

**IMAGE OBJECT SEARCH COMBINING COLOUR WITH  
GABOR WAVELET SHAPE DESCRIPTORS**

by

Darryl Anderson

B.Sc., University of Victoria, 1997

A THESIS SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF  
MASTER OF SCIENCE  
in the School  
of  
Computing Science

© Darryl Anderson 2004  
SIMON FRASER UNIVERSITY  
Fall 2004

All rights reserved. This work may not be  
reproduced in whole or in part, by photocopy  
or other means, without the permission of the author.

## APPROVAL

**Name:** Darryl Anderson  
**Degree:** Master of Science  
**Title of thesis:** Image Object Search Combining Colour with Gabor Wavelet  
Shape Descriptors

**Examining Committee:** Dr. Ghassan Hamarneh, Associate Professor  
Chair

---

Dr. Mark S. Drew, Associate Professor  
Senior Supervisor

---

Dr. Ze-Nian Li, Professor  
Supervisor

---

Dr. Stella Atkins, Professor  
Examiner

**Date Approved:** \_\_\_\_\_

# Abstract

An image and object search and retrieval algorithm is devised that combines colour and spatial information. Spatial characteristics are described in terms of Wiskott's jets formulation, based on a set of Gabor wavelet functions at varying scales, orientations and locations. Colour information is first converted to a form more impervious to illumination colour change, reduced to 2D, and encoded in a histogram. The histogram, which is based on a new stretched chromaticity space for which all bins are populated, is resized and compressed by way of a DCT. An image database is devised by replicating JPEG images by a set of transforms that include resizing, various cropping attacks, JPEG quality changes, aspect ratio alteration, and reducing colour to greyscale. Correlation of the complete encode vector is used as the similarity measure. For both searches with the original image as probe within the complete dataset, and with the altered images as probes with the original dataset, the grayscale, the stretched, and the resized images had near-perfect results. The most formidable challenge was found to be images that were cropped both horizontally as well as vertically. The algorithm's ability to identify *objects*, as opposed to just images, is also tested. In searching for images in a set of 5 classifications, the jets were found to contribute most analytic power when objects with distinctive spatial characteristics were the target.

*This thesis is dedicated to my wife Erika. Without her continuous support and encouragement this would have been the extent of my thesis. I would also like to acknowledge Amanda, our daughter, who was born during the first semester of this degree. Amanda is now over three years old and has told me that she is looking forward to my convocation. That makes two of us Amanda.*

*“I think I can, I think I can, I think I can!”*

– *The Little Engine That Could*, WATTY PIPER, 1978

# Acknowledgments

I acknowledge the support of the Science Council of British Columbia through its Graduate Research Engineering And Technology (GREAT) scholarship. This research has been supported by Imagis Technologies as the industrial sponsor for the GREAT scholarship.

# Contents

Approval	ii
Abstract	iii
Dedication	iv
Quotation	v
Acknowledgments	vi
List of Tables	ix
List of Figures	xii
List of Programs	xiv
<b>1 Introduction</b>	<b>1</b>
<b>2 Literature Review and Background</b>	<b>3</b>
2.1 Linear Correlation . . . . .	3
2.2 Colour Analysis . . . . .	3
2.2.1 Colour Histograms . . . . .	4
2.2.2 Illumination . . . . .	4
2.2.3 Chromaticity Histograms . . . . .	4
2.3 Spectral Analysis . . . . .	8
2.3.1 Transforms . . . . .	10
2.3.2 Wavelet Transforms . . . . .	11

2.3.3	Gabor Wavelet Filter . . . . .	12
2.3.4	Jets . . . . .	12
2.3.5	Gabor Filter Frequencies . . . . .	14
2.4	Combined Colour and Spectral Analysis . . . . .	14
<b>3</b>	<b>Procedure</b>	<b>18</b>
3.1	Frequency Analysis . . . . .	18
3.2	Colour Analysis and Illumination Invariance . . . . .	25
3.3	Image Encoding . . . . .	26
<b>4</b>	<b>Experiments</b>	<b>29</b>
4.1	Evolution of the Spectral Algorithmic Component . . . . .	31
4.1.1	16 x 16 Wavelet Grid . . . . .	33
4.1.2	10 x 10 Wavelet Grid . . . . .	35
4.1.3	8 x 8 Wavelet Grid . . . . .	38
4.1.4	6 x 6 Wavelet Grid . . . . .	41
4.1.5	5 x 5 Wavelet Grid . . . . .	44
4.1.6	Receiver Operator Curve . . . . .	50
4.1.7	Variance in Magnitude Coefficients Among Same Source Images . . . . .	50
4.2	Colour Analysis Evolution . . . . .	50
4.2.1	Single RGB Histogram Intersection . . . . .	55
4.2.2	Multiple RGB Histogram Intersection . . . . .	56
4.2.3	Multiple Compressed Chromaticity Coefficients . . . . .	59
4.3	Colour and Spectral Combined . . . . .	61
4.4	Similar Image Finding . . . . .	63
4.5	Example Searches . . . . .	67
<b>5</b>	<b>Conclusions and Future Work</b>	<b>79</b>
5.1	Conclusions . . . . .	79
5.2	Future Work . . . . .	81
<b>A</b>	<b>Generating Chromaticity Histograms in Matlab</b>	<b>82</b>
A.1	List of Programs . . . . .	82
	<b>Bibliography</b>	<b>86</b>



# List of Tables

3.1	Gabor Jet Evaluation Positions on Image 833009.jpg . . . . .	28
4.1	Format and motivation behind the image copies included in the test dataset.	30
4.2	Using a 16x16 Gabor Wavelet Grid, every original JPEG image was tested as a probe for finding each image variation in turn. The CMC score was recorded and the percentage of the variations returned, within each ranking category, is listed. . . . .	34
4.3	Using a 16x16 Gabor Wavelet Grid, every image variation was tested as a probe for finding each original JPEG in turn. The CMC score was recorded and the percentage of the original number of images returned, within each ranking category, is listed. . . . .	34
4.4	Using a 10x10 Gabor Wavelet Grid, every original JPEG image was tested as a probe for finding each image variation in turn. The CMC score was recorded and the percentage of the variations returned, within each ranking category, is listed. . . . .	39
4.5	Using a 10x10 Gabor Wavelet Grid, every image variation was tested as a probe for finding each original JPEG in turn. The CMC score was recorded and the percentage of the original number of images returned, within each ranking category, is listed. . . . .	39
4.6	Using an 8x8 Gabor Wavelet Grid, every original JPEG image was tested as a probe for finding each image variation in turn. The CMC score was recorded and the percentage of the variations returned, within each ranking category, is listed. . . . .	42

4.7	Using an 8x8 Gabor Wavelet Grid, every image variation was tested as a probe for finding each original JPEG in turn. The CMC score was recorded and the percentage of the original number of images returned, within each ranking category, is listed. . . . .	42
4.8	Using a 6x6 Gabor Wavelet Grid, every original JPEG image was tested as a probe for finding each image variation in turn. The CMC score was recorded and the percentage of the variations returned, within each ranking category, is listed. . . . .	45
4.9	Using a 6x6 Gabor Wavelet Grid, every image variation was tested as a probe for finding each original JPEG in turn. The CMC score was recorded and the percentage of the original number of images returned, within each ranking category, is listed. . . . .	45
4.10	Using a 5x5 Gabor Wavelet Grid, every original JPEG image was tested as a probe for finding each image variation in turn. The CMC score was recorded and the percentage of the variations returned, within each ranking category, is listed. . . . .	48
4.11	Using a 5x5 Gabor Wavelet Grid, every image variation was tested as a probe for finding each original JPEG in turn. The CMC score was recorded and the percentage of the original number of images returned, within each ranking category, is listed. . . . .	48
4.12	Average Within Class Scatter for Frequency Level 1 Per Image Tile . . . . .	53
4.13	Average Within Class Scatter for Frequency Level 2 Per Image Tile . . . . .	53
4.14	Average Within Class Scatter for Frequency Level 3 Per Image Tile . . . . .	54
4.15	Average Within Class Scatter for Frequency Level 4 Per Image Tile . . . . .	54
4.16	Using a RGB Color Histogram Intersection, every original JPEG image was tested as a probe for finding each image variation in turn. The CMC score was recorded and the percentage of the variations returned, within each ranking category, is listed. . . . .	57
4.17	Using a RGB Color Histogram Intersection, every image variation was tested as a probe for finding each original JPEG in turn. The CMC score was recorded and the percentage of the original number of images returned, within each ranking category, is listed. . . . .	57

4.18	Using a 5 x 5 RGB Non-Overlapping Color Histogram Intersection, every original JPEG image was tested as a probe for finding each image variation in turn. The CMC score was recorded and the percentage of the variations returned, within each ranking category, is listed. . . . .	58
4.19	Using a 5 x 5 RGB Non-Overlapping Color Histogram Intersection, every image variation was tested as a probe for finding each original JPEG in turn. The CMC score was recorded and the percentage of the original number of images returned, within each ranking category, is listed. . . . .	58
4.20	Using a 5 x 5 RGB Overlapping Color Histogram Intersection, every original JPEG image was tested as a probe for finding each image variation in turn. The CMC score was recorded and the percentage of the variations returned, within each ranking category, is listed. . . . .	60
4.21	Using a 5 x 5 RGB Overlapping Color Histogram Intersection, every image variation was tested as a probe for finding each original JPEG in turn. The CMC score was recorded and the percentage of the original number of images returned, within each ranking category, is listed. . . . .	60
4.22	Using a 5 x 5 Compressed Chromaticity, every original JPEG image was tested as a probe for finding each image variation in turn. The CMC score was recorded and the percentage of the variations returned, within each ranking category, is listed. . . . .	62
4.23	Using a 5 x 5 Compressed Chromaticity, every image variation was tested as a probe for finding each original JPEG in turn. The CMC score was recorded and the percentage of the original number of images returned, within each ranking category, is listed. . . . .	62
4.24	Using a 50% weighted combination of Colour and Spectral components, every original JPEG image was tested as a probe for finding each image variation in turn. The CMC score was recorded and the percentage of the variations returned, within each ranking category, is listed. . . . .	64
4.25	Using a 50% weighted combination of Color and Spectral components, every image variation was tested as a probe for finding each original JPEG in turn. The CMC score was recorded and the percentage of the original number of images returned, within each ranking category, is listed. . . . .	64
4.26	Results from similar image finding tests on five classifications of similar images.	67

# List of Figures

2.1	Sample image: Abstract Color 833009.jpg . . . . .	6
2.2	Binarized Linear Chromaticity Histogram from image 833009.jpg . . . . .	7
2.3	Binarized Spherical Chromaticity Histogram from Image 833009.jpg . . . . .	7
2.4	Binarized Stretched Chromaticity Histogram from Image 833009.jpg . . . . .	9
2.5	Colourized Chromaticity Image from 833009.jpg . . . . .	9
2.6	Sample image: Visual representation of splitting an image into sections using Abstract Color 833009.jpg . . . . .	11
2.7	Diagram Relating Pixel Width Evaluation to Frequency Level of the Wavelet Transform Demonstrating the Dynamic Resolution used in a Wavelet Trans- form. . . . .	13
3.1	Depiction of four Gabor filters evaluated in a single location at $0^\circ$ , $45^\circ$ , $90^\circ$ and $135^\circ$ . . . . .	19
3.2	A 5x5 Gabor wavelet grid evaluating frequency 2 at orientation 0 (horizontal). . . . .	21
4.1	Sample Thumbnails from the Corel Photo Library "Doors of Paris" Directory. . . . .	32
4.2	False positives and negatives from probe image 549001.jpg with a 16x16 grid size . . . . .	36
4.3	False positives and negatives from probe image 549001.jpg with a 10x10 grid size . . . . .	40
4.4	False positives and negatives from probe image 549001.jpg with a 8x8 grid size . . . . .	43
4.5	False positives and negatives from probe image 549001.jpg with a 6x6 grid size . . . . .	46
4.6	False positives and negatives from probe image 549001.jpg with a 5x5 grid size . . . . .	49
4.7	Average percentage return of the top ranked position using the original image as the probe. . . . .	51

4.8	Receiver Operator Curve for Grid Sizes 16x16 and 10x10. . . . .	51
4.9	Receiver Operator Curve for Grid Sizes 10x10 and 8x8. . . . .	52
4.10	Receiver Operator Curve for Grid Sizes 8x8 and 6x6. . . . .	52
4.11	Receiver Operator Curve for Grid Sizes 6x6 and 5x5. . . . .	53
4.12	Sample image: Parisian Door . . . . .	65
4.13	Sample image: Museum Dolls . . . . .	65
4.14	Sample image: Cards . . . . .	66
4.15	Sample image: Duck Decoys . . . . .	66
4.16	Sample image: Easter Egg . . . . .	69
4.17	Sample probe image: 423000.jpg . . . . .	69
4.18	Search Results from probe image: 423000.jpg . . . . .	70
4.19	Sample probe image: 282041.jpg . . . . .	71
4.20	Search Results from probe image: 282041.jpg . . . . .	72
4.21	Sample probe image: 549001.jpg . . . . .	73
4.22	Search Results from probe image: 549001.jpg . . . . .	74
4.23	Sample probe image: 655017.jpg . . . . .	74
4.24	Search Results from probe image: 655017.jpg . . . . .	75
4.25	Sample probe image: 702083.jpg . . . . .	76
4.26	Search Results from probe image: 702083.jpg . . . . .	77
4.27	Sample probe image: 408019.jpg . . . . .	77
4.28	Search Results from probe image: 408019.jpg . . . . .	78

# List of Programs

3.1	Generation of a Set of Filters Forming a Jet . . . . .	22
3.2	Generation of Rotated Planar Wave and Gaussian Coordinates . . . . .	23
3.3	Generation of Planar Wave Values . . . . .	24
3.4	Generation of Gaussian Envelope . . . . .	24
3.5	Generation of Gabor Components . . . . .	25
3.6	Generation of Gabor Magnitudes . . . . .	25
A.1	Matlab Script to Generate Linear Chromaticity Histogram . . . . .	83
A.2	Matlab Script to Generate Spherical Chromaticity Histogram . . . . .	84
A.3	Matlab Script to Generate Stretched Chromaticity Histogram . . . . .	85

# Chapter 1

## Introduction

Image content-based search and retrieval has the potential to be at least as useful, if not more so, than traditional text-based searching. Increases in processing power, bandwidth and storage capability have increased the availability of multimedia data. These collections of multimedia data need to be organized based on content. As image processing is typically computationally expensive, the need for efficient and scalable algorithms to retrieve image content is apparent. An efficient approach is to analyze an image and generate a signature based on distinguishing information. Images are then correlated based on their signatures. This method provides a fast and scalable method of image recognition, because the signature is generated offline as a preprocessing step and stored in a database. Typical image and object recognition algorithms analyze colour and shape information.

A traditional colour-based object recognition approach uses colour histogram information to compare objects. Though colour information is a powerful indicator in object recognition, it has several difficulties. First, this approach discards information about the objects' spatial properties, which is another powerful indicator of object similarity. It is very likely that two different objects could have similar colour decompositions and distinctly different shapes, resulting in an incorrect false positive identification. Second, colour information varies under different lighting conditions and with different cameras. Although the colours of the objects remain the same, the colours captured by the camera can vary dramatically. Fortunately, there has been significant progress in the field of colour invariance to overcome this difficulty. In this study, an illumination-invariant chromaticity method based on work by Drew and Au [7] is used to provide the colour analysis.

Shape-based object recognition compares objects based on their measurable shape. One

edge-based approach is to generate a zero crossing signature. The second derivative of the object's edge is used to reveal the number and placement of the zero crossings (where the second derivative changes sign). Alternatively, zero crossings of curvature can be mapped across scale changes. Another approach is to infer the texture or 3D shape from an image. This is typically done with a form of frequency analysis. Shape-based object recognition has at least one major drawback. Objects that have a complex 3D shape yield a very different shape analysis with only slight changes in rotation, pitch or yaw. In this study, a variation of Wiskott's Gabor wavelet filter based Jets [32] are used to provide the frequency analysis.

There are two main focuses of this paper. First is the improvement of current image recognition and content based image retrieval algorithms by using colour analysis combined with frequency analysis. Colour and shape indicators are used to recognize images, thus improving on the results of either method used individually. Secondly, the algorithm developed is intended to be applied to object recognition and this thesis serves as a proof-of-concept for an object recognition algorithm discussed in section 5.2. Experiments were conducted that show the effectiveness of the method developed for content-based image retrieval, indexing and recognizing images and objects. In addition, the method developed resulted in an efficient signature size and layout, which allowed searching to be performed at a high rate of speed.



## Chapter 2

# Literature Review and Background

### 2.1 Linear Correlation

Linear correlation is the most widely used measure of association between variables that are ordinal or continuous, rather than nominal [26]. Given two arrays,  $x$  and  $y$  of length  $N$ , having pairs of quantities  $(x_i, y_i)$ ,  $i = 1, \dots, N$ , the linear correlation coefficient  $r$  is given by the formula:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}} \quad (2.1)$$

where  $\bar{x}$  is the mean of the  $x_i$  and  $\bar{y}$  is the mean of the  $y_i$ .

The value of  $r$  is in the range  $[-1,1]$ . Complete positive correlation, with  $r = 1$ , occurs when the data values lie on a perfectly straight line with a positive slope (with  $x$  and  $y$  increasing together). Similarly, a value of  $-1$ , complete negative correlation, occurs when the data values lie on a perfectly straight line with a negative slope (with  $x$  and  $y$  decreasing together). The closer  $r$  is to zero, the more uncorrelated the data values are.

### 2.2 Colour Analysis

In this section, colour analysis is discussed starting with the seminal work of Swain and Ballard [29]. Subsequently, some techniques for the colour constancy problem are covered, as well as the use of chromaticity histograms and their ability to be significantly compressed.

### 2.2.1 Colour Histograms

Research on image content indexing and retrieval started by focusing on extraction and correlation of global image feature vectors. In one of the early works on image retrieval, Swain and Ballard [29] used *histogram intersection* to correlate colour histograms of two images. First, a colour histogram  $\mathbf{H}_i$  is generated for each image  $i$  in the database. The histogram is then normalized, and stored in the database. For a model image from the database, its histogram  $\mathbf{H}_m$  is intersected with all database image histograms  $\mathbf{H}_i$  according to the equation:

$$\sum_{j=1}^n \min(H_i^j, H_m^j), \quad (2.2)$$

where superscript  $j$  denotes histogram bin  $j$ , and each histogram has  $n$  bins. The closer the intersection value is to 1, the better the images match. Computing the intersection value is fast, but it is sensitive to colour quantization. Moreover, there is a problem that arises because of the effect of changing illumination on images of colour objects [12] or images of coloured objects captured with a different camera [22]. This is the colour constancy problem.

### 2.2.2 Illumination

In an attempt to address the colour constancy problem on images of coloured objects, Drew et al. [9] perform a normalization step on the colour channels in an image. The normalization step was originally the first step in the ‘‘Colour Angles’’ method of Finlayson et al. [13]. Finlayson et al. perform this normalization step followed by subtracting the mean from the images. Drew et al. [9] demonstrate that the normalization step promotes the illumination invariance of the method and that one can do better if further information is preserved.

Other methods exist for the complex problem of colour constancy; however the method of Drew et al. [9] is a very simple and has been shown to adequately discount illumination for many tasks [22, 7, 4, 8].

### 2.2.3 Chromaticity Histograms

Chromaticities are ratios of colours. Linear chromaticity, for example, is calculated using Equation 2.3.

$$(r, g) = (R, G)/(R + G + B) \quad (2.3)$$

As an alternative to using histogram intersection [29] on the normalized colour channels, chromaticity histograms can be used that still capture the essential colour information and sacrifice only image intensity [9]. Chromaticity histograms have been shown to be an improvement on colour histograms [9]. There are two advantages of using a chromaticity colour space. First, chromaticity colour space reduces the dimensionality of colour to 2, which among other things significantly reduces the size of the histogram. Second, because chromaticity is a ratio of colour bands, it has the effect of removing shading [7], which contributes to illumination invariance - the ability to retrieve the same colour information from objects under different lighting or capturing conditions.

The chromaticity histograms, being 2 dimensional, can be viewed as images. As such, they can be compressed as images by the use of wavelet compression followed by a second step of going to the frequency domain and keeping only the low frequency DCT coefficients [9]. This histogram compression is desirable as the histogram is typically sparse and this method preserves most of the distinguishing information [9]. Originally, Drew et al. introduced this technique on chromaticity histograms with normalized bin counts for size invariance [9]. It has been applied successfully in object recognition [21] and multimedia applications [22, 7].

