CANADIAN THESES

THÈSES CANADIENNES

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ MICROFILMÉE TELLE QUE NOUS L'AVONS REÇUE

The University of Alberta

Invariance Coding in Image Processing

by

Nicola Joy Ferrier

A thesis

submitted to the Faculty of Graduate Studies and Research

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Computing Science

Edmonton, Alberta

Spring, 1987

ISBN 0-315-37742-9

The University of Alberta

Release Form

Name of Author: Nicola Joy Ferrier

Title of Thesis: *Invariance Coding in Image Processing*

Degree for Which Thesis was Presented: Master of Science

Year This Degree Granted: 1987

Permanent Address: 13807 - 91 Avenue
Edmonton, Alberta
T5R 4X8, Canada

Dated 23 December 1988

The University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled *Invariance Coding in Image Processing* submitted by Nicola Joy Ferrier in partial fulfillment of the requirements for the degree of Master of Science.

W. A. Davis
.....
Supervisor

Mulcelh —
.....

M. M. Armstrong
.....

Z. J. Koles
.....

Xiaobo Li
.....

Date 16 Apr 87

Abstract

This thesis studies the mathematical problem of representing a two dimensional image so that the representation is invariant with respect to certain geometric transformations. Previous work on the problem of recognition under translation, scaling and rotation is reviewed and analyzed. A new image decomposition is introduced having scale and rotation invariance properties. A set of experiments is presented and are used to compare and evaluate various invariance coding methods.

Acknowledgements

As with any major project, the completion of this thesis would not be possible without the help and time of many others. The author wishes to acknowledge gratefully her supervisor, Terry M. Caelli for his guidance and encouragement in the preparation of this thesis; committee members, Wayne Davis, Zoley Koles, William Armstrong and Xiabo Li for their time and suggestions. The author thanks Martin Dubetz, for providing support software for image manipulation on the International Imaging System and for his continued interest throughout; to Ted Bentley for his help in textformatting the document; and to Roderick Johnson, George Labahn, and John Kmet for proofreading. The author expresses appreciation to the late Professor J. R. Sampson for his encouragement to pursue graduate work in computer science and to her parents and Roderick Johnson for their patience, encouragement and support. Financial support was given by the Department of Computing Science, University of Alberta, and the Natural Sciences and Engineering Reseach Council of Canada, through a postgraduate scholarship.

Table of Contents

List of Tables

List of Figures

List of Photographic Plates

Notational Conventions

A function denoted, $f(x,y)$ in Cartesian coordinates will be denoted by $f(r,\theta)$ in polar coordinates.

$f^*(x,y)$ denotes the complex conjugate of $f(x,y)$.

$f(x,y) * g(x,y)$ denotes the convolution of the functions $f(x,y)$ and $g(x,y)$.

$\mathbf{FT}[f(x,y)]$ represents the two dimensional Fourier Transform of $f(x,y)$. $\mathbf{FT}_x[f(x,y)]$ represents the one dimensional Fourier Transform, in the variable x, of $f(x,y)$.

On the photographic plates, black represents small values, and white represents large values.

Chapter 1

Introduction

This thesis is concerned with ways of representing digital images such that the representation is independent of as many parameters as possible. Specific parameters include: position, scale, orientation, etc. Parameter independent representation is usually termed "invariance coding". Invariance coding, that is, the problem of converting an image in a form which is invariant to certain geometric and affine transforms, is of interest in digital image processing. By utilizing image transformations, one may convert an image to such a form in which certain properties are more easily comparable. This thesis will look for representations which allow recognition of rotated or scaled objects.

1.1. Definition of the Problem

The location of an object within a scene is a relatively simple task for a human to perform. However, this problem of image recognition is difficult for a computer to perform as a digital image processing task, and appears in many forms. The notion of *recognition* in itself is not clearly defined. In what ways can the viewer recognize an image? How is it decided if two images are the same? One form of recognition involves deciding in which particular class, of a given set of image classes, an input image belongs. Usually the classes have been predefined and the assignment to a class may be based on predefined rules. Early computer recognition systems such as those of Uhr and Vossler [19] are of this type. Recognition of typewritten characters also falls in this type of recognition problem.

Another form of image recognition could be that of classifying *abstract* images. Abstract images are characterized by a property or group of properties. Examples of this include recognizing geometric shapes such as circles, squares, trapezoids and triangles. A similar method is that of classification by function. Objects may be

recognized as belonging to a class by the function they perform rather than a specific appearance, or shape. Chairs, lamps, etc. are objects recognized by function [11].

Other images are recognized by a complete structural analysis. This type of recognition represents an image as a set of primitives, and the relationships between these elements. Handwritten characters are not of a specific form, as typewritten characters are, but they are still recognizable through an analysis of their constituent components (or features) and the relative position and size of these components. This type of recognition has been explored by many [2,9].

In each of the above descriptions, recognition does not depend on the observational conditions, but rather on the object itself. The size, position, and orientation was not a factor. The recognition process is invariant under projective and affine transforms [19]. The question occurs: Can this same invariant recognition be achieved in digital image processing? A predefined model of the image can be used or assumptions can be made about the image so that a normalization process can be applied to a given image in order to centre the image at a specific location, size and orientation. Such normalization requires *a priori* knowledge about the image and thus is not general enough for a recognition system. The problem is to find a recognition process to determine if two images are similar (within a geometric or affine transform) without *a priori* knowledge about the images.

In 1964, Van der Lugt [21] introduced the idea of matched filtering to locate a signal embedded in a scene. The matched filter provides shift-invariant pattern recognition which is limited to finding a pattern only when the size and orientation of the image is known. When the size and orientation are not known, however, the problem becomes much more difficult. Systems have been proposed offering rotation or size invariance [7] [17]. These systems usually forfeit uniqueness, or are sensitive to noise, or are subject to error. Indeed the "invariance coding problem" is precisely concerned

with developing matching techniques which are rotation, scale and shift invariant and, while preserving uniqueness of the pattern, are not sensitive to noise effects.

1.2. Applications

The applications of a recognition system offering both rotation and scale invariance are the same as for existing recognition systems. Without the constraint of exact size and orientation, recognition is possible without preprocessing to register two images. Text written with different font sizes should still be recognizable by a character recognition system. Aerial photographs cannot realistically be taken from a fixed distance so their analysis would require scale invariance to overcome the size changes due to the different distances at which photos are taken. Navigation and computer guidance systems could use invariant transforms to match maps drawn at different scales. Medical imaging systems perform x-ray analysis, such as tumour detection. Robotics, and vision systems in industry requiring the recognition of objects would benefit, since many objects used by such systems should be recognizable regardless of size and orientation.

1.3. The Invariance Coding Problem

The invariance coding problem requires the development of matching techniques which are rotation, scale and shift invariant, and which preserve uniqueness. This section is intended to establish the approach to this problem that the thesis will take. The invariance coding problem will be considered using a formal mathematical approach.

All forms of recognition require some form of description or representation of an image. Such representations include: a description of features; a semantic net connecting features, their relation to each other and their relation to other objects in the image [2]; and a mathematical function. In this thesis, functional representation will

be used to formalize the recognition problem. An image is defined as a function, $f(x,y)$ where each value $f(x,y)$ represents the light intensity at the point (x,y). In an image processing application such as in this thesis, some assumptions can be made about this function [23]. Typically $f(x,y)$ is bounded, non-zero only within a finite region (optical systems have a finite field of view), and image functions are assumed to be analytically well-behaved, that is, they are integrable, differentiable, have no infinite discontinuities and only a finite number of discontinuities. Furthermore, in any finite neighbourhood they have a finite number of maxima and minima. These assumptions are reasonable for an image.

1.3.1. Measurement of Match

Using the functional representation, and computing some distance metric between the two functions (or between a transform of the two functions) it is possible to measure how "close" the two functions are and hence determine how alike the two images are. Given a reference pattern (or signal, or image), $f(x,y)$ and a test pattern $g(x,y)$ one would like a method of determining how closely $g(x,y)$ matches $f(x,y)$. Typically, $g(x,y)$ is smaller than $f(x,y)$ and it would be desirable to know *where* within $f(x,y)$ does $g(x,y)$ match the closest. There are various match criteria used over the decades [23] including (within a region **A**):

$$\min_{\mathbf{A}} |f-g| : \iint_{\mathbf{A}} |f-g| : \iint_{\mathbf{A}} (f-g)^2. \qquad 1.1$$

All of the above compare the images on a pixel by pixel basis. The third match criterion gives rise to cross correlation techniques, as one can decide if a test image, $g(x,y)$, matches an image $f(x,y)$ at any position. The cross correlation is defined by:

$$C_{fg}(u,v) = \iint f(x,y)g(x+u,y+v)\,dx\,dy. \qquad 1.2$$

From the mean square measure of match: $\Delta = \iint_{\mathbf{A}} (f-g)^2$ the following is obtained:

$$\Delta = \iint_A (f-g)^2 = \iint_A f^2 + \iint_A g^2 - 2\iint_A fg. \qquad 1.3$$

Which gives $\Delta=0$ if and only if $f(x,y)$ and $g(x,y)$ match exactly. Observe that Δ is smallest when the last term, $\iint_A fg$ is largest. The Cauchy-Schwarz inequality gives, for nonnegative f and g:

$$\iint_A fg \le \sqrt{\iint_A f^2 \iint_A g^2}, \qquad 1.4$$

with equality attained if and only if $f(x,y)=cg(x,y)$, c being a scalar [23]. Since $\iint_A g^2$ is not, in general, constant, the "normalized" (pre-whitened) cross correlation function N_{fg} is:

$$N_{fg} = \frac{C_{fg}(u,v)}{\sqrt{\iint_A g^2}}, \qquad 1.5$$

which, again, is a measure of the match between f and g which takes into account the energy of g.

This match measure is optimal when the images $f(x,y)$ and $g(x,y)$ differ by a translation plus statistically independent additive Gaussian noise or when the match criterion is the minimization of the mean-squared error of two images [23]. However, if the images do not have the same orientation and size this method degrades rapidly. Since each of the above measures of match is based on a pixel-by-pixel comparison, scaled and rotated images could not be expected to match since there is only one point in a scaled or rotated image that does not change from the original image.

The matched filter and cross correlation techniques provide a method for recognizing images at a specific orientation and size. At the time these techniques were introduced, Doyle [12] claimed that it is "impractical to include distortions such as rotations and dilations" in a recognition system. Since that time, various techniques have been used to make it possible to recognize objects under various geometric

transformations. This thesis will review techniques for image recognition which approach the problem on a functional basis.

Invariance coding deals with the problem of finding a *representation* of an image such that the representation is invariant to certain transforms; in particular, translation, rotation, and scale. Thus by transforming the image to a new representation which possesses invariant properties, comparisons of images can be made in this new space. This allows a conclusion to be drawn about the original image in the spatial domain. All techniques considered are *magnitude-invariant*, that is, the representation of the image in a complex plane possesses a constant magnitude for the mappings of geometrically transformed images. With constant magnitude in the complex plane, the phase component can vary. Thus, the combined magnitude and phase of a transform can uniquely represent an image in the complex plane.

This thesis will study the mathematical problem of invariant encoding of images in two dimensions. Based on the properties of the techniques reviewed, a new decomposition of an image is introduced having inherent scale and rotation invariant properties. This transform provides the representation of the image in a form which allows us to determine if images in the spatial domain are the same, within scaling and rotation.

1.4. Outline of the Thesis

Chapter 2 contains background material on which all further work is based and from which all comparisons and evaluations are drawn. Current techniques for recognition under image translation, scaling and rotation are presented.

Chapter 3 proposes a new integral transform for scale and rotation invariant encoding of images. This transform is seen to be an extension of the existing transforms presented in Chapter 2. Properties of this transform are explored.

Chapter 4 presents experiments in which the methods for invariance coding discussed in the preceding two chapters are compared and evaluated.

Chapter 5 presents the conclusions and offers a summary.

Chapter 2

Background

This chapter looks at various techniques currently employed in invariant recognition of images. All methods reviewed here consider the invariance coding problem as described in the first chapter: Find a representation of the image, either by one transformation or a series of transformations such that geometric transformations performed on the original image will not affect the representation.

2.1. Introduction

There have been many investigations into invariant operators with the property that scaling the image and then performing the transform or transforming the image and then scaling the transform are equivalent[18]. This does not agree with the definition of invariance used in this thesis, thus these transforms will not be considered. The matched filter is discussed as it was one of the first comparison techniques and still is the most frequently used criterion. Next, circular symmetric filters and circular harmonic image decompositions are reviewed. These techniques are rotation invariant mechanisms for recognition. The Fourier-Mellin techniques also have rotation invariance and in addition to this have scale and position invariance but lose uniqueness. Each technique is explained to allow comparisons on a theoretical basis and to provide the background necessary for the discussions in Chapter 4.

2.2. Matched Filters

The matched filter theorem states that for finding matches between a template, g, and an image, f, by cross correlating an *arbitrary* filter with f, the *best* filter to use is g itself when f consists of a signal, g, translated to any position, plus stationary white noise, (n), ie., $f = g + n$. The term *best* indicates optimal in a mean-square error sense, or optimal in maximizing the ratio of signal power to expected noise power (see [23]).

However, if g is a rotated version of f then the theorem does not hold. In fact for rotations larger than ± 5 to $10°$ the decay is too great, and no match will be found using these cross correlation techniques.

Since using the normalized cross correlation, N_{fg} for matching a template $g(x,y)$ with an image $f(x,y)$ is very sensitive to changes in scale and orientation, it cannot be used for matching for arbitrary rotations and scale sizes. One solution is to use many templates for $f(x,y)$, of different orientations and sizes. Using many templates for each pattern is not feasible, as storage requirements for many templates become unreasonable and computation time increases with the number of templates. This approach lacks elegance: It is a method of overcoming the short-comings of the matched filter, but it is not a solution to the problem of invariant recognition.

2.3. Filtering with Circularly Symmetrical Functions

The idea of circular symmetrical filtering [29] is to use the optimum circular symmetric function instead of $f(x,y)$ as input to the conventional matched filter. With the conventional filter, if $g(x,y)=f(x-a,y-b)$ then the cross correlation:

$$C_{gf}=f(x-a,y-b)*f(x,y), \tag{2.1}$$

yields a peak at the point (a,b). The function $f(x,y)$ can be expressed as:

$$f(x,y) = h(x,y) + k(x,y), \tag{2.2}$$

where $h(x,y)$ is a circular symmetric function. Now the cross correlation is

$$C_{gf} = f(x-a,y-b) * h(x,y) + f(x-a,y-b)*k(x,y). \tag{2.3}$$

If a proper $h(x,y)$ is chosen then

$$C_{gf}\approx f(x-a,y-b)*h(x,y), \tag{2.4}$$

will recognize images under rotations. This filter uses a proper circular symmetric component of the pattern of interest, $h(x,y)$, rather than f itself. The optimized

$h(x,y)$ is called the *optimum circular symmetric filter* (OCSF). The resulting filter is claimed to be rotation and space invariant. These claims, however, are subject to some clarification, as there are some questions which arise from the above discussion: (1) Is $h(x,y)$ required to be circularly symmetric to gain rotation invariance? (2) What defines an optimal $h(x,y)$? That is, which circular symmetric components should be utilized? (3) Around which point should $h(x,y)$ be circular symmetric? (that is, where is the centre of expansion?).

In answer to the first problem above, it can be shown that the transfer function of a space-invariant and rotation invariant linear system *must* be circularly symmetric (see [29]).

Using mean-square error to define optimal match and assuming that the centre has been found (Question 3 above has been solved), then let the centre be (0,0) in our coordinate system. Our optimization criterion is:

$$\int_{-\infty}^{\infty}\int |f-h|^2 dxdy = minimum. \qquad 2.5$$

The notation $\tilde{f}(r,\theta)$ will be used for polar representation. Using $\tilde{f}(r,\theta)=\tilde{h}(r)+\tilde{k}(r,\theta)$ Yang *et al.* [29] prove that the required h is:

$$\tilde{h}(r) = \frac{1}{2\pi}\int_0^{2\pi} \tilde{f}(r,\theta)d\theta. \qquad 2.6$$

This formula can be used to calculate the circular symmetric components, $h(r)$ if the required centre is found.

To solve Question (3) above, the centre of symmetry is the point (a,b) at which

$$\int_0^{2\pi}\int_0^{\infty} |\tilde{f}(r,\theta)-\tilde{h}(r)|^2 r dr d\theta = minimum. \qquad 2.7$$

where $\tilde{f}(r,\theta)=f_o(r\cos(\theta)+a, r\sin(\theta)+b)$ for f_o in the original coordinate system. Using the result from above,

$$h(r)=\frac{1}{2\pi}\int_0^{2\pi} f_o(r\cos\theta+a, r\sin\theta+b)d\theta, \qquad 2.8$$

The conditions found for the centre of symmetry are (see [29] for the derivation):

$$\frac{\partial}{\partial a}\int_0^{\infty}\left|\int_0^{2\pi} f_o(r\cos\theta+a, r\sin\theta+b)d\theta\right|^2 r\,dr=0, \qquad 2.9$$

$$\frac{\partial}{\partial b}\int_0^{\infty}\left|\int_0^{2\pi} f_o(r\cos\theta+a, r\sin\theta+b)d\theta\right|^2 r\,dr=0. \qquad 2.10$$

Without an analytic expression for $f_o(x,y)$, however, it is difficult to find the centre using this derivation.

Another problem to consider is that the peak of C_{fh} is not, in general, at the origin of the output coordinate system even if the optimum symmetric centre of $f(x,y)$ coincides with the origin of the input coordinate system. Rotating the pattern $f(x,y)$ around its optimum centre of symmetry (0,0), the peak of C_{fh} is not only rotated but also shifted. Thus the peak denotes *recognition*, but not the position at which recognition occurred.

The claim that this method is shift-invariant must be taken with a grain of salt. Finding the centre of the rotated image and then expanding about this centre is equivalent to converting the coordinate systems so that the images are aligned. Finding the centre is an important part of the invariance coding problem.

