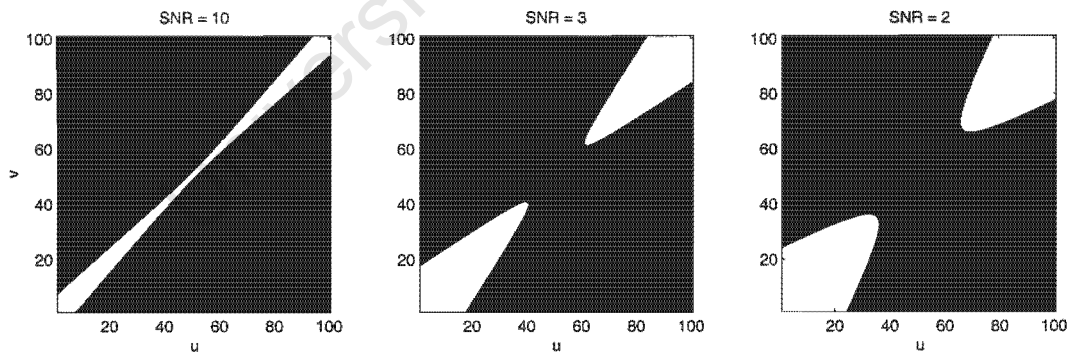


Figure 8-4: Critical regions for optimal scalar matching.

be used to select control points — if some images have no potential matching counterparts in the partition induced by the rejection hypothesis test, then using them for registration is pointless. This condition can be used to screen control points according to whether the blocks around them have potential counterparts in the second image.

Blocks $\mathbf{v}$ fall into this category if the rejection hypothesis is supported for all possible

counterparts $\mathbf{u}$ and positions k, that is, if $s(\mathbf{u},\mathbf{v}) \leq \min_k \acute{\lambda}_R$ for all $\mathbf{u}$, where

$$\begin{aligned}\min_k \acute{\lambda}_R &= \log\left(\frac{|\mathbf{K}_{\mathbf{w}_1}|}{|\mathbf{K}_{\mathbf{w}_0}|}\right) + 2\log\frac{\lambda_R}{1-\lambda_R} + \min_k\left[2\log\frac{1-P(R_k)}{P(R_k)}\right] \\ &= \log\left(\frac{|\mathbf{K}_{\mathbf{w}_1}|}{|\mathbf{K}_{\mathbf{w}_0}|}\right) + 2\log\frac{\lambda_R}{1-\lambda_R} + 2\log\frac{1-\max_k P(R_k)}{\max_k P(R_k)}.\end{aligned}$$

An equivalent statement of this condition is

$$\max_{\mathbf{u}} s(\mathbf{u},\mathbf{v}) \leq \min_k \acute{\lambda}_R,$$

or, written in terms of the LRT statistic for whitened images $\acute{\mathbf{u}}$ and $\acute{\mathbf{v}}$, it is

$$\max_{\acute{\mathbf{u}}} s(\acute{\mathbf{u}},\acute{\mathbf{v}}) \leq \min_k \acute{\lambda}_R.$$

Now the statistic can be written as the sum of n^2 statistics on the individual pixel pairs

$$\max_{\acute{\mathbf{u}}} s(\acute{\mathbf{u}},\acute{\mathbf{v}}) = \sum_{i=1}^{n^2} \max_{\acute{u}_i} s_i(\acute{u}_i,\acute{v}_i)$$

where

$$s_i(\acute{u}_i,\acute{v}_i) = \beta_i(\acute{u}_i - m_{\acute{u}_i})(\acute{v}_i - m_{\acute{v}_i}) - \alpha_i\left((\acute{u}_i - m_{\acute{u}_i})^2 + (\acute{v}_i - m_{\acute{v}_i})^2\right)$$

and where α_i and β_i are defined in Section 5.2. Since the statistic is a sum of independent pixel-pairs, the pixel values $\acute{u}_i$ that maximize $s(\acute{\mathbf{u}},\acute{\mathbf{v}})$ can be found separately using elementary calculus. First, taking the derivative of $s_i(\acute{u}_i,\acute{v}_i)$ with respect to $\acute{u}_i$:

$$\frac{d}{d\acute{u}_i} s_i(\acute{u}_i,\acute{v}_i) = \beta_i(\acute{v}_i - m_{\acute{v}_i}) - 2\alpha_i(\acute{u}_i - m_{\acute{u}_i})$$

Equating the result to zero and solving for $\acute{u}_i$, it is seen that

$$\acute{u}_i = \frac{\beta_i}{2\alpha_i}(\acute{v}_i - m_{\acute{v}_i}) + m_{\acute{u}_i}$$

maximizes $s_i(\acute{u}_i, \acute{v}_i)$. Using the definitions of α_i and β_i this maximum is

$$\begin{aligned}\max_{\acute{u}_i} s_i(\acute{u}_i, \acute{v}_i) &= \left(\frac{\beta_i^2}{4\alpha_i} - \alpha_i\right)(\acute{v}_i - m_{\acute{v}_i})^2 \\ &= \frac{\rho_1 - \rho_0}{\rho_1 + \rho_0}(\acute{v}_i - m_{\acute{v}_i})^2.\end{aligned}$$

Combining the independent maxima, the overall maximum of s is seen to be

$$\max_{\acute{\mathbf{u}}} s(\acute{\mathbf{u}}, \acute{\mathbf{v}}) = \frac{\rho_1 - \rho_0}{\rho_1 + \rho_0}\sum_{i=1}^{n^2}(\acute{v}_i - m_{\acute{v}_i})^2$$

and therefore the condition that $\acute{\mathbf{v}}$ must satisfy in order to ensure that it has a potential counterpart is

$$\frac{\rho_1 - \rho_0}{\rho_1 + \rho_0}\sum_{i=1}^{n^2}(\acute{v}_i - m_{\acute{v}_i})^2 > \acute{\lambda}_R. \tag{8.23}$$

For convenience the screening procedure for image $\mathbf{v}$ is written as the test $g(\acute{\mathbf{v}}) \leq \underline{\lambda_g \text{ with}}$ the test statistic

$$g(\acute{\mathbf{v}}) = \sum_{i=1}^{n^2}(\acute{v}_i - m_{\acute{v}_i})^2$$

and decision threshold

$$\lambda_g = \frac{\rho_1 + \rho_0}{\rho_1 - \rho_0}\left[\log\left(\frac{|\mathbf{K}_{\mathbf{w}_1}|}{|\mathbf{K}_{\mathbf{w}_0}|}\right) + 2\log\frac{\lambda_R}{1 - \lambda_R} + 2\log\frac{1 - \max_k P(R_k)}{\max_k P(R_k)}\right].$$

8.3.2 Effectiveness of Absolute Screening

From the point of view of computational efficiency, the effectiveness of the screening procedure will depend on the number of blocks that it can eliminate. This is related to the probability of eliminating a random block, which is derived here. If the random variables x_i are normal with mean m_i and variance σ_i^2, and are independent of each other, then

$$z = \sum_{i=1}^{k}\left(\frac{x_i - m_i}{\sigma_i}\right)$$

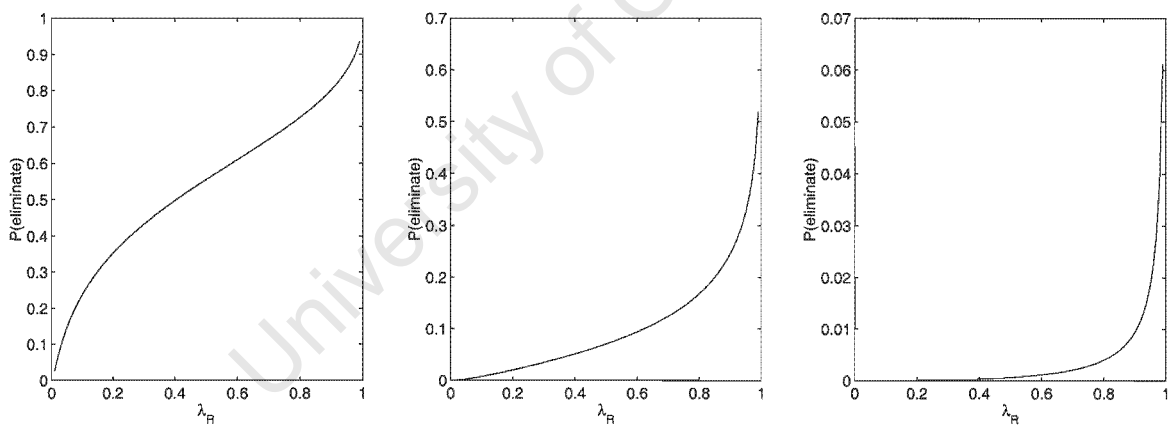
has a chi-square distribution with k degrees of freedom [61, p. 242]. Hence, noting that the whitened pixels $\acute{v}_i$ have unit variance, the screening test statistic $g(\acute{\mathbf{v}})$ has a chi-square

distribution with n^2 degrees of freedom. The probability that an image will be discarded is then

$$P\left(g\left(\hat{\mathbf{v}}\right) \leq \lambda_g\right) = \int_{-\infty}^{\lambda_g} p_{\chi^2}\left(x\right)\ dx = \int_{0}^{\lambda_g} p_{\chi^2}\left(x\right)\ dx,$$

where $p_{\chi^2}\left(x\right)$ is the chi-square pdf. This probability depends on the threshold λ_g, which is a function of the match and mismatch hypothesis conditional image-pair covariance matrices, the rejection threshold λ_R, and the maximum *a priori* probability of correct register for all positions in the search area — $\max_k P\left(R_k\right)$.

Figure 8-5 plots the probability of eliminating a randomly chosen block over λ_R. Recall that λ_R is the minimum *a posteriori* registration probability that is accepted by the rejection hypothesis. It appears that this screening procedure will only be effective for eliminating blocks in the extreme case of small images. An effective procedure for larger images must be based on a stricter condition, or one that is not absolute, but rather based on a relative comparison of registration suitability amongst the available control points.



(a) 4 × 4 blocks. (b) 5 × 5 blocks. (c) 6 × 6 blocks.

Figure 8-5: Probability of elimination for a block screening procedure ($\rho = 0.95$, SNR $= 3$, $\rho_1 = 0.6$, $\rho_0 = 0.0$, $\sigma_d = 20$).

8.3.3 Relative Control Point Comparison

The quantity $\max_{\mathbf{u}} s\left(\mathbf{u}, \mathbf{v}\right)$ can also be used to assess the registration suitability of $\mathbf{v}$ in a relative sense. If $\max_{\mathbf{u}} s\left(\mathbf{u}, \mathbf{v}_1\right) \gg \max_{\mathbf{u}} s\left(\mathbf{u}, \mathbf{v}_2\right)$, then it can be argued that $\mathbf{v}_1$ is more likely

than $\mathbf{v}_2$ to find a good match in any search area, since $s(\mathbf{u}, \mathbf{v})$ is monotonically increasing with the likelihood of match between $\mathbf{u}$ and $\mathbf{v}$. Figures 8-6 and 8-7 show the screening statistic for each position in a real and a synthetic image, respectively. The blocks with the maximum and minimum statistic are extracted and displayed alongside these images, illustrating that the former contain more useful information for matching purposes.

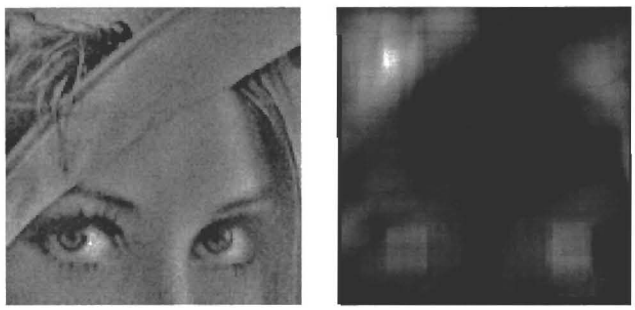
A simple procedure for selecting control points on the basis of this information would select a well-spaced group of points that report high values for the screening statistic in their vicinity. Figure 8-8 shows the result of registration performed for two identical scenes with additive noise (SNR = 2). Subfigure (a) shows the result for a uniform grid of control points, whereas in (b) the best control points were chosen for local areas using the screening statistic. Since the scenes are identical, the correct result for each control point is zero translation. It is clear that choosing control points using the screening statistic reduces the number of registration errors — the only two significant errors for the best control points are obvious examples of ambiguity in the scene (see the two erroneous displacement vectors on the brim of the hat in Figure 8-8). This improvement is achieved by avoiding areas that do not have distinguishing features, such as the forehead in Figure 8-8.

8.4 Monte Carlo Experiments

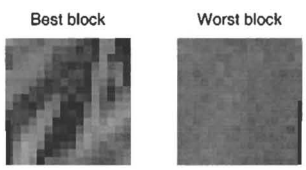
The block matching procedure based on hypothesis testing is now analyzed and compared to the standard methods using Monte Carlo simulation experiments. Three types of experiment are performed. The first computes the average match surface of a similarity statistic over many trials, which provides a qualitative indication of matching performance. The second experiment analyses the registration errors in simulated block matching for a quantitative measure of matching performance. The third adds two different samples of a synthetic noise field to a single real image in order to produce an image pair, and performs block matching for randomly selected control points.

8.4.1 Match Surfaces

Figure 8-9 compares the match surfaces of several registration statistics that have been averaged over 1000 trials. The surfaces have been normalized to make them comparable. Denoting the variance in peak amplitude as σ_p^2 and the mean and variance of the background level as m_b and σ_b^2 respectively, the normalized match surfaces $\hat{\mathbf{f}}$ are calculated from the original match

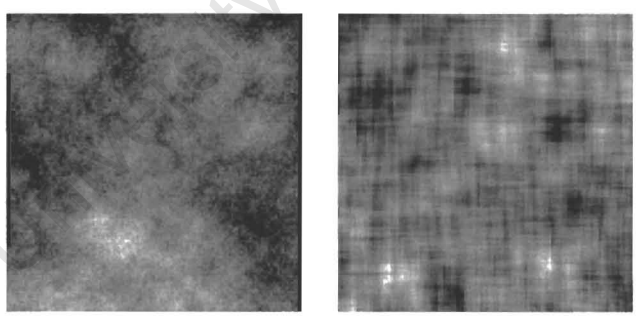


(a) Original image. (b) Screening statistic.

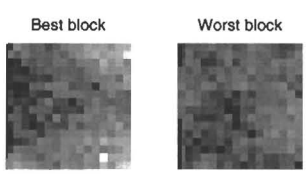


(c) Blocks.

Figure 8-6: Control point comparisons for a real image.

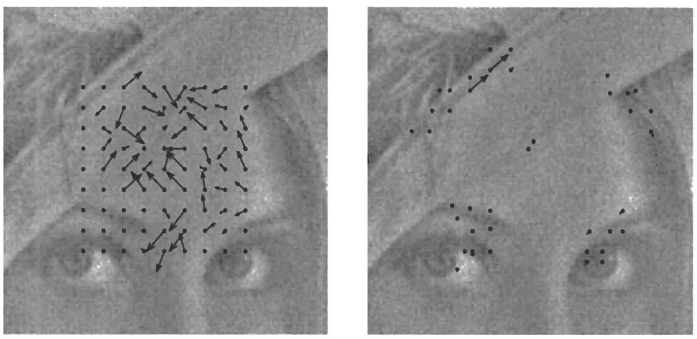


(a) Original image. (b) Screening statistic.



(c) Blocks.

Figure 8-7: Control point comparisons for a synthetic image.



(a) Uniform grid of control points.

(b) Best control points according to the screening statistic.

