On Extending the Capability of the Image Ray Transform

by

Ah Reum Oh

Thesis for the degree of Doctor of Philosophy

June 2016
A conventional approach to image analysis is to separately perform feature
extraction at a low level (such as edge detection) and follow this with high level
feature extraction to determine structure (e.g. by collecting edge points) using
the Hough transform. The original Image Ray Transform (IRT) demonstrated
capability to emphasise structures at a low level. Here we extend the IRT to add
shape specificity that makes it select specific shapes rather than just edges; the
new capability is achieved by addition of a single parameter that controls which
shape is selected by the extended IRT. The extended approach can then perform
low-and high-level feature extraction simultaneously. We show how the IRT
process can be extended to focus on chosen shapes such as lines and circles.
We also suggest an additional extension of IRT to detect shapes of chosen
colours. The new approach uses the CIEL*a*b* colour model within the IRT’s light
ray analogy. The capability of the extended IRT using colour information is
evaluated for correct shape location by conventional methods such as the Hough
Transform. We analyse performance with images from the Caltech-256 dataset
and show that the new approach can indeed select chosen shapes and colours.
We also show how the new approach has the capability to detect objects with
specific shape and colour. Further research will aim to capitalise on the new
extraction ability to extend descriptive capability.
# Table of contents

ABSTRACT........................................................................................................................................ i

Table of contents ......................................................................................................................... i

List of tables ................................................................................................................................... v

List of figures .................................................................................................................................... vii

DECLARATION OF AUTHORSHIP................................................................................................. xi

Acknowledgements....................................................................................................................... xiii

Definitions and abbreviations ......................................................................................................... xv

Chapter 1:  Introduction ................................................................................................................ 1

1.1 Context ................................................................................................................................. 1

1.2 Contributions ....................................................................................................................... 1

1.3 Publications ......................................................................................................................... 2

Chapter 2:  Background .............................................................................................................. 5

2.1 Analogies ............................................................................................................................ 5

2.1.1 Water flow ..................................................................................................................... 5

2.1.2 Heat ............................................................................................................................ 8

2.1.3 Force .......................................................................................................................... 12

2.1.4 Pressure ..................................................................................................................... 16

2.1.5 Other approaches ........................................................................................................ 19

2.2 Basic methods for feature extraction.................................................................................... 20

2.2.1 Sobel operator .................................................................................................................. 20

2.2.2 Hough Transform ......................................................................................................... 21

2.2.3 Colour models ............................................................................................................... 22

2.2.4 SIFT, SURF and CSIFT ............................................................................................... 24

2.2.5 Earth Mover’s Distance and other distance metrics ...................................................... 26

2.3 Conclusions ........................................................................................................................ 28

Chapter 3:  Image Ray Transform (IRT) ..................................................................................... 29

3.1 Overview ............................................................................................................................. 29
3.2 The Image ray transform .......................................................... 29
3.3 Image ray transform analysis ..................................................... 32
3.4 On extending the image ray transform ....................................... 34

Chapter 4: Adding shape selectivity to the IRT ............................... 35
4.1 Overview ................................................................................. 35
4.2 Extending the IRT .................................................................... 36
4.3 Variation of the shape factor ....................................................... 37
4.4 Analysing images by the extended IRT ....................................... 47
4.5 Applying the shape factor in real images ................................. 52
4.6 Conclusions ............................................................................ 56

Chapter 5: Detecting objects using colour information .................... 59
5.1 Adding colour within the IRT .................................................... 59
5.2 Comparing different colour spaces .......................................... 60
5.3 Detecting coloured objects ....................................................... 62
5.4 Adjustment .............................................................................. 66
5.5 Conclusions ............................................................................ 71

Chapter 6: Combining shape and colour sensitivity in the IRT .... 73
6.1 Overview ................................................................................. 73
6.2 Operation ................................................................................ 73
6.3 Contrast .................................................................................. 88
6.4 Conclusions ............................................................................ 94

Chapter 7: Conclusions and future work ...................................... 95
7.1 Conclusions ............................................................................ 95
7.2 Future work ........................................................................... 96

Appendix A ................................................................................... 99
A.1 Real images in database Caltech-256 Object Category Database .................................................................................. 99
A.2 Real images ............................................................................ 100

Appendix B .................................................................................. 101
B.1 The result of synthesised images: Rectangle .......................... 101
B.2 The result of synthesised images: Curve ............................... 102
B.3 The result of synthesised images: Mixed a circle, a rectangle and a curve .......................................................... 103
List of tables

Table 1 Comparison results of approaches using the Hough Transform for line ................................................................. 57
Table 2 Comparison results of approaches using the Hough Transform for circle ............................................................... 57
Table 3 The performance value $P$ of different colour models .................. 62
Table 4 The error values of the extended IRT ......................................................... 84
Table 5 The matching rate of SURF and coloured SURF .............................. 91
Table 6 Comparison of the error value by extended IRT and Sobel operator 93
Table 7 EMD results of correlation of Figure 7.1 ................................. 97
## List of figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>An MRI example of three-dimensional segmentation using water flow</td>
<td>8</td>
</tr>
<tr>
<td>2.2</td>
<td>Extraction of moving edge map</td>
<td>9</td>
</tr>
<tr>
<td>2.3</td>
<td>Examples of segmentation of human lung image</td>
<td>12</td>
</tr>
<tr>
<td>2.4</td>
<td>Force field calculation example at each pixel in an image</td>
<td>13</td>
</tr>
<tr>
<td>2.5</td>
<td>Potential energy surfaces of a human ear image</td>
<td>14</td>
</tr>
<tr>
<td>2.6</td>
<td>The force field transform results</td>
<td>14</td>
</tr>
<tr>
<td>2.7</td>
<td>The Divergence of a vector field</td>
<td>15</td>
</tr>
<tr>
<td>2.8</td>
<td>Convergence field</td>
<td>16</td>
</tr>
<tr>
<td>2.9</td>
<td>Examples of two synthetic disks</td>
<td>18</td>
</tr>
<tr>
<td>2.10</td>
<td>Rendering of the segmented image</td>
<td>19</td>
</tr>
<tr>
<td>2.11</td>
<td>Templates for the Sobel operator</td>
<td>20</td>
</tr>
<tr>
<td>2.12</td>
<td>The polar consideration of a line</td>
<td>21</td>
</tr>
<tr>
<td>2.13</td>
<td>Constructing the accumulator space for the HT for circles</td>
<td>22</td>
</tr>
<tr>
<td>2.14</td>
<td>Examples of colour models [58]</td>
<td>23</td>
</tr>
<tr>
<td>2.15</td>
<td>Example step for scale space [45]</td>
<td>25</td>
</tr>
<tr>
<td>2.16</td>
<td>A keypoint descriptor [45]</td>
<td>25</td>
</tr>
<tr>
<td>2.17</td>
<td>The simple examples using EMD</td>
<td>27</td>
</tr>
<tr>
<td>3.1</td>
<td>Ray Propagation</td>
<td>29</td>
</tr>
<tr>
<td>3.2</td>
<td>The result of changing the number of iterations $M$</td>
<td>33</td>
</tr>
<tr>
<td>3.3</td>
<td>Examples of detecting mixed shapes in images</td>
<td>33</td>
</tr>
<tr>
<td>4.1</td>
<td>Illustrating results of the original Image Ray Transform (IRT) and the extended IRT, with the Hough transform results superimposed showing that the circle and the line have been emphasised by different selection of the new shape factor.</td>
<td>35</td>
</tr>
<tr>
<td>4.2</td>
<td>The Ray Propagation in the Standard and Extended IRT</td>
<td>36</td>
</tr>
<tr>
<td>4.3</td>
<td>An Example of the error boxplot</td>
<td>38</td>
</tr>
<tr>
<td>4.4</td>
<td>The result according to varying the shape factor in Figure 3.3 (a)</td>
<td>40</td>
</tr>
<tr>
<td>4.5</td>
<td>The box plot of a synthesised image (Figure 4.4)</td>
<td>41</td>
</tr>
<tr>
<td>4.6</td>
<td>The result according to varying the shape factor for Figure 3.3 (c)</td>
<td>41</td>
</tr>
<tr>
<td>4.7</td>
<td>The box plot of a synthesised image (Figure 4.6)</td>
<td>42</td>
</tr>
<tr>
<td>4.8</td>
<td>Results of extended IRT by $F_S$ in a synthesised image</td>
<td>43</td>
</tr>
<tr>
<td>4.9</td>
<td>Analysing the outputs for circle detection</td>
<td>44</td>
</tr>
</tbody>
</table>
Figure 4.10 Analysing the outputs ($\rho$) for line detection .................................. 45
Figure 4.11 Analysing the outputs ($\phi$) for line detection ................................. 46
Figure 4.12 Basic Histogram Structure ................................................................. 48
Figure 4.13 The histograms of objects in various images ................................. 48
Figure 4.14 Histogram patterns by shape and iteration number when $F_s = 1.0$ .................................................................................................................. 49
Figure 4.15 Histogram patterns by shape factor variation when $M = 10000$ .... 50
Figure 4.16 Histogram patterns by object position when $F_s = 1.0$ .............. 51
Figure 4.17 Histogram patterns by object size when $F_s = 1.0$ ...................... 51
Figure 4.18 Results of extended IRT by $F_s$ in a real image ............................. 53
Figure 4.19 Analysing the IRT outputs for a real image ..................................... 54
Figure 4.20 The results with the error value of an example image. ............... 55
Figure 4.21 Examples of detection on real images ............................................ 56
Figure 5.1 Comparison of colour models to detect each coloured shape ...... 61
Figure 5.2 Results of extended IRT by $\lambda c$ in synthesised image ............... 63
Figure 5.3 Analysing the outputs by $\lambda c$ variation in synthesised image ....... 64
Figure 5.4 Results of extended IRT by $\lambda c$ in a real image ............................ 65
Figure 5.5 Thresholding results of extended IRT by $\lambda c$ in real image ......... 65
Figure 5.6 Detecting a red object with different colour spaces ..................... 67
Figure 5.7 The example outputs by the colour factor ..................................... 67
Figure 5.8 Analysing the outputs by the colour factor variation ..................... 68
Figure 5.9 Examples of detection by the colour factor on real images .......... 69
Figure 5.10 Analysing the outputs by $\lambda c$ variation in real image ............... 70
Figure 5.11 Results of extended IRT by $\lambda c$ in a real image .......................... 71
Figure 6.1 Results of extended IRT by $F_s$ and $\lambda c$ in a real image ............. 74
Figure 6.2 Analysing the outputs when $\lambda c$ is $-0.7$ ........................................ 76
Figure 6.3 Analysing the outputs when $\lambda c$ is $-0.7$ ........................................ 77
Figure 6.4 Analysing the outputs when $\lambda c$ is $-0.7$ ........................................ 78
Figure 6.5 Analysing the outputs when $\lambda c$ is $-0.7$ ........................................ 79
Figure 6.6 Analysing the outputs without the colour factor ......................... 80
Figure 6.7 Analysing the outputs without the colour factor ......................... 81
Figure 6.8 Analysing the outputs without the colour factor ......................... 82
Figure 6.9 Analysing the outputs without the colour factor ......................... 83
Figure 6.10 The results with the shape factor and the colour factor ............ 85
Figure 6.11 The result of finding red objects ....................................................... 86
Figure 6.12 The result of finding green objects .................................................... 87
Figure 6.13 Matching rate of original SURF examples ................................................. 89
Figure 6.14 Matching rate of colour SURF examples .................................................. 90
Figure 6.15 Example test images and samples ............................................................... 91
Figure 6.16 The result of test samples ............................................................................ 92
Figure 6.17 The result example of Sobel operator ......................................................... 93
Figure 7.1 Shape extraction and description in real images ......................................... 97
DECLARATION OF AUTHORSHIP

I, Ah Reum Oh, declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

On Extending the Capability of the Image Ray Transform

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given.
   With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

Signed:...................................................................................................................

Date:........................................................................................................................
Acknowledgements

Firstly, I would like to thank my supervisor Professor Mark S. Nixon for his help throughout my whole PhD course. All of his advice and guidance in my course and writing were always inestimable even the time when I strayed with disappointing since the outcome was seriously far from the certain. I could not imagine a better advisor for my PhD study than him. I am extremely grateful to him for encouraging me always.

Also, thanks to all colleagues from VLC and ECS for their assistance and friendship during my study.

Finally, I would like to thank my family, especially my father, for every their support. I really appreciate all of them.
# Definitions and abbreviations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta_i )</td>
<td>The angle of incidence</td>
</tr>
<tr>
<td>( \theta_c )</td>
<td>The critical angle</td>
</tr>
<tr>
<td>( n_t )</td>
<td>Refractive index of a ray at step ( t ) (time)</td>
</tr>
<tr>
<td>( n )</td>
<td>Refractive index of a medium</td>
</tr>
<tr>
<td>( \theta_l )</td>
<td>The angle of reflection</td>
</tr>
<tr>
<td>( \theta_r )</td>
<td>The angle of refraction</td>
</tr>
<tr>
<td>( n_{max} )</td>
<td>The maximum refractive index</td>
</tr>
<tr>
<td>( n_i )</td>
<td>Refractive index at intensity ( i )</td>
</tr>
<tr>
<td>( N )</td>
<td>Normals</td>
</tr>
<tr>
<td>( E_{dir} )</td>
<td>The direction of the Sobel operator</td>
</tr>
<tr>
<td>( \psi )</td>
<td>Normal direction</td>
</tr>
<tr>
<td>( x )</td>
<td>The position at a point in horizontal axis</td>
</tr>
<tr>
<td>( y )</td>
<td>The position at a point in vertical axis</td>
</tr>
<tr>
<td>( \varphi_0 )</td>
<td>The initial direction of the ray</td>
</tr>
<tr>
<td>( p_0 )</td>
<td>The initial position vector of the ray</td>
</tr>
<tr>
<td>( w )</td>
<td>The width of an image</td>
</tr>
</tbody>
</table>
\( h \)  The height of an image

\( p_t \)  The position vector of the ray at step \( t \)

\( \varphi_t \)  The direction of the ray at step \( t \)

\( V_t \)  The direction vector of the ray at step \( t \)

\( t \)  A step of the ray in progress

\( R_t \)  The reflection vector of the reflected ray

\( R_r \)  The refraction vector of the refracted ray

\( M \)  The iteration number

\( \text{maxPathIndex} \)  The max number of iterations

\( F_s \)  The shape factor

\( \lambda_c \)  The colour factor

\( d \)  Orientation (degree)

\( N_d \)  The normalised values at each orientation in a moving average window

\( T \)  The number of samples in a moving average window

\( C_t \)  The summation of angles between two points of a ray

\( \text{HT} \)  The Hough Transform

\( \text{IRT} \)  The Image Ray Transform
**RGB**  Red, green and blue

**HSV**  Hue, saturation and value

**HSI**  Hue, saturation and intensity

**CIE L’ a' b’**  A colour model using $L^*, a^*$ and $b^*$ by the international commission on illumination (Commission international de l’éclairage)

$L'$  The lightness

$a^*$  The digitised chrominance of red/green

$b^*$  The digitised chrominance of yellow/blue

$x_n, y_n, z_n$  Tristimulus values of the reference in the $xyz$ model.