2.4. Circular Harmonic Decomposition

Hsu, Arsenault and April [17] use circular harmonic expansion based on work originally done in image reconstruction [16]. Their method decomposes the function into circular harmonic components of the form:

$$f(r,\theta)=\sum_{m=-\infty}^{\infty} f_m(r)e^{jm\theta}, \qquad 2.11$$

where

$$f_M(r)=\frac{1}{2\pi}\int_0^{2\pi}\hat{f}(r,\theta)e^{-jM\theta}d\theta. \qquad 2.12$$

A target rotated by an angle α would be represented by:

$$\hat{f}(r,\theta+\alpha)=\sum_{m=-\infty}^{\infty} f_M(r)e^{jM\theta}e^{jM\alpha}. \qquad 2.13$$

In Cartesian coordinates $\hat{f}(r,\theta)$ and $\hat{f}(r,\theta+\alpha)$ are expressed as $f_o(x,y)$ and $f_\alpha(x,y)$, respectively. Using a conventional matched filter the output is:

$$C_\alpha(x,y)=\int_{-\infty}^{\infty}\int f_\alpha(\xi,\eta)f_o^*(\xi-x,\eta-y)d\xi d\eta, \qquad 2.14$$

where * denotes complex conjugation. When $\alpha=0$ this is the autocorrelation of f with the peak at $x=0,y=0$. For an arbitrary value of α, the centre value ($x=0,y=0$) may not be the correlation peak. Hsu *et al.* [17] define $C(\alpha)$ such that

$$C(\alpha)=C_\alpha(0,0)=\int_0^{\infty} r\,dr\int_0^{2\pi}\hat{f}(r,\theta+\alpha)\hat{f}^*(r,\theta)d\theta. \qquad 2.15$$

using equations 2.11 and 2.13 this reduces to:

$$C(\alpha)=2\pi\sum_{m=-\infty}^{\infty}e^{jm\alpha}\int_0^{\infty} r|f_m(r)|^2\,dr. \qquad 2.16$$

This centre correlation value is a sum depending on contributions from all circular harmonic components and in general *varies* with α. Therefore, if more than one circular harmonic component of the target is used then the centre correlation value is ***rotation-variant***. Using only one component, however, this value does not change.

One circular harmonic component is defined as:

$$f_m(r,\theta)=f_m(r)e^{jm\theta}. \qquad 2.17$$

Observe that $f_0(r,\theta)$ is the OCSF defined in the previous section. The cross correla-

tion of $f(x,y)$ with a single component $f_m(r,\theta)$ yields

$$C_m(\alpha) = Ae^{jm\alpha}, \qquad 2.18$$

where:

$$A = 2\pi \int_0^\infty r|f_m(r)|^2 dr, \qquad 2.19$$

and $|C_m(\alpha)|^2 = |A|^2 = constant$ for any α. Thus the amplitude of the centre cross correlation value is invariant for rotations in the original image. This depends heavily on the centre of expansion, which was assumed to be the origin in the above discussion. The proper centre can be determined analytically such that the centre yields $|C_m|$ as a maximum when a match is made as discussed in the previous section (see [29]). In practice this centre cannot be found as firstly, it requires an analytic form of f and secondly, the centre depends on m, the component used, so will be different for each component. Also, this matches the test pattern to only *one* of the circular harmonic components of the target. This brings forth the question of *which* harmonic order should be used. Experiments performed by Arsenault *et al.*, find that the harmonic order to be used depends on the images being used. Also, to discriminate between similar objects, such as the letters **E** and **F**, knowledge about the objects can be used to increase recognition capabilities [1]. To discriminate between the letters **E** and **F**, the fact that E has some symmetry while F does not was used and this enabled the experimenters to eliminate all even order components from testing. Thus, this method could be useful if recognizing classes of objects: prior tests can be used to decide which harmonics should be used to help distinguish between particular classes. This does not provide a general method to decide which harmonics to use for arbitrary images.

As with the OCSF, this method claims to be shift invariant. This is true if both the image and the test pattern are expanded about their proper expansion centres.

Once again, locating the centre of expansion is non-trivial.

A modified version of the above method was employed by Wu and Stark [28]. Their technique removed the need to find the expansion centre by using a common centre for all components (the geometric centre could be used), and considered N harmonic references. Their vector signature approach used a reference vector

$$\mathbf{R}=(|R_1|,|R_2|,\ldots,|R_N|), \tag{2.20}$$

where

$$R_n=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\, f_n'(x,y)\,dx\,dy, \tag{2.21}$$

(same as Hsu et al. [17]) and the vector

$$\mathbf{C}=(|C_1|,|C_2|,\ldots,|C_N|), \tag{2.22}$$

where

$$C_n=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f_n'(x,y)\,dx\,dy, \tag{2.23}$$

for the test scene $g(x,y)$. Then by using $\mathbf{X}=\mathbf{R}-\mathbf{C}$ and taking the norm $||\mathbf{X}||=(\mathbf{X}'\mathbf{X})^{\frac{1}{2}}$ the test criterion becomes

$$\begin{aligned} ||\mathbf{X}|| &\leq T \quad \textit{target present} \\ &> T \quad \textit{target absent}. \end{aligned} \tag{2.24}$$

This approach removed the need to find the expansion centre. However, the components, which must be used to match, must be determined by tests. Experiments by Wu and Stark [28] showed improved results using more than one circular harmonic component even though the proper expansion centre was not used. Also, this vector signature approach had better results in a noisy image.

Again, the question arises of which harmonics to use. This also requires $2N$ cross correlations in addition to calculating N harmonic components. It is questionable whether this provides any improvement over using a conventional matched filter and rotating the test pattern.

2.5. Fourier-Mellin Transforms

The circular harmonic decomposition does provide rotation invariant recognition and preserves uniqueness (Equation 2.11). Using the properties of the conventional Fourier transform and geometric coordinate transforms, Casasent and Psaltis [7] and Cavanagh [9], obtain encoding of images which is invariant to translational, rotational and scale changes but forfeits uniqueness. This technique uses some properties of the Fourier and Mellin transforms. The properties used in the following discussion are presented here (for further reference see [4, 13]).

2.5.1. Properties of the Fourier and Mellin Transforms

Given an image $f(x,y)$ the *Fourier Transform* is defined by:

$$\mathbf{FT}[f(x,y)] = F(\omega_x,\omega_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)exp[-j2\pi(\omega_x x+\omega_y y)]dxdy, \qquad 2.25$$

where j is $\sqrt{-1}$ and ω_x and ω_y are the spatial frequencies measured in cycles per picture [22].

The *shift* property of the Fourier Transform is:

$$\mathbf{FT}[f(x-\alpha,y-\beta)] = F(\omega_x,\omega_y)exp[-j2\pi(\omega_x\alpha+\omega_y\beta)], \qquad 2.26$$

which gives:

$$|\mathbf{FT}[f(x,y)]|^2 = |\mathbf{FT}[f(x-\alpha,y-\beta)]|^2. \qquad 2.27$$

Thus, the Fourier Power Spectrum, $|\mathbf{FT}[f(x,y)]|^2$, is invariant to shifts in the input image. For scaling and rotation this is not true. We have,

$$\mathbf{FT}[f(\mathbf{a}x,\mathbf{a}y)] = \frac{1}{|\mathbf{a}|^2}F(\frac{\omega_x}{\mathbf{a}},\frac{\omega_y}{\mathbf{a}}). \qquad 2.28$$

Similarly, a rotation of the image $f(x,y)$ rotates the spectral magnitude $|F(\omega_x,\omega_y)|$.

Given a function, $f(x)$, the *Mellin Transform* in x is defined by:

$$\mathrm{MT}[f(x)] = M(\omega_x) = \int_0^\infty f(x)x^{-j\omega_x-1}dx. \qquad 2.29$$

The magnitude of this transform is invariant to scale changes:

$$|\mathrm{MT}[f(x)]|^2 = |\mathrm{MT}[f(ax)]|^2. \qquad 2.30$$

Obviously, by a logarithmic scaling, $x = e^u$ we have:

$$M(\omega_x) = FT[f(e^t)]. \qquad 2.31$$

2.5.2. Use of the Fourier and Mellin Transforms in Matching

Using the properties above, Psaltis and Casasent [7] and Cavanagh [9] obtain invariant coding of images through a series of transforms.

The power spectrum $|F(\omega_x,\omega_y)|^2$ is invariant to shifts in the input image $f(x,y)$; however, scaling the input image scales the power spectrum and rotations in the input image rotate the power spectrum. Therefore, by changing the power spectrum to polar coordinates, $r = \sqrt{\omega_x^2+\omega_y^2}$ and $\theta = \tan^{-1}(\frac{\omega_y}{\omega_x})$, scale and rotational changes are now each isolated to one dimension. A function denoted, $\overline{F}(r,\theta)$, is obtained which is the polar representation of the power spectrum. (Note: this is a real function since the power spectrum is real, not to be confused with the Fourier transform, $F(\omega_x,\omega_y)$).

2.5.3. Scale Invariance of the Fourier-Mellin Techniques

A scale change in $|F(\omega_x,\omega_y)|$ by a does not effect the θ coordinate of $\overline{F}(r,\theta)$. Only the r coordinate is effected by scaling. By performing a one dimensional Mellin transform in r on $\overline{F}(r,\theta)$ yields a completely scale invariant transform:

$$M(\omega_\rho,\theta) = \mathrm{MT}[\,\overline{F}(r,\theta)\,]$$

$$= \int_0^\infty \overline{F}(r,\theta)r^{-j\omega_\rho-1}dr. \qquad 2.32$$

If $M_1(\omega_\rho,\theta)=\mathrm{MT}[\bar{F}(r,\theta)]$ and $M_2(\omega_\rho,\theta)=\mathrm{MT}[\bar{F}(ar,\theta)]$ then $M_2(\omega_\rho,\theta)=a^{-j\omega_\rho}M_1(\omega_\rho,\theta)$ So $|M_2|=|M_1|$ Now, by using the change of coordinates: $\rho=\ln(r)$, $\bar{F}(r,\theta)$ becomes $\bar{F}(e^\rho,\theta)$ and

$$M(\omega_\rho,\theta) = \int_0^\infty \bar{F}(r,\theta)r^{-j\omega_\rho}\frac{dr}{r}$$

$$= \int_{-\infty}^{\infty} \bar{F}(e^\rho,\theta)e^{-j\omega\rho}d\rho \qquad 2.33$$

$$= \mathrm{FT}[\bar{F}(e^\rho,\theta)].$$

Thus the Fourier transform of the logarithmic scaling in r of $\bar{F}(r,\theta)$ yields a 1-D Mellin transform in r.

2.5.4. Rotation Invariance of the Fourier-Mellin Techniques

As with the scale invariance, a rotation of $f(x,y)$ by α rotates the power spectrum $|F(\omega_x,\omega_y)|$ thus affecting only the θ coordinate of $\bar{F}(r,\theta)$. The result of a rotation will be a shift in the transform space but the shift is *not* the same for all parts of the image. Part of the image is shifted by $-(2\pi-\alpha)$. (See figure 2.1.). The rest of the image is shifted by α. Consider the image as two parts, $f(x,y)=f_1(x,y)+f_2(x,y)$. where f_1 is rotated by α, f_2 by $2\pi-\alpha$; obtaining their power spectra followed by the polar coordinate change gives:

$$\bar{F}(r,\theta) = \bar{F_1}(r,\theta) + \bar{F_2}(r,\theta). \qquad 2.34$$

with G denoting the one dimensional Fourier transform in the θ direction:

$$G(r,\omega_\theta) = G_1(r,\omega_\theta) + G_2(r,\omega_\theta). \qquad 2.35$$

For

$$\bar{F'}(r,\theta) = \bar{F_1'}(r,\theta+\alpha) + \bar{F_2'}(r,\theta+2\pi-\alpha). \qquad 2.36$$

we get

$$G'(r,\omega_\theta) = G_1(r,\omega_\theta)e^{-j\omega_\theta\alpha} + G_2(r,\omega_\theta)e^{j\omega_\theta(2\pi-\alpha)} \qquad 2.37$$

(from the shift property of the Fourier Transform). This is the "cyclic" wrap-around property (see [23]).

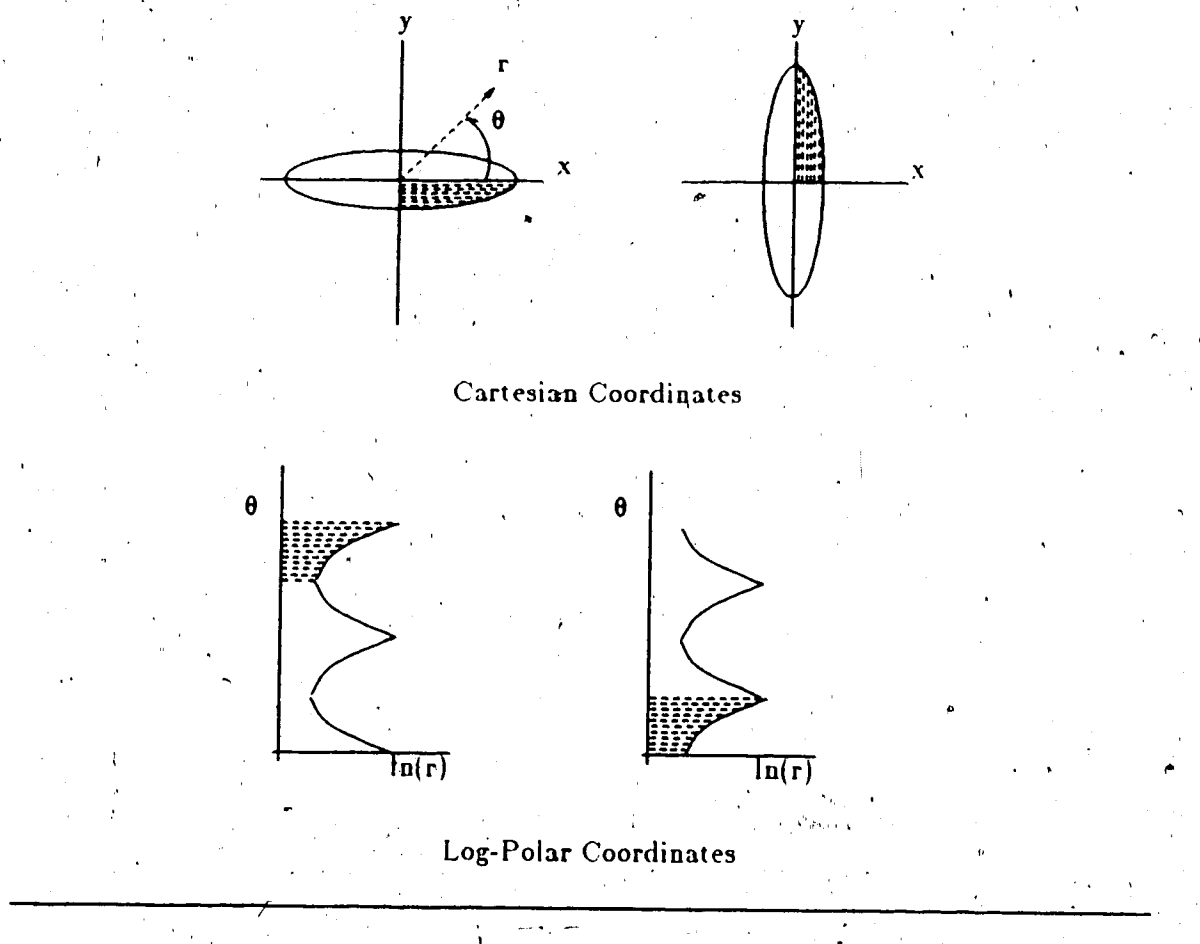


Figure 2.1 Effect of Rotation on Polar Representation

Scale invariance arises from a one-dimensional Fourier transform of $\overline{F}(e^{\rho},\theta)$ (a one-dimensional Mellin transformation of $\overline{F}(r,\theta)$) and rotation invariance arises from a one dimensional Fourier transform in θ. Equivalently this is a two dimensional Fourier transform of $\overline{F}(e^{\rho},\theta)$, resulting in both scale and rotation invariance. This will be denoted $M(\omega_\rho,\omega_{\tilde{\theta}})$. (The use of the Mellin transform notation helps to distinguish it as

a Mellin transform). Thus the Fourier-Mellin (FM) transforms are:

$$\mathbf{FT}_\theta[\mathbf{MT}_r[\bar{F}(r,\theta)]] = M(\omega_\rho,\omega_\theta) \qquad 2.38$$

$$= M_1(\omega_\rho,\omega_\theta)+M_2(\omega_\rho,\omega_\theta)$$

and

$$\mathbf{FT}_\theta[\mathbf{MT}_r[\bar{F}(ar,\theta+\alpha)]] = M'(\omega_\rho,\omega_\theta). \qquad 2.39$$

$$= M_1(\omega_\rho,\omega_\theta)e^{-j(\omega_\rho \ln a+\omega_\theta\alpha)}+M_2(\omega_\rho,\omega_\theta)e^{-j(\omega_\rho \ln a+\omega_\theta(2\pi-\alpha))}$$

The Mellin-type cross correlation, $\mathbf{FT}^{-1}[M'M']$, contains the scale factor and rotation angle information in the cross correlation peaks. There are two cross correlation peaks: the peak of $\bar{F}(e^\rho,\theta)*\bar{F}_1(e^\rho,\theta)$ is located at $\rho'=\ln a,\theta'=\alpha$ and the peak of $\bar{F}_2(e^\rho,\theta)*\bar{F}(e^\rho,\theta)$ is located at $\rho'=\ln a,\theta'=2\pi+\alpha$ (see [5,6]). The location of the peaks permits the retrieval of the scale factor, a, and the rotational angle, α. Psaltis and Casasent [7] claim that this information can be used to normalize the "test" image and then follow with a conventional matched filter. The information obtained, however, $\ln a$ and α, must first be used to scale the original power spectra and the phase spectra (which to this point has never been used). Then, the inverse Fourier transform of the spectra gives the "unrotated" and "unscaled" version of the test image. This may now be used in a conventional matched filter. Note, however, that in order to normalize the test image, in any other way, it is necessary to know about which centre the rotation and scaling was performed. This information is *not* available from the above derivation.

