A Hybrid Colour Image Segmentation Scheme

by

Nicolaos Ikonomakis

A thesis submitted in conformity with the requirements for the degree of

Master of Applied Science

Graduate Department
Electrical and Computer Engineering
University of Toronto

©Copyright by Nicolaos Ikonomakis, 1999
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-45988-8
A Hybrid Colour Image Segmentation Scheme
Master of Applied Science, 1999
Nicolaos Ikonomakis
Department of Electrical and Computer Engineering
University of Toronto

Abstract

A nonautonomous hybrid-based segmentation scheme has been developed that efficiently and robustly partitions a colour image into different regions that are homogeneous with respect to colour. As with human visual perception, image segmentation is an important aspect of any type of image or scene analysis. The scheme is based on the HSI colour space. It combines both region- and pixel-based techniques. A pixel classification algorithm is used to categorise the pixels in the image as either chromatic or achromatic. A seed determination algorithm is then used to find seed pixels in the image. The region growing algorithm starts with the set of seed pixels and from these grows regions by appending to each seed pixel those neighbouring pixels that satisfy a homogeneity criterion. After regions are grown they are further processed with a region merging algorithm. Regions that are similar in colour are merged.
Acknowledgments

I would like to express my thanks and gratitude to everyone who has helped me make this thesis possible. I would like to thank Professor Anastasios N. Venetsanopoulos and Professor Kostas N. Plataniotis for their help and supervision throughout my research. I would also like to thank my defence committee for their input and help in my thesis.

I would like to thank Vas, Pete, and Jim. I don’t think that I would have had any fun the past two years without them. It’s been a blast.

My sincerest thanks also go out to my family and friends who have provided me with constant support. I would not have been able to finish my work without their help and love.
# Contents

Abstract

Acknowledgments

List of Figures

List of Tables

## 1 Introduction

1.1 The Role of Segmentation in Digital Image Processing ....................................... 1
1.2 The Image Segmentation Problem ........................................................................... 2
1.3 Why Colour? ........................................................................................................... 3
1.4 Applications .......................................................................................................... 3
1.5 Contributions ....................................................................................................... 4

## 2 Colour Spaces

2.1 RGB Colour Space ................................................................................................. 6
2.2 XYZ Colour Space ................................................................................................. 8
2.3 Opponent Colour Spaces ........................................................................................ 8
2.3.1 YUV and YIQ Television Colour Spaces ............................................................ 9
2.4 Ohta's I1I2I3 Colour Space ...................................................................................... 10
2.5 HSI Family of Perceptual Colour Spaces ............................................................... 10
2.6 Perceptually Uniform Colour Spaces .................................................................... 11
2.6.1 CIE L*u*v* Colour Space .................................................................................. 13
2.6.2 CIE L*a*b* Colour Space .................................................................................. 14
2.6.3 Cylindrical L*u*v* and L*a*b* Colour Spaces .................................................. 14
2.7 Munsell Colour Space ............................................................................................. 15
2.8 Summary ............................................................................................................... 17
3 Colour Image Segmentation Techniques

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Pixel-based Techniques</td>
<td>18</td>
</tr>
<tr>
<td>3.1.1 Histogram Thresholding</td>
<td>19</td>
</tr>
<tr>
<td>3.1.2 Clustering</td>
<td>21</td>
</tr>
<tr>
<td>3.2 Edge-based Techniques</td>
<td>28</td>
</tr>
<tr>
<td>3.2.1 Techniques Extended From Monochrome Edge Detection Techniques</td>
<td>30</td>
</tr>
<tr>
<td>3.2.2 Vector Space Approaches</td>
<td>32</td>
</tr>
<tr>
<td>3.3 Region-based Techniques</td>
<td>35</td>
</tr>
<tr>
<td>3.3.1 Region Growing</td>
<td>36</td>
</tr>
<tr>
<td>3.3.2 Split and Merge</td>
<td>38</td>
</tr>
<tr>
<td>3.4 Model-based Techniques</td>
<td>41</td>
</tr>
<tr>
<td>3.4.1 The MAP Method</td>
<td>42</td>
</tr>
<tr>
<td>3.4.2 The Adaptive MAP Method</td>
<td>44</td>
</tr>
<tr>
<td>3.5 Physics-based Techniques</td>
<td>45</td>
</tr>
<tr>
<td>3.6 Hybrid Techniques</td>
<td>46</td>
</tr>
<tr>
<td>3.7 Summary</td>
<td>50</td>
</tr>
</tbody>
</table>

4 Proposed Colour Image Segmentation Scheme

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Region Growing</td>
<td>55</td>
</tr>
<tr>
<td>4.2 Achromatic/Chromatic Pixel Classification</td>
<td>56</td>
</tr>
</tbody>
</table>

5 Seed Determination

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Local Variance Seed Determination</td>
<td>61</td>
</tr>
<tr>
<td>5.2 Histogram Seed Determination</td>
<td>65</td>
</tr>
<tr>
<td>5.3 Summary</td>
<td>68</td>
</tr>
</tbody>
</table>

6 Colour Distance Measures

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Achromatic Distance Measure</td>
<td>71</td>
</tr>
<tr>
<td>6.2 Chromatic Distance Measures</td>
<td>72</td>
</tr>
<tr>
<td>6.2.1 Minkowski Distance Metric</td>
<td>73</td>
</tr>
<tr>
<td>6.2.2 Canberra Distance Metric</td>
<td>74</td>
</tr>
<tr>
<td>6.2.3 Cylindrical Distance Metric</td>
<td>77</td>
</tr>
<tr>
<td>6.2.4 Summary</td>
<td>78</td>
</tr>
<tr>
<td>6.3 Other Colour Spaces and Distance Measures</td>
<td>81</td>
</tr>
<tr>
<td>6.3.1 RGB Colour Space</td>
<td>81</td>
</tr>
<tr>
<td>6.3.2 L<em>a</em>b* Colour Space</td>
<td>81</td>
</tr>
<tr>
<td>6.3.3 The Cylindrical L<em>a</em>b* Colour Space</td>
<td>83</td>
</tr>
</tbody>
</table>
6.4 Summary ........................................................................ 84

7 Region Merging ................................................................ 86
  7.1 Hue Difference Metric .................................................. 87
  7.2 Cylindrical Distance Metric .......................................... 89
  7.3 Summary .................................................................... 89

8 Results ............................................................................. 91

9 Conclusions .................................................................... 99

References .......................................................................... 101
List of Figures

1.1 An example of an image analysis system. Parts inspection on an assembly line. 2

2.1 Representation of colour in the RGB colour space. 7

2.2 The colour models of the HSI family of colour spaces. 12

2.3 Representation of colour in the Munsell colour space. 16

3.1 Results of the Shafarenko et al. segmentation algorithm. 24

3.2 Fuzzy c-means segmentation results of Lim and Lee algorithm. 29

3.3 Edge detection by derivative operators. 30

3.4 Split and merge algorithm. 39

3.5 Results of the Zugaj and Lattuati segmentation algorithm. 48

3.6 Flowchart of Saber’s et al. segmentation algorithm. 49

3.7 The segmentation results of the Luo et al. algorithm. 50

4.1 Proposed colour image segmentation scheme. 54

4.2 The HSI colour model. 54

4.3 The region growing algorithm. 55

4.4 The HSI cone with achromatic region in yellow. 57

4.5 Results of achromatic/chromatic pixel classification algorithm. 58

4.6 Results of achromatic/chromatic pixel classification algorithm. 58

4.7 Results of achromatic/chromatic pixel classification algorithm. 58

4.8 Results of achromatic/chromatic pixel classification algorithm. 59

5.1 Artificial image with level 1, 2, and 3 seeds indicated. 62

5.2 Results of local variance seed determination algorithm. 64

5.3 Results of local variance seed determination algorithm. 65

5.4 Results of peak detection algorithm. 67

5.5 Results of peak detection algorithm. 68

5.6 Results of histogram seed determination algorithm. 69
6.1 Results with the achromatic distance measure. .......................... 71
6.2 Results with the achromatic distance measure. .......................... 72
6.3 Manual seeds chosen for three images. ..................................... 72
6.4 Results with the Minkowski distance metric. .............................. 75
6.5 Results with the Minkowski distance metric. .............................. 75
6.6 Results with the Canberra distance metric. .............................. 76
6.7 Results with the Canberra distance metric. .............................. 76
6.8 Results with the Canberra distance metric. .............................. 76
6.9 The chromatic plane of the HSI colour model. ............................ 77
6.10 Results with the Cylindrical distance metric. ........................... 78
6.11 Results with the Cylindrical distance metric. ........................... 79
6.12 Results with the Cylindrical distance metric. ........................... 79
6.13 Results with the Euclidean distance metric in the RGB colour space. ................................. 82
6.14 Results with the Canberra distance metric in the RGB colour space. ................................. 82
6.15 Results with the Euclidean distance metric in the L*ab* colour space. ................................. 83
6.16 Results with the Godlove distance metric in the HCV colour space. ................................. 84

7.1 Results of the merging algorithm with the Hue difference metric. ................................. 88
7.2 Results of the merging algorithm with the Hue difference metric. ................................. 88
7.3 Results of the merging algorithm with the Cylindrical distance metric. ................................. 90
7.4 Results of the merging algorithm with the Cylindrical distance metric. ................................. 90

8.1 Colour image segmentation scheme. ......................................... 92
8.2 Segmentation results with humans in image. ................................ 93
8.3 Segmentation results with humans in image. ................................ 94
8.4 Segmentation results with humans in image. ................................ 95
8.5 Segmentation results with images of flowers. ................................ 96
8.6 Segmentation results with images of buildings. ................................ 97
8.7 Segmentation results with images of landscapes. ................................ 98
List of Tables

2.1 Colour Spaces .............................................. 17
3.1 Colour Image Segmentation Techniques ....................... 52
6.1 Comparison of Chromatic Distance Measures ................... 80
Chapter 1

Introduction

IMAGE segmentation refers to partitioning an image into different regions that are homogeneous with respect to one or several image features. The process of segmenting an image is easy to define but difficult to develop. The research presented in this thesis is on the design of an autonomous hybrid colour image segmentation scheme. More specifically an unsupervised, efficient, robust, and relatively computationally inexpensive colour image segmentation scheme is presented.

1.1 The Role of Segmentation in Digital Image Processing

Digital colour images occur very frequently in the world today. All images on the Internet are in digital form; most images seen in magazines and newspapers have been in digital form at some point before publication; and many films have been converted to a digital format for remastering. Digital images are processed simply to improve the quality of the image, or they may be processed to extract useful information, such as the position, size or orientation of an object.

Image analysis is an area of image processing that deals with techniques for extracting information from an image. In the simplest form, this task could be reading an address on a letter or finding defective parts on an assembly line (Figure 1.1). More complex image analysis systems measure quantitative information and use it to make a sophisticated decision, such as trying to find images with a specified object in an image database. The various tasks involved in image analysis can be broken down into conventional (low-level) techniques and knowledge-based (high-level) techniques. Image segmentation falls into the low-level category and is usually the first task of any image analysis process. All subsequent tasks, such as feature extraction and object recognition rely heavily on the quality of the segmentation. For example, in Figure 1.1, if the segmentation algorithm did not partition the image correctly the recognition and interpretation algorithm would not recognize the object...
1.2 THE IMAGE SEGMENTATION PROBLEM

Figure 1.1: An example of an image analysis system. Parts inspection on an assembly line.

and a good stirring wheel may be classified as defective. Over-segmenting an image will divide an object into different regions. Under-segmenting the image will group various objects into one region. The segmentation step determines the eventual success or failure of the image analysis process. For this reason, considerable care is taken to improve the probability of a successful segmentation.

1.2 The Image Segmentation Problem

Image segmentation is an important aspect of the human visual perception. Humans use their visual sense to effortlessly partition their surrounding environment into different objects to help recognize the objects, guide their movements, and for almost every other task in their lives. It is a complex process that includes many interacting components that are involved with the analysis of colour, shape, motion, and texture of objects. For the human visual system, the segmentation of images is a spontaneous, natural activity. For example, if a scene contained a person wearing a blue shirt with a clear blue sky in the background a human would be able to segment the blue shirt from the blue sky. Unfortunately, it is not easy to create artificial algorithms whose performance is comparable to that of the human visual system. An artificial image analysis system may classify the sky and the shirt as one region. As Marr has suggested [1], one of the major obstacles to the successful development of theories of segmentation has been a tendency to underestimate the complexity of the problem exactly because the human performance is mediated by methods which are largely subconscious. Because of this, segmentation of images is weakened by various types of uncertainty making most
simple segmentation techniques ineffective.

1.3 Why Colour?

Most attention on image segmentation has been focused on grey-scale or monochrome images. A common problem in the segmentation of grey-scale images occurs: when an image has a background of varying grey level, such as gradually changing shades; when regions assume some broad range of grey levels; or with the presence of intensity changes due to shadows and surface curvatures in the grey-scale image. This problem is inherent since intensity is the only available information from monochrome images. It is known that the human eye can detect only in the neighbourhood of one or two dozen intensity levels at any point in a complex image, but can differentiate thousands of colour shades and intensities [2]. For example, some colours, when converted to grey-scale, produce the same grey levels; whereas, seen in colour, they will be easily separable by the human visual system and by a digital image processor.

1.4 Applications

From the applications point of view, digital photography, electronic imaging, digital television, image libraries, and multimedia have given image segmentation a 'push' so that the field has become a principal area of research. Image segmentation is not only in electrical engineering, but also in other academic disciplines, such as: computer science, geography, medical imaging, criminal justice, and remote sensing. Image segmentation has taken a central place in numerous applications, including, but not limited to:

- video-conferencing: increase transmission speed by compression
- multimedia video databases, video-on-demand: track objects in video which allows for better scene detection and scene indexing
- colour image and video transmission over the Internet, digital broadcasting, interactive TV: use of segmentation to track objects for modeling in MPEG-4 and MPEG-7 video standards
- computer-based training
- distance education
- tele-medicine

Many reasons can be cited for the success of the field. There is a strong underlying analytical framework based on mathematics, statistics, and physics. Thus, well-founded, robust algorithms
that eventually lead to consumer applications can be designed. The field has also been helped tremendously by the advances in computer and memory technology, enabling faster processing of images.

A hybrid-based colour image segmentation scheme was developed that gives efficient, robust and relatively computationally inexpensive results. By hybrid-based, the scheme uses both region- and pixel-based techniques. At the heart of the scheme is a region growing algorithm which is a region-based technique. Pixel-based techniques are used for pixel classification and for determining seeds for the region growing algorithm. The proposed colour image segmentation scheme is based on the HSI colour space. It can be split into four general algorithms:

1. Classify the pixels in the image as either chromatic or achromatic by examining their colour values.
2. Determine the seed pixels in the image.
3. Employ the region growing algorithm on the chromatic and achromatic pixels separately starting from the seeds.
4. Merge regions that are similar in colour.

The first algorithm uses pixel-based techniques to classify pixels as either having chromatic or achromatic information. The seed determination algorithms (2), use pixel-based techniques to find starting points for the region-based region growing algorithm (3).

1.5 Contributions

Many algorithms and ideas have been developed during the process of developing a colour image segmentation scheme. These include:

- A method of classifying pixels and regions as achromatic or chromatic for further analysis. This can be used in all types of colour image processing algorithms and not only image segmentation.
- Two methods of seed determination which find pixels or regions that are in the centre of regions. These techniques can be used, not only in region growing segmentation algorithms, but in any segmentation technique where starting points are needed (i.e. clustering algorithms).
- A full analysis of colour similarity measures in the HSI and other popular colour spaces.
- A merging algorithm that merges regions that are similar in colour.

The thesis is structured into seven parts. The first deals with colour science principles. It gives an outline of the most common colour spaces used in colour image segmentation research. Chapter 3 outlines the general techniques developed for colour image segmentation by researchers. Chapter 4 outlines the proposed region- and pixel-based segmentation scheme with emphasis on the first step of the scheme. Chapter 5 deals with the seed determination step. Colour similarity measures are dealt with in Chapter 6. Chapter 7 explains the region merging step. Results of the scheme are shown and discussed in Chapter 8.
Chapter 2

Colour Spaces

Segmented grey-scale images are affected by various situations to give poor results. They might be negatively affected by the presence of intensity changes due to shadows and surface curvature in the grey-scale images. If a shadow falls on a homogeneous surface then a segmentation algorithm based on grey-scale information will most likely partition the surface into two parts; the homogeneous surface might be considered as two surfaces with different grey-scale values. Also, an algorithm based on intensity information might not be able to detect the difference between two adjacent different coloured regions, reflecting the same amount of light (i.e. having the same intensity but different chromaticity).

Because of the advantages of considering colour, segmentation algorithms based on colour information are becoming very common. Currently, many colour spaces are being used for segmentation. Each of them describe colour by means of three colour features. For pictures acquired by digital cameras the most popular is the RGB model. It is represented by red, green, and blue images. The RGB colour space is based on the additive colour mixture principle which means that a range of colours is produced by the mixture of the three primary colours with corresponding intensities. Almost all other colour spaces used for segmentation are transformed from the RGB space.

Several colour spaces that are popular amongst colour image segmentation researches are described in the sections that follow. The proposed segmentation scheme uses the HSI colour space (Section 2.5).

2.1 RGB Colour Space

The RGB colour space is the most frequently used colour space for image processing, because most cameras, scanners and displays are provided with direct RGB signal input or output. The colour
2.1. RGB COLOUR SPACE

Figure 2.1: Representation of colour in the RGB colour space.

The model is based on a Cartesian coordinate system and can be represented by the cube in Figure 2.1. A colour value is represented by a point in or on the cube surface. All grey-scale values extend along the main diagonal of the cube from black (at the origin) to white. A disadvantage of the RGB colour space is that there is a high correlation between its components \((R - B, R - G, G - B)\). A 10% change in \(R\) may change the \(G\) or \(B\) values. In terms of segmentation, the RGB colour space is usually not preferred because it is psychologically non-intuitive and non-uniform. That is, it is hard to visualize a colour based on \(R\), \(G\), and \(B\) components and it is impossible to evaluate the perceived differences between colours on the basis of distance in the RGB colour space.

By dividing the \(R\), \(G\), and \(B\) coordinates by their total sum, the \(r\), \(g\), and \(b\) quantities are obtained, yielding the rgb colour space.

\[
\begin{align*}
    r &= \frac{R}{R + G + B}, \\
    g &= \frac{G}{R + G + B}, \\
    b &= \frac{B}{R + G + B} = 1 - r - g.
\end{align*}
\]  

This colour space is normalized with respect to intensity and therefore the influence of illumination intensity is eliminated. Since each of the normalized colours is linearly dependent, the rgb colour space may be represented by two normalized colours. Values of rgb colour coordinates are much more stable with changes in illumination than RGB colour coordinates [3].

Nicolao Ikonomakis
2.2 XYZ Colour Space

There exists no unique set of three primary colours which can describe all perceived colours, and thus, the RGB colour space is not perfect in describing colours. Because of this downfall, an international colour standard was established in 1931 by the Commission International de l'Eclairage (C.I.E.). The chosen primary colours are imaginary colour. They're too saturated to be produced or seen by the human eye. The advantage of the colour space, is that the primary colours can describe each perceived colour mathematically. These standard primary colours are denoted by X, Y, and Z, giving the XYZ colour space. They can be produced from the RGB tristimulus coordinates by a linear transformation that is determined empirically and shown in the following equations [4]:

\[
X = 0.490R + 0.310G + 0.200B , \quad (2.4) \\
Y = 0.177R + 0.812G + 0.011B , \quad (2.5) \\
Z = 0.000R + 0.010G + 0.990B . \quad (2.6)
\]

As is done with the RGB colour space, the XYZ colour space can be normalized to yield the xyz colour space.

The CIE XYZ colour space is highly non-uniform, and therefore, it is not appropriate for quantitative manipulations involving colour perception and is seldom used in colour image processing applications [5, 6]. However, as will be shown, it is used indirectly to construct various other colour spaces used today.

2.3 Opponent Colour Spaces

Opponent colour spaces have been inspired by the physiology of the human visual system [7]. In the late 19th century the German physiologist Ewald Hering proposed the colour opponency theory which explained some perceptual colour phenomena that could not be explained by classical trichromatic theory [8]. During his colour naming experiments, Hering observed that the reddish-green and yellowish-blue colours were not identified. This led Hering to believe that red must be the opposite colour to green and blue must be opposite to yellow. He proposed the existence of three opponent channels in the human visual system: red-green (RG), yellow-blue (YB), and achromatic white-black or intensity (WhBl or I). Hering's opponent theory was one of the first approaches to separate luminance from chrominance. Equations (2.7-2.9) shows how RGB colour coordinates are transformed to the three channels [7].

\[
RG = R - G , \quad (2.7)
\]
2.3.1 YUV and YIQ Television Colour Spaces

The YUV and YIQ colour spaces are opponent colour spaces because they define a luminance and two chrominance components based on colour difference signals $R-Y$ and $B-Y$. They were designed to minimize the bandwidth of signals to enable transmission to colour-difference signals within the bandwidth of black and white television [2, 7]. The human visual system is far less sensitive to spatial details in chrominance than in luminance and thus the bandwidth of chrominance signals are smaller than that of luminance signals [2, 7].