$s$  The junction point for colour transform

$n_{ij}$  The value in refractive materials’ matrix

$\Gamma$  The detected object

$A$  A measure of area

$P(\Gamma)$  The performance of

$E_\rho$  The error value of $\rho$ for a line

$E_\phi$  The error value of $\phi$ for a line

$E_c$  The error value of a circle

$\rho$  The distance between the origin and the line by the Hough Transform
The orthogonal vector angle from the origin by the Hough Transform

The radius of a circle

The value from the Hough Transform

The original value
Chapter 1: Introduction

1.1 Context

As computer vision continues to proliferate in industrial and domestic life, effective and generic methods are needed for image processing systems. Feature extraction is one of the principal stages in computer vision, so many approaches have been developed for low- and high-level processing. Using physical analogies as a basis for algorithms has been exemplified in low-level feature extraction. For instance, edge detection [1] was refined by anisotropic diffusion [2] to find moving-edges using the analogy of heat flow in the temporal dimension. Force [3] and magnetic field analysis [4] have also been applied for image segmentation. Apart from these methods, there is a variety of feature extraction techniques using physical properties, such as water flow [5] and time [6]. These operators can be deployed for low-level image segmentation prior to higher level shape analysis and classification.

The Image Ray Transform (IRT) is a new strategy to extract structures in 2-D data by mimicking the transmission of light and has been used in the pre-processing stages of image based systems within ear biometrics and retinal analysis [7,8]. The IRT is focused on the analogy to refraction and reflection of a light ray with refractive materials’ indices derived from image information, calculated from grey-level pixels. The IRT is a powerful technique for low-level feature extraction and is suited to emphasising curved objects, as shown in detecting ears for biometric purposes.

1.2 Contributions

We describe two novel extensions to the IRT. The first adds a new parameter: a shape factor, which allows edge detection and object selection at the same time. This appears to be the first approach to combine low-level analysis with high-level structure. We then extend the approach to derive descriptive capability by using colour information to allow edge detection and object selection at the same time while other extended approaches [9] employed grey level images. This gives low-level sensitivity to shape: the IRT can be arranged to find boundaries from different shapes. In this way, the operation of the transform
Chapter 1

characterises different shapes. Methods have been developed to detect objects with various approaches, such as: fuzzy set theory [10]; morphological gradient [11]; monogenic signal [12]; colour distributions [13] and more recently the Hilbert transform [14]. Our new approach differs from traditional edge based systems that incorporate colour, since the new approach focusses on detecting the boundaries of smoothly curving shapes [7,8]. Further, the IRT can be extended to detect chosen shapes [9]. Here, a parameter, called the colour factor, is added to the original IRT to detect shapes of a specific colour. The colour factor can be used to construct the refractive materials’ matrix for the IRT, and the CIE *L’ a*b* colour model appears most appropriate for this task. The Hough Transform [15] can be used to corroborate the quality of the new feature extraction process. Accordingly, adding the shape parameter is a new way to extract features of objects and extends the performance of the IRT. Application of the Hough Transform shows that the extended IRT can lead to more general capability, with superior performance, to that of the original IRT.

In this thesis, chapter 2 describes the principles of some physical analogies in computer vision and the fundamental techniques used in this thesis. This is followed by an introduction of the original IRT at first, and the new extended IRT with the shape factor is described and compared on synthesised and real images with confirming analysis by the Hough Transform together in chapter 4. Then, the analysis and the comparison of colour models for the IRT are demonstrated in chapter 5. Chapter 6 leads to the combination of shape and colour analysis for shape detection and extraction, and the abilities of the two operators are compared with confirming analysis together, before concluding in chapter 7.

1.3 Publications

The publications arising from this work concern the contributions described in Chapters 3, 4, 5 and 6. The extension of the IRT, analysis of the shape factor and histogram patterns for the shape detection and extraction, and the extended IRT using colour information for detecting coloured objects have been published or submitted as:


Chapter 2: Background

2.1 Analogies

Various approaches using basic analogies have been developed to extract low and high-level features in image processing. Template matching is a basic approach to extract features whilst mathematical morphology is also a fundamental technique to analyse and modify images for feature extraction in computer vision. While traditional methods have already proved their effectiveness, approaches from natural phenomena have been investigated to develop a variety of systems. For example, the systems in biometrics and the medicine might require high-specification cameras and more sophisticated and stable techniques for feature extraction.

Using physical analogies to natural phenomena is one of the latest technologies to express image content in a new way. By their physical basis, these approaches are more easily understood by researchers, and more clear to parameterise since their processes are already well understood in the natural world. The approaches using natural physical analogies [16] can have a basis in diverse visual systems and provide alternative methods to conventional techniques. These started with basic natural analogies, such as water flow, heat and force. The newer approaches, pressure and light, are less common and have the potential to improve feature extraction in image processing in new ways. The continuing development of these approaches can support the sustained advances in computer vision.

2.1.1 Water flow

Conventional approaches to use water for feature extraction arose from region growing methods which find similar pre-defined properties in neighbouring regions. However, region growing is sensitive to similarity criteria, such as irregular boundaries, holes and corruption by noise. Another developed approach is the morphological watershed-based region growing technique [5,17]. This technique uses the intensity or gradient of image pixels and focuses on the analogy to geographical features in images. It is also based on region growing methods, therefore, has similar problems to those with previous approaches.
Chapter 2

The latest approach based on water flow did not use complex finite element models since these do not offer the possibility of finding shapes with low computational complexity. Water has fluidity and surface tension which can lead to topological adaptability and geometrical flexibility like contour smoothness [18]. Water flow is comprised of features that depend on pressure, surface tension and adhesion/capillarity [18,19]. The final shape of a water flow boundary should be influenced by the contour of the related object. The velocity and direction of flow is determined by flow rate, resistance and pressure. The flow is calculated according to a framework which includes region growing and active contour techniques. Equation 2.1, which is a basis of this approach, shows the relationship between three elements: the flow rate $f_r$, the flow resistance $R$ and the pressure differences between the pressure inflow $P_i$ and the pressure outflow $P_o$.

$$f_r = \frac{P_i - P_o}{R}$$ \hspace{1cm} 2.1

The pressure difference controls the flow and the relation between the cross-sectional area $A$ of the flowing water and the effective flow velocity $V_{effective}$ is given by

$$f_r = A \cdot V_{effective}$$ \hspace{1cm} 2.2

These two equations provide the basis to simulate water flow in images. The water element from an initial point can flow in any direction when the computed velocity is positive and sufficiently large in that direction. Dynamic formulae are applied to combine the flow velocity and forces which leads to an equation describing the conservation of energy. When an element with positive velocity $v$ on a particular position is affected by a force $F$, it can arrive at the direction-related position, then the energies are related according to Equation 2.3

$$\frac{mv_f^2}{2} = F S + \frac{mv^2}{2}$$ \hspace{1cm} 2.3

where $v_f$ is the final scalar velocity with fixed displacement $S$ and assumed mass $m$. The force $F$ has a positive scalar magnitude when the force is consistent with velocity $v$, and is negative otherwise. The movement can be determined by the right hand side summation, for example, if $F$ is negative, the summation can be negative and the equality in Equation 2.3 cannot be satisfied.
By disparity of Equation 2.3 that it is possible for flow to move outwards. By Equation 2.2 the flow velocity $V$ can be calculated when other factors of water, such as turbulence, are ignored. If the velocity points move to the exterior of the water, the element is presumed to leave the initial point. Then, an image force $F$ is merged at this step and movement of flow can be determined from the result. A more detailed equation is shown in Equation 2.4 whilst $m$ and $S$ are constants in Equation 2.3.

$$J = \lambda \left( \frac{F_D + F_A}{R(x,y)} \right)^2 + F$$

where $F_D$ is driving force modified by surface tension, $F_A$ is adhesion and $R$ is the edge response at point $(x,y)$. $\lambda$ is a regularization parameter which controls the trade-off between the two energy terms, and it can be chosen by the user. In addition, it can be determined by combining mass $m$, displacement $S$ and area $A$ and it affects smoothness of image noise: if there is much noise contamination, the value of $\lambda$ should be larger. A positive direction is defined from the origin by the scalar components on the movement of $F_A$ and $F_D$. The velocity information inside the brackets can influence the movement decision and $J$ is the ultimate kinetic energy. For example, if the velocity component and $J$ are both positive, water will flow in the direction of the target point.

This water-flow method can be extended to a three-dimensional model for volume segmentation and application to various medical images containing atypical and complex structures is shown in Figure 2.1. The example size is a 181 X 217 X 181 MR image volume of a human brain. The method segments most parts of the lateral ventricles in Figure 2.1 (a) and lower and upper cross-sectional slices of the selected part are shown in Figure 2.1 (b) (as highlighted in white).
Figure 2.1 An MRI example of three-dimensional segmentation using water flow

2.1.2 Heat

Another approach using natural analogies is heat which has a functional similarity to water flow. Both methods aim to grow regions starting from a seed point and merge neighbouring points. Anisotropic diffusion for image enhancement was originally developed using the heat analogy in image processing. An anisotropic diffusion pyramid [20] and an anti-geometric heat-flow model [21] were suggested to extract smooth shapes for several region-based segmentation.

Recently Direkoglu and Nixon introduced an expanded technique combining anisotropic and isotropic heat flow to determine binary moving edges [22]. This method calculates heat flow by non-maxima suppression and hysteresis thresholding in each image of a sequence of frames, of edge data extracted using the Sobel operator. The approach shows better performance than the basic 2-D Sobel operator, for all levels of Gaussian noise, without using background subtraction. As the result of moving edge map extraction demonstrate in Figure 2.2, this approach extracts moving objects effectively. Figure 2.2 (a) shows images of a walking subject from a video sequence; the results of isotropic or linear heat analysis are shown in Figure 2.2 (c). In an iterative process, the discrete form of the isotropic equation is given by

\[
E_t^n = E_t^{n-1} + k \Delta E_t^{n-1} = E_t^{n-1} + k (E_{t+1}^{n-1} + E_{t-1}^{n-1} - 2E_t^{n-1})
\] 2.5
where $E^{n-1}_t$, $E^{n-1}_{t-1}$ and $E^{n-1}_{t+1}$ are images at iteration $n - 1$ resulting from the process using the Sobel operator and anisotropic diffusion are shown in Figure 2.2 (b). And then, the total amount of heat flow is calculated between an initial scale 0 (zero) and a final scale $n$. The process of whole scale is described by Equation 2.6.

$$E^n_t = E^0_t + k \sum_{i=0}^{n-1} \Delta E^i_t$$  \hspace{1cm} 2.6

(a) Frame$_{i+1}$, Frame$_i$ and Frame$_{i-1}$

(b) $E_{i+1}$, $E_i$ and $E_{i-1}$

(c) total heat flow

(d) HFO (–)

Figure 2.2 Extraction of moving edge map
Chapter 2

Then, the total heat flow can be expressed as follows

\[ |E^i - E^0| = k \sum_{i=0}^{n-1} |\Delta E^i| \]  

This equation expresses heat flow in and heat flow out together as shown in Figure 2.2 (c). Since the total heat flow out (HFO) is considerable in the reference frame \( E_t \) the following is the conditional equation and the moving edge map is shown in Figure 2.3 (d). These processes can extract the human subject successfully whilst the background is ignored.

\[ HFO = k \sum_{i=0}^{n-1} |\Delta E^i| \quad \Delta E^i < 0 \]  

There are three models of heat flow: conduction, convention and radiation. The conduction model is chosen to find shapes and to prove the usage of the theory. The temperature change over time at each point is given by Equation 2.9 using the general heat conduction or diffusion equation.

\[ \frac{dT}{dt} = \alpha \left( \frac{d^2T}{dx^2} + \frac{d^2T}{dy^2} \right) + Q = \alpha \nabla^2 T + Q \]  

where \( \nabla \) represents gradient and \( \alpha \) is thermal diffusivity of the material. The term \( Q \) is the source that applies internal heating. This equation provides the temperature distribution over a material body and it depends on time, distance, heat source, properties of material and specified conditions of initial and boundaries.

Now, the above equation is transformed to a two dimensional formula for pixel-based images. Initial and boundary conditions are given by \( T(x, t = 0) = 0 \) and \( T(x, t) = 0 \), which means the temperature inside the body is initially zero and the boundary condition is Dirichlet in that the temperature at the boundaries is zero. If the initial heat source at a point inside the body is positive and continuous constant, heat can diffuse from the initial point to other points. This means that the temperature of all points inside body will be not zero. All temperature values of the objects and background on the grey scale image are retained in another image symbolised \( I \) and its initial condition for the whole image is zero, \( I(x, t = 0) = 0 \). The following control equations in the heat conduction equation can be satisfied above condition.
\[
\frac{dI(x,t)}{dt} = CF(x,t) \left( \alpha \nabla^2 I(x,t) \right) + Q(x) \tag{2.10}
\]

where \(I(x,t)\) is an image pixel value which represents the temperature at each point and time. \(\alpha\) has a specific interval \(0 \leq \alpha \leq 0.25\) for the numerical scheme to be stable in a two dimensional system and \(Q(x)\) is the source term. \(CF(x,t)\) is the Control Function (CF), which is obtained from the region statistics of source location on a grey-level image. Equations 2.9 and 2.10 are similar to each other, then, the logical decision at each iteration in each position is shown in Equation 2.11.

\[
CF(x,t) = \begin{cases} 
1, & \sigma_1(x,t) \leq \sigma_2(x,t) \\
0, & \sigma_1(x,t) > \sigma_2(x,t)
\end{cases} \tag{2.11}
\]

where \(\sigma\) is the image variance, \(\sigma_1\) and \(\sigma_2\) denote the inside and the outside of the object. Heat diffusion can be generated inside the target object by the control function, and the proposed Dirichlet condition that the temperature on the boundaries is zero can be kept.

Geometric heat flow (GHF), which is a form of anisotropic diffusion for image denoising and enhancement, is used at the next stage of finding shapes. It decreases the curvature of shapes in the images while removing noise. Edge directions are affected by the tangents of the feature boundaries of an image \(B\). While \(\eta\) is the direction normal to the feature boundary through a point as the gradient direction and \(\tau\) is the tangent direction, these two variables are orthogonal. Therefore, the rotationally invariant Laplacian operator can be transformed to the summation of the second order spatial derivatives \(B_{\eta\eta}\) and \(B_{\tau\tau}\).

\[
\frac{dB}{dt} = \alpha \nabla^2 B = \alpha \left( B_{\eta\eta} + B_{\tau\tau} \right) \tag{2.12}
\]

The GHF equation eliminating the normal diffusion and preserving the tangential diffusion is shown in Equation 2.13.

\[
\frac{dB}{dt} = \alpha B_{\tau\tau} = \alpha \frac{B_{xx}B_y^2 - 2B_{xy}B_xB_y + B_{yy}B_x^2}{B_x^2 + B_y^2} \tag{2.13}
\]

Figure 2.3 shows the segmentation of a human lung image (123 X 118) by CF + GHF, together with the result by the Active Contour Without Edges (ACWE) [23] for comparison. Figure 2.3 (a) displays the initial contour for ACWE and the
Chapter 2

(a) initial contour and the source position  (b) segmentation by CF + GHF  (c) segmentation by ACWE

Figure 2.3 Examples of segmentation of human lung image

source position for CF. Figure 2.3 (b) and (c) show the segmentation by CF + GHF with the final result by the white contour and ACWE. These results demonstrate that CF + GHF is better for extracting boundaries of features, which are in the middle and the bottom of the image, than ACWE. These concept of heat analogy can be extended to multiscale Fourier-based descriptions [24] for feature extraction in computer vision.