This method matches power spectra of images. Since the power spectrum is not unique, all images with the same power spectrum are recognized as being similar (up to translation, rotation and scaling). The loss of uniqueness makes this technique inherently subject to error in recognition.

2.6. Summary

Current techniques for rotation and scale invariant recognition have been presented. These techniques establish a basis for comparison. Analysis of these methods provides insights into possible improvements and objectives for an invariant recognition system.

Chapter 3

Log-Polar Circular Harmonic Decomposition

This Chapter explores a new integral transform. The transform represents an image in a rotation and scale invariant form. Other properties of the transform are investigated. It is shown that this transform can be expanded and, in this higher dimensional form, the representation also achieves shift invariance.

3.1. Introduction

In the previous chapter, different existing methods to obtain representation of objects invariant to rotations and scaling were explored. These methods had common features which are desirable in a recognition system. Most methods exhibited invariance in the modulus of the transform. The following has been seen:

$$|\mathbf{FT}[f(x-a,y-b)]| = |\mathbf{FT}[f(x,y)]|, \quad |\mathbf{MT}[f(ax)]| = |\mathbf{MT}[f(x)]|,$$

and from the circular harmonics, $|C_m(\alpha)| = |A|$, and from the Fourier-Mellin derivation:

$$|\mathbf{FT}\left\{LP\left[|\mathbf{FT}[f]|\right]\right\}| = |\mathbf{FT}\left\{LP\left[|\mathbf{FT}[f']|\right]\right\}|, \tag{3.1}$$

where LP represents the log-polar coordinate change, and f' is a rotated and scaled version of the function f. Thus, these mappings to the complex plane represent an image such that the magnitude is invariant to certain changes in the original image, and, in the case of the Fourier and Mellin transforms, the phase component retains enough information to uniquely represent the image. Hence, these transforms are invertible.

This gives a reference point from which it is possible to establish a set of criteria that, if satisfied, will yield an invariant transformation. This chapter accomplishes such a task. By satisfying a set of criteria, a new transform emerges which has specific

invariant properties. Additional properties of this transform are investigated, including an extension of the transform to increase its recognition capabilities to include shift invariance.

3.2. Criteria for Invariant Recognition

A set of criteria are now established that include all the desired invariant properties. As described by Braccini [3], a general integral transformation of the image function $f(x,y)$ can be defined and specific properties on the kernel of the integral can be imposed. Formally,

$$g(u,v) = \int_{-\infty}^{\infty}\int f(x,y)\omega(u,v;x,y)dxdy \qquad 3.2$$

for some weighting function $\omega(u,v;x,y)$. If $\omega(u,v;x,y)$ can be found such that the new representation of $f(x,y)$, $g(u,v)$, allows easy comparisons of properties of spatial images then this representation can be used to determine matches. Call $g(u,v)$ the response of $f(x,y)$ to $\omega(u,v;x,y)$. Define, $g_{a,\alpha}(u,v)$ as the response of $f(ar,\theta+\alpha)$ to $\omega(u,v;x,y)$. Thus, $g(u,v)=g_{1,0}(u,v)$; $g_{1,\alpha}(u,v)$ is the response to a rotation by α and $g_{a,0}(u,v)$ denotes response to a scale change by $a>0$. The criteria for rotation and scale invariance require that the responses for rotated and scaled images be expressed as:

$$(C1)\quad g_{a,0}(u,v) = \gamma(u,v)g_{1,0}(\beta u,\beta v). \qquad 3.3$$

$$(C2)\quad g_{1,\alpha}(u,v) = \xi(u,v)g_{1,0}(\eta u,\eta v). \qquad 3.4$$

for complex constants η,β and real $a>0$, with:

$$(C3)\quad |g_{1,0}(u,v)| = |g_{1,\alpha}(u,v)| = |g_{a,0}(u,v)|. \qquad 3.5$$

(C3) implies $|\gamma(u,v)|=|\xi(u,v)|=1$. Further, with (C4) $\beta=\eta=1$, otherwise the transforms map scaled images to scaled transforms [18] contrary to the goal. As a final criterion, (C5), the transform should be invertible. These criteria formally state that $f(x,y)$, a scaled version of $f(x,y)$ and a rotated version of $f(x,y)$ all uniquely map to

functions in the complex plane which have the same magnitude at any point.

3.3. Sufficient Conditions for Invariance Coding

The nature of the weighting function that satisfies the above criteria is now investigated.

First, scale invariance:

$$g_{a,0}(u,v) = \int_{-\infty}^{\infty}\int f_1(ax,ay)\omega(u,v;x,y)\,dx\,dy$$

letting $\rho = ax$ and $\tau = ay$ this yields:

$$= \int_{-\infty}^{\infty}\int f_1(\rho,\tau)\omega(u,v;\frac{\rho}{a},\frac{\tau}{a})\frac{1}{a^2}d\rho\,d\tau. \qquad 3.6$$

Using criterion (C1):

$$g_{a,0}(u,v) = \frac{1}{a^2}\int_{-\infty}^{\infty}\int f_1(\rho,\tau)\omega(u,v;\frac{\rho}{a},\frac{\tau}{a})\,d\rho\,d\tau$$

$$= \gamma(u,v)\int_{-\infty}^{\infty}\int f_1(\rho,\tau)\omega(\beta u,\beta v;\rho,\tau)\,d\rho\,d\tau$$

$$= \gamma(u,v)g_{1,0}(\beta u,\beta v), \qquad 3.7$$

which implies:

$$(*)\quad \frac{1}{a^2}\omega(u,v;\frac{\rho}{a},\frac{\tau}{a}) = \gamma(u,v)\,\omega(\beta u,\beta v;\rho,\tau). \qquad 3.8$$

For rotation invariance:

$$g_{1,\alpha}(u,v) = \int_{-\infty}^{\infty}\int f_1(x\cos\alpha - y\sin\alpha, x\sin\alpha + y\cos\alpha)\omega(u,v;x,y)\,dx\,dy \qquad 3.9$$

changing to polar coordinates, $r = \sqrt{x^2+y^2}$ and $\theta = \tan^{-1}(\frac{y}{x})$, this yields:

$$g_{1,\alpha}(u,v) = \int_0^{2\pi}\int_0^{\infty} f_1(r,\theta+\alpha)\tilde{\omega}(u,v;r,\theta)r\,dr\,d\theta, \tag{3.10}$$

letting $\phi = \theta + \alpha$:

$$= \int_{\alpha}^{2\pi+\alpha}\int_0^{\infty} f_1(r,\phi)\tilde{\omega}(u,v;r,\phi-\alpha)r\,dr\,d\phi, \tag{3.11}$$

using periodicity of f:

$$= \int_0^{2\pi}\int_0^{\infty} f_1(r,\phi)\tilde{\omega}(u,v;r,\phi-\alpha)r\,dr\,d\phi. \tag{3.12}$$

Using criterion (C2):

$$g_{1,\alpha}(u,v) = \int_0^{2\pi}\int_0^{\infty} f_1(r,\theta)\tilde{\omega}(u,v;r,\theta-\alpha)r\,dr\,d\theta$$

$$= \xi(u,v)\int_0^{2\pi}\int_0^{\infty} f_1(r,\theta)\tilde{\omega}(\eta u,\eta v;r,\theta)r\,dr\,d\theta \tag{3.13}$$

$$= \xi(u,v)g_{1,0}(\eta u,\eta v),$$

which gives,

$$(**)\quad \tilde{\omega}(u,v;r,\theta-\alpha) = \xi(u,v)\tilde{\omega}(\eta u,\eta v;r,\theta). \tag{3.14}$$

To summarize the existing conditions on ω: with criterion (4) $\eta = \beta = 1$:

$$(*)\quad \frac{1}{a^2}\,\omega\left(u,v;\frac{\rho}{a},\frac{\tau}{a}\right) = \gamma(u,v)\,\omega(\beta u,\beta v;\rho,\tau)$$

$$= \gamma(u,v)\,\omega(u,v;\rho,\tau) \tag{3.15}$$

$$(**)\quad \tilde{\omega}(u,v;r,\theta-\alpha) = \xi(u,v)\tilde{\omega}(\eta u,\eta v;r,\theta)$$

$$= \xi(u,v)\tilde{\omega}(u,v;r,\theta) \tag{3.16}$$

with (***) $|\gamma(u,v)| = |\xi(u,v)| = 1$.

There are numerous functions which may satisfy any of the above criteria. The condition (***) suggests looking at complex exponential functions of the form $\exp[j\lambda(u,v)]$, for a real function, $\lambda(u,v)$.

3.4. Proposed Transform

The functions:

$$(A) \quad \omega(u,v;x,y) = \frac{1}{x^2+y^2}\exp\left[-j\kappa u \ln\left(\sqrt{x^2+y^2}\right)\right] \qquad 3.17$$

and

$$(B) \quad \omega(u,v;x,y) = e^{-j\zeta v \tan^{-1}\left(\frac{y}{x}\right)} \qquad 3.18$$

with real constants κ,ζ, are investigated. Observe that for (A):

$$\omega\left(u,v;\frac{x}{a},\frac{y}{a}\right) = \frac{a^2}{x^2+y^2} \times e^{-j\kappa u \ln\left(\sqrt{x^2+y^2}\right)+j\kappa u \ln(a)}$$

$$= e^{-j\kappa u \ln(a)} \cdot a^2 \cdot \omega(u,v;x,y). \qquad 3.19$$

Thus, (*) is satisfied, and $\gamma(u,v) = e^{-j\kappa u \ln(a)}$ which satisfies (***). However,

$$\tilde{\omega}(u,v;r,\theta+\alpha) = \frac{1}{r^2} \times e^{-j\kappa u \ln(r)}$$

$$= \tilde{\omega}(u,v;r,\theta); \qquad 3.20$$

satisfying (**) but is not unique as $\xi(u,v) = 1$. Similarly for (B):

$$\tilde{\omega}(u,v;r,\theta+\alpha) = e^{-j\zeta v(\theta+\alpha)}$$

$$= e^{-j\zeta v\alpha}\tilde{\omega}(u,v;r,\theta). \qquad 3.21$$

Thus, (**) is satisfied, and $\xi(u,v) = e^{-j\zeta v\alpha}$ which satisfies (***). But,

$$\omega\left(u,v;\frac{x}{a},\frac{y}{a}\right) = e^{-j\zeta v\tan^{-1}\left(\frac{ay}{ax}\right)}$$

$$= e^{-j\zeta v\tan^{-1}\left(\frac{y}{x}\right)} \qquad 3.22$$

$$= \omega(u,v;x,y),$$

so (*) is not satisfied.

These functions, however, can be combined so that all conditions are satisfied:

$$\omega(u,v;x,y) = \frac{1}{x^2+y^2}\exp\left[-j\kappa u\ln\left(\sqrt{x^2+y^2}\right) - j\zeta v\tan^{-1}\left(\frac{y}{x}\right)\right],\dagger \qquad 3.23$$

or in polar coordinates:

$$\omega(u,v;r,\theta) = \frac{1}{r^2}\times e^{-j\kappa u\ln(r)-j\zeta v\theta}. \qquad 3.24$$

This function ω satisfies criteria (C1) through (C4). Then see:

$$\omega\left(u,v;\frac{x}{a},\frac{y}{a}\right) = \frac{a^2}{x^2+y^2}\exp\left[-j\kappa u\left[\ln\left(\sqrt{x^2+y^2}\right)-\ln(a)\right] - j\zeta v\tan^{-1}\left(\frac{ay}{ax}\right)\right]$$

$$= a^2 e^{j\kappa u\ln(a)}\frac{1}{x^2+y^2}\exp\left[-j\kappa u\ln\left(\sqrt{x^2+y^2}\right) - j\zeta v\tan^{-1}\left(\frac{y}{x}\right)\right] \qquad 3.25$$

$$= a^2 e^{j\kappa u\ln(a)}\,\omega(u,v;x,y),$$

and

$$\tilde{\omega}(u,v;r,\theta-\alpha) = \frac{1}{r^2}\exp\left[-j\kappa u\ln(r) - j\zeta v[\theta-\alpha]\right]$$

$$= e^{j\zeta v\alpha}\frac{1}{r^2}\exp\left[-j\kappa u\ln(r) - j\zeta v\theta\right] \qquad 3.26$$

† The $\tan^{-1}$ function is taken as the principal branch, $[0, 2\pi]$ in accord with standard practice with polar forms of transform domain coordinates and complex function values [23].

$$= e^{j\zeta v\theta}\,\omega(u,v;r,\theta).$$

Thus (*), and (**) are satisfied. Further, $\xi(u,v)=e^{j\zeta v\theta}$ and $\gamma(u,v)=e^{j\kappa u\ln(a)}$ both satisfy (***). This function also satisfies criterion(C5), invertibility, as will be shown next.

3.4.1. Invertibility of the Transform

Using the above criteria, (with a scaling factor $\frac{1}{2\pi}$), the response of the image function, $f(x;y)$ is:

$$y(u,v) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int f(x,y)\omega(u,v;x,y)\,dx\,dy \tag{3.27}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int f(x,y)\frac{1}{x^2+y^2}\exp\left[-j\kappa u\ln(\sqrt{x^2+y^2})-j\zeta v\tan^{-1}\left(\frac{y}{x}\right)\right]dx\,dy \tag{3.28}$$

$$= \frac{1}{2\pi}\int_{0}^{2\pi}\int_{0}^{\infty} f(r,\theta)\frac{1}{r^2}\times e^{-j\kappa u\ln(r)-j\zeta v\theta}\,r\,dr\,d\theta \tag{3.29}$$

$$= \frac{1}{2\pi}\int_{0}^{2\pi}\left[\int_{0}^{\infty} f(r,\theta)e^{-j\kappa u\ln(r)}\frac{dr}{r}\right]e^{-j\zeta v\theta}\,d\theta \tag{3.30}$$

$$= \frac{1}{2\pi}\int_{0}^{2\pi}\left[\int_{0}^{\infty} f(r,\theta)r^{-j\kappa u-1}\,dr\right]e^{-j\zeta v\theta}\,d\theta \tag{3.31}$$

$$= \frac{1}{2\pi}\int_{0}^{2\pi} M(\kappa u,\theta)e^{-j\zeta v\theta}\,d\theta \tag{3.32}$$

where

$$M(\kappa u,\theta)=\int_{0}^{\infty} f(r,\theta)r^{-j\kappa u-1}\,dr. \tag{3.33}$$

The quantity in square brackets in Equation 3.31 is the Mellin transform in r, $M(\kappa u,\theta)$; this transform is followed by a circular harmonic decomposition as described by Arsenault *et al.*, which is actually the calculation of the Fourier series coefficients for $M(\kappa u,\theta)$. Both the Mellin transform and the Fourier series are invertible, hence $f(x,y)$ can be recovered from $g(u,v)$.

The inverse Mellin transform is given by [4]:

$$f(r,\theta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} M(\kappa u,\theta) r^{j\kappa u}\, d(\kappa u), \qquad 3.34$$

and the inverse of the Fourier series is given by:

$$M(\kappa u,\theta) = \sum_{\zeta v=-\infty}^{\infty} g(u,v) e^{j\zeta v\theta}. \qquad 3.35$$

Thus the inversion formula for $g(u,v)$ is:

$$f(r,\theta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \left[\sum_{\zeta v=-\infty}^{\infty} g(u,v) e^{j\zeta v\theta}\right] r^{j\kappa u}\, d(\kappa u), \qquad 3.36$$

which now satisfies criterion (C5). The kernel, $\omega(u,v;x,y)$, satisfies all the specified criteria † and therefore the integral transform with ω as the kernel provides rotation and scale invariance. Using the results that Mellin transforms in r are Fourier transforms in $ln(r)$ and that the decomposition in θ is circular harmonic, the decomposition in Equation 3.27 will be called the Log Polar Circular Harmonic Decomposition, LPCHD.

† The transform exhibits all normal convergence properties of the Fourier transform (see [4] for discussion on the convergence of Fourier and Mellin transforms).

3.4.2. Shift Invariance

This is a transform mapping the image to a representation in the complex plane such that the spectral magnitude of an image and a target image are invariant to changes in either scale or orientation of the images. This transform, however, is not invariant to shifts in the input images. This technique also has a proper centre, which was arbitrarily denoted (0,0) in the above derivation. In fact, criteria C1 and C2 required rotation and scale invariance about the origin. Hence the derivation inherently imposed shift *variance*. Any point, $(x_0\ y_0)$ can be used as the centre. Following the above derivation obtain ω':

$$\omega'(u,v;x,y) = \frac{1}{x^2+y^2} \times e^{-j u \ln\left(\sqrt{x^2+y^2}\right) - j v \tan^{-1}\left[\frac{y}{x}\right]} \qquad 3.37$$

where $x = x - x_0$ and $y = y - y_0$. The centre $(x_0\ y_0)$ is again a *proper expansion centre*. The problem of obtaining invariance about any arbitrary point will be discussed later. Rotation and scale invariance can be achieved once the proper expansion centre is known, however, it cannot be claimed that this is a shift invariant transform since the problem of finding the centre has yet to be solved.

3.5. Properties of the Transform

The kernel, $\omega(u,v;x,y)$, derived in the preceding section yields an integral transform of a function, $f(x,y)$, which has the desired invariant properties and is invertible, hence it is unique for each image. Now, properties of this transform are investigated.

What does this kernel look like? The kernel of the transform, ω, is a complex function. The real and imaginary parts of the kernel are depicted in Plate 3.1, and Plate 3.2. Plate 3.1 shows the functions $RE[r^2 \cdot \omega(u,v;r,\theta)] = \cos[-u\ ln(r) - v\theta]$ and Plate 3.2 shows the functions $IM[r^2 \cdot \omega(u,v;r,\theta)] = \sin[-u\ ln(r) - v\theta]$ for u = 0,1,2,4,8,16,31 and v = 0,1,2,4,8,16,31, where u varies along the horizontal axis and v

varies along the vertical axis. The maxima and minima of the kernels are the parts of the function which are attenuated through the filter or blocked out.