**YUV Colour Space**

In Europe the YUV colour space is used in the coding of television signals. The transformation from the RGB signal is calculated as [7]:

\[
Y = 0.299R + 0.587G + 0.114B ,
\]

\[
U = -0.147R - 0.289G + 0.437B = 0.493(B - Y) ,
\]

\[
V = 0.615R - 0.515G - 0.100B = 0.877(R - Y) .
\]

The $Y$ (luminance) component is identical to the $Y$ component in the XYZ colour space.

**YIQ Colour Space**

The YIQ colour space is the basis used in the coding of television signals in North America. The transformation from the RGB signal is calculated as [2, 7]:

\[
I = 0.596R - 0.274G - 0.322B = 0.74(R - Y) - 0.27(B - Y) ,
\]

\[
Q = 0.211R - 0.523G + 0.312B = 0.48(R - Y) + 0.41(B - Y) ,
\]

with the $Y$ component identical to the $Y$ component in the YUV colour space.
2.4 Ohta’s I1I2I3 Colour Space

Ohta et al. defined, through experimentation, a colour space that is obtained with the following transformations [9]:

\[
\begin{align*}
I_1 & = \frac{(R+G+B)}{3}, \\
I_2 & = (R-B), \\
I_3 & = \frac{(2G-R-B)}{2}.
\end{align*}
\]

(2.15)  
(2.16)  
(2.17)

The I1 component represents the intensity information, while the I2 and I3 components represent chromatic informations. Since a feature is said to have large discriminant power if its variance is large, they tried to derive colour features with large discriminant power by the Karhunen-Loeve (KL) transformation. At every step of their segmentation algorithm, calculation of the new colour features was done by the KL transform of RGB data. There experimentation is described in more detail in section 3.1.1.

2.5 HSI Family of Perceptual Colour Spaces

Although colour receptors in the human eye (cones) absorb light with the greatest sensitivity in the blue, green, and red part of the colour spectrum, the signals from the cones are further processed in the visual system [10]. Because of this perception process, a human can easily recognise basic attributes of colour: intensity (lightness or brightness), hue, and saturation. The hue component represents the impression related to the dominant wavelength of the colour stimulus. The saturation component corresponds to relative colour purity. Colours with no saturation are grey-scale colours. Intensity is the amount of light in a colour. Maximum intensity is sensed as pure white, while minimum is sensed as pure black. The HSI (or IHS) colour space is calculated from formulas approximately expressing the psychophysical sense of these notions. The transformation formulas are given in the literature in different forms because of a trade in accuracy and computational simplicity. The original formula, derived by Tenenbaum et al. [11] and accepted as basic, is given as:

\[
\begin{align*}
H & = \arccos \left( \frac{1/2 ((R-G) + (R-B))}{\sqrt{(R-G)(R-G) + (R-B)(G-B)}} \right), \\
& \text{if } B > G \text{ then } H = 360^\circ - H \\
S & = 1 - \frac{3 \min(R,G,B)}{R+G+B}.
\end{align*}
\]

(2.18)  
(2.19)
2.6. PERCEPTUALLY UNIFORM COLOUR SPACES

\[ I = \frac{R + G + B}{3}. \] (2.20)

Figure 2.2(a) shows the HSI colour model.

The two other colour spaces in the HSI family are the HSV (hue, saturation, value) and HLS (hue, lightness, saturation) colour spaces. They differ from the HSI space in the formulas for the values of intensity and saturation. Their respective colour models are shown in Figure 2.2 (b) and (c). The transformation formulas for the HSV colour space are as follows:

\[ V = \max(r, g, b), \] (2.21)
\[ S = \begin{cases} \frac{\max(r, g, b) - \min(r, g, b)}{\max(r, g, b)} & \text{if } \max(r, g, b) \neq 0, \\ 0 & \text{otherwise} \end{cases}. \] (2.22)

The transformation formulas for the HLS colour space are as follows:

\[ L = \frac{\max(r, g, b) + \min(r, g, b)}{2}, \] (2.23)
\[ S = \begin{cases} \frac{\max(r, g, b) - \min(r, g, b)}{\max(r, g, b) + \min(r, g, b)} & \text{if } L \leq 0.5, \\ \frac{2 - \max(r, g, b) - \min(r, g, b)}{2 - \max(r, g, b) - \min(r, g, b)} & \text{otherwise} \end{cases}. \] (2.24)

These colour models demonstrate that colours become less saturated when the intensity (value or lightness) approaches minimal or maximal levels. That is, unlike the HSI colour model, the saturation values in these models are not proportional to the distance from P to the center of the intensity line.

The important advantages of the HSI family of colour spaces over the other colour spaces are:

- good compatibility with human intuition
- separability of chromatic values for achromatic values
- the possibility of using only one colour feature (hue) for segmentation

Many image segmentation researchers take advantage of the third point. They perform their segmentation scheme on one colour feature (hue) instead of all three, allowing the use of much faster algorithms.

2.6 Perceptually Uniform Colour Spaces

In image segmentation, it is of particular interest to have a perceptually uniform colour space where a small perturbation in a component value is approximately perceptible across the range of that
Figure 2.2: (a) The HSI colour model (b) The HSV colour model (c) The HLS colour model.
value. The colour specification systems discussed in the previous sections are far from uniform. Although the HSI colour spaces are perceptual, they are based on simplifying assumptions and do not constitute perceptually uniform colour spaces. The CIE recommended two colour spaces in 1976, that are perceptually uniform. These are the CIE L*a*b* and the CIE L*u*v* colour spaces. They are slightly different because of the different approaches to their formulation [4, 5, 12, 13]. Nevertheless, both spaces are equally good in perceptual uniformity and provide very good estimates of colour difference between two colour vectors.

To obtain the colour values, the linear RGB components are first transformed to CIE XYZ components using the appropriate matrix and then these components are transformed to the L*a*b* and L*u*v* colour components. Both systems are based on the perceived lightness $L^*$. According to the CIE 1976 standard, the perceived lightness of a standard observer is assumed to follow the physical luminance (a quantity proportional to intensity) according to a cubic root law. Therefore, the lightness $L^*$ is defined by the CIE as:

$$L^* = \begin{cases} 
116\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16 & \text{if } \frac{Y}{Y_n} > 0.008856 \\
903.3\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} & \text{if } \frac{Y}{Y_n} \leq 0.008856
\end{cases} \quad \text{(2.25)}$$

where $Y_n$ is the physical luminance of the white reference point. The CIE definition of $L^*$ applies a linear segment with a slope of 903.3 near black for $(Y/Y_n) \leq 0.008856$. This linear segment is unimportant for practical purposes [5]. The range of values for $L^*$ is from 0 to 100. The white reference point $[X_n, Y_n, Z_n]$ is usually the linear $RGB = [1, 1, 1]$ values converted to the XYZ values using the following transformation:

$$\begin{bmatrix} X_n \\ Y_n \\ Z_n \end{bmatrix} = \begin{bmatrix} 0.4125 & 0.3576 & 0.1804 \\ 0.2127 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9502 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{(2.26)}$$

The $u^*$ and $v^*$ components in L*u*v* space and the the $a^*$ and $b^*$ components in L*a*b* space are representative of chrominance. In addition, both are device independent colour spaces. Both these colour spaces are, however, computationally intensive to transform to and from the linear as well as non-linear RGB spaces. This is a disadvantage if real-time processing is required or if computational resources are at a premium.

### 2.6.1 CIE L*u*v* Colour Space

The first uniform colour space, standardized by CIE, is the L*u*v*. The lightness component $L^*$ was given in Equation 2.26. Computation of $u^*$ and $v^*$ involves intermediate $u'$, $v'$, $u'_n$, and $v'_n$.
quantities defined as:

\[
\begin{align*}
    u' &= \frac{4X}{X + 15Y + 3Z}, \\
    v' &= \frac{9Y}{X + 15Y + 3Z}, \\
    u'' &= \frac{4X_n}{X_n + 15Y_n + 3Z_n}, \\
    v'' &= \frac{9Y_n}{X_n + 15Y_n + 3Z_n}. 
\end{align*}
\]

(2.27) (2.28)

Finally, \( u^* \) and \( v^* \) are computed as:

\[
\begin{align*}
    u^* &= 13L^*(u' - u'_n), \\
    v^* &= 13L^*(v' - v'_n). 
\end{align*}
\]

(2.29) (2.30)

2.6.2 CIE L*a*b* Colour Space

The L*a*b* colour space is the second uniform colour space, standardized by CIE. The lightness \( L^* \) component is the same as in the L*u*v* space defined in Equation 2.26. The \( a^* \) and \( b^* \) components are given by:

\[
\begin{align*}
    a^* &= 500 \left[ \left( \frac{X}{X_n} \right)^{\frac{1}{3}} - \left( \frac{Y}{Y_n} \right)^{\frac{1}{3}} \right], \\
    b^* &= 200 \left[ \left( \frac{Y}{Y_n} \right)^{\frac{1}{3}} - \left( \frac{Z}{Z_n} \right)^{\frac{1}{3}} \right], 
\end{align*}
\]

(2.31) (2.32)

with the constraint that \( \frac{X}{X_n}, \frac{Y}{Y_n}, \frac{Z}{Z_n} > 0.01 \). This constraint will be satisfied for most practical purposes [5]. Hence, the modified formulas described in [4] for cases that do not not satisfy this constraint can be ignored in practice [5, 6].

2.6.3 Cylindrical L*u*v* and L*a*b* Colour Spaces

A colour represented in the rectangular colour spaces L*u*v* and L*a*b*, can also be expressed in terms of cylindrical coordinates with the perceived lightness \( L^* \) and the psychometric correlates of chroma and hue [7].

The chroma of the cylindrical coordinated systems are defined as [4]:

\[
\begin{align*}
    C'_{uv} &= \left( (u^*)^2 + (v^*)^2 \right)^{\frac{1}{2}}, \\
    C'_{ab} &= \left( (a^*)^2 + (b^*)^2 \right)^{\frac{1}{2}}, 
\end{align*}
\]

(2.33) (2.34)

with \( C_{uv}^* \) and \( C_{ab}^* \) representing the chroma of the cylindrical L*u*v* and L*a*b* spaces, respectively.
The hue angle in the $L^*u^*v^*$ space is denoted as $H^*_{uv}$ and in the $L^*a^*b^*$ space as $H^*_{ab}$. They are defined as [4]:

\begin{align}
H^*_{uv} &= \arctan \left( \frac{v^*}{u^*} \right), \\
H^*_{ab} &= \arctan \left( \frac{b^*}{a^*} \right).
\end{align}

The cylindrical $L^*u^*v^*$ and $L^*a^*b^*$ colour spaces are usually referred to as $L^*H^*C^*$ or HCV in literature.

## 2.7 Munsell Colour Space

The Munsell colour space represents the earliest attempt to organize colour perception into a colour space [4]. The Munsell space is defined as a comparative reference for artists. Its general shape is that of a cylindrical representation with three dimensions roughly corresponding to the perceived lightness, hue and saturation. However, contrary to the HSI family of colour models where the colour solid are parameterized by hue, saturation, and perceived lightness, the Munsell space uses the method of the \textit{colour atlas}, where the perception attributes are used for sampling. The fundamental principle behind the Munsell colour space is that of equality of visual spacing between each of the three attributes. Hue is scaled according to some uniquely identifiable colour. It is represented by a circular band divided into ten sections. The sections are defined as: red, yellow-red, yellow, green-yellow, green, blue-green, blue, purple-blue, purple, and red-purple. Each section can be furthered divided into ten subsections if finer divisions of hue are necessary. A chromatic hue is described according to its resemblance to one or two adjacent hues. A value in the Munsell colour space refers to a colour's lightness or darkness and is divided into eleven sections numbered zero to ten. Value zero represents black while a value of ten represents white. The chroma defines the colour's strength. It is measured in numbered steps starting at one for weak colours having low chroma values. The maximum possible chroma depends on the hue and the value being used.

Although the Munsell book of colours can be used to define or name colours, in practice it is not used directly for image processing applications. Visually stored image data, most often in RGB format, are converted to the Munsell coordinates using either lookup tables or closed formulas. The conversion from RGB components to the Munsell Hue, Value and Chroma (HVC) coordinates can be achieved by using the following mathematical algorithm [14]:

\begin{align}
x &= 0.620R + 0.178G + 0.204B, \\
y &= 0.299R + 0.587G + 0.144B,
\end{align}
2.7. MUNSELL COLOUR SPACE

Figure 2.3: Representation of colour in the Munsell colour space.

\[ z = 0.056G + 0.942B . \]  
(2.37)

A nonlinear transformation is applied to the intermediate values as follows:

\[ p = f(x) - f(y) , \]  
(2.38)

\[ q = 0.4(f(z) - f(y)) , \]  
(2.39)

where \( f(r) = 11.6r^{\frac{1}{3}} - 1.6 \). Further the new variables are transformed to:

\[ s = (a + b\cos(\theta))p , \]  
(2.40)

\[ t = (c + d\sin(\theta))q , \]  
(2.41)

where \( \theta = \tan^{-1}(\frac{q}{p}) \), \( a = 8.880 \), \( b = 0.966 \), \( c = 8.025 \), and \( d = 2.558 \). Finally, we obtain the requested HVC values as:

\[ H = \arctan\left(\frac{s}{t}\right) , \]  
(2.42)

\[ V = f(y) , \]  
(2.43)

\[ C = (s^2 + t^2)^{\frac{1}{2}} . \]  
(2.44)

In summary, the Munsell colour system is an attempt to define colour in terms of hue, chroma and lightness parameters based on subjective observations rather than direct measurements or controlled perceptual experiments.

Nicolaos Ikonomakis
2.8 Summary

The basic colour sensing properties of the human visual system and the CIE standard colour specification system XYZ were described in this chapter.

Colour specification models are of paramount importance in applications where efficient manipulation and communication of images and video frames are required. A number of colour specification models are in use today. Examples include: RGB, rgb, Opponent, YIQ, YUV, Ohta, HSI family, L'*u'*v*, L*a*b*, and Munsell colour spaces. The colour model is a mathematical representation of spectral colours in a finite dimensional vector space. In each one of them the actual colour is reconstructed by combining the basis elements of the vector spaces, the so called primary colours. By defining different primary colours for the representation of the system different colour models can be devised. One important aspect is the colour transformation, the change of coordinates from one colour system to another. Such a transformation associates to each colour in one system a colour in the other model.

Each colour model came into existence for a specific application in colour image processing. Unfortunately, there is no technique for determining the optimum coordinate model for all image processing applications. For a specific application the choice of a colour model depends on the properties of the model and the design characteristics of the application. Table 2.1 summarizes the most popular colour spaces and some of their applications. As the table shows, the RGB

<table>
<thead>
<tr>
<th>Colour Space</th>
<th>Transform</th>
<th>Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGB</td>
<td>non linear</td>
<td>display</td>
</tr>
<tr>
<td>rgb</td>
<td>linear</td>
<td>storage, processing, analysis</td>
</tr>
<tr>
<td>XYZ</td>
<td>linear</td>
<td>colourimetric calculations</td>
</tr>
<tr>
<td>Opponent</td>
<td>linear</td>
<td>segmentation</td>
</tr>
<tr>
<td>YUV &amp; YIQ</td>
<td>linear</td>
<td>coding TV signals</td>
</tr>
<tr>
<td>ITI23</td>
<td>linear</td>
<td>segmentation</td>
</tr>
<tr>
<td>HSI Family</td>
<td>non linear</td>
<td>segmentation, human colour perception</td>
</tr>
<tr>
<td>L'<em>u'<em>v</em> &amp; L</em>a<em>b</em></td>
<td>non linear</td>
<td>colour management, segmentation</td>
</tr>
<tr>
<td>Cylindrical</td>
<td>non linear</td>
<td>segmentation</td>
</tr>
<tr>
<td>Munsell</td>
<td>non linear</td>
<td>human visual system</td>
</tr>
</tbody>
</table>
Chapter 3

Colour Image Segmentation

Techniques

Because of the uncertainty problems encountered while trying to model the human visual system, there are currently a large number of image segmentation techniques that are available. The early attempts at segmentation were on grey-scale images. They are based on three techniques: pixel-based, region-based, and edge-based. These early techniques have been adapted for colour images, and even though these techniques were introduced three decades ago [15, 16, 17], they still find great attention in colour image segmentation research today. Three of the major techniques that have been introduced recently include: physics-based, model-based, and hybrid-based colour image segmentation techniques. The following sections will survey the six major techniques of colour image segmentation used today.

3.1 Pixel-based Techniques

Pixel-based techniques do not consider the spatial context but only decide solely on the basis of the colour features at individual pixels. This attribute has its advantages and disadvantages. Simplicity of the algorithms are an advantage, while lack of spatial constraints make them susceptible to noise in the images. Model-based techniques alleviate the disadvantage of pixel-based techniques by utilizing spatial interaction models to model images. These techniques are discussed in Section 3.4
3.1.1 Histogram Thresholding

The simplest technique of pixel-based segmentation is histogram thresholding. It is one of the oldest and most popular techniques for image segmentation. If an image is composed of distinct regions, the colour histogram of the image usually shows different peaks, each corresponding to one region, separated by a valley. For example, if an image has a distinct object on a background, the colour histogram is likely to be bimodal with a deep valley. In this case, the bottom of the valley is taken as the threshold so that pixels that belong above and below this value on the histogram are grouped into different regions. This is called bilevel thresholding [18]. For multithresholding, the image is composed of a set of distinct regions. In this case, the histogram has one or more deep valleys and the selection of the thresholds becomes a problem of detecting valleys which is not a trivial job.

In the late 1970's, Ohlander et al. [19] proposed a method of colour image segmentation which incorporated the histograms of nine colour features from the three colour spaces RGB, HSI, and YIQ: three from each colour space. The most dominant peak of the nine histograms determines the intervals of the subregion. Pixels falling in this interval create one region and pixels falling out of it other ones. The dominant peak selection is driven by a priority list of seven. The general segmentation algorithm consists of the following steps:

1. Put the image domain into initially empty region list.
2. Determine the nine histograms of the region being considered. (The original region is the whole image)
3. Locate all peaks in the set of histograms.
4. Select the best peak in the list of peaks using the priority list of seven. If none then output this uniform region and go to step 2.
5. Determine and apply threshold.
6. Regions produced are then added to the list of region.
7. Go to step 2.

Ohlander et al. suggested several improvements for the basic method, such as, removing of small regions and adding textural features (e.g. density of edges).

At the same time of Ohlander's work, Ohta et al. [9] attempted to derive a set of effective colour features by systematic experiments of region segmentation. The new colour features were presented in Section 2.4. They applied an Ohlander type segmentation algorithm for the experiment. They explained that the colour feature which has the deep valleys on its histogram and has the largest discriminant power to separate the clusters in a given region need not be the R, G, and B colour features. Since, they explained, a feature is said to have large discriminant power if its variance is
large they tried to derive colour features with large discriminant power by the Karhunen-Loeve (KL) transformation. They developed the I1I2I3 colour space (Equation (2.15-2.17)). They compared their new colour features with the RGB, XYZ, YIQ, L*a*b*, U*V*W*, Irg, and HSI colour spaces. The results of their experiments showed that the I1I2I3 colour space has only a slight advantage over the other seven colour spaces. They justify the choice of their colour space because of the simplicity of transforming to this space from the RGB colour space.

In the early 1980's, Holla [20] modeled the human visual system and describes its use as a preprocessor for scene analysis. His model of the human visual system yields a pair of opponent colours as a 2-dimensional feature for further scene analysis. He describes the efficiency of this model for image segmentation. He determines the rough segmentation only upon the base of the 2-D histogram of the opponent colours. His procedure starts by transforming the RGB values to the opponent colour pairs red-green (RG), yellow-blue (YB), and the intensity feature (I). Then the three channels are smoothed by applying band-pass filters. Then peaks and valleys in the 2-D RG-YB histogram are searched for. Holla states that his method leaves some non-attachable pixels in the image, of which he does nothing about, but claims that his method is superior to Ohlander's et al. [19] and Ohta's et al. [9].