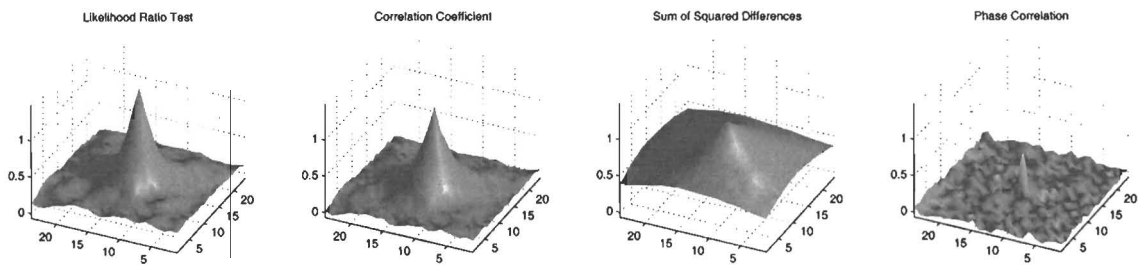
Figure 8-8: Control point selection in a registration experiment.

surfaces $\mathbf{f}$ using

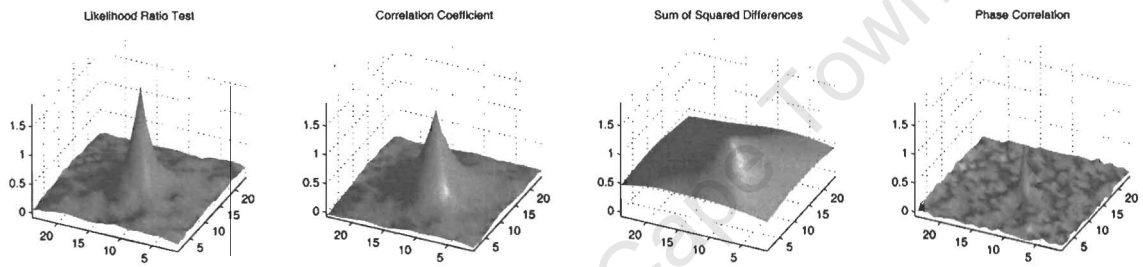
$$\hat{\mathbf{f}} = \frac{\mathbf{f} - m_b \cdot \mathbf{1}}{\sqrt{\sigma_p^2 + \sigma_b^2}}.$$

The values of m_b and σ_b^2 are obtained from separate Monte Carlo trials where no match for the block is present in the search area.

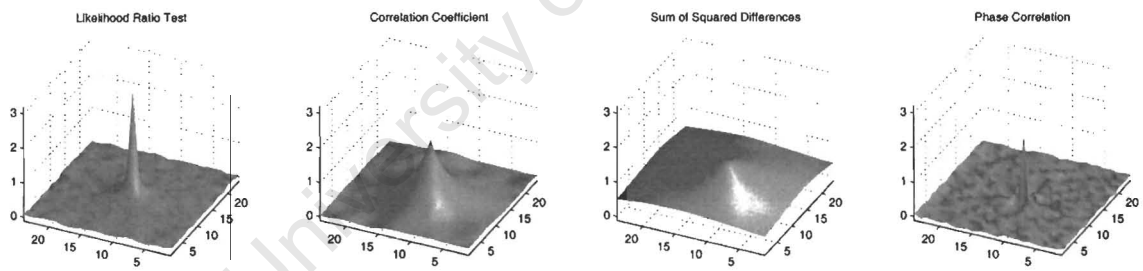
The results given in Figure 8-9 indicate that the optimal registration statistic has a more prominent peak than the suboptimal measures do. Phase correlation is investigated for the first time in this experiment, and exhibits poor performance for low SNR in Figure 8-9(a). This is expected, since the images are corrupted by additive white noise, whereas phase correlation is purportedly effective in narrowband noise [18]. One might expect that this technique would be effective for high SNR and low match correlation coefficient, because the difference between matching images for $\rho_1 < 1$ could be viewed as narrowband noise. Even in this case, however, phase correlation is no better than the correlation coefficient in Figure 8-9(c), and is excluded from further experiments. The next section provides more quantitative evidence of the optimal test's superiority over standard similarity statistics in the block matching application.



(a) SNR = 2.



(b) SNR = 3.



(c) SNR = 10.

Parameter	*Value*
Image size (n)	8
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
Scene one-step correlation (ρ)	0.95

(d) Simulation parameters.

Figure 8-9: Normalized Monte Carlo match surfaces for various similarity statistics.

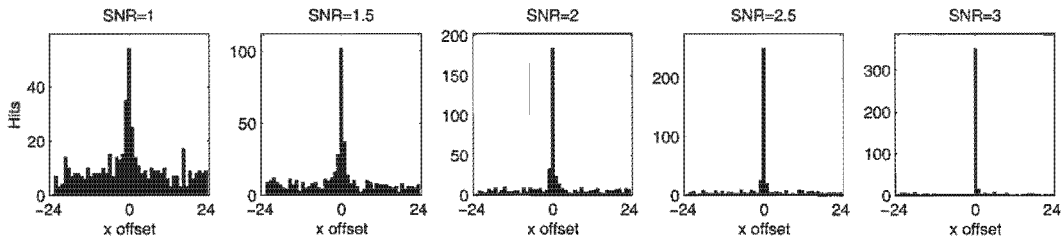
8.4.2 Block Matching Registration Error

Registration errors fall into two categories. First, there is the minor misalignment caused when the match surface peak is positionally inaccurate. Second, there is the *miss*, or completely spurious position of correct register. The former is inaccuracy of correct registration, whereas the latter leads to incorrect registration with an arbitrary position in the search area. In the experiments it is necessary to distinguish between these two types of error. For example, Svedlow, McGillem and Anuta regard any "substantial deviation" as an unsuccessful registration attempt [35], and Bhat and Nayar assume that any error greater than pixel-size constitutes a miss [25]. The latter approach is used here to count the number of registration misses in an experiment. Accuracy of correct registration is not investigated.

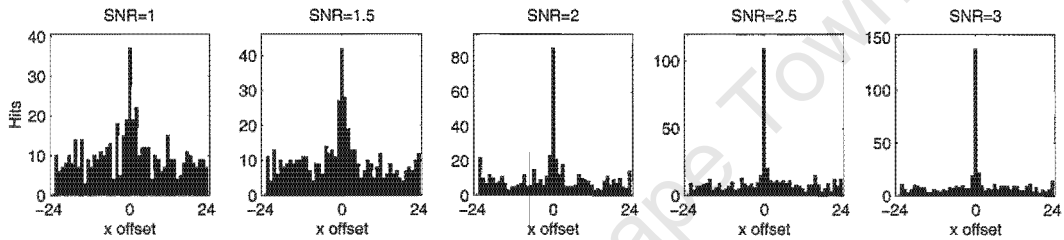
Standard Block Matching

This is the simplest form of translational image registration. The block is translated to every position in the search area and the position that reports the largest registration statistic is chosen as the position of correct register. Each Monte Carlo trial simulates this scenario with the true position of correct register at the centre of the search area. A random $n_s \times n_s$ search area $\mathbf{s}$ is generated. An $n \times n$ block is extracted from the centre of the scene component of the search area and used in conjunction with the image pair synthesis equations and the specified match correlation coefficient to create the matching block $\mathbf{v}$. The match surface is generated for $\mathbf{s}$ and $\mathbf{v}$ using each similarity statistic and the position of the peak is stored in each case. After T trials, the number of misses is counted and used to calculate the probability of a miss for the block matching algorithm based on each similarity statistic. The registration statistics compared here are the optimal registration statistic t, the correlation coefficient r, and the sum of squared differences d_2.

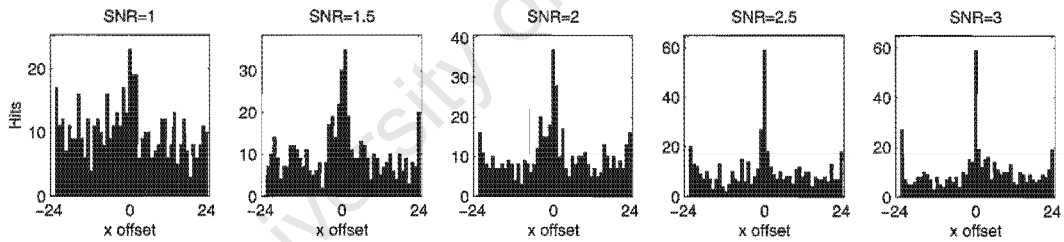
Figures 8-10 and 8-11 show the distribution of horizontal offsets from the correct position that were reported by the chosen registration statistics for a range of SNR and match correlation coefficients. Hits at any position other than zero offset represent registration errors, or misses. The block matching algorithm based on the optimal statistic is clearly superior in these results, with fewer misses in every case. A more concise comparison of the different methods is obtained by plotting the number of misses that were reported during a series of Monte Carlo trials. Figures 8-12 and 8-13 show the miss-rate against block size and SNR respectively. It is evident that for low match correlation coefficient ($\rho_1 = 0.6$) the optimal hypothesis test is orders of magnitude better than procedures based on the standard statis-



(a) Optimal block matching.



(b) Correlation coefficient.

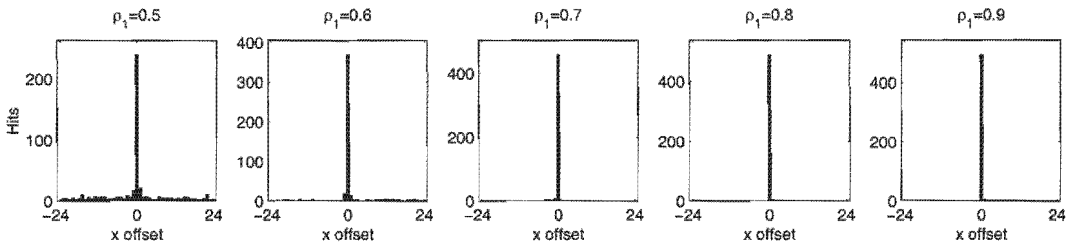


(c) Sum of squared differences.

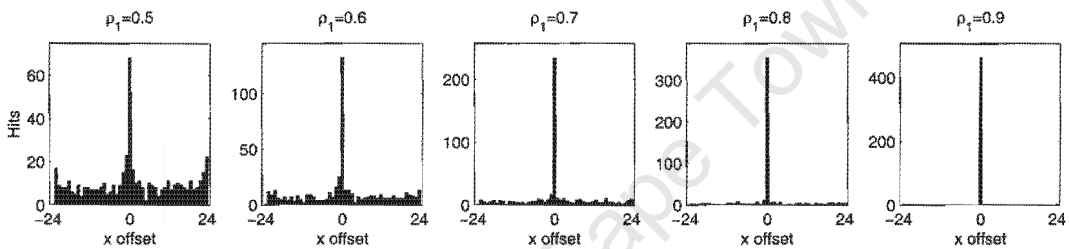
Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
Scene one-step correlation (ρ)	0.95

(d) Simulation parameters.

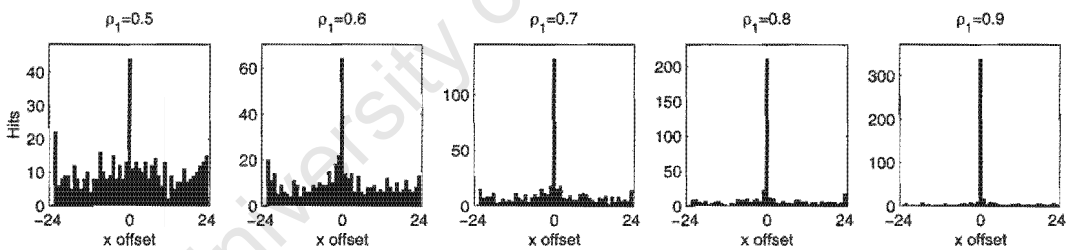
Figure 8-10: Distribution of horizontal offsets from the true position of correct register for a range of SNR values.



(a) Optimal block matching.



(b) Correlation coefficient.



(c) Sum of squared differences.

Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Scene one-step correlation (ρ)	0.95
SNR	3

(d) Simulation parameters.

Figure 8-11: Distribution of horizontal offsets from the true position of correct register for a range of match correlation coefficients.

tics. For match correlation coefficient approaching unity the advantage is less prominent, but still significant. These conclusions are also supported by the results of an experiment that investigates miss-rate as a function of match correlation coefficient in Figure 8-14. Figure 8-15 shows the miss-rate as a function of the degree of spatial correlation in the images. The trends here are similar to those that were reported for the image matching experiments in Chapter 6 in that the miss rate increases with increasing spatial correlation. Again the performance advantage of the optimal test is most prominent for low match correlation coefficient.

Block Matching with a Rejection Hypothesis

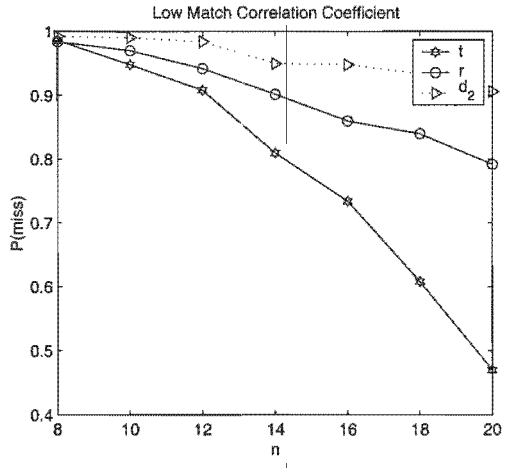
The value of a test with the rejection hypothesis is investigated in this section. Figure 8-16 shows the horizontal offset distribution of positions chosen by the optimal test. Figure 8-16(a) shows the hits accepted and rejected by the rejection hypothesis when the true position of correct register is in the center of the search area. Here the hits with non-zero offset are false alarms elsewhere in the search area. Figure 8-16(b) shows the same information for the case where there is no position of correct register in the search area. Here all of the hits are false alarms. In both (a) and (b) the distributions are given for a range of λ_R, the rejection threshold for the *a posteriori* probability of correct registration.

These results show that the number of false alarms can be significantly reduced by using the rejection threshold. Consider, for example, the column of results for $\lambda_R = 0.3$ in Figure 8-16. In (a), where the position of correct register is the center of the search area, the number of false alarms is reduced to a fraction of its former value at the cost of approximately one third of the true hits obtained without using λ_R. In (b) the false hits are reduced to a few percent of the number that would be obtained without using a rejection hypothesis.

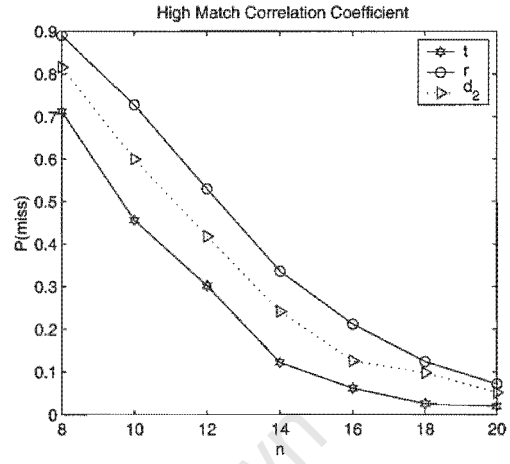
In Figure 8-16 the rejection hypothesis at $\lambda_R = 0.5$ almost eliminates false hits, at the same time sacrificing one third to one half of the true hits. In an image registration application, this strategy will yield fewer registered control points, each with a greater likelihood of being correct.

Block Matching with a Positional Prior

Prior knowledge about the mechanism of translational distortion in a block matching application can be incorporated into the test by specifying $P(R_k)$, the *a priori* probability that $\mathbf{p}_k$ is the position of correct register. Up to this point it has been assumed that $P(R_k)$ is uniform across valid positions in the search area. An alternative is to use another pdf for the

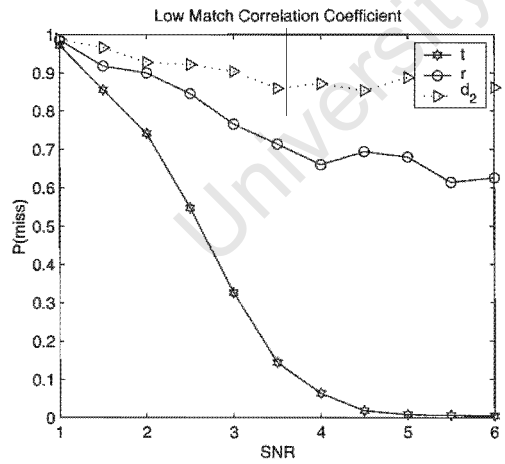


Parameter	*Value*
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
Scene one-step correlation (ρ)	0.95
SNR	2

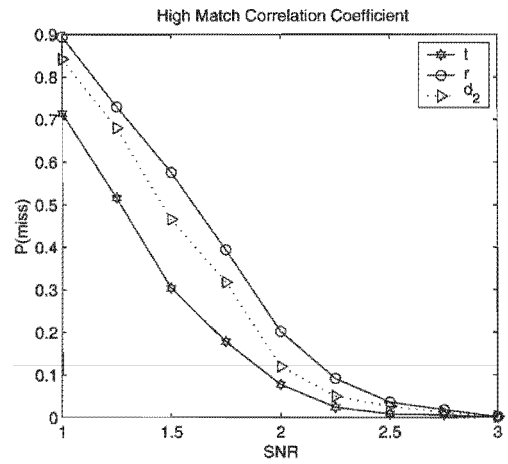


Parameter	*Value*
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.99
Scene one-step correlation (ρ)	0.95
SNR	2

Figure 8-12: Monte Carlo investigation of miss rate versus block size ($T = 500$).