The relationship:

$$g(u,v) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int f(x,y)\omega(u,v;x,y)dxdy, \qquad 3.38$$

will be denoted $f(x,y)\leftrightarrow g(u,v)$. Given the relationships; $f_1(x,y)\leftrightarrow g_1(u,v)$ and $f_2(x,y)\leftrightarrow g_2(u,v)$, Table 3.1 shows properties of this transform. Proofs of these properties are given in the appendices.

3.6. Calculation of the Transform

The relationship:

$$g(u,v) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int f(x,y)\frac{1}{x^2+y^2}e^{-j\kappa u\ln(\sqrt{x^2+y^2})-j\zeta v\tan^{-1}(\frac{y}{x})}dxdy \qquad 3.39$$

or in polar coordinates:

$$g(u,v) = \frac{1}{2\pi}\int_{0}^{2\pi}\int_{0}^{\infty} f(r,\theta)\frac{1}{r^2}e^{-j\kappa u\ln(r)-j\zeta v\theta}rdrd\theta, \qquad 3.40$$

uses real constants κ,ζ. For $\kappa=2\pi,\zeta=2\pi$ and letting $\rho=\ln(r)$ in Equation 3.40:

$$g(u,v) = \frac{1}{2\pi}\int_{0}^{2\pi}\int_{0}^{\infty} f(e^{\rho},\theta)e^{-j2\pi u\rho-j2\pi v\theta}d\rho d\theta$$

$$= \frac{1}{2\pi}\mathrm{FT}[f(e^{\rho},\theta)]. \qquad 3.41$$

hence Fourier techniques could be employed in the calculation.

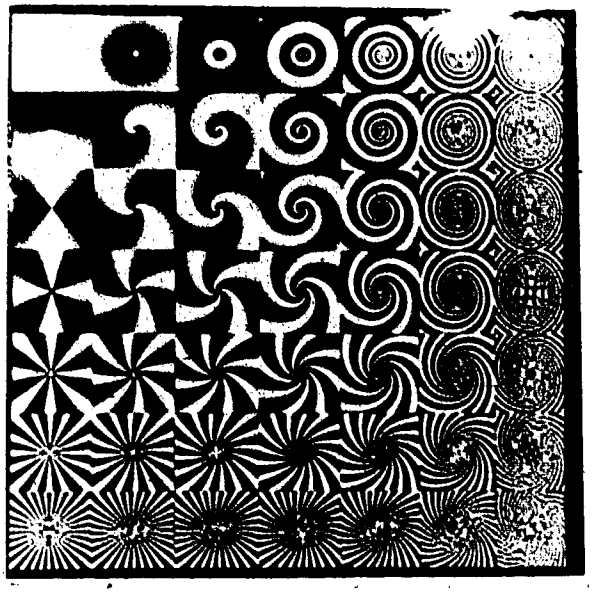


Plate 3.1 Real parts of the basis functions

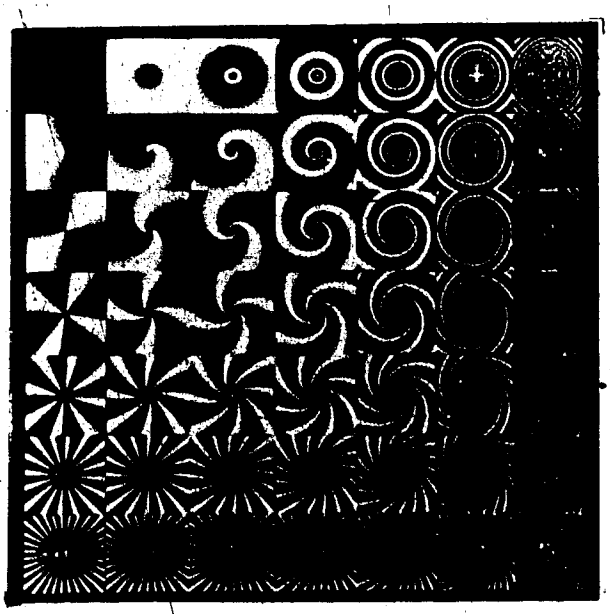


Plate 3.2 Imaginary part of the basis functions

Table 3.1 Properties of Proposed Transform

	$f(x,y)$	$g(u,v)=\frac{1}{2\pi}\int\int f\omega\, dxdy$
Linearity †	$cf_1(x,y)+bf_2(x,y)$	$cg_1(u,v)+bg_2(u,v)$
Scaling †	$\tilde{f}(ar,\theta)$	$e^{j\kappa \ln a}g(u,v)$
Rotation †	$\tilde{f}(r,\theta+\alpha)$	$e^{j\zeta\alpha}g(u,v)$
Differentiation	$-j\kappa\, ln(r)\tilde{f}(r,\theta)$	$\frac{\partial}{\partial u}g(u,v)$
	$-j\zeta\theta\tilde{f}(r,\theta)$	$\frac{\partial}{\partial v}g(u,v)$
	$-\kappa\zeta\theta\, ln(r)\tilde{f}(r,\theta)$	$\frac{\partial^2}{\partial v\partial u}g(u,v)$
Modulation †	$\tilde{f}(r,\theta)e^{j\lambda\kappa lnr+j\mu\zeta\theta}$	$\tilde{g}(u-\lambda,v-\mu)$
Conjugation	$\tilde{f}^*(r,\theta)$	$[g(-u,-v)]^*$
Real and	$RE[\tilde{f}(r,\theta)]$	$\frac{1}{2}[g(u,v)+g(-u,-v)^*]$
Imaginary Parts	$IM[\tilde{f}(r,\theta)]$	$\frac{1}{2}[g(u,v)-g(-u,-v)^*]$

† for complex c,b and real a,α,λ,μ

This will not, however, yield the same results as direct computation. The problem lies in the conversion to the log-polar coordinate system. Points close to the origin are sampled more than points farther from the origin (see Figure 3.1). This higher resolution at the origin places greater weight on these values. This makes the computation much more sensitive to the centre of expansion and a slight deviation of the centre (as little as 4 pixels in a 64 by 64 image) greatly reduces the matching abilities. Sampling in log-polar coordinates is discussed in [10,27].

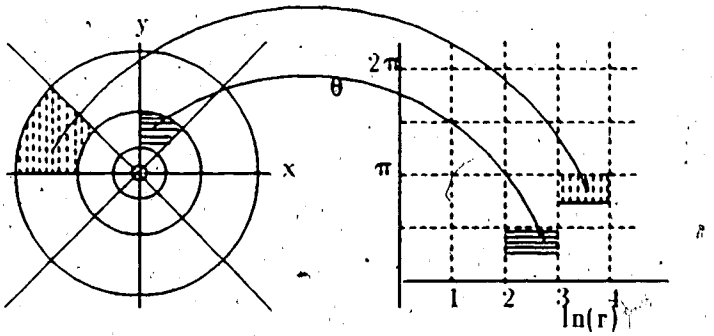


Figure 3.1 Log-Polar Coordinate Transform

3.7. Commutativity of Transformations

Although computing the log-polar representation of the powerspectrum and computing the powerspectrum of the log-polar representation of the image both yield rotation and scale invariant properties, these transformations do not commute. In general

$$|\mathbf{FT}[LP(Image)]|^2 \neq LP(|\mathbf{FT}[Image]|^2). \qquad 3.42$$

If the logpolar transform is performed using the proper centre (i.e. the centre of rotation or scale change) then rotations and scale changes in the original image become

shifts in the log-polar representation of the image. The magnitude of the Fourier transform of this log-polar image is invariant to shifts in the input image, hence the powerspectrum of the log-polar image is the same for an image, and a rotated, scaled version of the image. The proper centre, however, **must** be used or this does not hold.

No proper centre is required for converting to log-polar coordinates in the Fourier domain as the powerspectrum is centred about the origin. Rotations and scalings in the image domain still manifest themselves as rotations and scalings in the Fourier domain. Taking the log-polar representation of the power spectrum now converts the rotations and scalings to shifts. Hence this representation is **not** the same for an image and a rotated, scaled version of the image. To eliminate the effects of the shifts, an additional Fourier or other shift-invariant transform is required. Also, although no proper centre is required, the position information of the original image was contained in the phase component of the Fourier transform which was discarded at the first step. All uniqueness is lost as $|\mathbf{FT}[f(x,y)]|^2$ does not contain enough information to reconstruct f unless the input image has zero phase (which means that the input image must be circular symmetric which, in general, is not true).

3.8. Lack of Shift Invariance with LPCHD in 2-D

The proposed transform lacks shift invariance. Recall $g_{a,\alpha}(u,v)$ is defined as the response of $f(ar,\theta+\alpha)$ to $\omega(u,v;x,y)$. The original derivation was based on the criteria:

$$(C1)\quad g_{a,0}(u,v) = \gamma(u,v)g_{1,0}(\beta u,\beta v) \qquad 3.43$$

$$(C2)\quad g_{1,\alpha}(u,v) = \xi(u,v)g_{1,0}(\eta u,\eta v) \qquad 3.44$$

for complex constants η,β, with:

$$(C3)\quad |g_{1,0}(u,v)| = |g_{1,\alpha}(u,v)| = |g_{a,0}(u,v)| \qquad 3.45$$

or,

$$|\gamma(u,v)| = |\xi(u,v)| = 1 \qquad 3.16$$

and, (C4) $\beta = \eta = 1$, (C5) the transform should be invertible. Observe that conditions (C1) and (C2) require rotation and scale invariance about the origin. Hence, the derivation implicitly *imposed* the lack of shift invariance. One cannot discuss rotations and scaling without specifying a "centre" for these geometric transformations. Now consider the centre as part of the calculation. Chosing an arbitrary point (x_c, y_c) as the centre of rotation and scaling then the same result will be derived with x and y replaced by $x - x_c$ and $y - y_c$, as described in Section 3.3.2. This still lacks shift invariance.

To gain shift invariance, the original criterion must be extended to include an arbitrary point as the centre. Define, $g_{a,\alpha,x_c,y_c}(u,v)$ as the response of $f(ar, \theta+\alpha)$ to $\omega(u,v;x,y)$, where $r = \sqrt{(x-x_c)^2+(y-y_c)^2}$ and $\theta = \tan^{-1}\left[\frac{y-y_c}{x-x_c}\right]$. Now, $g_{a,0,x_c,y_c}(u,v)$ denotes response to a scale change by a ; $g_{1,\alpha,x_c,y_c}(u,v)$ denotes response to a rotation by α where both these changes are centred at any point (x_c, y_c). Now, the criteria become:

$$(C1') \quad g_{a,0,x_1,y_1}(u,v) = \gamma(u,v) g_{1,0,x_c,y_c}(\beta u, \beta v) \qquad 3.17$$

$$(C2') \quad g_{1,\alpha,x_2,y_2}(u,v) = \xi(u,v) g_{1,0,x_c,y_c}(\eta u, \eta v) \qquad 3.18$$

for complex constants η, β, with:

$$(C3') \quad |g_{1,0,x_c,y_c}(u,v)| = |g_{a,0,x_1,y_1}(u,v)| = |g_{1,\alpha,x_2,y_2}(u,v)| \qquad 3.19$$

with $|\gamma(u,v)| = |\xi(u,v)| = 1$ and (C4') $\beta = \eta = 1$, (C5') the transform should be invertible.

This new criterion would allow one to have rotations and scalings about any arbitrary point. Also, note that $g_{1,0,x_1,y_1}(u,v)$ and $g_{1,0,x_2,y_2}(u,v)$ represent rotation by 0°

and scaling by a factor of 1 where these "identity" operations are performed about different points, thus criterion (C3') also specifies shift invariance. It is not clear, however, that invertibility is possible with these conditions. Since γ and ξ do not depend on the centre, and condition (C3') implies that the magnitude is invariant to shift, rotation and scale changes about any point, therefore, the argument, $Arg\left[g_{a,\alpha,x_c,y_c}(u;v)\right]$, must contain enough information to uniquely represent the four parameters a,α,x_c and y_c. Since γ and ξ are only functions of two variables u and v, the above four parameters could not uniquely be represented by these functions. Thus the criterion (C5') could not be satisfied, given (C1'), (C2'), and (C3').

Consider the centre as a parameter in the proposed method and allow ω,γ and ξ to be functions of x_c,y_c:

$$\omega(u,v;x_c,y_c;x,y)=\frac{e^{-j\kappa u\ln\left(\sqrt{(x-x_c)^2+(y-y_c)^2}\right)-jv\tan^{-1}\left[\frac{y-y_c}{x-x_c}\right]}}{(x-x_c)^2+(y-y_c)^2} \qquad 3.50$$

then the response of $f(x,y)$ to $\omega(u,v;x_c,y_c;x,y)$ is:

$$g(u,v,x_c,y_c)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\frac{e^{-j\kappa u\ln\left(\sqrt{(x-x_c)^2+(y-y_c)^2}\right)-jv\tan^{-1}\left[\frac{y-y_c}{x-x_c}\right]}}{(x-x_c)^2+(y-y_c)^2}dxdy. \qquad 3.51$$

So if $|g_1(u,v,0,0)|^2 = |g_2(u,v,x_0,y_0)|^2$ then:

$$f_1(x,y)=f_2(a(x-x_0)\cos\alpha-a(y-y_0)\sin\alpha,(a(x-x_0)\sin\alpha+a(y-y_0)\cos\alpha) \qquad 3.52$$

for some $a>0,\alpha\in R$. This would solve the problem of rotation, scale and shift invariant encoding of images, but this solution to rotation, scale and shift invariant recognition does not lie in 2 dimensions. It appears that uniqueness can only be guaranteed when higher dimensions are considered. Once the proper expansion centre has been found then rotation and scale changes can be dealt with using methods described earlier in this chapter. Finding the proper expansion centre, however, is a large part of

the problem of invariance coding.

3.9. Summary

A new integral transform with inherent rotational and scale invariant properties has been introduced. This transform was derived using criteria which were established in reviewing existing techniques. The resulting transform, as it turns out, can be viewed as an extension of these existing techniques. This transform exhibits properties analogous to the Fourier transform and can in fact be viewed as a Fourier transform in the log-polar coordinate space. This transform is invertible and hence provides a unique representation for each image. The transform is also extendible to include shift-invariance; however, to do so requires forfeiting uniqueness or considering higher dimensions. Thus a theoretical solution to the problem of finding a representation of an image which is invariant to rotational and scaling changes has been provided.

Chapter 4

Experiments

In the previous chapters, techniques for invariant recognition were introduced. This chapter will examine the performance of those methods for image recognition. Each method is compared using different criteria:

- Is the method easily and quickly computable?
- What information can be obtained from the image transform: Is the method invertible?
- How well does this method work in recognizing objects?
- Are invariant forms of images obtained?
- What limitations does this method have in pattern recognition?

This chapter will consider each of these questions in turn. The first section looks at the recognition capabilities: determining how well the transforms work in recognizing objects, discriminating between similar objects, and recognizing images in the presence of noise. The first section will also consider the question of whether or not the method obtains a truly invariant form of the image. As previously stated, an invariant form of the image enables an ease of comparisons of different properties of the test images. The form obtained should not require excessive computation to determine if a match has been found. The next section will discuss the image information contained in the transform space, in particular, whether the transform retains enough image information to reconstruct the original image. The next section will consider the time and space requirements for computing each transform. The last section discusses the uses and limitations of each method in pattern recognition based on the results of these experiments.

4.1. Recognition Capabilities

How to determine how well a matching technique works? Optimally, a recognition system that will successfully match occurrences of an object, regardless of position, size or orientation, and "mismatch" with other objects is desired. Various tests have been performed to determine how closely the techniques under review meet this criteria. First each method will be tested to confirm that the object matches successfully with scaled and rotated versions of itself. Next, the ability to discriminate between the object and dissimilar objects is tested. The third tests investigate recognition between the object and objects similar to the reference object, such as the letter H compared with N, E compared with F, and C compared with G and O. Finally the transforms are tested in the presence of noise. These tests should indicate how well a technique works.

In order to provide a measure of how well each transform performs, the matching criterion used must be considered. For the vector signature method, a threshold value, T, must be found, which maximizes the number of matches while minimizing the number of false alarms. Both the Fourier-Mellin technique and the techniques proposed in Chapter 3 are magnitude-invariant to rotations and scalings in the spatial domain. Thus the normalized cross correlation (see Section 1.4), N_{fg}, would indicate how close the spectrum match. The cross correlation operation considers whether the functions match in any position and the only required value is that of the origin, which indicates whether the images match in the current position. Also, the correlation techniques will yield a peak if $f = cg$ for any scalar c. Since the techniques are magnitude invariant, a required match occurs for $c = 1$ only. Therefore, a mean-square difference between the magnitude spectra will be used as an additional measure of match between the magnitude spectra. Define $M_{T(f)}(u,v)$ as the magnitude spectra for the transform of f, then the mean square difference between the transform of f and that of g is:

$$\delta = \frac{\sqrt{\sum_u \sum_v \left[M_{T(f)}(u,v) - M_{T(g)}(u,v) \right]^2}}{UV} \qquad 4.1$$

This would give a measure of how "close" the mappings of the two images are and therefore a measure of "similarity" between the image, f, and the test pattern, g.

4.2. Results

Tests were performed on various simple images such as geometric figures: circles, squares, rectangles, a Maltese-type cross, 2 parallel bars; and the alphabetic characters: H, N, E, F, O, Z, C, G, D, I, L, T. The images were tested at various orientations, scale sizes, and positions. The results of the tests are summarized below. These tests determine whether the transforms could invariantly recognize an object under scale and rotation changes and whether there was a distinction between an object and different objects. Tests were run to determine the ability of each method in discriminating between similar objects such as parallel bars, ||, and the letter H; the letters E and F; the letter O and C; and other similarities. Further tests investigate the performance in the presense of noise. Two types of "noise" were added to the image: (1) Gaussian white noise with zero mean, and variances 10, 20, 30, and 40; and (2) presence of more than one object in the scene. Then each method was tested to determine the sensitivity to small displacements of the centre. The results for each method are now considered.