Later, an additional step was added whereby the chromaticity histograms were first binarized [4, 8]. This technique attempts to concisely record the chromaticities in an image in the form of a signature, where the chromaticity bin counts are binarized as is shown in Fig. 2.2. Instead of counting the actual chromaticity bin counts, only the existence of a chromaticity value is recorded by producing a binary image where the pixel values either confirm or deny the existence of a particular chromaticity value within the image. Appendix A.1 shows the implementation of this idea in Matlab. This approach limits the information stored in the image signature and lends itself well to image compression. Moreover the binarized chromaticity signature is like a statement that a particular colour is or is not in the image colour palette much like image file formats such as GIF that use a colour palette.

The calculation of the linear chromaticity in Equation 2.3 introduces a problem. As ratios are not evenly distributed, they do not fully utilize the evenly spaced histogram bins. Drew and Au [7] point out that because linear chromaticity obeys  $r + g \leq 1$ , there exists a straight diagonal edge in a chromaticity space histogram. The binarized chromaticity space



Figure 2.1: Sample image: Abstract Color 833009.jpg  
 Image is from the Corel Gallery and is copyright Corel. All rights reserved.

histogram generated from Figure 2.1, using the linear chromaticity computation, is shown in Figure 2.2. The above mentioned diagonal edge is readily apparent in the image.

To overcome the negative effects of the diagonal edge, Drew and Au used a spherical chromaticity space of the form described in Equation 2.4, to mitigate a ringing effect in the Fourier domain caused by the diagonal edge. Spherical chromaticity does not eliminate the edge, but improves matters by replacing it with a circular edge as shown in Fig. 2.3. In addition, spherical chromaticity space improved upon the linear model by utilizing more of the histogram bins. Fig. 2.3 is the binarized chromaticity histogram that was created using the spherical chromaticity computation with  $p = 2$  in Equation 2.4. The Matlab script implementation of generating a binarized spherical chromaticity space histogram is shown in Appendix A.2.

$$(r, g) = (R, G) / \sqrt[p]{R^p + G^p + B^p} \quad (2.4)$$

Here, a new technique, proposed in [1], is used. This approach introduces the use of a 2D *stretched* chromaticity space similar to that in [24], shown in Equation 2.5, which utilizes all

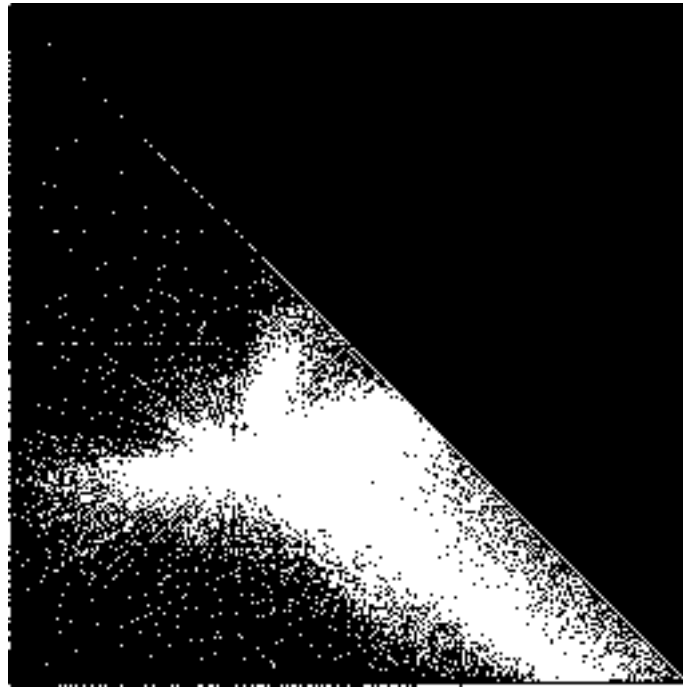


Figure 2.2: Binarized Linear Chromaticity Histogram from image 833009.jpg

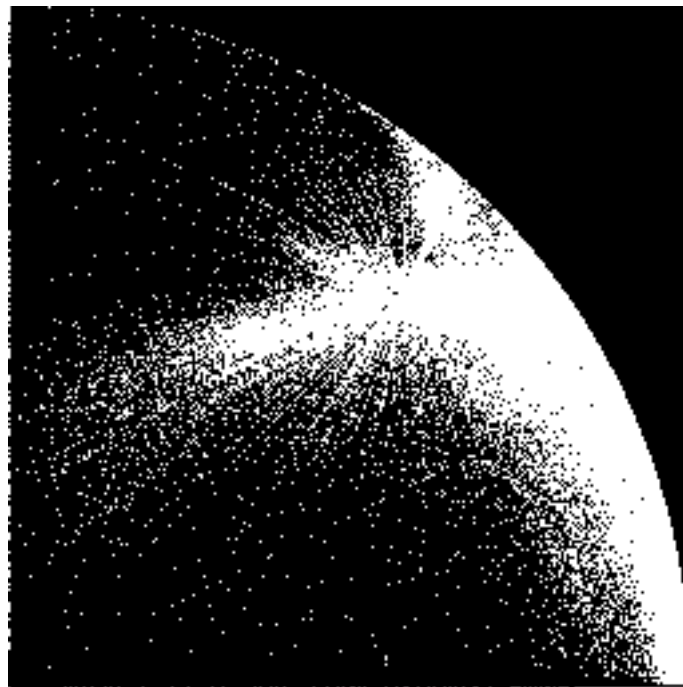


Figure 2.3: Binarized Spherical Chromaticity Histogram from Image 833009.jpg

of the histogram bins and eliminates the edge effect altogether.<sup>1</sup> Using the stretched chromaticity is a reasonable approach as any edge in the chromaticity histogram compromises the subsequent DCT very strongly. Chromaticity by itself is already a highly nonlinear transformation, in that it is a projective transform of RGB. The stretching operation does not change the proximity relationships of colours, but simply fills up the 2D colour space. Utilizing all the histogram bins has the desired effect of eliminating the edge, but also eliminates the inefficient use of space taken up by the empty histogram bins. Figure 2.5 shows the sample image, Figure 2.1, in a colourized form. This form is achieved by computing the stretched chromaticity values using Equation 2.4, then using the  $r$  and  $g$  chromaticities as  $R$  and  $G$  and then setting  $B$  equal to 1. Figure 2.5 shows that some different RGB colours have approximately equal chromaticities. For example, consider a white area and a black area in Figure 2.1, and then examine the corresponding areas in the chromaticity image, Figure 2.5. The original white and black colours are now the same, or very similar, chromaticity. This is the effect of removing the dependence on lighting: White and black are essentially the same colour with the difference being the illumination intensity.

Fig. 2.4 is the binarized stretched chromaticity histogram generated from Fig. 2.1. The Matlab script implementation of generating a binarized stretched chromaticity space histogram is shown in Appendix A.3.

$$(r, g) = (R, G)/(R + G + B)$$

$$r = \begin{cases} r + g & : \text{if } r \geq g \\ 2r & : \text{otherwise} \end{cases} \quad (2.5)$$

$$g = \begin{cases} 2g & : \text{if } r \geq g \\ r + g & : \text{otherwise} \end{cases}$$

## 2.3 Spectral Analysis

The motivation behind using a wavelet based approach for the spectral analysis is discussed in this section. This section begins by covering transforms in general, as an appreciation for

---

<sup>1</sup>This chromaticity space is like that in [24], but properly produces the range  $\{[0..1],[0..1]\}$ , rather than  $\{[0..1],[0..2]\}$ .

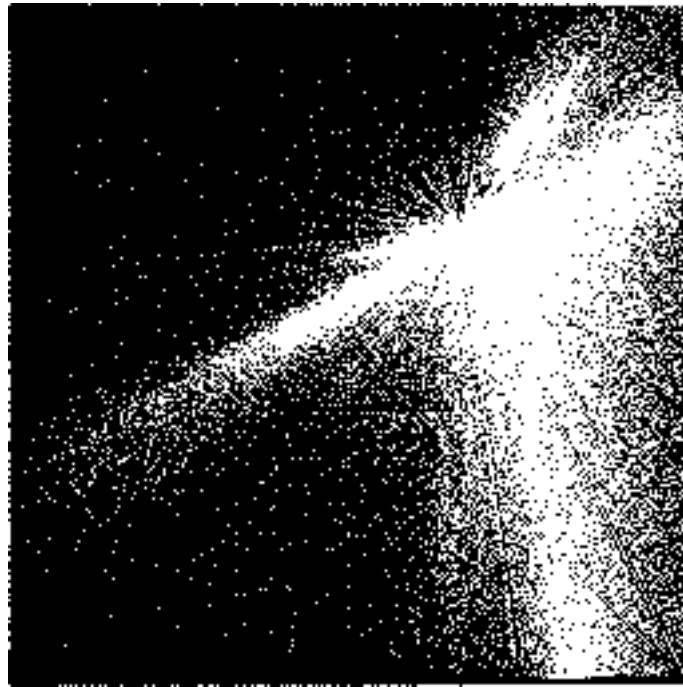


Figure 2.4: Binarized Stretched Chromaticity Histogram from Image 833009.jpg

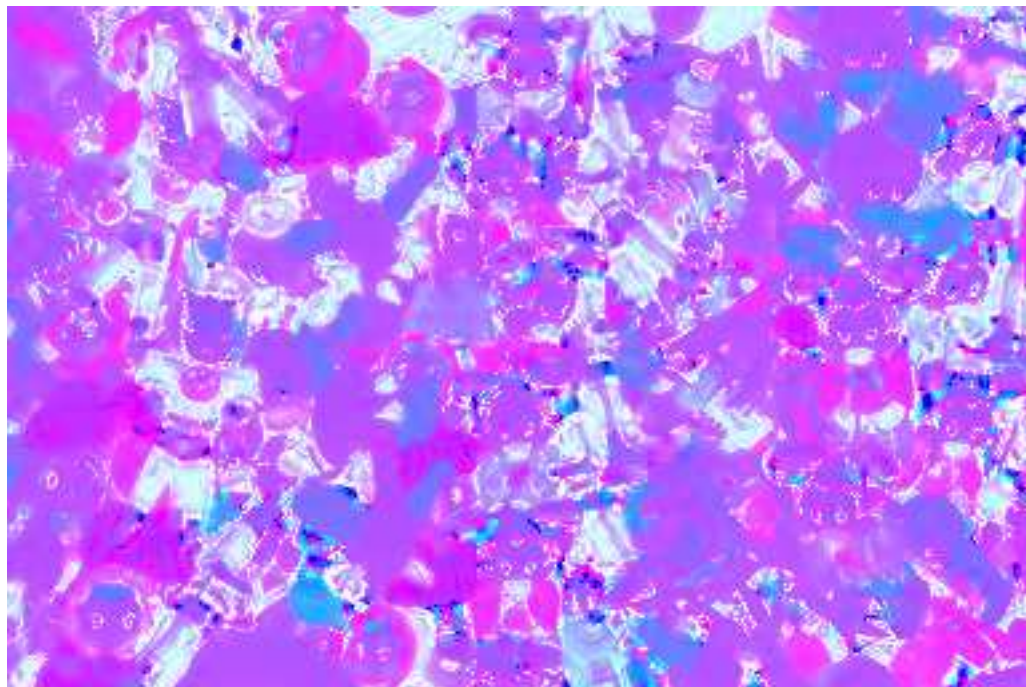


Figure 2.5: Colourized Chromaticity Image from 833009.jpg

the basic ideas and problems is necessary in understanding why wavelets are used. Moreover, the choice of using Gabor Wavelets in particular is also covered.

### 2.3.1 Transforms

Transforms are a method to convert raw signals into another domain where additional information may be apparent or further operations can be performed. Commonly, signals are in the time domain, where the signal varies with respect to time; an example is an electrocardiography (ECG) signal. Signals can also be in a spatial domain, as in the case of imagery. There exist many types of transforms available for a wide range of applications, each having benefits and drawbacks. The most common transform is the Fourier Transform.

When comparing various transforms, it is important to understand the difference between stationary and non-stationary signals. Stationary signals are signals for which the frequency components in the signal exist throughout the entire signal and do not start and stop over time. Many signals are non-stationary. A greyscale image is an example of a 2D signal where the frequency components can vary over the image extents. The Fourier Transform is generally more applicable to stationary signals. The Fourier Transform can be used on images where the signal is non-stationary, however, if one is only concerned with the frequency components within the image and not where in the image those frequency components occur. The implication of this is that a Fourier Transform will only provide information about what frequencies exist and will not reveal at what locations they existed. In imagery, the frequency components within a scene can vary dramatically over the image extents and can be a significant source of distinguishing information. With this idea in mind, a method was devised whereby the frequency components within discrete locations could be compared directly to the frequency components from the same areas in other images. The objective of performing this kind of analysis is to strengthen the discriminatory power of the spectral analysis and add the ability to find similar images based on the spatial distribution of frequency components. This increases the recognition ability of the algorithm and its ability to find similar looking images.

Applying a Fourier Transform to an entire image does not reveal any spatial information identifying frequency components that exist in that image at certain locations. This spatial component could be achieved by dividing the image into sections (windows) and using a Fourier transform on each section, as depicted in Fig. 2.6. In this case, a fixed resolution (image area) is chosen and the image is transformed in areas instead of as a whole. This





Figure 2.6: Sample image: Visual representation of splitting an image into sections using Abstract Color 833009.jpg

approach would reveal more information about what frequency components exist within each image area. However, this approach has a problem that is introduced by the fixed resolution that is chosen. For any given resolution or image area, because of the Nyquist theorem, the lower frequencies have sufficient samples, but the higher frequencies do not. This allows the lower frequencies to be better resolved than the higher frequencies. Wavelets were developed as an alternative to this approach as they, in part, overcome this problem of resolution.

Wavelets handle the resolution problem through the use of a Gaussian envelope, which is used as an effective window on the signal. The Gaussian envelope is parameterized by the frequency, so that lower frequencies have a narrower window than higher frequencies. This allows the higher frequencies to have the additional samples they need in order to be resolved, and at the same time restricts the samples for the lower frequencies that are not needed in order for the lower frequencies to be resolved.

### 2.3.2 Wavelet Transforms

Wavelets are a relatively new tool for transforms that partially avoid the problem of resolution that other transforms have when dealing with non-stationary signals. In accordance

with Heisenberg's uncertainty principle, we cannot discern what spectral components exist at any given point or instant. We can, however, determine what spectral components exist over a given interval. Choosing this interval can introduce a resolution problem for transforms that have fixed resolutions. Smaller intervals have better time resolution, but poorer frequency resolution. Wider intervals have better frequency resolution, but poorer time resolution — there is a tradeoff. Wavelets have variable resolutions that make them more desirable to use in this scenario. For low-frequency spectral components the spatial resolution is increased and the frequency resolution decreased. For higher-frequency spectral components, the spatial resolution is decreased and the frequency resolution is increased. In this manner, the wavelet transform adapts in order to mitigate the effects of the resolution problem. Figure 2.7 depicts this graphically with the frequency resolution on the  $y$  axis and the spatial resolution on the  $x$  axis.

### 2.3.3 Gabor Wavelet Filter

The use of the 2D Gabor filter in computer vision was introduced by Daugman in the late 1980s [6, 5]. Since that time it has been used in many computer vision applications including image compression [5], edge detection [19], texture analysis [20], object recognition [14] and facial recognition [17, 16, 33, 32, 31].

The general form for a complex-valued 2D Gabor function is a planar wave attenuated by a Gaussian envelope:

$$\Psi(x, k, \sigma) = \frac{k^2}{\sigma^2} \exp\left(-\frac{k^2 x^2}{2\sigma^2}\right) [\exp(ikx) - \exp\left(-\frac{\sigma^2}{2}\right)] \quad (2.6)$$

In order to render the filters insensitive to the overall level of illumination, the term  $\exp\left(-\frac{\sigma^2}{2}\right)$  is subtracted. The multiplicative factor  $k^2$  ensures that filters tuned to different spatial frequency bands have approximately equal energies.

### 2.3.4 Jets

Wiskott uses jets extensively in his facial recognition and scene analysis applications [34, 33, 32, 31]. Jets are a grouping of wavelets in varying orientations and frequencies evaluated at a single point. A jet is a condensed and robust representation of a local grey value distribution, termed a *local expert* [32]. Wiskott further describes a jet as being based on a Gabor wavelet transform, which is a convolution with a family of complex Gabor wavelets

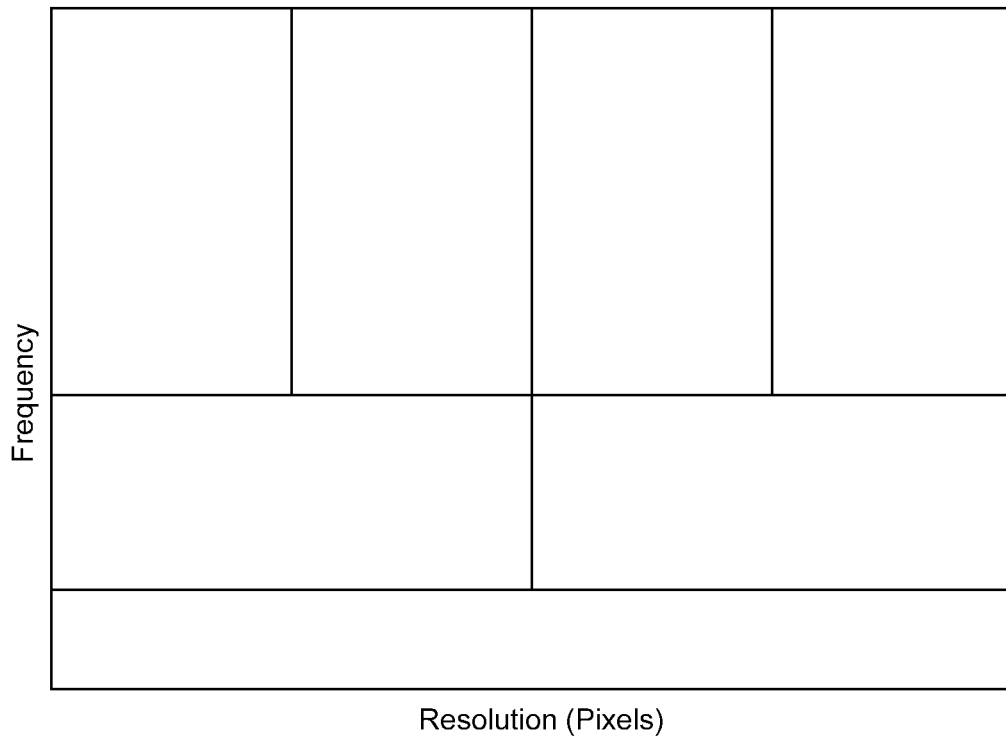


Figure 2.7: Diagram Relating Pixel Width Evaluation to Frequency Level of the Wavelet Transform Demonstrating the Dynamic Resolution used in a Wavelet Transform.

having the shape of plane waves restricted by a Gaussian envelope function. The wavelets are similar in the sense that they can all be generated from a mother wavelet by rotation and scaling. All complex coefficients of the transform taken at one image location form a jet.

A small displacement may lead to very different coefficients. However, the magnitudes vary slowly and can be used directly for comparison [32]. Wiskott also states that jets are robust with respect to illumination variations, scaling, translation and distortion. Another benefit of Gabor wavelets is that they are a good model for the receptive fields of complex cells in primary visual cortex of primates [32].

### 2.3.5 Gabor Filter Frequencies

When performing a frequency analysis it is important to consider what frequencies should be evaluated. Low frequency information remains more stable across images. Nestares et al. [23] selected  $\frac{Nyquist}{2}$  as the highest central frequency of their implemented Gabor filter banks. In accordance with Nyquist's theorem, lower resolution images have fewer useful frequencies that can be used. High frequency information has been shown to be a differentiating factor in texture analysis in high-resolution images [20], but the majority of images are not high resolution. In addition, high frequency information in images is often associated with edges and noise [28] and we are not directly concerned with edge information. For this implementation, we therefore use only the low frequency information.

## 2.4 Combined Colour and Spectral Analysis

Using a combination of colour, texture and shape in Content-based Image Retrieval (CBIR) systems has been attempted in a variety of ways. Systems covered in this section include IBM's QBIC system [18], Blobworld [3], C-BIRD [35], a project by Tian et al. [30] and an earlier iteration of the algorithm presented here [1]. Other systems include Virage [2], Photobook [25] and Amore [27].