2.1.3 Force

One of physical analogies for feature extraction in computer vision is a transform which uses an analogy to gravitational force based on pixel brightness and distance between pixels. The transform, called the force field transform [3,25], generates a force field from vector operations and highlights the contour shape of objects. This approach focuses on a novel representation for object boundaries in an active contour model while a solution to dependency initialisation is still an open challenge [3]. This approach was developed for ear biometrics where identification aims to recognise the shape of people’s ears, which does not change with age.

There are two novel stages. The first step is that the values of an image are converted with a force field transformation. The force field transform can be generated at each pixel of an image by the following equation

\[ F_i(r_j) = P(r_i) \frac{r_i - r_j}{|r_i - r_j|^3} \]  \hspace{1cm} 2.14

where the force \( F_i(r_j) \) is calculated between two position vectors \( r_i \) and \( r_j \) at points with coordinates \((x_i, y_i)\) and \((x_j, y_j)\) and pixel intensity \( P(r_i) \). A total force
with vector lengths indicating magnitude at a single position as illustrated in Figure 2.4. This potential energy field can be calculated by the following equations, which are similar to the force field Equations 2.14

\[
E(r_j) = \sum_{i \in 0,N-1 | i \neq j} E_i(r_j) = \sum_{i \in 0,N-1 | i \neq j} \frac{P(r_i)}{|r_i - r_j|}
\]

2.15

The potential energy field at a particular position generates a potential energy surface, which is smoothly varying because of the principles of equipotential. For example, the potential energy surface for an ear is shown in Figure 2.5. The smooth surface of energy allows a test source to reach extrema where the gradient is zero, and these extrema can be the descriptive features of the image.

Figure 2.6 (a) shows an original image of an ear and the magnitude of the force field transform of the image is illustrated in Figure 2.6 (b). One clear advantage of the force field transform is that it offers a combination of edge detection with smoothing. This can be seen in Figure 2.6 (a) where the edges of the shape are clearly shown, and the effects of the hair are reduced by the smoothing process. Difficulties with noise affecting the detected positions of extrema led to enhancement by convergence feature extraction using a ‘convergence map’, which has more information than force field lines, from the divergence of a vector field. This process is the second step in this approach.

The divergence of a vector field is a differential operator that produces a scalar field at each field point. The vector force field is given by
Figure 2.5 Potential energy surfaces of a human ear image

(a) original  (b) result

Figure 2.6 The force field transform results

\[
div \mathbf{F}(\mathbf{r}) = \lim_{\Delta V \to 0} \frac{\oint \mathbf{F}(\mathbf{r}) \cdot d\mathbf{S}}{\Delta V}
\]

where \(d\mathbf{S}\) is the outward normal to a closed surface \(S\) enclosing an incremental volume \(\Delta V\). It can be formulated in two-dimensional coordinates as following

\[
div \mathbf{F}(\mathbf{r}) = \nabla \cdot \mathbf{F}(\mathbf{r}) = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y}
\]

where \(F_x\) and \(F_y\) are components of \(\mathbf{F}\). The concept of divergence is graphically described in Figure 2.7. The directions of the arrows in Figure 2.7 (a-c) show positive, negative and zero divergence.

Two important concepts of this method, potential energy wells and channels representing peaks and ridges in the function value, can create a more general description mathematically. The force field could be transformed to a scalar field
by a new function called the force direction convergence field $C(r)$ as following

$$
C(r) = -\text{div } f(r) = -\lim_{\Delta A \to 0} \frac{\oint f(r) \cdot dl}{\Delta A} = -\nabla \cdot f(r) = -\left( \frac{\partial f_x}{\partial x} + \frac{\partial f_y}{\partial y} \right)
$$

2.18

where $f(r) = F(r)/|F(r)|$, $\Delta A$ is the incremental area and $dl$ is the (outward normal to the boundary. This function has merit since both positive and negative values are meaningful: negative values correspond to force direction divergence. The Marr-Hildreth operator (LoG) is similar in operation, but the Marr-Hildreth operator is linear whilst the convergence operator (based on force direction rather than force [26]) is nonlinear. This nonlinearity causes the order of the operation to be unchangeable: $\text{div}(\text{grad}/|\text{grad}|) \neq (\text{div } \text{grad})/|\text{grad}|$.

The convergence field of an ear image is shown in Figure 2.8. The field lines and the convergence map are displayed in Figures 2.8 (a) and (b), and a small magnified version of the force direction field is shown in Figure 2.8 (c) from the region highlighted in Figure 2.8 (a) and (b). The values in the convergence map range from 0 to 255, so that negative convergence values corresponding to anti-channels are close to black while the positive values in channels are close to white. Figure 2.8 (c) shows the patterns in the force field direction by correlation between the convergence function and the field line features, with the correspondence between the converging arrows and the white ridges and between the diverging arrow and the black ridges. It supports the notion that the convergence map contains more information, which is demonstrated more in the negative version of wells and channels or anti-wells and anti-channels,
than the earlier field lines. The superposition of the field line map from the field line features in Figure 2.8 (a) and the convergence map in Figure 2.8 (b) is shown in Figure 2.8 (c). It illustrates that the channels correspond to white ridges in the convergence map while the potential wells correspond to the convergence peak.

2.1.4 Pressure

One of the approaches using a physical analogy was formulated to separate touching objects in intersecting regions for 3D image analysis, for example, to counting the number of coins inside a jar which was imaged as a composite archaeological artefact. It is difficult to separate objects connected by a region of similar density in 2D-based images, hence motivating the use of pressure in 3D volumetric data [27]. This approach does not require a template or previous knowledge about the regions of interest, and traditional techniques based on connectivity by morphological analysis could not be sufficiently flexible for such a sensitive situation. The new approach can be applied to images directly and automatically.

The pressure analogy addresses the separation of objects which have similar density or distance between them. A one-sided threshold is applied to calculate the volume of the background and obtain a pressure domain, which describes the connected areas between objects.
\[ P_{x,y,z} = 1/ \sum_{x,y,z \in W} V_{x,y,z} \quad \forall V_{x,y,z} \in \text{background} \quad 2.19 \]

where \( W \) is a \( 5 \times 5 \times 5 \) sized window, \( V \) is the volume threshold to remove background data and \( P \) is the pressure domain. The pressure data is constructed using a high pass filter response to retain consistency and provide a smooth transform. The filter derives a mapping function extracted from a high pass filter with the pressure level \( p_f \) in ascending order as following

\[ V_o = v_i \frac{Z_R}{Z_R + Z_C} \quad 2.20 \]

where \( v_i = 180, Z_C = -j(1/2\pi p_f C), Z_R = 1500 \) and \( C = 1.06 \times 10^{-6} \) with a cut-off frequency \( p_f \text{ cut off} = 100 \) for \( p_f \in [1.125] \). This preserves and emphasises the high pressure value in the top 25 levels. The next threshold condition is obtained by \( P' = P(V_o) \).

\[ M_{x,y,z} = \begin{cases} 1 & \text{if } P'_{x,y,z} \geq \text{threshold} \\ 0 & \text{otherwise} \end{cases} \quad 2.21 \]

The labelling is the next step to retain each element data in 4D matrix. And then, dilation and erosion are repeated three times to obtain the final mask \( M_{final_{x,y,z}} \), which removes small data. The mask is combined with a logical thresholded volume to recover the volume of data in previous stage as follows

\[ R_{x,y,z} = M_{final_{x,y,z}} \times \text{logical}(V_{x,y,z}) \quad 2.22 \]

where \( R_{x,y,z} \) is the recovery volume and it only exists in the mask.

Figure 2.9 illustrates examples of two synthetic disks at each step of this approach. These disks are in contact at their middle in Figure 2.9 (a). Based on Equation 2.19 the pressure domain is calculated using local pressure, and the image slices and the rendered touching surface in the thresholded pressure domain are shown in Figure 2.9 (b) and (c), respectively. After that, component shapes are separated by multiplying the mask with the image using Equation 2.22. Figure 2.9 (d) show the result that the connecting objects have indeed been separated.
This approach can be extended with the pressure mask effect, logical operations and labelling for automatically separating connected structures which have similar appearance by image density in medical images [28], for instance, an example of final segmentation in a CT image is shown in Figure 2.10. In the extended stage, the labelled mask is expanded using the statistical mode, and a specific label is then obtained in object volume $F$ without data loss.

$$F_{x,y,z} = \text{Mode}(R_{x,y,z} \in W) \quad \forall R_{x,y,z} \in \text{foreground}$$  \hspace{1cm} (2.23)

The last stage is that morphological closing operation is applied to every labelled object three times, and it is also closing any open gaps in detected objects.

$$B_{\text{final}i} = F_i \cdot S \quad \forall i \in 1: \text{number of labels}$$  \hspace{1cm} (2.24)

where $B_{\text{final}}$ has volume data of a segmented object in a 4D matrix.
2.1.5 Other approaches

Another novel approach uses a physical elastic mesh model for image segmentation [29]. This model is based on the analogy of physical strain and stress of elastic objects, so two virtual forces, called the repulsion force (as strain) and the elastic force (as stress), are proposed. The repulsion force can be obtained from the difference in grey values between neighbouring pixels, and it builds $z$ coordinates. The elastic force is then calculated by considering the borders of objects by the differencing the $z$ coordinates while two positive coefficients are applied to each force. The values of the $z$ coordinates are calculated by two individual forces at each $x$ and $y$ coordinate, and the resultant force is constructed. The resultant force updates after each iteration and finishes when the condition is satisfied. It has functional similarity to a merging method and the performance is acceptable when segmenting regions in the images. Also, X. Zhuang and N. E. Mastorakis use electromagnetic interaction via ‘current-carrying wires’ to analyse the local changes of object edges between two related images [30]. The proposed virtual current can generate the virtual magnetic force field, and then the force affects the virtual force distribution which can
represent the local change of object position or the object’s shape in image sequences. This method is based on grey-scale images, so it needs further analysis for application to colour images.

## 2.2 Basic methods for feature extraction

This thesis includes several theoretical techniques, practical techniques and their results in image processing for feature extraction, thus, the techniques will be briefly explained in this chapter.

### 2.2.1 Sobel operator

Edges are a fundamental element in image interpretation for low-level feature extraction. Many approaches, which were focused on highlighting image contrast, were developed and applied for edge detection. The Sobel operator is one of the most popular edge detection techniques and is based on vertical and horizontal operators by first-order differentiation in Equation 2.25 and 2.26 [40].

\[
E_{x,y} = |P_{x,y} - P_{x+1,y}| \quad \forall x \in 1, N - 1; y \in 1, N
\]

\[
E_{y,x} = |P_{x,y} - P_{x,y+1}| \quad \forall x \in 1, N; y \in 1, N - 1
\]

The vertical edges \(E_x\) and the horizontal edges \(E_y\) in Image \(P\) can be combined to an operator \(E\) as in Equation 2.27.

\[
E_{x,y} = |2 \times P_{x,y} - P_{x+1,y} - P_{x,y+1}| \quad \forall x, y \in 1, N - 1
\]

These equations influenced the development of the Sobel edge-detection operator which consists of two masks to determine the edge in vector form. The 3×3 masks are shown in Figure 2.11.

![Figure 2.11 Templates for the Sobel operator](image)

Figure 2.11 Templates for the Sobel operator
The process of applying the Sobel operator is that the Sobel templates are first convolved with a target image. Then, the outputs of the process are the edge magnitude ($E_{mag}$) and direction ($E_{dir}$) in vector form. The edge direction can be arranged to be normal to the edge itself by swapping the Sobel templates, according to the convention required for the chosen shape or area.

The larger templates, which are 5×5, 7×7 or more, have increased smoothing to reduce noise. However, an increase in smoothing generally implies an increase in blurring and blurred edges are less likely to provide accurate edge information. Thus, basic 3×3 Sobel masks will be sufficient for this thesis.

2.2.2 Hough Transform

The Hough Transform (HT) [15] is an efficient version of template matching in high-level feature extraction wherein edge points vote for accumulator cells by using a voting function that describes the shape that is sought in the image.

We shall first use the polar HT for lines [55]. The point where this line intersects the line in the image is given by

$$\rho = x \cos \phi + y \sin \phi$$

2.28

where $\rho$ is the distance between the origin and the point where the lines intersect, and $\phi$ is the angle of the line normal to the line in the image. Figure 2.12 illustrates the polar consideration of a line.

The range of two parameters $\rho$ and $\phi$ can be set of the value $[0, \sqrt{N^2 + M^2}]$ and $[0, 180^\circ]$ respectively, where $N \times M$ is the image size. Thus, the parameter space

![Figure 2.12 The polar consideration of a line](image)
Chapter 2

is bounded which it is not in the basic HT (using Cartesian coordinates).

We shall then deploy the HT for circles by replacing the curve equation as following in explicit or parametric form. The explicit form is considered by the equation of a circle following Equation \[ (x - x_0)^2 + (y - y_0)^2 = r^2 \] and the parametric form is defined as Equation \[ x = x_0 + r \cos \phi; \quad y = y_0 + r \sin \phi \]

These equations can be used to transform the accumulator space from \( x - y \) coordinates to \( x_0 - y_0 \) coordinates by circle radius \( r \) as following Figure 2.13, which describes an example how four points match to four circles between two spaces.

Since the HT is only used to evaluate the precision of line and circle detection in this thesis, the standards of a line and a circle (\( \rho, \phi, x_0, y_0 \) and \( r \)) are given by values which are manually measured from the salient object in the image. Therefore, the proper range of standards will be provided to the HT for objects for any image that contains linear or circular objects.

### 2.2.3 Colour models

Various colour models are applied to refine feature extraction in computer vision when the information is insufficient in grey-scale images. Thus, it is important to select an appropriate colour model for the proposed application.

![Diagram](image)

Figure 2.13 Constructing the accumulator space for the HT for circles
The RGB colour model, which uses base colour components close to the wavelengths of visible light (red, green and blue), is an addictive model and one of the basic colour models. The RGB model can display colours as shown in a cube Figure 2.14 (a). This model focuses on representing colour in the reproduction system. For this reason, the combination of RGB components is not able to express colours consistent with human interpretation. The HSV colour model is arranged to be more similar to human vision than the RGB colour model. The three components H, S and V in this model are related to a colour’s hue, saturation and brightness. The relation between the RGB and the HSV colour model is shown in equations [40], and the schematic colour model is illustrated in Figure 2.14 (b). The range of H is set of 0—360° and both S and V are set from zero to one. Figure 2.14 (c) shows a more practical version of the HSV model as the colour is close to black if V is close to zero.

\[
\begin{align*}
M &= \max(R, G, B) \\
m &= \min(R, G, B) \\
D &= M - m
\end{align*}
\]

\[
H = \begin{cases} 
\frac{1}{6} \times \frac{G - B}{D} & \text{if } R = M \\
\frac{1}{3} + \frac{1}{6} \times \frac{B - R}{D} & \text{if } G = M \\
\frac{2}{3} + \frac{1}{6} \times \frac{R - G}{D} & \text{if } B = M 
\end{cases}
\]

\[
S = \frac{D}{M} \\
V = D
\]

The HSI colour model is similar to the HSV colour model. The three components, H (Hue), S (Saturation) and I (Intensity), are calculated from the RGB model as shown in Equation 2.32. If B is larger than G, H is rearranged to \(360° - H\).