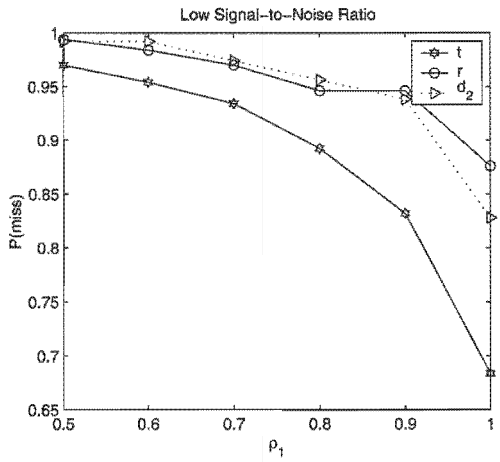


Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
Scene one-step correlation (ρ)	0.95

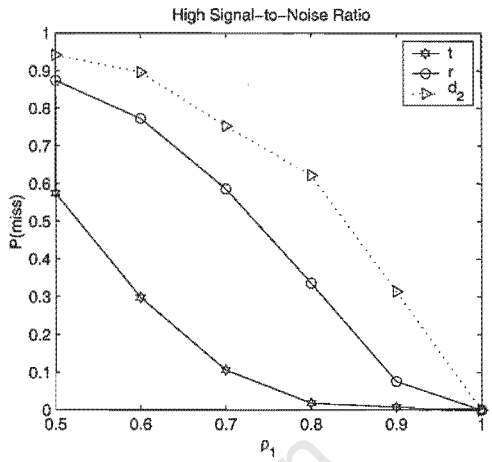


Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.99
Scene one-step correlation (ρ)	0.95

Figure 8-13: Monte Carlo investigation of miss rate versus image SNR ($T = 500$).

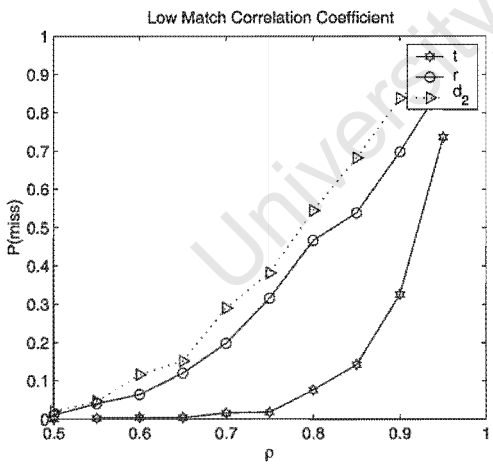


Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Scene one-step correlation (ρ)	0.95
SNR	1

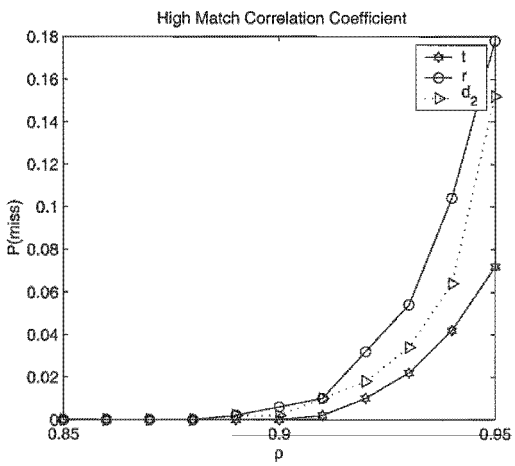


Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Scene one-step correlation (ρ)	0.95
SNR	3

Figure 8-14: Monte Carlo investigation of miss rate versus image match correlation coefficient ($T = 500$).

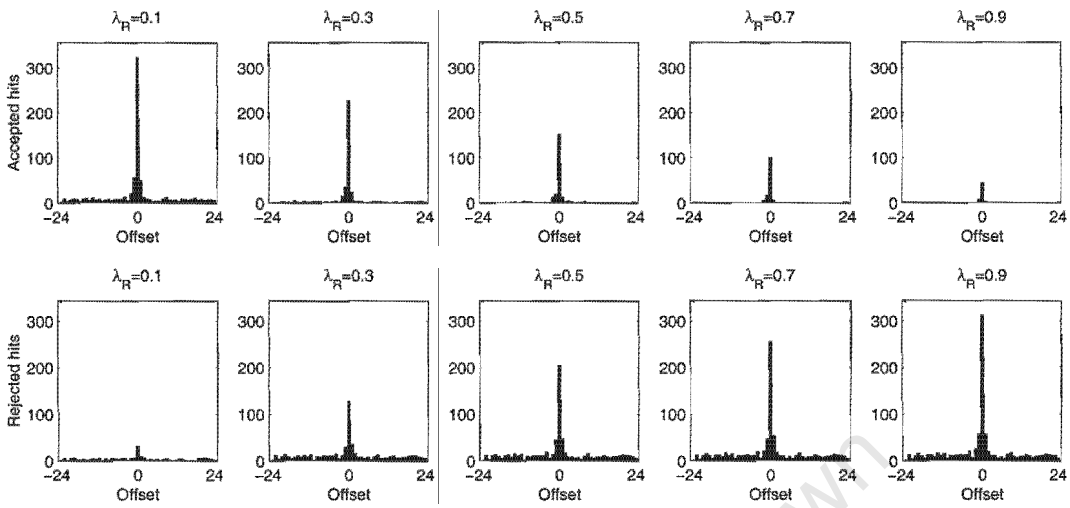


Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
SNR	2

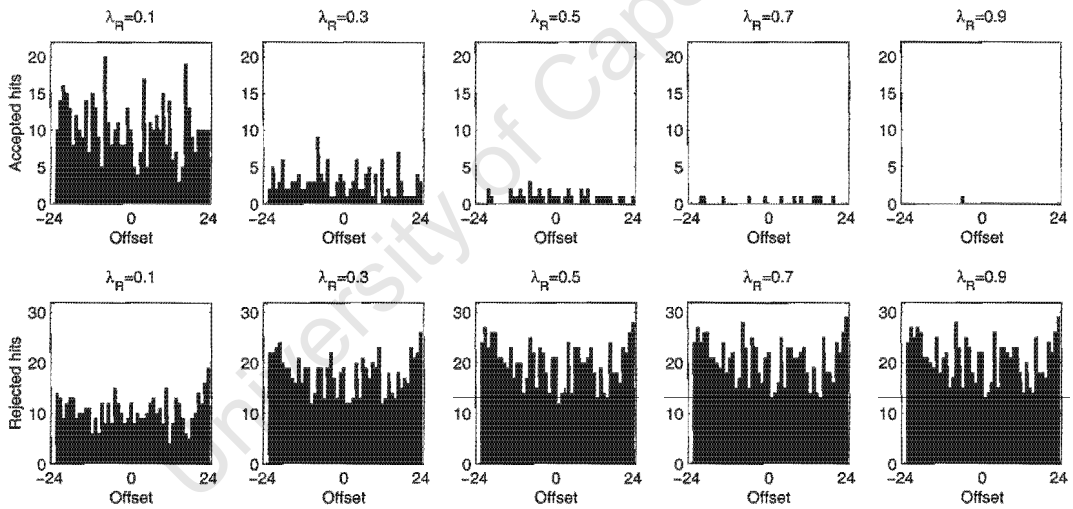


Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.99
SNR	2

Figure 8-15: Monte Carlo investigation of miss rate versus image spatial correlation coefficient ($T = 500$).



(a) Accepted (top row) and rejected (bottom row) hits with position of correct register at centre of search area.



(b) Accepted (top row) and rejected (bottom row) hits with position of correct register not present in search area.

Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
Scene one-step correlation (ρ)	0.95
SNR	2

(c) Simulation parameters.

Figure 8-16: Distribution of horizontal offsets from the true position of correct register for the test with a rejection hypothesis.

expected translation, such as the separable bivariate normal pdf proposed in Section 8.1.3. Assuming that the translation is separable and has equal variance in x and y, and that the most probable position is the center of the search area, the positional *a priori* probabilities can be written as

$$P(R_k) = \frac{1}{2\pi\sigma_d^2} \exp\left[\frac{{i_k}^2 + {j_k}^2}{2\sigma_d^2}\right], \tag{8.24}$$

where $\mathbf{p}_k = (i_k, j_k)$.

Figure 8-17 compares the results of a test that uses this positional prior (statistic denoted t_{PP}) to one that does not do so (statistic denoted t). Instead of putting the position of correct register at the center of the search area, the Monte Carlo trials chose a random position in the search area according to the prior of (8.24). The result shows that there is a marked improvement in error-rate after incorporating this knowledge into t_{PP}. Although this degree of accuracy in the prior knowledge of the distortion is unrealistic, the result does indicate that there is potential to improve block matching performance by incorporating a positional prior into the registration test.

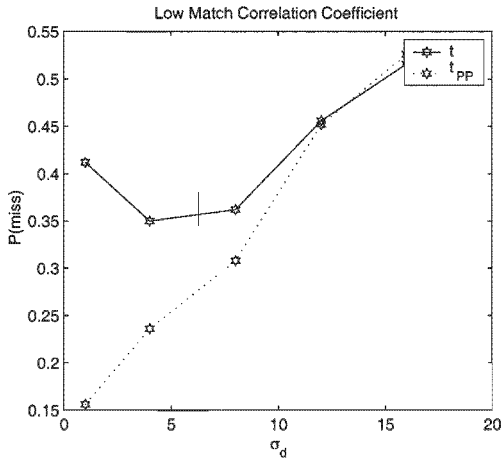
8.4.3 Real Images with Synthesized Noise

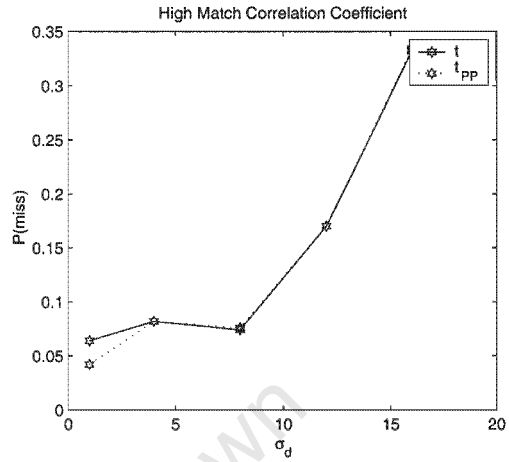
It is difficult to obtain conclusive results for real image data without targeting a specific application. It is also unrealistic to expect that experiments with a small number of images will provide information about matching performance for images in general. Rather, the intention here is to give some indication of what can be expected when applying the approach based on hypothesis tests to real images.

Experimental Procedure

An individual Monte Carlo trial in the block matching experiment consists of the following steps:

1. Generate two random noise fields with variance σ_η^2.
2. Add these to the original image to produce an image pair where the images differ only in the noise component.
3. Select a randomly positioned control point in the first image.

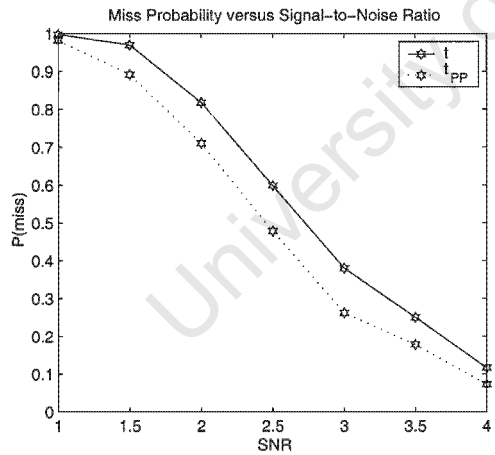


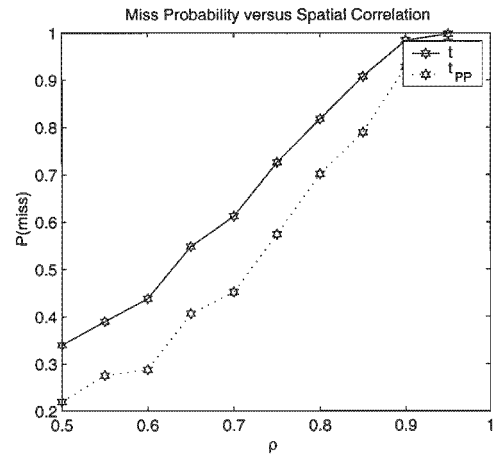


Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
Scene one-step correlation (ρ)	0.95
SNR	3

Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.99
Scene one-step correlation (ρ)	0.95
SNR	2

(a) Miss rate versus standard deviation of the translation prior.





Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
Scene one-step correlation (ρ)	0.95

Parameter	*Value*
Image size (n)	16
Mismatch correlation (ρ_0)	0
Match correlation (ρ_1)	0.6
SNR	1

(b) Miss rate versus image parameters.

Figure 8-17: Monte Carlo investigation of miss rate with a positional prior.

4. Perform block matching in a search area around the position of this point in the second image.

5. Establish whether the trial resulted in successful registration. Since the images were identical to start with, the correct registration offset is $(0, 0)$ and if the result is not this position or one of its four immediate neighbours, then it is counted as a miss.

The scene variance σ^2 and spatial one-step correlation coefficient ρ were estimated from the image data using standard ML estimators[2]. Experiments were performed for a range of noise variance (and therefore a range of SNR).

Images

The experiment was performed with two images: a photographic image of a crowd outdoors and an X-ray radiograph of a human head. In both cases the experiment was conducted for the original image and three other images:

1. **Gradient image** Several authors have found, both theoretically and experimentally, that a gradient preprocessor improves registration performance with the standard similarity statistics (e.g. Pratt [22], and Svedlow, McGillem and Anuta [35]). The gradient image used here is the combined magnitude result of preprocessing with both horizontal and vertical Sobel edge operators [17, p. 332].

2. **Stationary image** In order to better approximate an image with stationary first order statistics, local averages in 32×32 image windows were subtracted from the original image. Hunt and Cannon propose this simple procedure for enforcing stationarity [75] (see Appendix A for a discussion of this approach). A pointwise transform on this image can provide pixels that better resemble samples from a normally distributed random process (see Appendix A). This procedure, proposed by Chapple and Bertilone for simulating infra-red imagery [89], provides images that are empirically better samples of an MVN process.

3. **Markov equivalent image** since the Monte Carlo experiments in this and previous chapters were based on a synthetic Markov image, it is interesting to compare the results

[2]It should be mentioned that the issue of estimating the model parameters has been neglected in this research. Clearly, good estimates of the SNR and spatial correlation coefficient are desirable. This, however, is a topic all of its own that (in this author's opinion) has not received satisfactory attention in the literature and deserves further investigation.

for the stationary image to those obtained for a synthetic Markov image with the same image parameters.

Results

Figures 8-18 and 8-19 both provide block matching results for the original image, the stationary approximation and the synthetic equivalent. The optimal statistic reported the lowest error-rate in all of the experiments save one — the original 'skull' image. For both scenes, however, the best performance in the unaltered original image is significantly worse than that of the stationary image. Performance for all statistics is best in the synthetic equivalent image. In fact, the difference in error rate between the stationary image and the synthetic equivalent for both scenes is clearly an indication that the assumed model is inadequate. The ordering of different measures according to error-rate is the same, however, suggesting that qualitative aspects of the experimental results for synthetic images may still carry over to real images.