IBM has developed the QBIC (Query By Image Content) system [18] that indexes colour using colour histogram distances measured in two ways. The low-dimensional colour histograms are matched using a weighted Euclidean distance measure, which acts as a filter for the more comprehensive quadratic histogram distance. It is shown that using this match measure it is possible to first prune histogram matches by using a lower bound on the match

measure of highly reduced dimensional histograms. QBIC also measures texture features including coarseness, contrast and orientation. Shape features are made up of shape area, circularity, eccentricity, major axis orientation and a set of algebraic moment invariants. Texture and shape features are also matched using a weighted Euclidean distance calculation.

Belongie et al. [3] developed the image representation called “Blobworld” along with an image retrieval system based on that image representation. The system allows a user query for an image, or any number of objects within an image, by selecting a region for use as a sample image. Another significant feature of the “Blobworld” system is that the user can view the internal representation of the probe image and that of the query results. This allows for the refinement of queries and produces a certain level of user satisfaction as typical image retrieval systems return unintuitive results that are confusing to the casual observer. Segmentation in “Blobworld” is accomplished by the Expectation-Maximization (EM) algorithm that is used to group a large set of 6-D feature vectors. The algorithm first determines the polarity measure at every pixel location at various scales. The first scale in which the polarity doesn’t change (difference between successive values is  $< 2\%$ ) is the chosen one. Once a scale has been chosen, a 6-D vector is created for every pixel made up of three texture descriptors (orientation, anisotropy, and contrast) and three colour descriptors (based on the HSV colour space). Pixels are grouped based on the 6-D feature vectors using the Expectation-Maximization algorithm to determine the maximum likelihood parameters of a mixture of  $K$  Gaussians. The image matching score is the Mahalanobis distance between feature vectors, and is similar to the QBIC colour histogram matching for colour vectors. Experiments are performed on 2,000 natural images selected from the same commercial Corel photo collection used in Chapter 4. Their method typically outperforms global histogram matching, however, the precision reported is very low.

The C-BIRD system generates a feature descriptor and a layout feature for each image in the database. There are four vectors in the feature descriptor. Two of the feature vectors are a 512-bin histogram colour vector and the chromaticity vector found in [9]. The other two feature vectors are the centroids of the five most frequent colour regions and the centroids from the regions of the five most frequent edge orientations. The layout feature contains colour layout and edge layout vectors constructed from determining the most frequent colours and the number of edges for each orientation in each of the 64 image tiles. There are a variety of matching techniques used in C-Bird depending on the feature

being measured. The distance between chromaticity vectors is the  $L_2$  distance and histogram intersection is used for the texture orientation histograms and colour histograms. When querying for an object Li et al. [35] first localize the regions of the most frequent colours as well as computing the area, the centroid and the eccentricity for each region. Using this as a filter, images are selected that share a number of colour regions with the query image. This generates a short list of similar objects which is then subjected to texture and shape similarity matching. The 2D texture histogram measures orientation and edge separation from the grey level image. A shape verification based on the Generalized Hough Transform is performed if there is sufficient similarity in the texture between the query object and the area in the short list of database images where the similar objects were identified.

Tian et al. [30] developed a CBIR system using wavelets. Their approach was to detect the salient features in the image using a Haar wavelet-based salient point extraction algorithm. Colour features and texture features are extracted at the salient points. Colour features are extracted by interpreting the colour distribution of an image as a probability distribution, which can then be characterized by its moments. The first three low-order moments in the HSV colour space make up the 9 dimensional colour feature vector. Texture features are extracted by considering the Gabor filters as orientation and scale adjustable edge and line detectors. Statistics were generated for those microfeatures at the salient points and the low-order moments used to characterize the texture information.

There are two main differences in the use of wavelets here from that in the Tian et al. [30] paper. First, they use Haar wavelets to detect the salient points in the image. Typically they detected 50-100 such points. For a simplification in this study, we do not attempt to determine the areas of distinguishing information, but rather simply use a coarse grid in which to evaluate the wavelet filters. Second, although they use Gabor filters to extract the texture features from the point locations, they use the coefficients to generate statistics on the microfeatures present. Those statistics are then characterized by the low-order moments. This study uses the magnitudes generated from the Gabor filters. In facial recognition studies [32], Wiskott points out that magnitudes vary slowly and can be used directly for comparison.

An earlier iteration of the algorithm presented here appeared in [1]. There are three main differences that have been implemented in this study from the previous iteration. The colour analysis now includes further refinement and compression. In [1] only global colour information was retrieved. In the new algorithm colour information is extracted separately

from each of the 25 tiles in the image. Moreover, instead of using the chromaticity bin counts for comparison, the chromaticity histogram is now binarized and resized into a 16 by 16 pixel image. The image is then moved into the frequency domain via a DCT and the top 8 coefficients are selected. This compression technique allowed for the additional granularity in the colour analysis. The second difference is in the spectral analysis. In [1], two wavelet grids were used. A 10 x 10 wavelet grid was overlapped by a 9 x 9 wavelet grid and at each location the Gabor filters were only evaluated at 1 or 2 frequencies and in only two orientations. This configuration was chosen empirically and has since been refined to a single 5 x 5 wavelet grid, evaluating the Gabor filters at each location in four orientations at four different frequencies in each orientation. This configuration was the result of extensive testing by iteratively making improvements and measuring the increasing accuracy by the use of a benchmark. The third change was in the benchmark used to test the similar image retrieval capabilities. Previously there was only one data set used for both the benchmark results and the similar image finding capabilities. In this study a new dataset was used to test the similar image finding capabilities of the algorithm. The new data used is the commercially available Corel photo collection. The entire collection is included, where as in the previous iteration only a subset of the collection is used.

The work presented here is similar to that in Liapis and Tziritas [15] in that their paper covers image retrieval based on a combined approach of using colour and texture features using chromaticities and wavelets<sup>2</sup>. This work differs, however, in the methods used to extract and produce these feature vectors. The colour features in Liapis and Tziritas are described in terms of 2D or 1D histograms of CIE Lab chromaticity coordinates. The work presented here uses 2D stretched chromaticity histograms, based on the RGB colour space, for which all bins are populated. Texture information in Liapis and Tziritas' paper is extracted using a Discrete Wavelet Frames analysis, where they determine characteristics corresponding to each texture type, so they can uniquely define each texture pattern. The work presented here accomplishes feature extraction in a single step by evaluating Gabor wavelet filters without the initial step of determining where to evaluate the filters; the placement of the filters is made to be function of the size and aspect ratio of the image. This approach offers an attractive degree of simplicity, while retaining a high level of discriminatory power between images.

---

<sup>2</sup>Note that the work here was first reported in [1]; over a year before that in [15].

## Chapter 3

# Procedure

### 3.1 Frequency Analysis

For this application, a variation on Wiskott's jet was used that evaluated the Gabor filter in four different directions. The four directions were at 0 degrees (horizontal), at 45 degrees, at 90 degrees (vertical), and 135 degrees. Figure 3.1 depicts the idea behind the construction of a jet: Applying the same size filter in multiple orientations at a single location. The jet in this study used the four lowest frequencies at each of the four orientations. The purpose is to sample the significant frequency components in the four major directions. In this way the magnitudes computed from the evaluation of the wavelets becomes a record of the major frequency components in a specific location. By evaluating four frequencies at each of four directions, the feature vector retrieved from each image location contains 16 coefficients. By evaluating the jets in multiple locations spanning the entire image, we retrieve stable and significant information in the feature vectors that can be compared to the feature vectors from other images.

The jets must be evaluated at locations that will provide statistically significant features with high discriminatory power. As image content is not known *a priori*, the features with discriminating power also cannot be known beforehand. Therefore, in order to gather enough distinguishing information, locations covering the image extents were sampled. Four considerations were taken into account when devising this method:

1. As the content of the image was not known beforehand, no assumptions could be made as to the placement or orientation of the distinguishing content.



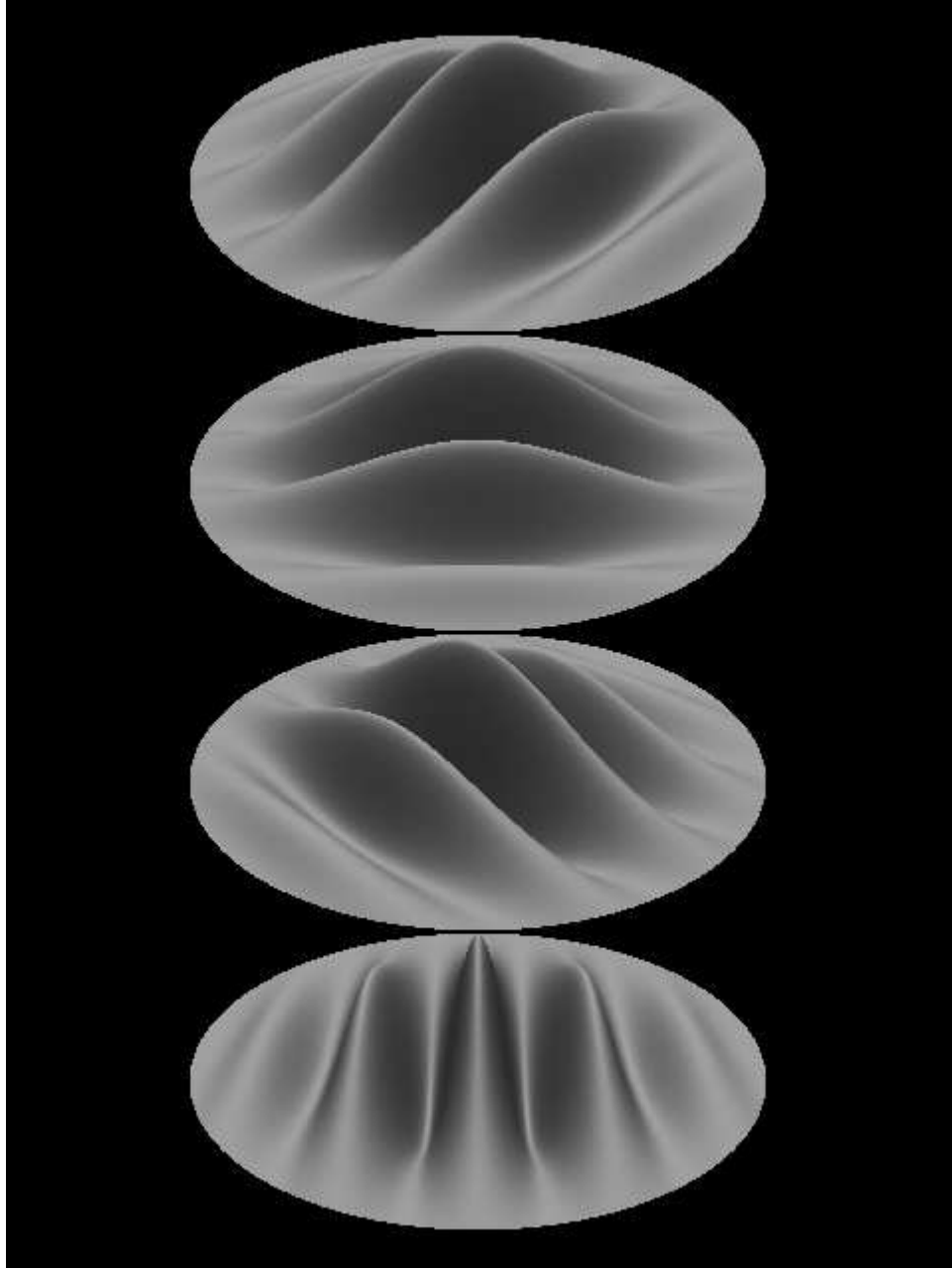


Figure 3.1: Depiction of four Gabor filters evaluated in a single location at  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  and  $135^\circ$ .

2. The wavelet placement must be predictable and repeatable.
3. The wavelet placement method must be size invariant as the image dimensions can vary.
4. The wavelet placement must be invariant to aspect changes in the imagery.

To achieve goal 1, a uniform grid of Gabor filters is evaluated over the entire image so as to retrieve feature information from the entire image. To achieve goal 2, the placement of the filters were made a function of the image extents. This makes the placement of the wavelets consistent as they are based on the image's dimensions and not the image content. In addition, by making the shape and size of the filters a function of the image extents as well, the filter size and shape were proportional to the size and shape of the image. This allows the wavelets to evaluate similar magnitudes as the wavelets cover the same area no matter how many pixels make up that area. This satisfies both goals 3 and 4.

Once the methodology of placing and sizing the filters was decided upon, the question of how many filters that should be used still needed to be answered. Various uniform grids with differing numbers of filters are attempted. Below, the evolution of the frequency algorithm is followed, with grid sizes of 16x16, 10x10, 8x8, 6x6 and 5x5 being tested. Each variation is benchmarked against a dataset in order to determine the effectiveness in retrieving the same source images and its ability to retrieve images with similar content. The results of these benchmarks are discussed in chapter 4. Ultimately a grid size of 5x5 was chosen as it was reasonably stable under minor spatial shifts allowing the same source images to be found with high accuracy, while still maintaining sufficient spatial constraints to aid in retrieving images with similar content. An image depicting the highlighted wavelets in the four corners of the 5x5 wavelet grid is shown in Figure 3.2. All 25 wavelets are shown, overlaid in Figure 3.2 — the image is layered with many wavelets so only a part of each wavelet can be seen. The wavelets are made to overlap following Wiskott [32]. If the wavelets did not overlap, then the areas that correspond to the highly attenuated edges of the wavelet would contribute little to the recognition of the images or objects. Figure 3.2 shows prominently the placement of the four corner wavelets in the image over a faded background comprised of all the wavelets on the image. The wavelets in the corners are displayed in Figure 3.2 for clarity. As they overlap and occlude each other, it is impossible to show all of the wavelets at the same time.

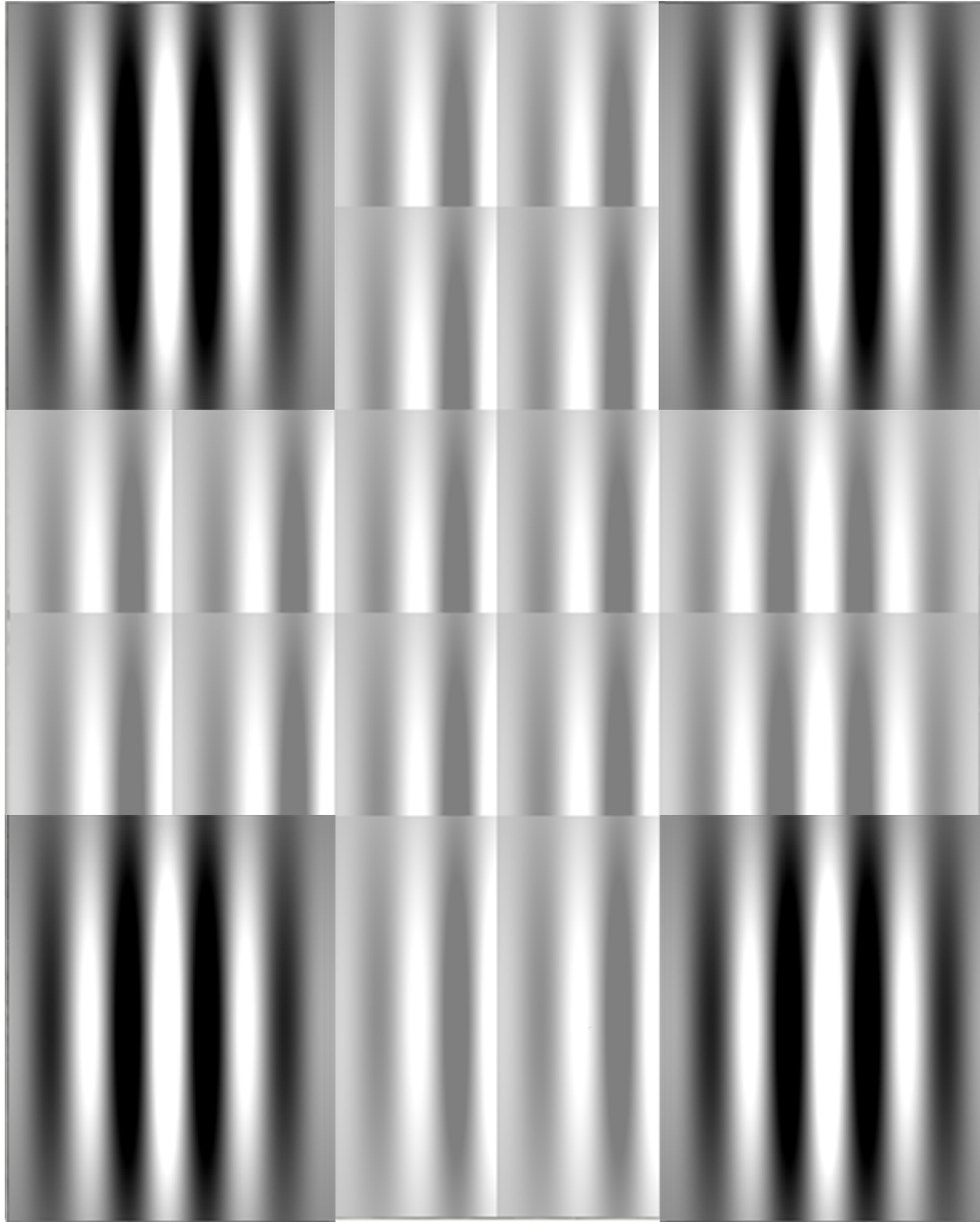


Figure 3.2: A 5x5 Gabor wavelet grid evaluating frequency 2 at orientation 0 (horizontal).

In order to accurately describe the implementation of the Gabor wavelet, the steps taken to compute a Gabor wavelet filter magnitude are described in detail. This section takes the reader through a complete algorithm, starting from the original image, and proceeding all the way down to the generation of the first Jet. For this section, Figure 2.1 is used.

When performing the spectral analysis on an image, the system first converts the image to greyscale as the wavelet filters work on intensity values, or greylevel values. The system analyzes the image dimensions in order to determine where to place the Gabor Jets to achieve the 5x5 grid. Figure 2.1 is 384 pixels in width and 256 pixels in height. With a 50% overlap in the Jets, this means that the system must evaluate a Jet in the 25 positions listed in Table 3.1, with each jet having a size of 128 pixels in the x direction and 84 pixels in the y direction; which gives the Jets the same aspect ratio as Figure 2.1.

After determining the size, shape and placement of the Gabor Jets, the system can move on to evaluating each Jet. When evaluating a jet at a specific location, the algorithm starts off by looping over the number of frequencies and the number of orientations as shown in Program 3.1. This will result in 16 calls to the function “EvaluateSingleFilter” for each Jet. The following paragraph will go into detail on what happens during the first call to this function.

```

iFirst0 = 0;
iLast0  = 3;
iFirstF = 1;
iLastF  = 4;

// Generate a Gabor Jet for this feature location according to
// the initialized number of orientations and frequencies

for( i0 = iFirst0; i0 <= iLast0; i0++)
{
    for( iF = iFirstF; iF <= iLastF; iF++ )
    {
        pdDescriptors[iCount++] = EvaluateSingleFilter(i0, iF);
    }
}

```

Program 3.1: Generation of a Set of Filters Forming a Jet

The first time “EvaluateSingleFilter” is called it is tasked with evaluating the Gabor

filter in orientation 0 (horizontal) and frequency 1. The frequency number refers to the number of complete planar wave cycles that cover the width of the wavelet filter. The four orientations correspond to the angles 0, 45, 90 and 135.