![Figure 2.14 Examples of colour models](image-url)

(a) RGB model  (b) HSV model (cylinder)  (c) HSV model (cone)

Figure 2.14 Examples of colour models [58]
\[ H = \cos^{-1}\left\{ \frac{1}{2}\left(\frac{(R - G) + (R - B)}{\sqrt{(R - G)^2 + (R - B)(G - B)}}\right) \right\} \]
\[ S = 1 - \frac{3}{(R + G + B)} [\min(R, G, B)] \]
\[ I = \frac{1}{3}(R + G + B) \]

Many other colour models have been researched to express colours in a manner similar to human vision. We shall compare these colour models to choose a suitable model when studying colour-feature extraction.

### 2.2.4 SIFT, SURF and CSIFT

Since the Sobel operator is one of classic approaches to extract edge features in image processing, other more recent methods are required to compare the performance with the Sobel operator in this thesis. These three approaches are represented by low-level localised feature extraction.

The Scale-Invariant Feature Transform (SIFT) [45] has been introduced to extract local keypoints with difference-of-Gaussians in scale space. Difference-of-Gaussian images are derived by subtracting adjacent Gaussian images from repeated convolution in scale space as shown in Figure 2.15. Local maxima and minima can be SIFT keypoints from comparing neighbour scales in a Gaussian pyramid. Each SIFT keypoint is described by a 128 element feature vector that contains a 4x4 array of orientation histograms with 8 orientation bins as shown in Figure 2.16. The gradient and orientation are calculated at each image point and weighted by Gaussian window within a circle in Figure 2.16 (a). These point values are accumulated into orientation histograms in Figure 2.16 (b) and the histograms with bins can be SIFT keypoints. These keypoints are analogous to other keypoints when the image is rotated, scaled or illuminated. However, the volume of computation is large and only grey-scale information is used to extract keypoints.

Speeded up robust features (SURF) [49] has been improved to reduce the amount of computation in SIFT. For this, there are three procedures; using the integral image approach with a Hessian matrix, smaller detectors and descriptors based on a summation of Harr wavelet (with fewer demensions) and a simple matching
using the sign of the Laplacian function. These methods successfully find features much faster, however the accuracy is lower than SIFT. SURF also uses only grey-scale image information. For this reason that other approaches using colour information have been investigated, aimed to extract coloured objects.

Coloured SIFT (CSIFT) [52] focuses on extracting objects of specific colour using a local invariant descriptor in colour images. The one of suggested colour
Chapter 2

Invariant descriptors [52] is calculated in accordance with the incoming spectrum $E(\lambda)$ containing three spectral measurements $\hat{E}$, $\hat{E}_\lambda$ and $\hat{E}_\lambda\lambda$ in Gaussian colour model based on Gaussian weighted spectral energy distribution [48]. The components can be simply computed from Gaussian colour model in RGB terms and the transform is shown in Equation 2.33.

$$
\begin{bmatrix}
\hat{E} \\
\hat{E}_\lambda \\
\hat{E}_\lambda\lambda
\end{bmatrix} =
\begin{bmatrix}
0.06 & 0.63 & 0.27 \\
0.3 & 0.04 & -0.35 \\
0.34 & -0.6 & 0.17
\end{bmatrix}
\begin{bmatrix}
R \\
G \\
B
\end{bmatrix}
$$

We shall arrange the descriptor in the Gaussian colour model with SURF for performance comparison with a traditional operator (Sobel) in this thesis.

2.2.5 Earth Mover’s Distance and other distance metrics

It is often necessary to compare the similarity of distributions. The Earth Mover’s Distance (EMD) reflects the minimal cost of work that must be performed to transform between two signatures [53]. The first stage of the EMD is to find an optimal solution of transportation problem, and the result can show numerical value which indicates how similar two clusters are, using the weights representing the clusters. The following sets are the features of two histograms:

$$P = \{(p_1, \omega_{p_1}), \cdots, (p_m, \omega_{p_m})\}, \quad Q = \{(q_1, \omega_{q_1}), \cdots, (q_n, \omega_{q_n})\}$$

where $P$ and $Q$ denote each cluster and $p_i$ is the $i$-th element of a cluster and $\omega_{p_i}$ is the value of the index in the cluster $P$ (of size $m$). Also, $q_j$ and $\omega_{q_j}$ are set in the cluster $Q$ (of size $n$) in a similar manner. The ground distance matrix $D_{ij}$ is determined by the ground distance $d_{ij}$ between $p_i$ and $q_j$. In the case of a flow matrix $F_{ij}$, we choose the minimum value $f_{ij}$ between $\omega_{p_i}$ and $\omega_{q_j}$, and there are the following conditions for calculating the EMD.

$$f_{ij} \geq 0 \quad 1 \leq i \leq m, 1 \leq j \leq n$$

$$\sum_{j=1}^{n} f_{ij} \leq \omega_{p_i} \quad 1 \leq i \leq m$$

$$\sum_{i=1}^{m} f_{ij} \leq \omega_{q_j} \quad 1 \leq j \leq n$$

26
The summation of flow amount $f_{ij}$ from cluster $P$ to $Q$ when the index of $P$ is $i$ cannot be over the weight $p_i$ in Eqn. 3.5, whilst all flow $f_{ij}$ is above zero in Eqn. 3.4. In the same way, the summation of flow amount $f_{ij}$ from cluster $Q$ to $P$ when the index of $Q$ is $j$ cannot be over the weight $q_j$ in Eqn. 3.6. Thus, the total flow amount $f_{ij}$ is the minimum of the weight amount in $P$ and $Q$ in Eqn. 3.7, and Eqn. 3.8 shows the result of EMD. A simple example of two histograms using EMD is shown in Fig. 3.9. When there are two clusters $P$ to $Q$ the (a) result of EMD is smaller than (b) result. The EMD is affected by clusters’ weights and the optimal flow so that the suitable optimal problem algorithm should be determined first.

The Bhattacharyya distance, measures similarity of two discrete or continuous probability distributions using Bhattacharyya coefficient [56]. The coefficient is calculated as

$$Bhattacharyya(H_1, H_2) = \sum_{i=1}^{n} \sqrt{p_i q_i}$$

![Diagram](image-url)
Chapter 2

for two histograms $H_1 = \{p_1, \ldots, p_n\}$ and $H_2 = \{q_1, \ldots, q_n\}$; $p_i$ and $q_i$ are the probability at index $i$. The Bhattacharyya coefficient is close to one when the similarity of two histograms is high, whereas the coefficient is zero when there is no similarity between them. This is one of the methods to evaluate the similarity among clusters, therefore, a suitable measure should be chosen according to the data under investigation. The bin-by-bin similarity measures (Bhattacharyya coefficient, the sum of squared distance, etc.) are insensitive to colour variations and sensitive to bin size. On the other hand, cross-bin similarity measures (EMD) hardly suffer from cross-bin information to measure the similarity [57]. Any distance measure is appropriate to a particular application and the EMD will be applied to analysing histogram patterns in coloured object segmentation in the final chapter in this thesis.

2.3 Conclusions

As mentioned previously physical analogies using natural properties have many advantages: the theory is already well established, the techniques are easily understood, and have simple formulations. Some are rather cumbersome, using forms of region growing whereas others require Fourier analysis for implementation. Most of all, the concept of using an analogy to light appears highly intuitive. Cummings [7,8] used Snell’s law applied on grey scale image directly which appeared to perform well in comparison with the Sobel operator, which is also well-known and one of the simplest methods. This concept and the property of the rays can extend to further approach in various ways, therefore, we will develop the ray analogy further and propose a novel approach to extend the analogy in a new way.

Also, several fundamental techniques for feature extraction in computer vision have been outlined and will be deployed later for analysis in this thesis. These approaches will be applied across the board for both improvement of the basic thesis and for comparison with improved versions.
Chapter 3: Image Ray Transform (IRT)

3.1 Overview

Cummings and Nixon [8] suggested the modelling of propagation of light through image data as a new method of low level feature extraction. The basic Image Ray Transform (IRT) analogises the progression of light via Snell’s law, in pixel-based images and is an approach unique in its basis on a natural phenomenon and with demonstrated efficacy.

3.2 The Image ray transform

The IRT uses Snell’s law within images using reflection and refraction of a light ray as shown on Figure 3.1. The amount of reflection and refraction depends on the materials’ refractive indices and the incoming ray’s angle of incidence $\theta_i$. If two media $n_t$ and $n_{t+1}$ have different refractive index the ray will reflect or refract on the boundary of the two media according to the critical angle $\theta_c$ where $\theta_c$ is given by Snell’s law. These are given by following Equations 3.1 and 3.2.

$$\theta_c = \sin^{-1}\left(\frac{n_{t+1}}{n_t}\right)$$

3.1

where $\theta_l$ and $\theta_r$ represent the angles of reflection and refraction, respectively. When the incoming ray meets a boundary where the material indices change and $\theta_c$ is larger than $\theta_i$, the ray refracts at an angle $\theta_r$ given by Equation 3.2.

Figure 3.1 Ray Propagation
\[
\frac{\sin \theta_i}{\sin \theta_r} = \frac{n_{t+1}}{n_t}
\]  

When \( \theta_c \) is smaller than \( \theta_i \), the ray reflects at an angle \( \theta_i \) that has the same magnitude as \( \theta_i \). However, the direction of \( \theta_i \) is opposite to that of \( \theta_i \), according to a normal that is orthogonal to the boundary. These theories relating to reflection and refraction of rays are the basic rules of the IRT.

The material indices \( n_t \) and \( n_{t+1} \) are replaced by the gray level of 2-D images which have values from 0 to 255, and a maximum refractive index \( n_{\text{max}} \) is usually set to 256 (just beyond the maximum gray level). Equation 3.3 defines the refractive material indices \( n_i \) in the images as

\[
n_i = \left( \frac{i}{255} \right) \times (n_{\text{max}} - 1)
\]  

One of the vital parameters for the IRT is the normals \( (N) \) that are used to calculate directions when the ray refracts or reflects. The normal direction \( \psi \) can be estimated using the direction of the Sobel operator \( (E_{\text{dir}}) \) by Equation 3.4

\[
N = (\sin \psi, \cos \psi), \quad \psi = E_{\text{dir}} + \frac{\pi}{2}
\]  

The above equations give the foundations to calculate the magnitude and directions for reflection and refraction in the basic IRT. After calculating the normals for initialising the materials’ refractive indices matrix of the whole image, the initial position \( (x_0,y_0) \) and direction \( \varphi_0 \) of each ray are selected at random. The bounds on these selections are shown in Equation 3.5

\[
0 \leq x_0 < w, \quad 0 \leq y_0 < h, \quad 0 \leq \varphi_0 < 2\pi
\]  

where the width of the image is \( w \) and the height of the image is \( h \), respectively. The angle \( \varphi_t \) describes the direction of the ray. An initial position vector \( p_0 \) and direction vector \( V_t \) are:

\[
p_0 = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}, \quad V_t = \begin{pmatrix} \sin \varphi_t \\ \cos \varphi_t \end{pmatrix}
\]  

and these are used to calculate the position of the ray at the following iteration, the position \( p_t \) at time \( t \):

\[
p_t = p_{t-1} + V_t
\]
When a ray meets a boundary where the material refractive indices change, we assign the angle of incidence $\theta_i$ and $n$:

$$\cos \theta_i = N \cdot V_t, \quad n = \frac{n_t}{n_{t+1}} \tag{3.8}$$

As the iteration number increases, position vectors and direction vectors are updated by material indices and the incidence angle $\theta_i$. Next Equation 3.9 describes how to determine the reflection vector $R_l$ and refraction vector $R_r$ according to $n_t, n_{t+1}, \theta_c$ and $\theta_i$ using Equations 3.1-3.8.

$$V_{t+1} = \begin{cases} R_l & \text{if } n_t > n_{t+1} \text{ and } \theta_i > \theta_c \\ R_r & \text{if } n_t > n_{t+1} \text{ and } \theta_i < \theta_c \\ V_t & \text{otherwise} \end{cases} \tag{3.9}$$

If $n_t > n_{t+1}$ and $\theta_i > \theta_c$ the ray reflects with reflection vector $R_l$:

$$R_l = V_t - 2(N \cdot V_t)N \tag{3.10}$$

When the ray is in the condition $n_t > n_{t+1}$ and $\theta_i < \theta_c$ it refracts with refraction vector $R_r$:

$$R_r = nV_t + (n(N \cdot V_t) - \cos \theta_r)N \tag{3.11}$$

where $\cos \theta_r$ is

$$\cos \theta_r = \sqrt{1 - n^2(1 - N \cdot V_t)} \tag{3.12}$$

Pseudo code for the operation of the IRT is given in Function 1. Here, the process operates for $M$ random initialisations. The number of initialisations $M$ can be set from 1 (for a single trajectory) to 10,000 (generally used in this thesis) or further. A single trajectory which has a randomised initial start progresses until the max number of iterations `maxPathIndex` unless it meets a boundary of image. The `maxPathIndex` can vary according to image size and user setting and it is usually set to 800. Rays reflect or refract according to the refractive indices calculated using Equation 3.3, with reflection controlled by Equation 3.9 to include shape sensitivity. The resulting image is normalised for display, and it can emphasise the area where objects exist much more than the background.
Function 1

\[ \text{Normal} \leftarrow \text{SobelDirection} + \pi/2 \]

\[ \text{RefractiveMatrix} \left( \text{InputImage}, \text{nMax} \right) \]

\[
\text{for } t \leftarrow 0 \text{ to } M \text{ do }
\]

\[
\begin{align*}
x_0 & \leftarrow \text{Rands}(0, \text{width}), \ y_0 \leftarrow \text{Rands}(0, \text{height}), \ \varphi_0 \leftarrow \text{Rands}(0, 2\pi) \\
V_{0,x} & \leftarrow \cos \varphi, \ V_{0,y} \leftarrow \sin \varphi
\end{align*}
\]

\[
\text{for } i \leftarrow 1 \text{ to } \text{maxPathIndex} \text{ do }
\]

\[
\begin{align*}
\text{if } n_1 > n_2 \text{ and } \theta_i > \theta_c \text{ then } & \ x_i, y_i, V_i \leftarrow \text{Reflect}(x_{i-1}, y_{i-1}, V_{i-1}) \\
\text{else if } n_1 > n_2 \text{ and } \theta_i < \theta_c \text{ then} & \ x_i, y_i, V_i \leftarrow \text{Refract}(x_{i-1}, y_{i-1}, V_{i-1}) \\
\text{else} & \ x_i \leftarrow x_{i-1} + V_{i-1,x}, y_i \leftarrow y_{i-1} + V_{i-1,y}, V_i \leftarrow V_{i-1}
\end{align*}
\]

\[ \text{ResultAccumulator} \left( \text{result}, x_i, y_i, t \right) \]

\[ t = t + 1 \]

end

\text{normalise} \left( \text{result} \right)

\text{return result}

---

### 3.3 Image ray transform analysis

The performance of the IRT is influenced by the grey level of the image and the number of iterations. Figure 3.2 demonstrates how the result of the basic IRT in an image of a synthetic circle \((w \times h): 512 \times 512 \text{ pixel, radius is 100})\) varies when the iteration number changes. As the number of iterations increases, the IRT tends to distinguish objects more clearly from the background in the image as the paths in the background are removed after normalisation. Also, the results for other objects - e.g. a rectangle and a curve - are similar to those for a circle.