Reimers and von Stein [21] improved on Holla's procedure by merging pixels that are not attached to a region. Spatial neighbourhood relations are used for the merging criterion. Reimers' and von Stein's improvements consists of an additional refinement process that is employed to the segmentation results obtained with Holla's technique. If one or more pixels of the eight neighbours of an unassigned pixel are assigned to the same region, the unassigned pixel is marked for assignment to this region. Nothing is done if none of the neighbourhood pixels is assigned or if several pixels in the neighbourhood belong to different regions. After the entire image is scanned the marked pixels are assigned to the corresponding regions. This procedure is applied five time to the intermediate results. Reimer and von Stein claim that, while 30% to 80% of the pixels in an image are assigned to regions when employing Holla's approach, less than 10% are unassigned to regions when using their modification.

Tominaga [22] developed a histogram thresholding technique that is similar to Ohlander's et al. [19] work. The difference in his technique is that he attempted to detect the peaks of the three histograms in the hue, value, and chroma (HVC) components of the Munsell space. As mentioned is Section 2.7, conversion from the RGB to the HVC colour space is based on a table [4]. His segmentation algorithm consists of the following steps:

1. The histogram, of the region under consideration, is computed for each of the colour features (HVC). Initially the entire image is regarded as the region. The histograms are smoothed by an average operator.

2. The most dominant peak in either of the three histograms is found. The peak selection is
3.1. PIXEL-BASED TECHNIQUES

based on the shape analysis of each peak under consideration. First, some clear peaks are selected. Next, the following criterion function is calculated for each candidate peak:

\[ f = \frac{S_p}{T_a F_p} \times 100 \quad (3.1) \]

where \( S_p \) denotes a peak area between two valleys, \( F_p \) is the full-width at half the maximum of the peak, and \( T_a \) is the total number of pixels in the specified region; the area of the histogram.

3. The two thresholds, one on each side, of the most dominant peak of the three histograms, are derived. Applying the thresholds, partition the region into two sets of subregions: one consists of subregions corresponding to the colour attributes within the threshold limits, and the other is a set of subregions with the remaining attribute values.

4. The threshold process is repeated for the extracted subregions. If all the histograms become monomodal, a suitable label is assigned to the latest extracted subregions.

5. Steps 1 through 4 are repeated for the remaining regions. The segmentation is terminated when the areas of the regions are sufficiently small in comparison to the original image size or no histogram has significant peaks.

The remaining pixels, which have not been assigned to a region, are merged into the neighbouring regions of similar colour.

Histogram thresholding is one of the simplest methods of image segmentation. This attribute lends it great consideration in current segmentation research when a rough segmentation of the image is needed. Many of the current image and video database systems [23] employ histogram thresholding for image and video retrieval. A seed determination technique which utilizes peak detection in histograms is developed for the proposed segmentation scheme. A common drawback of histogram thresholding is that it often produces unsatisfactory segmentation results on colour images of natural scenes.

3.1.2 Clustering

Clustering is another pixel-based technique that is extensively used for image segmentation. The rationale of the clustering technique is that, typically, the colours in an image tend to form clusters in the histogram, one for each object in the image. In the technique, a histogram is first formed with the colour values at all pixels and the shape of each cluster is found. Each pixel in the image is then assigned to the cluster that is closest to the pixel colour. Many different clustering algorithms
are in existence today [24, 25]. Among these, the K-means and the fuzzy K-means algorithms have received extensive attention [26, 27, 28, 29, 30, 31].

In the early 1990's, Tominaga extended his early work on histogram thresholding [22] to include clustering [32]. His approach consisted of two steps. The first is a modification of the algorithm introduced in [22] and reviewed in the previous section (Section 3.1.1). He modifies his earlier work to overcome the problem of handling overlapping clusters. The modification consists of computing the principal components axes in the CIE L* a* b* colour space for every region to be segmented (i.e. colour features have been transformed onto the principal component axes). The principal component coordinates are obtained from the eigenvalues and eigenvectors of the covariance matrix of colour data vectors. Peaks and valleys are searched for in the three 1-dimensional histograms of the three coordinate axis. The second step consists of reclassifying the pixels on a colour distance measure. Suppose a set of K representative colours \{\overline{m}_1, \overline{m}_2, \ldots, \overline{m}_K\} are extracted from the image. The first cluster center \(\overline{a}_1\) in the colour space is chosen as \(\overline{a}_1 = \overline{m}_1\). Next, the colour difference from \(\overline{m}_2\) to \(\overline{a}_1\) is computed. If this difference exceeds a given threshold \(T\), a new cluster center \(\overline{a}_2 = \overline{m}_2\) is created. Otherwise \(\overline{m}_2\) is assigned to the domain of the class \(\overline{a}_1\). In a similar fashion, the colour difference from each representative colour \(\overline{m}_3, \overline{m}_4, \ldots\) to every established cluster center is computed and thresholded. A new cluster is created if all of these distances exceed \(T\), otherwise the colour is assigned to the class to which it is closest. In his paper, Tominaga does not mention the colour difference measure he uses.

Celenk [33, 34] used a method of detecting clusters by fitting to them circular-cylindrical decision elements in the CIE L* a* b* uniform colour coordinate system. This estimates the clusters' colour distributions without imposing any constraints on their forms. Boundaries of the decision elements are formed with constant lightness and chromaticity loci. Each boundary is obtained using only 1-D histograms of the L* H* C* cylindrical coordinates. The Fisher linear discriminant method [35] is then used to simultaneously project the detected colour clusters onto a line. For two clusters \(w_1\) and \(w_2\) the Fisher line \(W\) is given by:

\[
W = (K_1 + K_2)^{-1}(M_1 - M_2),
\]

where \((K_1, K_2)\) and \((M_1, M_2)\) are the covariance matrices and the mean vectors, respectively, of the two clusters. The colour vectors of the image points, which are the elements of clusters \(w_1\) and \(w_2\), are then projected onto this line using the equation \(d(C) = W^TC\), where \(C\) is a colour vector in one of the clusters and is the linear discriminant function. The 1-D histogram is calculated for the projected data points and thresholds are determined by the peaks and valleys in the histogram. Projecting the estimated colour clusters onto a line permits utilization of all the property values of clusters for segmentation and inherently recognizes their respective cross correlation. This way, the
region acceptance is not limited to the information available from one colour component. This gives the method an advantage over the multidimensional histogram thresholding techniques presented in Section 3.1.1.

Shafarenko et al. [36] have recently proposed a colour segmentation algorithm that uses the watershed algorithm to segment the three-dimensional colour histogram of an image. An explanation of the morphological watershed transform can be found in [37]. Shafarenko et al. utilize the \( L^*u^*v^* \) colour space. Shafarenko et al. state that the nonlinearity of the transformation from the RGB space to the \( L^*u^*v^* \) space transforms the homogeneous noise in the RGB space to inhomogeneous noise. Even if the RGB data is smoothed prior to the transformation, any small residual amount of noise may be significantly amplified due to the nonlinearity of the transform. To this end, they employ an adaptive filter that removes noise from a 3-D colour histogram in the \( L^*u^*v^* \) colour space with subsequent perceptual coarsening. There algorithm is as follows:

- calculate the colour histogram of the image
- filter it for noise reduction
- perform perceptual coarsening
- perform clustering using the watershed algorithm in the 3-D \( L^*u^*v^* \) colour space

Shafarenko et al. give an illustration that both noise filtering and perceptual coarsening are essential for good segmentation. This is presented in Figure 3.1 [36].

A new segmentation algorithm based on mathematical morphology has been recently presented by Park et al. [38]. The algorithm employs the idea of thresholding the difference of two Gaussian smoothed 3-D histograms that differ only in the standard deviation used, to get the initial seeds for clustering. They then use an adaptive dilation and a closing operation to extract the number of clusters and their representative values. Through experimentation on various colour spaces (RGB, XYZ, YIQ, U*V*W*, \( I_1I_2I_3 \)), they concluded that the proposed algorithm yields almost identical segmentation results in any colour space. The algorithm works independently of the choice of colour space.

Among the most popular clustering algorithms in existence today [24, 25], the K-means and the fuzzy K-means algorithms have received extensive attention [26, 27, 28, 29, 30, 31]. A survey of segmentation techniques that have been developed and that utilize these clustering algorithms will presented next.

**K-means algorithm**

The K-means algorithm for cluster-seeking is based on the minimization of a performance index which is defined as the sum of the squared distances from all points in a cluster domain to the
3.1. PIXEL-BASED TECHNIQUES

Figure 3.1: Results of the Shafarenko et al. segmentation algorithm [36]. (a) Original image. (b) Segmentation using both noise filtering and perceptual coarsening. (c) Segmentation without perceptual coarsening. (d) Segmentation without noise filtering.

cluster center. This algorithm consists of the following steps [25]:

1. Determine or choose K initial cluster centers \( c_1(1), c_2(1), \ldots, c_K(1) \). Here \( c_1(1) \) is the colour features of the first cluster center during the first iteration.

2. At the \( k \)th iteration each pixel \( \vec{a} \) is assigned to one of the \( K \) clusters \( C_1(k), \ldots, C_K(k) \), where \( C_j(k) \) denotes the set of pixels whose cluster center is \( c_j(1) \). \( \vec{a} \) is assigned to cluster \( C_j(k) \) if:

\[
\| \vec{a} - c_j(k) \| \leq \| \vec{a} - c_i(k) \| \\
\forall i, j = 1, 2, \ldots, K, \quad i \neq j
\]  

3. From the results of step 2, compute the new cluster centers \( c_j(k+1), j = 1, 2, \ldots, K \), such that the sum of the squared distances from all points in \( C_j(k) \) to the new cluster center is minimized.
In other words, the new cluster center $c^*_j(k + 1)$ is computed so that the performance index

$$J_j = \sum_{\bar{a} \in C_j(k)} \|\bar{a} - c^*_j(k + 1)\|^2, \quad j = 1, 2, \ldots, K,$$

is minimized. The new cluster center which minimizes this is the sample mean of $C_j(k)$. Therefore, the new cluster center is given by:

$$c^*_j(k + 1) = \frac{1}{N_j} \sum_{\bar{a} \in C_j(k)} \bar{a}, \quad j = 1, 2, \ldots, K,$$

where $N_j$ is the number of pixels of cluster $C_j(k)$.

4. If $c^*_j(k + 1) = c^*_j(k)$ for $j = 1, 2, \ldots, K$, the algorithm has converged and the procedure is terminated. Otherwise, go to step 2.

The determination of initial cluster centers plays a crucial part, because the better the initial partition is, the faster the algorithm will converge.

Gevers and Groen [26], compared the K-means clustering algorithm technique to a region-based technique. They compared the two algorithms in seven colour spaces: RGB, XYZ, HSI, rgb, xyz, L*a*b*, and II1213. In their clustering algorithm, they determined the initial cluster centers of the image by first generating the n-dimensional histogram of the image and then determining the dominant peaks in the histogram. The $K$ initial cluster centers correspond to $K$ dominant peaks in the histogram. Their results showed that the K-means clustering technique didn't perform as well as the region-based technique.

Weeks and Hague [27] proposed a segmentation method which uses the K-means algorithm to locate clusters within the HSI colour space. The hue, saturation and luminance components of the image are determined and used to form a three-dimensional vector that represents the colour of any pixel within the image. They treat each colour within the image simply as a three-dimensional vector. $K$ initial cluster centers are initially chosen at random. K-means clustering is then implemented, in the three-dimensional vector space, with the Euclidean distance as a metric to distribute the pixels (Step 2 in K-means algorithm above). Weeks and Hague modified their algorithm by separately segmenting the hue feature and the two-dimensional saturation and luminance features. They modified their algorithm because the hue colour feature corresponds with that of human visual perception and this approach biases the segmentation process toward a hue colour value.

Heisele et al. [28] proposed an algorithm for tracking non-rigid, moving objects in a sequence of coloured images. They use a parallel K-means clustering algorithm to track the centroids of the clusters. Their tracking algorithm is robust with respect to shape variations and partial occlusions.
of the objects.

**Fuzzy K-means algorithm**

Fuzzy sets and fuzzy logic were first introduced by Zadeh [39] in 1965. They take into account situations where the classification into disjoint sets (regions) does not reflect reality in a satisfactory way. Fuzzy sets are an attempt to deal with uncertainty induced, not by the stochastic nature of images, but rather by the imprecision of the existing regions in which the pixels have to be classified.

A fuzzy set \( A \) is represented as:

\[
A = \{ \mu_A(x_i)/x_i, i = 1, 2, \ldots, n \},
\]

where \( \mu_A(x_i) \) gives the degree of belonging of the elements \( x_i \) to the set \( A \).

The fuzzy K-means algorithm, which is also referred to as the fuzzy c-means algorithm, was first generalized by Bezdek [40, 41]. The algorithm uses an iterative optimization of an objective function based on a weighted similarity measure between the pixels in the image and each of the \( K \) cluster centers. A local extremum of this objective function indicates an 'optimal' clustering of the input data. The objective function that is minimized is given by:

\[
J_m(U, v) = \sum_{k=1}^{n} \sum_{i=1}^{K} (\mu_{ik})^m (d_{ik})^2,
\]

where: \( \mu_{ik} \) is the fuzzy membership value of pixel \( k \) in cluster center \( i \), \( d_{ik} \) is any inner product induced norm metric (i.e. the Euclidean norm), \( m \) varies the nature of clustering with hard clustering at \( m = 1 \) and increasingly fuzzier clustering at higher values of \( m \), \( v \) is the set of \( K \) cluster centers, and \( U \) is the fuzzy K-partition of the image. The algorithm relies on the appropriate choices of \( U \) and \( v \) to minimize the objective function given above.

Huntsberger et al. [29] proposed an iterative algorithm based on Bezdek's original algorithm [40, 41] to segment colour images in the RGB and I1I2I3 colour spaces. For the given data points \( \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in \mathbb{R}^p \):

1. Fix the number of clusters \( K \), \( 2 \leq K < n \), where \( n \) is the number of pixels. Fix \( m, 1 \leq m < \infty \).

   Choose any inner product induced norm metric \( \| \cdot \| \).

2. Initialize the fuzzy K-partition, \( U^{(b)} \in \text{all possible fuzzy partitions, with } b = 0 \text{ initially} \).

3. Calculate the \( K \) cluster centers \( \{ \bar{v}_i \} \) with \( U^{(b)} \) and:

\[
\bar{v}_i = \frac{\sum_{k=1}^{n} (\mu_{ik})^m \mathbf{x}_k}{\sum_{k=1}^{n} (\mu_{ik})^m}, \quad i = 1, \ldots, c.
\]
4. Update $U^{(b)}$. Let $d_{ik} = \| x_k - \mathbf{u}_i \|$:

\[
\mu_{ik} = \frac{1}{\sum_{j=1}^{K} \left( \frac{d_{ik}}{d_{jk}} \right)^{2/(m-1)}}
\]  

(3.9)

else, $\mu_{ik} = 0$.

5. Compare $U^{(b)}$ and $U^{(b+1)}$ in a matrix norm: if $\| U^{(b)} - U^{(b+1)} \| \leq \varepsilon$, stop; otherwise, set $b = b + 1$ and return to step 3.

There are a number of parameters that need to be set in the system before the algorithm can be used. These are: $K, m, \varepsilon, U^{(0)}$, the inner product induced norm metric, and the number of items in the data set $n$. Huntsberger et al. decided that, due to the large amount of data items $n$ being processed at any one time, a randomly chosen sample of 2400 pixels taken form the input picture will be initially clustered. They also choose the number of clusters $K$ to be four. The cluster center of this sample are used to calculate membership functions for all of the pixels in the image using (Equation (3.9)) above. These membership functions are examined and any pixel with a membership above a pre-defined threshold, called an $\alpha$-cut, is assigned to the feature space cluster of that membership function. All of the pixels that remain are put back into the algorithm and the process is repeated until either all or a pre-determined amount of the pixels are identified as belonging to the clusters that were found during each iteration. Experiments were done in both the RGB and the I1I2I3 colour spaces. Huntsberger et al. claim that the difference in results between the two is minimal. This type of algorithm will produce spherical or ellipsoidal shaped clusters in the feature space. Huntsberger et al. state that human visual colour matching for constant chromaticity has been shown to follow the spherical or ellipsoidal shaped cluster pattern [29].

Based on the fuzzy clustering principle, Trivedi and Bezdek [30] proposed a segmentation algorithm for aerial images. The method utilizes region growing concepts and pyramidal data structure for hierarchical analysis. Segmentation of the image at a particular processing level is done by the fuzzy K-means algorithm. Four values are replaced by their mean value to construct a higher level in the pyramid. Starting from the highest level, regions are created by pixels that have their fuzzy membership value above $\alpha$-cut. If the homogeneity test fails, regions are split to form the next level regions which are again subjected to the fuzzy K-means algorithm. This algorithm is a region splitting algorithm.

Lim and Lee [31] proposed a colour image segmentation algorithm based upon histogram thresholding and fuzzy K-means techniques. The segmentation technique can be considered as a kind of coarse-to-fine technique. The main reason they adopt this strategy is to reduce the computational complexity required for the fuzzy K-means algorithm. The coarse segmentation stage attempts to segment by using histogram scale space analysis [42, 43]. This analysis enables reliable detection of
dominant peaks in the given histogram and the intervals around those peaks. The bounds of the intervals are found as zero-crossings of the second derivative for a $\tau$-scaled version of the histogram. The $\tau$-scaling of the histogram $h(z)$ is defined by the convolution of $h$ with a Gaussian function which has a mean of zero and the standard deviation equal to $\tau$. The second derivative of the scaled function can be computed by the convolution with the second derivative of the Gaussian function. Those pixels which are not segmented by the coarse segmentation are further segmented using the fuzzy K-means algorithm, proposed by Bezdek [40, 41], in the fine segmentation stage with the pre-determined clusters. Lim and Lee tested their technique in the RGB, XYZ, YIQ, U*V*W*, and U12U3 colour spaces. Figure 3.2 [31], presents their segmentation results for a colour image in the five colour spaces. They concluded that their proposed algorithm yields best results in the U12U3 colour space.

It is widely recognized that the clustering technique to image segmentation suffers from problems related to (1) adjacent clusters frequently overlap in colour space, causing incorrect pixel classification; and (2) clustering is more difficult when the number of clusters is unknown, as is typical for segmentation algorithms [31].

The pixel-based segmentation techniques surveyed above do not consider spatial constraints which make them susceptible to noise in the images. The resulting segmented images often contain isolated, small regions that are not present in noise-free images. In the past decade, many researchers have included spatial constraints in their pixel-based segmentation techniques using statistical models. These techniques will be surveyed in the model-based techniques section (Section 3.4).

### 3.2 Edge-based Techniques

Edge-based techniques all consist of detecting the edges in an image. Edge detection can be considered similar to image segmentation. Instead of finding the regions associated with various objects, the goal of edge detection is to find the discontinuity of a region in the image. Once the edges have been found, the interior can be filled in to obtain the region. Most edge detection techniques are based on finding maxima in the first derivative of the image function or zero-crossings in the second derivative of the image function. Figure 3.3 illustrates this concept for a grey-level image [2]. The figure shows that the first derivative of the grey-level profile is positive at the leading edge of a transition, negative at the trailing edge, and zero in homogeneous areas. The second derivative is positive for that part of the transition associated with the dark side of the edge, negative for that part of the transition associated with the light side of the edge, and zero in homogeneous areas.
Figure 3.2: Fuzzy c-means segmentation results of Lim and Lee algorithm [31] in five colour spaces (a) Original (b) RGB (c) XYZ (d) YIQ (e) U*V*W* (f) U*V*W*.
Since edges are local features, they are determined based on local information. Once the edges within an image have been identified, the image can be segmented into different regions based upon these edges.

3.2.1 Techniques Extended From Monochrome Edge Detection Techniques

In a monochrome image, an edge is defined as an intensity discontinuity. Early approaches to colour edge detection comprise of extensions from monochrome edge detection techniques. These techniques were applied to each of the colour components independently and then the results were combined using certain logical operations [44].

Sobel operator

The first derivative at any point in an image is obtained using the magnitude of the gradient at that point. This can be done using various operators including the Sobel, Prewitt, and Roberts operators [2, 45]. However, the Sobel operator have the advantage of providing both a differencing
and a smoothing effect [2]. Because derivatives enhance noise, the smoothing effect is a particularly attractive feature of the Sobel operator.

Koschan [44] compares various types of edge detectors including the Sobel operator. He implements the Sobel operator by convolving a pixel and its eight neighbours with the following two $3 \times 3$ convolution masks [2, 45, 44]:

\[
M_x = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{pmatrix}, \quad M_y = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}.
\]  

(3.10)

The two masks are applied to each colour channel independently and the sum of the squared convolution results states an approximation of the magnitude of the gradient in each channel. A pixel is regarded as an edge point if the mean of the gradient magnitude values in the three colour channels exceeds a given threshold. Koschan found that the Sobel operator produces very thick edges that have to be thinned.

