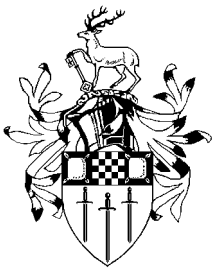
Colour-based Object Recognition

J. Matas

Submitted for the Degree of
Doctor of Philosophy
from the
University of Surrey



Vision, Speech and Signal Processing Group
Department of Electronic and Electrical Engineering
University of Surrey
Guildford, Surrey GU2 5XH, U.K.

May 1996

© J. Matas 1996

Summary

This thesis studies the use of colour information for object recognition. A new representation for objects with multiple colours - the colour adjacency graph (CAG) - is proposed. Each node of the CAG represents a single chromatic component of the image defined as a set of pixels forming a unimodal cluster in the chromatic scattergram. Edges encode information about adjacency of colour components and their reflectance ratio.

The CAG is related to both the histogram and region adjacency graph representations. It is shown to be preserving and combining the best features of these two approaches while avoiding their drawbacks. The proposed approach is tested on a range of difficult object recognition and localisation problems involving complex imagery of non rigid 3D objects under varied viewing conditions with excellent results.

Acknowledgements

I would like to thank my supervisor Josef Kittler for his user-friendly guidance during the course of this work. "Very spatial thanks" go to Radek Mařík. It would have been impossible to carry out all the experiments reported in this thesis without the software libraries he developed and programs he implemented. Lastly and most importantly I would like to thank my wife Romana for her support and understanding.

The work was supported under ESPRIT project 7108 *Vision as process.*

Contents

1 Introduction **1**

1.1 Background 1

1.2 Objective 7

1.3 Outline 9

2 Performance Characterisation **19**

3 Colour-based Recognition: A Survey **25**

3.1 Introduction 25

3.2 Recognition of Single-Coloured Objects 26

3.3 Histogram Based Methods 29

3.4 Region Adjacency Graph 35

4 The Colour Adjacency Graph **39**

4.1 Introduction 39

4.2 Reflection Model 41

4.3 Building the Colour Adjacency Graph 43

4.3.1 CAG Nodes. Graph-theoretical clustering 43

4.3.2 Edges. Reflectance Ratio 48

4.4 The Image, Single View and Object CAG 50

4.4.1 Link to other representations 51

4.5 Properties: Final Remarks. 53

4.5.1 CAG extensions 54

5 CAG Matching **59**

5.1 Introduction 59

5.2 Simple *CAG* matching 61
5.3 Localisation and Identification 62
5.4 Matching by Probabilistic Relaxation 63

6 Recognition Experiments **65**
6.1 Introduction 65
6.2 Design of the model database 65
6.3 Acquisition hardware. Experimental set-up. 66
6.4 Experiments 67
6.5 Computational Complexity and Timing. 69
6.6 Conclusions 71

7 Invariant Recognition and Attention **79**
7.1 Introduction 79
7.2 Surface reflectance, geometry and illumination 81
7.3 Overall structure 85
7.4 Experiments 87
7.5 Conclusions 90

8 Conclusions and Future Work **95**
8.1 Summary and Conclusions 95
8.2 Lessons from Past and Ideas for Future Research 97

A Spatial and Feature Space Clustering **101**
A.1 Introduction 101
A.2 Algorithm 103
A.3 Test Cases 105
A.4 Experiments 109
A.5 Conclusions 113

Chapter 1

Introduction

1.1 Background

Consider the four colour images depicted in figures 1.1a-d. Very few observers will doubt that figures 1.1a-d depict the same object, regardless of the fact that different parts of the object - ICPR proceedings - are imaged at significantly different scales and, in the case of image 1.1b, the visible part of the ICPR proceedings does not overlap any of the parts shown in 1.1a, 1.1b and 1.1d [1]. More importantly, the imaged object can be readily localised in the cluttered scenes of figs. 1.1e and 1.1f using only description of the object derived from any of the figs. 1.1a-d. Recognition and localisation of the proceedings in figures 1.1e and 1.1f poses few problems and requires little effort if colour information about the scenes is available; the same task is undoubtedly more complex if only intensity images of the same scenes, shown in figs. 1.2e-f, are presented to the observer.