When implementing the Gabor wavelet filter, described in Equation 2.6, for use at different orientations, the waveform is fixed in the x direction and the x-y sampling area is rotated. In order to calculate the planar wave value at any point for the current orientation angle, the algorithm rotates the x-y points instead of rotating the waveform — which ultimately is the same thing. Rotating the x-y points around the center of the wavelet is done by using a simple 2D transform shown in Program 3.2. `dOri` is the wave propagation direction in radians, `(iX,iY)` is the current pixel location and `(iXC,iYC)` is the center of the wavelet. The rotated point, `(dRotX,dRotY)` is used in calculating the planar wave and gaussian values.

```
dRotX = (double)(iX-iXC)*cos(dOri) + (double)(iY-iYC)*sin(dOri);
dRotY = -(double)(iX-iXC)*sin(dOri) + (double)(iY-iYC)*cos(dOri);
```

Program 3.2: Generation of Rotated Planar Wave and Gaussian Coordinates

The system iterates over each pixel in turn. The first pixel in this image has intensity 171, which is read directly from the image. For each pixel, the system needs to determine the planar wave value for that position. This value depends not only on the pixel location, but the wavelet rotation as well. The first time “EvaluateSingleFilter” is called with orientation 0 and frequency 1, the values of `dRotX` and `dRotY` are equal to `iX-iXC` and `iY-iYC` respectively, as there is no rotation in the wavelet at orientation 0. If there were a rotation angle, the values of `dRotX` and `dRotY` would be different and they would correspond to the position on the planar wave that this particular pixel would lie on, if it was indeed rotated. Once the correct planar wave position is computed, the planar wave value at this position can now be computed using the code snippet shown in Program 3.3. `dFreq` is the frequency of the waveform, `dRotX` is the rotated x-y point calculated above, `dRadWaveCov` is the radial coverage of the waveform in the x direction and `2*PI` converts the `dRotX/dRadWaveCov` ratio into radians. With no rotation, the radius of the filter, `dRadWaveCov`, is equal to half of the width of the filter as determined above. When the wavelet rotates, `dRadWaveCov` will change and the purpose of the code snippet listed in Program 3.3, is to always maintain the exact number of planar wave cycles `dFreq` dictates

```

dWaveMod = 2*PI*dFreq*dRotX/dRadWaveCov;
dPWaveCos = cos(dWaveMod);
dPWaveSin = sin(dWaveMod);

```

Program 3.3: Generation of Planar Wave Values

```

dGauss = dFreq * exp(-0.5*( dRotX*dRotX/dSigX2 + dRotY*dRotY/dSigY2));
dGaussSum += dGuass;

```

Program 3.4: Generation of Gaussian Envelope

in the horizontal direction. For our current example, `dRadWaveCov` is 64 and `dFreq` is 1. As we are starting at the edge of the filter, `dRotX` is `-64`, which is the far left edge. Finally, `dRadWaveCov` is 64 as the radius in the  $x$  direction is half the width of the filter when the filter is not rotated. When we run the values through the code snippet in Program 3.3, we end up with the value of  $2\pi$  for the variable `dWaveMod`. Intuitively, this is correct as this is the first pixel and it is at the edge of the planar wave. Program 3.3 finishes for this pixel by calculating the sin and cos values for `dWaveMod`. Now that the pixel value and the planar wave value is known at this location, the only remaining item to compute is the gaussian value at this pixel location. As this is an edge pixel, we would expect a small gaussian value here as the farther a pixel is from the center, the less it contributes to the result of the filter.

The jet uses  $\sigma_x$  and  $\sigma_y$  to determine the gaussian width in the  $x$  and  $y$  directions. In many applications, these two values are equal, and hence yield a circular filter.  $\sigma_x$  and  $\sigma_y$  are used to produce elliptical shaped filters, *with the same aspect ratio* as the images on which they were evaluated. Elliptical filters are used to analyze foreshortened objects or texture patterns. Images that have been stretched, or have otherwise had the aspect ratio altered, are analogous to a foreshortened view. Binding the shape of the filter to the aspect ratio of the image produces an algorithm that is invariant to aspect changes. At each rotation the waveform is stretched or compressed to ensure the foreshortened x-y information generates the same coefficient magnitudes as the unforeshortened view. This was accomplished by the C++ code snippets shown in Program 3.3 and Program 3.4.

The planar waves are attenuated by a gaussian envelope generated by Program 3.4.

```
dGaborReal += dGauss * dPWaveCos * (double)iIntensity;
dGaborImag -= dGauss * dPWaveSin * (double)iIntensity;
```

Program 3.5: Generation of Gabor Components

```
dGaborMag = (hypot(dGaborReal, dGaborImag) / sqrt(dGaussSum));
```

Program 3.6: Generation of Gabor Magnitudes

After having computed the planar wave and gaussian values we can calculate a running sum of the Gabor components which is the product of pixel intensity value multiplied by the planar wave attenuated by the gaussian envelope. The C++ code snippet is shown in Program 3.5. `dGaussSum` is used later when we need to normalize the Gabor magnitude.

When the algorithm is finished evaluating the filter we compute the magnitude as the hypotenuse of the real and imaginary components and then normalize the Gabor magnitude by dividing by `dGaussSum`. The C++ code snippet in Program 3.6 shows how this is accomplished.

The  $\exp(-\frac{\sigma^2}{2})$  factor, listed in Equation 2.6, was not included. This factor is usually subtracted to yield a zero DC response, since the integral of the cosine is always larger than the sine. This factor makes the response magnitudes invariant under different lighting contrast conditions. Since the outer ends are severely attenuated by the gaussian, this factor will have little effect on the overall response.

## 3.2 Colour Analysis and Illumination Invariance

The colour analysis was performed using a method similar to a method used by Drew and Au [7]. As with their method, the colour channels were normalized before moving into chromaticity space. This greatly attenuates dependence on both luminance and lighting colour [7]. Normalization is accomplished by first dividing each colour channel by its mean, then normalizing each pixel's RGB colour vector to length 1 by dividing by the square root of the sum of the squares of the RGB values. Iterating in this manner has been shown to provide convergence after five iterations [10, 11]; however, good results can still be achieved in one iteration with far less computation.

In order to increase the algorithm's similar-image finding ability, the colour analysis was performed on each of the same 25 image tiles as in the spectral analysis. This added a spatial component to the colour analysis giving the algorithm the ability to not only find images with similar colours, but also add discrimination for finding images having the same colours in similar locations.

For each image tile, the raw chromaticity values are used to create a binarized 2D stretched-chromaticity histogram using the Matlab script shown in Appendix A.3, which was based on Equation 2.5. The 2D stretched-chromaticity histogram was then reduced into a 16x16 image by means of an image resizing operation. The resized 2D histogram was then compressed by means of a 16x16 DCT operation. The first 8 DCT coefficients were then appended to the filter coefficients to complete the encode vector. Linear correlation was used as a measure of difference between the resulting feature vectors.

This approach has a significant level of illumination independence. First, normalization of the colour channels, before moving into a chromaticity space, greatly attenuates dependence on both luminance and lighting colour [7]. Moreover, because chromaticity is a ratio of colour bands, moving over to chromaticity space has the effect of removing shading [7], which also contributes to illumination invariance.

### 3.3 Image Encoding

The image was encoded in the following manner:

1. A Gabor wavelet filter grid was used to find discriminating information from multiple locations over the image. The magnitude coefficients, produced by evaluating the Gabor filters at each location, were added to an encode vector.
2. A colour decomposition was performed by creating a binarized 2D stretched chromaticity histogram image for each of 25 image tiles. The histogram was resized and then compressed with a DCT. The top 8 DCT coefficients were added to the encode vector for each image tile.

The data from the frequency analysis and chromaticity histogram was combined to form an encode vector which acted as a signature. The similarity between any two encode vectors was calculated using linear correlation. Linear correlation produces a number in



the range of  $[-1, 1]$ . Zero represents no correlation, negative unity represents a complete negative correlation, and unity represents a complete correlation. The absolute value of the correlation was used to rank the images — this is valid because here it is the distance from zero that represents correlation, as described in Section 2.1.

Because the values in the encoded feature vector are always positive, the results were stored in an unsigned 16-bit integer. The small data type was used to increase the speed of the correlation and to reduce the data that was stored in the database. If the results of the Gabor filter were greater than  $2^{16}$  then the result was truncated and the coefficient was assigned to be  $2^{16}$ . In the dataset used for this experiment this condition never occurred.

The encode vector is made to be efficient for comparison in a number of ways. First, the overall signature size is small (800 bytes), which allows for smaller storage requirements and faster searching. Second, the layout of the signature allows for correlating frequency and colour separately, without fragmentation in the signature, allowing for an efficient implementation of the correlation routine. Finally, by using simple data types like unsigned integers, we can correlate signatures faster than we could using floating point data types. Benchmarks performed on an Intel P4 2.8 GHz machine revealed that the searching could be performed at a rate of 300,000 images per second.

Table 3.1: Gabor Jet Evaluation Positions on Image 833009.jpg

(x,y)	1	2	3	4	5
1	(64,42)	(128,42)	(192,42)	(256,42)	(320,42)
2	(64,85)	(128,85)	(192,85)	(256,85)	(320,85)
3	(64,128)	(128,128)	(192,128)	(256,128)	(320,128)
4	(64,170)	(128,170)	(192,170)	(256,170)	(320,170)
5	(64,213)	(128,213)	(192,213)	(256,213)	(320,213)

## Chapter 4

# Experiments

An image data set was assembled for testing the image recognition algorithm. 2,708 JPEG images of varying content were chosen from a photo library <sup>1</sup> of 41,510 images. The JPEG images were selected by choosing five images, if available, from each directory within the photo library. The purpose of these images was to obtain varying content in which the image recognition capabilities could be tested. In order to test the robustness of various aspects of the algorithm, nine variations of each image were generated programmatically using the Victor Image Processing Library. The format and motivation behind the image variations are listed in Table 4.1. The image variations include resizing, colour changes, distorting the aspect ratio and cropping. It is important to note that for the three image variations involving cropping, the image center is unaffected. This point is emphasized here in order to explain why cropping attacks are a challenging recognition task for this algorithm. When cropping occurs, not only is information removed, but the relative positions of objects within the image change. The effect of this shift on the spectral analysis is that the wavelet filters are evaluated in slightly different positions, thus effecting the magnitudes calculated from the filters. Although Wiskott points out that the magnitudes vary slowly with spatial shifts [32], the results of this study indicate that this effect is still significant in image recognition. All the original images and their nine variations were loaded into a database along with the encode vectors that were produced from those images. With ten variations (original included) of each image, the total number of images in the dataset was 27,080. The dataset was then used to benchmark the algorithm's recognition ability by testing its ability

---

<sup>1</sup>Corel Gallery Photo Library.

to recover the same source image, including variations by cropping and aspect changes, from the database. The benchmark also provided a subjective method to test every iteration of the algorithm and determine the relative recognition contributions of each algorithmic step. The performance was measured by the use of the Cumulative Match Characteristic (CMC). The CMC score is the cumulative count of the correct number of returns. It is shown as a percentage of the total number of correct images expected. A score of 80% achieved by looking at only the first return ( $n = 1$ ) from each search indicates that 80% of the images returned the correct image variation in the top ranked position. The CMC was measured at  $n = 1, 2, 5, 10$  and  $25$ . In addition, this study reports on the receiver operating characteristic (ROC) curves, which are used to evaluate the results of a prediction. A ROC curve is a graphical plot of the number of true positives versus the number of false positives for a binary classifier system as its discrimination threshold is varied. Consider for example, that the algorithm was set up to declare any two images a match if their correlation exceeded 90%. The ROC curve could be inspected to determine the number of true positives versus the number of false positives at 90%. The ROC curve plots this information for all correlation values.

Table 4.1: Format and motivation behind the image copies included in the test dataset.

Format of Image Copy	Purpose
Original JPG image	Original
Greyscale image	Colour invariance
70% resized image	Invariance to small size change
30% resized image	Invariance to large size change
24% height cropped image	Spatial shift invariance
24% width cropped image	Spatial shift invariance
24% height and width cropped image	Spatial shift invariance
20% quality JPG image	Pixelation and colour invariance
30% width stretched image	Aspect change invariance
30% height stretched image	Aspect change invariance

A second dataset was assembled for testing the similar image finding abilities of the

algorithm. This dataset was comprised of all 41,510 images from the same photo library <sup>2</sup>, which is divided into directories based on context and content. Directories that had very similar visual content were selected to measure the similar image finding capabilities. Figure 4.1 shows an example of a selection of thumbnails from one such directory called "Doors of Paris". In this directory, all of the pictures have the same content and layout. They are all pictures of doors that take up a similar portion of the image, have similar shapes and similar backgrounds, but are in fact all pictures of different doors. One measure of the success of this algorithm is to use it to find the remaining pictures of doors based on a single probe image. The purpose of this dataset is to obtain content in which the similar image finding capabilities could be tested as well as getting a sense for the potential object recognition capabilities of the algorithm.

Testing began by isolating each image variation in the dataset, so that each search was restricted to the 2,708 images of the same variation. Using the original image as the probe, each image variation was searched for in turn. This allowed the system's performance to be measured under each of the conditions produced by the image variations. The top 25 appropriate matches from each search were analyzed. The search results were used to calculate the cumulative match characteristic score. The CMC scores for each image variation search are summarized for each iteration of the algorithm.

## 4.1 Evolution of the Spectral Algorithmic Component

During this study various versions and stages of the algorithm were tested. The results of these tests were analyzed and new iterations of the algorithm were designed based on the results. This process, as it was applied to the spectral component, is summarized in this section. The evolution of the colour and chromaticity component is summarized in section 4.2. Spectral content alone is studied in this section. When correlating image signatures during a search, the colour information was ignored and the signatures were based on the original images.

The first version of the algorithm used a 16 x 16 wavelet grid to obtain the spectral components. After analyzing the results of the 16 x 16 wavelet grid, the algorithm progressed to use a 10 x 10 wavelet grid. This study then tested a 8 x 8 wavelet grid, followed by a

---

<sup>2</sup>Corel Gallery Photo Library.



549000.jpg



549001.jpg



549002.jpg



549003.jpg



549004.jpg



549005.jpg



549006.jpg



549007.jpg



549008.jpg



549009.jpg



549010.jpg



549011.jpg

Figure 4.1: Sample Thumbnails from the Corel Photo Library "Doors of Paris" Directory.  
*Images are from the Corel Gallery and are copyright Corel. All rights reserved.*

6 x 6 wavelet grid and then finally a 5 x 5 wavelet grid. The algorithmic progression and associated results are discussed in detail in this section. At each stage, the results of the benchmark are discussed.

#### 4.1.1 16 x 16 Wavelet Grid

Table 4.2 summarizes the CMC results for the 16 x 16 wavelet grid benchmark using the original image as the probe image. The system found the Greyscale, 70% resized, 20% Quality images, 30% X stretch and 30% Y stretch with over 97% accuracy in the top ranked position. The images cropped in one direction, 24% X crop and 24% Y crop, had reasonable rates of returns with 81.28% and 76.33% respectively. The 24% X and Y cropped image had poor returns, achieving only 23.26% returns in the top ranked position. The 30% resized image was essentially never found using this version of the algorithm.

The dataset was then restricted to the 2,708 original images. All of the image variations were then used as probes to search for the original images. The CMC scores generated by using each image variation as the probe image are summarized in Table 4.3. The results indicate similar relative strengths in the system. Using the 30% resized image as the probe gave near zero returns. Cropping in both directions had very poor results as well, with only 23.30% correct returns in the top ranked position. Cropping in a single direction produced better results, with 61.89% and 85.75% in the X and Y direction respectively. Using the aspect altered images as probes provided returns of 99.04% for the X direction and 84.97% for the Y direction.

Two results are apparent in this data set. The first result is that the 30% resized test case shows virtually no recognition capabilities. This is due to a combination of a violation of the Nyquist-Shannon sampling theorem and of an implementation issue. In order to understand the implementation issue, it is first important to understand how the Nyquist-Shannon sampling theorem is violated. With 16 wavelets being evaluated in each direction, each with a 50% overlap, means that each individual wavelet covers 12.5% of the image, or 1/8th. The first four frequencies are analyzed, meaning that for the fourth frequency, a minimum of 16 pixels are required to satisfy the Nyquist-Shannon sampling theorem. When resized to 30%, most images did not satisfy this requirement. The effect of this is that the magnitudes being generated at the higher frequencies are unreliable. The magnitudes generated under this condition are significantly larger. This impacted the implementation of the benchmark program. In order to perform this number of tests against this number of

Table 4.2: Using a 16x16 Gabor Wavelet Grid, every original JPEG image was tested as a probe for finding each image variation in turn. The CMC score was recorded and the percentage of the variations returned, within each ranking category, is listed.

Image Variation	n = 1	n = 2	n = 5	n = 10	n = 25
Original	100	100	100	100	100
Greyscale	99.48	99.48	100	100	100
70% resized	99.04	99.04	99.34	100	100
30% resized	0.04	0.04	0.04	0.04	0.04
24% Y crop	76.33	78.95	81.32	87.22	91.62
24% X crop	81.28	83.75	86.19	90.36	94.28
24% Y & X crop	23.26	25.74	29.36	37.59	47.78
20% quality	99.63	99.82	100	100	100
30% X stretch	99.04	99.04	99.82	100	100
30% Y stretch	97.78	97.97	98.71	99.48	99.67

Table 4.3: Using a 16x16 Gabor Wavelet Grid, every image variation was tested as a probe for finding each original JPEG in turn. The CMC score was recorded and the percentage of the original number of images returned, within each ranking category, is listed.

Image Variation	n = 1	n = 2	n = 5	n = 10	n = 25
Original	100	100	100	100	100
Grayscale	99.48	99.67	100	100	100
70% resized	99.15	99.63	100	100	100
30% resized	0.04	0.04	0.04	0.04	0.04
24% Y crop	85.75	88.29	90.69	93.65	96.42
24% X crop	61.89	64.48	67.61	70.46	73.14
24% Y & X crop	23.30	26.44	30.58	39.48	48.97
20% quality	99.63	99.63	100	100	100
30% X stretch	99.04	99.56	100	100	100
30% Y stretch	84.97	85.64	86.71	87.19	87.81



images, the benchmark application was designed to run all the tests simultaneously so as to avoid performing too many searches. The results for each individual test were determined by inspecting the attributes of the returns. If a particular return was pertinent to a given test then it was used in the calculation of the CMC values, otherwise it was ignored. The total number of returns for each search was set to 2000. No returns were inspected beyond the first 2000 images. The implication of this is that if none (or very few) of the first 2000 returns are applicable to a given test, then the results of that test are incomplete. The number of returns was adequate for all the tests with the exception of the following scenario: with each 30% resized image producing a large spike in the magnitudes at the higher frequencies they all correlated unnaturally low and as a consequence, the 2000 returns inspected was not sufficient to complete this test. The solution to this problem is to run the tests in isolation or by increasing the number of returns inspected. In both cases the benchmark application became unusable as the length of time taken to complete was unmanageable.

The second result we observe is that the cropping attacks produce large reductions in recognition capability, particularly when cropping is in both directions. There are two reasons for this. The first reason is that when images are cropped, information is removed and as a result any algorithm that extracts that information will be affected. The second reason has to do with the placement of the wavelets. Because the placement of the wavelets is a function of the image extents, when the image is cropped the positioning of the wavelets is subject to a spatial shift. The magnitudes generated from the wavelets are sensitive to spatial shifts, especially when the shift is large relative to the wavelet's size.

Figure 4.2 is the top 12 returns at this grid size and it shows a similar result to the CMC results. The system found the Original, Greyscale, 20% Quality and 70% resized in the top four returns. The 30% X stretched image and 30% Y stretched image were next in the number 5 and number 6 positions respectively. The only cropped image to make the top 12 returns was the Y cropped image in position 12. Positions 7 through 11 are all false positives as they were ranked higher than the images generated from the probe.

#### 4.1.2 10 x 10 Wavelet Grid

Table 4.4 summarizes the CMC results for the 10 x 10 wavelet grid benchmark using the original image as the probe image. The system produced similar or better results in all categories. The Greyscale, 70% resized, 20% Quality images produced similar returns of over 99% accuracy in the top ranked position. The aspect altered images, 30% X stretch and



Figure 4.2: False positives and negatives from probe image 549001.jpg with a 16x16 grid size

30% Y stretch, also achieved similar good returns with 98.97% and 98.49% respectively. The cropped images all showed significant increases in recognition rates ranging from a 14.26% increase for the images cropped in both directions, a 9.27% increase for the X cropped images and a 12.33% increase for the images cropped in the Y direction. Although substantially better, the images that were cropped in both directions still had a poor recognition rate of only 37.52% in the top ranked position. The 30% resized image was the most significant improvement, by achieving a 98.56% recognition rate up from 0.04% observed in table 4.2.

The CMC scores generated by using each image variation as the probe image, searching for the original image, are summarized in Table 4.5. The results indicate similar relative strengths in the system as reported in table 4.4. The 30% resized image had the biggest increase in recognition over the previous version of the algorithm, bringing the CMC score up from 0.04% to 98.74% in the top ranked position. The cropped images all showed significant increases in recognition when used as probe images. This increase ranged from 8.27% in the Y cropped images to 18.5% for the images cropped in both directions. The recognition rates of the images cropped in both directions remains relatively poor at 41.80%. Using the X stretched images as probes provided a negligible difference in recognition rates with a difference of less than 0.1%. The Y stretched image however yielded a 13.22% increase in recognition in the top ranked position.