4.2.1. Circular Harmonic Techniques

The vector signature method performed well. Images and rotated versions of the image could be correctly recognized and comparisons of an image with a different object can be "mismatched" for correctly chosen threshold value. Experiments had to be performed to determine a "good" threshold value. False alarms were infrequent for correctly chosen components and threshold values. This method however, missed objects in the presence of noise and was very sensitive to the centre. Table 4.1 shows some typical results using this technique. Plate 4.1 shows images and their circular harmonic components.

The work of Hsu, Arsenault *et al.*, [17] suggests that recognition is possible using only one circular harmonic component. If the proper expansion centre is found and the *correct* component is chosen, then this may be so; however, both finding the centre and chosing the correct component are difficult tasks to perform in the general case when no assumptions are made about the image. The vector signature approach proposed by Wu and Stark [28], gives improved results in noise and eliminates the problem of finding the proper expansion centre, hence this approach was thoroughly tested. The circular harmonic cross correlation technique depends strongly on the components and threshold value used. The choice of components could be determined by tests; however, this requires knowledge about the image. To distinguish an E from a dissimilar object such as a C is possible with various components and a high threshold value. To distinguish an E from an F is improved if only even components are used. Using the notation in Section 2.3 and Table 4.1, where E is the reference image and F is the test pattern, the value of $||X \cdot X||^2$ is 1352 using all even components $m < 16$, and 1085 using all odd components $m < 16$. In contrast, to distinguish an I from a pair of parallel bars: ||, recognition is improved if only odd components are used. Using the I as the reference image and the bars as the test pattern, the value of $||X \cdot X||^2$ is 372 using all even components $m < 16$, and 721 using all odd components $m < 16$. In addition to

determining which coefficients, there is also the problem in determining a threshold value, T. Using all components, $m<10$, a threshold value of $T=1000$ will correctly identify all occurrences of an E at various rotations including those with white noise added, and will correctly eliminate all occurrences of an F. This value of T will not identify a square and a rotated square. Hence there is no general rule that can be used to determine which components to use and an appropriate value of T to use with those components.

Table 4.1 Vector Signature Results

Image 1	Image 2	α†	$\|\|X-X\|\|^2$ using components, m:			
			0-15	even	0-5	odd
E	E	45	291.3	240.6	142.7	164.4
E	E	60	407.0	287.0	106.3	288.9
E	E	90	0.0	0.0	0.0	0.0
E	E	180	0.0	0.0	0.0	0.0
E	F	0	1531.6	1352.5	1408.0	1085.8
Square	Square	45	2965.0	2118.4	3108.6	2677.3
Bars	Bars	45	644.2	637.1	340.6	128.5
Bars	I	0	788.1	247.6	564.3	754.3
F	F	45	391.5	103.5	366.3	377.5
Square	Rectangle	0	3082.6	2189.8	2879.9	2820.7

† α is the rotation angle, in degrees, between the images.

Plate 4.1 shows the first through sixth circular harmonic components of the test image, and the cross correlation of the reference image with these components.

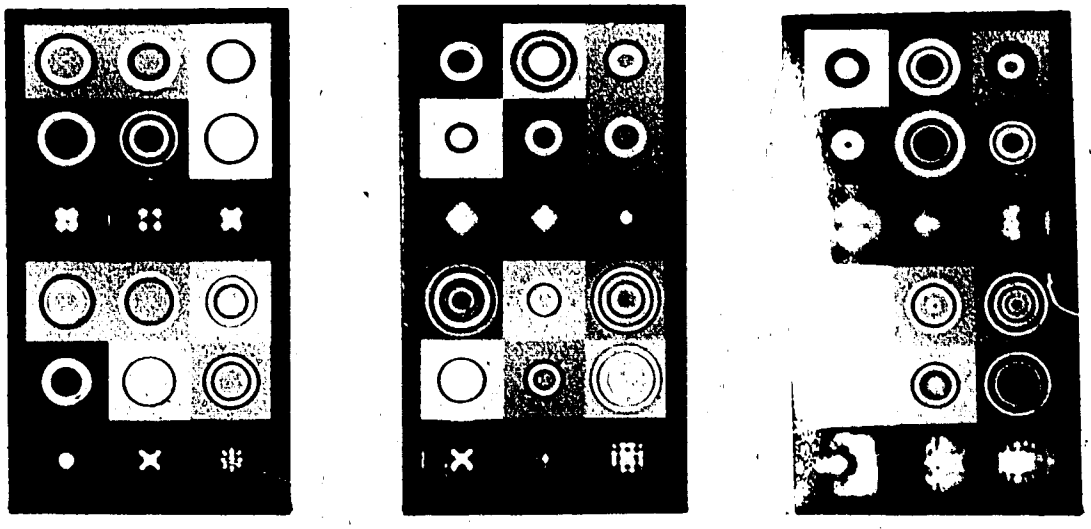


Plate 4.1 Circular Harmonic Components

ROW:
1/2: real/imaginary part of circular harmonic components 1,2 and 3
3: cross correlation of circular harmonic components with image file
4/5: real/imaginary part of circular harmonic components 4,5 and 6
6: cross correlation of circular harmonic components with image file

The real and imaginary parts of the components are shown since these are complex functions, and the magnitude of the cross correlation is shown. In the first image, a square is used as a reference image and a square rotated 45° is used as the test image, hence this is the crosscorrelation of an image and its own circular harmonic components. In the middle image a square is used as a reference image and a parallel bars, ||, are used as the test image. In the right picture, a letter E is used as a reference image and a letter E rotated 60° is used as the test image.

4.2.2. Fourier-Mellin Techniques

The Fourier-Mellin method recognized objects at various rotations and scale sizes. This method had the highest frequency of false alarms. For example, on 50 tests using the letter E compared to other letters and objects, and using a 90% criterion for the cross correlation peaks, 7 objects were missed and 7 false alarms occurred. Typically, the objects missed involved comparisons between objects with a large difference in the relative scale (4×) and noisy images. The false alarms occurred between objects that had less than a 90% correlation when conventional cross correlation techniques were applied. False alarms occurred between the letters F and N, E and H, and E and F; and between the geometric figures of a square and rectangle, and the letter O compared to the square. However, this method showed no sensitivity to the centre.

Although the Fourier-Mellin techniques are not sensitive to the centre, scaling problems occur in the digital implementation. The logpolar spectra of the power spectra of the images will differ by shifts, if the original images differ by rotations or scale changes. This difference can be eliminated by the final Fourier transform used; however, this transform does not account for the variation in the energy in the images. The logpolar images have variations due to the scaling of the power spectrum. Referring to Plate 4.3, as the images are scaled down, the power spectrum grows (see Equation 2.29) and hence the energy in the power spectrum is greater. Cavanagh [9] men-

tioned this briefly in a description of the computation of this method, and suggested converting to zero-mean amplitude. Converting to a fixed mean, however, does not overcome the scaling problem. The resulting Fourier-Mellin transform cannot be measured with δ_s (Equation 4.1), without some form of normalization. Using a cross correlation of spectra will work, but as mentioned earlier in the chapter, this extra computation should not be necessary, as a truly invariant form of the image should be comparable pixel by pixel.

Further problems arise due to the aperture effect [4,15]. If the logpolar image is too large then the power spectra of the logpolar image is "interesting" (i.e., non-zero) close to the origin. In a digital computation, values of points between the pixels can only be interpolated and hence once mapped to the frequency domain, all values of the power spectra are mapped to a few pixels. If the logpolar image is compressed to overcome these aperture effects (as suggested by [9]) the information is lost in the compression. This trade-off is "tricky" to balance. Further problems are encountered when deriving the values of the relative scaling a and the relative orientation α between two images. The peaks should occur at $(\ln a, 2\pi + \alpha)$ and $(\ln a, \alpha)$. In the process of scaling, the scale changes performed must be kept track of in order to determine where the peaks lie. For small values of a, such as $a < 1$, the exact value of a was not derivable without using a small sampling rate for r. The overhead here is in (1) determining how to scale the logpolar image of the power spectrum (i.e., the sampling rate in r) and (2) maintaining a record of these changes in order to locate the cross-correlation peaks and (3) performing the scalings (each is $O(n^2)$ operation). If the required sampling rate in r is high, the size of the image must be increased.

4.2.3. Fourier Transforms of Log-polar Images

The Fourier transform of the log-polar image provides accurate matching. The method performs well in discriminating between different objects. False alarms occur with similar objects such as the letters E and F, H and ||. The Fourier transform of the logpolar image has varying results in the presence of noise. For more than one object in the image, the results are poor, and for images with Gaussian noise added the results varied depending on the image. For some images, such as the letter E, noise has less effect than with the image of a square (See Table 4.2). This method, however, is very sensitive to the centre. For example, moving the centre of one image up 2 pixels and right 2 pixels caused the auto correlation peak to drop to 64%.

As with the Fourier-Mellin techniques, this method can not be measured with δ without some form of normalization performed on the log-polar images. Cross correlation of log polar images could be used, but again, this is not a truly invariant *representation* of the image. Plate 4.4 shows images with their log-polar representation and the cross correlation of the log-polar images.

4.2.4. Log-Polar Circular Harmonic Decomposition

The direct computation of the method proposed in Chapter 3 identified rotated and scaled versions of the same image and successfully eliminated objects differing from the image. False alarms occurred with similar images. This method performed well in noise but is sensitive to the centre.

This method mapped images into a form such that their spectra could be compared without any normalization or scaling. As seen in Table 4.2, the spectra of rotated or scaled images did not vary greatly.

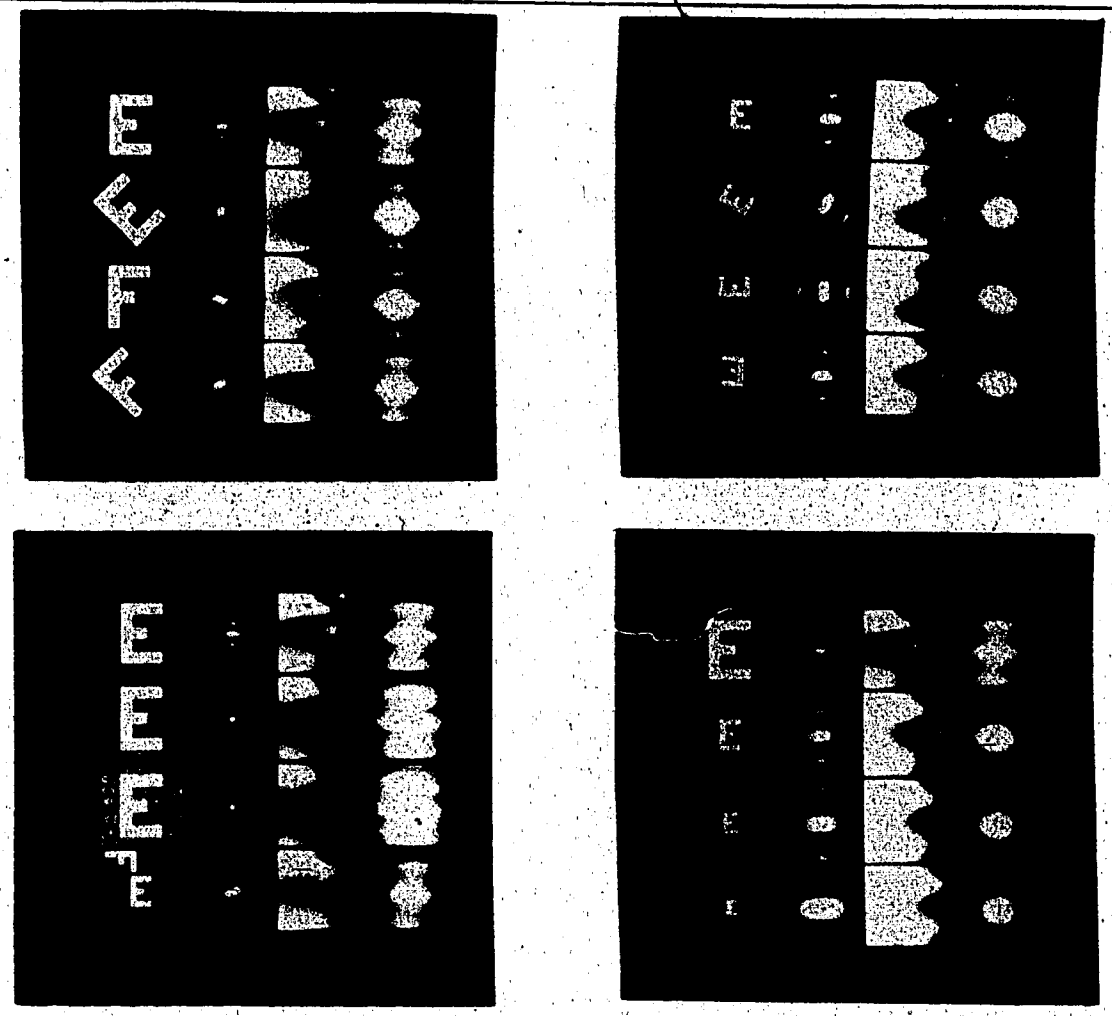


Plate 4.2 Fourier-Mellin Transforms

Each row contains: (a) the image; (b) the power spectrum of the image (c) the log-polar coordinate representation of the power spectrum; and (d) the Fourier power spectrum of the log-polar image (which is the Fourier-Mellin transform of the original image). For the logpolar images, the horizontal axis is $\ln(r)$ and the vertical axis is θ.

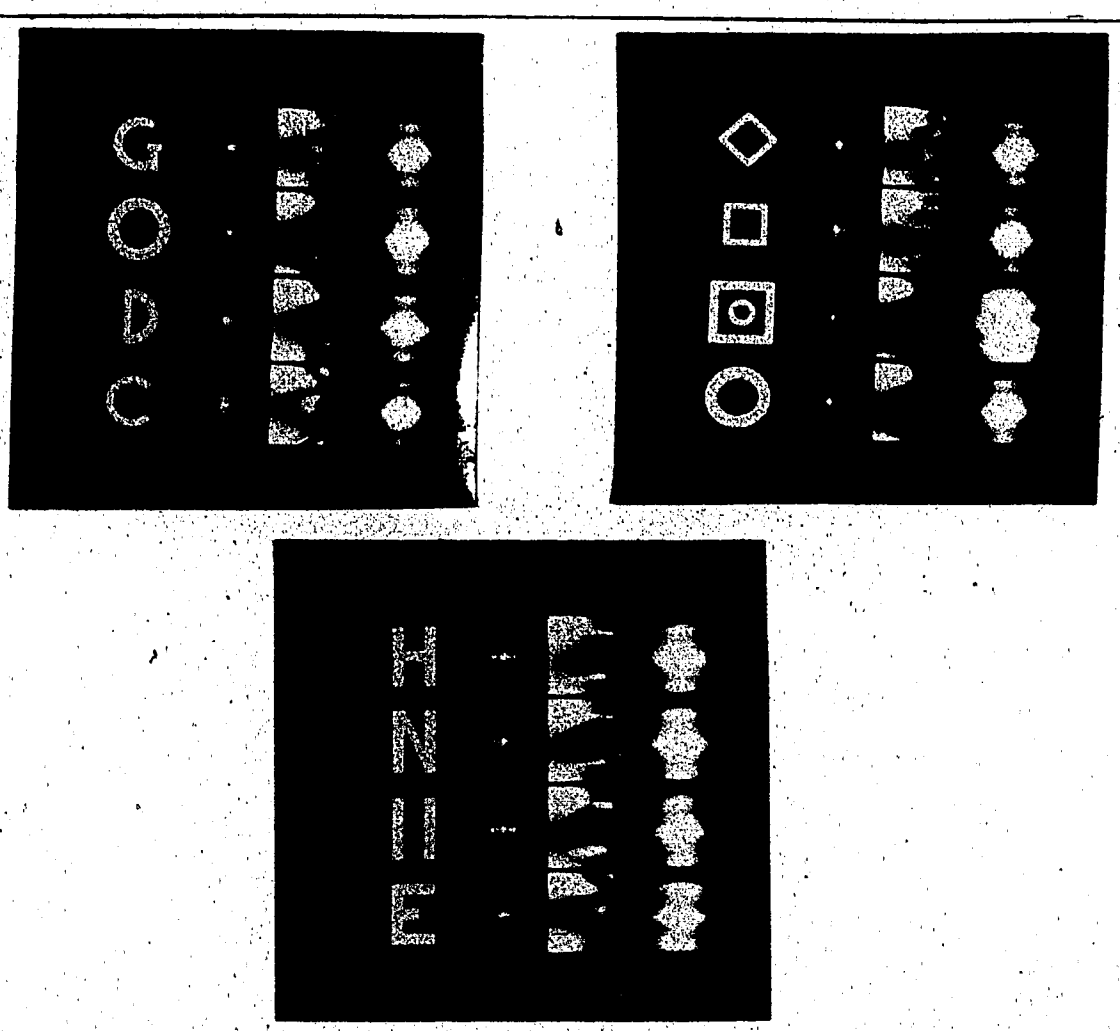


Plate 4.3 Fourier-Mellin Transforms

Each row contains: (a) the image; (b) the power spectrum of the image (c) the log-polar coordinate representation of the power spectrum; and (d) the Fourier power spectrum of the log-polar image, which is the Fourier-Mellin transform of the original image.