In this thesis we investigate the problem of recognition of coloured objects, especially objects with complex colour structure. At a first glance, it may not seem difficult to design and implement a recognition system capable of emulating the almost effortless ability of human observers to build a colour-based model of an object, often from a single image, and to identify instances of the object in a complex scene as shown in the example given above. It may therefore be rather surprising that research in recognition-by-colour received, after playing an important role in the early image interpretation systems like [HR78c] and [NM80], very lit-

[1] Colour images analysed in the thesis are available by anonymous ftp to ftp.ee.surrey.ac.uk, directory /pub/vision/images/cag-iccv95.

(a) (b) (c) (d) (e) (f)

Figure 1.1: ICPR proceeding (in colour). (a) Detailed view, front. (b) Distant view, back. (c) Full view, front. (d) Full frontal view with a specularity. (e)(f) Scenes with occluded views of the proceedings. Colour and adjacency graphs of (a)-(d) are depicted in fig. 1.3. Recognition results for scene (f) are shown in fig. 6.6.

(a) (b)

(c) (d)

(e) (f)

Figure 1.2: ICPR proceeding (intensity image). (a) - (f) See fig. 1.1. Compared with the colour images 1.1(e) and (f), localisation of the ICPR proceedings in figure (e) above requires increased effort; in figure (f) it cannot be performed reliably at all.