After reviewing the results of this benchmark it is apparent that the 30% resized images can now be used with confidence. The reduction in the number of wavelets from 16 to 10 causes each individual wavelet to cover more of the image. Using 10 wavelets in each direction means that each wavelet covers 20%, or 1/5th of the image. This increase in resolution, allowed the 30% resized images to be accurately transformed as they now had sufficient resolution to satisfy the Nyquist-Shannon sampling theorem. This eliminated the large spikes in the magnitudes that caused the benchmark application to be unreliable for this image category in the previous section. Moreover, this additional resolution reduced the overall sensitivity to spatial shifts, allowing for an increase in the recognition rates involving the cropped images.

Another important effect of reducing the number of wavelets in the wavelet grid is that the size of the image signature is dramatically reduced. Instead of containing 4096 magnitude coefficients,<sup>3</sup> the signature has now been reduced to 1600 magnitude coefficients.

---

<sup>3</sup>(16\*16 wavelets in the grid \* 4 orientations \* 4 frequencies = 4096)

This reduction in data size not only saves storage space but allows for faster searches as less data needs to be correlated.

Taking into account the significant increases in recognition in some categories, with no significant recognition decreases in other categories, and the reduced data overhead of the smaller grid size, a further reduction in the number of wavelets was tested in the next section.

Figure 4.3 is the top 12 returns for this grid size using the probe image, Figure 4.12. Figure 4.3 shows very similar results, not only to the CMC results, but also to Figure 4.2. The system found the Original, Greyscale, 20% Quality and 70% resized in the top four positions as before. The difference comes in with the addition of the 30% resized image in position 6. The 30% X stretched image and 30% Y stretched image came in the number 5 and number 7 position respectively. The Y cropped image took position 11 and position 12 saw the new addition of the X cropped image. There were 3 false positives in this search occupying positions 8 through 10.

### 4.1.3 8 x 8 Wavelet Grid

Table 4.6 summarizes the CMC results for the 8 x 8 wavelet grid benchmark using the original image as the probe image. The system produced similar or better results in all categories compared to the previous iteration. The Greyscale, 70% resized and 20% Quality images produced similar returns of over 99% accuracy in the top ranked position. The 30% resized image also maintained a high rate of return that dropped slightly to 98.38%. The aspect altered images also maintained similar returns, with 98.93% and 98.52% for the X stretched and Y stretched images respectively. The returns of the cropped images continued to improve with increases in recognition rates ranging between a 1.37% increase for the Y cropped images to a 6.61% increase for the images cropped in both directions.

The CMC scores generated by using each image variation as the probe image, searching for the original image, are summarized in Table 4.7. The results indicate similar relative differences as compared to the returns reported in table 4.6. The Greyscale, 70% Resized, 30% Resized and 20% Quality probe images produced virtually identical returns as compared to the previous iteration. The cropped image probes all showed increases in recognition rates ranging from 1.29% in the Y cropped images to 7.2% for the images cropped in both directions. The recognition rates of the images cropped in both directions remains poor at 49.00%. Using the aspect altered images as probes produced small increases in recognition,

Table 4.4: Using a 10x10 Gabor Wavelet Grid, every original JPEG image was tested as a probe for finding each image variation in turn. The CMC score was recorded and the percentage of the variations returned, within each ranking category, is listed.

Image Variation	n = 1	n = 2	n = 5	n = 10	n = 25
Original	100	100	100	100	100
Greyscale	99.48	99.48	100	100	100
70% resized	99.04	99.04	100	100	100
30% resized	98.56	98.63	98.71	99.85	99.89
24% Y crop	88.66	90.25	91.69	94.61	96.49
24% X crop	90.55	92.06	92.80	95.72	97.56
24% Y & X crop	37.52	41.47	45.75	54.91	65.88
20% quality	99.63	99.82	100	100	100
30% X stretch	98.97	98.97	99.52	100	100
30% Y stretch	98.49	98.60	98.74	99.89	99.96

Table 4.5: Using a 10x10 Gabor Wavelet Grid, every image variation was tested as a probe for finding each original JPEG in turn. The CMC score was recorded and the percentage of the original number of images returned, within each ranking category, is listed.

Image Variation	n = 1	n = 2	n = 5	n = 10	n = 25
Original	100	100	100	100	100
Grayscale	99.48	99.67	100	100	100
70% resized	99.19	99.59	100	100	100
30% resized	98.74	99.19	99.71	99.82	99.89
24% Y crop	94.02	95.02	96.60	98.34	98.97
24% X crop	93.09	94.20	95.83	97.49	98.30
24% Y & X crop	41.80	46.79	51.92	60.93	70.75
20% quality	99.63	99.63	100	100	100
30% X stretch	98.97	99.56	100	100	100
30% Y stretch	98.19	98.93	99.34	99.56	99.56



Figure 4.3: False positives and negatives from probe image 549001.jpg with a 10x10 grid size

when compared to the previous iteration, with a difference of 0.07% for the 30% X stretched probe and a 0.52% difference for the 30% Y stretched probe.

After reviewing the results of this benchmark it is apparent that the reduction in the number of wavelets from 100 (10 x 10) to 64 (8 x 8) causes an overall improvement in recognition. Figure 4.7 shows the increase in the average returns for each iteration. This increase in recognition is largely due to increases in the cropped image categories. This signifies that the increase in image area covered by the larger wavelets is still reducing the effect of the spatial shift by significant amounts.

Reducing the number of wavelets in the wavelet grid from 100 (10 x 10) to 64 (8 x 8) continues to reduce the size of the image signature. The signature is reduced from 1600 magnitude coefficients to 1024 magnitude coefficients. The reduction in data and the overall increase in recognition indicate that the algorithm is heading in the right direction. A reduction to a 6 x 6 wavelet grid is tested in the next section.

Figure 4.4 is the top 12 returns for this grid size using the probe image, Figure 4.12. Figure 4.4 has the same relative ordering of the true positive results as Figure 4.3. The only difference is that the Y cropped image and the X cropped image came before the three false positive images which now take up positions 10 through 12.

#### 4.1.4 6 x 6 Wavelet Grid

Table 4.8 summarizes the CMC results for the 6 x 6 wavelet grid benchmark using the original image as the probe image. The system produced similar results in all categories compared to the previous iteration, however, some categories increased slightly while other categories decreased. Greyscale, 70% resized and 20% Quality images produced similar returns of over 99% accuracy in the top ranked position. The 30% resized image also maintained a high rate of return that increased slightly to 98.78%. The aspect altered images maintained similar returns with 98.93% and 98.34% for the X stretched and Y stretched image probes respectively. The returns of the cropped images remained similar with an increase in the Y cropped direction and small decreases in the X and X&Y cropped directions.

The CMC scores generated by using each image variation as the probe image, searching for the original image, are summarized in Table 4.9. The results indicate similar relative differences as compared to the returns reported in table 4.8. The Greyscale, 70% Resized, 30% Resized and 20% Quality probe images produced virtually identical returns as compared

Table 4.6: Using an 8x8 Gabor Wavelet Grid, every original JPEG image was tested as a probe for finding each image variation in turn. The CMC score was recorded and the percentage of the variations returned, within each ranking category, is listed.

Image Variation	n = 1	n = 2	n = 5	n = 10	n = 25
Original	100	100	100	100	100
Grayscale	99.48	99.48	100	100	100
70% resized	99.00	99.00	99.93	99.96	99.96
30% resized	98.38	98.49	98.60	99.67	99.74
24% Y crop	90.03	91.43	93.02	95.98	97.27
24% X crop	92.76	93.65	94.65	97.30	98.60
24% Y & X crop	44.13	48.38	53.40	61.00	70.20
20% quality	99.63	99.82	100	100	100
30% X stretch	98.93	98.93	99.52	100	100
30% Y stretch	98.52	98.71	98.93	99.96	99.96

Table 4.7: Using an 8x8 Gabor Wavelet Grid, every image variation was tested as a probe for finding each original JPEG in turn. The CMC score was recorded and the percentage of the original number of images returned, within each ranking category, is listed.

Image Variation	n = 1	n = 2	n = 5	n = 10	n = 25
Original	100	100	100	100	100
Grayscale	99.48	99.67	100	100	100
70% resized	99.08	99.56	99.96	99.96	99.96
30% resized	98.78	98.86	99.45	99.85	99.85
24% Y crop	95.31	96.09	97.45	98.74	99.34
24% X crop	94.39	95.09	96.31	97.23	97.78
24% Y & X crop	49.00	53.47	59.05	66.91	75.52
20% quality	99.63	99.63	100	100	100
30% X stretch	99.04	99.56	99.96	100	100
30% Y stretch	98.71	99.34	99.59	99.67	99.67





Figure 4.4: False positives and negatives from probe image 549001.jpg with a 8x8 grid size

to the previous iteration, with either no differences or modest increases. The cropped image probes all showed small changes in recognition rates ranging from a 0.77% decrease in the Y cropped images to 0.52% for the Y cropped images. Using the aspect altered images as probes produced a small reduction in recognition for the X stretched probes and for the Y stretched probe images.

After reviewing the results of this benchmark it is apparent that the reduction in the number of wavelets from 8 to 6 does not cause a significant overall reduction in recognition. Figure 4.7 shows the trend in the average returns for each iteration. This lack of change in the recognition rates inspired one more reduction in the number of wavelets.

Reducing the number of wavelets in the wavelet grid from 64 (8 x 8) to 36 (6 x 6) continues to reduce the size of the image signature. The signature is reduced by 43.75%, by reducing the 1024 magnitude coefficients to 576 magnitude coefficients. The significant reduction in data compared to the small overall decrease in recognition suggests that this is an acceptable compromise. A reduction to a 5 x 5 wavelet grid is tested in the next section.

Figure 4.5 is the top 12 returns for this grid size using the probe image, Figure 4.12. There were no changes in the ordering or positioning of the true positives or false positives for this grid size. The system found the Original, Greyscale, 20% Quality and 70% resized in the top four positions. This was followed by the 30% X stretched image, 30% resized image and the 30% Y stretched image in positions 5, 6 and 7. The Y cropped image and the X cropped image are in positions 8 and 9, followed by 3 false positives in positions 8 through 10.

#### 4.1.5 5 x 5 Wavelet Grid

Table 4.10 summarizes the CMC results for the 5 x 5 wavelet grid benchmark using the original image as the probe image. The system produced similar results in all categories compared to the previous iteration, however, there was a slight overall increase in recognition rates. The Greyscale, 70% Resized and 20% Quality images produced similar returns of almost 99% accuracy or higher in the top ranked position. The 30% resized image also maintained a high rate of return that decreased slightly to 98.67%. The aspect altered images maintained similar returns, with 98.71% and 98.45% for the X stretched and Y stretched images respectively. The top ranked return of the 30% X stretched image drop by 1.26% from the previous iteration and the 30% Y stretched returns dropped by 0.19% in the returns for the top ranked position. The returns of the cropped images also remained

Table 4.8: Using a 6x6 Gabor Wavelet Grid, every original JPEG image was tested as a probe for finding each image variation in turn. The CMC score was recorded and the percentage of the variations returned, within each ranking category, is listed.

Image Variation	n = 1	n = 2	n = 5	n = 10	n = 25
Original	100	100	100	100	100
Grayscale	99.63	99.63	100	100	100
70% resized	99.04	99.04	99.96	99.96	99.96
30% resized	98.78	98.78	98.97	99.82	99.82
24% Y crop	90.55	91.69	92.91	95.42	97.08
24% X crop	90.92	92.25	93.35	95.75	97.19
24% Y & X crop	43.87	48.38	52.88	59.01	68.58
20% quality	99.63	99.67	100	100	100
30% X stretch	98.93	98.93	99.45	99.96	99.96
30% Y stretch	98.34	98.45	98.63	99.78	99.89

Table 4.9: Using a 6x6 Gabor Wavelet Grid, every image variation was tested as a probe for finding each original JPEG in turn. The CMC score was recorded and the percentage of the original number of images returned, within each ranking category, is listed.

Image Variation	n = 1	n = 2	n = 5	n = 10	n = 25
Original	100	100	100	100	100
Grayscale	99.63	99.67	100	100	100
70% resized	99.11	99.52	99.96	99.96	99.96
30% resized	98.82	99.37	99.78	99.89	99.96
24% Y crop	94.54	96.01	97.08	98.01	98.71
24% X crop	93.87	95.24	96.23	96.86	97.08
24% Y & X crop	47.08	52.18	57.20	63.59	71.64
20% quality	99.63	99.63	100	100	100
30% X stretch	98.89	99.52	99.96	99.96	99.96
30% Y stretch	98.26	99.00	99.37	99.59	99.59



Figure 4.5: False positives and negatives from probe image 549001.jpg with a 6x6 grid size

similar with an increase in the Y cropped returns of 3.02% and an increase in the X cropped returns of 1.95%. There was a notable decrease in the X&Y cropped image returns of 2.7% for the top ranked position.

The CMC scores generated by using each image variation as the probe image, searching for the original image, are summarized in Table 4.11. The results indicate similar relative differences as compared to the returns reported in table 4.10, however the increase in recognition identified in table 4.10, is in part negated by a drop in recognition reported by table 4.11. All categories produced virtually identical returns as compared to the previous iteration. The cropped probes produced the largest increase of 0.77%, and the aspect altered images remained relatively the same.

After reviewing the results of this benchmark it is apparent that the reduction in the number of wavelets from 6 to 5 does not alter the overall recognition ability of the algorithm significantly. Figure 4.7 shows the slight increase in the average returns for this iteration. As the number of wavelets in each direction is an odd number, there exists a wavelet in the image that is centered around the center of the image. It is hypothesized that, statistically speaking, the differentiating content in an image is generally centered in the image, and thus this center wavelet contributes greatly to the overall recognition abilities. The magnitude coefficient variance is studied in section 4.1.7.

Reducing the number of wavelets in the wavelet grid from 36 (6 x 6) to 25 (5 x 5) continues to reduce the size of the image signature. The signature is reduced by 30.56%, by reducing the 576 magnitude coefficients to 400 magnitude coefficients. The significant reduction in data compared to the small overall increase in recognition reinforces this new wavelet grid size as the best choice.

Figure 4.6 is the top 12 returns for this grid size using the probe image, Figure 4.12. Figure 4.6 shows all 10 image variations being returned in the top 10 positions. The ordering is the same as in the previous iterations of the grid size, with the addition of the X&Y cropped image in position 10. As all ten image variations are now returned in the top ten positions, the images in positions 11 and 12 are no longer considered false positives. The ordering of the images being found are as follows: The Original image, Greyscale image, 20% Quality image, 70% resized image, 30% X stretched image, 30% resized image, 30% Y stretched and finally the X&Y cropped image. This ordering is consistent for most probe images.

Table 4.10: Using a 5x5 Gabor Wavelet Grid, every original JPEG image was tested as a probe for finding each image variation in turn. The CMC score was recorded and the percentage of the variations returned, within each ranking category, is listed.

Image Variation	n = 1	n = 2	n = 5	n = 10	n = 25
Original	100	100	100	100	100
Grayscale	99.59	99.59	100	100	100
70% resized	98.97	98.97	99.96	99.96	99.96
30% resized	98.67	98.71	98.93	99.74	99.78
24% Y crop	93.57	94.68	95.94	98.34	99.08
24% X crop	92.87	94.05	95.31	97.75	98.52
24% Y & X crop	41.17	45.16	50.37	55.95	64.25
20% quality	99.63	99.67	100	100	100
30% X stretch	97.67	97.90	98.52	99.48	99.52
30% Y stretch	98.15	98.34	98.71	99.89	99.93

Table 4.11: Using a 5x5 Gabor Wavelet Grid, every image variation was tested as a probe for finding each original JPEG in turn. The CMC score was recorded and the percentage of the original number of images returned, within each ranking category, is listed.

Image Variation	n = 1	n = 2	n = 5	n = 10	n = 25
Original	100	100	100	100	100
Grayscale	99.59	99.63	100	100	100
70% resized	99.00	99.52	99.96	99.96	99.96
30% resized	98.86	99.00	99.59	99.96	99.96
24% Y crop	90.36	92.87	94.79	96.94	98.08
24% X crop	90.77	93.28	95.16	97.38	98.04
24% Y & X crop	43.54	48.52	54.39	59.71	66.69
20% quality	99.63	99.63	100	100	100
30% X stretch	98.71	99.11	99.70	99.96	99.96
30% Y stretch	98.45	99.08	99.67	99.85	99.89



Figure 4.6: False positives and negatives from probe image 549001.jpg with a 5x5 grid size

### 4.1.6 Receiver Operator Curve

Another important consideration when evaluating a recognition system is the Receiver Operator Curve (ROC). The ROC is a plot of the number of correctly identified images (true positives) versus the number of falsely identified images (false positives). The ROCs comparing each iteration of the spectral analysis algorithm are shown. Figure 4.8 is the ROC showing the difference between a grid size of 16x16 and 10x10, which shows a substantial increase in recognition. The differences in the ROC curves in the subsequent grid size changes are shown in Figure 4.9, Figure 4.10 and Figure 4.11. The ROC curves show a similar result that the CMC results did, in that the recognition ability between a grid size of 16x16 and 10x10 increased significantly, followed by more modest increases in recognition from grid sizes 10x10 to 8x8, then onto 6x6 and finally 5x5.

### 4.1.7 Variance in Magnitude Coefficients Among Same Source Images

An analysis was conducted of the Gabor Wavelet magnitudes in an effort to determine where the recognition capabilities come from. Every signature was exported into csv files for import into Matlab. The Gabor magnitude coefficients were extracted from each signature and images from the same source were grouped together. The within-class scatter was calculated on each of the 2708 groups of 10 image variations. These variations were then averaged for each of the 25 image tiles at each frequency level. Table 4.12 shows the result of this analysis for the first frequency level. Although Table 4.12 does not have a clear pattern, Table 4.13, 4.14, and 4.15 all show smaller variances in the center of the image than at the edges suggesting that the contributions from the center of the image contribute more to the overall recognition of the algorithm. This could be a result of the desire to center the item of interest in the image, thus placing more distinguishing information in the center of the image. Table 4.12 suggests that the low frequency information is present throughout the image and is not concentrated in the centers as the higher frequency information is.

## 4.2 Colour Analysis Evolution

In this section, the use of colour alone is studied, independent of the spatial frequency by correlating only the colour content of a feature vector during a search. The algorithm started by using a single RGB histogram with 8 bins for each colour channel for a total of



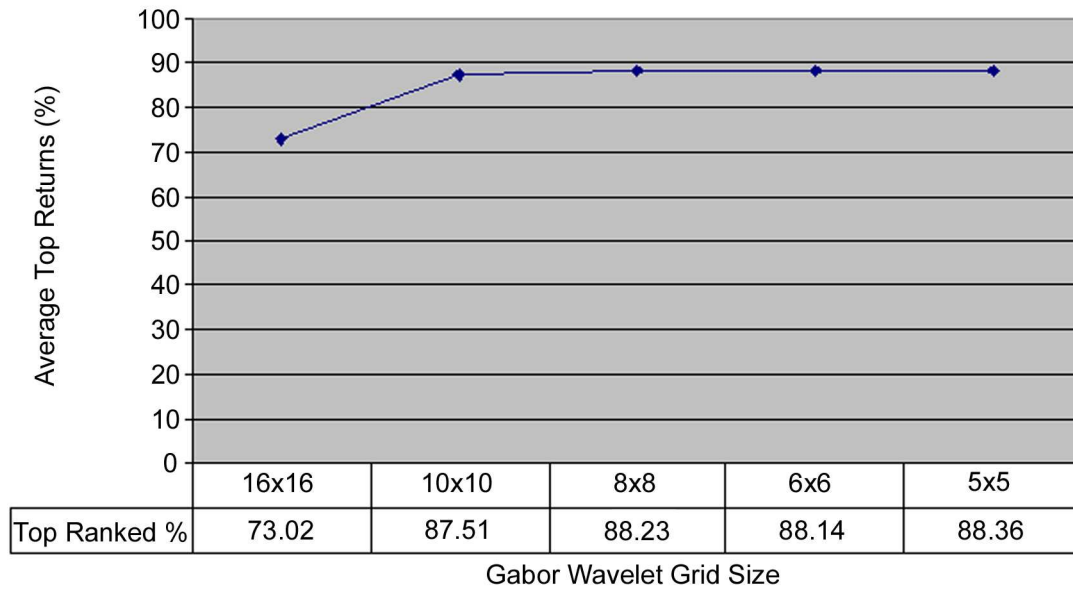


Figure 4.7: Average percentage return of the top ranked position using the original image as the probe.