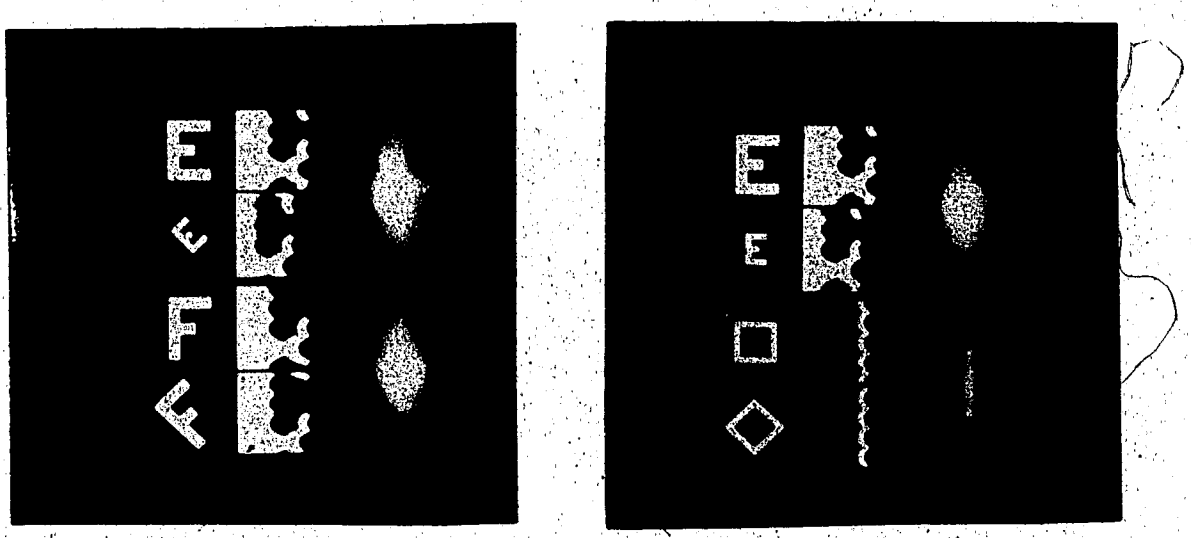


Plate 4.4 Crosscorrelation of Log-Polar Transformed Images

Logpolar transform of each image followed by a conventional crosscorrelation. The images were centred about the geometric centre which was used as the centre for the log polar coordinate transformation which has ln(r) as the horizontal axis and the vertical axis is θ.

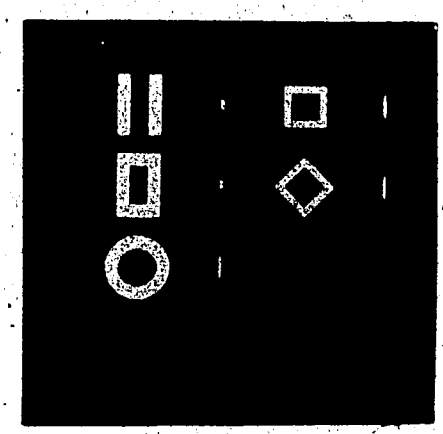


Plate 4.5 Spectral Magnitude using the Proposed Method

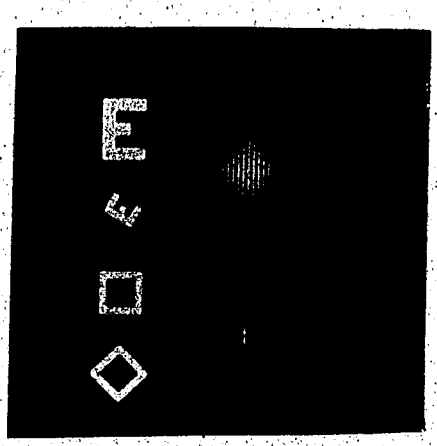


Plate 4.6 Cross Correlation of Spectral Magnitude

The spectral magnitude of each image is obtained using the transform proposed in Chapter 3. These spectra were then used as input to a conventional crosscorrelation.

Table 4.2 Comparison of Fourier-Mellin and Proposed Transform

Image 1	Image 2	Relative Scale	Relative Orientation	Difference in Spectra δ	
		a	α^o	FM †	LPCHD
E	E	0.5	0	274.2	3.8
E	E	0.33	0	399.0	4.1
E	E	1.0	45	74.9	1.0
E	E	0.5	45	226.5	3.6
E	E	1.0	60	90.9	0.9
E	E	0.5	60	217.1	3.6
E	E	1.0	90	99.4	0.0
E	E	0.5	90	236.3	3.8
E	E	1.0	180	0.0	0.0
E	F	1.0	0	107.4	0.1
E	F	0.5	0	176.6	3.8
E	rectangle	1.0	0	43.7	35.2
E	bars	1.0	0	49.2	35.2
square	square	1.0	45	21.2	0.5
rectangle	rectangle	1.0	45	66.9	1.5
rectangle	bars	1.0	0	43.1	2.9
rectangle	square	1.0	0	90.3	2.1

† This uses the normalized FM

Table 4.3 Effect of Noise on Recognition

I1	I2	Noise	LPCH	LPF	FM	Vector Signature Methods		
		σ^2	δ	R	δ	$\|\|X-X\|\|^2$ using components:		
						0-15	0-5	10-15
E	E	F	21.7471	0.7930	32.913	2319.4883	1093.4539	2436.8249
E	E	20	14.3620	0.9823	33.866	600.0417	473.7680	112.5997
E	E45	20	15.8082	0.8685	51.867	816.1863	681.7677	122.4457
E	F	20	14.4541	0.9702	43.808	1068.7153	1055.8096	1311.5840
E	E	40	37.0293	0.9365	50.950	807.1304	118.1128	281.6892
E	E45	40	36.3115	0.8449	57.020	738.8744	337.3830	354.6499
O	O	□	10.8949	0.7391	75.626	3782.9392	1912.8863	3443.0802
□	□	O	14.1743	0.7227	146.278	4209.1541	2346.7518	4513.9754
□	□15	20	22.7115	0.5341	130.588	3547.7596	2305.9012	3909.5543
□	□15	40	34.8841	0.6427	148.951	3266.2260	1642.8628	3512.9005

I = image; $\delta = \sqrt{\sum (A_i - B_i)^2}$ Δ displacement of centre $R = \frac{N_{fg}}{\sqrt{\int\int f^2}}$

4.3. Information Contained in Transform Space

This section will consider what information is obtained in the transform space. The first question considered is whether the transform contains the information to reconstruct the original image. The second section will look at other information contained in the transform space.

4.3.1. Uniqueness

The invertibility of a transform is crucial if the transform is to be used in pattern recognition. A transform lacking uniqueness cannot give accurate results since more than one image in the spatial domain will map to the same function in the transform space. If the transform space contains enough information to reconstruct the original image, then the transform space representation can be treated as an *alternate* form of the image. The Fourier transform is such a transform. In the discrete case, this transform can represent exactly an $n \times n$ image. Operations can be performed in either the spatial (image) domain or the frequency (Fourier) domain. Either form of the image is an equivalent representation since one form can be converted to the other without loss of information. This type of uniqueness is desirable in a transform. Now consider each of the transforms reviewed in Chapters 2 and 3.

Circular harmonic matching techniques are based on circular harmonic reconstruction techniques where an image $f(r,\theta)$ is represented as:

$$f(r,\theta) = \sum_{m=-\infty}^{\infty} f_m(r) e^{jm\theta}. \qquad 4.2$$

where

$$f_m(r) = \frac{1}{2\pi} \int_0^{2\pi} f(r,\theta) e^{-jm\theta} d\theta. \qquad 4.3$$

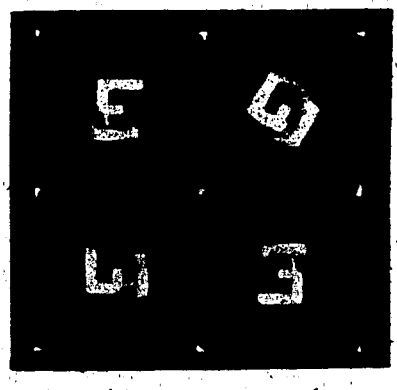
This transform is invertible. It should be noted that since this decomposition is a Fourier series in θ for each value of r, the number of coefficients $f_m(r)$ that are

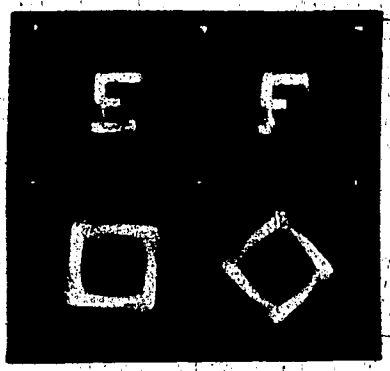
required to reconstruct an image may exceed the space requirements to store the original image. Reconstruction tests show that for circular symmetric images as few as 8 coefficients will be sufficient to reconstruct the images; for non symmetric images, the image cannot be accurately reconstructed with 128 coefficients. Hence it is possible to map to a domain in which all image information is represented, however the size of the alternate representation is large.

Fourier-Mellin techniques are not invertible. Since only the power spectrum of the image is used, there is not sufficient information to reconstruct the image. The phase information was "discarded" at the first step. It may be argued that phase information is still available and thus the image is recoverable. However, in the final Mellin representation, the only way to rebuild the image is to reverse each step. Since the intermediate steps involve scaling, to reduce aperture effects, and normalization, to equalize grey levels, each step can only be reversed if one keeps track of the parameters used in each of these steps. The transformed image itself does not contain all the necessary image information.

The log-polar Fourier method used to calculate the transform in Chapter 3 is invertible since each step (logpolar coordinate transform and Fourier transform) is invertible.

The method proposed in Chapter 3 is invertible with the inverse transform given in Section 3.4.1. As with the circular harmonic decomposition, the use of a Fourier series may require many coefficients to rebuild the original image. Plate 4.7 shows the synthesis of images from the transform space. This synthesis made use of the conjugacy properties of the transform (since the original image is real), in order to use fewer coefficients. The top row uses 64 values of $g(u,v)$ and 64 values of $g(u,-v)$ to reconstruct the images. The bottom row uses 16 values of $g(u,v)$ and 16 values of $g(u,-v)$ to reconstruct the images. The images in the lower row of Plate 4.7 are very "similar"





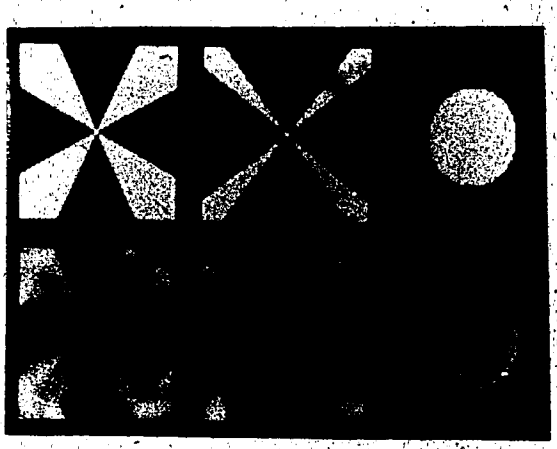


Plate 4.7 Image Synthesis Using Proposed Method

to the basis functions and hence they can be reconstructed with fewer coefficients.

4.3.2. Other Information

In addition to containing information to reconstruct the image, one would like to extract additional information, such as the relative orientation and size between two objects. As seen in Section 1.4, the cross correlation function, $C_{fg}(u,v)$ yields a peak at (u,v) if and only if $f(x,y) \mp cg(x+u,y+v)$ for scalar c. Thus the relative displacement between two images can be derived from the location of the cross correlation peak. Consider similar properties for each of the transforms and check the results from tests to see how accurate the information is.

In the circular harmonic methods, the centre correlation value between an image $f(r,\theta)$ and one circular harmonic component of $f(r,\theta+\alpha)$ is given by:

$$C_m(\alpha) = Ae^{jm\alpha} \quad \text{where} \quad A = 2\pi \int_0^\infty r\,|f_m(r)|^2\,dr. \tag{4.4}$$

For any α, $|C_m(\alpha)|^2 = |A|^2 = constant$. Thus using any m, the value of α can be extracted from Equation 4.4. Given two centre values:

$$C_m(\alpha_1) = C_m(0)e^{jm\alpha_1} \quad \text{and} \quad C_m(\alpha_2) = C_m(0)e^{jm\alpha_2}, \tag{4.5}$$

solve the right hand side of the equation above for $C_m(0)$ and substitute it into the left hand equation:

$$C_m(\alpha_1) = \left[C_m(\alpha_2)e^{-jm\alpha_2}\right]e^{jm\alpha_1} = C_m(\alpha_2)e^{jm(\alpha_1-\alpha_2)}. \tag{4.6}$$

Rearranging this:

$$e^{jm(\alpha_1-\alpha_2)} = \frac{C_m(\alpha_1)C_m^*(\alpha_2)}{|C_m(\alpha_2)|^2}, \tag{4.7}$$

which yields:

$$m(\alpha_1-\alpha_2) = \arg\left[\frac{C_m(\alpha_1)C_m^*(\alpha_2)}{|C_m(\alpha_2)|^2}\right]. \tag{4.8}$$

Since arg[z] is a many-valued function, this requires a choice of which branch to use and this information is not available from the above expression. Tests indicated that the harmonic, m=1, gave most accurate results using the principal branch, Arg[z] (Any complex analysis text, such as [20], will discuss the complex exponential). An example of these derivations is given in Table 4.4 for cross correlations of the letter E, and circular harmonic components of rotated versions of E. This table shows a fairly accurate extraction of the relative orientation of two images.

Table 4.4 Information Extraction Using Circular Harmonics

Image: E using $\alpha_1 = 0, C_1(0) = 33.04733 - 23.37405j$			
m	α	Centre Value $C_m(\alpha)$	Derived α
1	45°	6.802149-40.587307j	44.275
-1	60°	-4.533462-38.729923j	65.704

4.3.2.1. Fourier-Mellin Techniques

Although this technique lacks uniqueness, information is available in the transform space. The location of the cross correlation peaks contain the relative orientation and scale of two images. Psaltis and Casasent fully describe this derivation in [7]. The Fourier transform of the product of $M_1^*(\omega_\rho,\omega_\theta)$ and $M_2(\omega_\rho,\omega_\theta)$, $FT[M_1^* M_2]$ is the Mellin-type cross correlation. If two images are rotated or scaled versions of each other then the peaks of the Mellin-type cross correlation occur at $(u,v)=(ln(a),2\pi+\alpha)$ and $(u,v)=(ln(a),\alpha)$. Thus by analyzing the cross correlation peaks, the relative scale factor a and the relative orientation α can be derived. As indicated in the previous section, there are some difficulties in extracting this information. The scale used in the log polar conversion must be kept track of and this scale must be fine enough to

extract these values. Typical values of a such as $\frac{1}{2} \leq a < 1$ have $-.69 < ln(a) < 0$ hence the extraction of $ln(a)$ requires a very small scale for sampling in r. In order to represent the image and accurately extract the value of $ln(a)$ requires enlarging the digital representation of the image to compensate for the high sampling frequency in r. Test results had $ln(a)=0$ for all scaling. These tests, however, did not increase the size of the image to compensate for this problem.

4.3.2.2. Logpolar Fourier Techniques

Similarly to the Cartesian cross correlation, the logpolar-Fourier calculation of the proposed transform in Chapter 3 contains information in the location of the cross correlation peaks. As described above, a peak occurs at (u,v) if and only if $f(x,y)=cg(x+u,y+v)$ for scalar c, thus indicating the relative displacement. Since the logpolar representation of the image converts rotations and dilations into shifts in $ln(r)$, and θ, the peaks of the cross correlation of the logpolar images indicates the relative changes in $ln(r)$ and θ. Thus a peak at (u,v) occurs if and only if $f(ln(r),\theta)=cg(ln(r)+u,\theta+v)$ for scalar c. A displacement of u in the $ln(r)$ direction of the logpolar image represents a scale change from r to $e^u r$ in the Cartesian coordinates and a displacement of v in the θ direction of the logpolar image represents a rotation from θ to $v+\theta$ in the Cartesian coordinates (see Figure 4.1). Plate 4.8 shows images, and the cross-correlation of the log polar image of the upper right hand image with the log polar image of each of the images. The peak moves along the horizontal axis, which corresponds to the $ln(r)$ axis, indicating a scaling in the images.

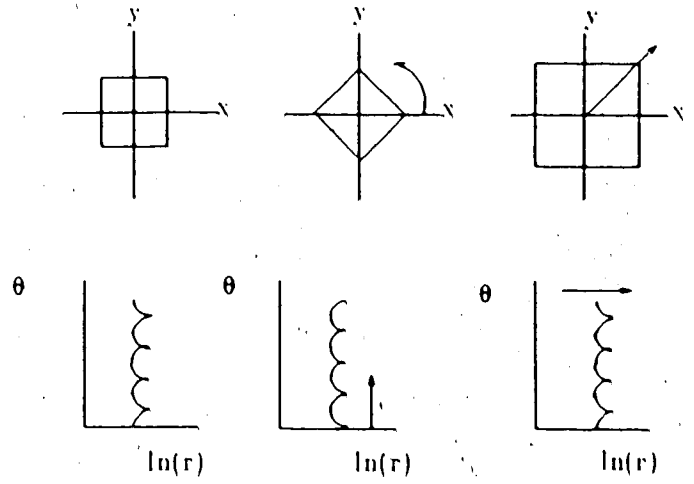


Figure 4.1 Effect of Rotation and Scaling on Logpolar Images

Table 4.5 Information From Logpolar Cross Correlation Peaks

I_1	I_2	Image Relation		Peak Location,	Corresponding Changes	
		a	α^o	[lnr,θ]	a	α^o
rect	rect	1.0	45	$[0,8(\frac{2\pi}{64})]$	1.00	45.0
E	E	2.0	0	[-11τ,0]	1.94	0.0
E	E	-2.0	45	$[-10\tau,7(\frac{2\pi}{64})]$	1.83	39.4

$\tau = \frac{ln(maxr)}{size}$, for size = 63 and maxr = $31.5\sqrt{2}$

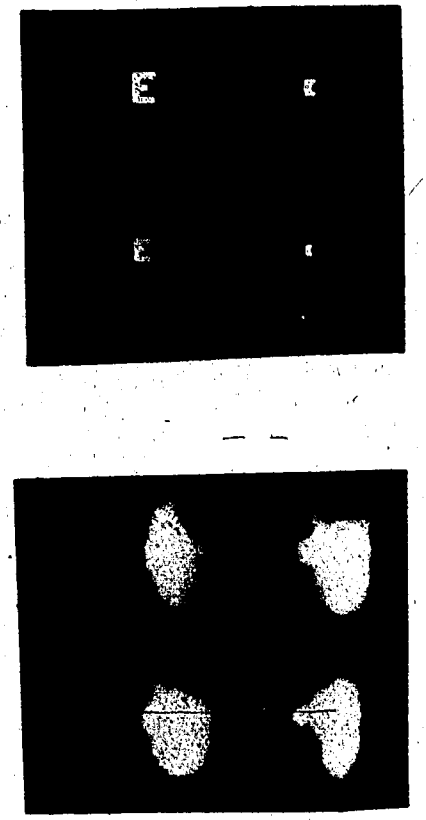


Plate 4.8 Crosscorrelation of Log-Polar Images

Information extraction using crosscorrelation of Log-Polar images where the peak locations indicate the relative orientation and scaling between images.