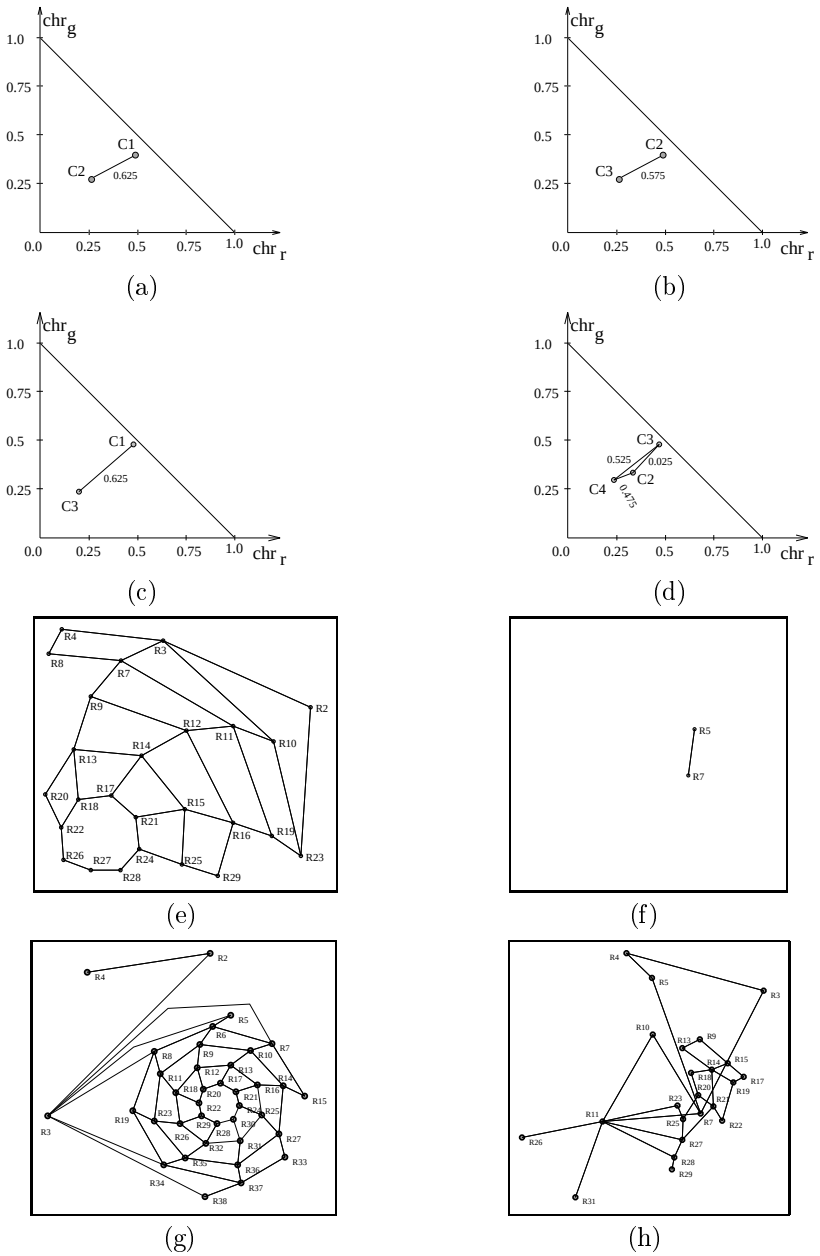


Figure 1.3: ICPR proceedings. Colour (a-d) and region (e-h) adjacency graphs for images shown in fig.1.1a-d. The difference in stability of the *CAG* and *RAG* representation is striking. The *CAG*s are embedded in the chromatic plane to show chromaticity values associated with nodes. *RAG*s are embedded in image plane with nodes positioned at region barycenters. *CAG* edges are attributed with reflectance ratios. Subimages (d) and (h) show the influence of a large specularity on the two representations. The effect is identical to an occlusion by a white object. If the *CAG* of (d) is matched against any of (a)(b)(c), than node C2 of (d) is left unmatched.

tle attention of the computer vision community - a trend reversed only recently [Wix90b, Fat92, EM93, SS94, GROK94] in the wake of Swain's influential work [SB90, Swa90].

The evolutionary pattern may have a number of reasons. Swain [SB90] considers the fact that colour is not directly linked to object class identity (and functionality) in the way shape is to be the fundamental cause of the neglect of colour as a recognition cue. The opinion is supported by quotations from influential works of Biederman [Bie85] and Ullman [Ull86]. It is true that while it is possible to imagine a general description of a shape of a car or a cup, colour (or colours) of cars and cups are in general not constrained. On the other hand, daily experience at a car park shows the value of the colour cue for locating a *known* object, namely your car. The usefulness of colour-related information for object categorisation and generic or functional recognition is therefore limited; but this does not reduce the power of colour as an indexing cue in a *model-based* setting. Unsurprisingly, the approach to colour-based recognition presented in this thesis (and other recently published colour related work, eg. [EM93, GROK94]) is purely model-based.[2]

Another reason for the temporary decline in research related to colour recognition may be linked to the shift of attention towards the study of the image formation process. Early approaches to colour-based recognition [HR78b] did not attempt to model the image formation process. Separation of the influence of surface reflectance of objects from the effects of viewing geometry, illumination, noise, etc. on the observed colour was treated as a standard pattern recognition problem solvable by a proper choice of decision surfaces for classification[AMA79] or a choice of colour space [OKS80, HR78b]. A black box approach without an underlying physical model is incapable of predicting effects caused by phenomena not represented in the training set. Given the number of factors that influence colour measurements (discussed at length in chapter 2) it is hardly feasible to obtain a general and representative data set. The unavoidable consequence being the loss of performance in some previously unencountered (illumination, viewing, etc.) conditions.

The physics-based approach pioneered by Shafer [Sha84] changed the perspective. Rather then analysing shapes of clusters in some colour space, Shafer studied

[2] In fact, regardless of the potential for generic recognition, virtually all shape-based recognition systems reported in the literature [Bin82, BH86, Bro83, Low87] rely on pre-stored non-parametric models.

basic properties of light reflection and concluded that reflectance of a large class of materials can be approximated by the dichromatic reflection model. The introduction of the dichromatic model triggered a wave of interest and both practical experiments [KSK87, KSK88, GJT87, Kli93, Mal86, Tom93] and theoretical analyses [LEC90, HB87, Hea89] confirmed the model as being simple enough to stay practical and yet sufficiently general to be commonly applicable. At the same time, significant progress has also been made in the understanding of specular reflection [GJT87, BB88, ZGB99, LB90] and of the effects of changing spectral power distribution of illumination on observed colour. A number of methods for compensation of the effect, known as colour constancy algorithms, have been proposed [MB86, For88, Bri90, DI93, Fin94, FF94, FB95].

The return of interest in colour-based recognition is more than a mere completion of an evolutionary cycle or a fashion-driven swing. The 'classical' [OPR78, HR78c, NM80] and recent approaches [SB90, Fat92, GROK94] differ in two fundamental aspects: the top-level objective and the overall structure of the system within which colour information is processed. Symbolic interpretation, ie. association of a general class label with *every* region of a *prestored* image, was the ultimate goal of systems like VISIONS [HR78c]. To have a chance of achieving this ambitious goal, large amounts of tailored information about objects and specific environments were exploited by a sophisticated control strategy. Impressive results were reported, but the approach has a number of drawbacks. Firstly, it is not easy to obtain all the required a priori scene knowledge. Secondly, performance becomes necessarily highly domain-specific. Moreover, analysis of results is complicated due to the complex control mechanism. Consequently, it is difficult to evaluate the quality of individual components that contribute to the overall performance. Finally, as the proponents of purposive vision might argue [Baj88] [Alo90], symbolic interpretation is not at all necessary for most of the purposes a vision system may serve and may be likened to an attempt to run before being able to walk.