The IRT can extract various types of objects from the background more clearly as seen in Figure 3.3 (All synthesised images have 512×512 pixel with 8 bit depth. A radius of (a):100, \(w \times h:\) of (b): 200×200, a radius of (c):60, \(w \times h:\) of (c):100×70, a radius of (d):75, \(w \times h:\) of (d):150×150). In particular, the IRT appears better suited to finding curves than it is to finding circles or squares although the result is sufficient to find circular and rectangular objects in the image. Therefore, the IRT appears suited to ear biometrics and determining retinal structure by its ability to detect curves.
Figure 3.2 The result of changing the number of iterations $M$

(a) Original
(b) $M = 1$
(c) $M = 10$
(d) $M = 100$
(e) $M = 1000$
(f) $M = 10000$

Figure 3.3 Examples of detecting mixed shapes in images

(a) a circle
(b) a rectangle
(c) a circle and a rectangle
(d) a circle, a rectangle and a curve
3.4 On extending the image ray transform

So far there have been no extensions made to the original IRT. The focus of this thesis concerns adapting properties of the rays to achieve different selectivity. In the original IRT the rays were either singular or multiple and rigidly adhered to the Snell's law formulation. As the original IRT only employs an analogy, it is possible to alter the analogy and change the way the rays' progress through an image. We shall show how this can be used to make the IRT specific to shape and to colour: to select different objects based not only on how the rays interact with image data, but also to include the effect of different shape and colour.
Chapter 4: Adding shape selectivity to the IRT

4.1 Overview

This chapter describes the new extension to the IRT adding capability for detection and characterisation of intrinsic image structure. Figure 4.1 illustrates an example of the aim of this extension, and uses a synthetic image containing a square and a circle in Figure 4.1(a). The original IRT provides the result in Figure 4.1(b), showing the edges of the square and of the circle - and the result of applying the Hough transform is to find neither the line nor the circle, precisely. By introducing the new shape factor, the process can be arranged to

![Figure 4.1 Illustrating results of the original Image Ray Transform (IRT) and the extended IRT, with the Hough transform results superimposed showing that the circle and the line have been emphasised by different selection of the new shape factor.](image)

(a) original image  
(b) original IRT  
(c) extended IRT with shape factor = 1.5  
(d) extended IRT with shape factor = 1.9
select circles, as in Figure 4.1(c) – where the circle is correctly detected, or lines, as in Figure 4.1(d). This is achieved by changing the value of a single parameter – the new shape factor.

4.2 Extending the IRT

For wider application, further analyses are required to improve the performance of the basic IRT. We propose an extension to the IRT that modifies its principal operations to imbue capability to extract chosen shape.

The extension is derived by quite a simple process. It appears that addition of a single parameter introduces an ability to select between different objects, highlighting one shape – e.g. a circle – in preference to another – e.g. a square. This ability is not shared by other techniques: to extract a circle by curvature, one would need to determine areas of constant curvature and this implies a two-step process. Snakes can be predisposed to be attracted to objects of a different shape, by suitable selection of priors; the Hough transform for circles and evidence gathering for lines differ clearly in the basis equations for describing the shape, rather than in the way the evidence gathering process is defined. As such, we shall show an extension which although simple, appears likely to have significant impact on the generality and use of the image ray transform.

![Figure 4.2 The Ray Propagation in the Standard and Extended IRT](image-url)
The extension is focused on the reflection stage of the IRT since reflection influences emphasis of the edge of objects more effectively than refraction. The basic propagation compared with the standard IRT and the extended IRT is described above in Figure 4.2 when the reflection angle $\theta_l$ is changed by the shape factor. The extended Equation 4.1 is a modified version of Equation 3.10 that controls the amount of reflection, so it is definitely important to choose a suitable shape factor $F_s$ for images that are comprised of different objects. The modified version including the shape factor is given in Equation 4.1, for the same symbols as in chapter 3.

$$R_l = F_s \times V_t - 2(N \cdot V_t)N$$ \hspace{1cm} (4.1)

Most setup options of the extended IRT are similar to the original ones, however, $n_{\text{max}}$ is different. The original IRT focuses on highlighting curved objects more than other shaped or linear objects [7], therefore, it is suitable for ear biometrics and extraction of retinal blood vessels. In the contrast, since this thesis aims to detect various type of objects with shape sensitivity, the original value of $n_{\text{max}}$ can lead to difficulty in shape discrimination. Thus, the value of $n_{\text{max}}$ will be set to 256 in this thesis.

### 4.3 Variation of the shape factor

A new parameter, the shape factor, is a fundamental point in this thesis since it can influence the quality of the extended IRT. Thus, analysing a variety of the shape factors needs to be studied in images which include a circle, a rectangle or a mixture of shapes. In this section, synthesised images are used to analyse the possible range of values for the shape factor.

Analysis of selected images was repeated 100 times with the same parameter settings, each time with different randomly selected initialisations. The results for finding a circle by using the standard Hough Transform are shown with a box plot of the error analysis. The errors in estimating the accuracy of the extended IRT as a pre-processing stage are derived from the difference in the values of the parameters used to describe the shapes as:

$$E_\rho = |\rho_m - \rho_o|; \quad E_\phi = |\phi_m - \phi_o|$$ \hspace{1cm} (4.2)

$$E_c = \sqrt{(x_m - x_o)^2 + (y_m - y_o)^2 - (r_m - r_o)^2}$$ \hspace{1cm} (4.3)
when $E_\rho$, $E_\phi$ and $E_c$ describe the errors in the values estimated for $\rho$, $\phi$ (for a line) and a circle, respectively. The parameter $\rho_m$ is the distance between the origin and the line, and $\phi_m$ is the orthogonal vector angle from the origin by the Hough transform. $x$, $y$ and $r$ are the $x$ center position, the $y$ center position and radius of the circle by the Hough Transform. The values for the subscript $m$ denote the result of the Hough Transform and the values of the subscript $o$ are the original values used in image synthesis. And then, the boxplots using Equation 4.2 and 4.3 by Matlab are utilized to demonstrate the capability of the shape factor. Figure 4.3 is an example of the boxplot. The horizontal axis shows the shape factor value from one to two, and the vertical axis displays the error value from zero to 90 (according to the image chosen for analysis). The elements enlarged from part of the box plot inside the green square illustrate principal statistical variables, which are the median (a red line in a blue box), the 25% quantile (bottom line of a blue central box), the 75% quantile (top line of a blue central box), and the remaining data range (the maximum and minimum values denoted by vertical broken lines extending from a blue central box) and outliers (red crosses). These elements of the boxplot can display the properties of test

Figure 4.3 An Example of the error boxplot
results more easily. For example, two green ellipses describe the status of error. The vertical ellipse in the left part of the boxplot shows that an error value is too flexible to determine a representative value, because the interquartile range of an example box is too wide. Therefore, the error value is uncertain in this interval and it shows lower stability. Conversely, the error values converge from left to toward right during the whole shape factor range and became stable as shown in the horizontal ellipse. It means the shape factor within this interval is an optimal solution. Thus, we decide to use boxplots for displaying the results of the extended IRT in this thesis.

Figure 4.4 shows examples of results where the shape factor varies from 0.25 to 5 in an image including an ideal circle. The result can be acceptable to emphasise the edge of a circle when the factor is between one and two, however, the ray tends to be divergent if the factor is under one and over two. The other results including divergence of rays are inadequate to accentuate the shape edge, so we consider only the period where the shape factor is between one and two. The example of rectangle and mixture of different shapes are shown in Appendix B. Fig 4.5 demonstrates a box plot of the result when the factor changes between one and two in the image. According to this box plot the result is acceptable when the shape factor is over 1.70, where the error is consistently small. Figures 4.6 and 4.7 show that the result of the image which combines a circle and a rectangle is similar to the result in Figures 4.4 and 4.5. The error value in Figure 4.7 tends to increase rapidly when the factor is around two, whilst the error values of the image mixed objects are larger than the image including only one synthesised circle. Except for the confusion interval, which the error value increases dramatically when the shape factor is around two in Figure 4.7, the error value patterns of two images in Figure 4.5 and 4.7 are similar and minimum value periods where $F_s$ is between 1.70 and 1.90 also correspond approximately to each other.

As the number of iterations increases, the extended IRT tends to distinguish objects more clearly from the background in the image as the paths in the background are removed after image normalization. The new shape factor $F_s$ tends to reduce the amount of reflection so that rays trace better the edge of the image object. If the factor is set between zero and one, the results are unacceptable for extraction of the circle’s boundary; conversely the rays scatter
Figure 4.4 The result according to varying the shape factor in Figure 3.3 (a)
Figure 4.5 The box plot of a synthesised image (Figure 4.4)

Figure 4.6 The result according to varying the shape factor for Figure 3.3 (c)
Chapter 4

Figure 4.7 The box plot of a synthesised image (Figure 4.6) when $F_s$ exceeds two. However, the result can be acceptable if the shape factor is set between one and two. After comparing the results of the original IRT (where $F_s = 1$), and the extended IRT, the object in the image detected using the extended IRT reveals the edges of constituent objects more clearly. The study of combining low and high level methods for feature extraction and segmentation has been accomplished in computer vision, for example [31,32].

The most appropriate technique for comparison is the Hough Transform (HT) as this is known to provide optimum results. To be able to compare results after applying the HT in the extended IRT results we have developed a heuristic method which is suitable for analysis of synthesised images. The heuristic techniques highlights the ability to find appropriately selected points, the HT shows how well shapes can be matched to the data. The heuristic method operates after thresholding the result of the extended IRT, and the position and radius of a circle are derived by averaging the coordinates of the points in the thresholded image. A line is detected by applying least squares analysis to the points resulting from the IRT. This provides results which can be compared with the result of applying the standard HT, to be used in Equation 4.2 or 4.3.
Figure 4.8 shows the effect of choosing different values for the shape factor. Figure 4.8 (a) and (d) are the original images containing a circle and a line, and Figure 4.8 (b,c,e,f) show the results of basic IRT and extended IRT with a shape factor $F_s=1.0$ and $F_s=1.8$, respectively. Superimposed are the results of applying the Hough Transform for lines and for circles. The result of the shape factor 1.0 is the same as the result of basic IRT and shows ability to detect the line rather than the circle: one shape appears in the output of the IRT and so the line is correctly extracted and the circle is determined at an arbitrary position. When $F_s=1.8$ the circle is detected as the size of the circle in the image is close to the original circle while the detected line is incorrect. The results in Figures 4.9, 4.10 and 4.11 describe how the error value by the Hough transform and the heuristic method change with the shape factor. The plot shows the shape factor variation from one to two along the horizontal axis and the error value on the vertical axis. In the case of finding an ideal circle, the minimum interval of the error is between $F_s=1.7$ and 2.0 in Figure 4.9 (a), and minimum interval of the error value in finding a line is between $F_s=1.0$ and 1.3 in Figure 4.10 (a) and 4.11 (a), thus motivating the values selected previously. It is important to assign a suitable radius commensurate with the shape type of object in advance for finding specific objects in the image. On the other hand, the minimum intervals

![Figure 4.8 Results of extended IRT by $F_s$ in a synthesised image](image-url)
Figure 4.9 Analysing the outputs for circle detection
Chapter 4

Figure 4.10 Analysing the outputs ($\rho$) for line detection

(a) error ($\rho$) for line detection by HT

(b) error ($\rho$) for line detection by the heuristic method
Figure 4.11 Analysing the outputs ($\phi$) for line detection

(a) error ($\phi$) for line detection by HT

(b) error ($\phi$) for line detection by the heuristic method
of the error value using the heuristic method to find a circle and a line are \( F_s = 1.0 \sim 1.3 \) and \( 1.77 \sim 1.78 \) in Figure 4.9 (b) and two plots (Figures 4.10 (b) and 4.11(b)), respectively. Comparing the result of the Hough transform and the heuristic method, the results of Hough transform are closer to a zero error value and the interval of the error value with near zero is longer than the results for the heuristic method. However, there is a definite difference in computation time between two methods. Although the result of the Hough transform is more stable the time taken by the Hough transform is much greater, especially for a circle.

### 4.4 Analysing images by the extended IRT

The histogram of the reflection angles after applying extended IRT shows different patterns for different object shapes [33]. For the histogram, the horizontal axis describes the orientation between \( d = 0^\circ \) and \( d = 180^\circ \) and the vertical axis is the normalised values at each orientation. The normalised values \( N_d \) is defined in Equation 4.4 and 4.5 as a moving average to emphasise where discontinuities occur as following

\[
N_d = \frac{1}{T} \sum_{t=1}^{T} C_t
\]  

where \( T \) is the number of samples in the moving average window and where \( C_t \) is the summation of angles between two successive points of a light ray.

\[
C_t = 180 - (\theta_{lt} + \theta_{rt})
\]  

Each histogram is described by the structure in Figure 4.12 and is repeated for \( 0^\circ \) to \( 90^\circ \) and underneath for \( 90^\circ \) to \( 180^\circ \). For angles of \( 0^\circ \) and \( 180^\circ \) degrees the rays progress without in change in direction. If either of them is significantly large this could affect to the normalisation of the result where other values present the feature of objects. Thus, histograms represent the value of orientation from \( 1^\circ \) to \( 179^\circ \). Figure 4.13 shows that the histogram patterns vary with objects in the same image. The histogram of a circle has a particular pattern related to its constant curvature whilst the histogram for a rectangle has a uniform distribution. The process follows the principle of superposition in that separate shapes in the same image produce the same histogram as that derived from combining the analysis of the objects in separate images. The histograms
preserve a shape's own characteristics even though there are two objects in the same image. Thus, analysing a histogram of an image can show which shapes are present and how many objects are in the image. By this we suggest a novel extension to the image ray transform which confers object specificity into the feature extraction process.

The performance of the extended IRT using histogram pattern analysis is analysed for invariance with respect to different selections for the shape factor and for experimental factors, such as number of iterations $M$, the shape factor

![Figure 4.12 Basic Histogram Structure](image)

![Figure 4.13 The histograms of objects in various images](image)
$F_s$, and object position, orientation and size. The analysis here is by visual patterns of the histogram to illustrate how the technique operates, as no analytical evidence is possible by the nature of the IRT. We shall later move to use distance metrics for comparison purposes.

In order to study the effect of the number of iterations, we first fix the shape factor since it influences the quality of the extracted object. The factor tends to reduce the amount of reflection so that rays trace better the edge of the image object. After comparing the results of the basic IRT and the extended IRT in Figure 4.4, the object in the image using the extended IRT reveals the edge more clearly. Figure 4.14 shows how the histogram patterns are similar for different numbers of iterations ($M$), when the shape factor $F_s$ is 1.0 for iteration numbers varying from $M = 100$ to $M = 10000$. In the case of a circle and a line the patterns are similar and consistent with the number of iterations. Even

![Figure 4.14 Histogram patterns by shape and iteration number when $F_s = 1.0$](image-url)
though there are several scale differences, larger feature histogram values still exist at the same points. Although the patterns for a rectangle tend to be unstable when the number of iterations is low, the patterns show consistency when the iteration number increases.

Though the histogram patterns change little with a larger number of iterations, the value of the shape factor can shift the pattern and scale of histogram values. Figure 4.15 shows the variation in histogram patterns with change in the shape factor. The patterns are shifted and stretched as $F_s$ increases, and the values of the histogram are sufficient to distinguish the pattern of the object from the background. In this Section the patterns of a rectangle also tend to be unstable when the shape factor varies, the resulting patterns showing consistency for the same shape factor.