8.5 Discussion

The block matching sub-problem of image registration has been formulated as a hypothesis testing procedure and the test has been derived. The test includes a rejection hypothesis and can incorporate a prior for the expected displacements between images. The registration statistic incorporates the LRT statistic for image matching that was derived in Chapter 5.

Block matching is computationally expensive whether the optimal test or the standard similarity statistics are used. It is shown that the standard approach to speeding up correlation operations using FFTs can be extended and — together with a fast algorithm for calculating the pixel intensity sums in local image windows — used to develop fast algorithms for computing the match surface. Both the LRT and the standard similarity statistics can benefit from this approach.

An insight gained from the scalar matching test derived in Chapter 3 — that some scalars have no potential matching counterparts — was used to derive a screening test for identifying blocks that have no chance of finding a match in any search area. These blocks can be eliminated before registration begins, thereby reducing the required computation and reducing the number of registration misses. This *absolute* condition for selecting control points is only practical for very small images, and a more generally applicable strategy selects the best

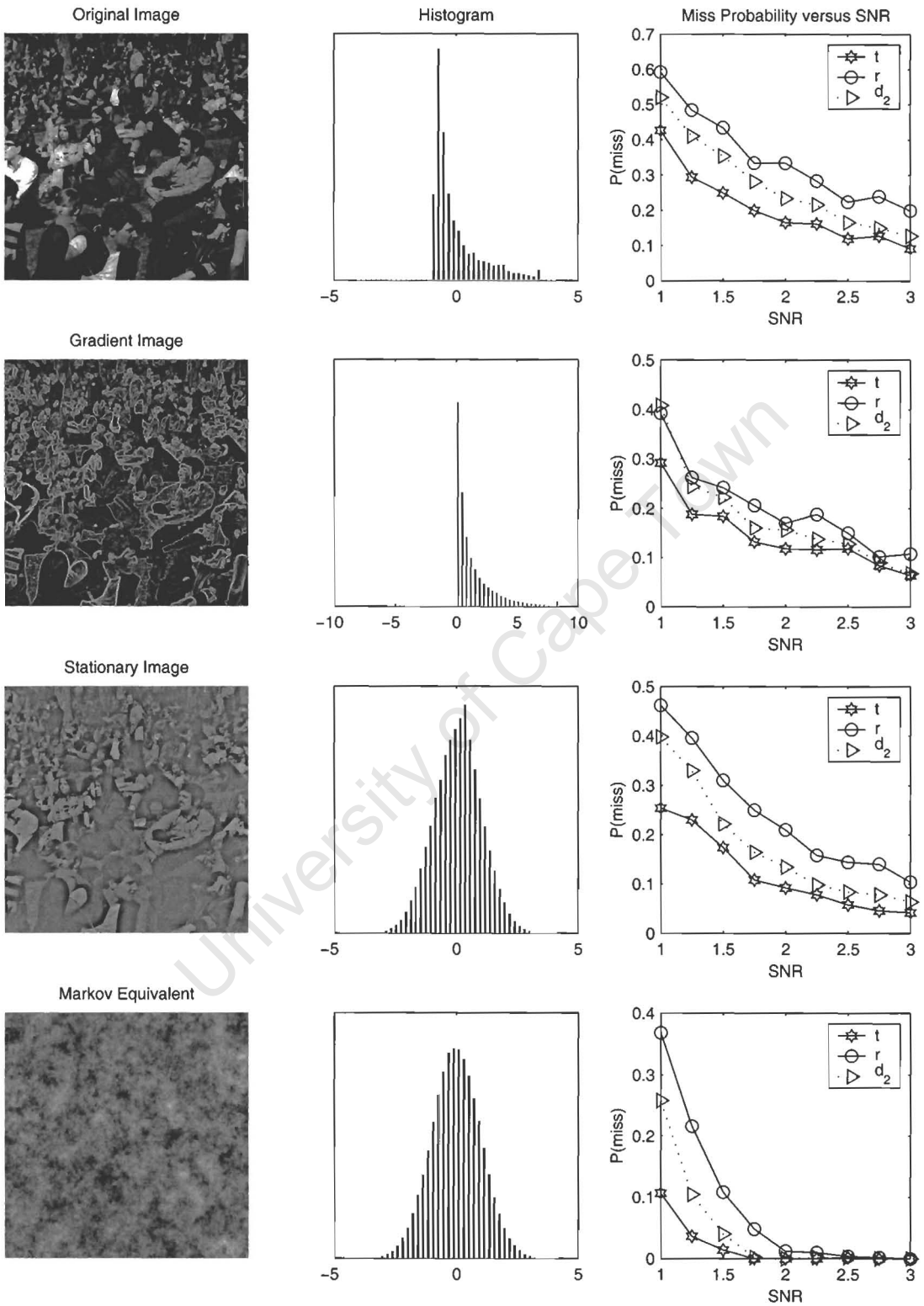


Figure 8-18: Monte Carlo matching results for real 'crowd' images corrupted by artificial noise ($n = 8$, $n_s = 32$, $T = 500$).

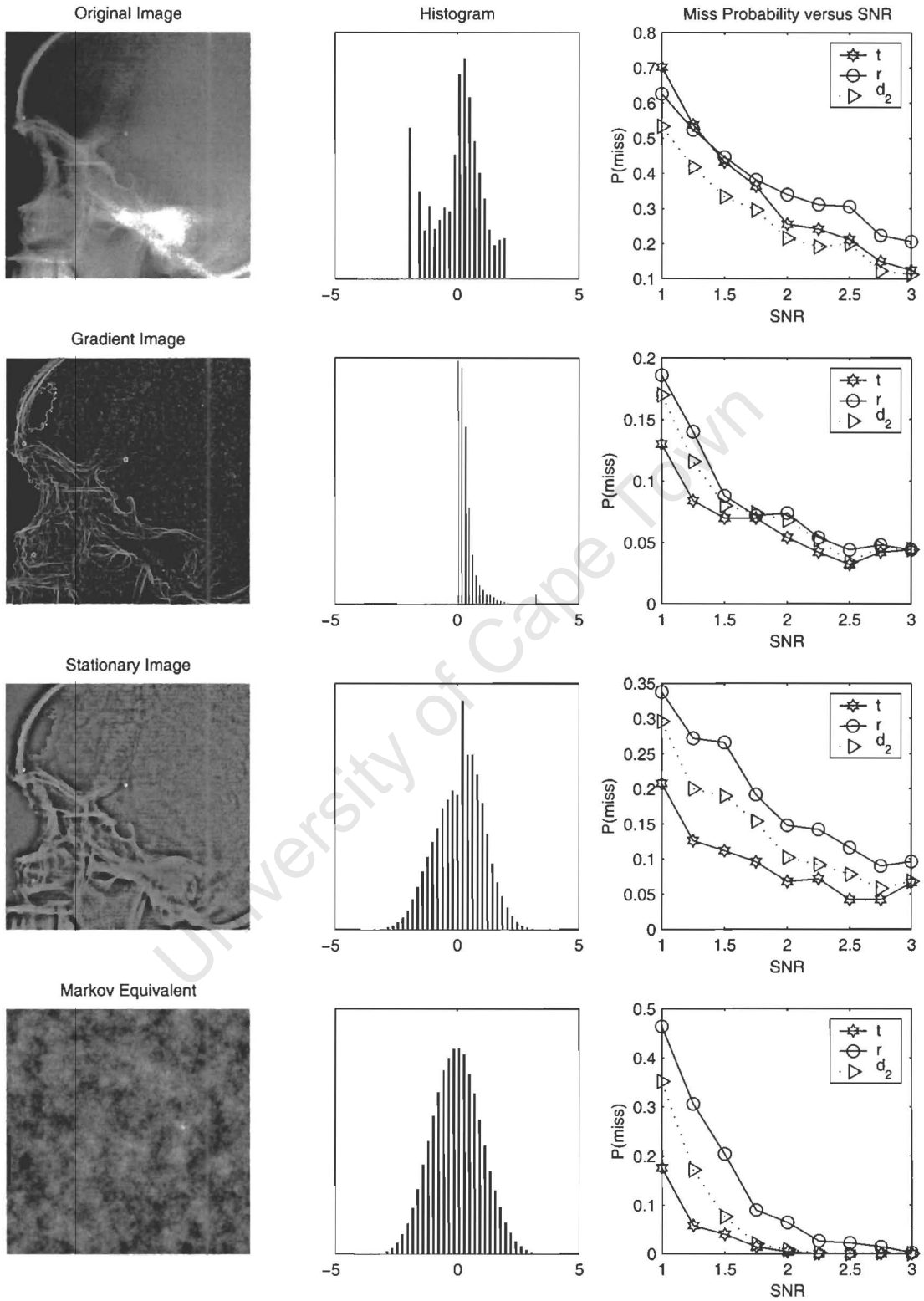


Figure 8-19: Monte Carlo matching results for real X-ray images corrupted by artificial noise ($n = 8$, $n_s = 32$, $T = 500$).

control points using *relative* comparisons of the screening statistic for the blocks that surround them.

Monte Carlo experiments show that the block matching search based on hypothesis testing has the same performance advantage that was found for the LRT and image matching in Chapter 6. This advantage is carried over to experiments with real images that are corrupted by synthetic noise. Working with real images does reveal, however, that the success of the tests will depend on the adequacy of models and the accuracy of model parameter estimates.

Chapter 9

Conclusion

Direct image matching is one of the oldest image processing problems, and one for which the literature proposes many and varied solutions. A single unifying formulation of the problem, however, is absent. The work presented in this dissertation proposes a strategy for formulating the problem — hypothesis testing based on a probabilistic joint image-pair model — and follows this strategy through for the simple multivariate normal case. The results are image matching and registration algorithms that are demonstrably superior to other solutions in terms of probability of error under a wide variety of conditions.

The research is now consolidated and future directions contemplated. Section 9.1 summarizes the contribution made. This work is only the beginnings of a rigorous approach to image matching and Section 9.2 suggests directions for further research that may be fruitful. Final remarks are made in Section 9.3.

9.1 Summary of the Contribution

The contribution made in the dissertation is summarized here in terms of the key insights and their consequences, the main theoretical results, the implementation strategies developed, and the results of experiments conducted.

9.1.1 Key Insights

The following key insights are the foundation of this research:

1. **Matching in a decision theoretic framework** The matching problem is one of classifying an observation of the real world into one of two classes, match or mismatch.

On this basis it is deduced that the concept of similarity between images is a diversion.

The fields of detection theory and pattern recognition provide a wealth of literature on the topic of making decisions based on observations of the real world. In order to formulate image matching as a decision problem, performance criteria must be specified for the solution and a representation must be found for the observation. For most problems the probability of error, or a more general cost function, will suffice as the performance criterion. The requirement for representing the observation leads to the next key insight.

2. **The image-pair as unit of observation** The unit of observation in the image matching problem is the image pair. Although there are examples of previous work where the rationale for a direct image matching solution includes an implicit reference to the image pair, in these examples the rationale came first. This "rationale" is, in fact, evidence of the designer introducing subjectivity into the form of the solution. By contrast, the research presented in previous chapters made no assumptions regarding the solution, but rather embodied the assumptions in a model of the observed image pair. The subjectivity in the design is thus restricted to the model and the performance criteria.

   As a consequence of this perspective, a joint image-pair model must be developed in order to proceed. If the *a priori* information is assumed to be probabilistic, then this model is simply the joint pdf of all pixels in both images.

9.1.2 Theoretical Results

The important theoretical results that have emerged from this research are now summarized.

1. **The image-pair joint covariance matrix** The simplifying multivariate normal assumption reduces the problem of specifying the image-pair pdf to one of specifying the joint image-pair mean vector and covariance matrix. Assuming that the image pair is multivariate normal, that the individual images share intra-image correlation structure to within a scaling factor, and that the images are corrupted by additive white noise, then a covariance matrix of the form

$$\mathbf{K_w} = \begin{bmatrix} \sigma_a^2\mathbf{R}+\sigma_\mu^2\mathbf{I} & \sigma_a\sigma_b\rho_{ab}\mathbf{R} \\ \sigma_a\sigma_b\rho_{ab}\mathbf{R} & \sigma_b^2\mathbf{R}+\sigma_\nu^2\mathbf{I} \end{bmatrix}.$$

is shown to describe the image pair $\mathbf{w}$. Both a correlation-based and a difference-based model of the image pair lead to this result.

2. **The optimal test for image matching** The optimal test for match in an image pair $\mathbf{w} = \left[\mathbf{u}^T, \mathbf{v}^T\right]$ consists of the likelihood ratio and a decision threshold, written as the likelihood ratio test (LRT)

$$l(\mathbf{w}) \underset{H_1}{\overset{H_0}{\gtrless}} \lambda.$$

Equivalently, the LRT can be written in terms of a modified statistic and decision threshold as

$$s(\mathbf{w}) \underset{H_1}{\overset{H_0}{\gtrless}} \acute{\lambda}.$$

A more convenient representation is obtained if the images are whitened beforehand. In this case the equivalent LRT for whitened images $\acute{\mathbf{u}}$ and $\acute{\mathbf{v}}$ is $s(\acute{\mathbf{u}}, \acute{\mathbf{v}}) \underset{H_1}{\overset{H_0}{\gtrless}} \acute{\lambda}$, where

$$s(\acute{\mathbf{u}}, \acute{\mathbf{v}}) = \sum_{i=1}^{n^2} \beta_i \left(\acute{u}_i - m_{\acute{u}_i}\right)\left(\acute{v}_i - m_{\acute{v}_i}\right) - \alpha_i \left(\left(\acute{u}_i - m_{\acute{u}_i}\right)^2 + \left(\acute{v}_i - m_{\acute{v}_i}\right)^2\right),$$

for

$$\alpha_i = \frac{k_i^2\left(\rho_1^2 - \rho_0^2\right)}{\left(1 - k_i^2\rho_0^2\right)\left(1 - k_i^2\rho_1^2\right)} \quad \text{and} \quad \beta_i = 2\frac{k_i\left(\rho_1 - \rho_0\right)\left(1 + k_i^2\rho_0\rho_1\right)}{\left(1 - k_i^2\rho_0^2\right)\left(1 - k_i^2\rho_1^2\right)},$$

and

$$\acute{\lambda} = \log\lambda^2 + \sum_{i=1}^{n^2} \log\left(\frac{1 - k_i^2\rho_1^2}{1 - k_i^2\rho_0^2}\right).$$

The quantity λ is specified according to whether the ideal observer or the Neyman-Pearson test is to be used.

3. **The optimal compaction of inter-image correlation structure in the image pair** In the same way that the top n principal components of $\mathbf{u}$ are an optimal compaction of the information in $\mathbf{u}$ (in a mean-square sense), the top n whitened pixel pairs $\{\acute{u}_i, \acute{v}_i\}$ (ordered by decreasing magnitude of their correlation coefficient) represent an optimal compaction of the correlation structure *between* $\mathbf{u}$ and $\mathbf{v}$. This is the case be-

cause $\{\acute{u}_i, \acute{v}_i\}$ are the canonical variables of $\{\mathbf{u}, \mathbf{v}\}$. Furthermore, $\acute{u}_i$ and $\acute{v}_i$ are the i-th (normalized) principal components of $\mathbf{u}$ and $\mathbf{v}$, respectively.

In summary, under the proposed model, the principal components of $\mathbf{u}$ and $\mathbf{v}$ capture the most important information regarding match and mismatch between them.

4. **The optimal test for block matching in image registration** Block matching is the process of translating a subimage block from one image over a search area in another, searching for the position that correctly registers the subimage to the second image. Although this is one of the oldest image processing operations, it is still an important part of many contemporary algorithms. Given multivariate normal image models, the optimal test for the position of correct register $\mathbf{p}$ is the one that maximizes the *a posteriori* probability of correct register over the $k \in \{1, 2, \ldots, N_p\}$ positions in the search area, and that exceeds a rejection threshold with this maximum. If the $n \times n$ block is represented by vector $\mathbf{v}$ and the k, $n \times n$ subimages of the search area are represented by $\mathbf{u}_k$, then $\mathbf{p} = \mathbf{p}_m$ is the position of correct register if

$$m = \arg\max_k \left[t\left(\mathbf{u}_k, \mathbf{v}\right)\right] \quad \text{and} \quad s\left(\mathbf{u}_m, \mathbf{v}\right) > \acute{\lambda}_R,$$

where $t\left(\cdot\right)$ is the registration statistic and $s\left(\cdot\right)$ is the LRT statistic for image matching.