In comparison, most recent systems are designed as light-weight, near real-time modular components of an active vision system. For example, Wixson [Wix90a] uses colour for fast object search in low-resolution images. The search reduces the amount of data that need to be considered by subsequent shape-based recognition, performed on high-resolution images acquired by a sensor pointed to the area se-

lected in the search step. Wixson analyses the efficiency improvements gained by the two step strategy. Grimson [GROK94] addresses a similar problem, but the focus of attention is driven by colour and stereo. Swain's histogram intersection method [SB90] is complementary to the techniques mentioned above; given an image of an already localised object, it is capable of real-time recognition by comparing colour histograms. Ennesser [EM93] uses a modified version of histogram backprojection, a method introduced in [SB90], to localise image regions likely to contain instances of an object.

1.2 Objective

Despite the recent increase in the amount of published work related to colour-based recognition, very limited attention has been given to the central issue investigated in this thesis: how to exploit colour in the context of the *general recognition problem*, which we define as follows:

Given an image and a set of colour-based descriptions of objects in a model database

1. determine which objects appear in the image. (what?)

2. localise the parts of image in which they appear. (where?)

As briefly touched upon in the previous section and described in detail in chapter 3, with the exception of [Fat92], the reported approaches either assume that a region corresponding to a single object has been selected (segmented) before processing of colour information begins or, alternatively, that a pose of an object with known identity has been established. Therefore the use of colour information serves *either* the recognition process *or* as an attention mechanism. Although the introduction of colour into a recognition strategy will generally improve performance, an important question - what is achievable using colour-related information only - is not directly addressed.

In the following chapters we focus on the issue of purely colour-based recognition and present a system for localisation, identification and general recognition. Some parts of the system, eg. the matching module, implement standard solutions; other components, eg. feature and spatial clustering for automatic model acquisition,

improve upon standard techniques. But the core of the system, a new representation for objects with multiple colours, the *colour adjacency graph* (*CAG*), is an original concept promising significant advantages over previously published work [MMK95b] [MMK95a].

The *CAG* is organised in the following way. Each node of the *CAG* represents a single colour (chromatic) component of the image (or object); a component is defined as a set of pixels forming a unimodal cluster in the chromatic histogram. Edges of the *CAG* contain information about adjacency of colour components. Each edge has as an attribute the reflectance ratio, a photometric invariant proposed by Nayar and Bolle [NB93]. The *CAG* is related to both the histogram [SB90] and region adjacency representations [Fat92]. Nodes of *CAG* correspond to *modes* (peaks) of the histogram. Location of the maxima, unlike the associated bin *counts* are extremely stable under viewpoint change and are robust to occlusion. Surprisingly, *CAG*s of many man-made and natural objects have only one or two aspects! In other words, the number of coloured regions and their relative size may change with the viewpoint, self-occlusion and non-rigid transformation; but the set of visible colours and their adjacency relationship remains stable. This is exactly the case for the ICPR proceedings in figures 1.1a-f; but think of a tiger in motion, a bush of roses, the books on your shelf or even yourself, as you are dressed now! But in general object *CAG* does change under the transformations discussed above. We therefore define an object *CAG* as a *union* of all observed image *CAG*s for the object. After sufficient number of observations, the set of nodes in the union should be equivalent to the set of colours on the surface of the object and the set of edges identical to the set of all colour boundaries.

The problem of localisation and recognition of objects can be posed as a search for subgraphs of object *CAG*(s) and the image *CAG*. The *CAG* subgraph matching does not suffer from the complexity problems of the region adjacency graph (*RAG*) matching for two reasons. First, edge matches, requiring agreement of five attributes (two chromaticities and the reflectance ratio), are rarely ambiguous. Moreover, in most of our experiments, the complexity of *CAG*s was at least an order of magnitude smaller than that of *RAG*s (see examples in chapter 4, eg. fig. 5.1).

Although investigation of a particular approach to recognition that is *purely* based on chromatic and topological properties of colour components of objects is the

primary objective of the thesis, it does not mean that we do not see the benefits if an integrated system, especially an active-vision system, combines cues from a number of sources, like motion, shape, texture and colour. On the contrary, the suitability of the proposed approach for operation in a cooperative, active and purpose-driven environment has been considered important not least because simultaneously with its development we actively participated in the effort to build a continuously-operating integrated system [CC95], [MRKI94], [RKMI95]. We regard the efforts to manage the complexity of recognition by exploiting cooperation of multiple cues and by improved understanding of individual ones as complementary and mutually beneficial. Thus we present in chapter 7 a description of a colour-based attention mechanism intended for cooperative operation.