**Laplacian operator**

The second derivative at any point in an image is obtained by using the Laplacian operator. The basic requirement in defining the Laplacian operator is that the coefficient associated with the center pixel be positive and the coefficients associated with the outer pixels be negative [2]. The sum of the coefficients has to be zero. Koschan [44] applies an eight-neighbour Laplacian operator using the following convolution mask:

\[
M = \begin{pmatrix} 10 & 22 & 10 \\ 22 & 128 & 22 \\ 10 & 22 & 10 \end{pmatrix}.
\]  

(3.11)

The Laplacian mask is applied to the three colour channels independently and the edge points are located by thresholding the maximum gradient magnitude.

As a second-order derivative, the Laplacian operator typically is unacceptably sensitive to noise. This operator also produces many gaps and double edges in the image.

**Mexican Hat operator**

The second derivative at any point in an image can also be obtained by using the Mexican Hat operator [44, 46]. The operator is defined by the negative Laplacian derivative of the Gaussian distribution [44]:

\[
-\nabla^2 G(x, y) = \frac{x^2 + y^2 - 2\sigma^2}{2\pi\sigma^6} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right).
\]  

(3.12)
The convolution masks are generated based on the equation above. The masks are applied to the three colour channels and a pixel is regarded as an edge point if a zero-crossing occurred in any of the three colour channels.

### 3.2.2 Vector Space Approaches

One common problem with the approaches mentioned previously is that they fail to take into account the correlation among the colour channels, and as a result, they are not able to extract certain crucial information revealed by colour. For example, they tend to miss edges that have the same strength but in opposite direction in two of the three colour channels. Consequently, the approach to treat the colour image as vector space has been proposed [47, 48, 49, 50, 51, 52, 53, 54, 55]. A colour image can be viewed as a two-dimensional three-channel vector field which can be characterized by a discrete integer function \( f(x, y) \). The value of this function at each pixel is defined by a three dimensional vector in a given colour space.

**Vector Gradient Operators**

The vector gradient operator employs the concept of a gradient operator on a three channel colour vector space. Di Zenzo [48] proposed a combination of three chromatic gradients for getting a global gradient. He implemented this operator in the RGB colour space.

Let the image be a vector function \( \vec{f}(x, y) = (R(x, y), G(x, y), B(x, y)) \), and let \( \vec{r}, \vec{g}, \vec{b} \) be the unit vectors along the \( R, G, B \) axes, respectively. The horizontal and vertical directional operators can be defined as:

\[
\vec{u} = \frac{\partial R}{\partial x} \vec{r} + \frac{\partial G}{\partial x} \vec{g} + \frac{\partial B}{\partial x} \vec{b},
\]

\[
\vec{v} = \frac{\partial R}{\partial y} \vec{r} + \frac{\partial G}{\partial y} \vec{g} + \frac{\partial B}{\partial y} \vec{b}.
\]

\[
g_{xx} = \vec{u} \cdot \vec{u} = \left( \frac{\partial R}{\partial x} \right)^2 + \left( \frac{\partial G}{\partial x} \right)^2 + \left( \frac{\partial B}{\partial x} \right)^2,
\]

\[
g_{yy} = \vec{v} \cdot \vec{v} = \left( \frac{\partial R}{\partial y} \right)^2 + \left( \frac{\partial G}{\partial y} \right)^2 + \left( \frac{\partial B}{\partial y} \right)^2,
\]

\[
g_{xy} = \frac{\partial R}{\partial x} \frac{\partial R}{\partial y} + \frac{\partial G}{\partial x} \frac{\partial G}{\partial y} + \frac{\partial B}{\partial x} \frac{\partial B}{\partial y}.
\]

Di Zenzo shows that the maximum rate of change of \( \vec{f} \) and the direction of the maximum contrast
3.2. EDGE-BASED TECHNIQUES

can be calculated as:

\[ \theta = \frac{1}{2} \arctan \left( \frac{-2g_{xy}}{g_{xx} - g_{yy}} \right) \],

(3.18)

\[ F(\theta) = \frac{1}{2} \left\{ (g_{xx} + g_{yy}) + \cos 2\theta (g_{xx} - g_{yy}) + 2g_{xy} \sin \theta \right\} \].

(3.19)

Edges can be obtained by thresholding \( \sqrt{F(\theta)} \). Di Zenzo computed the image derivatives along the \( x \) and \( y \) directions by convolving the vector function \( \vec{f} \) with the two Sobel spatial masks (Equation (3.10)).

Unlike the gradient operator extended from monochrome edge detection mentioned above, the vector gradient operator can extract more colour information from the image because it considers the vector nature of the colour image. On the other hand, the vector gradient operator is very sensitive to small texture variations [51]. This may be undesirable in some cases since it can cause confusion in identifying the real objects.

Chapron [49] developed an edge detection technique that is an extension of Di Zenzo's [48] technique. Instead of using the Sobel gradient to compute the elementary gradients; \( \frac{\partial R}{\partial x}, \frac{\partial R}{\partial y} \), etc. Chapron used the Canny-Deriche gradient [56]. The Canny-Deriche's gradient mask is programmed in a recursive way. Several scannings of the image are necessary. Chapron's technique is an improvement over Di Zenzo's in that good results are obtained even when the images are corrupted by Gaussian and impulse noise.

Entropy operator

The entropy operator can be used for both the monochrome and colour image edge detection [52]. The operator yields a small value when the colour chromaticity in the local region is uniform and a large value when there are drastic changes. The entropy in a processing window (i.e. \( 3 \times 3 \) window) centered on pixel \( \vec{v}_0 = (r_0, g_0, b_0) \) is defined as:

\[ H = q_R H_R + q_G H_G + q_B H_B \],

(3.20)

where \( H_R, H_G, H_B \) denote the entropies in the \( R, G, B \) directions, respectively, and

\[ q_R = \frac{r_0}{r_0 + g_0 + b_0} \],

(3.21)

\[ q_G = \frac{g_0}{r_0 + g_0 + b_0} \],

(3.22)

\[ q_B = \frac{b_0}{r_0 + g_0 + b_0} \].

(3.23)
Let $X_0, X_1, \ldots, X_N, (X = R, G, B)$ denote the values in each corresponding colour channel inside the processing window which has $N - 1$ pixels. $H_X$ is defined as:

$$
H_X = -\sum_{i=1}^{N} p_{X_i} \log(p_{X_i}) \log(N),
$$

(3.24)

$$
p_{X_i} = \frac{X_i}{\sum_{j=1}^{N} X_j},
$$

(3.25)

Edges can be extracted by detecting the change of entropy $H$ in a window region. Since the presence of noise can disturb the local chromaticity in an image, the entropy operator is sensitive to noise [52].

**Second Derivative Operators**

Recently, Cumani proposed an edge detection technique which involves second derivative operators [54]. Given a vector field $\mathbf{f}(x, y)$ for a colour image, the squared local contrast of $\mathbf{f}$ at pixel $P = (x, y)$ in the direction of the unit vector $\mathbf{n}(n_1, n_2)$ is defined as [44, 47]:

$$
S(P, \mathbf{n}) = En_1^2 + Fn_2 + Gn_2^2,
$$

(3.26)

where

$$
E = \frac{\partial^2 f}{\partial x^2} \frac{\partial f}{\partial x} = \frac{\partial R}{\partial x} \frac{\partial R}{\partial x} + \frac{\partial G}{\partial x} \frac{\partial G}{\partial x} + \frac{\partial B}{\partial x} \frac{\partial B}{\partial x},
$$

(3.27)

$$
F = \frac{\partial^2 f}{\partial x \partial y} \frac{\partial f}{\partial y} = \frac{\partial R}{\partial x} \frac{\partial G}{\partial y} + \frac{\partial G}{\partial x} \frac{\partial R}{\partial y} + \frac{\partial B}{\partial x} \frac{\partial B}{\partial y},
$$

(3.28)

$$
G = \frac{\partial^2 f}{\partial y^2} \frac{\partial f}{\partial y} = \frac{\partial R}{\partial y} \frac{\partial R}{\partial y} + \frac{\partial G}{\partial y} \frac{\partial G}{\partial y} + \frac{\partial B}{\partial y} \frac{\partial B}{\partial y}.
$$

(3.29)

The eigenvalues of the $2 \times 2$ matrix $\begin{pmatrix} E & F \\ F & G \end{pmatrix}$ coincide with the extreme values of $S(P, \mathbf{n})$ and are attained when $\mathbf{n}$ is the corresponding eigenvector. The extreme value is:

$$
\lambda_{\pm} = \frac{E + G \pm \sqrt{(E - G)^2 + 4F^2}}{2},
$$

(3.30)

The two corresponding eigenvectors $\mathbf{n}_+$ and $\mathbf{n}_-$ are given as:

$$
\mathbf{n}_\pm = (\cos \theta_\pm, \sin \theta_\pm),
$$

(3.31)
\[ \theta_+ = \begin{cases} 
\frac{\pi}{4} & \text{if } (E - G) = 0 \text{ and } F > 0, \\
-\frac{\pi}{4} & \text{if } (E - G) = 0 \text{ and } F < 0, \\
\text{undefined} & \text{if } E = F = G = 0, \\
\frac{1}{2} \arctan \left( \frac{2F}{E-G} \right) + k\pi & \text{otherwise} 
\end{cases} \] 
(3.32)

\[ \theta_- = \theta_+ \pm \frac{\pi}{2} \] 
(3.33)

Possible edge pixels are considered as pixel \( P \) where the first directional derivative \( D_x(P, \vec{n}) \) of maximal squared contrast \( \lambda_+(P) \) is zero in the direction of maximal contrast \( \vec{n}_+(P) \). The directional derivative is defined as:

\[ D_x(P, \vec{n}) = \nabla \lambda_+ \cdot \vec{n}_+ \] 
(3.34)

\[ = \frac{\partial \lambda_+}{\partial x} n_1 + \frac{\partial \lambda_+}{\partial y} n_2 \]

\[ = E_x n_1^3 + (2F_x + E_y)n_2 n_1^2 + (G_x + 2F_y)n_1 n_2^2 + G_y n_2^3. \]

The edge pixels are determined by computing the zero-crossings of \( D_x(P, \vec{n}) \). Since the local directional contrast needs to be a maximum or minimum, the sign of \( D_x \) along a curve tangent at \( P \) in the direction of \( \vec{n}_+ \) is checked and the edge pixel is located if it is found to be a maximal pixel.

The ambiguity of the gradient direction, in the above method, causes some difficulties in locating edge pixels. Cumani [54] suggested the subpixel technique with bilinear interpolation to solve this problem. Alshatti and Lambert [55] suggested a modification in solving the ambiguities by estimating the eigenvector \( \vec{n}_+ \) directly, which can avoid the computational costly subpixel approximation.

Koschan [44] improved on Cumani's, Alshatti's and Lambert's techniques by employing different sized convolution masks based on the derivatives of the Gaussian distribution in the computation process instead of the set of fixed-sized 3 × 3 masks. Koschan found a considerable increase in the quality of the results when the Gaussian masks were employed.

Similar to the vector gradient operator, the second-order derivative operators are very sensitive to texture variations and impulsive noise, but produce thinner edges.

### 3.3 Region-based Techniques

Region-based techniques focus on the continuity of a region in the image. Segmenting an image into regions is directly accomplished through region-based segmentation which makes it one of the most popular techniques used today. Unlike pixel-based, region-based techniques consider both colour distribution in colour space and spatial constraints. Standard techniques include region growing, and split and merge techniques. Region growing is the process of grouping neighbouring pixels or a
collection of pixels of similar properties into larger regions. The split and merge technique constitutes iteratively splitting the image into smaller and smaller regions and testing to see if adjacent regions need to be merged into one. The process of merging pixels or regions to produce larger regions is usually governed by a homogeneity criterion, such as a distance measure linked to colour similarity.

### 3.3.1 Region Growing

Region growing is the process of grouping neighbouring pixels or a collection of pixels of similar properties into larger regions. Testing for similarity is usually achieved through a homogeneity criterion. Quite often after an image is segmented, the regions produced are further merged for improved results. A region growing algorithm typically starts with a number of seed pixels in an image and from these grows regions by iteratively adding unassigned neighbouring pixels, that are similar in colour to the seed pixels, with the existing region of the seed pixel. If the pixel is assigned to the region, the pixel set of the region is updated to include this pixel. Region growing techniques differ in choice of homogeneity criterion and choice of seed pixels. Several homogeneity criteria linked to colour similarity or spatial similarity can be used to analyze if a pixel belongs to a region. These criteria can be defined from local, regional, or global considerations. The choice of seed pixel can be supervised or unsupervised. In a supervised method the user chooses the seed pixels while in an unsupervised method the choice is made by an algorithm.

Gauch and Hsia [57] compared a region growing segmentation algorithm against an edge detection algorithm and a split and merge algorithm. They tested all algorithms in the RGB, YIQ, HLS, and L*a*b* colour spaces. Their region growing algorithm is a supervised one where the seed pixels and threshold values are chosen by a user. They use the Euclidean distance in colour space to determine which pixels in the image satisfy the homogeneity condition. If the colour of the seed pixel is given as \( S = (s_1, s_2, s_3) \) and the colour of a pixel in consideration is \( P = (p_1, p_2, p_3) \), all pixels which satisfy

\[
(s_1 - p_1)^2 + (s_2 - p_2)^2 + (s_3 - p_3)^2 < T^2,
\]

(3.35)

would be included in the region. Here \( T \) is the threshold value which is chosen by the user. The algorithm can be summarized with the following steps:

1. Choose next seed pixel. This seed pixel is the first pixel of the region.
2. Test to see if the four neighbouring pixels (vertical and horizontal neighbours) of the pixel belong to the region with condition (Equation (3.35)).
3. If any of the four neighbouring pixels satisfy the condition, they are assigned to the region, step 2 is repeated, and their four neighbours are considered and tested for homogeneity.
4. When the region is grown to its maximum (i.e. no neighbours of the pixels on the edge of the region satisfy (Equation (3.35))), go to step 1.

Gauch and Hsia found that their region growing algorithm performed best in the HLS and L*a*b* colour spaces. They also suggest that instead of comparing the unassigned pixel to the seed pixel, to compare it to the mean colour of the set of pixels already assigned to the region. Every time a pixel is assigned to the region the mean value is updated. They never conducted any experiments with this new homogeneity criterion.

Tremeau and Borel [58] have recently proposed a colour segmentation algorithm which combines region growing and region merging processes. This algorithm starts with the region growing process which is based on three homogeneity criteria that take into account colour similarity and spatial proximity. The resulting regions are then merged on the basis of a homogeneity criterion that takes into account colour similarity only. The three criteria they used for the region growing approach include:

1. The *Local Homogeneity Criterion*, which corresponds to a local comparison between adjacent pixels;

2. The first *Average Homogeneity Criterion*, which corresponds to a local and regional comparison between a pixel and its neighbourhood, considering only the region under study;

3. And the second *Average Homogeneity Criterion*, which corresponds to a global and regional comparison between a pixel and the studied region.

For a visual point of view, they consider that regions which present similar colour properties belong to the same class, even if they are spatially disconnected. Consequently, these regions are merged using a *Global Homogeneity Criterion* which corresponds to a global comparison of the average colour features representative of the two regions under study. They have also considered that regions, which are spatially dispersed in the image (such as details, edges, or high-frequency noise), have to be merged to the other regions either locally pixel by pixel, or globally. Choice of seed pixels is somewhat arbitrary. The algorithm starts in the top left corner of the image, choosing that pixel as the seed, and continues through the image. All colour comparisons are accomplished using the Euclidean distance measure in the RGB colour space. Threshold values are computed according to an adaptive process relative to the colour distribution of the image. Tremeau and Borel claim that their algorithm can be extended to other uniform colour spaces but new thresholds have to be defined.

Vlachos and Constantinides [59] proposed an image segmentation algorithm which is considered in a graph-theoretic context. The algorithm is based on region growing in the RGB and L*a*b* colour spaces using the Euclidean distance metric to measure the colour similarity between pixels. The
suppression of artificial contouring is formulated as a dual graph-theoretic problem. A hierarchical classification of contours is obtained which facilitates the elimination of the undesirable contours. Regions are represented by vertices in the graph and links between geometrically adjacent regions have weights that are proportional to the colour distance between the regions they connect. The link with the smallest weight determines the regions to be merged. At the next iteration of the algorithm the weights of all the links that are connected to a new region are recomputed before the minimum weight link is selected. The links chosen in this way define a spanning tree on the original graph and the order in which links are chosen defines a hierarchy of image representations. Vlachos and Constantinides found that no clear advantage was gained when employing the L*a*b* colour space instead of the RGB space.

### 3.3.2 Split and Merge

As opposed to the region growing technique of segmentation, where a region is grown from a seed pixel, the split and merge technique subdivides an image initially into a set of arbitrary, disjointed regions and then merge and/or split the regions in an attempt to satisfy a colour homogeneity criterion between the regions. Gonzales and Wood [2] describe a split and merge algorithm that iteratively works toward satisfying these constraints. They describe the split and merge algorithm for grey-scale images proposed by Horowitz and Pavlidis [60]. It will be described here for colour images.

The image is subdivided into smaller and smaller quadrant regions so that for each region a colour homogeneity criterion holds. That is, if for region $R_i$ the homogeneity criterion does not hold, divide the region into four subquadrant regions, and so on. This splitting technique may be represented in the form of a so-called quadtree (a tree in which each node has exactly four descendants), as shown in Figure 3.4. The quadtree data structure is the most commonly used data structure in split and merge algorithms because of its simplicity and computational efficiency [61]. Note that the root of the tree corresponds to the entire image. Merging of adjacent subquadrant regions is allowed if they satisfy a homogeneity criterion. The procedure may be summarized as:

1. Split into four disjointed quadrants any region where a homogeneity criterion does not hold.
2. Merge any adjacent regions that satisfy a homogeneity criterion.
3. Stop when no further merging or splitting is possible.

Most split and merge approaches to image segmentation follow this simple procedure with varying approaches coming from different colour homogeneity criteria.

As mentioned in the previous section, Gauch and Hsia [57] compared a split and merge segmentation algorithm against an edge detection algorithm and a region growing algorithm. They tested all algorithms in the RGB, YIQ, HLS, and L*a*b* colour spaces. They used statistical properties
of the image regions to determine when to split and when to merge. They use the trace of the covariance matrix for a region to determine how homogeneous a given region is. If the mean colour in a region with \( n \) pixels is:

\[
M = (m_1, m_2, m_3) = \left( \frac{\sum_i c_{1i}}{n}, \frac{\sum_i c_{2i}}{n}, \frac{\sum_i c_{3i}}{n} \right),
\]

where \((m_1, m_2, m_3)\) and \((c_1, c_2, c_3)\) represent the three colour features of the mean of the region and of a pixel, respectively, then the trace of the covariance matrix is equal to:

\[
T = v_{11} + v_{22} + v_{33} = \left( \sum_i (c_{1i} - m_1)^2 + \sum_i (c_{2i} - m_2)^2 + \sum_i (c_{3i} - m_3)^2 \right) / n.
\]

If the trace is above a user specified threshold, the region is recursively split. Otherwise, the rectangular region is added to a list of regions to be subsequently merged.

Gauch and Hsia have experimented with two statistical measures for merging regions. The first is based on the trace of the covariance matrix of the merged region. This value is calculated for the two regions that are being considered. If this value is below the specified threshold, then the two regions are merged. Otherwise, they are not. The second method considers the Euclidean colour distance between the means of the two regions to be merged. As with their region growing method, the two regions are merged when this distance is below the specified threshold and not otherwise.