The position of objects in images cannot influence the histogram patterns, as shown in Figure 4.16. When the object’s sizes are similar, the histogram patterns do not vary with change in the objects’ positions. Similarly, the patterns are

![Histogram patterns](image)

Figure 4.15 Histogram patterns by shape factor variation when $M=10000$
hardly affected by the orientation of the objects as shown in Figure 4.16. The scales of the histogram in each image change slightly when the curve orientation varies, but the patterns preserve their properties. In case of size there is a small difference with change in position and orientation. The histogram values are scaled by object size instead of shifting the histogram patterns. As the radius of a circle increases the magnitudes of the histogram values decrease in inverse proportions, thus the magnitudes $N_d$ decrease with increase in radius, as shown.
Chapter 4

in Figure 4.17. The positions of the elements of the histogram \( N_d \) do not change, reflecting retention of the circle’s structure. All detailed histograms are shown in Appendix C.

4.5 Applying the shape factor in real images

Synthesised images and real images from the Caltech 256 dataset [34] that contains multiple images of 256 object categories for evaluation of object recognition in computer vision, and real images are used to demonstrate the ability of the extended IRT (and later the adaptive colour factor). Other tested images are shown in Appendix A. The effects of applying the extended IRT to real images are more distinct than the results in synthetic images. Figure 4.17 (a) is the original image which \( w \times h \) is 420x420 with 24 bit depth. This type of colour image can be transformed to grey level image for applying the extended IRT using Equation 4.6

\[
grey\ value = 0.4124\ r + 0.3576\ g + 0.1805\ b \quad \text{4.6}
\]

where \( r, g \) and \( b \) are the transformed values from each colour channel of red, green and blue in RGB colour model.

Figure 4.18 (b-d) are the normalised outputs of the IRT with differing values of the shape factor \( F_s \), showing the emphasis of different shapes. By Figure 4.18 (c) choosing \( F_s = 1.5 \) highlights the boundary of the circle for clearly than Figure 4.18 (b) and omits the text on the book’s cover. When \( F_s \) is changed to 1.9 the boundary of the book’s spine is much clearer than the boundary of the golf ball (which is now greatly reduced). The line markings on the book’s cover are also highlighted when \( F_s = 2 \). Figure 4.18 (e-g) are the results of applying the HT superimposed on the output of the IRT. Although finding a line and a circle looks effective when \( F_s = 1.5 \), the two images Figure 4.18 (e) and (g) show discrepancy between the result using Hough Transform and visual results (the result by visual inspection). Figure 4.19 (a-c) show that the shape factor interval for a line and a circle can again be separated, where the minimum error value interval of line is over 1.2 and the minimum error value interval of circle is under 1.6. This is the inverse of that experienced with synthesized images and appears to be due to the nature of the brightness variation.
On the other hand, the result in Figure 4.20 (b) shows that a circle and a line cannot be found when the shape factor is 1.0. The minimum error value interval of the line and the circle are between 1.5 and 1.9 and between 1.4 and 1.6, respectively. Figure 4.20 (c) shows an example where a circle and a line can be detected when the shape factor is set within the determined minimum intervals for a line (Figures 4.20 (e) and (f)) and a circle (Figure 4.20 (g)). If the factor is set outside of this interval the possibility of missing the objects will increase as shown in Figure 4.20 (d). This suggests that there is an appropriate shape factor for each object if the objects are contained in a single image. Other examples of the extended IRT applied to real images are shown in Figure 4.21. The radius for HT in the result of Figure 4.21 (b-d) is the same for both circles and the radius Figure 4.21 (f-g) is set to the small circle near the bottom of the image (the top of the yellow golf tee).
Figure 4.19 Analysing the IRT outputs for a real image.
Figure 4.20 The results with the error value of an example image.
To verify the performance of the extended IRT, the Sobel operator as a classic approach is chosen with the original IRT. Table 1 and 2 shows the result of the Sobel operator, the original IRT and the extended IRT using the shape factor by the Hough Transform. The results are evaluated by Equations 4.2 and 4.3, and tested images are shown in Appendix A. Although the result of the Sobel is acceptable to find shape objects, the extended IRT result using the suitable shape factor is also reasonable for detecting each shapes and several results (for example, for line: image number 088_0047, for circle: tomato2 and tomato 3) are even better than the Sobel result. It can validate the shape capability of the extended IRT using the shape factor.

4.6 Conclusions

The results of applying the new shape factor show the possibility of using the Image Ray Transform to search for the correct position of a particular shape by appropriate choice of the shape factor. The value of the shape factor depends not only on the shape, but also on the brightness of the image. The process appears to be able to determine the edges of chosen shapes faster than conventional feature extraction techniques. Also, it appears possible to describe objects by their histogram pattern. We shall continue to refine the analysis of its abilities in real images, doubtless requiring refinement of the techniques’ basis.
Table 1 Comparison results of approaches using the Hough Transform for line

<table>
<thead>
<tr>
<th></th>
<th>$\rho$ (pixels)</th>
<th>$\phi$ (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sobel</td>
<td>origin IRT</td>
<td>extended IRT</td>
</tr>
<tr>
<td>088_0047</td>
<td>52.7</td>
<td>252</td>
</tr>
<tr>
<td>061_0088</td>
<td><strong>1.43</strong></td>
<td>346</td>
</tr>
</tbody>
</table>

Table 2 Comparison results of approaches using the Hough Transform for circle

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>circle</td>
<td>Sobel</td>
<td>original IRT</td>
</tr>
<tr>
<td>088_0047</td>
<td><strong>2.00</strong></td>
<td>189</td>
</tr>
<tr>
<td>061_0088</td>
<td><strong>1.41</strong></td>
<td>6.35</td>
</tr>
<tr>
<td>tomato1</td>
<td><strong>2.65</strong></td>
<td>53.9</td>
</tr>
<tr>
<td>tomato2</td>
<td>5.10</td>
<td>29.2</td>
</tr>
<tr>
<td>tomato3</td>
<td>173</td>
<td>23.3</td>
</tr>
</tbody>
</table>
Chapter 5: Detecting objects using colour information

5.1 Adding colour within the IRT

Techniques using colour information have been researched for some time to detect features in computer vision. Colour information can be a constructive element to extract features of objects in various fields of computer vision [35,36] and represented by different colour spaces. For example, Chuang [37] used a ratio histogram of colour images of videos for object detection. Most of all, the usage of colour models has pervaded preprocessing and analysis stages for feature extraction, and the importance of the colour model selection [38,39] is naturally important when analysing colour images. The basic IRT uses only grey values of the image after converting the colour image to a grey image or choosing a channel in the RGB colour space. Thus, it is key which colour space should be chosen and how it applies to an extended IRT system.

The CIE $L^*a^*b^*$ colour model is similar to human vision and it is composed of three planes; $L^*$ shows the lightness and $a^*$ and $b^*$ planes are digitised chrominance of red/green and yellow/blue, respectively [40]. In order to obtain values of CIE $L^*a^*b^*$ model, the process employs a transform from the RGB model to xyz model, the definition is given by [54]

\[
\begin{align*}
    x &= 0.4124 r + 0.3576 g + 0.1805 b \\
    y &= 0.2126 r + 0.7152 g + 0.0722 b \\
    z &= 0.0193 r + 0.1192 g + 0.9505 b
\end{align*}
\]  

where $r$, $g$ and $b$ are the transformed values from each colour channel of red, green and blue. The transformed $xyz$ values can changed to CIE $L^*a^*b^*$ model by Equation 5.2 and 5.3, while $x_n$, $y_n$ and $z_n$ are the tristimulus values of the reference in the $xyz$ model.

\[
\begin{align*}
    L^* &= 116 f \left( \frac{y}{y_n} \right) - 16 \\
    a^* &= 500 \left[ f \left( \frac{x}{x_n} \right) - f \left( \frac{y}{y_n} \right) \right] \\
    b^* &= 200 \left[ f \left( \frac{y}{y_n} \right) - f \left( \frac{z}{z_n} \right) \right]
\end{align*}
\]  

5.1
\[ f(s) = \begin{cases} \sqrt[3]{s}, & s > 0.008856 \\ 7.787s + \frac{16}{116}, & s \leq 0.008856 \end{cases} \]

5.3

For discrimination by colour we propose an approach to use planes containing only the colour information, which can be used to constitute materials’ refractive indices in the IRT. The refractive materials’ value \( n_{ij} \), of size \( i \times j \) (width \( \times \) height), is determined by a weighted combination of the colour spaces as

\[ n_{ij} = (1 - \lambda_c)a_{ij} + \lambda_c b_{ij} \]

5.4

where \( a_{ij} \) is the \( a^* \) plane and \( b_{ij} \) is of \( b^* \) plane of the CIE \( L^*a^*b^* \) model that are combined using the parameter \( \lambda_c \), which is called the colour factor. If the colour factor is negative, the refractive matrix value \( n_{ij} \) emphasises blue shapes. On the contrary, \( \lambda_c > 1 \) means that refractive values are close to green. This means that the refractive materials’ value after conversion with Equation 5.4 can be used to distinguish almost all colours by the colour factor \( \lambda_c \) expect achromatic ones. The range of values for the colour factor is demonstrated in the results.

5.2 Comparing different colour spaces

Various colour spaces have been applied in computer vision systems with different attributes. The following figure shows a comparison of the performance of different colour models. In order to detect each colour circle three threshold values are used in the RGB model, and the colour factor was applied to the HSV, HSI and the CIE \( L^*a^*b^* \) models before the IRT stage. In this stage, we use the \( H \) and \( S \) planes for applying the colour factor in the HSV and HSI model. Hence, Equation 5.4 will be applied to the HSV and HSI colour models in the same ways as following

\[ n_{ij} = (1 - \lambda_c)H_{ij} + \lambda_c S_{ij} \]

5.5

where \( H_{ij} \) is the \( H \) plane and \( S_{ij} \) is the \( S \) plane in two models. \( V \) plane of the HSV model and \( I \) plane of the HSI model are not considered from the same reason that \( L^* \) plane of CIE \( L^*a^*b^* \) model is not applied. Figure 5.1 (a) is a synthesised image (\( w \times h: 512 \times 512 \), bit depth: 24) containing a red, green, blue and a white circle. The threshold values in the RGB model are \( r > 100, g < 100, b < 100 \) in Figure 5.1 (b). Also, the threshold values of (f) and (j) are set by \( r < 100, g < \)
100, b > 100 and r < 100, g > 100, b < 100, respectively. These three results show that each colour object can be detected precisely exactly by controlling the parameter selection, however, three parameters need to be chosen at the same time. Meanwhile, the results of the HSV and HSI model are inappropriate to detect the red circle in Figure 5.1 (c-d) while the circle is emphasised more than other circles in CIE L*a*b* model result (e). Moreover, the results for detecting the green circle in Figure 5.1 (k-m) are similar to the result for the red circle. On the other hand, when the colour factor is negative the IRT result with the HSI model can extract only the blue circle whilst there is no result of the HSV model in Figure 5.1 (g). Although the other circles appear in the result of CIE L*a*b* model the blue circle is most emphasised in the IRT result. These results demonstrate that CIE L*a*b* model has an advantage to find a specific coloured object with the IRT.

![Figure 5.1 Comparison of colour models to detect each coloured shape](image-url)
One performance measure is to determine the average grey level in the detected object $\Gamma$, compared with the average level for the background. For a manually labelled object $\Gamma$ in an image $I$ (and a measure of area $A$), the performance $P(\Gamma)$ measure is then

$$P(\Gamma) = \frac{\sum_{x,y \in \Gamma} I_{x,y} / A(\Gamma)}{\sum_{x,y \notin \Gamma} I_{x,y} / A(\Gamma)}, \quad 0 < P(\Gamma) < \infty$$

and larger values of the performance value indicate good performance. The compared performance between CIE $L^*a^*b^*$ colour model and other models is demonstrated in Table 3. The RGB colour model has the best performance while the HSV, HSI colour models are inconsistent in the detection of each coloured object. However, there are three operators to detect only one colour in the RGB model (for example, white has a high value in each plane). Other colour models use only one factor for extraction, and CIE $L^*a^*b^*$ model has better performance than the HSV and HSI models for both red and green.

Table 3 The performance value $P$ of different colour models

<table>
<thead>
<tr>
<th></th>
<th>RGB model</th>
<th>HSV model</th>
<th>HSI model</th>
<th>CIE $L^*a^<em>b^</em>$</th>
</tr>
</thead>
<tbody>
<tr>
<td>red</td>
<td>62.70</td>
<td>4.31</td>
<td>5.98</td>
<td>13.59</td>
</tr>
<tr>
<td>blue</td>
<td>58.99</td>
<td>0</td>
<td>59.87</td>
<td>9.19</td>
</tr>
<tr>
<td>green</td>
<td>60.12</td>
<td>4.21</td>
<td>5.85</td>
<td>8.09</td>
</tr>
</tbody>
</table>

### 5.3 Detecting coloured objects

The results of the IRT using a colour factor varying from minus one to two are shown in Figure 5.2 and Figure 5.3. These examples demonstrate that there is a suitable colour factor to detect a specific coloured object. Figure 5.2 (a) in a synthesised image containing a red, blue, green and a white circle. As the result of not using the $L^*$ plane, we can predict that it is difficult to choose an appropriate value of the colour factor to detect a white circle. Blue can be emphasised when the colour factor is negative as shown in Figure 5.2 (b), while red and green are evident between zero to one and one to two in Figure 5.2 (c-d) respectively. If the colour factor is less than minus one, the results are similar to the result for $\lambda_c = -1$. Similarly, results where the colour factor exceeds two are comparable to the ones when $\lambda_c = 2.0$. 

62
Figures 5.2 (a-c) describe the error value according to the variation of the colour factor with each colour. For the blue circle the interval between -1 and 0 is the best while the error value of the red circle and the green circle are close to zero between zero to one and one to two, respectively. These box plots show that a suitable colour factor can be chosen so as to detect an object that has a specific colour. And then, there are real image examples of the colour factor by the previous box plots in Figure 5.4. Figure 5.4 (b) shows that the brightest part is within a blue tee when the colour factor $\lambda_c$ is -0.8. However, (c-d) show that further processing stage is needed for extracting coloured objects more clearly although it is possible to recognise target objects.

Figure 5.5 shows the results of choosing $r$ channel and thresholding by each $r$, $g$ and $b$ factor with the shape factor $F_s$ in the $RGB$ colour model. The shape factor $F_s$ is chosen as an example value. It is unsuited to extracting a red tee using the $r$ channel as shown in Figure 5.5 (b-d) because the red value of the area of white golf balls is as high as the value of the red tee area. Therefore, we apply thresholding for comparison purposes. Three threshold values are needed to extract object of specific colour. The threshold values for Figure 5.5 (e) are $r > 100, g < 100, b < 100$. The threshold values of (i) and (m) are set by $r < 100, g < 100, b > 100$ and $r < 100, g > 100, b < 100$ in similar way.

As shown in Figure 5.5 the shape factor does not influence the resulting threshold. Even though the results of thresholding by three factors in the $RGB$
Chapter 5

(a) error for a blue circle detection by HT

(b) error for a red circle detection by HT

(c) error for a green circle detection by HT

Figure 5.3 Analysing the outputs by $\lambda_c$ variation in synthesised image.
Figure 5.4 Results of extended IRT by $\lambda_c$ in a real image.