4.3.2.3. Direct Computation of Proposed Method

Similarly to the above two techniques, the relative orientation and scale changes can be derived from the transform of the image. Using the rotation and scale invariant properties from table 3.1, one does not need the cross correlation of two transforms in order to extract the above information. As with the circular harmonics, only two points of the transform are required. For $T[f(r,\theta)]=g(u,v)$:

$$T[f(ar,\theta+\alpha)]=e^{ju\kappa ln(a)+jv\zeta\alpha}g(u,v). \qquad 1.9$$

Thus using only two points, (u_1,v_1) and (u_2,v_2) for functions $f_1(r,\theta)=f(a_1r,\theta+\alpha_1)$ and $f_2(r,\theta)=f(a_2r,\theta+\alpha_2)$:

$$T[f_1(r,\theta)] = g_1(u,v) = e^{ju\kappa ln(a_1)+jv\zeta\alpha_1}g(u,v) \qquad 1.10$$

and

$$T[f_2(r,\theta)] = g_2(u,v) = e^{ju\kappa ln(a_2)+jv\zeta\alpha_2}g(u,v). \qquad 1.11$$

Rearranging the above yields:

$$g(u,v) = g_2(u,v)e^{-ju\kappa ln(a_2)-jv\zeta\alpha_2} \qquad 1.12$$

and

$$g(u,v) = g_1(u,v)e^{-ju\kappa ln(a_1)-jv\zeta\alpha_1}. \qquad 1.13$$

Equating the right hand sides above:

$$g_2(u,v)e^{-ju\kappa ln(a_2)-jv\zeta\alpha_2} = g_1(u,v)e^{-ju\kappa ln(a_1)-jv\zeta\alpha_1} \qquad 1.14$$

or, rearranging this, yields:

$$e^{ju\kappa ln(\frac{a_1}{a_2})+jv\zeta(\alpha_1-\alpha_2)} = \left[\frac{g_1(u,v)g_2^*(u,v)}{|g_2(u,v)|^2}\right] \qquad 1.15$$

$$u\kappa[ln(\frac{a_1}{a_2})] + v\zeta[\alpha_1-\alpha_2] = arg\left[\frac{g_1(u,v)g_2^*(u,v)}{|g_2(u,v)|^2}\right], \qquad 4.16$$

Using two points, (u_1,v_1) and (u_2,v_2), and using the above equation one can solve for the relative scale change, $\frac{a_1}{a_2}$ and the relative orientation $\alpha_1-\alpha_2$. Note, the two points required must satisfy $\frac{u_1}{v_1}\neq\frac{u_2}{v_2}$ in order to yield a unique solution for the resulting system of equations. As with the circular harmonics, arg[z] is a many valued function. however tests show that using the points (0,1) and (1,0) and the principal branch Arg, we get accurate results. Table 4.6 show the results of extracting information from the transform, where α and a are the relative orientation and scale size, respectively, between image 1 and image 2.

Table 4.6 Information Extraction Using Proposed Method

Image1	Image2	α	a	derived α^o	derived a
E	E	45	1	45.000002	1.002547
E	E	0	0.5	0.000000	0.502134
E	E	60	0.5	59.381980	0.496877
E½	E½	60	1	59.897535	0.996606
E½	E	45	2	44.603965	1.999494

4.4. Sensitivity of Recognition to the Centre

Tests were performed to determine the sensitivity of the methods to slight displacements of the centre. The Fourier-Mellin technique is not sensitive to the centre since it uses the power spectrum, which is invariant to shifts in the image, instead of the image itself. Table 4.7 summarizes typical results, where it can be seen that the vector signature methods will require a high threshold value, $T > 1000$, in order to recognize an image that has not been rotated; this may cause many false alarms.

Table 4.7 Sensitivity of Recognition to Displacements of the Centre

		LPCH	LPF	FM	Vector Signature Methods		
I	Δ	δ	R	δ	$\|\|X-X\|\|^2$ using components:		
					0-15	0-5	10-15
E	2	4.2389	0.961414	0.00003	953.9039	355.1031	627.7757
E	4	5.89978	0.920233	0.00001	1135.7088	235.3299	1009.3060
E	$2\sqrt{2}$	10.1197	0.8763	0.000021	1260.1815	731.63750	1092.6251
E	$4\sqrt{2}$	101.6762	0.6450	0.000015	1473.3353	792.62400	1455.9899
□	2	1.836608	0.7726	0.00002	1607.4871	1862.4199	2015.5788
□	4	3.730506	0.5092	0.00002	2507.5182	1927.7291	2949.8986
□	$2\sqrt{2}$	2.2301	0.5701	0.000018	2233.9659	1993.4712	2496.8991
□	$4\sqrt{2}$	1.2597	0.5389	0.000019	1903.2332	1853.6911	2546.7338

I = image: $\delta = \sqrt{\sum (A_i - B_i)^2}$; Δ = displacement of centre: $R = \dfrac{\int_{\theta_0}^{\theta_0+\pi} f}{\sqrt{\int_0^{2\pi} f^2}}$

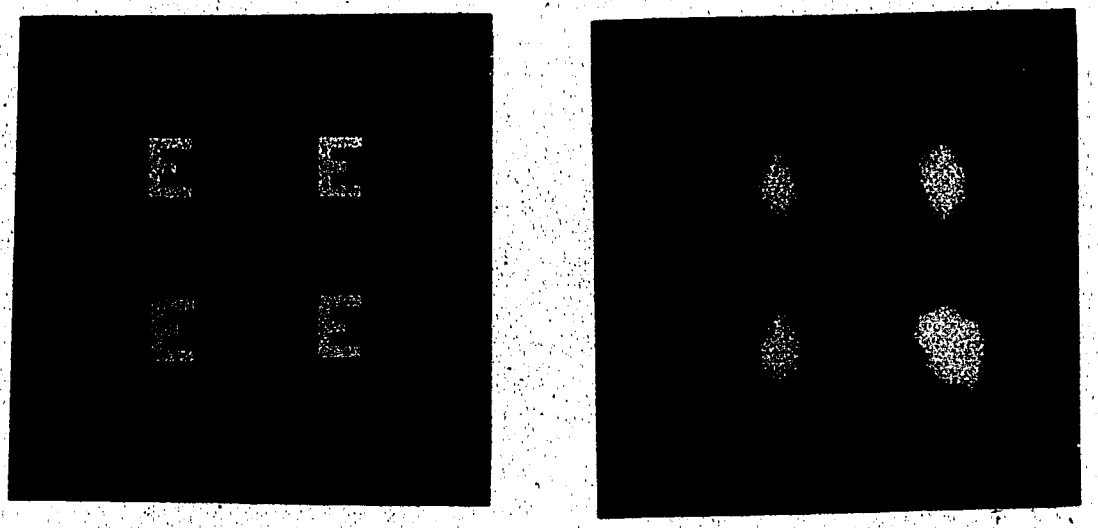


Plate 4.9 Effect of Centre with Log-Polar Fourier Transforms

Logpolar transform of each image followed by a conventional crosscorrelation. The second image shows the crosscorrelation of the letter E (centred) and the letters E with the centre moved. The displacements (in pixels) are:

$$\begin{matrix} 0 & 2\sqrt{2} \\ 2 & 4\sqrt{2} \end{matrix}$$

4.5. Computer Time and Space Requirements

The time and space requirements in order to compute the methods described in Chapters 2 and 3 are now considered. The emphasis of this thesis is on solving the problem of invariance coding not algorithm analysis; however, the methods must be "computable" to be applicable in a realistic digital image processing situation. In the following sections, all coordinate transformations were performed using bilinear interpolation as described in [23] which is an $O(n^2)$ algorithm and fast Fourier techniques as described in [13,14,26] are employed in computing Fourier transforms and cross correlations, which require time $O(n^2\log_2(n))$. All calculations are performed using floating point arithmetic on a VAX 11/780 running UNIX.†

4.5.1. Calculating Circular Harmonic Components

This method requires, for each m: (1) the computation of the circular harmonic component $f_m(r)e^{jm\theta}$ of the reference image; (2) the autocorrelation of the reference image with its m^{th} circular harmonic component (3) the computation of the circular harmonic component $g_m(r)e^{jm\theta}$ of the test pattern; (4) the cross correlation of the reference image with the circular harmonic component of the test pattern; and (5) the comparison between the centre values of the autocorrelation function and the cross correlation function.

The cross correlation of the image, $f(x,y)$ and the m^{th} circular harmonic component were calculated using conventional Fast Fourier techniques. Using real multiplications as a measure of time and assuming the images are $n \times n$, steps (1) and (3) require time $O(n^2)$ and steps (2) and (4) require $O(n^2\log_2(n))$. Step (5) requires constant time. To utilize the vector signature method, step (1)-(4) must be performed for each m, and the comparison $||X-X||^4$ requires $O(m)$ multiplications. Thus, the total

† Registered trademark of AT&T in the USA and other countries.

time required to calculate one comparison using the vector signature approach is: $O(mn^2\log_2(n)+mn^2+m)$.

For storage, each component is complex and requires space $O(2n\times 2n)$. The components, however, need not be stored, only the centre cross correlation is required.

4.5.2. Calculating Fourier-Mellin Transforms

The Fourier-Mellin transform was computed much as described by [7,9]. This method requires:

(1) the power spectrum of the reference image;

(2) the logpolar conversion of the powerspectrum;

(3) appropriate scaling and normalization;

(4) the power spectrum of the logpolar image;

(5) the power spectrum of the test pattern;

(6) the logpolar conversion of the powerspectrum;

(7) appropriate scaling and normalization;

(8) the power spectrum of the logpolar image; and

(9) comparison of the spectra either by cross correlation or pixel by pixel comparison.

This technique requires time: $O(n^2\log_2 n+n^2)$ and space $O(4n^2)$ for the complex images.

4.5.3. Calculating Proposed Method Using Direct Calculation

Direct calculation was used in the Cartesian coordinate system to avoid problems with resolution differences. The basis functions:

$$B_{uv}(x,y)=\frac{1}{x^2+y^2}\times e^{-j\kappa u\ln(\sqrt{x^2+y^2})-j\zeta v\tan^{-1}(\frac{y}{x})} \qquad 4.17$$

were computed for $-\frac{U}{2}\le u<\frac{U}{2}, -\frac{V}{2}\le v<\frac{V}{2}$ and stored. This computation required time $O(n^2UV)$ and storage $O(2n^2UV)$ to store the bases. Given that the basis

functions need only be computed once the direct computation of the method proposed in Chapter 3 requires $O(n^2(V)$ to compute $g(u,v)$ for $-\frac{U}{2} \le u < \frac{U}{2}, -\frac{V}{2} \le v < \frac{V}{2}$ and storage $O(2UV)$ to store $g(u,v)$.

4.5.4. Calculating Proposed Method Using Fourier Techniques

For comparison, the Fourier transform of the log-polar image $f(\rho,\theta)=f(e^{\rho},\theta)$ where $\rho=\ln(r)$ was computed and cross correlations of log-polar images were evaluated. As discussed in Chapter 3, this calculation will differ from the proposed technique due to the resolution difference at the origin.

Computation using the Fourier transform of the log-polar representation of the image takes time $O(n^2)$ to convert to log-polar form and $O(n^2\log_2(n))$ for the calculation of the Fourier transform. Storage requirements are $O(n^2)$ to store the image.

4.6. Summary

In previous chapters, techniques for invariant recognition were introduced. The tests in this chapter provide a basis for comparison of the use of the techniques described in Chapters 2 and 3. The experiments reveal problems in the digital implementation which are not apparent from the theory. The following observations have been made earlier in the chapter:

(1) the circular harmonic techniques require tests to determine which components to use and to find an appropriate threshold value;

(2) the circular harmonic techniques and the LPCHD proposed in Chapter 3 are sensitive to displacements of the centre, that is, they have no shift invariance;

(3) the complete "set" of harmonic components uniquely describes the image; however, using only a few harmonics is *not* a unique representation of the image;

(4) including more than one object in an image caused recognition problems for *all* techniques;

(5) In order to compare the Fourier-Mellin representations, a normalization must be performed or comparisons must be made using the normalized cross correlation, N_{fg}, therefore, this is not an invariant form of the image;

(6) The Fourier-Mellin representation is not unique, this causes inaccuracy in matching results;

(7) The Fourier-Mellin representation requires a high sampling frequency in r to accurately extract information and to accurately represent all information, therefore the image size must be increased or some information must be discarded;

(8) The transform proposed in Chapter 3, LDCHD, is an invariant form without any normalization;

(9) The LPCHD works well in noise;

(10) LPCHD either requires large amounts of space (to store the basis functions) or much more computer time, to calculate the transform;

(11) Both the Fourier-Mellin representation and the proposed transform have difficulty discriminating between similar objects, while the vector signature methods performed well when using a "good" choice of harmonics and threshold value;

(12) The Fourier-Mellin technique experiences more false alarms, including false alarms using objects that were not close.

From these observations, some conclusions can be drawn about the uses of each of these methods.

Circular harmonic techniques will work well in an application that requires classification of objects into predefined classes. Under these circumstances, the choice of harmonics and the threshold value which "best" identifies each class could be determined beforehand through tests. However, without position information or scale invariance, and considering that the vector signature method requires computing 2N circular harmonic components and then performing 2N cross correlations, the question arises as to whether this technique is an improvement over using a conventional

matched filter and rotating each test image.

With the Fourier-Mellin techniques, the lack of uniqueness implies possible inaccuracy in matching. In addition, with a digital implementation, errors are added with each computation: Fourier transform and power spectra calculation, logpolar coordinate conversion, compression to reduce aperture effects, and again, Fourier transform and power spectra calculation. Also, as mentioned earlier, the Fourier-Mellin power-spectrum:

$$|\mathrm{FT}\{L^2[|\mathrm{FT}(\mathit{Image})|^2]\}|^2 \qquad 4.18$$

is not a truly invariant form of the image, one still must use cross correlation or some form of normalization to compare this representation. This method is indeed faster than using multiple templates at various orientations and sizes. Here, the question arises as to how accurate the results are and how can the results be compared without having to perform additional cross correlations.

In contrast, the transform proposed in Chapter 3 is an invariant form of a rotated or scaled image; however, it lacks translation invariance and is slow to compute. Comparing $g_1(u,v)$ directly with $g_2(u,v)$ for each (u,v) determines whether a match has occurred. In order to have accurate results, the images must be centred about the invariant point (that is, the point which does not change after performing a rotation or scaling of the image). This is a non-trivial task. Also, the computation is slow to compute many values of $g(u,v)$; however, the number of values needed to establish a match is not high (usually 16 will do).

The tests have shown that under certain conditions, when the position of an object is known, the proposed transform enables the mapping of an image to a representation that will allow comparison of the spectra, and one can determine whether objects match (up to scaling or rotation). Indeed, under such conditions,

using such a technique would surpass the use of a conventional matched filter requiring many templates. In any real application, the constraint of finding the proper centre of an object seriously inhibits the use of such a method.

Chapter 5

Summary and Conclusions

The intent of this research was to explore the problem of encoding images in a form which is invariant to geometric transforms in the spatial domain. This thesis investigated image transforms which produce such an encoding. After reviewing current techniques in the literature, the desired invariant properties were used to develop a kernel for an integral transform such that this transform satisfied specific invariant properties. This technique was then tested to insure that it was digitally feasible and tests were performed to compare it to existing methods.

Results of the tests indicate that the method works well in recognizing objects but has weaknesses in discriminating between similar objects and has no position invariance due to the dependence of the method on the centre. The dependence on the centre, however, seems "natural" for rotation and scale invariance as it does not make sense to discuss these transforms without mention of the centre. In addition, the proposed method can be expanded to include position invariance (see Section 3.8), but using magnitude invariance and uniqueness as criteria, the solution does not appear to lie in 2−D. Without a priori knowledge, this solution works as well as existing techniques. In addition, the proposed technique mathematically solves the problem of unique invariant encoding of rotated and scaled images.

5.1. Further Work

Thus the problem of finding an encoding of an image in 2−D which is invariant to geometric transforms has not yet been solved. The expansion of the method discussed in Section 3.8 to include the centre does solve the problem; however, the complexity is increased since the solution is in a four dimensional space. Further work in this area would include determining if the problem is solvable in 2 dimensions, and determining what necessary conditions are required for invariance coding (Chapter 3

considered only sufficient conditions). Also, the approach used in Chapter 3 could be considered in determining a recognition transform for higher dimensions.

The tests performed in Chapter 4 indicate that work could be done in developing fast algorithms for the computation of the transform. The difficulties with similar objects might be improved through the use of edge-only information. Since this improves the performance of a conventional matched filter, perhaps a similar technique will improve the performance of this transform.

Further study is also possible in the area of human vision. The literature [8, 24, 25] suggests that the human eye performs a log-polar transform. There have also been arguments as to whether or not the eye performs a Fourier transform [8]. It might be an interesting problem to study the feasibility of the proposed transform since it is essentially a combination log-polar and Fourier transform.

References

[1] H. H. Arsenault and Yuan-Neng Hsu, "Rotation-invariant discrimination between almost similar objects", *Applied Optics*, Vol. 22, No. 1, January,1983, pp. 130-132.

[2] D. Ballard and C. Brown, *Computer Vision*, Prentice-Hall, Inc., Englewood Clifs, New Jersey, 1982.

[3] C. Braccini, "Scale-invariant image processing by means of scaled transforms or form-invariant, linear shift-variant filters", *Optics Letters*, Vol. 8, No. 7, July,1983, pp. 392-394.