The registration statistic and rejection threshold are given by

$$t\left(\mathbf{u}_k, \mathbf{v}\right) = s\left(\mathbf{u}_k, \mathbf{v}\right) + 2\log P\left(R_k\right)$$

and

$$\acute{\lambda}_R = \log\left(\frac{|\mathbf{K}_{\mathbf{w}_1}|}{|\mathbf{K}_{\mathbf{w}_0}|}\right) + 2\log\frac{\lambda_R}{1-\lambda_R} + 2\log\frac{1-P\left(R_k\right)}{P\left(R_k\right)},$$

respectively.

Note that the registration statistic is written in terms of the optimal LRT statistic $s\left(\mathbf{u}_k, \mathbf{v}\right)$. The test also incorporates the *a priori* probability of registration for each position in the search area $P\left(R_k\right)$, which can be specified using prior knowledge about the mechanism causing translational differences between the two images.

9.1.3 Implementation Strategies

Efficient implementation strategies have been developed for the proposed methods.

1. **Fast computation of the LRT statistic** The theoretically optimal test for matching involves the computationally expensive calculation of a statistic based on the image-pair covariance matrix. Three methods have been proposed for reducing the computation required. First, the optimal compaction of the correlation structure in the canonical variable pairs and the fact that the statistic is expressed in term of these variables suggests a strategy of using only a subset of the canonical variables to calculate the statistic. Second, a separable model can be used for images that are nonseparable with little compromise in error rate. Further economy is gained if the spatial correlation in the images is high, since in this case the whitening transform can be approximated using the DCT. Third, large images can be partitioned into non-overlapping blocks and the LRT statistic calculated for each block-pair. An approximation of the LRT statistic for the full image pair is then simply the sum of the block-pair LRT statistics.

2. **Fast filtering with the LRT statistic** For the purposes of translational block matching, it is necessary to calculate the similarity statistic between an image and overlapping positions in the larger search area of another image. This filtering of an image with a smaller image using a similarity statistic is a computationally intensive operation. It has been shown that the LRT filter can be performed more efficiently by viewing the basis expansion of the whitening operation at each position in the search area as a correlation operation, which can be performed efficiently over the entire search area for each basis image using FFTs. The component of the filter output associated with each basis image (and therefore associated with each pixel in the whitened smaller image) is calculated in this way and they are all summed together to provide the overall result.

3. **Fast filtering with standard similarity statistics** It is well known that filtering with the correlation function can be performed efficiently using FFT operations, and some authors have claimed that this is an advantage of the correlation function over other, more complicated measures such as the correlation coefficient. It has been shown in this dissertation, however, that the windowed calculation of other similarity statistics can also be decomposed into FFT-based operations. A new method for calculating windowed sums efficiently has been introduced and is used in conjunction with FFT-based correlation operations to speed up the calculation of match surfaces for the correlation

coefficient and the sum of squared differences.

4. **Selection of control points for block matching** A statistic for evaluating the suitability of blocks for image registration has been derived. The statistic can be used in an absolute sense to eliminate blocks that have no matching counterparts (i.e. they are not represented in the critical region for matching image pairs under the optimal test), but this is only practical for very small images. In a more generally applicable strategy, the same statistic can be used to make relative comparisons of registration suitability amongst image blocks. The control points can then be selected according to the suitability of the image blocks in their vicinity. This approach can be used to reduce the number of control points required (and therefore the computation), but can also improve overall registration performance by eliminating unsuitable blocks.

9.1.4 Experimental Results

The availability of a stochastic model for the image pair makes it possible to do extensive experimentation using Monte Carlo simulation methods. The important results are summarized here.

1. **Monte Carlo matching experiments** The error-rate superiority of the optimal test over the standard similarity statistics under the assumed model suggests that there is significant scope for improving on current methods. With respect to the standard methods, a general rule emerges: difference-based statistics (e.g. sum of squared or absolute differences) are superior when the scenes are identical under the match hypothesis, whereas the correlation-based methods (e.g. correlation coefficient) are superior when the scenes are not identical under the match hypothesis.

   The LRT shows a reasonable degree of insensitivity to model parameter inaccuracies and deviations from the assumed noise model, but is very sensitive to occlusion compared to the standard approaches. Here the advantage of a nonparametric measure, such as the stochastic sign change criterion, becomes evident.

2. **Monte Carlo registration experiments** Monte Carlo experiments show that the block matching search based on hypothesis testing has the same performance advantage that was found for the LRT and image matching. The facility to incorporate a prior for the position of correct register and a rejection hypothesis are shown to enhance registration performance. Experiments with real images and synthetic noise also

demonstrate the superiority of the approach based on hypothesis tests, but working with real images does reveal that the success of the tests will depend on the adequacy of models and the accuracy of model parameter estimates.

9.2 Directions for Further Research

There are several avenues of further research that could follow from the work summarized in the previous section.

9.2.1 Image-Pair Models

The proposed approach to the image matching problem has its subjectivity in the selection of the image-pair model. In comparison to this modelling stage, the derivation of the actual test, although challenging, is a mechanistic mathematical process. A relatively simple model was chosen for this research, but there is almost unlimited potential for using more sophisticated models. Some options in this regard are generalized Gaussian and higher order models that relax the assumption of normality, and multi-resolution models such as scale-space and wavelets. These models may reflect the characteristics of real images more accurately and if this is the case, solutions based on them will be closer to optimal for the actual images.

As important as the ability to develop a representative model for the image pair is the ability to accurately estimate the free parameters in such a model from real image data. Even with the relatively simple multivariate normal image model used in this dissertation, this issue emerges when one deals with real images. Further research, therefore, is required on the topic of model parameter estimation in image pairs.

9.2.2 Robust Tests for Matching

Robust statistical methods are designed to operate in the vicinity of a parametric model that describes most of the data. Outliers with respect to this model are tolerated and do not compromise performance to the extent that they would for a purely parametric method. With the borderline applicability of the multivariate normal model to image data, it seems that a robust test, designed to be effective in the vicinity of this model, might be an improvement over the parametric test proposed in this dissertation. Care would have to be taken, however, that the data considered to be outliers in this formulation do not contain important information for discriminating between match and mismatch. As Hampel, Ronchetti, Rousseeuw and Stahel

put it: "Not all outliers are 'bad' data caused by gross errors; sometimes they are the most valuable datapoints in the whole set" [30, p. 12].

9.2.3 Image Matching Applications

The effectiveness of the proposed approach must be confirmed in its application to a real-world problem. One possibility is face recognition, where one of the most effective algorithms is already based on a linear model and principal component representation of the images [73]. It may be that the optimal weighting of principal components and the facility to specify a degree-of-match parameter, both of which are provided by the proposed test, will improve the recognition rate.

Another potential application area is medical imaging. Digital subtraction angiography (DSA) [108], for example, is improved if the images taken before and after the administration of a contrast medium are registered for accurate subtraction [57]. The dangers associated with radiation dose have increasingly become a source of concern among medical practitioners [109] and new international standards for radiation safety have emerged during the last decade [110]. Minimizing dose, however, compromises the image quality in terms of contrast and SNR, emphasizing the importance of *optimal* matching algorithms. It should also be noted that recent developments in low-dose digital radiography systems have revived interest in the 2D modality — a new low-dose full-body X-ray scanner, for example, is undergoing medical trials at the Groote Schuur Hospital in Cape Town, South Africa [111].

9.2.4 Combined Detection and Matching

There are many applications where a reference image is used to aid the detection of a target in a more recent image of the same scene. In lung cancer screening, radiologists use previous images of the same lung as a point of reference. In subtraction angiography, a reference image is subtracted in order to see more clearly the contrast media that is present. In both cases there is (1) the problem of matching the two images, and (2) the problem of detecting a certain target in one of them. The image-pair model proposed in this dissertation offers a way of formulating the problem that combines these two sub-problems.

Considering the multivariate normal image pair $\mathbf{w}^T = \begin{bmatrix} \mathbf{u}^T & \mathbf{v}^T \end{bmatrix}$, $\mathbf{v}$ can be viewed as the reference and $\mathbf{u}$ can be viewed as the current image that might contain the known deterministic target $\mathbf{p}$. The presence of target $\mathbf{p}$ in image $\mathbf{u}$ can be modelled as a component in the mean vector, and in this case $\mathbf{m_u} = \mathbf{m_a} + \mathbf{p}$. If the target is not present, then $\mathbf{m_u} = \mathbf{m_a}$ as

before. The combined matching and detection problem has the hypotheses shown in Table 9.1. The optimal test for this scenario selects the hypothesis with the maximum *a posteriori*

	Images mismatch ($\rho_{ab} = \rho_0$)	Images match ($\rho_{ab} = \rho_1$)
Target not present ($\mathbf{m_u} = \mathbf{m_a}$)	$H_0 \iff \begin{cases} \mathbf{m_u} = \mathbf{m_a} \\ \rho_{ab} = \rho_0 \end{cases}$	$H_1 \iff \begin{cases} \mathbf{m_u} = \mathbf{m_a} \\ \rho_{ab} = \rho_1 \end{cases}$
Target present ($\mathbf{m_u} = \mathbf{m_a} + \mathbf{p}$)	$H_3 \iff \begin{cases} \mathbf{m_u} = \mathbf{m_a} + \mathbf{p} \\ \rho_{ab} = \rho_0 \end{cases}$	$H_2 \iff \begin{cases} \mathbf{m_u} = \mathbf{m_a} + \mathbf{p} \\ \rho_{ab} = \rho_1 \end{cases}$

Table 9.1: Hypotheses for a combined matching and detection problem.

probability.

The framework of Table 9.1 encapsulates common image matching and detection problems. Image matching, the topic of this dissertation, is concerned with hypotheses H_0 (mismatch) and H_1 (match). A detection scheme using correlated reference images that is proposed by Margalit, Reed and Gagliardi [112] is concerned with hypotheses H_1 (matching reference images with no target present) and H_2 (matching reference images with target present). A DSA registration algorithm that must cope with the presence of contrast media might use hypotheses H_0 (mismatch with no contrast media present), H_1 (match with no contrast media present), and [H_2 OR H_3] (contrast media present).

9.3 Final Remarks

The problem of direct image matching has been approached in a new way, viewing the observed image pair as a unit, developing a statistical model for this unit, and deriving the optimal test for making the match versus mismatch decision. This test has been evaluated using Monte Carlo simulation techniques and is seen to compare favourably with the similarity measures that are commonly used. Methods have been developed to allow efficient computation of the test statistic. The hypothesis testing approach has also been applied to the block matching sub-problem of image registration and, here too, the results are compelling. Several potential avenues for further research have been identified.

The success of the proposed matching technique in practice will depend on the adequacy

of the multivariate normal image model in any particular application. Where this model is inadequate, alternative models can be used within the same general decision theoretic framework to derive an alternative test of the match and mismatch hypotheses. In either case, the solution will be based on explicit assumptions and will be optimal with respect to the assumed model. It is hoped that this emphasis on the modelling aspect of the problem will facilitate future advances in image matching research.

Appendix A

Simplified Random Field Models for Images

Much statistical analysis of images relies on the assumption of ergodicity. Without it, a representative ensemble of images is required in order to estimate the parameters of a statistical image model, and obtaining this ensemble is impractical in many applications. Even if the ensemble were available, dropping the ergodicity assumption also sacrifices the assumption of spatial stationarity. Stationarity simplifies the mathematics and leads to efficient, spatially invariant algorithms.

Another common assumption is that the image can be modelled as a multivariate normal (MVN) random field. The normal pdf has very convenient mathematical properties — analysis using other distributions generally leads to intractable problems [29, p. 1].

Although these two assumptions are often violated by real-world images, some authors have shown that relatively simple methods can be used to overcome this problem by designing a transform $T(\mathbf{u})$, such that the statistics in $\acute{\mathbf{u}} = T(\mathbf{u})$ are approximately stationary and MVN. Optimal algorithms can then be designed for $\acute{\mathbf{u}}$. Some of these methods are now discussed in more detail.

A.1 Nonstationary Mean

The MVN pdf for the $n \times n$ image $\mathbf{u}$ is

$$p_{\mathbf{u}}(\mathbf{u}) = \frac{1}{\sqrt{(2\pi)^{n^2} |\mathbf{K}_{\mathbf{u}}|}} \exp\left[-\frac{1}{2}(\mathbf{u}-\mathbf{m}_{\mathbf{u}})^T \mathbf{K}_{\mathbf{u}}^{-1}(\mathbf{u}-\mathbf{m}_{\mathbf{u}})\right], \tag{A.1}$$

which is characterized completely by the mean vector, $\mathbf{m_u}$, and covariance matrix, $\mathbf{K_u}$. Hence the popular abbreviated notation: $p_{\mathbf{u}}(\mathbf{u}) = N(\mathbf{u}; \mathbf{m_u}, \mathbf{K_u})$[1]. An ergodic model forces the mean, $\mathbf{m_u}$, to be a constant vector. Hunt and Cannon point out that the image ensemble of a particular class of images will not have a stationary mean in most applications and propose a model that has a nonstationary mean component and a stationary MVN component fluctuating around the mean [75]. Typically, an ensemble of images is not available, and so the ensemble mean is approximated by using a spatial "blurring" filter on the single available image. Hunt and Cannon use a Gaussian point-spread function, but Margalit, Reed and Gagliardi have subsequently used simple windowed averaging [112]. The mean in (A.1), $\mathbf{m_u}$, is estimated using

$$\bar{\mathbf{m}}_{\mathbf{u}} = \mathbf{u} * \mathbf{h},$$

where $*$ denotes convolution and $\mathbf{h}$ is the blurring filter kernel. The MVN component is then given by $\acute{\mathbf{u}} = \mathbf{u} - \bar{\mathbf{m}}_{\mathbf{u}}$. Experiments reveal that the histogram after removing the ensemble mean estimate is indeed more symmetrical than the histogram of the original image and it was this observation that was the initial motivation for the MVN model of the stationary component [75]. Figure A-1 shows histograms of the stationary component of a lung radiograph where the mean image was calculated using a range of kernel sizes. A question remains: how does one select the best kernel size? Margalit, Reed and Gagliardi use the size that minimizes the magnitude of the skewness of the pixel intensities in the modified image, since this should be zero for data under a normal distribution [112].

At this point a concern must be raised with this approach. It is not satisfactory that a technique for enforcing *stationarity* by estimating the ensemble mean is optimized using the *skewness* of the resulting distribution. With the possibility of stationarity being sacrificed in favour of a low third moment, the validity of the ensemble mean estimate comes into question. Even if the kernel size *is* optimal, it is questionable whether there is enough data in the image for an adequate estimate of the ensemble mean, particularly in regions where the ensemble mean is changing quickly. If the "blurred" image is *not* an adequate estimate of the ensemble mean, then valuable information will be lost when it is subtracted from the original image to produce the stationary component. To reinforce this last point, consider the MVN pdf of

[1]Strictly speaking, the notation $N(\mathbf{m}, \mathbf{K})$ is commonly used to represent the multivariate Normal *distribution* with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{K}$, but the presentation here will also use the notation $N(\mathbf{x}; \mathbf{m}, \mathbf{K})$ to represent the associated Normal *probability density function*, where $\mathbf{x}$ is the independent variable.