As mentioned in Section 3.1.2, Gevers and Groen [26] compared the quadtree split and merge algorithm to the K-means clustering algorithm. They compared the two algorithms in seven colour spaces: RGB, XYZ, HSI, rgb, xyz, L^*a^*b^*, and I1I2I3. They tested the quadtree split and merge algorithm explained above with two homogeneity criteria: a homogeneity criterion based on func-
3.3. REGION-BASED TECHNIQUES

tional approximation and the mean and variance homogeneity criterion. The functional approximating criterion assumes that the colour over a region may either be constant or variable due to intensity changes caused by shadows and surface curvatures. They used low-order bivariate polynomial approximating functions as the set of approximating functions, because these functions detect useful information (i.e. abrupt changes in the colour features) relatively well and ignore misleading information (i.e. changes in intensity caused by shadows and surface curvature) when the order is not too high. The set of low-order polynomials can be written as:

\[ f_m(x, y) = \sum_{i+j \leq m} a_{ij} x^i y^j, \quad m \leq 4 \quad (3.38) \]

The vector \( \vec{a} \) is calculated with a least-square solver. The fitting error \( \epsilon \) for a region \( R \) is:

\[ \epsilon = \sqrt{\frac{\sum_{(x,y) \in R} (g(x,y) - f_m(x,y))^2}{n}}, \quad (3.39) \]

where \( n \) is the number of pixels in \( R \) and \( g(x,y) \) is the pixel value at coordinates \((x,y)\). The fitting error \( \epsilon \) is compared to the mean noise variance in the region and is considered homogeneous if it is less than this value. The mean and variance homogeneity criterion assumes that the colour of the pixels, discarding noise, over a region is constant and is based on the mean and variance of a region, which is the case for \( m = 0 \). That is:

\[ f_0 = a_{00} \quad (3.40) \]

The fitting error is calculated and compared to the mean noise variance of the region, as before. Their found that the split and merge method of image segmentation outperforms the K-means clustering method.

A major drawback of quadtree-structured split and merge algorithms is the inability to adjust their tessellation to the underlying structure of the image data because of the rigid rectilinear nature of the structure. Gevers and Kajcovski [62] have proposed an image segmentation algorithm to reduce this drawback. Their split and merge algorithm employs the incremental Delaunay triangulation as a directed region partitioning technique which adjusts the image tessellation to the semantics of the image. A Delaunay triangulation of a set of points is a triangulation in which the circumcircle of any of its triangles does not contain any other point in its interior [63]. The homogeneity criterion is the same used by Gevers and Groen [26], explained above. Gevers and Kajcovski tested their algorithm in the L*rg colour space on two synthetic images and obtained excellent results.
Region-based techniques of colour image segmentation are very common today because of their simplicity and computational simplicity. This lends them great attention when hybrid segmentation techniques are created. Region-based techniques are often mixed with other techniques, such as edge detection. These hybrid techniques will be described in Section 3.6.

3.4 Model-based Techniques

Recently, much work has been directed toward stochastic model-based techniques [64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76]. In such techniques, the image regions are modeled as random fields and the segmentation problem is posed as a statistical optimization problem. Compared to previous techniques, the stochastic model-based techniques often provide more precise characterization of the image regions. In fact, various stochastic models can be used to synthesize colour textures that closely resemble natural colour textures in real-world natural images [68]. This characteristic, along with the optimization formulation, provides better segmentation when the image regions are complex and otherwise difficult to discriminate by simple low-order techniques. Most of the techniques introduced use spatial interaction models like Markov Random Field (MRF) or Gibbs Random Field (GRF) to model digital images. Although interest in MRF models for tackling image processing problems can be traced back to the work of Abend et al. [77], only recently have the applicable mathematical tools for exploitation of the full power of MRF in image segmentation found their way into image processing literature. The reports by Cross and Jain [68], Geman and Geman [69], Cohen and Cooper [70, 71], Derin and Elliott [72], Lakshmanan and Derin [73], Panjwani and Healey [74], Liu and Yang [66], Pappas [75], and Chang et al. [76] all make use of the Gibbs distributions for characterizing MRF.

Stochastic model-based colour image segmentation techniques can be either supervised or unsupervised. In a supervised approach, the model parameters are obtained from training data, whereas in an unsupervised approach, the model parameters have to be estimated directly from the observed colour image. Hence, the unsupervised segmentation problem can be considered as a model-fitting problem where a random field model is fitted to an observed image. The unsupervised approach is often necessary in many practical applications where training data is not available, for example, when only one image is available.

Panjwani and Healey [74], developed an unsupervised segmentation algorithm which uses Markov random field models for colour textures. These models characterize a texture in terms of spatial interaction within each colour plane and interaction between different colour planes. The algorithm consists of a region splitting phase and an agglomerative clustering phase and is performed in the RGB colour space. In the region splitting phase, the image is partitioned into a number of square regions that are recursively split until each region satisfies a homogeneity criterion. The agglomer-
ative clustering phase is divided into a conservative merging process followed by a stepwise optimal merging process. Conservative merging uses colour mean and covariance estimates for the efficient processing of local merges. The stepwise optimal merging process maximizes a global performance functional based on the conditional pseudolikelihood of the colour image at each iteration. The stepwise optimal merging process is stopped using a test based on rapid changes in the pseudolikelihood of the image.

The following is a review and survey of the maximum a posteriori (MAP) probability approach to image segmentation, which belongs to the class of Bayesian methods. The form of the outline is in accordance with Tekalp's [78] review of the area but with emphasis on colour image segmentation.

### 3.4.1 The MAP Method

The maximum a posteriori (MAP) approach is motivated by the desire to obtain a segmentation that is spatially connected and robust in the presence of noise in the image. The MAP criterion functional consists of two parts, the class conditional probability distribution, which is characterized by a model that relates the segmentation to the data; and the a priori probability distribution, which expresses the prior expectations about the resulting segmentation.

Derin and Elliott [72] have proposed a MAP approach of the segmentation of monochromatic images, and have successfully used GRF’s as a priori probability models for the segmentation of labels (regions). The GRF prior model expresses the expectation about the spatial properties of the segmentation. In order to eliminate isolated regions in the segmentation that arise in the presence of noise, the GRF model can be designed to assign a higher probability for segmentation results that have contiguous, connected regions. Thus, estimation of the segmentation is not only dependent on the image colour, but also constrained by the expected spatial properties imposed by the GRF model.

The observed colour image channels are denoted by a $3N$-dimensional vector $[y_1,y_2,y_3]^t$. Each individual pixel colour is denoted by $[y_1,s,y_2,s,y_3,s]^t$, where $s$ denotes the pixel location. A segmentation field, denoted by the $N$-dimensional vector $x$ and is consistent with all 3 channels of data, is obtained by assigning labels to each pixel site in the image. A label $x_s = i$, $i = 1, \ldots, K$, implies that the site $s$ belongs to the $i$'th class among the $K$ classes. The desired estimate of the segmentation label field is defined as the one that maximizes the a posteriori pdf $p(x|y)$ of the segmentation label field, given the observed image $y$. Using the Bayes rule,

\[ p(x|y) \propto p(y_1,y_2,y_3|x)p(x) \]
\[ \propto p(y_1|x)p(y_2|x)p(y_3|x)p(x) , \]

Nicolaos Ikonomakis
where $p(y_i|x), i = 1, 2, 3$ represents the conditional pdf of the data given the segmentation labels, i.e., the class-conditional pdf. The term $p(x)$ is the \textit{apriori} probability distribution that can be modeled to impose a spatial connectivity constraint on the segmentation.

\textbf{A priori probability model: GRF}

A spatial connectivity constraint on the segmentation field can be imposed by modeling it as a discrete-valued GRF. Detailed discussion of GRF models can be found in [64, 69, 72]. Using a neighbourhood system of four or eight nearest neighbours and considering only pairwise cliques, the \textit{a priori} probability $p(x)$ can be modeled as a Gibbs distribution,

$$p(x) = \frac{1}{Z} \exp[-U(x)/T] ,$$  \hspace{1cm} (3.42)

where the normalizing constant $Z$ is called the partition function, $T$ is the temperature constant, and $U(x)$, the Gibbs potential (Gibbs energy), is defined by

$$U(x) = \sum_{C \in C} V_C(x) ,$$  \hspace{1cm} (3.43)

where $C$ is the set of all cliques, and $V_C$ is the individual clique potential. Spatial connectivity of the segmentation is imposed by assigning the following clique potential,

$$V_C(i, j) = \begin{cases} -\beta & \text{if } x_i = x_j, \\
+\beta & \text{otherwise} \end{cases}$$  \hspace{1cm} (3.44)

This potential assignment implies higher probability for pixel pairs with identical labels and lower probability for pairs with distinct labels, thus encouraging spatially connected regions.

\textbf{Conditional probability model}

The original image $f$ is modeled by a mean colour function, denoted by the vectors $\mu^1, \mu^2, \mu^3$, plus a zero-mean white Gaussian residual process $r$,

$$f = \mu + r ,$$  \hspace{1cm} (3.45)
Usually a noise-contaminated version of the image is observed, given by

\[ y = f + v, \]

where \( v \) denotes the noise, which is assumed to be a zero-mean white Gaussian process. Therefore,

\[ y = \mu + n, \]

is obtained, where \( n = r + v \) is a zero-mean Gaussian process. This combined term \( n \) can be referred to as the additive noise term. The image is modeled as consisting of \( K \) distinct region, where the \( i \)'th region has the uniform mean colour represented by \((\mu_1^i, \mu_2^i, \mu_3^i)\). Based on the model of the observed image in (3.47), the conditional probability distribution is expressed as

\[ p(y|x) \propto \exp \left\{ \sum_s \left[ \frac{1}{2\sigma_j^2} (y_{j,s} - \mu_j^i) \right] \right\}, \]

where \( y_{j,s} \) represents the intensity data in channel \( j \) at site \( s \), and \( \sigma_j^2 \) denotes the variance of the combined additive noise for the \( j \)'th colour channel. Thus, a pixel represented by the colour triplet \((y_{1,s}, y_{2,s}, y_{3,s})\) is assigned to the region characterized by the class mean \((\mu_1^i, \mu_2^i, \mu_3^i)\) according to the single segmentation label \( z_s = i \).

Note that this is the probability distribution used in the case of estimating the segmentation on the basis of the maximum likelihood (ML) criterion. It should be observed that the MAP estimation follows a procedure that is similar to that of the \( K \)-means algorithm (i.e. start with an initial estimate of the class means and assign each pixel to one of the \( K \) classes by maximizing, then update the class means using these estimated labels, and iterate between these two steps until convergence.)

### 3.4.2 The Adaptive MAP Method

Pappas [75] proposed an adaptive clustering algorithm for monochrome image segmentation that is based on improving the conditional probability model proposed by Derin and Elliott [72], for grey-scale images. The uniform region mean intensity \( \mu_{z_s} \) is not adequate in modeling actual images intensities. Pappas proposed using a space variant intensity \( \mu_{z,s} \) to model each region as a slowly varying function of the site location \( s \). Chang, Sezan, and Tekalp [76] extended Pappas' [75] adaptive clustering algorithm developed for monochrome image segmentation to colour segmentation using an adaptive MAP framework in the \( L^*u^*v^* \) and \( L^*a^*b^* \) colour spaces. They assume that each region \( i \) in the colour image has a distinct space-variant mean colour, denoted by \((\mu_{1,i}^z, \mu_{2,i}^z, \mu_{3,i}^z)\) for each

Nicolaos Ikonomakis
The modified conditional probability becomes:

\[
p(y|x) \propto \exp \left\{ \sum_s \left[ \sum_{j=1}^{3} \frac{1}{2\sigma_j^2} (y_{js} - \mu_{s,xs})^2 \right] \right\} .
\]  

(3.49)

The posterior probability distribution for estimating an adaptive colour segmentation is now:

\[
p(x|y) \propto \exp \left\{ \sum_s \left[ \sum_{j=1}^{3} \frac{1}{2\sigma_j^2} (y_{js} - \mu_{s,xs})^2 \right] + \sum_{C\in C} V_C(x) \right\} .
\]  

(3.50)

Note that a nonadaptive colour clustering algorithm, similar to K-means, is obtained when both the spatial dependence of the class means and the prior probability distribution are ignored.

Results obtained when using most stochastic model-based techniques for colour image segmentation are favourable in most cases, especially with natural scenes. The problems encountered with these techniques is there complexity. These techniques are computationally intensive. There is a trade-off between complexity and segmentation results.

3.5 Physics-based Techniques

Physics-based segmentation techniques use the underlying physical models of the colour image formation process in developing colour difference metrics. The objective of these techniques is to segment a colour image at object boundaries and not at the edges of highlights and shadows in an image [79]. Physics-based techniques allow the segmentation of colour images based on physical models of image formation. The basic mathematical methods used by these techniques are often similar to those already discussed in the previous sections. They differ regarding the reflection models employed for segmenting colour images. For example, Healey [80] uses region splitting guided by preliminary edge detection to classify regions.

Klinker et al. [81, 82] investigated the influence of highlights, shading, and camera properties (e.g. colour clipping, colour balancing, and chromatic lens aberration) on the results of colour image segmentation. They classify physical events with measured colour variation in the image by employing the Dichromatic Reflection Model from dielectrics. Shafer's Dichromatic Reflection Model (DRM) [83] describes the light, \( L(\lambda, i, e, g) \), which is reflected from a point on a dielectric nonuniform material as a mixture of the light \( L_s(\lambda, i, e, g) \) reflected at the material surface and the light \( L_b(\lambda, i, e, g) \) reflected from the material body. The parameters \( i, e, g, \lambda \) denote the angle of incident light, the angle of emitted light, the phase angle, and the wavelength, respectively. The
Using this classification, Klinker et al. developed a hypothesis-based segmentation algorithm. The algorithm searches for colour clusters from local image areas that show the characteristic features of the body and surface reflection processes in a bottom-up manner. When a promising cluster is found in an image area, a hypothesis is generated which describes the object colour and/or highlight colour in the image area and the shading and highlight components of every pixel in the area is determined. The new hypothesis is then applied to the image using a region growing approach. This determines the exact extent of the image area to which the hypothesis applies. This step verifies the applicability of the hypothesis. Accurate segmentation results are presented by Klinker et al. [81, 82] for images of plastic objects.

There are many rigid assumptions of the DRM (e.g. the illumination conditions, the type of materials). For most realistic scenes, these assumptions do not hold. Therefore, the DRM can be used to segment colour scenes taken only within a controlled environment.

3.6 Hybrid Techniques

Recently, there is a sparked interest in hybrid colour image segmentation techniques [84, 85, 86, 87, 88, 89]. These techniques combine the benefits of the various techniques mentioned in previous sections and masks the disadvantages of others.

Tseng and Chang [85] proposed a segmentation scheme that first splits the colour image into chromatic and achromatic regions and then employs a histogram thresholding technique to the two regions, separately. The proposed algorithm is employed in the L*H°C* colour space due to its close relation to human colour perception. In their research, they refer to this colour space as the IHS space with I, H, and S corresponding to L*, H*, and C* respectively. The scheme can be summarized into the following steps:

1. Convert RGB colour values to IHS colour values.

2. Define the effective ranges of hue and saturation in the IHS space, and determine chromatic and achromatic regions in the image.

3. Use hue, saturation, and/or intensity one-dimensional histogram thresholdings to further segment the image.

4. Detect and recover over-segmentation regions using a region merging technique.

Tseng and Chang suggest to split up the colour image into chromatic and achromatic regions to
determine effective ranges of hue and saturation. The criteria for achromatic areas were measured by experimental observation of human eyes and are defined as follows:

1. \( (\text{intensity}>95) \) or \( (\text{intensity}\leq 25) \),
2. \( (81<\text{intensity}\leq 95) \) and \( (\text{saturation}<18) \),
3. \( (61<\text{intensity}\leq 81) \) and \( (\text{saturation}<20) \),
4. \( (51<\text{intensity}\leq 61) \) and \( (\text{saturation}<30) \),
5. \( (41<\text{intensity}\leq 51) \) and \( (\text{saturation}<40) \),
6. \( (25<\text{intensity}\leq 41) \) and \( (\text{saturation}<60) \),

while intensity is re-scaled from 1 to 100, and saturation is variable with a maximal value of 180. In step 3, chromatic regions are segmented using hue histogram thresholding and achromatic regions are segmented using intensity histogram thresholding. The histogram thresholding they employ in their algorithm is the one proposed by Tominaga [22], which is described in Section 3.1.1. Over-segmented regions are recovered using an \( 8 \times 8 \) mask. The mask is evenly divided into sixteen \( 2 \times 2 \) sub-masks. If there is at least a chromatic and an achromatic pixel in the \( 2 \times 2 \) sub-mask, then the sub-mask has a vote to the dispersion of the mask. A special label is assigned the \( 8 \times 8 \) region if the mask possesses more than seven votes. After convoluting the mask throughout the image, region merging is used to merge the labelled regions with the segmented region or to form some new regions.

Zugaj and Lattuati [84] proposed a segmentation technique that combines edge-based segmentation results with region-based segmentation results. In their algorithm they utilize the RGB colour components for its simplicity, in spite of its disadvantages. The algorithm consists of the gradient operator proposed by Di Zenzo [48], for edge detection and the region growing algorithm for region-based segmentation results. They obtain an accurate superposition between the edge pixels supplied by the gradient operator and the contours provided by the region growing pixels approach. This good correlation performs a significant matching for both edges and contour images, improves linkage between dislocated edges, and closes pixel elements of contours. Closing is achieved by a local operation which combines an iterative labeling method associated with a probabilistic relaxation approach. Results of their algorithm are shown in Figure 3.5[84].

Moghaddamzadeh and Bourbakis [87] proposed two colour image segmentation algorithms that employ a fuzzy region growing technique and an edge detection technique in the RGB colour space. One of the proposed algorithms is used for fine segmentation towards compression and coding of image and the other for coarse segmentation towards other applications like object recognition and image understanding. Edge detection and region growing approaches are combined to find large and
Figure 3.5: Results of the Zugaj and Lattuati segmentation algorithm [84]. (a) Original image (b) Segmented image

crisp segments for coarse segmentation. Segments can grow or expand based on two fuzzy criteria. The fuzzy region growing and expanding approaches use histogram tables for fine segmentation.

Ito et al. [88] proposed an image segmentation technique which is based both on edge detection and region extraction. Suitable fuzzy sets representing the colour information of a given image are automatically generated, which are then used for the intuitive edge detection and region extraction (pixel classification) approaches. The method is based on doing both in tandem in order to make up for the disadvantages inherent in applying them singly. In this technique, the HLS colour space is used by employing fuzzy sets to quantify them. Membership functions are based on histograms of hue and lightness. Membership functions are generate by the following steps:

1. The histograms of lightness and hue are smoothed. If the pixel's saturation is too low, the hue attribute of that pixel is ignored.

2. A triangular membership function is set up around a mountain on the histogram, with the apex of the triangular membership function corresponding to the peak of the mountain, and the slope according to the frequencies at peaks and valleys. The lightness or hue attribute corresponding to the peak is called the 'representative colour' of that fuzzy set. Trapezoidal membership functions are assigned when the frequencies are too large.

3. Among the fuzzy sets generated at step 2, those in which representative colour cannot be distinguished are merged into one set.

They proposed a method of combining the results of the two segmentation processes which have been run in parallel, with neither procedure providing the bulk of the segmentation process. The method is based on classifying a pixel as an edge pixel based on four cases which take into account
Saber et al. [86] proposed a new method for colour image segmentation that combines edge detection, a split and merge algorithm, and the model-based technique proposed by Chang, Sezan, and Tekalp [76]. The approach rests on the principle that the segmentation map which is an indicator of 'similarities' between pixels must be consistent with the edge map which represents 'discontinuities' between pixels. The method is summarized in Figure 3.6[86]. First, an initial colour segmentation map is computed where labels form spatially contiguous regions. Then, region labels are optimized by split and merge procedures to enforce consistency with the edge map. There method is performed in the YES luminance-chrominance colour space. However, they state that there method can be easily applied in any other suitable colour space.

Luo et al. [89] have recently proposed an algorithm which integrates a physically meaningful colour space and the corresponding colour difference metric into the adaptive Bayesian k-means framework presented by Pappas [75] and Chang et al. [76] in an effort towards physics-based segmentation of photographic colour images. The algorithm uses a physics-based distance metric to generate regular partitioning of the Lst colour space. The colour difference measure is defined with respect to the mode (peak) of each cluster, \( m \), and a given point \( c \), in the \( st \) plane. The difference measure is given as:

\[
d = \sqrt{w_l \cdot (\Delta L)^2 + w_s \cdot (\Delta a)^2 + w_h \cdot (\Delta h)^2} ,
\]  

\( (3.52) \)
where $w_l, w_s, w_h$ are weighting factors; $\Delta lum$ is the difference in $L$; and $\Delta sat$ and $\Delta hue$ are the saturation and hue components, respectively. Results of their algorithm are shown in Figure 3.7\[89\].

Hybrid-based colour image segmentation techniques have the advantage that they utilize many pixel-based, edge-based, region-based, etc. techniques in tandem and overcome the problems inherent when each of them is applied singly. Because of this inherent advantage, most of the segmentation algorithms developed today are hybrid-based.

### 3.7 Summary

Image segmentation is one of the most important parts of any image analysis process. The segmentation step determines the eventual success or failure of the analysis. It is not easy to create segmentation algorithms whose performance is comparable to that of the human visual system. This is because segmentation of image is weakened by various types of uncertainty making most simple segmentation techniques ineffective. This uncertainty has led to the development of a large number
of different colour image segmentation techniques.