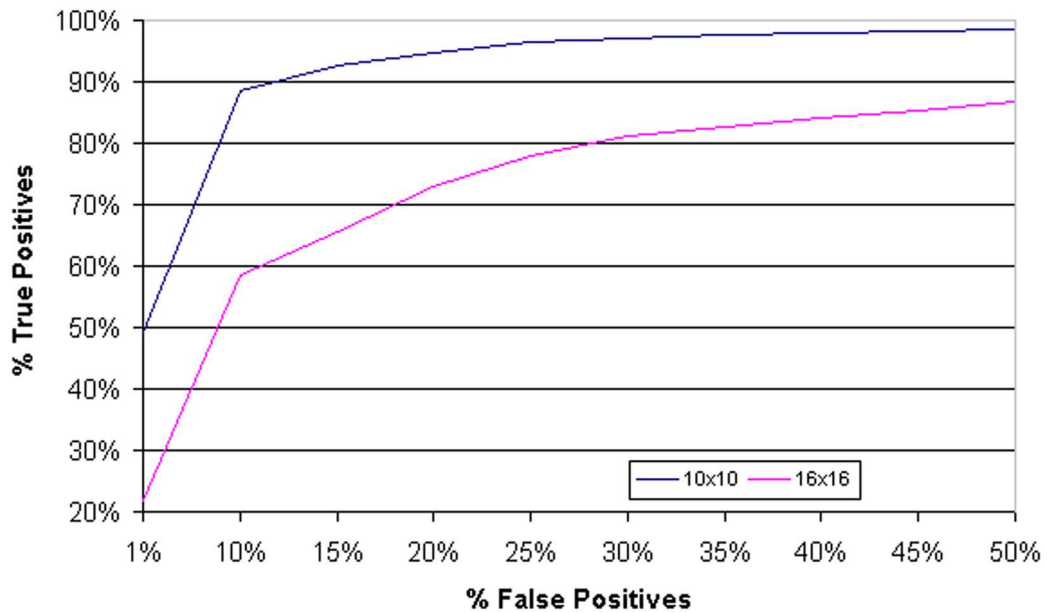


Figure 4.8: Receiver Operator Curve for Grid Sizes 16x16 and 10x10.

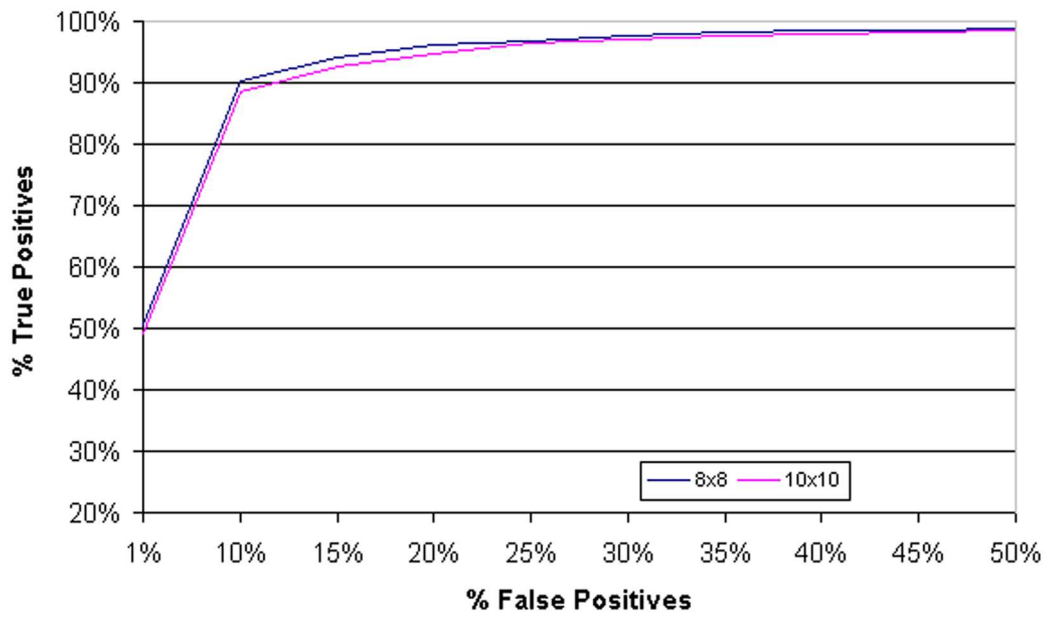


Figure 4.9: Receiver Operator Curve for Grid Sizes 10x10 and 8x8.

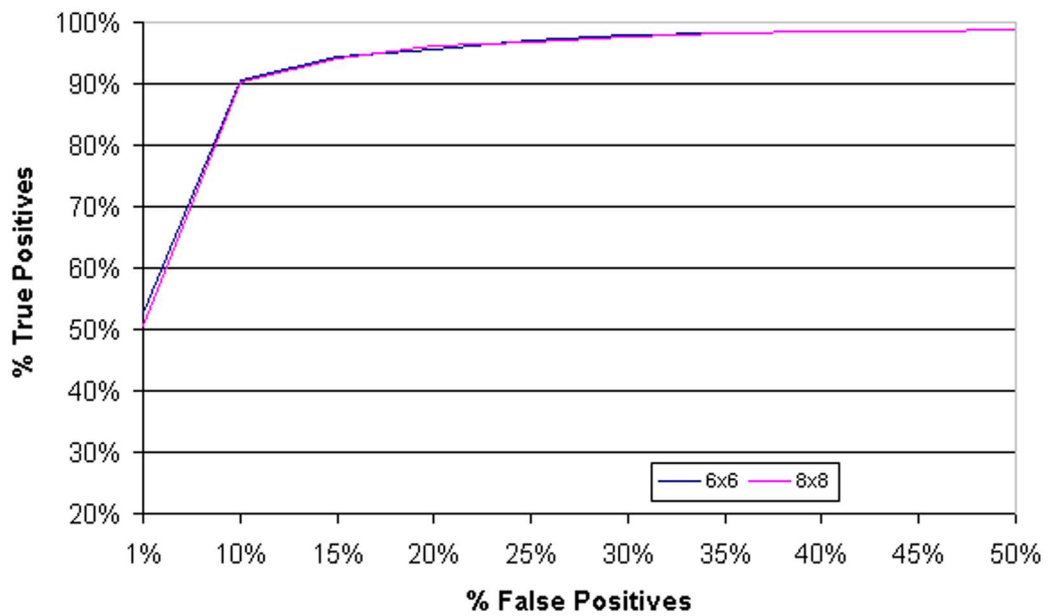


Figure 4.10: Receiver Operator Curve for Grid Sizes 8x8 and 6x6.

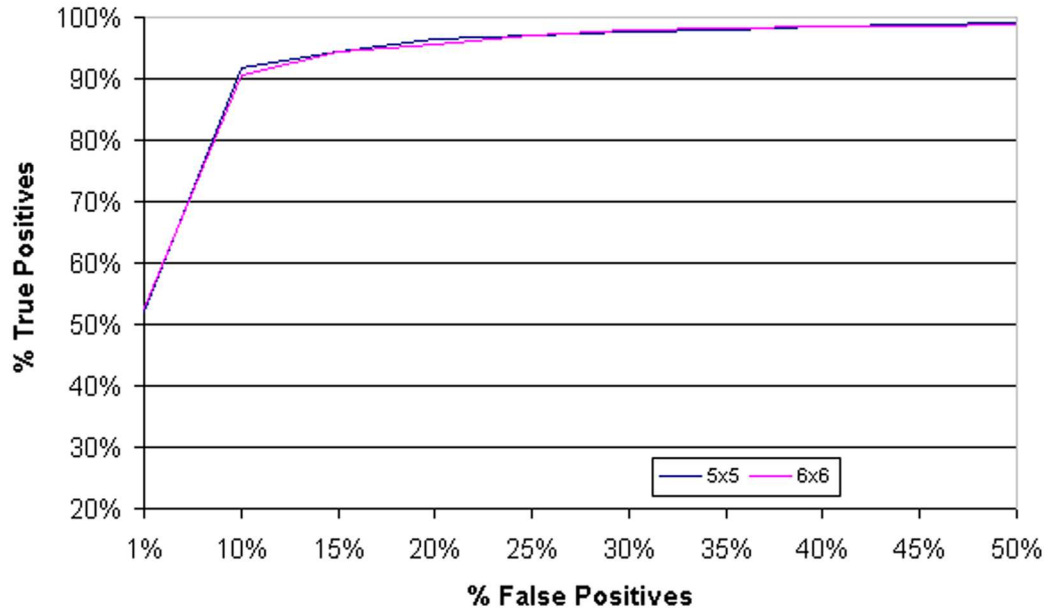


Figure 4.11: Receiver Operator Curve for Grid Sizes 6x6 and 5x5.

Table 4.12: Average Within Class Scatter for Frequency Level 1 Per Image Tile

Image Tile	1	2	3	4	5
1	1.0500	1.0941	1.0602	1.0326	0.9696
2	1.0738	1.0911	1.0527	1.0283	0.9838
3	1.0797	1.0717	1.0413	1.0183	0.9760
4	1.0860	1.0873	1.0566	1.0467	0.9925
5	1.0641	1.0825	1.0630	1.0501	0.9721

Table 4.13: Average Within Class Scatter for Frequency Level 2 Per Image Tile

Image Tile	1	2	3	4	5
1	0.1492	0.1752	0.1836	0.1877	0.1714
2	0.1564	0.1882	0.1984	0.2057	0.1821
3	0.1602	0.1933	0.2048	0.2058	0.1885
4	0.1597	0.1872	0.2005	0.2037	0.1866
5	0.1485	0.1688	0.1798	0.1878	0.1739

Table 4.14: Average Within Class Scatter for Frequency Level 3 Per Image Tile

Image Tile	1	2	3	4	5
1	0.0751	0.0894	0.0960	0.0994	0.0903
2	0.0776	0.0961	0.1011	0.1056	0.0960
3	0.0814	0.1011	0.1052	0.1074	0.0986
4	0.0802	0.0962	0.1040	0.1058	0.0996
5	0.0755	0.0876	0.0979	0.0975	0.0899

Table 4.15: Average Within Class Scatter for Frequency Level 4 Per Image Tile

Image Tile	1	2	3	4	5
1	0.0472	0.0595	0.0651	0.0642	0.0571
2	0.0495	0.0627	0.0685	0.0705	0.0640
3	0.0545	0.0632	0.0723	0.0710	0.0685
4	0.0513	0.0622	0.0681	0.0685	0.0653
5	0.0486	0.0582	0.0612	0.0645	0.0587

512 bins. It was desirable at this point to add a spatial component to the colour analysis. To this end, the algorithm was adapted to create an RGB histogram for each of the non-overlapping 25 image tiles generated from dividing the image into 5 equal sections in both directions. This allowed for the ability to discriminate between different areas of the image when correlating colour components. The algorithm progressed from there to use the same partially overlapping image tiles that were used in the spectral analysis. This allowed for a more efficient algorithm as well as adding more resolution to the RGB histogram. Finally in order to reduce the significant size of the signature the algorithm was redeveloped to use a compressed chromaticity signature. Ratios of normalized colour bands were taken and a binarized 2D chromaticity histogram was created. The two level histogram was compressed as an image resulting in a substantial data compression. At each stage a benchmark was run indicating the relative success of each algorithm step. In each algorithmic step, the performance of retrieving the greyscale images is reported. The retrieval is low, as would be expected as there is no colour information; however, it is not zero as there are a few original greyscale images in the image database.

#### 4.2.1 Single RGB Histogram Intersection

This section examines using only a single RGB histogram over the entire image. There are eight equal sized bins in the histogram for each colour band resulting in a histogram of 512 bins ( $8 * 8 * 8 = 512$ ). The bin counts were calculated as a percentage of the total number of pixels and then normalized into an unsigned 2 byte integer data type.

Table 4.16 summarizes the CMC results for the RGB histogram intersection benchmark using the original image as the probe image. The system found the Original, 70% resized, 20% quality and stretched images with over 98.60% accuracy or higher in the top ranked position. The 30% resized, X and Y cropped images had slightly lower CMC curves starting at 95.24% or lower in the top ranked position. The X&Y cropped image had the poorest CMC results with 79.32% in the top ranked position.

The CMC scores generated by using each image variation as the probe image, searching for the original image, are summarized in Table 4.17. Similar relative differences as compared to the returns reported in Table 4.16 are seen.

It is hypothesized that the poor results from the X&Y cropped images is the result of there being distinguishing colour information at the image extremities that is discarded during the cropping operation, thereby altering the properties of the histogram significantly,

lowering recognition of this image group.

### 4.2.2 Multiple RGB Histogram Intersection

As an alternative to the single RGB histogram, here we study using a separate RGB colour histogram for each image tile in the non-overlapping 5x5 grid. Table 4.18 summarizes the CMC results for the 5 x 5 RGB histogram intersection benchmark using the original image as the probe image. The Original image, 70% resized, 20% quality and stretched images all had moderate gains in recognition starting at 99.15% accuracy or higher in the top ranked position. The 30% resized, X and Y cropped images had significant increases in recognition. The CMC results for these image variations were 98.71% or higher in the top ranked position. The X&Y cropped image had the largest increase. The recognition of the X&Y cropped images jumped 16.21% to 95.53% in the top ranked position as compared with the previous benchmark.

The CMC scores generated by using each image variation as the probe image are summarized in Table 4.19. The X&Y cropped probe images did slightly better relative to the results in table 4.18, but otherwise the benchmark revealed similar relative recognition abilities.

When analyzing the histograms generated in this iteration of the algorithm it was apparent that the histograms were sparsely populated. This is due to the already small images being divided into 25 separate areas. In an attempt to provide more information to the histograms, the same partially overlapping image tiles used in the spectral analysis were also used for the histograms in the next section.

When the algorithm moved to the tiled RGB histograms, they were using a small image area to fill the histogram. As the number of pixels in these areas is quite small, the size of the tiles is increased. The tile size is the same size as the tiles in the spectral analysis. This made for a more efficient implementation and provided additional resolution for the histograms. The overlapping histogram approach is hypothesized to be more robust to the cropping attacks as the histograms should become more stable. Table 4.20 summarizes the CMC results for the 5 x 5 overlapping RGB histogram intersection benchmark using the original image as the probe. All CMC scores remained the same or were slightly higher, except for a small decline for the 30% resized image which fell to 98.6%.

The CMC results generated by using each image variation as the probe image are summarized in Table 4.21. Compared to the results in Table 4.20, the results are almost identical. The 30% resize probe also had a slightly lower CMC curve, but the X&Y cropped image

Table 4.16: Using a RGB Color Histogram Intersection, every original JPEG image was tested as a probe for finding each image variation in turn. The CMC score was recorded and the percentage of the variations returned, within each ranking category, is listed.

Image Variation	n = 1	n = 2	n = 5	n = 10	n = 25
Original	100	100	100	100	100
Grayscale	1.00	1.00	1.18	1.18	1.18
70% resized	98.60	98.60	99.08	99.59	99.74
30% resized	95.24	96.01	96.64	98.23	98.89
24% Y crop	94.87	95.20	95.83	97.82	98.78
24% X crop	95.16	95.61	96.31	97.64	98.63
24% Y & X crop	79.32	81.28	83.64	86.89	90.03
20% quality	99.04	99.15	99.56	99.67	99.78
30% X stretch	98.74	98.74	99.41	99.67	99.78
30% Y stretch	98.67	98.71	99.30	99.67	99.78

Table 4.17: Using a RGB Color Histogram Intersection, every image variation was tested as a probe for finding each original JPEG in turn. The CMC score was recorded and the percentage of the original number of images returned, within each ranking category, is listed.

Image Variation	n = 1	n = 2	n = 5	n = 10	n = 25
Original	100	100	100	100	100
Grayscale	1.00	1.00	1.18	1.18	1.18
70% resized	98.74	98.97	99.34	99.67	99.82
30% resized	95.83	96.75	97.97	98.89	99.37
24% Y crop	95.24	95.90	96.71	97.97	99.04
24% X crop	97.45	97.86	98.23	98.86	99.26
24% Y & X crop	85.49	87.74	89.96	93.17	95.61
20% quality	99.04	99.04	99.45	99.67	99.78
30% X stretch	98.74	98.97	99.48	99.67	99.74
30% Y stretch	98.63	98.97	99.30	99.45	99.56

Table 4.18: Using a 5 x 5 RGB Non-Overlapping Color Histogram Intersection, every original JPEG image was tested as a probe for finding each image variation in turn. The CMC score was recorded and the percentage of the variations returned, within each ranking category, is listed.

Image Variation	n = 1	n = 2	n = 5	n = 10	n = 25
Original	100	100	100	100	100
Grayscale	1.00	1.00	1.18	1.18	1.18
70% resized	99.15	99.15	99.78	99.96	99.96
30% resized	98.71	98.71	98.82	99.78	99.82
24% Y crop	98.86	98.86	98.89	99.71	99.78
24% X crop	98.93	98.93	99.04	99.93	99.93
24% Y & X crop	95.53	95.79	96.16	97.49	98.12
20% quality	99.63	99.82	100	100	100
30% X stretch	99.26	99.26	99.71	100	100
30% Y stretch	99.26	99.26	99.67	100	100

Table 4.19: Using a 5 x 5 RGB Non-Overlapping Color Histogram Intersection, every image variation was tested as a probe for finding each original JPEG in turn. The CMC score was recorded and the percentage of the original number of images returned, within each ranking category, is listed.

Image Variation	n = 1	n = 2	n = 5	n = 10	n = 25
Grayscale	1.00	1.00	1.18	1.18	1.18
Grayscale	99.59	99.63	100	100	100
70% resized	99.26	99.52	99.93	100	100
30% resized	99.15	99.19	99.63	100	100
24% Y crop	98.78	98.86	99.34	99.82	99.82
24% X crop	99.00	99.19	99.63	99.74	99.78
24% Y & X crop	97.97	98.12	98.49	99.11	99.23
20% quality	99.63	99.63	100	100	100
30% X stretch	99.26	99.56	99.93	100	100
30% Y stretch	99.19	99.41	99.78	99.82	99.82



probe increased by 2.07% to 98.56% in the top ranked position.

Although the results from this approach are excellent, the size of the signature for this method limits the number of practical applications. In the next section we try to compress the colour signature significantly in order to make it available for deployment in many different scenarios.

### 4.2.3 Multiple Compressed Chromaticity Coefficients

In order to reduce the size of the image signature, chromaticities were calculated instead of RGB values. The first significant reduction in signature size came because the problem was reduced to 2 dimensions. Second, instead of recording bin counts in the histogram, we use a binary histogram whereby the existence of a particular chromaticity is recorded and not the quantity of that chromaticity. As this severely curtails the amount of information collected the number of bins was increased to partially offset reduced information collected from the image. Finally, the binary histogram is treated like an image and resized to a 16x16 image and then compressed with a DCT. The top 8 coefficients are stored as the signature for that image tile. With 25 tiles, and 8 (2 byte) coefficients for each tile, the colour signature was reduced to 400 bytes. This algorithm is described in more detail in Section 3.2.

Table 4.22 summarizes the CMC results for the 5 x 5 compressed chromaticity benchmark. When using the original image as the probe, the algorithm found the Original, 20% quality, 70% resized, X and Y cropped images and the stretched images with results of 98.49% correct returns or higher in the top ranked position. The 30% resized images and the X&Y cropped image were the two problem areas with 93.10% and 93.5% returns respectively for the top ranked position.

The CMC results generated by using each image variation as the probe image are summarized in Table 4.23. Compared to the results in Table 4.22 the results suggest the same relative strengths and weakness. The Original, 20% quality, 70% resized, X and Y cropped images and the stretched images had results of 98.49% correct returns or higher in the top ranked position. The 30% resized image probe and the X&Y cropped image probe had top ranked returns of 96.34% and 95.31% respectively in the top ranked position.

The benchmark produced results that were slightly worse in all categories when compared to the results in Table 4.21 and in Table 4.20. However, when compared to the original single RGB histogram reported in table 4.16 and table 4.17, the results are higher in the case of the 20% quality, the X and Y stretched and all three cropped image categories. The results

Table 4.20: Using a 5 x 5 RGB Overlapping Color Histogram Intersection, every original JPEG image was tested as a probe for finding each image variation in turn. The CMC score was recorded and the percentage of the variations returned, within each ranking category, is listed.