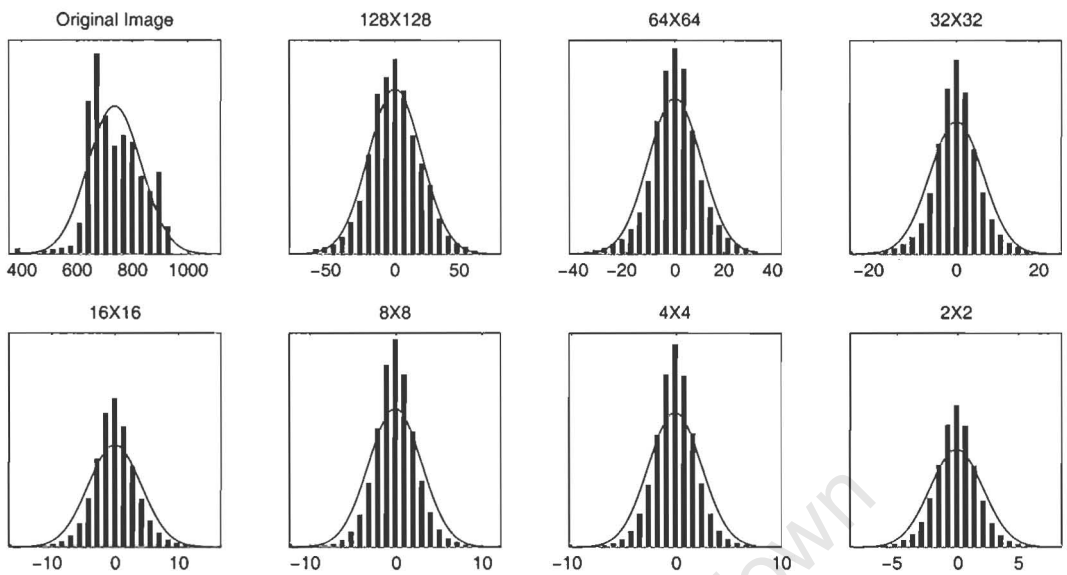


Figure A-1: Histograms for the stationary component of a lung radiograph for a range of kernel sizes.

(A.1). The procedure uses estimate $\bar{\mathbf{m}}_{\mathbf{u}}$ to approximate the ensemble mean vector $\mathbf{m}_{\mathbf{u}}$, which is a constant feature of the ensemble and therefore has no image specific information. The image-specific information is now assumed to be in the stationary component $\acute{\mathbf{u}}$, which has the pdf

$$p_{\acute{\mathbf{u}}}(\acute{\mathbf{u}}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{K}_{\mathbf{u}}|}} \exp\left[-\frac{1}{2}(\acute{\mathbf{u}})^T \mathbf{K}_{\mathbf{u}}^{-1}(\acute{\mathbf{u}})\right]. \tag{A.2}$$

Detection or matching algorithms now only need consider $\acute{\mathbf{u}}$ and its pdf. However, if the estimate $\bar{\mathbf{m}}_{\mathbf{u}}$ is not an ensemble mean, then valuable image specific information will be lost when it is subtracted from $\mathbf{u}$. The algorithms based on $\acute{\mathbf{u}}$ will be suboptimal.

Figure A-2 uses the 'LAX' image as a pathological example of this problem. Figure A-2(a) tabulates the magnitude of the skewness for a range of kernel sizes and suggests that a kernel size of 2×2 is optimal. Figure A-2(b) confirms that the 2×2 kernel provides a pdf that is more symmetrical than the original histogram, but the nonstationary component contains little of the information in the original image. Figure A-2(c) shows the results for a 16×16 kernel. The pdf is clearly skewed, but the image captures more scene information.

This procedure, therefore, should be viewed purely as an approximate decomposition into

N	**2**	4	8	16	32	64
Skewness	**0.1898**	0.8286	1.1801	1.3026	1.3225	1.3469

(a) Skewness of the stationary component.

(b) Ensemble mean estimate, stationary component and histogram - 2 × 2 kernel.

(c) Ensemble mean estimate, stationary component and histogram - 16 × 16 kernel.

Figure A-2: Skewness as a basis for kernel size selection.

stationary and nonstationary image components. The utility in the decomposition is that white noise is almost completely contained in the stationary component. This could be the basis of an approach that uses statistical models to good effect on the stationary component and noise-sensitive deterministic approaches on the nonstationary component. If it turns out that the stationary model has a non-normal pdf, then this can be dealt with as a separate issue (using, for example, the methods outlined in Section A.3).

A.2 Nonstationary Variance

Hunt has also investigated enforced stationarity of the covariance matrix, by normalizing the standard deviation in local windows and using a spatial warp in local areas to standardize the correlation structure [88]. Margalit, Reed and Gagliardi model the image as a random field that is piecewise stationary — the image is divided into smaller subimages and the maximum likelihood estimate of the covariance matrix is calculated for each subimage [112].

A.3 Non-MVN Distributions

Although MVN models are convenient, they are rarely accurate models of the information in an image scene, as the examples in Figure 4-1 illustrate. However, having extracted a stationary component using the method described in Section A.1, it is possible to transform the image in such a way that the result approximates a realization of a stationary MVN random field. Chapple and Bertilone propose the use of such a transform for simulating stationary, non-normal infrared imagery [89]. Given a positivity constraint, it can be shown that transforming the samples x of the non-normal random field by

$$F(x) = \sqrt{2}\,\mathrm{erf}^{-1}\left[2\int_0^x p_x(\acute{x})d\acute{x} - 1\right] \tag{A.3}$$

yields samples of a new random field with normal marginal pdfs. Given sample image data (a single image or an ensemble of images), Chapple and Bertilone transform the image intensities of the original images using (A.3) and treat the resulting images as MVN random fields. They then generate synthetic images using MVN parameters estimated from the transformed real images. Since F is an invertible transform, realistic test images can now be generated by applying the inverse transform, F^{-1}, to the synthetic MVN images. This method is simpler and more computationally efficient than direct methods of simulating non-normal random

fields [89] and has also been used to simulate synthetic aperture radar images [113].

The same approach can simplify the derivation of detection and matching algorithms based on non-normal models. The image pixels can be preprocessed using the transform in (A.3) and then treated as MVN samples. In practice $p_x(x)$ is approximated by the histogram of the original image, with points between histogram bin centers approximated using interpolation. This method was used on the stationary component of the 'LAX' image that is shown in Figure A-2(c). First the pixel values were given a constant offset in order to satisfy the positivity constraint. A histogram with 100 bins was used to approximate $p_x(x)$. Figure A-3 shows the skewed histogram of the original image, the pointwise transform and a histogram of the result after using this transform on the pixel values.

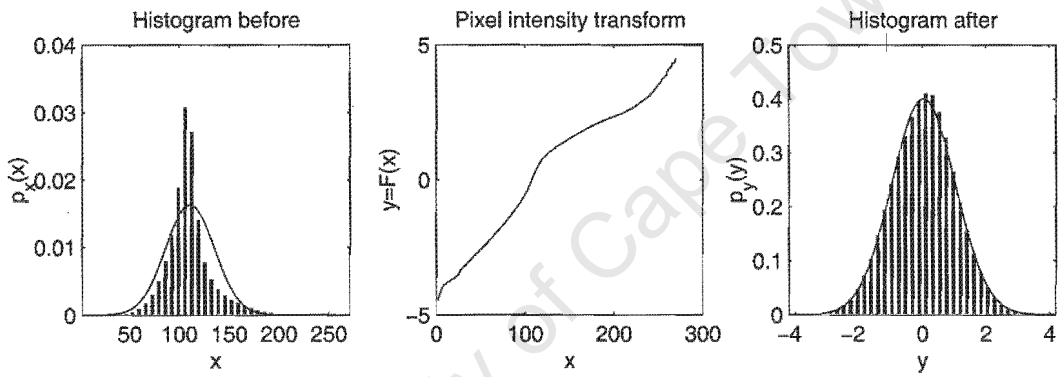


Figure A-3: Pointwise MVN transform.

Standard tests are available for establishing whether the resulting marginal pdf is indeed normal [28, p. 254]. It should be noted that although the transformed images are *treated* as MVN random fields, only the marginal pdfs of the individual pixels are actually normal. Strictly speaking, this is not a sufficient condition for the joint pdf to be MVN [29, p. 7], but in practice this distinction is often ignored.

This fairly ad-hoc procedure for transforming images so that they better resemble realizations of a MVN random field is not intended to be rigorous. There are no guarantees that it will be successful with images in general, but the empirical evidence suggests that the transformed images are far better approximations of stationary MVN random field realizations than the original images were. The use of this procedure in any specific application, however, must be preceded by experiments that confirm its effectiveness with a representative sample of test images.

Appendix B

Mathematical Derivations

In order to maintain the flow of the main text in the dissertation, several mathematical derivations have been placed in this appendix.

B.1 Covariance Matrix of the Sum or Difference of Two Random Vectors

Consider two random column vectors $\mathbf{x}$ and $\mathbf{y}$, which have the covariance matrices $\mathbf{K_x}$ and $\mathbf{K_y}$, respectively. The covariance matrix of their sum $\mathbf{z} = \mathbf{x} + \mathbf{y}$, is given by

$$\mathbf{K_z} = \mathbf{K_x} + \mathbf{K_y} + 2\mathbf{K_{xy}},$$

where $\mathbf{K_{xy}}$ is the cross-covariance matrix between $\mathbf{x}$ and $\mathbf{y}$.

Proof. Since the sum of the mean vectors of $\mathbf{x}$ and $\mathbf{y}$ is simply the sum of the individual elements

$$\mathbf{m_z} = \mathbf{m_x} + \mathbf{m_y}$$

is the mean vector of $\mathbf{z}$ [61, p. 178]. The covariance matrix is defined as

$$\mathbf{K_z} = E\left[(\mathbf{z} - \mathbf{m_z})(\mathbf{z} - \mathbf{m_z})^T\right].$$

Substituting $\mathbf{z} = \mathbf{x} + \mathbf{y}$ and manipulating

$$\begin{aligned}
\mathbf{K}_\mathbf{z} &= E\left[(\mathbf{z} - \mathbf{m}_\mathbf{z})(\mathbf{z} - \mathbf{m}_\mathbf{z})^T\right] \\
&= E\left[((\mathbf{x} + \mathbf{y}) - (\mathbf{m}_\mathbf{x} + \mathbf{m}_\mathbf{y}))((\mathbf{x} + \mathbf{y}) - (\mathbf{m}_\mathbf{x} + \mathbf{m}_\mathbf{y}))^T\right] \\
&= E\left[((\mathbf{x} - \mathbf{m}_\mathbf{x}) + (\mathbf{y} - \mathbf{m}_\mathbf{y}))((\mathbf{x} - \mathbf{m}_\mathbf{x}) + (\mathbf{y} - \mathbf{m}_\mathbf{y}))^T\right] \\
&= E\left[(\mathbf{x} - \mathbf{m}_\mathbf{x})(\mathbf{x} - \mathbf{m}_\mathbf{x})^T\right] + E\left[(\mathbf{y} - \mathbf{m}_\mathbf{y})(\mathbf{y} - \mathbf{m}_\mathbf{y})^T\right] + 2E\left[(\mathbf{x} - \mathbf{m}_\mathbf{x})(\mathbf{y} - \mathbf{m}_\mathbf{y})^T\right] \\
&= \mathbf{K}_\mathbf{x} + \mathbf{K}_\mathbf{y} + 2\mathbf{K}_{\mathbf{xy}}
\end{aligned}$$

as required. ■

It can be shown in the same way that for the difference, $\mathbf{z} = \mathbf{x} - \mathbf{y}$,

$$\mathbf{K}_\mathbf{z} = \mathbf{K}_\mathbf{x} + \mathbf{K}_\mathbf{y} - 2\mathbf{K}_{\mathbf{xy}}.$$

Note that if $\mathbf{x}$ and $\mathbf{y}$ are uncorrelated, then

$$\mathbf{K}_{\mathbf{x}\pm\mathbf{y}} = \mathbf{K}_\mathbf{x} + \mathbf{K}_\mathbf{y}.$$

B.2 Optimal Threshold for the Scalar Squared-Difference Test

Consider the random scalars $u = a + \mu$ and $v = b + \nu$, where a and b are $N\left(m, \sigma^2\right)$, and where noise scalars μ and ν are $N\left(0, \sigma_\mu^2\right)$ and $N\left(0, \sigma_\nu^2\right)$ respectively. The optimal decision threshold for deciding between match and mismatch of a and b using the scalar squared difference statistic,

$$s(u, v) = -(u - v)^2,$$

is derived here. The pdf of the difference

$$\begin{aligned}
d(u, v) &= u - v \\
&= a - b + \mu - \nu
\end{aligned}$$

is given by

$$p_d(d) = N\left(d; 0, 2\sigma^2(1-\rho) + \sigma_\mu^2 + \sigma_\nu^2\right).$$

Given a normally distributed random variable x, where

$$p_x(x) = N\left(x; 0, \sigma_x^2\right) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left[-\frac{x^2}{2\sigma_x^2}\right],$$

the pdf of the random variable $y = x^2$, is given by [114, p. 108]

$$p_y(y) = \frac{1}{\sqrt{2\pi\sigma_x^2 y}} \exp\left[-\frac{y}{2\sigma_x^2}\right] \ \forall\, y > 0.$$

The pdf of the squared difference, conditioned on ρ, is therefore given by

$$p_s(s) = \frac{1}{\sqrt{-2\pi s\left(2\sigma^2(1-\rho) + \sigma_\mu^2 + \sigma_\nu^2\right)}} \exp\left[\frac{s}{2\left(2\sigma^2(1-\rho) + \sigma_\mu^2 + \sigma_\nu^2\right)}\right] \ \forall\, s < 0.$$

For the match hypothesis $H_1 \iff \rho = 1$ and the mismatch hypothesis $H_0 \iff \rho = 0$, the ideal observer test is

$$\frac{p_s(s|\rho=1)}{p_s(s|\rho=0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P_0}{P_1}$$

and by re-arranging this expression, the condition for a match is seen to be

$$s(u, v) > \lambda,$$

where the RHS of this inequality is the optimal decision threshold for the squared difference test, given by

$$\lambda = 2\frac{\left(\sigma_\mu^2 + \sigma_\nu^2\right)\left(2\sigma^2 + \sigma_\mu^2 + \sigma_\nu^2\right)}{2\sigma^2} \log\left[\sqrt{\frac{\sigma_\mu^2 + \sigma_\nu^2}{2\sigma^2 + \sigma_\mu^2 + \sigma_\nu^2}}\frac{P_1}{P_0}\right].$$

If $\sigma_\mu^2 = \sigma_\nu^2$, then the threshold becomes

$$\lambda = \frac{4\sigma_\nu^2\left(\sigma^2 + \sigma_\nu^2\right)}{\sigma^2} \log\left[\sqrt{\frac{\sigma_\nu^2}{\sigma^2 + \sigma_\nu^2}}\frac{P_1}{P_0}\right].$$

B.3 The Whitening Transform

The random image $\mathbf{x}$ with covariance matrix $\mathbf{K}$ can transformed to a random image with independent, identically distributed (iid) pixels using the transformation

$$\mathbf{Tx} = \Lambda_K^{-\frac{1}{2}} \mathbf{V}^T \mathbf{x}, \tag{B.1}$$

where $\mathbf{V}$ is a matrix with columns that are the eigenvectors of $\mathbf{K}$ and Λ_K is a matrix with the corresponding eigenvalues on the diagonal.

Proof. Since $\mathbf{K}$ is a real symmetric matrix, it can be diagonalised by the similarity transformation

$$\mathbf{V}^T \mathbf{K} \mathbf{V} = \Lambda_K, \tag{B.2}$$

where $\mathbf{V}$ is a matrix with columns that are the eigenvectors of $\mathbf{K}$ and Λ_K is a matrix with the corresponding eigenvalues on the diagonal [115, p. 269]. Since $\mathbf{V}$ is an orthogonal matrix, it follows that $\mathbf{K} = \mathbf{V}\ \Lambda_K \mathbf{V}^T$. Substituting this into the pdf for $\mathbf{x}$ it is seen that $\acute{\mathbf{x}} = \Lambda_K^{-\frac{1}{2}} \mathbf{V}^T \mathbf{x}$ is a unit-variance, random vector with iid elements, as required. ■

Equation B.1 is referred to as a *whitening transform* on $\mathbf{x}$, and $\acute{\mathbf{x}}$ itself is referred to as the *whitened* image.

B.4 Eigenvalue Relationships in the Image Covariance Matrix

If random image $\mathbf{x}$ has the covariance matrix

$$\mathbf{K} = \sigma^2 \mathbf{R} + \sigma_\eta^2 \mathbf{I}, \tag{B.3}$$

then the eigenvalues of $\mathbf{K}$ are related to those of $\mathbf{R}$ by

$$\Lambda_K = \sigma^2 \Lambda_R + \sigma_\eta^2 \mathbf{I},$$

where the elements in the diagonal matrices Λ_K and Λ_R are the eigenvalues of $\mathbf{K}$ and $\mathbf{R}$ respectively.