The techniques surveyed in this chapter cover colour image segmentation techniques. The various classes of techniques are summarized in the following table (Table 3.1).
### Colour Image Segmentation

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pixel-based</strong></td>
<td>- colour regions are determined by thresholding peak(s) in the histogram(s)</td>
</tr>
<tr>
<td></td>
<td>- simple to implement</td>
</tr>
<tr>
<td></td>
<td>- no spatial considerations</td>
</tr>
<tr>
<td></td>
<td>- decide on the basis pixel colour</td>
</tr>
<tr>
<td></td>
<td>- no spatial constraints</td>
</tr>
<tr>
<td></td>
<td>- simplicity of algorithms</td>
</tr>
<tr>
<td></td>
<td>- Clustering</td>
</tr>
<tr>
<td></td>
<td>- many clustering algorithms</td>
</tr>
<tr>
<td></td>
<td>- K-means and fuzzy K-means most popular</td>
</tr>
<tr>
<td></td>
<td>- pixels in image are assigned to the cluster that is similar in colour</td>
</tr>
<tr>
<td></td>
<td>- adjacent clusters frequently overlap in colour space, causing incorrect pixel assignments</td>
</tr>
<tr>
<td><strong>Edge-based</strong></td>
<td>- monochrome edge detection techniques applied to each colour component of image independently and then results are combined</td>
</tr>
<tr>
<td></td>
<td>- many first and second derivative operators can be used</td>
</tr>
<tr>
<td></td>
<td>- Sobel, Laplacian, and Mexican Hat operators are most popular</td>
</tr>
<tr>
<td></td>
<td>- Clustering</td>
</tr>
<tr>
<td></td>
<td>- views colour image as a vector space</td>
</tr>
<tr>
<td></td>
<td>- Vector Gradient, Entropy, and Second Derivative operators</td>
</tr>
<tr>
<td></td>
<td>- sensitive to noise</td>
</tr>
<tr>
<td></td>
<td>- Techniques extended from monochrome techniques</td>
</tr>
<tr>
<td><strong>Region-based</strong></td>
<td>- process of growing neighbouring pixels or a collection of pixels of similar colour properties into larger regions</td>
</tr>
<tr>
<td></td>
<td>- further merging of regions is usually needed</td>
</tr>
<tr>
<td></td>
<td>- Region Growing</td>
</tr>
<tr>
<td></td>
<td>- iteratively splitting the image into smaller and smaller regions and merging adjacent regions that satisfy a colour homogeneity criterion</td>
</tr>
<tr>
<td></td>
<td>- octree data structure is most common used data structure in algorithms</td>
</tr>
<tr>
<td></td>
<td>- Split and Merge</td>
</tr>
<tr>
<td></td>
<td>- regions modelled as random fields</td>
</tr>
<tr>
<td></td>
<td>- most techniques use the spatial interaction models like Markov Random Field or Gibba Random Field</td>
</tr>
<tr>
<td></td>
<td>- maximum a posteriori (MAP) approach is most common</td>
</tr>
<tr>
<td></td>
<td>- high complexity</td>
</tr>
<tr>
<td><strong>Model-based</strong></td>
<td>- segmentation of colour images based on physical models of image formation</td>
</tr>
<tr>
<td><strong>Physics-based</strong></td>
<td>- basic methods are similar to traditional methods above</td>
</tr>
<tr>
<td></td>
<td>- most empty the Dichromatic Reflection Model (DRM) from dielectrics</td>
</tr>
<tr>
<td></td>
<td>- many assumptions made</td>
</tr>
<tr>
<td><strong>Hybrid</strong></td>
<td>- best results for images taken in controlled environment</td>
</tr>
<tr>
<td></td>
<td>- combine the advantages of different techniques</td>
</tr>
<tr>
<td></td>
<td>- most common techniques of colour image segmentation today</td>
</tr>
</tbody>
</table>

Table 3.1: Colour Image Segmentation Techniques
Chapter 4

Proposed Colour Image Segmentation Scheme

The hybrid colour image segmentation scheme that has been developed gives efficient, robust, and relatively computationally inexpensive results. By hybrid, the scheme utilizes pixel- and region-based techniques to segment colour images. At the core of the scheme is the region-based region growing algorithm mentioned in Section 3.3.1. Pixel-based techniques are used for pixel classification and determining seeds for the region growing algorithm. The proposed segmentation scheme is shown in Figure 4.1. The scheme is based on the HSI colour space so the first step in the scheme is to convert the RGB colour values of the pixels to the HSI values using the transformation formulae presented in Section 2.5. For clarity, the formulae are:

\[
H = \arccos\left( \frac{1/2((R-G) + (R-B))}{\sqrt{(R-G)(R-G) + (R-B)(G-B)}} \right), \quad \text{if } B > G \text{ then } H = 360^\circ - H \tag{4.1}
\]

\[
S = 1 - 3\frac{\min(R, G, B)}{R + G + B}, \tag{4.2}
\]

\[
I = \frac{R + G + B}{3}. \tag{4.3}
\]

The HSI colour model is redisplayed in Figure 4.2. The hue component of the colour space represents the impression related to the dominant wavelength of the colour stimulus. The saturation corresponds to relative colour purity. Colours with no saturation are grey scale colours. Intensity is the amount of light in a colour. Maximum intensity is sensed as pure white, while minimum is sensed as pure black. Some advantages of the HSI colour spaces over other colour spaces are:
Figure 4.1: Proposed colour image segmentation scheme.

- good compatibility with human intuition
- separability of chromatic values for achromatic values
- the possibility of using one colour feature (hue) only for segmentation

The third point is taken advantage of in the seed determination algorithms of the proposed segmentation scheme.

Figure 4.2: The HSI colour model.

Even though many images have been test, in the following sections and in the following chapters, results for three standard images will be shown. The images are stills from multimedia video sequences. More specifically, they are video-phone type images. The *Claire*, *Canberra*, and *Mother_Daughter* images are shown in Figure 4.5(a), Figure 4.8(a), and Figure 4.8(c), respectively.
4.1 Region Growing

As mentioned in Section 3.3.1, the region growing algorithm starts with a set of seed pixels and from these grows regions by appending to each seed pixel those neighbouring pixels that satisfy a homogeneity criterion. In the proposed scheme, the homogeneity criterion is based on colour similarity. The algorithm is summarized in the Figure 4.3. The first seed pixel, which is classified as the first region, is compared to its 8-connected neighbours: eight neighbours of the seed pixel. Any of the neighbouring pixels that satisfy a homogeneity criterion are assigned to the first region. This neighbour comparison step is repeated for every new pixel assigned to the first region until the region is completely bounded by the edge of the image or by pixels that do not satisfy the criterion. Once the first region is fully grown, the colour of each pixel assigned to the region is changed to the mean colour of the all the pixels in the region. The average colour of the pixels in the region is calculated by determining the average value of the red, green, and blue colour channels. The process is repeated for the next and each of the remaining seed pixels until every pixel in the image is assigned to a region.

![Figure 4.3: The region growing algorithm.](image)

The major problems encountered in region growing algorithms are determining good seed pixels...
and finding a homogeneity criterion that compares colours successfully. These issues will be discussed in the next two chapters.

4.2 Achromatic/Chromatic Pixel Classification

As shown in Figure 4.1, certain steps of the proposed segmentation scheme require the comparison of colour features. For example, during the region merging algorithm regions are compared with one another to test for similarity. When comparing the colours of two regions or pixels, a problem is encountered when one or both of the regions or objects have no or very little chromatic information. That is, a grey-scale object can not successfully be compared to an object that has substantial chromatic information. For this reason, all the pixels in the image are classified as either chromatic or achromatic pixels. Pixels or regions that have very little or no chromatic information are referred to as achromatic pixels or regions, respectively. Achromatic pixels or regions are never compared to chromatic pixels or regions, respectively.

Classifying the pixels as either chromatic or achromatic can be considered a crude form of segmentation since the image is segmented into two groups. Although this form of segmentation does have an affect on the segmentation results in the latter algorithms there is no change in the pixel colours. The chromatic/achromatic information is used, in the latter algorithms, as an indication of whether two colours can be compared.

As mentioned in the introduction of this chapter, the HSI colour model corresponds closely to the human perception of colour [2]. The hue value of the pixel has the greatest discrimination power among the three values because it is independent of any intensity attribute. Even though hue is the most useful attribute, there are two problems in using this colour value: hue is meaningless when the intensity is very low or very high; and hue is unstable when the saturation is very low [2]. These problems are realized when examining the HSI cone in Figure 4.2. At very low saturation values the hue is difficult to define and at low and high intensity values the hue value cannot accurately discriminate amongst colours. Figure 4.4 shows the HSI cone with the hue problem areas in yellow. Because of the hue attributes the image is first divided into chromatic and achromatic regions by defining effective ranges of hue, saturation, and intensity values.

Since the hue value of a pixel is meaningless when the intensity is very low or very high the achromatic pixels in the image are defined as the pixels that have low or high intensity values. Pixels can also be categorized as achromatic if their saturation value is very low, since hue is unstable for low saturation values. From the concepts discussed above, the pixels in the image with low saturation, low intensity, or high intensity values are classified as achromatic. These threshold values are defined as: SATLOW, INTLOW, and INTHIGH. It was found that achromatic pixels are best defined
4.2. ACHROMATIC/CHROMATIC PIXEL CLASSIFICATION

Figure 4.4: The HSI cone with achromatic region in yellow.

as follows:

achromatic pixels: \[ \text{saturation (S)} < SATLOW = 10\% \text{ of maximum} \] \hspace{1cm} (4.4)
\[ \text{or} \]
\[ \text{intensity (I)} < INTLOW = 10\% \text{ of maximum} \] \hspace{1cm} (4.5)
\[ \text{or} \]
\[ \text{intensity (I)} > INTHIGH = 90\% \text{ of maximum} \] \hspace{1cm} (4.6)

These threshold values were determined by experimental human observation. Pixels that do not fall into the achromatic category are categorized as chromatic pixels. Figures 4.5-4.7 show images with the chromatic pixels in blue and the achromatic pixels as they are in the original image. In all the figures, and in the remainder of the dissertation, the saturation and intensity values will be given on a scale of 0 to 100. Figure 4.5 compares the results when only the \( SATLOW \) threshold value changes, while Figure 4.6 compares the results when only the \( INTLOW \) threshold value changes and Figure 4.7 compares the results when only the \( INTHIGH \) threshold value changes. It can be observed, in all three scenarios, that having to low threshold values classifies achromatic pixels as chromatic and having high values classifies chromatic pixels as achromatic. It may be noted that most colour images do not have many achromatic pixels, as is observed in Figure 4.8.
4.2. ACHROMATIC/CHROMATIC PIXEL CLASSIFICATION

Figure 4.5: Pixel classification with chromatic pixels in blue and achromatic pixels in the original colour. (a) Original image. Achromatic pixels have intensity < 10, intensity > 90, or (b) saturation < 5 (c) saturation < 10 (d) saturation < 15.

Figure 4.6: Pixel classification with chromatic pixels in blue and achromatic pixels in the original colour. (a) Original image. Achromatic pixels have saturation < 10, intensity > 90, or (b) intensity < 5 (c) intensity < 10 (d) intensity < 15.

Figure 4.7: Pixel classification with chromatic pixels in blue and achromatic pixels in the original colour. (a) Original image. Achromatic pixels have saturation < 10, intensity < 10, or (b) intensity > 85 (c) intensity > 90 (d) intensity > 95.
Figure 4.8: (a) Original image (b) Pixel classification with chromatic pixels in red and achromatic pixels in the original colour. (c) Original image. (d) Pixel classification with chromatic pixels in tan and achromatic pixels in the original colour.
Chapter 5

Seed Determination

Determining seed pixels is of great importance to a region growing algorithm since these seeds are the starting points of the growing process. Good starting points will lead to successful segmentation results. Finding the 'best' seed pixels encompasses finding the pixels that are: dominant in colour in the region and predominantly in the spatial center of the region. Seeds can be determined autonomously or with user-supervision.

An interactive algorithm was developed to find the seed pixels in the image [90]. The seed pixels are selected by a user. Determining the best seed pixels with a supervised algorithm is easily accomplished since the human visual system is the best segmentation algorithm. The algorithm is used in the analysis of the colour distance measures (Chapter 6) and in the merging algorithm (Chapter 7).

Although supervised seed determination ultimately gives the best results, the development of an automatic algorithm was the goal set out from the beginning. Autonomous seed determination is very difficult to achieve since finding good seed pixels requires a somewhat segmented version of the image in order for the regions to be defined. Because of the difficulty of autonomous seed determination, currently, very little research is being done in the area. As mentioned in Section 3.3.1, Gauch and Hsia [57] avoided automation and used a supervised technique, like the one mentioned above, to determine the seeds in their region growing algorithm. Tremeau’s and Borel’s [58] autonomous choice of seed pixels is somewhat arbitrary. Their region growing algorithm starts by choosing the pixel in the top left corner of the image as the seed pixel and continues arbitrarily choosing seeds as the algorithm travels through the image.

As referred to in Figure 4.1, the seed determination algorithm requires knowledge from the achromatic/ chromatic pixel classification algorithm (Section 4.2) because only seeds that are chromatic are considered. Since, as mentioned in Section 4.2, most colour images do not have many achromatic
regions, not considering them will not affect the results.

A simple form of automatic seed determination would be to randomly select pixels in the image as seeds. Obviously, this form of seed determination does not attempt to find the best seed pixels.

Two techniques of autonomous seed determination were developed and examined. The first, attempts to find the seeds by calculating the local colour variance in $3 \times 3$ masks in a hierarchy system [91]. The second, finds the seeds by determining the dominant colours in the image with the use of histograms. Both of the techniques are pixel-based. They are presented in the next two sections.

5.1 Local Variance Seed Determination

The local variance seed determination technique of seed determination employs colour variance mask to the image on different levels [91]. The $3 \times 3$ mask calculates the variance, in colour, of the nine pixels. The basic idea behind the technique is that if the variance of the 9 pixels is less than a predefined threshold then the spatially center pixel is considered as a seed pixel. For complexity reasons, only one colour channel is considered when calculating the variance. Because hue (H) is the most significant colour channel that can be used to detect uniform colour regions [2, 7, 92], it is used to calculated the colour variance.

The seed determination algorithm works in a hierarchical scheme. All the pixels in the image are first considered as level zero seed pixels. At level one, a $3 \times 3$ non-overlapping mask is applied to the chromatic pixels in the image. The mask determines the variance in hue values of the nine level zero pixels. If the variance is less than a certain threshold and the nine level zero pixels in the mask are chromatic pixels then the center pixel of the mask is considered as a level one seed pixel. All level one seed pixels will have their hue values changed to the average hue value of their respective 8 neighbours. The first level seeds represent regions of $3 \times 3$ pixels in the image. In the second level, the non-overlapping mask is applied to the level one seed pixels in the image. Once again, the mask determines the variance in the average hue values of the nine level one seed pixels. If the variance is less than a certain threshold then the center pixel of the mask is considered as a level two seed pixel and the eight other level one seed pixels are disregarded as seeds. All the level two seed pixels have their hue values changed to the average hue value of their respective level one seed pixels. The second level seeds represent regions of $9 \times 9$ pixels. The process is repeated for successive level seed pixels until the variance threshold is not met by any of the higher level seed pixels. Figure 5.1 shows an example of an image with level 1, 2, and 3 seeds. The algorithm is summarized in the following steps, with $a$ representing the level:
1. All chromatic pixels in the image are set as level 0 seed pixels. Set $a$ to 1.

2. Shift the level $a$ mask to the next nine pixels (beginning of image, if just increased $a$).

3. If the mask reaches the end of the image increase $a$ and go to step 2.

4. If all the seed pixels in the mask are of level $a - 1$, continue. If not, go to step 2.

5. Determine the hue variance of the nine level $a - 1$ seed pixels in the $3 \times 3$ mask. The variance is computed by considering the hue values of the nine pixels, if $a = 1$, and the average hue values of the level $a - 1$ seed pixels, otherwise.

6. If the variance is less than a threshold $T_{VAR}$ then the centre level $a - 1$ seed pixel is changed to a level $a$ seed pixel and the other eight level $a - 1$ seed pixels are no longer considered as seeds.

7. Go to step 2.

Although the image is not altered in the algorithm it can be considered as a crude segmentation of the image since the image is divided into regions.

Since hue is considered as a circular value, the variance and average values of a set of hues cannot be calculated using standard linear equations. To calculate the average and variance of a set
of hue values, first the sum of the cosine and the sine of the nine pixels must be determined [93]:

\[
\begin{align*}
C &= \sum_{k=1}^{9} \cos(H_k) , \\
S &= \sum_{k=1}^{9} \sin(H_k) .
\end{align*}
\]

The average hue, \( \text{AVGHUE} \), of the nine pixels is then defined as:

\[
\text{AVGHUE} = \begin{cases} 
\arctan(S/C) & \text{if } S > 0 \text{ and } C > 0 , \\
\arctan(S/C) + \pi & \text{if } C < 0 , \\
\arctan(S/C) + 2\pi & \text{if } S < 0 \text{ and } C > 0 .
\end{cases}
\]

The variance, \( \text{VARHUE} \), of the nine pixels is determined as follows:

\[
\text{VARHUE} = \begin{cases} 
0 & \text{if } R > 0 , \\
(-2\log(R))^{1/2} & \text{otherwise ,}
\end{cases}
\]

where \( R \) is the radiance of the hue and is defined as:

\[
R = \frac{1}{9} (C^2 + S^2)^{1/2} .
\]

If the value of \( \text{VARHUE} \) is less than the threshold \( TVAR \) then the centre level \( a - 1 \) pixel is changed to a level \( a \) seed. The value of \( TVAR \) varies linearly with the level. The threshold value for each level is determined with the following formula:

\[
TVAR = VAR \cdot a ,
\]

where \( a \) and \( VAR \) are the level and an initial variance threshold value, respectively. Once the different level seeds are determined, the region growing algorithm can commence starting with the highest level seeds.

The original images of Claire and Carphone are displayed in Figures 5.2(a) and 5.3(a), respectively. The automated seed determination algorithm finds seeds in the image that are in the centre of the human defined regions in the image. It was found that increasing the variance threshold \( TVAR \) linearly with the level (i.e. \( TVAR = VAR \cdot a \)) produced the best seed pixels. It was also found that setting \( VAR \) to 0.2 gives the best results, with no undesirable seeds, for all the images.
tested. Figure 5.2(b) shows the original Claire image with the level 3 and higher seed pixels found indicated as white pixels. Here VAR was set at 0.1. In particular, 1 level 4 and 42 level 3 seed pixels were found. Figure 5.2(c) shows the Claire seeds image with VAR = 0.2. Here 1 level 4 and 43 level 3 seed pixels were found. Figure 5.2(d) shows the Claire seeds image with VAR = 0.3. The same number of level 1 and three more (46) level 3 seed pixels were found. Several of the seed pixels found are located at the edges of regions which is undesirable. For example a level 3 pixel was found at the left shoulder of Claire. This seed pixel will not produce a good region.

![Results of local variance seed determination.](image)

Figure 5.2: Results of local variance seed determination. (a) Original Claire Image (b) Image showing seeds with VAR = 0.1 (c) Image showing seeds with VAR = 0.2 (d) Image showing seeds with VAR = 0.3.

The same results were observed with the other images tested. Figure 5.3(b) shows the original Claire image with the level 2 and higher seed pixels found indicated as white pixels. Here VAR was set at 0.1 and there were 4 level 2 seed pixels found. Figure 5.3(c) shows the Claire seeds image with VAR = 0.2. Here 19 level 2 seed pixels were found. Figure 5.3(d) shows the Claire seeds image with VAR = 0.3. An increase to 34 level 2 seed pixels was the result. Once again, undesirable seed pixels were found at the edge of regions.
Figure 5.3: Results of local variance seed determination algorithm. (a) Original Carphone Image (b) Image showing seeds with $VAR = 0.1$ (c) Image showing seeds with $VAR = 0.2$ (d) Image showing seeds with $VAR = 0.3$.

5.2 Histogram Seed Determination

Pixels that are dominant in colour, in their respective regions, can be considered as candidates for seed pixels. This idea can be exploited with colour histograms. Colour histograms are used to determine the dominant colours, and consequently, the seed pixels in the image. As mentioned in the histogram thresholding section (Section 3.1.1) of the segmentation techniques chapter, if an image is composed of distinct regions, the colour histogram of the image shows different peaks, each corresponding to one or several regions. The peaks correspond to the dominant colours in the regions. Once the peaks are found, the seeds can be defined as the pixels in the image that correspond to these peaks.