Figure 5.5 Thresholding results of extended IRT by $\lambda_c$ in real image.
colour model are more distinct than the result of CIE L*a*b* colour model, there is a requirement to choose three threshold values (and results are sensitive to that choice) for extracting specific coloured objects in the former method while the latter method requires choice of only the colour factor. Also, it is difficult to select suitable values because of the invariance to brightness. Brightness also affects the CIE L*a*b* colour model, however, the amount of this affect is less than other colour models since the formulation for the colour factor does not contain the L* plane, which shows the lightness. Thus, the colour factor has an advantage to control the value and elicit colour information unaffected by brightness.

5.4 Adjustment

After comparing colour models we confirm the use of the CIE L*a*b* colour space for extending the IRT. Example results for detecting a red object are displayed in Figure 5.6. Figure 5.6 (a) is a real image from the Caltech 256 dataset and others are the result of the IRT used to detect the red (golf) tee in the image, (b) is the IRT result after thresholding with the red colour in the pre-processing stage. The colour factor is applied to the HSV, HSI and the CIE L*a*b* colour models in Figure 5.6 (c-e) respectively. Figure 5.6 (c) shows that the tee is emphasised enough for recognition, however other objects (a yellow tee and a white ball) are highlighted more than the red tee when the colour factor $\lambda_c$ is 1.0. In case of the HSI model it is difficult to determine the red object in Figure 5.6 (d). On the other hand, the most emphasised object in the image based on CIE L*a*b* model is the red tee with $\lambda_c = 0.3$ as shown Figure 5.6 (e). Although (e) is less clear than (b), the RGB-based approach needs the thresholding stage for each colour. In contrast the approach using the CIE L*a*b* is simpler since quality is controlled by only the – single – colour factor, and the result in CIE L*a*b* colour space is more reasonable to detect a specific coloured object than the result in the HSV model. If the colour factor is set within an appropriate range in the CIE L*a*b* model, it can be deployed to extract proper objects of differing colour.
Chapter 5

Figure 5.6 Detecting a red object with different colour spaces

One potential area for application of this new technique is systems that can automatically detect ripe fruit; another is to select differently coloured objects of the same shape. The application to ripe fruit detection is illustrated in Figure 5.7 which shows it is possible to select between ripe (Figure 5.7 (b)) or unripe (Figure 5.7 (c)) tomatoes by different choice of the colour factor. Were the system applied to determine the proportion of ripe fruit, $\lambda_c$ could be selected to have a value under 0.8 as shown boxplots in Figure 5.8 (a), whereas a value exceeding 1.1 is more suited to detection of unripe fruit in Figure 5.8 (b).

Figure 5.7 The example outputs by the colour factor.
(d) error for a red circle detection by HT

(e) error for a green circle detection by HT

Figure 5.8 Analysing the outputs by the colour factor variation.
Figure 5.9 shows another example result concerning colour factor variation and Figure 5.10 demonstrates how the colour factor can influence on the result. To find a blue object $\lambda_c$ should be set to a negative value, such as $-0.8$, and the error value of Figure 5.9 (b) is lowest when the shape factor is between 1.4 and 1.6 while generally whole range is close to zero error value in Figure 5.10 (a). Also, the suitable value of $\lambda_c$ is between zero and one to find a red object in Figure 5.9 (c) and the whole range of the shape factor is close to zero error value in Figure 5.10 (b). In the case of a green object the result is similar to blue and read results although the error value is higher than others when the shape factor is around one in Figure 5.10 (c). However, the case of the green object can depend on the image’s coloured objects is illustrated in Figures 5.9 (d) and (h). If there is a yellow object in an image, the result of extended IRT tends to find a yellow object.

The another examples in Figure 5.11 show the result of the extended IRT using the colour factor, with the line and circle extracted by the Hough Transform in an image that contains red, blue, yellow and green objects. The results in Figure 5.11 (b-d) show examples when the colour factor $\lambda_c$ is $-0.7$, 0.3 and 1.9, respectively (the shape factor $F_s$ 1.0 is applied at all examples). When the factor $\lambda_c$ is $-0.7$ and 0.3, the red ball and red croquet stick are detected correctly even though the background appears noisy in Figure 5.11 (c). On the other hand, the yellow ball and croquet stick can be extracted when $\lambda_c$ is 1.9. These results show
Chapter 5

Figure 5.10 Analysing the outputs by $\lambda_c$ variation in real image.
Figure 5.11 Results of extended IRT by $\lambda_c$ in a real image.

the difference in comparison with the above result, because the chroma and shadows influence the result. Analysing the statistical result of this example can explain the differences of the method either combining the shape factor and the colour factor or using a single factor. This implies a need for more study of the two factors and their adjustments will discussed in the next chapter.

5.5 Conclusions

Clearly, adding colour sensitivity confers greater capability to any operator. The \textit{CIEL*a*b*} colour space can provide better discrimination about colour than other colour models and it improves the performance of the extended IRT with the colour factor.

In order to demonstrate those factors in real images, we need a dataset comprised of the factor that the technique is designed for. As this technique is now sensitive to colour, we need real images of coloured shapes. This becomes more difficult, especially for outdoor scenes, since as yet the use of colour in image segmentation is not widespread and as yet there appear to be no
Chapter 5

appropriate databases containing various type of shape and colour at the same image, that can be used for analysis of coloured object detection.
6.1 Overview

To extract features more effectively in computer vision, especially since it is more difficult to ascertain objects from the background in natural images, studies combining several techniques have been examined. One scheme combined colour and shape information for object recognition [41] and another model used local variation and boundaries [42], Colour combination using a top-down colour attention method [43] and the database model called SEEMORE [44] are examples of approaches to improve the performance of feature extraction. Therefore, in this chapter we combine shape with colour perception aiming to detect specific colours of a specific shape, for example to extract a red circular shape, in a single combined technique.

Figure 6.1 shows the result of the new combined technique with the same image of Figure 5.11 (a). The shape factor can influence to find a circle of a croquet ball in a correct position in Figure 6.1 (b) and (d) when \( F_s = 1.1 \), and then a specific coloured object can be selected by the colour factor \( \lambda_c = 0.3 \) for a red ball in Figure 6.1 (b) and \( \lambda_c = 1.6 \) for a yellow ball in Figure 6.1 (d)). It is the same the case of detecting a line. When the shape factor can extract correct positions of a croquet stick while \( F_s \) is around 1.8, the colour factor can judge which colour croquet stick should be selected by its variation in Figure 6.1 (c) and (e). In this chapter analysing the statistical result of this example can explain the differences of the method either combining the shape factor and the colour factor or using a single factor.

6.2 Operation

To combine shape specificity with colour perception we suggest that the extended IRT combining both the shape factor and the colour factor is used. Naturally, even though either the shape factor or the colour factor can be used to extract shapes when deployed alone, the combined method should derive greater performance than a single independent method for coloured object
extraction. In addition, the two factors can be applied as a preprocessing stage for feature extraction. We now show results for a small set of images of croquet equipment as these images contain examples different coloured shapes: coloured circles (the croquet balls) and coloured lines (the croquet sticks) which allows for study of the effect of the combined techniques.

The combination process focuses on the separate two stages. The first stage is to change the colour space of the target image to the CIE $L^*a^*b^*$ colour space with the colour factor for colour sensitivity, and which is followed by application of the extended IRT, with the shape factor for shape adaptivity. This sequence given in Function 2 shows the entire stage of the extended IRT including the stage of IRT detailed at Function 1 and equations in previous chapter 3-5. This step is simple and independent, therefore, it can be straightforward to control
the factors. The examples below are the result of the combination using two factors by the Hough Transform in real images.

<table>
<thead>
<tr>
<th>Function 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>Load(InputImage)</code></td>
</tr>
<tr>
<td><code>ConvertRGB_CIELAB(r, g, b)</code></td>
</tr>
<tr>
<td><code>SetConvertedMatrix(a^*, b^*, \lambda_c)</code></td>
</tr>
<tr>
<td><code>IRT(convertedMatrix, F_s)</code></td>
</tr>
</tbody>
</table>

As shown in Figures 6.2 to 6.5 boxplots display how the shape factor can influence the error value while the colour factor is \(-0.7\). The error values of two red objects are calculated by the extended IRT using the shape factor and the colour factor in Figure 6.2. The red line and circle can be found at the correct positions when the shape factor is over 1.3 and under 1.5, respectively. The proper range of the error value, which is calculated for red objects detected by the HT in Figure 6.2, is much closer to zero than the error value for yellow objects and green objects in Figure 6.4 and 6.5. However, the extended IRT could not find the blue objects at the correct position as shown in Figure 6.3. The suitable value of the colour factor for detecting blue objects appears to be between \(-1\) and zero with the synthesised image in the previous chapter (Figure 5.3 (a)). This implies that there is the difference between synthesised images and outdoor images since the outdoor images are affected by natural light and background, and the colours are not necessarily consistent.

To compare with the result of using only the shape factor, boxplots are shown in Figures 6.6 to 6.9. There is no boxplot that shows zero error value except in two cases, which involve finding the red line when the shape factor is under 1.6 (Figure 6.6 (a)) and the yellow circle when the shape factor is under 1.1 (Figure 6.8 (b)). From these charts it appears that the shape factor has discriminative capability for shape, but no capability for colour. Thus, the combination of the shape factor and the colour factor is a simple and efficient way to extract specific coloured objects in images.

The result of combining two factors can be represented much clearly in Figure 6.10. When only the shape factor is applied to the image in Figures 6.10 (b) and (c), a circle whether ripe tomato or unripe tomato is detected. This shows that the shape factor can influence the detection of an object’s shape. However, the
Figure 6.2 Analysing the outputs when $\lambda_c$ is $-0.7$. 

(a) error for a red line detection by HT 

(b) error for a red circle detection by HT
Figure 6.3 Analysing the outputs when $\lambda_c$ is $-0.7$. 

(a) error for a blue line detection by HT

(b) error for a blue circle detection by HT
(a) error for a yellow line detection by HT

(b) error for a yellow circle detection by HT

Figure 6.4 Analysing the outputs when $\lambda_c$ is −0.7.
Figure 6.5 Analysing the outputs when $\lambda_c$ is $-0.7$. 

(a) error for a green line detection by HT 

(b) error for a green circle detection by HT
Chapter 6

(a) error for a red line detection by HT

(b) error for a red circle detection by HT

Figure 6.6 Analysing the outputs without the colour factor
Figure 6.7 Analysing the outputs without the colour factor
Figure 6.8 Analysing the outputs without the colour factor
Figure 6.9 Analysing the outputs without the colour factor
results with applying both the shape factor and the colour factor show that this combination can classify ripe and unripe tomatoes distinctly in Figure 6.10 (d-g). At the first stage the colour factor accentuates objects with a specific colour, and then the shape factor emphasises specific shape of objects at the next stage. Even though the shape factor is the same, the result clearly differs according to the selection of the colour factor (Figures 6.10 (d) and (f), and (e) and (g)). Figure 6.11 illustrates the results of how to find red objects, which include ripe tomatoes and unripe tomatoes (the same image as Figure 5. 7 (a)) when only the shape factor is applied and the shape factor and the colour factor (0.5) are applied together, respectively. The result with the colour factor (Figure 6.11 (b)) is much more stable than the result without the colour factor (Figure 6.11 (a)), and the number of outliers is much smaller. Figure 6.11 also shows the result of finding green objects in the image. The means and quantiles of the result without the colour factor are significantly greater than zero in Figure 6.12 (a) while the statistical variables of the result with two factors in Figure 6.11 (b) are much closer to zero. These boxplots show that the colour factor is beneficial to the extraction of specific coloured objects in images.

Table 4 demonstrates the error value of the extended IRT without the colour factor $\lambda_c$ and the extended IRT with the colour factor $\lambda_c$ from Figure 6.11 and 6.12. The error value is calculated by Equation 4.3. As shown in chapter 4 the result of using only the shape factor is good enough to extract specific shapes, however, the error value increases by the standard which is set to specific coloured objects. The error value when the standard is set for a red circle is much less than the error by a standard green circle, because the result is

<table>
<thead>
<tr>
<th></th>
<th>The extended IRT without $\lambda_c$ (pixel)</th>
<th>The extended IRT with $\lambda_c$ (pixel)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F_s &lt; 1.6$</td>
<td>$F_s &gt; 1.6$</td>
</tr>
<tr>
<td>Red circle</td>
<td>54.5</td>
<td>47.9</td>
</tr>
<tr>
<td>($\lambda_c = 0.4$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green circle</td>
<td>169</td>
<td>170</td>
</tr>
<tr>
<td>($\lambda_c = 1.5$)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 6.10 The results with the shape factor and the colour factor
Figure 6.11 The result of finding red objects
Figure 6.12 The result of finding green objects
Chapter 6

focused on the ripe tomatoes that have greater intensity differences from the background than unripe tomatoes by image intensity when the colour factor is not applied. On the other hand, the error values for the red and green circles dramatically decrease after applying the colour factor. The colour factor is a more efficient way to reduce the error value for recognising coloured objects than the performance without the colour factor. In addition, these results for the shape factor also accord closely with the result in chapter 4 that the suitable shape factor to extract a circle in real images is under 1.6.

6.3 Contrast

There are many methods to extract local features for coloured objects in images. The Scale-Invariant Feature Transform (SIFT) [45] is one of the most prominent methods and many variations of SIFT have been researched for the last decade, including the study of SIFT for coloured objects. The Coloured SIFT (CSIFT) which is constructed in a colour invariant space [46] extends the conventional SIFT from grey level images, and the Multispectral SIFT descriptor [47] was suggested to improve performance for scene recognition. Also, colour invariant descriptors [48] have been proposed for image category recognition.

Whilst SIFT is a powerful technique to extract keypoints in images it requires some time to match keypoints because there are many features. This is the reason that other approaches have been developed to reduce the time needed for matching. Speeded Up Robust Features (SURF) [49] is one of the most well-known approaches. SURF also uses only grey level images like the original SIFT approach, therefore, the extension of SURF using colour has been issued in various realm of feature extraction. Colour invariant SURF (CSURF) [50] has been applied to track objects in a discriminative framework and a local kernel colour histogram could be a part of SURF [51]. These latest techniques are appropriate to compare the performance of the processes of combined shape and colour capabilities.

The way used to compare the performance in this thesis is that the extended IRT without the colour factor (using only the shape factor) and with the colour factor are compared to SURF and the colour-based SURF. Before this comparison the results of SURF and the colour-based SURF are demonstrated in advance.
The result of original SURF is shown in Figure 6.13. Several keypoints on the part of a ripe tomato are missed when the sample is detached or scaled from the original image in Figures 6.13 (b) and (c). In addition, SURF is not focussed on the colour because a colour image is changed to a grey image during the process. For this reason, the coloured feature of the CSIFT in chapter 2.2.4 is applied to SURF for another comparison and the results are shown in Figure 6.14. The colour image is changed for each measurement plane $\hat{E}$, $\hat{E}_\lambda$ and $\hat{E}_{\lambda\lambda}$ using Equation 2.33 as the first step, and then SURF is processed in each plane. However, the keypoints with CSIFT features lead to difficulty in detecting matching pairs in $\hat{E}_\lambda$ and $\hat{E}_{\lambda\lambda}$ planes in Figure 6.14 (e) and (f). Therefore, an additional plane, $\hat{E}_{\text{total}}$, is generated by overlapping three planes. The SURF results of $\hat{E}$ and $\hat{E}_{\text{total}}$ are displayed in Figure 6.14 (a)-(d). Even though the colour property of CSIFT is applied to SURF pairs of matching keypoints are missed.