Proof. If the $n \times n$ matrix $\mathbf{A}$ has eigenvalues $\lambda_1, \ldots, \lambda_p$, then [29, p. 584]

1. The matrix $k\mathbf{A}$ has eigenvalues $k\lambda_1, \ldots, k\lambda_p$ and

2. the matrix $\mathbf{A} - k\mathbf{I}$ has eigenvalues $\lambda_1 - k, \ldots, \lambda_p - k$.

Combining these properties, the matrix $a\mathbf{A} + b\mathbf{I}$ has eigenvalues $a\lambda_1 + b, \ldots, a\lambda_p + b$. Applying this new property to (B.3), the eigenvalues of $\mathbf{K}$, denoted λ_i, are seen to be

$$\lambda_i = \sigma^2\omega_i + \sigma_\eta^2,$$

where ω_i are the eigenvalues of $\mathbf{R}$. Writing the eigenvalues of $\mathbf{K}$ and $\mathbf{R}$ as the diagonal elements of diagonal matrices Λ_K and Λ_R respectively, the required result is obtained. ■

B.5 Shared Covariance Matrix Eigenvectors

The covariance matrices $\mathbf{K_x} = \sigma_a^2\mathbf{R}+\sigma_\mu^2\mathbf{I}$ and $\mathbf{K_y} = \sigma_b^2\mathbf{R}+\sigma_\nu^2\mathbf{I}$ share the eigenvectors of $\mathbf{R}$.

Proof. The characteristic polynomial of $\mathbf{K_x}$ is

$$p_{\mathbf{x}}(\lambda) = \det(\mathbf{K_x} - \lambda\mathbf{I}).$$

The i-th eigenvector of $\mathbf{K_x}$ is the solution of $p_x(\lambda_i^{\mathbf{x}}) = 0$, where $\lambda_i^{\mathbf{x}}$ denotes the i-th eigenvalue of $\mathbf{K_x}$. Using the result in Section B.4,

$$\lambda_i^{\mathbf{x}} = \sigma_a^2\omega_i + \sigma_\mu^2, \tag{B.4}$$

where ω_i is the i-th eigenvalue of $\mathbf{R}$. Substituting (B.4) into the characteristic polynomial and manipulating,

$$\begin{aligned} p_{\mathbf{x}}(\lambda_i^{\mathbf{x}}) &= \det\left(\mathbf{K_x} - \left(\sigma_a^2\omega_i + \sigma_\mu^2\right)\mathbf{I}\right) \\ &= \det\left(\sigma_a^2\mathbf{R}+\sigma_\mu^2\mathbf{I} - \left(\sigma_a^2\omega_i + \sigma_\mu^2\right)\mathbf{I}\right) \\ &= \det\left(\sigma_a^2\mathbf{R} - \sigma_a^2\omega_i\mathbf{I}\right) \\ &= \sigma_a^2\det(\mathbf{R} - \omega_i\mathbf{I}). \end{aligned}$$

So a vector that solves $\det(\mathbf{R} - \omega_i\mathbf{I}) = 0$ also solves $p_{\mathbf{x}}(\lambda_i^{\mathbf{x}}) = 0$, and therefore the i-th eigenvector of $\mathbf{K_x}$ is the i-th eigenvector of $\mathbf{R}$. Similarly, it can be shown that the i-th eigenvector of $\mathbf{K_y}$ is the i-th eigenvector of $\mathbf{R}$. It follows that $\mathbf{K_x}$ and $\mathbf{K_y}$ share the eigenvectors of $\mathbf{R}$, as required. ■

B.6 Block Diagonalizing the Image-Pair Covariance Matrix

Denote the pair of images $\mathbf{x}$ and $\mathbf{y}$ as $\mathbf{w}^T = [\mathbf{x}^T, \mathbf{y}^T]$. Let the joint covariance matrix be

$$\begin{aligned}\mathbf{K_w} &= \begin{bmatrix} \mathbf{K_x} & \mathbf{K_{xy}} \\ \mathbf{K_{xy}} & \mathbf{K_y} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_a^2\mathbf{R}+\sigma_\mu^2\mathbf{I} & \sigma_a\sigma_b\rho_{ab}\mathbf{R} \\ \sigma_a\sigma_b\rho_{ab}\mathbf{R} & \sigma_b^2\mathbf{R}+\sigma_\nu^2\mathbf{I} \end{bmatrix}.\end{aligned}$$

If independent whitening transforms[1] are applied to the images $\mathbf{x}$ and $\mathbf{y}$ then the joint covariance matrix will have diagonal partitions according to

$$\mathbf{K_{\hat{w}}} = \begin{bmatrix} \mathbf{I} & \mathbf{D_{\hat{x}\hat{y}}} \\ \mathbf{D_{\hat{x}\hat{y}}} & \mathbf{I} \end{bmatrix},$$

where $\mathbf{D_{\hat{x}\hat{y}}}$ has the diagonal elements

$$\mathbf{D_{\hat{x}\hat{y}}}[i,i] = \frac{\sigma_a\sigma_b\rho_{ab}\omega_i}{\sqrt{\left(\sigma_a^2\omega_i+\sigma_\mu^2\right)\left(\sigma_b^2\omega_i+\sigma_\nu^2\right)}}$$

and is zero elsewhere.

Proof. The whitening transforms on $\mathbf{x}$ and $\mathbf{y}$ are given by (see Section B.3)

$$\mathbf{T_x} = \Lambda_\mathbf{x}^{-\frac{1}{2}}\mathbf{V}^T \quad \text{and} \quad \mathbf{T_y} = \Lambda_\mathbf{y}^{-\frac{1}{2}}\mathbf{V}^T$$

where $\mathbf{V}$ is a matrix with columns that are the eigenvectors of $\mathbf{R}$ ($\mathbf{K_x}$ and $\mathbf{K_y}$ share the eigenvectors of $\mathbf{R}$ — see Appendix B.5). Using the result of Section B.4,

$$\Lambda_\mathbf{x} = \sigma_a^2\Lambda_R + \sigma_\mu^2\mathbf{I} \quad \text{and} \quad \Lambda_\mathbf{y} = \sigma_b^2\Lambda_R + \sigma_\nu^2\mathbf{I} \tag{B.5}$$

respectively.

Performing the independent transformations $\hat{\mathbf{x}} = \mathbf{T_x}\mathbf{x}$ and $\hat{\mathbf{y}} = \mathbf{T_y}\mathbf{y}$ is equivalent to the

[1] The result of a whitening transform on $\mathbf{x}$ is an image vector with statistically independent pixels that have unit variance.

transform

$$\acute{\mathbf{w}} = \mathbf{T}_\mathbf{w}\mathbf{w}, \text{ where } \mathbf{T}_\mathbf{w} = \begin{bmatrix} \mathbf{T}_\mathbf{x} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_\mathbf{y} \end{bmatrix},$$

on the image pair.

If a random vector $\mathbf{z}$ with covariance matrix $\mathbf{K}_\mathbf{z}$ is transformed by $\mathbf{B}$, then the covariance matrix of the transformed $\mathbf{z}$ is $\mathbf{B}\mathbf{K}_\mathbf{z}\mathbf{B}^T$ [29, p. 6, Theorem 1.2.6]. The covariance matrix of the transformed image pair is therefore

$$\begin{aligned}
\mathbf{K}_{\acute{\mathbf{w}}} &= \mathbf{T}_\mathbf{w}\mathbf{K}_\mathbf{w}\mathbf{T}_\mathbf{w}^T \\
&= \begin{bmatrix} \mathbf{K}_{\acute{\mathbf{x}}} & \mathbf{K}_{\acute{\mathbf{x}}\acute{\mathbf{y}}} \\ \mathbf{K}_{\acute{\mathbf{y}}\acute{\mathbf{x}}} & \mathbf{K}_{\acute{\mathbf{y}}} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{T}_\mathbf{x} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_\mathbf{y} \end{bmatrix} \begin{bmatrix} \sigma_a^2\mathbf{R}+\sigma_\mu^2\mathbf{I} & \sigma_a\sigma_b\rho_{ab}\mathbf{R} \\ \sigma_a\sigma_b\rho_{ab}\mathbf{R} & \sigma_b^2\mathbf{R}+\sigma_\nu^2\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{T}_\mathbf{x} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_\mathbf{y} \end{bmatrix}^T \\
&= \begin{bmatrix} \sigma_a^2\mathbf{T}_\mathbf{x}\mathbf{R}\mathbf{T}_\mathbf{x}^T+\sigma_\mu^2\mathbf{T}_\mathbf{x}\mathbf{T}_\mathbf{x}^T & \sigma_a\sigma_b\rho_{ab}\mathbf{T}_\mathbf{x}\mathbf{R}\mathbf{T}_\mathbf{y}^T \\ \sigma_a\sigma_b\rho_{ab}\mathbf{T}_\mathbf{y}\mathbf{R}\mathbf{T}_\mathbf{x}^T & \sigma_b^2\mathbf{T}_\mathbf{y}\mathbf{R}\mathbf{T}_\mathbf{y}^T+\sigma_\nu^2\mathbf{T}_\mathbf{y}\mathbf{T}_\mathbf{y}^T \end{bmatrix}
\end{aligned}$$

By the property of the diagonalizing transform

$$\mathbf{K}_{\acute{\mathbf{x}}} = \mathbf{K}_{\acute{\mathbf{y}}} = \mathbf{I}.$$

The cross covariance matrix $\mathbf{K}_{\acute{\mathbf{x}}\acute{\mathbf{y}}}$, is

$$\begin{aligned}
\mathbf{K}_{\acute{\mathbf{x}}\acute{\mathbf{y}}} &= \sigma_a\sigma_b\rho_{ab}\mathbf{T}_\mathbf{x}\mathbf{R}\mathbf{T}_\mathbf{y}^T \\
&= \sigma_a\sigma_b\rho_{ab}\Lambda_\mathbf{x}^{-\frac{1}{2}}\mathbf{V}^T\mathbf{R}\mathbf{V}\Lambda_\mathbf{y}^{-\frac{1}{2}} \\
&= \sigma_a\sigma_b\rho_{ab}\Lambda_\mathbf{x}^{-\frac{1}{2}}\Lambda_R\Lambda_\mathbf{y}^{-\frac{1}{2}}.
\end{aligned}$$

Using (B.5) and denoting the eigenvalues of $\mathbf{R}$ as ω_i, $\mathbf{K}_{\acute{\mathbf{x}}\acute{\mathbf{y}}}$ is the diagonal matrix $\mathbf{D}_{\acute{\mathbf{x}}\acute{\mathbf{y}}}$, with diagonal elements

$$\mathbf{D}_{\acute{\mathbf{x}}\acute{\mathbf{y}}}[i,i] = \frac{\sigma_a\sigma_b\rho_{ab}\omega_i}{\sqrt{\left(\sigma_a^2\omega_i+\sigma_\mu^2\right)\left(\sigma_b^2\omega_i+\sigma_\nu^2\right)}}.$$

Similarly, $\mathbf{K}_{\acute{y}\acute{x}}$ is also equal to $\mathbf{D}_{\acute{x}\acute{y}}$ and

$$\mathbf{K}_{\acute{w}} = \begin{bmatrix} \mathbf{I} & \mathbf{D}_{\acute{x}\acute{y}} \\ \mathbf{D}_{\acute{x}\acute{y}} & \mathbf{I} \end{bmatrix}$$

as required. ■

B.7 Type I and Type II Error Probabilities for Normal Hypotheses

In a binary hypothesis testing problem a false positive, or type I error occurs when H_1 is accepted erroneously. The false negative, or type II error occurs when H_1 is rejected erroneously. Given that the test statistic has the following hypothesis conditional pdfs: $p(s|H_0) = N(s; m_0, \sigma_0)$ and $p(s|H_1) = N(s; m_1, \sigma_1)$, where $m_1 > m_0$; and given that the decision threshold is $\acute{\lambda}$, the type I error probability is given by

$$\begin{aligned} P_I &= P_1 \cdot P\left(s < \acute{\lambda}|H_1\right) \\ &= P_1 \int_{-\infty}^{\acute{\lambda}} N(s; m_1, \sigma_1)\ ds \\ &= \frac{P_1}{\sqrt{2\pi\sigma_1^2}} \int_{-\infty}^{\acute{\lambda}} \exp\left[-\frac{1}{2}\frac{(s-m_1)^2}{\sigma_1^2}\right]\ ds \\ &= \frac{P_1}{2}\left[1 + \operatorname{erf}\left(\frac{m-\acute{\lambda}}{\sqrt{2}\sigma}\right)\right]. \end{aligned}$$

The type II error probability is given by

$$\begin{aligned} P_{II} &= P_0 \cdot P\left(s \geq \acute{\lambda}|H_0\right) \\ &= P_0 \cdot \left(1 - P\left(s < \acute{\lambda}|H_0\right)\right) \\ &= P_0 \left(1 - \int_{-\infty}^{\acute{\lambda}} N(s; m_0, \sigma_0)\ ds\right) \\ &= P_0 \left(1 - \frac{1}{\sqrt{2\pi\sigma_0^2}} \int_{-\infty}^{\acute{\lambda}} \exp\left[-\frac{1}{2}\frac{(s-m_0)^2}{\sigma_0^2}\right]\ ds\right) \\ &= \frac{P_0}{2}\left[1 - \operatorname{erf}\left(\frac{m_0-\acute{\lambda}}{\sqrt{2}\sigma_0}\right)\right]. \end{aligned}$$

Bibliography

[1] A. Tversky, "Features of similarity," *Psychological Review*, vol. 84, pp. 327–352, July 1977.

[2] R. C. Jain and T. O. Binford, "Ignorance, myopia, and naivete in computer vision," *CVGIP: Image Understanding*, vol. 53, pp. 112–117, Jan. 1991.

[3] M. A. Snyder, "A commentary on the paper by Jain and Binford," *CVGIP: Image Understanding*, vol. 53, pp. 118–119, Jan. 1991.

[4] Y. Aloimonos and A. Rosenfeld, "A response to 'Ignorance, myopia, and naivete in computer vision' by R. C. Jain and T. O. Binford," *CVGIP: Image Understanding*, vol. 53, pp. 120–124, Jan. 1991.

[5] T. S. Huang, "Computer vision needs more experiments and applications," *CVGIP: Image Understanding*, vol. 53, pp. 125–126, Jan. 1991.

[6] M. Kunt, "Comments on 'Dialog,' a series of articles generated by the paper entitled 'Ignorance, myopia, and naivete in computer vision'," *CVGIP: Image Understanding*, vol. 54, pp. 428–429, Jan. 1991.

[7] F. G. Stremler, *Introduction to Communications Systems.* Addison-Wesley series in Electrical Engineering, Addison-Wesley, 2nd ed., 1982.

[8] A. K. Jain, *Fundamentals of Digital Image Processing.* Prentice-Hall, 1989.

[9] J. G. Kawamura, "Automatic recognition of changes in urban development from aerial photographs," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-1, pp. 230–239, July 1971.

[10] A. Goshtasby, S. H. Gage, and J. F. Batholic, "A two-stage cross correlation approach to template matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, pp. 374–378, May 1984.

[11] P. E. Anuta, "Spatial registration of multispectral and multitemporal digital imagery using fast Fourier transform techniques," *IEEE Transactions on Geoscience Electronics*, vol. GE-8, pp. 353–368, Oct. 1970.

[12] A. V. Oppenheim and R. W. Schafer, *Discrete-time Signal Processing.* Signal Processing Series, Prentice Hall, 1989.

[13] P. E. Green, "The output signal-to-noise ratio of correlation detectors," *IRE Transactions on Information Theory*, vol. IT3, pp. 10–18, Mar. 1957.

[14] G. M. Roe and G. M. White, "Probability density functions for correlators with noisy reference signals," *IEEE Transactions on Information Theory*, vol. IT-7, pp. 13–18, Jan. 1961.

[15] L. C. Andrews, "Output probability density functions for cross correlators utilizing sampling techniques," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-10, pp. 78–81, Jan. 1974.

[16] V. N. Dvornychenko, "Bounds on (deterministic) correlation functions with application to registration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-5, pp. 206–213, Mar. 1983.

[17] B. Jähne, *Digital Image Processing.* Springer-Verlag, 4^{th} ed., 1997.