[4] R. N. Bracewell, *The Fourier Transform and Its Applications, Second Edition*, McGraw-Hill Book Company, USA, 1978.

[5] D. Casasent and D. Psaltis, "Scale Invariant Optical Transform", *Optical Engineering*, Vol. 15, No. 3, May-June, 1976, pp. 258-261.

[6] D. Casasent and D. Psaltis, "Position, Rotation, and Scale Invariant Optical Correlation", *Applied Optics*, Vol. 15, No. 7, July, 1976, pp. 1795-1799.

[7] D. Casasent and D. Psaltis, "New Optical Transforms for Pattern Recognition", *Proceedings of the IEEE*, Vol. 65, No. 1, January, 1977, pp. 77-84.

[8] P. Cavanagh, "Size and position invariance in the visual system", *Perception*, Vol. 7, 1978, pp. 167-177.

[9] P. Cavanagh, Image Transforms in the Visual System, in *Figural Synthesis*, T. M. Caelli and P. C. Dodwell (ed.), Press, pp. 185-218.

[10] G. M. Chaiken and C. F. R. Weiman, "Logarithmic Spiral Grids for Image Processing and Display", *Computer Graphics and Image Processing*, Vol. 11, 1979, pp. 197-226.

[11] P. C. Dodwell, Geometrical Approaches to Visual Processing, in *Analysis of Visual Behavior*, D. J. Ingle, M. A. Goodale and R. J. W. Mansfield (ed.), MIT Press, pp. 801-825.

[12] W. Doyle, "Operations Useful for Similarity-Invarjant Pattern Recognition", *ACM Journal*, Vol. 9, 1962, pp. 259-267.

[13] D. E. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1984.

[14] G-AE Subcommittee on Measurement Concepts, "What is the Fast Fourier Transform?", *IEEE Transactions on Acoustics, etc. (then Audio and Electroacoustics)*, Vol. AU-15, No. 2, June 1967, pp. 49-50.

[15] J. Goodman, *An Introduction to Fourier Optics*, McGraw-Hill Book Company, USA, 1968.

[16] E. W. Hansen, "Circular harmonic image reconstruction: Experiments", *Applied Optics*, Vol. 20, No. 13, July 1, 1981, pp. 2266-2274.

[17] Yuan-Neng Hsu, H. H. Arsenault and G. April, "Rotation-invariant Digital Pattern Recognition Using Circular Harmonic Expansion", *Applied Optics*, Vol. 21, No. 22, November 15, 1982, pp. 4012-4015.

[18] R. W. Klopfenstein and C. R. Carlson, "Theory of Shape-invariant Imaging Systems", *Journal of the Optical Society of America*, Vol. (A) 1, No. 10, October, 1984, pp. 1040-1053.

[19] V. A. Kovalevsky, *Image Pattern Recognition*, Springer-Verlag, New York, 1980.

[20] N. Levinson and R. M. Redheffer, *Complex Variables*, Holden-Day, Inc., San Francisco, CA, 1970.

[21] A. Vander Lugt, "Signal Detection By Complex Spatial Filtering", *IEEE Transactions on Information Theory*, Vol. 10, April, 1964, pp. 139-145.

[22] A. Papoulis, *Systems and Transforms with Applications in Optics*, McGraw Hill, New York, 1968.

[23] A. Rosenfeld and A. C. Kak, *Digital Picture Processing, Second Edition*, Academic Press, Orlando, Florida, USA, 1982.

[24] E. L. Schwartz, "Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding", *Vision Research*, Vol. 20, 1980, pp. 645-669.

[25] E. L. Schwartz, "Cortical anatomy, size invariance, and spatial frequency analysis", *Perception*, Vol. 10, 1981, pp. 455-468.

[26] R. C. Singleton, "On Computing the Fast Fourier Transform", *Communications of the ACM*, Vol. 10, No. 10, October, 1967.

[27] H. Stark, "Sampling theorems in polar coordinates", *J. Opt. Soc. Am.*, Vol. 69, No. 11, November,1979, pp. 1519-1525.

[28] R. Wu and H. Stark, "Rotation-invariant Pattern Recognition Using a Vector Reference", *Applied Optics*, Vol. 23, No. 6, March 15, 1984, pp. 838-840.

[29] Y. Yang, Yuan-Neng Hsu and H. H. Arsenault, "Optimum Circular Symmetrical Filters and their Uses in Optical Pattern Recognition", *Optica Acta*, Vol. 29, No. 5, 1982, pp. 627-644.

Appendix A1

Proofs of Properties

Table 3.1 listed properties of the transform of $f(x,y)$, given by:

$$g(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\omega(u,v;x,y)\,dx\,dy \qquad \text{A.1}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\frac{e^{-j\kappa u\ln(\sqrt{x^2+y^2})-j\zeta v\tan^{-1}(\frac{y}{x})}}{x^2+y^2}\,dx\,dy, \qquad \text{A.2}$$

or in polar coordinates:

$$= \int_{0}^{2\pi}\int_{0}^{\infty} f(r,\theta)e^{-j\kappa u\ln(r)-j\zeta v\theta}\,\frac{dr}{r}\,d\theta. \qquad \text{A.3}$$

In the following discussions, this relationship will be denoted $f(x,y) \leftrightarrow g(u,v)$.

The proofs of the properties in Table 3.1 follow from the equivalent properties of the Fourier transform (see [13, Chapter 2]). Linearity is obvious, scale invariance and rotation invariance are the criteria on which the transform was derived (Section 3.2), so they will not be repeated here. The other properties listed in the table are shown in the next sections.

1.1. Differentiation

The differentiation properties are:

$$-jk\ln r f(r,\theta) \leftrightarrow \frac{\partial}{\partial u}\left[g(u,v)\right] \qquad \text{A.4}$$

$$-j\zeta\theta f(r,\theta) \leftrightarrow \frac{\partial}{\partial v}\left[g(u,v)\right] \qquad \text{A.5}$$

$$-\kappa\zeta\theta\ln r f(r,\theta) \leftrightarrow \frac{\partial^2}{\partial v\partial u}\left[g(u,v)\right]. \qquad \text{A.6}$$

To show A.4 holds:

$$\frac{\partial}{\partial u}\left[g(u,v)\right] = \frac{\partial}{\partial u}\left[\int_0^{2\pi}\int_0^{\infty}\tilde{f}(r,\theta)e^{-j\kappa u\ln(r)-j\zeta v\theta}\frac{dr}{r}d\theta\right] \qquad A.7$$

taking the differentiation inside (assumptions about the image function in Chapter 1 justify this):

$$= \int_0^{2\pi}\int_0^{\infty}\tilde{f}(r,\theta)\frac{\partial}{\partial u}\left[e^{-j\kappa u\ln(r)-j\zeta v\theta}\right]\frac{dr}{r}d\theta \qquad A.8$$

$$= \int_0^{2\pi}\int_0^{\infty}\tilde{f}(r,\theta)[-j\kappa\ln(r)]e^{-j\kappa u\ln(r)-j\zeta v\theta}\frac{dr}{r}d\theta \qquad A.9$$

$$= \int_0^{2\pi}\int_0^{\infty}\left[-j\kappa\ln(r)\tilde{f}(r,\theta)\right]\tilde{\omega}(u,v;r,\theta)\frac{dr}{r}d\theta \qquad A.10$$

$$\leftrightarrow -j\kappa\ln r\tilde{f}(r,\theta). \qquad A.11$$

The second property, A.5, is similar:

$$\frac{\partial}{\partial v}\left[g(u,v)\right] = \frac{\partial}{\partial v}\left[\int_0^{2\pi}\int_0^{\infty}\tilde{f}(r,\theta)e^{-j\kappa u\ln(r)-j\zeta v\theta}\frac{dr}{r}d\theta\right]. \qquad A.12$$

Again, moving the differentiation inside the integral:

$$= \int_0^{2\pi}\int_0^{\infty}\tilde{f}(r,\theta)\frac{\partial}{\partial v}\left[e^{-j\kappa u\ln(r)-j\zeta v\theta}\right]\frac{dr}{r}d\theta \qquad A.13$$

$$= \int_0^{2\pi}\int_0^{\infty}\tilde{f}(r,\theta)[-j\zeta\theta]e^{-j\kappa u\ln(r)-j\zeta v\theta}\frac{dr}{r}d\theta \qquad A.14$$

$$= \int_0^{2\pi}\int_0^{\infty}\left[-j\zeta\theta\tilde{f}(r,\theta)\right]\omega(u,v;r,\theta)\frac{dr}{r}d\theta \qquad A.15$$

$$\leftrightarrow -j\zeta\theta\tilde{f}(r,\theta). \qquad A.16$$

The third property follows trivially from A.4 and A.5.

1.2. Modulation

The modulation property is:

$$\tilde{f}(r,\theta)e^{j\lambda\kappa\ln r + j\mu\zeta\theta} \leftrightarrow \tilde{g}(u-\lambda, v-\mu). \qquad \text{A.17}$$

for $\lambda,\mu \in \mathbf{R}$. This property is shown by direct substitution into the integral form of the transform:

$$\int_0^{2\pi}\int_0^{\infty}\left[\tilde{f}(r,\theta)e^{j\lambda\kappa\ln r + j\mu\zeta\theta}\right] e^{-j\kappa u\ln(r) - j\zeta v\theta}\,\frac{dr}{r}d\theta \qquad \text{A.18}$$

$$= \int_0^{2\pi}\int_0^{\infty}\tilde{f}(r,\theta)e^{-j\kappa|u-\lambda|\ln(r) - j\zeta|v-\mu|\theta}\,\frac{dr}{r}d\theta \qquad \text{A.19}$$

$$= \tilde{g}(u-\lambda, v-\mu). \qquad \text{A.20}$$

1.3. Conjugation

The conjugation property is:

$$\tilde{f}^{*}(r,\theta) \leftrightarrow [\tilde{g}(-u,-v)]^{*}. \qquad \text{A.21}$$

To show this property, consider the formula:

$$\tilde{g}(u,v) = \int_0^{2\pi}\int_0^{\infty}\tilde{f}(r,\theta)\; e^{-j\kappa u\ln(r) - j\zeta v\theta}\,\frac{dr}{r}d\theta. \qquad \text{A.22}$$

using the notation $\tilde{f}_R(r,\theta) = \mathrm{RE}\,[\tilde{f}(r,\theta)]$, $\tilde{f}_I(r,\theta) = \mathrm{IM}\,[\tilde{f}(r,\theta)]$, and for the basis functions, $B^{+}_{(u,v)}(r,\theta) = \mathrm{RE}\;\tilde{\omega}(u,v;r,\theta)$, $B^{-}_{(u,v)}(r,\theta) = \mathrm{IM}\;\tilde{\omega}(u,v;r,\theta)$ express A.22 as:

$$\tilde{g}(u,v) = \int_0^{2\pi}\int_0^{\infty}[\tilde{f}_R(r,\theta) + j\tilde{f}_I(r,\theta)][B^{+}_{(u,v)}(r,\theta) + jB^{-}_{(u,v)}(r,\theta)]\,dr\,d\theta. \qquad \text{A.23}$$

Then

$$g(u,v) = \int_0^{2\pi}\int_0^{\infty} [\tilde{f}_R(r,\theta)\, B^{+}_{(u,v)}(r,\theta) - \tilde{f}_I(r,\theta)\, B^{-}_{(u,v)}(r,\theta)] \qquad \text{A.24}$$

$$+ j\,[\tilde{f}_R(r,\theta)\, B^{-}_{(u,v)}(r,\theta) + \tilde{f}_I(r,\theta)\, B^{+}_{(u,v)}(r,\theta)]\, dr d\theta. \qquad \text{A.25}$$

Some simple observations:

$$B^{+}_{(u,v)}(r,\theta) = \text{RE}\ \tilde{\omega}(u,v;r,\theta) = \frac{\cos[-\kappa u \ln(r) - \zeta v\theta]}{r^2}$$

$$= \frac{\cos[\kappa u \ln(r) + \zeta v\theta]}{r^2} = B^{+}_{(-u,-v)}(r,\theta), \qquad \text{A.26}$$

and,

$$B^{-}_{(u,v)}(r,\theta) = \frac{\sin[-\kappa u \ln(r) - \zeta v\theta]}{r^2}$$

$$= -B^{-}_{(-u,-v)}(r,\theta), \qquad \text{A.27}$$

Hence to evaluate the transform of $\tilde{f}^{*}(r,\theta)$, replace $\tilde{f}_I$ with $-\tilde{f}_I$:

$$g(u,v) = \int_0^{2\pi}\int_0^{\infty} [\tilde{f}_R(r,\theta)\, B^{+}_{(u,v)}(r,\theta) + \tilde{f}_I(r,\theta)\, B^{-}_{(u,v)}(r,\theta)] \qquad \text{A.28}$$

$$+ j\,[\tilde{f}_R(r,\theta)\, B^{-}_{(u,v)}(r,\theta) - \tilde{f}_I(r,\theta)\, B^{+}_{(u,v)}(r,\theta)]\, dr d\theta \qquad \text{A.29}$$

$$= \int_0^{2\pi}\int_0^{\infty} [\tilde{f}_R(r,\theta)\, B^{+}_{(u,v)}(r,\theta) - \tilde{f}_I(r,\theta)\, [-B^{-}_{(u,v)}(r,\theta)]\,] \qquad \text{A.30}$$

$$- j\,[\tilde{f}_R(r,\theta)\, [-B^{-}_{(u,v)}(r,\theta)] + \tilde{f}_I(r,\theta)\, B^{+}_{(u,v)}(r,\theta)]\, dr d\theta. \qquad \text{A.31}$$

and using the observations A.26 and A.27:

$$= \int_0^{2\pi}\int_0^{\infty} [\tilde{f}_R(r,\theta)\, B^{+}_{(-u,-v)}(r,\theta) - \tilde{f}_I(r,\theta)\, B^{-}_{(-u,-v)}(r,\theta)] \qquad \text{A.32}$$

$$- j \left[\tilde{f}_R(r,\theta)\, B^{-}_{(-u,-v)}(r,\theta) + \tilde{f}_I(r,\theta)\, B^{+}_{(-u,-v)}(r,\theta) \right] dr d\theta \qquad \text{A.33}$$

$$= \int_0^{2\pi} \int_0^{\infty} \left[\tilde{f}_R(r,\theta) + j\, \tilde{f}_I(r,\theta) \right] \left[B^{+}_{(-u,-v)}(r,\theta) - j\, B^{-}_{(-u,-v)}(r,\theta) \right] dr d\theta \qquad \text{A.34}$$

$$= g^{*}(-u,-v). \qquad \text{A.35}$$

If the function $f(x,y)$ is real then $f^{*}(x,y) = f(x,y)$ so

$$g(u,v) = g^{*}(-u,-v). \qquad \text{A.36}$$

Similarly it is possible to demonstrate that:

$$g(-u,v) = g^{*}(u,-v). \qquad \text{A.37}$$

1.4. Real and Imaginary Parts

The claim is:

$$\mathrm{RE}\left[\tilde{f}(r,\theta)\right] \leftrightarrow \frac{1}{2}\left[g(u,v) + g(-u,-v)^{*}\right] \qquad \text{A.38}$$

$$\mathrm{IM}\left[\tilde{f}(r,\theta)\right] \leftrightarrow \frac{1}{2}\left[g(u,v) - g(-u,-v)^{*}\right]. \qquad \text{A.39}$$

From the above section we have:

$$g(u,v) = \int_0^{2\pi} \int_0^{\infty} \left[\tilde{f}_R(r,\theta) + j\, \tilde{f}_I(r,\theta) \right] \left[B^{+}_{(u,v)}(r,\theta) + j\, B^{-}_{(u,v)}(r,\theta) \right] dr d\theta \qquad \text{A.40}$$

Using the observations A.26 and A.27:

$$g(-u,-v) \qquad \text{A.41}$$

$$= \int_0^{2\pi} \int_0^{\infty} \left[\tilde{f}_R(r,\theta) + j\, \tilde{f}_I(r,\theta) \right] \left[B^{+}_{(-u,-v)}(r,\theta) + j\, B^{-}_{(-u,-v)}(r,\theta) \right] dr d\theta$$

$$= \int_0^{2\pi} \int_0^{\infty} \left[\tilde{f}_R(r,\theta) + j\, \tilde{f}_I(r,\theta) \right] \left[B^{+}_{(u,v)}(r,\theta) - j\, B^{-}_{(u,v)}(r,\theta) \right] dr d\theta \qquad \text{A.42}$$

$$g^*(-u,-v) = \int_0^{2\pi}\int_0^{\infty} [\tilde{f}_R(r,\theta) - j\,\tilde{f}_I(r,\theta)][B^+_{(u,v)}(r,\theta) + j\,B^-_{(u,v)}(r,\theta)]\,dr d\theta. \quad \text{A.43}$$

Now, adding $\frac{1}{2}[g(u,v) + g^*(-u,-v)]$:

$$\frac{1}{2}[g(u,v) + g^*(-u,-v)] \quad \text{A.44}$$

$$= \frac{1}{2}\int_0^{2\pi}\int_0^{\infty} [\tilde{f}_R(r,\theta) + j\,\tilde{f}_I(r,\theta)][B^+_{(u,v)}(r,\theta) + j\,B^-_{(u,v)}(r,\theta)]$$

$$+ [\tilde{f}_R(r,\theta) - j\,\tilde{f}_I(r,\theta)][B^+_{(u,v)}(r,\theta) + j\,B^-_{(u,v)}(r,\theta)]\,dr d\theta$$

$$= \frac{1}{2}\int_0^{2\pi}\int_0^{\infty} [2\,\tilde{f}_R(r,\theta)][B^+_{(u,v)}(r,\theta) + j\,B^-_{(u,v)}(r,\theta)]\,dr d\theta \quad \text{A.45}$$

$$= \int_0^{2\pi}\int_0^{\infty} [\tilde{f}_R(r,\theta)]\,[B^+_{(u,v)}(r,\theta) + j\,B^-_{(u,v)}(r,\theta)]\,dr d\theta. \quad \text{A.46}$$