1.3 Outline

We start with a short chapter on performance characterisation of a colour recognition system. In a model-based approach, recognition is achieved by matching pre-stored object descriptions against some representation derived from image measurements. The measurements, in our case the observed colour, depend not only on intrinsic properties of the imaged object, ie. its surface reflectance, but also on a number of scene parameters, eg. viewing geometry and illumination conditions. One of the most important characteristic of any recognition strategy is its ability to cope with changes due to factors not related to object properties. This characteristic, named *scope* by Grimson [Gri90], critically influences the time and space efficiency of localisation and identification. It is therefore necessary to consider the assumptions made about the environment and the variability of experimental conditions before comparing complexity or run-time efficiency. The influence of the size of model database and the properties of the background objects ('image clutter') is also discussed. We conclude the chapter by considering the properties of a recognition system desirable for cooperation in an integrated interpretation system.

Chapter 3 provides a literature survey of several approaches to colour-based object recognition with a focus on methods that exploit colour *structure* of multicoloured objects. We start by presenting a number of algorithms designed for recognition of single-coloured objects. Methods that represent multi-coloured ob-

jects as collections of single-coloured surface patches are included in the same section (section 3.2). Following the general framework laid down in chapter 2 we review in detail histogram-based methods (section 3.3) and a method for matching region adjacency-graphs (section 3.4)

In chapter 4 we present in detail the *colour adjacency graph* (CAG) representation for multicoloured objects, informally introduced in section 1.2. We start by discussing the approximation to surface reflectance adopted in the thesis - the unichromatic reflection model. We then show that the reflection model coupled with a general assumption about properties of noise predicts that observations corresponding to surface patches with identical reflectance form a unimodal cluster in the chromatic plane. Having established the approximate independence of the CAG representation on viewing and illumination geometry and intensity of illumination we turn our attention in sections 4.3.1 and 4.3.2 and in appendix A to implementational issues. In section 4.3.1 and in appendix A we describe two algorithms that implement detection of unimodal clusters, the first step of building the colour adjacency graph. Section 4.3.2 discusses computation of the reflectance ratio.

All three basic categories of recognition problems, ie. localisation ('where'), identification ('what') and interpretation ('what and where') (see chapter 2, page 21) can be solved by the matching of CAGs, which is the subject of chapter 5. In section 5.2 we present a matching algorithm performing a variant of a best-first search, used in recognition experiments reported in chapter 6. A more sophisticated algorithm for CAG matching based on probabilistic relaxation developed by Stoddart et al.[SPK95] is reviewed in section 5.4.

In chapter 6 we present a set of experiments that illustrate the robustness of the CAG recognition method under change of viewpoint, scale and illumination intensity and in the presence of occlusion and non-rigid transformation of shape, ie. most of the environmental changes affecting object appearance listed in table 2.1. We start with a short section (6.2) discussing the design of the model database and the properties of objects included in the database. Images of the model objects and the test images were acquired in laboratory conditions described in section 6.3. The section also discusses specifications of the acquisition equipment used in the experiments. The core of the chapter, section 6.4, describes a sequence of recognition experiments followed by an analysis of the recognition results. In section 6.5 we

discuss the efficiency of the *CAG* matching and compare theoretical predictions with measured execution times. We summarise the experimental results in section 6.6.

In chapter 7 we describe a colour-based attention system successfully tested in a continuously operating, integrated cooperative system [CC95]. In such a system the ability of a recognition engine to *exhaustively* search the image for all instances of known objects is often less important than the capability to focus computational resources to areas of competence, where reliable matching is accomplished at high speed. We present a system whose attention mechanism attempts to avoid analysing the whole image in the spirit of traditional methods; instead it tries to find and select for further processing areas where distinctive colour provides least ambiguous information. From the point of view of colour information processing, the main contribution of the work presented in this chapter lies in the suitability of the proposed approach for environments with *spectrally varying illumination*.

In chapter 8 we first summarise the contributions made in the thesis. We then focus on possible extension and discuss future directions of research.