For complexity and speed purposes, a one dimensional colour histogram is used to determine the seed pixels. As mentioned previously, the hue value of a pixel has the greatest discrimination power among the three HSI values because it is independent of any intensity attribute. Consequently, the hue values of the pixels are used to build the colour histogram of the image. Because a colour histogram is being used, the achromatic pixels of the image are not considered in the algorithm. Since most colour images have a small percentage of achromatic pixels (Section 4.2), this is not a problem.

To determine the peaks in the histogram, a simple but effective first derivative technique is
The algorithm is composed of four steps:

1. Build the hue histogram of the image.
2. Smooth the histogram using an averaging window.
3. Determine the first derivative of the smoothed histogram.
4. Determine the peaks of the histogram using the first derivative information.

Since hue is an angular measure, the histogram developed is a polar one. The difference in hue of a bin at 0° and one at 359° is 1° and not 358°.

The purpose of smoothing is to remove noise spikes and to smoothen peaks. An averaging technique is used, in which, an averaging window is passed through the histogram. Since smoothing may alter the location of peaks, the size of the window is kept low at 9. That is, the bin value of the histogram at angle $H$ is defined as:

$$bin(H) = \frac{\sum_{i=H-4}^{H+4} bin(i)}{9},$$  \hspace{1cm} (5.7)

where $bin(H)$ is the bin count at hue angle $H$ of the smoothed histogram. Figure 5.4 (a) and (b) show the histogram and the smoothed histogram of the Claire image in Figure 5.2(a), respectively. Figure 5.5 (a) and (b) shows the histogram and its smoothed version of the Jelly Bean image in Figure 5.5 (d), respectively. The Jelly Bean image is used here to show the many peaks that could be existent in a histogram.

Finding the first derivative of the smoothed histogram will help determine the peaks. As in 2-D graphs, the first derivative gives the slope of a curve. The slope at the peak will be close to zero while the slope just before and just after the peak will be positive and negative, respectively. The first derivative of the smoothed histogram is calculated using a differencing technique. The first derivative, $first$, at angle $H$ is:

$$first(H) = \frac{(bin(H - 1) - bin(H + 1)) + (bin(H - 2) - bin(H + 2))}{2}.$$  \hspace{1cm} (5.8)

Two values where taken on each side of the hue value, $H$, to avoid any noise that may have not been smoothed.

The peaks of the histogram are found by examining the first derivative histogram. Hue value,
$H$, is considered a peak if:

$$\text{first}(H - 3), \text{first}(H - 2), \text{first}(H - 1) > 0$$  \hspace{1cm} (5.9)

and

$$\text{first}(H + 3), \text{first}(H + 2), \text{first}(H + 1) < 0$$  \hspace{1cm} (5.10)

and

$$|\text{first}(H)| < |\text{first}(H - 1)|$$  \hspace{1cm} (5.11)

and

$$|\text{first}(H)| < |\text{first}(H + 1)|.$$  \hspace{1cm} (5.12)

The first two conditions (Equation (5.9) and (5.10)) check the slope 3 bins just before and 3 bins just after the peak. The last two conditions (Equation (5.11) and (5.12)) make sure the peak has a slope closer to zero when compared to the slopes of the two bins on either side of it. Figure 5.4(c) shows the first derivative histogram of the Claire image. Peaks were detected at hue values of $34^\circ, 125^\circ, 164^\circ, 206^\circ,$ and $238^\circ$. Figure 5.5(c) shows the first derivative histogram of the Jelly Bean image. Here six peaks were detected at hue values of $112^\circ, 133^\circ, 175^\circ, 202^\circ, 314^\circ,$ and $336^\circ$. As can be seen, this peak detection algorithm accurately finds the dominant peaks in the histogram.

Once the peaks are known, all the pixels in the image that correspond to the peaks are classified as seed pixels. Figures 5.6 shows the original Claire and Jelly Bean images with seeds found using the algorithm in white. As can be observed, the seeds found are not necessarily in the the spatial
Figure 5.5: (a) Hue histogram of Jelly Bean image (b) Smoothed histogram (c) First derivative (d) Original Jelly Bean image.

centre of the regions. This may affect the results of the region growing algorithm.

5.3 Summary

Seed pixels are used in the region growing algorithm as starting points. They are, essentially, like anchors for growing the regions. Good seed pixels may not lead to perfect segmentation results, but, poor seed pixels will definitely lead to poor results. Finding the 'best' seed pixels encompasses finding the pixels that are: dominant in colour in the region and predominantly in the spatial centre of the region. Two different autonomous seed determination algorithms were developed. The local variance algorithm attempts to find the seeds by calculating the local colour variance in $3 \times 3$ mask
5.3. **SUMMARY**

Figure 5.6: Original image with seeds, found with the histogram seed determination algorithm, in white. (a) *Claire* (b) *Jelly Bean*

The algorithm accurately finds seeds that are spatially in the centre of the regions in the image. The histogram seed determination algorithm finds the seeds by determining the dominant colours in the image with the use of histograms. This algorithm, although finding the dominant colours in the image, does not find the seeds that are spatially in the centre of the regions. Because all the regions may not be represented as peaks in the colour histogram, not all the seed pixels of the regions are found. The local variance algorithm produces the better seeds of the two algorithms.
Chapter 6

Colour Distance Measures

The region growing algorithm, described in Chapter 4, starts with a set of seed pixels and from these grows regions by appending to each seed pixel those neighbouring pixels that satisfy a homogeneity criterion. The homogeneity criterion is used to assess if a pixel belongs to a growing region. This criterion includes a colour distance measure which is used to compare the similarity between two colours. An accurate distance measure will lead to good segmentation results. If the value of the distance measure used is less than some predefined threshold then the colours being compared are considered similar. The homogeneity criterion used for the proposed scheme is defined as:

if the colour distance between the seed pixel and the pixel under consideration is less than a predefined threshold then the pixel is assigned to the seed pixel’s region.

It will be shown that the Cylindrical distance measure used in the HSI colour space compares colours most accurately.

The distance measures described in this chapter are also considered for the merging algorithm described in Chapter 7. Regions are compared for similarity using distance measures. Any regions that are similar in colour are merged into one.

For the proposed segmentation scheme, two different distance measures have to be used: one for the achromatic pixels and one for the chromatic pixels. The next section will describe and evaluate achromatic distance measures. Section 6.2 describes and evaluates chromatic distance measures for the HSI colour space. Section 6.3 considers other colour spaces and distance measures. These other colour spaces and distance measures were tested to confirm the choice of the HSI colour space and the Cylindrical metric.
6.1 Achromatic Distance Measure

Achromatic pixels have no chromatic information and, therefore, only their intensity values are used for comparison. The homogeneity criterion used is: if the difference in the intensity values between an unassigned pixel and the seed pixel is less than a threshold value, $T_{achrom}$, than the pixel is assigned to the seed pixel's region. That is, if

$$|I_s - I_i| < T_{achrom},$$

then pixel $i$ would be assigned to the region of seed pixel $s$.

Figure 6.1 shows some of the results of the region growing algorithm, on the achromatic pixels, with the homogeneity criterion described above. In all the results shown, the threshold values; $SATLOW$, $INTLOW$, and $INTHIGH$, were set to the preferred values; 10, 10, and 90, respectively. Here again, intensity is taken on a scale of 0 to 100. In the original unsegmented achromatic Claire image (Figure 6.1(a)), there are 14 grey-scale levels. In Figure 6.1(b), with $T_{achrom} = 5$, the achromatic growing algorithm cut the number of grey levels by 1 to 13. With $T_{achrom} = 15$ (Figure 6.1(c)), the number of grey levels dropped to 10. And with $T_{achrom} = 25$ (Figure 6.1(d)), the number of grey levels dropped to 7. However, at this threshold level, regions that should be separate are considered as one, such as the collar of Claire's jacket get assigned to the shaded hair region. Through experimental observation, it was found that the best achromatic threshold value, $T_{achrom}$, is 15. Figures 6.2(a-d) show the results, with $T_{achrom}$ set to 15, for the Carphone and Mother Daughter images.

![Figure 6.1](image-url)

Figure 6.1: (a) Unsegmented achromatic original Claire image. (b-d) Segmented achromatic image with (b) $T_{achrom} = 5$, (c) $T_{achrom} = 15$, and (d) $T_{achrom} = 25$. 

Nicolaos Ikonomakis
6.2 Chromatic Distance Measures

The homogeneity criterion for the chromatic regions encompasses all three HSI colour values. Even though the hue value of the pixel has the greatest discrimination power among the three and the algorithm would therefore be computationally simple if only hue is considered, considering all three values will ensure that the comparison results are more accurate. The homogeneity criterion for the chromatic pixels is: if the value of the colour distance metric used to compare the unassigned pixel and the seed pixel is less than a threshold value $T_{chrom}$ than the pixel is assigned to the region. Varying the value of $T_{chrom}$ controls the degree of segmentation with a low value resorting to an over-segmented image and a high value to an under-segmented image. Three different colour distance measures were compared [94]: the generalized Minkowski metric, the Canberra metric, and the Cylindrical distance metric. In the experimental analysis that will follow, the supervised method of seed determination, mentioned in the previous chapter (Chapter 5), will be used. This ensures that the analysis of the distance measures is unbiased. The manual seeds used for three of the colour images tested are shown in white in Figure 6.3.

Figure 6.2: (a) Unsegmented achromatic original Carphone image. (b) Segmented achromatic image with $T_{achrom} = 15$. (c) Unsegmented achromatic original Mother-Daughter image. (b) Segmented achromatic image with $T_{achrom} = 15$.

Figure 6.3: Manual seeds chosen for the (a) Claire, (b) Carphone, and (c) Mother-Daughter images.
6.2.1 Minkowski Distance Metric

The most commonly used measure to quantify distance between two vectors is the generalized Minkowski \( L_p \) norm metric. It is defined as follows\[35\]:

\[
d_M(i,j) = \left( \sum_{k=1}^{p} |x_i^k - x_j^k|^p \right)^\frac{1}{p},
\]

where \( p \) is the dimension of the vector \( x \) and \( x_i^k \) is the \( k^{th} \) element of \( x \). Three special cases of the \( L_M \) metric are of particular interest. Namely:

1. The City-Block distance (\( L_1 \) norm) corresponding to \( M = 1 \). In this case, the distance between the two \( p \)-D vectors is considered to be the summation of the absolute values between their components:

\[
d_1(i,j) = \sum_{k=1}^{p} |x_i^k - x_j^k|.
\]

2. The Euclidean distance (\( L_2 \) norm) corresponding to \( M = 2 \). In the Euclidean model, the distance between the two \( p \)-D vectors is set to be the square root of the summation of the square distances among their components:

\[
d_2(i,j) = \left( \sum_{k=1}^{p} (x_i^k - x_j^k)^2 \right)^\frac{1}{2}.
\]

3. The Chess-board distance (\( L_{\infty} \) norm) corresponding to \( p = \infty \). In this case, the distance between the two \( p \)-D vectors is considered equal to the maximum distance among their components:

\[
d_{\infty}(i,j) = \max(|x_i^1 - x_j^1|, |x_i^2 - x_j^2|, ..., |x_i^p - x_j^p|).
\]

For the 3-dimensional HSI colour space, the Minkowski metric is defined as follows:

\[
d_M(s,i) = (|H_s - H_i|^a + |S_s - S_i|^b + |I_s - I_i|^c)^\frac{1}{2},
\]

where \( s \) and \( i \) refer to the seed pixel and the pixel being tested for assignment, respectively. Pixel \( i \) is assigned to the region represented by seed pixel \( s \) if the value of the distance metric \( d_M \) is less than the threshold \( T_{\text{chrom}} \).

An advantage of using different variants of the Minkowski metric for comparing colours is that more emphasis can be put on the three different HSI colour values. This can be accomplished by having different values for the powers \( a, b, \) and \( c \). For the City-Block (\( L_1 \) norm) and Euclidean (\( L_2 \) norm) cases, the powers are typically set to \( a = b = c = 1 \). However, for the Chess-board distance, the power is set to \( a = b = c = \infty \) to ensure the maximum distance is considered. The choice of powers can be tailored to the specific application, allowing for more nuanced distance calculations.
norm) distance measures the three powers are all set to 1 and 2, respectively. A disadvantage of the metric is that it considers the three HSI colour channels to be linearly related. That is, a difference of 10 in saturation may not correlate to a difference of 10 in intensity.

An array of values ranging from 1 to 5, including the City-Block and Euclidean distance values, were tested for the three power parameters: $a$, $b$, and $c$. These were tested against varying threshold values $T_{\text{chrom}}$. The parameter $d$ was set to: $3/(a + b + c)$, to justify the performance comparison of the varying powers $a$, $b$, and $c$.

Figures 6.4 and 6.5 show some of the results of the segmentation algorithm with the Minkowski distance measure. The results obtained with the City-block distance (Figures 6.4(a-c)) are not very good. In Figure 6.4(c), with a high threshold value of 25, the lips of Claire are lost while her jacket is segmented, not in one, but several regions. Figures 6.4(d-f) shows some of the segmentation results with the Euclidean distance measure. Figure 6.4(e) and (f) show that the same undesirable results as the ones obtained with the City-block distance are produced. Figures 6.4(g-i) show some of the results obtained with emphasis put on each of the HSI colour components, respectively. Experimental analysis revealed that no improvement in the segmentation results is obtained with emphasis put on any of the colour components. Overall, with the Minkowski distance metric, the better results obtained were with the Euclidean ($L_2$ norm) distance at a threshold of $T_{\text{chrom}} = 15$. Figure 6.5 shows some of the results obtained using the Euclidean distance on other images.

### 6.2.2 Canberra Distance Metric

The Canberra metric applies only to non-negative multivariate data, which is the case when colour vectors are considered. It is defined, for the HSI colour space, as follows[95]:

$$d_{\text{can}}(s, i) = \frac{|H_s - H_i|}{|H_s + H_i|} + \frac{|S_s - S_i|}{|S_s + S_i|} + \frac{|I_s - I_i|}{|I_s + I_i|},$$

(6.7)

where, once again, $s$ and $i$ refer to the seed pixel and the pixel being tested for assignment, respectively. A pixel is assigned to a region if the value of the metric $d_{\text{can}}$ is less than the chromaticity threshold $T_{\text{chrom}}$.

Figures 6.6-6.8 display the results of the proposed segmentation algorithm with the Canberra distance metric. In each of the figures, the segmented images for $T_{\text{chrom}}$ threshold values of 0.1, 0.15, and 0.2 are displayed. Through experimental observation, it was found that the best results that are produced, with the Canberra metric, are with a threshold of $T_{\text{chrom}} = 0.15$.

Unlike the Minkowski metric, a benefit of using the Canberra metric is that the ranges taken for the three HSI colour values do not affect the results of the colour comparisons. This is because each colour channel is normalized.
Figure 6.4: Segmented Claire image with the Minkowski distance measure. (a-c) $a = 1, b = 1, c = 1$ (City-block) and $T_{chrom} = 10, 20, 25$ (d-f) $a = 2, b = 2, c = 2$ (Euclidean) and $T_{chrom} = 10, 15, 25$ (g) $a = 2, b = 2, c = 3, T_{chrom} = 20$ (h) $a = 2, b = 3, c = 2, T_{chrom} = 20$ (i) $a = 3, b = 2, c = 2, T_{chrom} = 10$.

Figure 6.5: Segmented images with the Minkowski distance measure. $a = 2, b = 2, c = 2$ (Euclidean) (a) $T_{chrom} = 10$ (b) $T_{chrom} = 15$ (c) $T_{chrom} = 10$ (d) $T_{chrom} = 15$. 
Figure 6.6: Segmented Claire image with the Canberra distance measure. (a) $T_{chrom} = 0.1$ (b) $T_{chrom} = 0.15$ (c) $T_{chrom} = 0.2$.

Figure 6.7: Segmented Carphone image with the Canberra distance measure. (a) $T_{chrom} = 0.1$ (b) $T_{chrom} = 0.15$ (c) $T_{chrom} = 0.2$.

Figure 6.8: Segmented Mother Daughter image with the Canberra distance measure. (a) $T_{chrom} = 0.1$ (b) $T_{chrom} = 0.15$ (c) $T_{chrom} = 0.2$. 

Nicolaos Ikonomakis
6.2.3 Cylindrical Distance Metric

The Cylindrical distance metric computes the distance between two pixel points in the HSI cone. Hence, it can be considered as the ideal similarity measure for the HSI colour space. It is defined as follows [85]:

\[ d_{cyl}(s,i) = \left( (d_{achrom})^2 + (d_{chrom})^2 \right)^{\frac{1}{2}}, \]  

where \( d_{achrom} \) and \( d_{chrom} \) are the achromatic and chromatic components of the measure, respectively. These components are defined as:

\[ d_{achrom} = |I_s - I_i| \]  

and

\[ d_{chrom} = \left( (S_s)^2 + (S_i)^2 - 2S_s S_i \cos \theta \right)^{\frac{1}{2}}, \]

where

\[ \theta = \begin{cases} 
|H_s - H_i| & \text{if } |H_s - H_i| < 180^\circ, \\
360^\circ - |H_s - H_i| & \text{if } |H_s - H_i| > 180^\circ.
\end{cases} \]  

Once again, in all the equations (Equations 6.9-6.11), \( s \) and \( i \) represent the seed pixel and the pixel under consideration, respectively. A pixel is assigned to a seeds region if the value of the metric \( d_{cyl} \) is less than the threshold \( T_{chrom} \). The value of \( d_{chrom} \) is the distance between the 2-dimensional (hue and saturation) vectors of the seed pixel and the pixel under consideration, on the chromatic plane (Figure 6.9). Hence, \( d_{chrom} \) combines the (chromatic) hue and saturation components of the colour pixels. An examination of the metric (Equation 6.8) reveals that it can be considered as a cylindrical form of the Euclidean distance (\( L_2 \) norm) metric (Equation 6.4).

![Figure 6.9: The chromatic plane of the HSI colour model.](image)

Figure 6.10-6.12 display the results of the proposed segmentation algorithm with the Cylindrical
distance metric. As can be observed, the results are favourable. As will be seen (Chapter 7), with the region merging algorithm employed after these region growing results, the proposed algorithm with the Cylindrical distance measure will produce very good results. Through analysis of all the images, it was determined that the best threshold value with the Cylindrical metric is $T_{chrom} = 15$.

Figure 6.10: Segmented Claire image with the Cylindrical distance measure. (a) $T_{chrom} = 5$ (b) $T_{chrom} = 10$ (c) $T_{chrom} = 15$ (d) $T_{chrom} = 25$.

6.2.4 Summary

Of the three colour distance measures tested, the Cylindrical metric shows most promise [94]. With the Cylindrical metric, good results were obtained for all the types of images tested. A reason for this may be that the HSI colour space is a cylindrical colour space which correlates with the Cylindrical distance measure. On the contrary, the Canberra and Minkowski distance measures are not cylindrical and don't compensate for angular values. As Table 6.2.4 shows, the cylindrical distance measure is more discriminating, in colour difference, than the other two distance measures. Even though the second colour similarity test compares two colours that are visually similar, the Cylindrical distance between the colour is 3.43% of the maximum. This implies that the metric will be able to discriminate two colours that are virtually similar.
Figure 6.11: Segmented Carphone image with the Cylindrical distance measure. (a) $T_{\text{chrom}} = 10$ (b) $T_{\text{chrom}} = 15$ (c) $T_{\text{chrom}} = 20$.