Image Variation	n = 1	n = 2	n = 5	n = 10	n = 25
Original	100	100	100	100	100
Grayscale	1.00	1.00	1.18	1.18	1.18
70% resized	99.23	99.23	99.78	99.96	99.96
30% resized	98.60	98.63	98.89	99.78	99.78
24% Y crop	99.08	99.08	99.08	99.82	99.82
24% X crop	99.00	99.00	99.10	99.93	99.96
24% Y & X crop	96.49	96.71	96.97	98.23	98.71
20% quality	99.63	99.82	100	100	100
30% X stretch	99.26	99.26	99.71	100	100
30% Y stretch	99.26	99.26	99.67	100	100

Table 4.21: Using a 5 x 5 RGB Overlapping Color Histogram Intersection, every image variation was tested as a probe for finding each original JPEG in turn. The CMC score was recorded and the percentage of the original number of images returned, within each ranking category, is listed.

Image Variation	n = 1	n = 2	n = 5	n = 10	n = 25
Original	100	100	100	100	100
Grayscale	1.00	1.00	1.18	1.18	1.18
70% resized	99.26	99.52	99.93	100	100
30% resized	99.04	99.11	99.63	99.93	100
24% Y crop	99.08	99.15	99.56	99.96	99.96
24% X crop	99.04	99.26	99.71	99.78	99.82
24% Y & X crop	98.56	98.56	98.93	99.37	99.41
20% quality	99.63	99.63	100	100	100
30% X stretch	99.26	99.56	99.93	100	100
30% Y stretch	99.19	99.48	99.78	99.82	99.82

in the cropped image categories are all higher. The Original and 70% resized categories had the same results as in Table 4.16 and the 30% resized image category performed worse than the single RGB histogram approach.

The signature size was greatly reduced with this approach. For the single RGB histogram approach the colour signature was 1KB in size. For the tiled compressed chromaticity approach this was reduced to 400 bytes. Given that the benchmark had the same or better results in all categories except the 30% resize images, the 60% reduction in signature size seems to be an acceptable tradeoff for the lower results in the one category. Moreover, as the 30% resized image category performs well in the spectral analysis the poorer results for the colour analysis will be mitigated when the two techniques are combined into a single algorithm.

### 4.3 Colour and Spectral Combined

The spectral analysis and the colour analysis developed in the previous sections were combined into a single algorithm in this section. The results of the benchmark of this algorithm are discussed in this section as well as the complimentary results of the merged approach.

Table 4.24 summarizes the CMC results of the benchmark for the final algorithm, when using the original image as the probe image. The merged algorithm found the Original image 100% of the time in the top ranked position. The 70% resized and 30% resized had returns of 99.15% and 98.97% respectively. Searching for the greyscale image produced 98.85% correct returns in the top ranked position. The 20% quality image search gave 99.63% correct returns in the top ranked position. The X and Y cropped images had 98.86% and 96.82% returns for the top ranked position and the X&Y cropped image produced a CMC result of 93.94% for the top ranked position.

The CMC results generated by using each image variation as the probe image are summarized in Table 4.25. The results show similar relative strengths and weaknesses when compared to the results in table 4.24. The merged algorithm found the Original image 100% of the time in the top ranked position. Using the 70% resized and 30% resized images as probes found the original images 99.23% and 99.15% respectively in the top ranked position. Using the greyscale image as the probe produced 99.59% correct returns in the top ranked position. The 20% quality image probe gave 99.63% correct returns in the top ranked position. The X and Y cropped image probes had 96.75% and 98.56% returns for the

Table 4.22: Using a 5 x 5 Compressed Chromaticity, every original JPEG image was tested as a probe for finding each image variation in turn. The CMC score was recorded and the percentage of the variations returned, within each ranking category, is listed.

Image Variation	n = 1	n = 2	n = 5	n = 10	n = 25
Original	100	100	100	100	100
Grayscale	1.00	1.00	1.18	1.18	1.18
70% resized	98.60	98.60	99.11	99.45	99.59
30% resized	93.10	93.54	94.20	95.68	96.27
24% Y crop	98.49	98.56	98.60	99.30	99.45
24% X crop	98.63	98.74	98.89	99.67	99.78
24% Y & X crop	93.50	93.83	94.46	95.86	96.90
20% quality	99.34	99.52	99.71	99.74	99.82
30% X stretch	98.93	98.93	99.37	99.74	99.82
30% Y stretch	98.93	98.93	99.30	99.71	99.78

Table 4.23: Using a 5 x 5 Compressed Chromaticity, every image variation was tested as a probe for finding each original JPEG in turn. The CMC score was recorded and the percentage of the original number of images returned, within each ranking category, is listed.

Image Variation	n = 1	n = 2	n = 5	n = 10	n = 25
Original	100	100	100	100	100
Grayscale	1.00	1.00	1.18	1.18	1.18
70% resized	98.78	99.08	99.52	99.63	99.78
30% resized	96.34	96.64	97.71	98.60	98.89
24% Y crop	98.08	98.15	98.56	99.00	99.23
24% X crop	98.49	98.71	99.15	99.34	99.52
24% Y & X crop	95.31	95.83	96.64	97.45	97.97
20% quality	99.34	99.34	99.71	99.74	99.82
30% X stretch	98.93	99.15	99.59	99.74	99.85
30% Y stretch	98.86	99.04	99.45	99.56	99.67

top ranked position and the X&Y cropped image probe produced a CMC result of 95.02% for the top ranked position.

The overall performance increased for the merged algorithm. The colour analysis added additional recognition in the cropped image categories, where the spectral components were suffering from a spatial shift. The spectral analysis added additional recognition in the case of using the greyscale image as the probe or gallery category as there was no colour information available. In addition, the spectral analysis added recognition to the 30% resized image probe and gallery categories where the colour information was severely reduced. Overall the recognition results were excellent. All benchmark categories returned the correct image in the top ranked position with a 93.94% success rate or better.

As the benchmark database was automatically generated by performing various image conversions on an original image, there is one category that was never tested that will be introduced when the algorithm is adapted for object recognition. Colour constancy will be an issue for the algorithm when the same object is captured with different cameras or under different lighting conditions. As the benchmarks in this section show that the spectral analysis seems reasonably impervious to colour changes, it is hypothesized that this will be another strength of the spectral component of the algorithm in the domain of object recognition.

## 4.4 Similar Image Finding

Out of the entire Corel Photo Library of 42,510 images, 5 classifications were chosen and are used for *similar image testing*. The five classifications of visually similar images that were chosen were Parisian Door, Museum Dolls, Cards, Duck Decoys and Easter Eggs. Figures 4.12, 4.13, 4.14, 4.15 and 4.16 are the example images from these categories.

The similar images in each of the classifications were counted. There are 100 Parisian Door pictures, 100 Museum Dolls pictures, 51 single playing Card pictures, 100 Duck Decoy pictures and 100 Easter Egg pictures. Table 4.26 summarizes the results of searching for visual similar images using only one of the images as the probe. The number of similar scenes in each classification is listed in the second column of Table 4.26. The third column is the number of similar images returned in the top 12. The fourth column is the number of similar images that were returned in the top 24. The fifth column is the number of similar images that were returned in the top 48. The sixth column is the percentage of similar

Table 4.24: Using a 50% weighted combination of Colour and Spectral components, every original JPEG image was tested as a probe for finding each image variation in turn. The CMC score was recorded and the percentage of the variations returned, within each ranking category, is listed.

Image Variation	n = 1	n = 2	n = 5	n = 10	n = 25
Original	100	100	100	100	100
Grayscale	99.85	99.85	100	100	100
70% resized	99.15	99.15	99.93	99.93	99.93
30% resized	98.97	98.97	99.08	99.78	99.82
24% Y crop	96.82	96.96	97.34	98.41	98.63
24% X crop	98.86	98.86	98.93	99.93	99.93
24% Y & X crop	93.94	94.57	95.72	97.23	98.30
20% quality	99.63	99.82	100	100	100
30% X stretch	98.38	98.41	99.04	99.48	99.63
30% Y stretch	99.19	99.19	99.48	100	100

Table 4.25: Using a 50% weighted combination of Color and Spectral components, every image variation was tested as a probe for finding each original JPEG in turn. The CMC score was recorded and the percentage of the original number of images returned, within each ranking category, is listed.

Image Variation	n = 1	n = 2	n = 5	n = 10	n = 25
Original	100	100	100	100	100
Grayscale	99.59	99.63	100	100	100
70% resized	99.23	99.59	99.96	99.96	99.96
30% resized	99.15	99.19	99.78	99.96	99.96
24% Y crop	98.56	98.86	99.34	99.85	99.85
24% X crop	96.75	97.27	98.26	98.63	99.04
24% Y & X crop	95.02	96.16	97.34	98.23	98.82
20% quality	99.63	99.63	100	100	100
30% X stretch	99.15	99.52	99.93	100	100
30% Y stretch	97.86	98.30	98.97	99.41	99.63



Figure 4.12: Sample image: Parisian Door  
*Image is from the Corel Gallery and is copyright Corel. All rights reserved.*



Figure 4.13: Sample image: Museum Dolls  
*Image is from the Corel Gallery and is copyright Corel. All rights reserved.*



Figure 4.14: Sample image: Cards  
*Image is from the Corel Gallery and is copyright Corel. All rights reserved.*



Figure 4.15: Sample image: Duck Decoys  
*Image is from the Corel Gallery and is copyright Corel. All rights reserved.*



images that were returned out of the total, up to a maximum of 48. The top 12 returns were completely accurate using this method. The percentage of similar scenes in the top 48 ranged from 72.9% to 100%. These categories work particularly well as the images are of a single object. Images containing multiple objects or highly varied patterned areas of interest were noted to not perform as well. This is not unexpected, as the algorithm does not include any segmentation step. This is a key component to any object recognition system.

Table 4.26: Results from similar image finding tests on five classifications of similar images.

Classification	Number of sim. scenes	Number of top 12 sim.	Number of top 24 sim.	Number of top 48 sim.	% of total up to 48
Parisian Door	100	12	21	35	72.9%
Museum Dolls	100	12	20	37	77.1%
Cards	51	12	24	39	81.3%
Duck Decoys	100	12	24	42	87.5%
Easter Eggs	100	12	24	48	100%

The similar image testing reveals the system’s applicability to object recognition. The similar images tested are essentially all pictures of objects. All of the images had some similar qualities. The Parisian Doors, for example, are all inset in brick buildings and occupied the same proportion of the image. However, there are significant differences in the images as well. Some of the doors were rounded while some were square and some had elaborate detailing while others were quite simple. Still the system was able to pull the images of the doors to the top of the returns. In this respect, there is significant potential for the application of this algorithm in object recognition.

## 4.5 Example Searches

To demonstrate how the system performs, searches were conducted against the entire Corel Photo Library of 42,510 images. Six probe images were chosen and the top 12 returns from searching with those probe images are shown. The search results are ordered in ranking from left to right, top to bottom.

The first probe, shown in Figure 4.17, generated the 12 returns shown in Figure 4.18.

The results have a high degree of visual similarity to the probe image. In particular, return #9 is a close-up picture of the exact same plant that was in the probe image. The second probe image, Figure 4.19, is that of a person performing aerobics. A search with this probe image generated the top 12 returns shown in Figure 4.20. Seven of the top twelve returns are of people doing aerobics. The remaining five images do not seem to be contextually related, however there are similar visual characteristics. They all share the same colour background, and all but one have objects centered in the image.

Figure 4.21 is the same Parisian door seen earlier in this section. It is included again for easy comparison to the search results shown in Figure 4.22. As mentioned earlier in section 4.4, the 12 returns are all doors of slightly varying shape and level of detail. Figure 4.23 is labeled as a wooden duck decoy. When used as a probe image, the top 12 images returned by searching are all wooden duck decoys, shown in Figure 4.24. It should be noted that the dataset contains 100 of each of the Parisian doors and duck decoys. Figure 4.25 is a wedding photograph of which there are 7 similar photos in the image dataset, where the wedding couple in the picture are only photographed from the waist up. When used as a probe image, Figure 4.25 produced four of these photos in the top 12 returns shown in Figure 4.26.

Finally, Figure 4.27 is shown as an example of when the system starts to fail to return the images expected. Figure 4.27 is that of a zebra, and when used as a probe image it returns the 12 images shown in Figure 4.28. There are over 40 pictures that contain a zebra in this dataset and none of them appeared in the top 12 images returned. As can be seen in Figure 4.28, the colour analysis is certainly returning images that contain similar colours. Likewise, the spectral analysis is also returning images that have similar frequency components to those that were measured in the original probe. The problem, however, is that the algorithm only has 4% of the colour analysis and spectral analysis covering the zebra. The algorithm, in a sense, makes no attempt to single out the distinguishing information of the zebra, but instead attempts to maximize the entire correlation. The returns are the images that shared the most in common (colour and frequency) on average. A human observer would certainly expect to see some of the over 40 pictures of zebras in the search returns.



Figure 4.16: Sample image: Easter Egg  
*Image is from the Corel Gallery and is copyright Corel. All rights reserved.*



Figure 4.17: Sample probe image: 423000.jpg  
*Image is from the Corel Gallery and is copyright Corel. All rights reserved.*

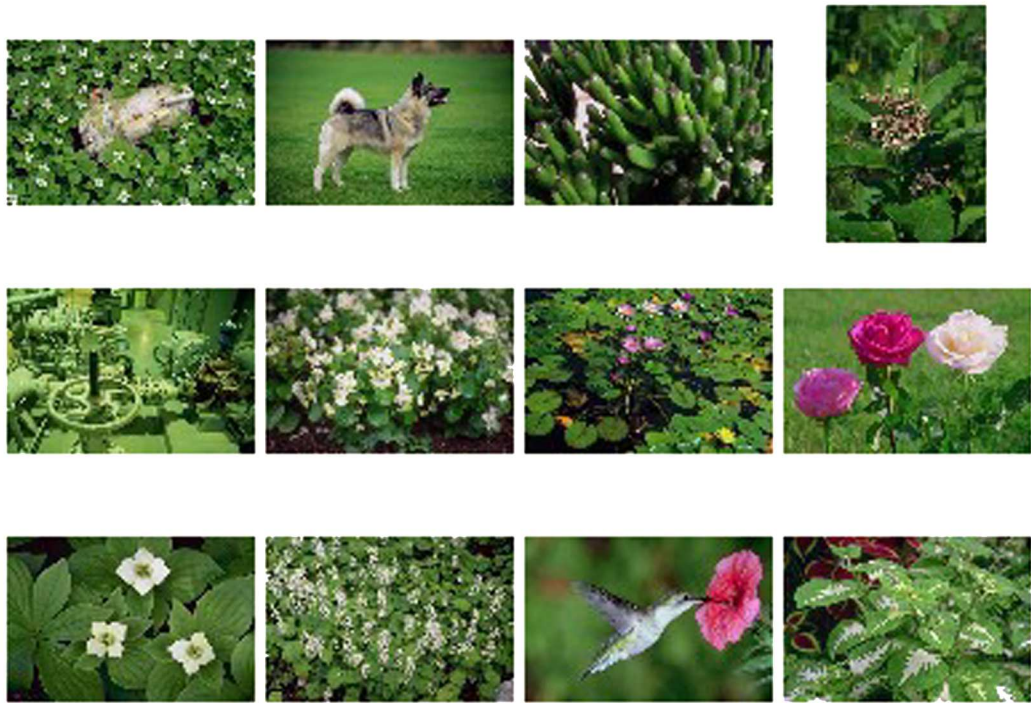


Figure 4.18: Search Results from probe image: 423000.jpg  
*Images are from the Corel Gallery and are copyright Corel. All rights reserved.*



Figure 4.19: Sample probe image: 282041.jpg  
*Image is from the Corel Gallery and is copyright Corel. All rights reserved.*

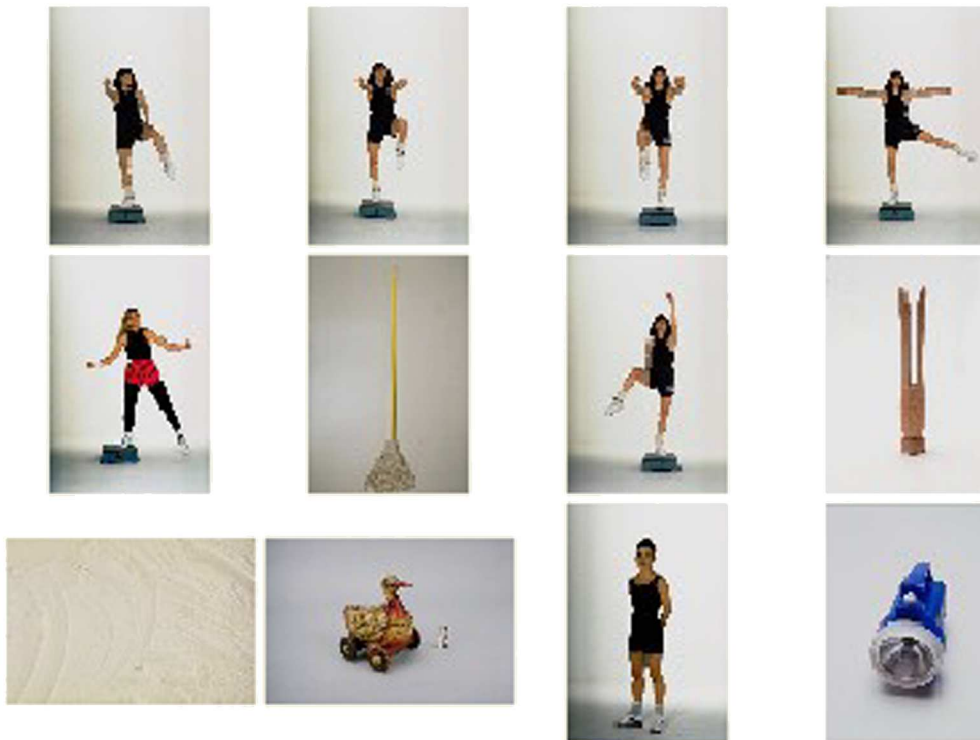


Figure 4.20: Search Results from probe image: 282041.jpg  
*Images are from the Corel Gallery and are copyright Corel. All rights reserved.*



Figure 4.21: Sample probe image: 549001.jpg  
*Image is from the Corel Gallery and is copyright Corel. All rights reserved.*



Figure 4.22: Search Results from probe image: 549001.jpg  
*Images are from the Corel Gallery and are copyright Corel. All rights reserved.*



Figure 4.23: Sample probe image: 655017.jpg  
*Image is from the Corel Gallery and is copyright Corel. All rights reserved.*



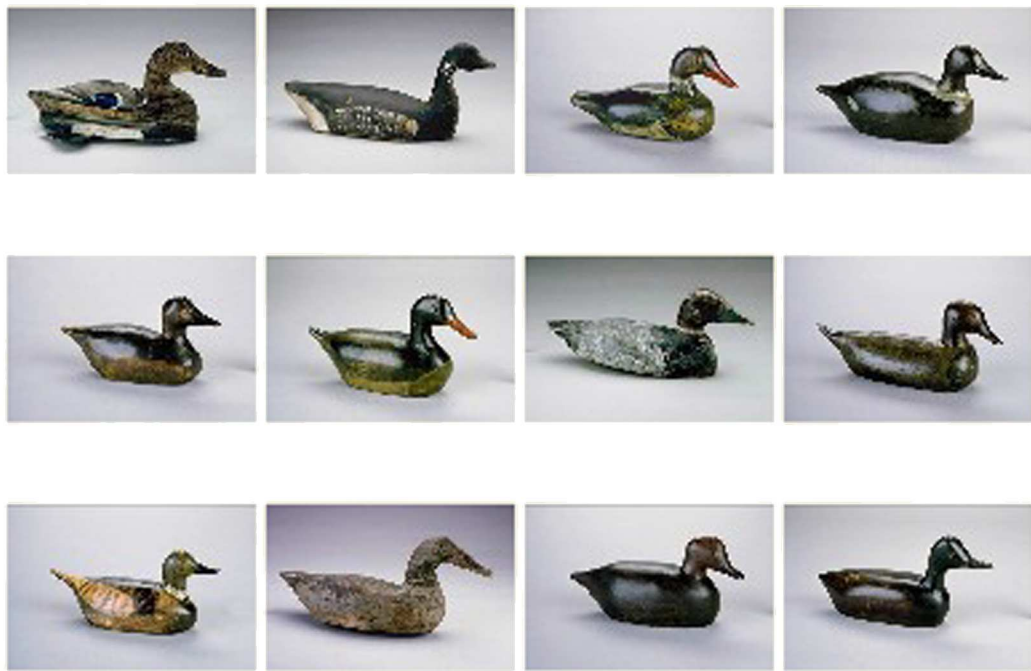


Figure 4.24: Search Results from probe image: 655017.jpg  
*Images are from the Corel Gallery and are copyright Corel. All rights reserved.*