[18] C. D. Kuglin and D. C. Hines, "The phase correlation image alignment method," in *Proceedings of the IEEE International Conference on Cybernetics and Society*, pp. 163–165, Sept. 1975.

[19] S. Alliney and C. Morandi, "Digital image registration using projections," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, pp. 222–233, Mar. 1986.

[20] E. De Castro and C. Morandi, "Registration of translated and rotated images using finite Fourier transforms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, pp. 700–703, Mar. 1987.

[21] F. Pla and M. Bober, "Estimating translation/deformation motion through phase correlation," *Lecture Notes in Computer Science*, vol. 1310, pp. 653–660, 1997.

[22] W. K. Pratt, "Correlation techniques of image registration," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-10, pp. 353–358, May 1974.

[23] M. Miyahara and K. Kotani, "Block distortion in orthogonal transform coding — analysis, minimization, and distortion measure," *IEEE Transactions on Communications*, vol. COM-33, pp. 90–96, Jan. 1985.

[24] T. M. Buzug, J. Weese, C. Fassnacht, and C. Lorenz, "Image registration: convex weighting functions for histogram-based similarity measures," in *First Joint Conference on Computer Vision, Virtual Reality and Robotics in Medicine and Medical Robotics and Computer-Assisted Surgery*, (Grenoble, France), pp. 203–212, 1996.

[25] D. N. Bhat and S. K. Nayar, "Ordinal measures for image correspondence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 415–423, Apr. 1998.

[26] G. P. Penny, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. Hawkes, "A comparison of similarity measures for use in 2D-3D medical image registration," in *First Conference on Medical Image Computing and Computer Assisted Intervention*, vol. 1496, (Cambridge, MA, USA), pp. 1153–1161, 1998.

[27] R. C. Gonzalez and R. E. Woods, *Digital Image Processing.* Addison-Wesley, 1992.

[28] E. Kreyszig, *Introductory Mathematical Statistics.* John Wiley and Sons, 1970.

[29] R. J. Muirhead, *Aspects of Multivariate Statistical Theory.* John Wiley and Sons, 1982.

[30] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions.* John Wiley and Sons, 1986.

[31] R. Brunelli and S. Messelodi, "Robust estimation of correlation with applications to computer vision," *Pattern Recognition*, vol. 28, pp. 833–841, June 1995.

[32] T. Radcliffe, R. Rajapakshe, and S. Shalev, "Pseudocorrelation: A fast, robust, absolute grey-level image alignment algorithm," *Medical Physics*, vol. 21, pp. 761–769, June 1994.

[33] R. N. Nagel and A. Rosenfeld, "Ordered search techniques in template matching," *Proceedings of the IEEE*, pp. 242–244, Feb. 1972.

[34] G. J. Vanderbrug and A. Rosenfeld, "Two-stage template matching," *IEEE Transactions on Computers*, vol. C-26, pp. 384–393, Apr. 1977.

[35] M. Svedlow, C. D. McGillem, and P. E. Anuta, "Image registration: Similarity measure and preprocessing method comparisons," *IEEE Transactions on Aerospace and Electronics Systems*, vol. AES-14, pp. 141–149, Jan. 1978.

[36] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, vol. 2. Academic Press, 2nd ed., 1982.

[37] G. S. Cox and G. de Jager, "Automatic registration of temporal image pairs for digital subtraction angiography," in *Medical Imaging*, vol. 2167, pp. 188–199, SPIE, Feb. 1994.

[38] P. J. Huber, *Robust Statistics*. John Wiley and Sons, 1981.

[39] M. Boninsegna and M. Rossi, "Similarity measures in computer vision," *Pattern Recognition Letters*, vol. 15, pp. 1255–1260, 1994.

[40] D. I. Barnea and H. F. Silverman, "A class of algorithms for fast digital image registration," *IEEE Transactions on Computers*, vol. C-21, pp. 179–186, Feb. 1972.

[41] J. B. Thomas, "Nonparametric detection," *Proceedings of the IEEE*, vol. 58, no. 5, pp. 623–631, 1970.

[42] A. Venot, J. F. Lebruchec, and J. C. Roucayrol, "A new class of similarity measures for robust image registration," *Computer Vision, Graphics, and Image Processing*, vol. 28, pp. 176–184, 1984.

[43] A. Venot and V. Leclerc, "Automated correction of patient motion and gray values prior to subtraction in digitized angiography," *IEEE Transactions on Medical Imaging*, vol. MI-3, pp. 179–186, Dec. 1984.

[44] A. Venot, J. Y. Devaus, M. Herbin, J. F. Lebruchec, L. Dubertret, Y. Raulo, and J. C. Roucayrol, "An automated system for the registration and comparison of photographic images in medicine," *IEEE Transactions on Medical Imaging*, vol. 7, pp. 298–303, Dec. 1988.

[45] M. Herbin, A. Venot, J. Y. Devaux, E. Walter, J. F. Lebruchec, L. Dubertret, and J. C. Roucayrol, "Automated registration of dissimilar images: application to medical imagery," *Computer Vision, Graphics and Image Processing*, vol. 47, pp. 77–88, 1989.

[46] A. Venot, L. Pronzato, and E. Walter, "Comments about the coincident bit counting (CBC) criterion for image registration," *IEEE Transactions on Medical Imaging*, vol. 13, pp. 565–566, Sept. 1994.

[47] J. Y. Chiang and B. J. Sullivan, "Coincident bit counting — a new criterion for image registration," *IEEE Transactions on Medical Imaging*, vol. 12, pp. 30–38, Mar. 1993.

[48] G. S. Garret, E. L. Reagh, and E. B. Hibbs Jr., "Detection threshold estimation for digital area correlation," *IEEE Transactions on Systems, Man, and Cybernetics*, pp. 65–70, Jan. 1976.

[49] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," in *International Conference on Computer Vision*, pp. 16–23, June 1995.

[50] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multi-modality image registration by maximization of mutual information," in *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pp. 14–22, June 1996.

[51] F. Maes, A. Collignon, D. Vermeulen, G. Marchal, and P. Seutens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging*, vol. 16, pp. 187–198, Apr. 1997.

[52] T. Buzug and J. Weese, "Similarity measures for subtraction methods in medical imaging," in *18th Annual International Conference of the IEEE EMBS*, p. 140, 1996.

[53] B. Moghaddam, C. Nastar, and A. Pentland, "A Bayesian similarity measure for direct image matching," in *International Conference on Pattern Recognition*, Aug. 1996.

[54] B. Moghaddam, T. Jebara, and A. Pentland, "Efficient MAP/ML similarity matching for visual recognition," in *International Conference on Pattern Recognition*, vol. 1, pp. 876–881, 1998.

[55] P. Aschwanden and W. Guggenbül, "Experimental results from a comparative study on correlation-type registration algorithms," in *International Workshop on Robust Computer Vision*, no. 2, pp. 268–289, Mar. 1992.

[56] J. L. Horner and P. D. Gianino, "Phase-only matched filters," *Applied Optics*, vol. 23, pp. 812–816, Mar. 1984.

[57] E. H. W. Meijering, W. J. Niessen, and M. A. Vergiever, "Retrospective motion correction in digital subtraction angiogrpahy: A review," *IEEE Transactions on Medical Imaging*, vol. 18, pp. 2–21, Jan. 1999.

[58] H. Mostafavi and F. W. Smith, "Image correlation with geometric distortion part I: Acquisition performance," *IEEE Transactions on Aerospace and Electronics Systems*, vol. AES-14, pp. 487–493, May 1978.

[59] H. Mostafavi and F. W. Smith, "Image correlation with geometric distortion part II: Effect on local accuracy," *IEEE Transactions on Aerospace and Electronics Systems*, vol. AES-14, pp. 494–500, May 1978.

[60] F. A. Sadjadi, "Performance evaluations of correlations of digital images using different separability measures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-4, pp. 436–441, July 1982.

[61] A. M. Mood, F. A. Graybill, and D. C. Boes, *Introduction to the Theory of Statistics.* McGraw-Hill, 3rd ed., 1974.

[62] C. W. Helstrom, *Elements of Signal Detection and Estimation.* Prentice Hall, 3rd ed., 1995.

[63] M. Akay, *Detection and Estimation Methods for Biomedical Signals.* Academic Press, 1996.

[64] D. Kazakos and P. Papantoni-Kazakos, *Detection and Estimation.* Computer Science Press, 1990.

[65] H. L. V. Trees, *Detection, Estimation, and Modulation Theory.* John Wiley and Sons, 1968.

[66] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis.* Wiley-Interscience, 1973.

[67] S.-T. Bow, *Pattern Recognition and Image Processing.* Marcel Dekker, 1992.

[68] C. M. Bishop, *Neural Networks for Pattern Recognition.* Oxford University Press, 1995.

[69] B. D. Ripley, *Pattern Recognition and Neural Networks.* Cambridge University Press, 1996.

[70] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[71] T. Kohonen, *Self-Organization and Associative Memory*. Springer Series in Information Sciences, Springer-Verlag, 3rd ed., 1989.

[72] M. A. Turk and A. P. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[73] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 586–591, June 1991.

[74] S. T. Barnard and W. B. Thompson, "Disparity analysis of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-2, pp. 333–340, July 1980.

[75] B. R. Hunt and T. M. Cannon, "Nonstationary assumptions for Gaussian models of images," *IEEE Transactions on Systems, Man, and Cybernetics*, pp. 876–882, Dec. 1976.

[76] F. O. Huck, R. Alter-Gartenberg, and Z. Rahman, "Image gathering and digital restoration for fidelity and visual quality," *CVGIP: Graphical Models and Image Processing*, vol. 53, pp. 71–84, Jan. 1991.

[77] S. J. Reeves, "Optimal space-varying regularization in iterative image restoration," *IEEE Transactions on Image Processing*, vol. 3, pp. 319–324, May 1994.

[78] W. H. Pun and B. D. Jeffs, "Adaptive image restoration using a generalised Gaussian model for unknown noise," *IEEE Transactions on Image Processing*, vol. 4, pp. 1451–1456, Oct. 1995.

[79] M. K. Tsatsanis and G. B. Giannakis, "Object and texture classification using higher order statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 733–750, July 1992.

[80] S. Krishnamachari and R. Chellapa, "Multi-resolution Gauss-Markov random field models for texture segmentation," *IEEE Transactions on Image Processing*, vol. 6, pp. 251–267, Feb. 1997.

[81] K. Popat and R. W. Picard, "Cluster-based probability model and its application to image and texture processing," *IEEE Transactions on Image Processing*, vol. 6, pp. 268–284, Feb. 1997.

[82] A. Gersho and R. M. Gray, *Vector Quantization And Signal Compression.* Communications and Information Theory, Kluwer Academic Publishers, 1992.

[83] J. W. Woods, "Image detection and estimation," in *Digital Image Processing Techniques* (M. P. Ekstrom, ed.), ch. 3, pp. 77–110, Academic Press, 1984.

[84] R. N. Strickland, "Tumor detection in nonstationary backgrounds," *IEEE Transactions on Medical Imaging*, vol. 13, pp. 491–499, Sept. 1994.

[85] M. Basseville, A. Benveniste, K. C. Chou, S. A. Golden, R. Nikoukhah, and A. S. Willsky, "Modeling and estimation of multiresolution stochastic processes," *IEEE Transactions on Information Theory*, vol. 38, pp. 766–784, Mar. 1992.

[86] R. W. Dijkerman and R. R. Mazumdar, "Wavelet representations of stochastic processes and multiresolution stochastic models," *IEEE Transactions on signal Processing*, vol. 42, pp. 1640–1652, July 1994.

[87] X. Zhuang, Y. Huang, K. Palaniappan, and Y. Zhao, "Gaussian mixture density modeling, decomposition and applications," *IEEE Transactions on Image Processing*, vol. 5, pp. 1293–1302, Sept. 1996.

[88] B. R. Hunt, "Nonstationary statistical image models (and their application to image data compression)," *Computer Graphics and Image Processing*, vol. 12, pp. 173–186, 1980.

[89] P. B. Chapple and D. C. Bertilone, "Stochastic simulation of infrared non-Gaussian terrain imagery," *Optics Communications*, no. 150, pp. 71–76, 1998.

[90] G. E. Johnson, "Constructions of particular random processes," *Proceedings of the IEEE*, vol. 82, pp. 270–285, Feb. 1994.

[91] F. Nicolls and G. S. Cox, "Synthesis of multivariate Gaussian images," Tech. Rep. 01-03-2000, DebTech Research: Machine Intelligence, Jan. 2000.

[92] M. Kendall and A. Stuart, *The Advanced Theory of Statistics.* Charles Griffin and Company, 4^{th} ed., 1979.

[93] A. Papoulis, *Probability, Random Variables and Stochastic Processes.* McGraw-Hill, 3rd ed., 1991.

[94] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods.* Methuen and Co., 1965.

[95] I. Selin, *Detection Theory.* Princeton University Press, 1965.

[96] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis.* John Wiley and Sons, 1958.

[97] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Journal of the Optical Society of America A*, vol. 4, pp. 519–523, Mar. 1987.

[98] M. Kirby and L. Sirovich, "Application of the Karhunen-Loève procedure for the characterization of human faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 103–108, Jan. 1990.

[99] L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, pp. 325–376, Dec. 1992.

[100] M. Moshfeghi, "Elastic matching of multimodality medical images," *CVGIP: Graphical Models and Image Processing*, vol. 53, pp. 271–281, May 1991.

[101] A. Goshtasby, "Piecewise linear mapping functions for image registration," *Pattern Recognition*, vol. 19, no. 6, pp. 459–466, 1986.

[102] J. Flusser, "An adaptive method for image classification," *Pattern Recognition*, vol. 25, no. 1, pp. 45–54, 1992.

[103] R. J. Althof, M. G. J. Wind, and J. T. Dobbins III, "A rapid and automatic image registration algorithm with subpixel accuracy," *IEEE Transactions on Medical Imaging*, vol. 16, pp. 308–316, June 1997.

[104] G. L. Bretthorst, "An introduction to parameter estimation using bayesian probability theory," in *Maximum Entropy and Bayesian Methods* (P. F. Fougere, ed.), pp. 53–79, Kluwer Academic Publishers, 1990.

[105] K.-I. Mori, M. Kidode, and H. Asada, "An iterative prediction and correction method for automatic stereocomparison," *Computer Graphics and Image Processing*, vol. 2, pp. 393–401, 1973.

[106] L.-W. Lee, J.-F. Wang, J.-Y. Lee, and J.-D. Shie, "Dynamic search window adjustment and interlaced search for block-matching algorithm," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, pp. 85–87, Feb. 1993.

[107] A. Rosenfeld and G. J. Vanderbrug, "Coarse-fine template matching," *IEEE Transactions on Systems, Man, and Cybernetics*, pp. 104–107, Feb. 1977.

[108] W. R. Brody, "Digital subtraction angiography," *IEEE Transactions on Nuclear Science*, vol. NS-29, pp. 1176–1180, June 1982.

[109] A. Oppelt, "Possibilities for dose reduction with modern X-ray systems," *Electromedica*, no. 2, pp. 58–61, 1997.

[110] A. J. Gonzalez, "Radiation safety: New international standards," *International Atomic Energy Agency Bulletin*, no. 2, pp. 2–11, 1994.

[111] S. J. Beningfield, J. H. Potgieter, P. Bautz, M. Shackleton, E. Hering, G. de Jager, G. Bowie, M. Marshall, G. S. Cox, G. Pagliari, and N. Coetzee, "Evaluation of a new type of direct digital radiography machine," *South African Medical Journal*, vol. 89, pp. 1182–1188, Nov. 1999.

[112] A. Margalit, I. S. Reed, and R. M. Gagliardi, "Adaptive optical target detection using correlated images," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-21, pp. 777–786, May 1985.

[113] P. B. Chapple, D. C. Bertilone, and S. Angeli, "Non-Gaussian model for analysis of automatic detection/recognition," in *International Workshop on Statistical Pattern Recognition*, pp. 897–904, Aug. 1998.

[114] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes.* Oxford Science Publications, 2nd ed., 1993.

[115] A. Jeffrey, *Linear Algebra and Ordinary Differential Equations.* Blackwell Scientific Publications, 1990.