Figure 6.12: Segmented Mother Daughter image with the Cylindrical distance measure. (a) $T_{\text{chrom}} = 10$ (b) $T_{\text{chrom}} = 15$ (c) $T_{\text{chrom}} = 20$.
6.2. CHROMATIC DISTANCE MEASURES

Table 6.1: Comparison of Chromatic Distance Measures

<table>
<thead>
<tr>
<th>Colour 1</th>
<th>Colour 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>R=35 O=19 B=205</td>
<td>R=18 O=153 B=210</td>
</tr>
<tr>
<td>D=43.44 S=0.71.43 E=30.065</td>
<td>D=2112 S=0.43.72 E=40.54.26</td>
</tr>
<tr>
<td>R=45 O=33 B=109</td>
<td>R=43 O=33 B=109</td>
</tr>
<tr>
<td>D=43.34 S=0.71.43 E=30.065</td>
<td>D=43.34 S=0.71.43 E=30.065</td>
</tr>
<tr>
<td>R=25 O=10 B=205</td>
<td>R=16 O=5 B=132</td>
</tr>
<tr>
<td>D=43.34 S=0.71.43 E=30.065</td>
<td>D=2112 S=0.43.72 E=40.54.26</td>
</tr>
<tr>
<td>R=35 O=19 B=205</td>
<td>R=19 O=244 B=210</td>
</tr>
<tr>
<td>D=43.44 S=0.71.43 E=30.065</td>
<td>D=2112 S=0.43.72 E=40.54.26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chromatic Distance Measures</th>
<th>Minkowski</th>
<th>Canberra</th>
<th>Cylindrical</th>
</tr>
</thead>
<tbody>
<tr>
<td>111 (max=3)</td>
<td>222 (max=3)</td>
<td>322 (max=1923.9)</td>
<td>232 (max=394.7)</td>
</tr>
<tr>
<td>38.60</td>
<td>99.97</td>
<td>137.23</td>
<td>184.09</td>
</tr>
<tr>
<td>24.64%</td>
<td>15.50%</td>
<td>7.09%</td>
<td>27.11%</td>
</tr>
<tr>
<td>40.072</td>
<td>7.676</td>
<td>5.337</td>
<td>12.59</td>
</tr>
<tr>
<td>2.65%</td>
<td>1.98%</td>
<td>0.280%</td>
<td>3.19%</td>
</tr>
<tr>
<td>32.275</td>
<td>38.189</td>
<td>22.686</td>
<td>69.259</td>
</tr>
<tr>
<td>13.75%</td>
<td>9.87%</td>
<td>1.17%</td>
<td>22.37%</td>
</tr>
<tr>
<td>12.552</td>
<td>91.285</td>
<td>243.148</td>
<td>88.492</td>
</tr>
<tr>
<td>33.94%</td>
<td>21.02%</td>
<td>12.36%</td>
<td>22.43%</td>
</tr>
</tbody>
</table>
6.3 Other Colour Spaces and Distance Measures

Other colour spaces and distance measures were tested to confirm the choice of the HSI colour space and the Cylindrical distance measure. As can be realized by reviewing the survey of the colour image segmentation techniques (Chapter 3), the RGB, L*a*b*, and Cylindrical L*a*b* colour spaces are among the most popular spaces used by image segmentation researchers. Their involvement in the proposed segmentation scheme are evaluated in the following sections.

6.3.1 RGB Colour Space

The RGB colour space is the most frequently used colour space for image processing because most cameras, scanners, and displays are provided with direct RGB signal input or output. Because of this, analysis of the colour space with the proposed algorithm was conducted. With the RGB colour space the Euclidean and Canberra distance measures were used to compare colours in the region growing algorithm. The Euclidean distance metric for the RGB colour space is defined as:

\[
d_{\text{euclidean.RGB}}(s,i) = \left( (R_s - R_i)^2 + (G_s - G_i)^2 + (B_s - B_i)^2 \right)^{\frac{1}{2}},
\]

and for the Canberra metric as:

\[
d_{\text{can.RGB}}(s,i) = \frac{|R_s - R_i|}{|R_s + R_i|} + \frac{|G_s - G_i|}{|G_s + G_i|} + \frac{|B_s - B_i|}{|B_s + B_i|}.
\]

Again a threshold, defined by \(T_{\text{chrom}}\), is used to assign a pixel to a region.

Even though the RGB colour space is usually not preferred because it is psychologically non-intuitive and non-uniform, the results obtained with this colour space are fairly good. Figures 6.13 and Figure 6.14 show some results with the Euclidean and Canberra distance measure, respectively. With the Euclidean distance measure, the best results are obtained with a threshold value of 50, while, a threshold value of 0.4 gives the best results with the Canberra metric. As can be seen, the Euclidean distance measure gives the better results in the RGB space.

6.3.2 L*a*b* Colour Space

The L*a*b* colour space was introduced in Section 2.6.2. It is a perceptually uniform colour space recommended by the Commission International de l'Eclairage (CIE) in 1976. The linear RGB components are first transformed to CIE XYZ components using the appropriate matrix and then these components are transformed to the L*a*b* colour components. The colour space is computationally intensive to transform to and from the linear as well as non-linear RGB spaces. This is a disadvantage.
Figure 6.13: Segmented images with the Euclidean distance measure in the RGB colour space. (a) $T_{chrom} = 30$ (b) $T_{chrom} = 50$ (c) $T_{chrom} = 50$ (d) $T_{chrom} = 50$

Figure 6.14: Segmented images with the Canberra distance measure in the RGB colour space. (a) $T_{chrom} = 0.2$ (b) $T_{chrom} = 0.4$ (c) $T_{chrom} = 0.4$ (d) $T_{chrom} = 0.4$
if real-time processing is required or if computational resources are at a premium.

The Euclidean distance is widely used, in image segmentation algorithms, as the perceptual colour distance metric in the \( L^*a*b* \) colour space. For the \( L^*a*b* \) colour space it is defined as:

\[
d_{\text{euclidean}_{L^*a*b*}}(s, i) = ((L_s^* - L_i^*)^2 + (a_s^* - a_i^*)^2 + (b_s^* - b_i^*)^2)^{\frac{1}{2}},
\]

(6.14)

![Figure 6.15: Segmented images with the Euclidean distance measure in the \( L^*a*b* \) colour space. (a) \( T_{\text{chrom}} = 5.0 \) (b) \( T_{\text{chrom}} = 7.5 \) (c) \( T_{\text{chrom}} = 10.0 \) (d) \( T_{\text{chrom}} = 7.5 \) (e) \( T_{\text{chrom}} = 10.0 \) (f) \( T_{\text{chrom}} = 7.5 \) (g) \( T_{\text{chrom}} = 10.0 \).](image)

Although the \( L^*a*b* \) colour space is perceptually uniform and the Euclidean distance metric is ideal for determining colour difference, as Figure 6.15 shows, the segmentation results obtained are not very good.

### 6.3.3 The Cylindrical \( L^*a*b* \) Colour Space

The cylindrical \( L^*a*b* \) colour space is referred to as the \( L^*H^*C^* \) colour space or, as it will be referred to here, as the HCV colour space. The colour space was introduced in Section 2.6.3. In 1951 Godlove [96], defined a colour difference metric for the HCV colour space. The Godlove metric
The Godlove metric is similar to the Cylindrical metric (Equations 6.8-6.11) defined in Section 6.2.3. The difference between the two is the colour space used and the factor of 16 in front of the value difference (Equation 6.15).

Figure 6.16, shows some of the results obtained with the HCV colour space and the Godlove distance measure. The results are satisfactory but not worth the computational complexity of transforming to this colour space.

![Figure 6.16](image.png)

Figure 6.16: Segmented images with the Godlove distance measure in the HCV colour space. (a) $T_{chom} = 20$ (b) $T_{chom} = 30$ (c) $T_{chom} = 40$ (d) $T_{chrom} = 30$ (e) $T_{chrom} = 40$ (f) $T_{chrom} = 20$ (g) $T_{chrom} = 30$.

### 6.4 Summary

From all the colour spaces and distance measures tested, the Cylindrical metric used in the HSI colour space compares colours most accurately. A reason for this may be that the HSI colour space is a cylindrical colour space which correlates with the Cylindrical distance measure. The
transformation from the RGB to the L*a*b* colour space is complex to even consider the L*a*b* space for segmentation. Plus, white reference points have to be specified for each image for accurate transformations to the L*a*b* space. The same problem is inherent in the HCV colour space since it is the cylindrical L*a*b* space. Even though the results obtained with the Euclidean distance metric in the RGB colour space are good, it is beneficial to transform to the HSI colour space and use the Cylindrical metric for the better results.

For the proposed segmentation scheme, the chromatic distance that produces the better results is the Cylindrical metric with a threshold of \( T_{\text{chrom}} = 15 \). The achromatic distance measure that gives the better results is a simple intensity difference metric with a threshold of \( T_{\text{achrom}} = 15 \).
Chapter 7

Region Merging

Region merging encompasses grouping existing regions in an image relative to their similarity in colour. It is considered a form of image segmentation but it is usually conducted after a segmentation process has been performed so that regions are produced and defined in the image. As shown if Figure 4.1, for the proposed segmentation scheme, the region merging process is administered after the region growing process.

In general, regions are merged: (1) if they are joined and similar in colour; or (2) if they are similar in colour but spatially not connected. The first constraint is used to merge joint regions that, for some reason, were not defined as one region in the pre-merging segmentation process. For the proposed scheme, low segmentation thresholds \( T_{achrom} \) and \( T_{chrom} \) will lead to some groups of connected regions that should be merged. The second constraint is used to decrease the number of colours in the image by making similar regions identical in colour. Merging regions helps in the control of the number of regions and colours in the image.

In the proposed merging scheme, colour distance measures (Chapter 6) are used to test the colour similarity in regions. The homogeneity criterion used for the scheme is defined as:

\[
\text{if the colour distance between two regions is less than a predefined threshold} \\
\text{then the regions are merged.}
\]

The merging threshold will be referred to as \( T_{merge} \). Merging regions encompasses making the regions involved identical in colour. The colour of the regions are changed to the mean colour of the regions involved. The size of the regions involved in any of the merging will have an affect on their contribution to the colour of the merged regions.

As refered to in Figure 4.1, the merging algorithm requires knowledge from the achromatic/chromatic pixel classification algorithm (Section 4.2) because only regions that are chromatic are
considered for merging. Since, as mentioned in Section 4.2, most colour images do not have many achromatic regions, not considering them will not affect the results.

The merging algorithm starts by determining the distance between each possible pair of regions in the image. The pair that are closest in colour (i.e. smallest distance) are then merged. The process is repeated until the distance between the remaining regions are greater than the threshold $T_{merge}$.

Two different colour distance metrics are examined in the next two sections. The first is a distance metrics that compares only the hue values of regions. The second is the Cylindrical distance metric described in Section 6.2.3.

### 7.1 Hue Difference Metric

For reasons of computational simplicity, a distance measure that compares only the hue values of the regions was tested. The distance measure is similar to the achromatic distance metric described in Section 6.1 with the hue values of regions used instead of the intensity. The metric is defined as:

$$d_{hue}(i, j) = \begin{cases} 
|H_i - H_j| & \text{if } |H_i - H_j| < 180^\circ, \\
360^\circ - |H_i - H_j| & \text{if } |H_i - H_j| > 180^\circ,
\end{cases}$$

(7.1)

where $H_i$ and $H_j$ are the hue values of regions $i$ and $j$, respectively. The hue value of the regions is chosen over the saturation and intensity, because of the discriminating characteristics of the colour channel.

Figures 7.1 and 7.2 show some of the merging results obtained with the Hue difference metric. In all the results the image is first segmented using the region growing algorithm with the cylindrical distance metric. Figure 7.1 shows the results of the merging algorithm on the Claire image. As can be observed, the results are very poor. Regions that should not be merged are merged. The same undesirable results are obtained with the Carphone and Mother_Daughter images in Figure 7.2. Clearly the benefit of simplicity in using the Hue difference metric does not compensate for the poor results obtained.
Figure 7.1: Results of the merging algorithm with the Hue difference metric on the segmented Claire image. (a) Unmerged image $T_{chrom} = 10$ (Figure 6.10(b)) (b) Merged image with $T_{chrom} = 10$ and $T_{merge} = 5$. (c) Merged image with $T_{chrom} = 10$ and $T_{merge} = 10$. (d) Unmerged image $T_{chrom} = 15$ (Figure 6.10(c)) (e) Merged image with $T_{chrom} = 15$ and $T_{merge} = 2.5$. (f) Merged image with $T_{chrom} = 15$ and $T_{merge} = 5$.

Figure 7.2: Results of the merging algorithm with the Hue difference metric on the segmented Carphone and Mother Daughter images. (a) Unmerged Carphone image $T_{chrom} = 10$ (Figure 6.11(b)) (b) Merged image with $T_{merge} = 5$. (c) Unmerged Mother Daughter image $T_{chrom} = 15$ (Figure 6.12(c)) (d) Merged image with $T_{merge} = 5$. 

Nicolaos Ikonomakis
7.2 Cylindrical Distance Metric

As was shown in Chapter 6, the Cylindrical distance measure is the better metric for comparing colours in the HSI colour space. The metric is redefined for the merging algorithm as follows:

\[
d_{cyt}(i, j) = \left[ |H_i - H_j|^2 + (S_i^2 + (S_j)^2 - 2S_iS_j \cos \theta)^2 \right]^{\frac{1}{2}},
\]

(7.2)

where

\[
\theta = \begin{cases} 
|H_i - H_j| & \text{if } |H_i - H_j| < 180^\circ, \\
360^\circ - |H_i - H_j| & \text{if } |H_i - H_j| > 180^\circ.
\end{cases}
\]

(7.3)

The results obtained with the Cylindrical metric are better than the ones with the Hue difference metric presented in the previous section. Figure 7.3 shows the results with the Claire image. Although the lips are lost in almost all the results, the algorithm accurately merges regions that are similar in colour. Since the lips are similar in colour to the face, lose of the lips can not be avoided. Figure 7.4 shows the results of the Carphone and Mother_Daughter images. Through analysis of all the images, it was determined that the best merging threshold value with the Cylindrical metric is \( T_{merge} = 17 \).

7.3 Summary

Merging regions ensures that joint regions that are similar in colour and were not defined as one region in the pre-merging segmentation algorithm are merged. It also ensures that regions that are similar in colour and disjoint from one another are merged. Hence, it helps in the control of the number of regions and colours in the image.

The benefit of simplicity in using the Hue difference metric does not warrant using the metric because of poor results obtained. The Cylindrical distance measure gives better results because the HSI colour space is a cylindrical colour space which correlates with the metric.

For the proposed merging scheme, the colour distance measure that produces the better results is the Cylindrical metric with a merging threshold of \( T_{merge} = 17 \) and segmentation threshold of \( T_{chrom} = 15 \) and \( T_{achrom} = 15 \).
Figure 7.3: Results of the merging algorithm with the Cylindrical distance measure on the segmented Claire image. (a) Unmerged image $T_{chrom} = 10$ (Figure 6.10(b)) (b) Merged image with $T_{chrom} = 10$ and $T_{merge} = 15$. (c) Merged image with $T_{chrom} = 10$ and $T_{merge} = 20$. (d) Unmerged image $T_{chrom} = 15$ (Figure 6.10(c)) (e) Merged image with $T_{chrom} = 15$ and $T_{merge} = 17$. (f) Merged image with $T_{chrom} = 15$ and $T_{merge} = 20$.

Figure 7.4: Results of the merging algorithm with the Cylindrical distance measure on the segmented Carphone and Mother_Daughter images. (a) Unmerged Carphone image $T_{chrom} = 10$ (Figure 6.11(b)) (b) Merged image with $T_{merge} = 15$. (c) Unmerged Mother_Daughter image $T_{chrom} = 15$ (Figure 6.12(c)) (d) Merged image with $T_{merge} = 17$. 
Chapter 8

Results

The colour image segmentation scheme with the algorithms and parameter values that give the better results, as analysed and explained throughout the thesis, is shown in Figure 8.1. The first step in the scheme converts the colour image from the RGB to the HSI colour space. The second step classifies the pixels in the image as either chromatic or achromatic depending on their HSI colour values. The achromatic pixels in the image are defined as:

\[
\begin{align*}
I & > 90\% \text{ of maximum} \\
\text{or} & \\
I & < 10\% \text{ of maximum} \\
\text{or} & \\
S & < 10\% \text{ of maximum}
\end{align*}
\]

This information is used in the remaining steps of the scheme. Next, seeds are determined using the Local Variance algorithm with the variance threshold \( T_{\text{var}} \) set to: \( 0.2 \times a \), where \( a \) is the level of the seed. The region growing algorithm with the Cylindrical distance measure is the next step in the scheme. Regions are grown starting from the seed pixels determined in the previous step. The threshold parameters are set to: \( T_{\text{achrom}} = 15 \) and \( T_{\text{chrom}} = 15 \). The final step merges regions that are similar in colour. The Cylindrical distance metric with a merging threshold parameter value of \( T_{\text{merge}} = 17 \) is used to compare the chromatic regions.

Results using the scheme described above and show in Figure 8.1 are displayed in the next few pages. Only the original and the final segmented images are shown. As can be seen, the results on a varied set of images using the segmentation scheme are very good. Figures 8.2-8.4 show the
segmentation results on images of people. Humans are very difficult to segment because of the detail involved in their appearances. However, the results obtained on these images are good. Figure 8.5, Figure 8.6, and Figure 8.7 show the results on images of flowers, buildings, and landscapes, respectively. The segmentation results on these type of images are very good. Some of the landscape images have a lot of detail and still results are good.
Figure 8.2: Segmentation results with humans in image.
Figure 8.3: Segmentation results with humans in image.
Figure 8.4: Segmentation results with humans in image.
Figure 8.5: Segmentation results with images of flowers.
Figure 8.6: Segmentation results with images of buildings.
Figure 8.7: Segmentation results with images of landscapes.
Chapter 9

Conclusions

An autonomous region- and pixel-based segmentation scheme has been developed that efficiently and robustly partitions a colour image into different regions that are homogeneous with respect to colour. As with human visual perception, image segmentation is an important aspect of any type of image or scene analysis. In the one case, humans use their visual sense to effortlessly partition their surrounding environment into different objects to help recognize the objects, guide their movements, and for almost every other task in their lives. In the other case, image segmentation is usually the first task of any artificial image analysis process and all subsequent tasks, such as feature extraction and object recognition rely heavily on the quality of the segmentation.

The colour image segmentation scheme is based on the HSI colour space. The benefits of considering the hue, saturation, and intensity colour values of pixels and regions include:

- good compatibility with human intuition
- separability of chromatic values for achromatic values
- the possibility of using only one colour feature (hue) for segmentation

All three of these benefits are taken advantage of in the segmentation scheme.

The hybrid-based colour image segmentation scheme combines both region- and pixel-based techniques. The pixel classification algorithm (Section 4.2) and the two seed determination algorithms (Chapter 5) utilize pixel-based techniques while the region growing algorithms (Section 4.1) and the region merging algorithms (Chapter 7) use region-based techniques.

Chapter 4 described the region-based region growing algorithm which is the heart of the segmentation scheme. It also described the achromatic/chromatic pixel classification algorithm. This algorithm alleviates the problem encountered when an achromatic and a chromatic pixel need to be compared in terms of their colour value.
Chapter 5 described and analysed two automated seed determination algorithms. Seed pixels are used in the region growing algorithm as starting points. Finding the 'best' seed pixels encompasses finding the pixels that are dominant in colour and predominantly in the spatial centre of the region. The first algorithm described, attempts to find the seeds by calculation the local colour variance in $3 \times 3$ masks in a hierarchy system. The second algorithm finds the seeds by determining the dominant colours in the image with the use of a 1-dimensional colour histogram. Both of the techniques are pixel-based. While the local variance algorithm finds seed pixels that are dominant in colour and that are in the spatial centre of the regions, the histogram algorithm only considers dominant coloured pixels. Hence, the local variance algorithm determines the better seeds.

Colour distance measures are described and analysed in Chapter 6. Distance measures are used to compare the colour similarity between pixels in the region growing algorithms (Section 4.1) and between regions in the region merging algorithm (Chapter 7). From all the colour spaces and distance measures tested, the Cylindrical metric used in the HSI colour space compares two colour most accurately. The HSI colour space is represented in a cylindrical coordinate-type space and, therefore, correlates well with the Cylindrical distance measure.

As explained in Chapter 7, regions are merged if they are: (1) joined and similar in colour; or (2) similar in colour but spatially not connected. Colour distance measures (Chapter 6) are used to test the colour similarity between regions. Once again, the Cylindrical distance measure gives the better results.

Many algorithms and ideas have been developed during the process of developing the hybrid colour image segmentation scheme. These algorithms can be used in other image processing and segmentation techniques and not only in the proposed scheme. The algorithm that classifies pixels as achromatic or chromatic can be used in image processing algorithms that have colour information. Seed determination is not only needed for region growing but in clustering and model-based segmentation techniques. Automated seed determination algorithms are avoided in segmentation research [57, 58] because of the complexity involved. Colour distance measures are used in image processing algorithms where colour comparison is needed. Merging algorithms are used in most segmentation schemes to further process the image. All the algorithms developed in the thesis help in many facets of image processing and not only in the proposed colour image segmentation scheme.

The segmentation scheme developed is more of a general scheme than an application specific one. Because of the good results obtained with the scheme on a varied set of images, it can be combined with specific applications and adjusted to give better results. For example, in multimedia image retrieval databases segmentation is used for defining and indexing objects. Segmentation is also used for second generation object-based coding techniques for image and video compression and transmission over the Internet. For specific applications, the parameters of the proposed segmentation scheme can be adjusted for better results.
References


Nicolaos Ikonomakis 103