Figure 4.25: Sample probe image: 702083.jpg  
*Image is from the Corel Gallery and is copyright Corel. All rights reserved.*

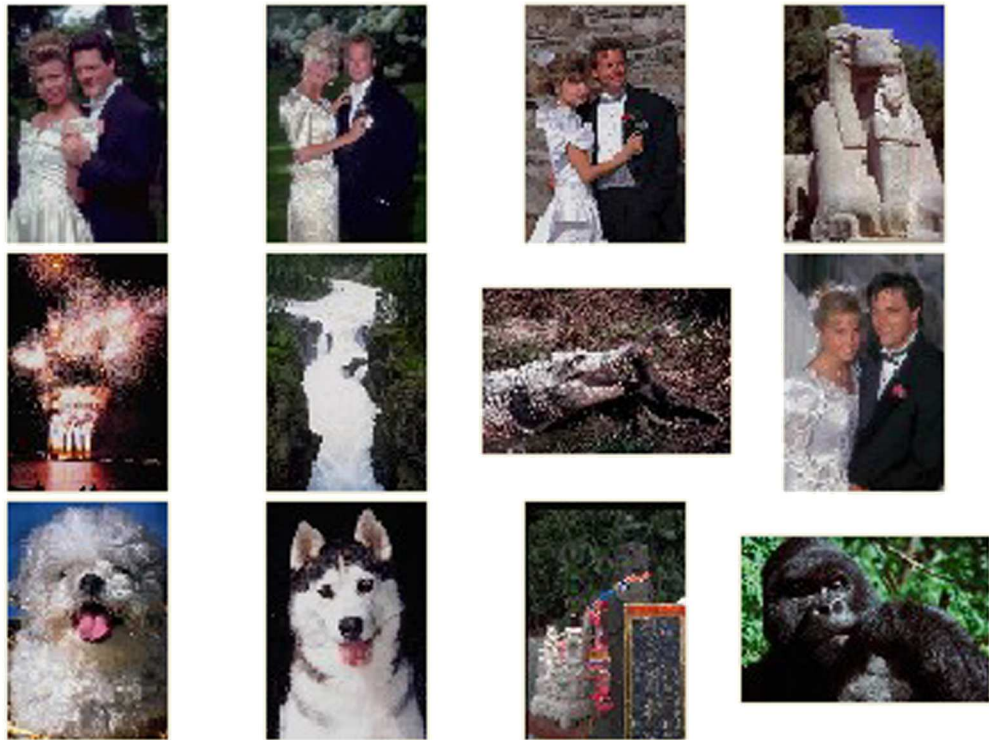


Figure 4.26: Search Results from probe image: 702083.jpg  
*Images are from the Corel Gallery and are copyright Corel. All rights reserved.*



Figure 4.27: Sample probe image: 408019.jpg  
*Image is from the Corel Gallery and is copyright Corel. All rights reserved.*

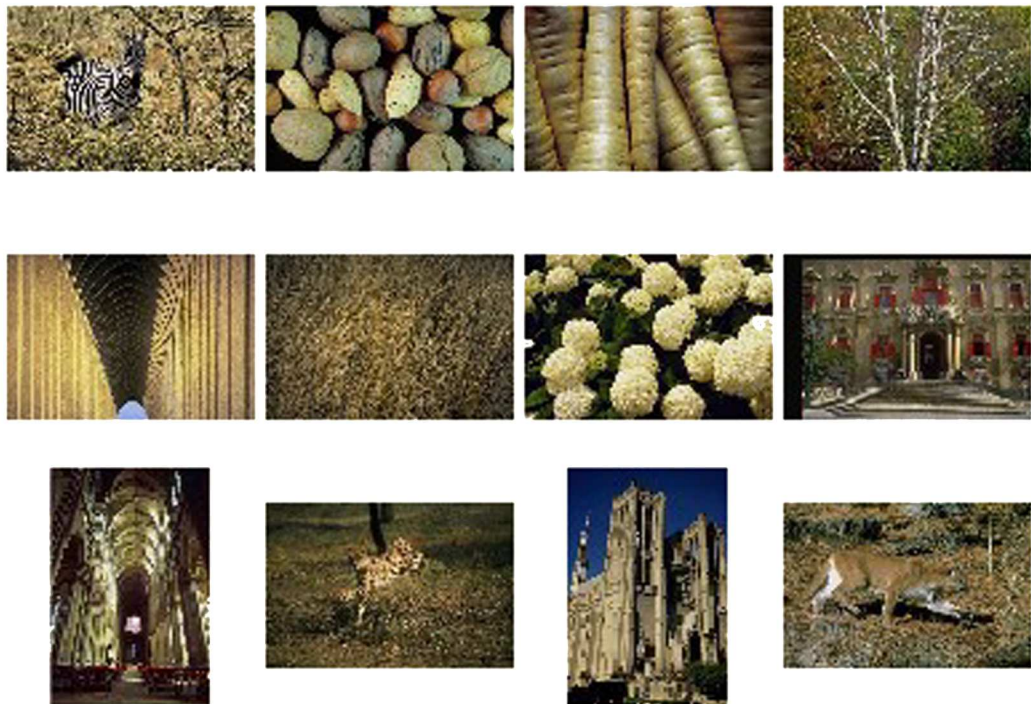


Figure 4.28: Search Results from probe image: 408019.jpg  
*Images are from the Corel Gallery and are copyright Corel. All rights reserved.*

## Chapter 5

# Conclusions and Future Work

### 5.1 Conclusions

This project was conducted as a proof of concept for a hybridized approach to object recognition. Based on the results of this project, there is a significant potential for this algorithm. Recognition rates for same-image recognition were near perfect, 93.94% or higher, for all image variations. Similar-image finding also produced encouraging results, with similar scenes being returned with between 72.9% and 100% success. The algorithm presented here has several attractive qualities including simplicity, compactness and speed.

The frequency analysis presented here has one step. This offers a significant simplification over the other approaches discussed in section 2.4. The approaches presented in [18, 3, 35, 30, 2, 25, 27] all have two or more steps in which they, at a basic level, try to determine the salient features or areas first. This is accomplished by some number of steps involving statistical, image processing or computer vision techniques. Following this step, the salient features are then summarized or otherwise represented by a small number of values, which then comprises the feature vector. The approach presented here removes the step of determining where the salient features exist and instead supposes that if we sample the entire image we will then sample the salient features as well. Although the approach is significantly simpler than other approaches, this study demonstrates that high levels of distinguishing information can still be extracted.

The colour analysis also offers a level of simplicity as it too involves a single step, without the prior steps of determining the salient areas of the image to sample. Moreover, this study presents a colour analysis in which the signature is highly compressed. The compression

happens at multiple stages. First, the use of chromaticities is a form of compression that discards information about illumination, which not only compresses the colour information, but also contributes to a more illumination invariant approach. Second, the binarized histogram is a more compact method of recording which chromaticities exist over the area being analyzed. Finally, compressing the binarized chromaticity by the use of a DCT offers a final reduction in data size to eight coefficients (two bytes each).

The image signature generated by this approach is very compact. For each area sampled the colour analysis produced 16 bytes of data (eight coefficients of two bytes each). In addition to the colour analysis, the frequency analysis also required 16 bytes of storage space (16 coefficients of one byte each generated from four orientations with four frequencies each). With 25 areas being sampled for both the colour analysis and the frequency analysis, the total signature is 800 bytes in length. This signature size has two main benefits. First, the small signature requires less storage space and second, a smaller signature can be correlated more quickly than a larger one. These results make the approach more adaptable for scaling out the algorithm for use with larger image repositories.

The combination of Gabor Wavelets and compressed chromaticities outperformed either method used in isolation as the strengths of one approach complimented the other approach's weaknesses. The effect of combining a colour analysis with a frequency analysis was an overall improvement in recognition. The frequency analysis is particularly sensitive to the spatial shifts generated by the cropping attacks. The colour analysis performed here proved to be much less sensitive to spatial shifts and therefore aided the performance of the overall algorithm with respect to the cropping attacks. Conversely, the colour analysis was rendered useless for the greyscale images. In this instance, the recognition was achieved by the frequency analysis.

The algorithm presented here has limitations that were revealed during testing. The algorithm performed poorly on images with a high number of objects or images that were otherwise very busy in appearance. Moreover, the algorithm is sensitive to spatial shifts, which has the effect of reduced recognition capabilities on images where the content has shifted. This includes pictures taken from different angles and cropping attacks where the image center is altered.

## 5.2 Future Work

There are several changes that would have to be made in order to apply this technique more successfully at the object level. First, the objects would have to be segmented from the imagery. Second, because objects can have irregular shaped edges, a different technique for distributing the wavelets would have to be implemented. One such approach could be implemented as follows. Instead of wavelets being placed in a grid pattern, they could be placed in a pattern that is a function of the object's edges and the object's center. First the object's center would be determined: this could be achieved by any method, so long as it was repeatable even under rotation of the object. A weighted center, or a center computed from the minimum bounding box could be used. Next, imaginary lines are drawn from the center of the objects to the edges in steps of some angle,  $\theta$ . Each line is divided into  $x$  pieces of equal length. The wavelets are evaluated at each intersection, and stored in the signature in a clockwise fashion starting with the magnitudes that were closest to the center, then second closest to the center and working out to the magnitudes closest to the object's edge. There are two problems with this approach that need to be discussed, along with tentative solutions. First, the object's shape may be such that a wavelet placement is outside the object. If this case occurs, it should be detected and its magnitude set to 0, or some other value reserved for this scenario. Second, a direct comparison of magnitudes will only work if the object has not been rotated. As this assumption is not practical, it is speculated that if the magnitudes were transformed using a Fourier Transform, that the coefficients from that operation could be correlated. In effect, it would be the pattern or shape of the magnitudes that would provide the recognition, instead of the magnitudes themselves. As long as a repeatable coordinate system could be applied to each object, this algorithm is postulated to be rotationally invariant. A coordinate system could be placed using a variety of techniques including an axis of major inertia.

## Appendix A

# Generating Chromaticity Histograms in Matlab

### A.1 List of Programs



```
%read input image and convert to double floating point data type
img = double(imread('833009.jpg','jpg'));
[width,height,channels] = size(img);

%Calculate the denominator: R+G+B. Avoid divide by zero.
%Note: If the denominator is zero then so is the numerator.
%      Thus, making this modification is reasonable.
denom = img(:,:,1) + img(:,:,2) + img(:,:,3);
back = denom==0; denom(back) = 1.0;
img(:,:,1) = img(:,:,1) ./ denom;
img(:,:,2) = img(:,:,2) ./ denom;

%bring the chromaticity values into the 1-256 range
img(:,:,1) = (img(:,:,1) .* 255) + 1;
img(:,:,2) = (img(:,:,2) .* 255) + 1;
img(:,:,2) = 257 - img(:,:,2); %flip the g value over the y axis
img = uint16(img);

histogram = zeros(256,256);
for i=1:width
    for j=1:height
        x = img(i,j,1);
        y = img(i,j,2);
        histogram(y,x) = 1; % binarized
    end
end
imwrite(histogram,'linear_hist_plot.tif','tif');
```

Program A.1: Matlab Script to Generate Linear Chromaticity Histogram

```
%read input image and convert to double floating point datatype
img = double(imread('833009.jpg','jpg'));
[width,height,channels] = size(img);

%Calculate the denominator: sqrt(R*R + G*G + B*B). Avoid divide by zero.
%Note: If the denominator is zero then so is the numerator.
%      Thus, making this modification is reasonable.
denom = (img(:,:,1) .* img(:,:,1));
denom = denom + (img(:,:,2) .* img(:,:,2));
denom = denom + (img(:,:,3) .* img(:,:,3));
denom = sqrt(denom);
back = denom==0; denom(back) = 1.0;
img(:,:,1) = img(:,:,1) ./ denom;
img(:,:,2) = img(:,:,2) ./ denom;

%bring the chromaticity values into the 1-256 range
img(:,:,1) = (img(:,:,1) .* 255) + 1;
img(:,:,2) = (img(:,:,2) .* 255) + 1;
img(:,:,2) = 257 - img(:,:,2); %flip the g value over the y axis
img = uint16(img);

histogram = zeros(256,256);
for i=1:width
    for j=1:height
        x = img(i,j,1);
        y = img(i,j,2);
        histogram(y,x) = 1;
    end
end
imwrite(histogram,'spherical_hist_plot.tif','tif');
```

Program A.2: Matlab Script to Generate Spherical Chromaticity Histogram

```

%read input image and convert to double floating point datatype
img = double(imread('833009.jpg','jpg'));
[width,height,channels] = size(img);

%Calculate the denominator: R+G+B. Avoid divide by zero.
%Note: If the denominator is zero then so is the numerator.
%      Thus, making this modification is reasonable.
denom = img(:,:,1) + img(:,:,2) + img(:,:,3);
back = denom==0; denom(back) = 1.0;
img(:,:,1) = img(:,:,1) ./ denom;
img(:,:,2) = img(:,:,2) ./ denom;

%alter the chromaticity values to achieve the stretched chromaticity space
for i=1:width
    for j=1:height
        if img(i,j,1) >= img(i,j,2)
            img(i,j,1) = img(i,j,1) + img(i,j,2);
            img(i,j,2) = img(i,j,2) * 2;
        else
            img(i,j,2) = img(i,j,1) + img(i,j,2);
            img(i,j,1) = img(i,j,1) * 2;
        end
    end
end

%bring the chromaticity values into the 1-256 range
img(:,:,1) = (img(:,:,1) .* 255) + 1;
img(:,:,2) = (img(:,:,2) .* 255) + 1;
img(:,:,2) = 257 - img(:,:,2); %flip the g value over the y axis
img = uint16(img);

histogram = zeros(256,256);
for i=1:width
    for j=1:height
        x = img(i,j,1);
        y = img(i,j,2);
        histogram(y,x) = 1;
    end
end
imwrite(histogram,'stretched_hist_plot.tif','tif');

```

Program A.3: Matlab Script to Generate Stretched Chromaticity Histogram

# Bibliography

- [1] D. Anderson and M.S. Drew. Image object search combining shape and color descriptors using gabor wavelets and illumination invariance. In *Storage and Retrieval for Media Databases 2003*, pages 44–53, January 2003.
- [2] J.R. Bach, C. Fuller, and A. Gupta. The virage search engine: An open framework for image management. *SPIE Storage and Retrieval for Image and Video Databases, IV*, Feb. 1996.
- [3] S. Belongie, C. Carson, H. Greenspan, and J. Malik. Color- and texture-based image segmentation using EM and its application to content-based image retrieval. *International Conference of Computer Vision*, 1998.
- [4] M.S. Drew C. Lu and J. Au. Classification of summarized video by using hidden markov models on compressed chromaticity signatures. In *ANNIE'01: Artificial Neural Networks In Engineering*, pages 645–650, November 2001.
- [5] J. G. Daugman. Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression. *IEEE Trans. on Acoustics Speech and Signal Processing*, 36(7):1169–1179, 1988.
- [6] J.G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by twodimensional visual cortical filters. *Journal of the Optical Society of America*, 2(A):1160–1169, 1985.
- [7] M.S. Drew and J. Au. Video keyframe production by efficient clustering of compressed chromaticity signatures. In *ACM Multimedia '00*, pages 365–368, 2000. <http://www.cs.sfu.ca/~mark/ftp/AcmMM00>.
- [8] M.S. Drew and J. Au. Clustering of compressed illumination-invariant chromaticity signatures for efficient video summarization. *Image and Vision Computing Journal*, 21(8):705–716, 2003.
- [9] M.S. Drew, J. Wei, and Z.N. Li. Illumination-invariant color object recognition via compressed chromaticity histograms of color-channel-normalized images. In *ICCV98*, pages 533–540. IEEE, 1998.

- [10] G.D. Finlayson and S.D. Hordley. Color constancy at a pixel. *J. Opt. Soc. Am. A*, 18(2):253–264, Feb. 2001. Also, UK Patent application no. 0000682.5. Under review, British Patent Office.
- [11] G.D. Finlayson, S.D. Hordley, and M.S. Drew. Removing shadows from images. In *ECCV 2002: European Conference on Computer Vision*, pages 4:823–836, 2002. Lecture Notes in Computer Science Vol. 2353.
- [12] B.V. Funt and G. Finlayson. Color constant color indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):522–529, May 1995.
- [13] S. Chatterjee G. Finlayson and B. Funt. Colour angular indexing. In *ECCV96 Fourth European Conference on Computer Vision*, volume II, pages 16–27, April 1996.
- [14] M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Wrtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42(3):300–311, March 1993.
- [15] S. Liapis and G. Tziritas. Colour and texture image retrieval using chromaticity histograms and wavelet frames. To appear, *IEEE Transactions on Multimedia*, 2004.
- [16] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expressions with gabor wavelets. In *Proceedings, Third IEEE International Conference on Automatic Face and Gesture Recognition*, pages 200–205, April 1998.
- [17] M. Lyons, J. Budynek, A. Plante, and S. Akamatsu. Classifying facial attributes using a 2-d gabor wavelet representation and discriminant analysis. In *Proceedings of the 4th International Conference on Automatic Face and Gesture Recognition*, pages 202–207, March 2000.
- [18] H. Sawhney M. Flickner and W. Niblack. Query by image and video content: The QBIC system. *IEEE Comput.*, pages 23–32, Sept. 1995.
- [19] R. Mehrotra, K.R. Namuduri, and N. Ranganathatn. Gabor filter based edge detection. *Pattern Recognition*, 25(12):1479–1494, 1992.
- [20] A. Monadjemi, B. T. Thomas, and M. Mirmehdi. Experiments on high resolution images towards outdoor scene classification. Technical report, University of Bristol, Department of Computer Science, February 2002.
- [21] J. Wei M.S. Drew and Z. Li. On illumination invariance in color object recognition. *Pattern Recognition*, 31(8):1077–1087, 1998.
- [22] J. Wei M.S. Drew and Z. Li. Illumination-invariant image retrieval and video segmentation. *Pattern Recognition*, 32:1369–1388, 1999.

- [23] O. Nestares, R. Navarro, J. Portilla, and A. Taberner. Efficient spatial-domain implementation of a multiscale image representation based on gabor functions. *Journal of Electronic Imaging*, 7(1):166–173, 1998.
- [24] D. Paulus, K. Horecki, and K. Wojciechowski. Localization of colored objects. In *International Conference on Image Processing: ICIP '00*, pages Vol III 492–495, 2000.
- [25] A. Pentland, R.W. Picard, and S. Sclaroff. Photobook: Tools for content-based manipulation of image databases. In *SPIE Storage and Retrieval for Image and Video Databases II*, volume 2, 185, pages 34–47, San Jose, CA, 1994.
- [26] W.H. Press, B.P. Flannery, S.A Teukolsky, and W.T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, New York, New York, USA, 1988.
- [27] K. Hirata S. Mukherjea and Y. Hara. Toward a multimedia world wide web information retrieval engine. In *Sixth International World Wide Web Conference*, Santa Clara, CA, April 1997.
- [28] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing Analysis and Machine Vision*. International Thomson Computer Press, London, 1996.
- [29] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–13, 1991.
- [30] Q. Tian, N. Sebe, M.S. Lew, E. Loupias, and T.S. Huang. Content-based image retrieval using wavelet-based salient points. In *Storage and Retrieval for Media Databases 2001*, volume 4315, pages 425–436, January 2001.
- [31] L. Wiskott. *Labeled Graphs and Dynamic Link Matching for Face Recognition and Scene Analysis*, volume 53 of *Reihe Physik*. Verlag Harri Deutsch, Thun, Frankfurt am Main, Germany, 1995. (PhD thesis).
- [32] L. Wiskott. Phantom faces for face analysis. *Pattern Recognition*, 30(6):837–846, 1997.
- [33] L. Wiskott. The role of topographical constraints in face recognition. *Pattern Recognition Letters*, 20(1):89–96, 1999.
- [34] L. Wiskott. Segmentation from motion: Combining Gabor- and Mallat-wavelets to overcome the aperture and correspondence problems. *Pattern Recognition*, 32(10):1751–1766, 1999.
- [35] O.R. Zaïane Z.N. Li and Z. Tauber. Illumination invariance and object model in content-based image and video retrieval. *Journal of Visual Communication and Image Representation*, 10:219–244, 1999.