The matching rates of SURF and SURF using colour invariance are shown in Table 5. In this thesis, we shall define that the matching rate is the percentage of finding both matching and mismatching keypoints pairs between the sample and the target image by SURF. The matching rate used in this thesis is considered matching of individual keypoints with untrained images and scaled images, while SIFT uses the clusters of keypoints (containing at least three keypoints) and training images for object recognition [45]. Although SIFT can better extract features than SURF, the computational time for SURF is much faster for matching keypoints than that for SIFT. Thus, the matching rate of SURF and SURF using

![Figure 6.13 Matching rate of original SURF examples](image)

(a) RGB and RGB : 100%
(b) A red tomato to RGB whole: 95.2%
(c) A red tomato (a quarter scale) to RGB whole: 50%
colour invariance in CSIFT [52] are more appropriate for comparison of results with the extended IRT.

We use $\hat{E}$ and $\hat{E}_{\text{total}}$ planes, which are more comparable than other planes ($\hat{E}_\lambda$ and $\hat{E}_{\lambda\lambda}$) in CSIFT, to compare with original SURF. Samples containing a coloured object were taken from a part of a whole image in Figure 6.15 (b) and compared to the whole images in Figure 6.15 (a). The sample images and detailed results are described in Figure 6.16 and Appendix D. While SURF results (95.1 percent in a whole part and 93 percent in a partial sample) are good to match whole or partial images CSIFT results appear less suited to matching. Even though one of planes in CSIFT can be possible to find more key features ($\hat{E}$ plane) it cannot affect the results. In addition, the result could depend on how to set the sample part in a whole image because all samples are manually selected. Furthermore,
Table 5 The matching rate of SURF and coloured SURF

<table>
<thead>
<tr>
<th></th>
<th>SURF (%)</th>
<th>SURF using colour invariant (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( E )</td>
</tr>
<tr>
<td>Whole</td>
<td>95.1</td>
<td>58.7</td>
</tr>
<tr>
<td>Object / Whole image</td>
<td>93.0</td>
<td>53</td>
</tr>
</tbody>
</table>

(a) whole images

(b) samples

Figure 6.15 Example test images and samples
the position of keypoints in SIFT and SURF could be outside the target objects. It can be profitable to categorise and search an image in the database, however, the matching between keypoints in the background can influence mistrust of the performance in extracting exact target objects.

As shown by the above results SURF appears to be unsuitable to compare the result with the extended IRT since the nature of the results of the two approaches is different and difficult to be recognisable at once. Hence, we compare performance with that of the Sobel operator which is a classical and basic method to extract edges in images. The Sobel operator is applied to grey-level images (without the colour factor) and CIE L*a*b* images under the same condition as the extended IRT with the colour factor. Table 6 shows the result that each value means the average of example image in Figure 6.15 (a) by each standard from a red and a green tomato. In the case without the colour factor, the result of the extended IRT with only the shape factor is more accurate than the result of Sobel operator on average. The Sobel operator is frequently used to detect edges in grey-scale images and it is suitable to extract a circle information by the Hough Transform as shown in Figure 6.17 (b), however, the wrong area is occasionally chosen by the Sobel operator as a circle in Figure 6.17 (e). It shows that the shape factor can suffice to extract object shapes independently. In addition, the performance is superior after applying the colour factor. The results with the colour factor are much better than the result without the factor even the case of Sobel operator. The Sobel operator can be used to detect edges in an image to which the colour factor has been applied in Figure 6.17 (c), however, there is no capability to select a specific coloured object. For example, the Sobel operator can find the red tomato instead of the green tomato when \( \lambda_c = 1.5 \) in Figure 6.17 (f). These results explain that not only individual
factor, the shape factor and the colour factor, but also that combing the two factors leads to a powerful and simple system. Furthermore, combination of the shape factor and the colour factor has greater capability to extract specific coloured object shapes than other approaches.

Table 6 Comparison of the error value by extended IRT and Sobel operator

<table>
<thead>
<tr>
<th></th>
<th>Sobel with grey channel (pixel)</th>
<th>Sobel with $\lambda_c$ (pixel)</th>
<th>The extended IRT without $\lambda_c$ (pixel)</th>
<th>The extended IRT with $\lambda_c$ (pixel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red circle ($\lambda_c = 0.4$)</td>
<td>96.4</td>
<td>14.1</td>
<td>51.8</td>
<td>5.61</td>
</tr>
<tr>
<td>Green circle ($\lambda_c = 1.5$)</td>
<td>248</td>
<td>84.8</td>
<td>169</td>
<td>17.0</td>
</tr>
</tbody>
</table>

Figure 6.17 The result example of Sobel operator

(a) original  (b) grey-level image  (c) $\lambda_c = 0.5$ (for red)

(d) original  (e) grey-level image  (f) $\lambda_c = 1.5$ (for green)
Chapter 6

6.4 Conclusions

The advantage of the extended IRT including the colour factor is that there are specific ranges of the factors according to the type of objects and colour, whereas it is difficult to set a sample for matching in SIFT. For instance, SIFT needs a sample of the object to be searched for in various categories and one consideration is which image is more suitable for searching. However, the extended IRT needs only two facts, object type and colour. It is a simple and new approach to extract features in image processing without searching database. Naturally, it needs experimental results to detect relevant ranges of two factors in each coloured object. Thus, building dataset for extended IRT can be a practical method to not only locate factor ranges but also compare the performance with other approaches.
Chapter 7: Conclusions and future work

7.1 Conclusions

We have extended the IRT by addition of the capability to detect selected shapes, to detect coloured shapes and the combination of shape and colour. The first of these contributions might first appear incidental, but it is actually a major extension of the analogy conferring new properties, especially since the extension has no known physical basis. The factor for the first extension, the shape factor, controls the amount of reflection in Snell’s law. Thus, it influences the highlighting of edges of objects with a particular shape. The results from the experiments show the consistency of the first extension by analysing histogram patterns and the shape factor has a suitable range to extract each shape of object as shown by studies on synthesised and on real images. The first extension focuses only on shape extraction using grey-level images, therefore, we have extended the IRT to detect coloured shapes using a colour factor in the CIEL*a*b* colour model. The colour factor controls the refractive materials’ matrix value to find specific coloured objects, and studies again showed detection using synthesised and real images. The capability of the colour factor is difficult to analyse for consistency with real images because there is no appropriate database for this, however, the results show acceptable performance when extracting coloured objects according to suitable range of the colour factor. The implementation using the RGB colour model can actually provide more accurate results, though it employs three parameters, whereas the implementation using the CIEL*a*b* colour model uses only a single parameter. In this way, the IRT can be extended for feature extraction to detect specific shapes with colour.

The addition of parameters has transformed a technique which was previously only for low-level extraction, into one with capability to simultaneously perform low- and high-level extraction. The analyses, using the Hough Transform, of images that include various types of objects verify how the extended IRT can perform on real images and where it can be applied. The shape factor that controls the amount of reflection of light rays is the one of important parameters that can determine the quality of performance in the extended IRT. In addition, the colour factor using colour information is verified that there is a specific range
of the factor for each colour and it can be advantageous for colour discrimination capacity in the extended IRT. In this way, we have a new technique that can be used to detect coloured objects with a specific shape.

7.2 Future work

Future work will continue to extend the IRT for combining low- and high-level feature extraction methods. We suggest that it can be applied as a pre-processing stage and can be expanded to extract features using colour information.

Another extension could be to analyse histogram patterns of the extended IRT to extract more information. For this extension, the Earth Mover’s Distance (EMD) [53] can be applied to measure the similarity between histogram patterns of the extended IRT. Figure 7.1 shows the results of applying the HT for circles and for lines to the extended IRT output for synthesised and real images. The best results for each object in the synthesised images according to chapter 4.3 are displayed in Figure 7.1 (b) and (h). Figure 7.1 (d-f) and (j-l) are the result of real images by different values for the shape factor. The resulting correlation between the histogram patterns using EMD is displayed in Table 7. The minimum distance was derived for the best result for a circle in the synthesised image (b), and one in a real image (e) in Table 7. Since the optimal shape factor interval for a circle is between 1.7 and 2.0, the EMD results (e) and (f) of Table 7 are similar. The resulting line is firmly proved by Table 7, where the minimum value of EMD for a line in the synthesised image (h) best matches the line in the real image (j).

When the shape factor increases, the edge of the golf tee still exist but the shape resembles a rectangle more than a line. This suggests why (k-l) present a distance which is much greater than expected in Table 7. The result on the real images confirm that the extended IRT can indeed intergrate shape selectivity with low level feature extraction, though the exact method for image feature extraction is yet to made genetic and it as yet specific images and shapes. If there are specific patterns of the extended IRT results by the shape factor and the colour factor, it will be more efficient to detect coloured object shapes in the image.
Figure 7.1 Shape extraction and description in real images

Table 7 EMD results of correlation of Figure 7.1

<table>
<thead>
<tr>
<th></th>
<th>(d)</th>
<th>(e)</th>
<th>(f)</th>
<th>(j)</th>
<th>(k)</th>
<th>(l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b)</td>
<td>69.3</td>
<td>11.0</td>
<td>16.7</td>
<td>73.3</td>
<td>26.7</td>
<td>17.2</td>
</tr>
<tr>
<td>(h)</td>
<td>5.08</td>
<td>53.6</td>
<td>52.0</td>
<td>0.607</td>
<td>11.4</td>
<td>10.8</td>
</tr>
</tbody>
</table>

More colour information of images could be expressed if the refractive indices were more elaborate or another vector field, such as the field using the force field transform. By making the refractive indices complex, the vector operations associated with the force field approach (chapter 2.4) could be deployed. One further extension could be applied to construct materials’ refractive indices for containing colour information. Most colour models have multi channels to express colour, and a channel of the materials’ refractive induces combining $a^*$ and $b^*$ planes of $CIE L^*a^*b^*$ colour model using the colour factor by Equation 5.4 is generated in this thesis. This expression will be able to expend in three channels of the refractive indices as following
\begin{align*}
n_{ij,\ell} &= \lambda_c L_{ij} \\
n_{ij,a} &= \lambda_c a_{ij} + (1 - \lambda_c)b_{ij} \\
n_{ij,b} &= (1 - \lambda_c)a_{ij} + \lambda_c b_{ij}
\end{align*}

when refractive materials’ value \(n_{ij,\ell}, n_{ij,a}\) and \(n_{ij,b}\) of each plane of CIE \(L^*a^*b^*\) model, which size is \(i \times j\) (width \(\times\) height), is determined by a weighted combination of the colour spaces using the colour factor. Thus, there are three different materials’ refractive indices and it need another equation of correlation between them. Other colour models could be also applied to generate the indices and more various combinations would be in this stage.

Another extended vector field will be developed using the force field transform. When the total force is calculating by Equation 2.14 the pixel intensity \(P(x_i)\) would be allowed to affect by colour elements from a colour space. One way to generate a force field is that the total force is calculated in each colour channel, similar to above suggestion using colour factor, or a pixel intensity at each pixel is influenced by combining all colour channels of a colour model like the approach in this thesis. It is sure that both ways are possible to represent colour information in the field and need sufficient study about them.

In addition, a database presenting multiple instances of higher order shapes is needed to be constructed since no established database is currently available that offers multiple examples of the same shape class at sufficient resolution. As further research progresses a more satisfactory database will be assembled to verify the performance for shape and colour capabilities concurrently. One of the problems with the current extensions is that as yet there is no unique value of the colour value for a particular shape which could be used to detect the same shape in every (real) image. This might appear optimistic, but an extension to achieve this would extend considerably the performance capabilities of the extended IRT.
Appendix A

A.1 Real images in database Caltech-256 Object Category Database

![Real images from Caltech-256 database]
Appendix A

A.2 Real images

tomato1
tomato2
tomato3
Appendix B

B.1 The result of synthesised images: Rectangle

\[
\begin{align*}
F_s &= 0.25 \\
F_s &= 0.50 \\
F_s &= 0.75 \\
F_s &= 1.00 \\
F_s &= 1.25 \\
F_s &= 1.50 \\
F_s &= 1.75 \\
F_s &= 2.00 \\
F_s &= 2.25 \\
F_s &= 2.50 \\
F_s &= 2.75 \\
F_s &= 3.00 \\
F_s &= 3.25 \\
F_s &= 3.50 \\
F_s &= 3.75 \\
F_s &= 4.00 \\
F_s &= 4.25 \\
F_s &= 4.50 \\
F_s &= 4.75 \\
F_s &= 5.00
\end{align*}
\]
Appendix B

B.2 The result of synthesised images: Curve

![Images showing curve results for different values of Fs]

- $F_s = 0.25$
- $F_s = 0.50$
- $F_s = 0.75$
- $F_s = 1.00$
- $F_s = 1.25$
- $F_s = 1.50$
- $F_s = 1.75$
- $F_s = 2.00$
- $F_s = 2.25$
- $F_s = 2.50$
- $F_s = 2.75$
- $F_s = 3.00$
- $F_s = 3.25$
- $F_s = 3.50$
- $F_s = 3.75$
- $F_s = 4.00$
- $F_s = 4.25$
- $F_s = 4.50$
- $F_s = 4.75$
- $F_s = 5.00$
B.3 The result of synthesised images: Mixed a circle, a rectangle and a curve

\[ F_s = 0.25 \quad F_s = 0.50 \quad F_s = 0.75 \quad F_s = 1.00 \]
\[ F_s = 1.25 \quad F_s = 1.50 \quad F_s = 1.75 \quad F_s = 2.00 \]
\[ F_s = 2.25 \quad F_s = 2.50 \quad F_s = 2.75 \quad F_s = 3.00 \]
\[ F_s = 3.25 \quad F_s = 3.50 \quad F_s = 3.75 \quad F_s = 4.00 \]
\[ F_s = 4.25 \quad F_s = 4.50 \quad F_s = 4.75 \quad F_s = 5.00 \]
Appendix C

Figure 4.12 (a) and (b)

Figure 4.12 (c) and (d)
Appendix C

Figure 4.13 (a-c)

![Histogram](image)

Figure 4.13 (d-f)

![Histogram](image)
Figure 4.13 (g-i)

Figure 4.14 (a-c)
Appendix C

Figure 4.14 (d-f)

Figure 4.14 (g-i)
Appendix C

Figure 4.15 (e-h)

![Histogram and Normalised Value](image1)

Figure 4.16 (e-h)

![Histogram and Normalised Value](image2)
Appendix D

D.1 Speeded Up Robust Features (SURF)

Example 02

RGB and RGB : 100%  
RGB and Grey : 93.8%

A red tomato to RGB whole : 90%  
A red tomato (a quarter scale) to RGB whole : 14.3%  
A red tomato to grey scale whole : 82.5%

Matching rate of original SURF examples
Appendix D

Example 03

RGB and RGB : 100%  RGB and Grey : 95.5%

A red tomato to RGB whole: 93.7%  A red tomato (a quarter scale) to RGB whole: 36.4%  A red tomato to grey scale whole : 84.8%

Matching rate of original SURF examples
D.2 Colour invariant SURF

Example 02

Matching rate of colour SURF examples
Appendix D

Example 03

RGB and $E : 64.4\%$

RGB and $E_{total} : 44\%$

A red tomato to $E : 54.5\%$

A red tomato to $E_{total} : 20.3\%$

Matching rate of colour SURF examples
References


12. G. Demarcq, L. Mascarilla, M. Berthier and P. Courtellemont, “The color monogenic signal: application to color edge detection and color optical