In Appendix A we present a novel and unifying approach to image segmentation, called *feature and spatial domain* (FSD) clustering . The algorithm is related to the work on colour-based recognition, it can be used for building the *CAG*, but the main contribution of this method lies in its generality. We therefore decided to stress the wide range of possible applications of the FSD CLUSTERING and to move the presentation to an appendix.

The method is devised to group pixel data by taking into account simultaneously both their feature space similarity and spatial coherence. We include examples of gray and colour image segmentation, edge and line detection, range data and motion segmentation. In comparison with existing segmentation approaches, the method can resolve image features even if their distributions significantly overlap in the feature space. It can distinguish between noisy regions and genuine fine texture. Moreover, if required, FSD CLUSTERING can produce partial segmentation by identifying salient regions only.

References

[Alo90] J. (Y.) Aloimonos. Purposive and qualitative active vision. In IUW90 [IUW90], pages 816–828.

[AMA79] M. Ali, W.N. Martin, and J.K. Aggarwal. Color-based computer analysis of aerial photographs. *Computer Graphics and Image Processing*, 9:282–283, 1979.

[Baj88] R. Bajcsy. Active perception. *Proceedings of the IEEE*, 76:996–1005, 1988.

[BB88] A. Blake and G. Brelstaff. Geometry from specularities. In ICCV88 [ICC88], pages 394–403.

[BH86] R. C. Bolles and P. Horaud. A three-dimensional part orientation system. *International Journal of Robotics Research*, 5:3–26, 3 1986.

[Bie85] I. Biederman. Human image understanding: Recent research and a theory. *Computer Vision, Graphics and Image Processing*, 32:29–83, 1 1985.

[Bin82] T. O. Binford. Survey of model-based image analysis systems. *International Journal of Robotics Research*, 1, 1982.

[Bri90] M.H. Brill. Image segmentation by object color. A unifying framework and connection to color constancy. *JOSA-A*, 7:2041–2047, 1990.

[Bro83] R. A. Brooks. Model-based three-dimensional interpretations of two-dimensional images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5:140–150, 2 1983.

[CC95] J Crowley and H I Christensen. *Vision as Process*. Springer-Verlag, Berlin, 1995.

[DI93] Michael D'Zmura and Geoffrey Iverson. Color constancy. I. Basic theory II. Results of two-stage linear recovery of spectral descriptions for lights and surfaces. *JOSA-A*, 10:2148–2180, October 1993.

[EM93] F. Ennesser and G. Medioni. Finding waldo, or focus of attention using local color information. In *CVPR'93 (IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York City, NY, June 15–17, 1993)*, pages 711–712, Washington, DC., Jun 1993. Computer Society Press.

[Fat92] S.T. Fathima. Data and model-driven selection using color regions. In *Image Understanding Workshop (San Diego, CA, Jan 26-29, 1992)*, pages 705–716. Morgan Kaufmann, 1992.

[FB95] W.T. Freeman and D.H. Brainard. Bayesian decision theory, the maximum local mass estimate and color constancy. In ICCV95 [ICC95], pages 210–217.

[FF94] B. V. Funt and G. D. Finlayson. Color constant color indexing. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17:522–529, May 1994.

[Fin94] G. Finlayson. Color constancy and changing illumination. In *SPIE, Feb. 94, San Jose CA*, 1994.

[For88] D.A. Forsyth. A novel approach to colour constancy. In ICCV88 [ICC88], pages 9–18.

[GJT87] R. Gershon, A.D. Jepson, and J.K. Tsotsos. Highlight identification using chromatic information. In ICCV87 [ICC87], pages 161–170.

[Gri90] W. Eric L. Grimson. *Object Recognition*. MIT Press, 1990.

[GROK94] W.E.L. Grimson, A. Lakshmi Ratan, P.A. O'Donnell, and G. Klanderman. An active visual attention system to "play where's waldo". In *DARPA 1994 Workshop*, 1994.

[HB87] G. Healey and T.O. Binford. The role and use of color in a general vision system. In *Image Understanding Workshop (Los Angeles, CA, Feb. 23-25, 1987)*, pages 599–613. Morgan Kaufmann, 1987.

[Hea89] G. Healey. Using color for geometry-insensitive segmentation. *Journal of the Optical Society of America*, 6:86–103, June 1989.

[HR78a] A. R. Hanson and E. M. Riseman. *Computer vision systems.* Academic Press Inc., Florida, 1978.

[HR78b] Allen R. Hanson and Edward M. Riseman. Segmentation of natural scenes. In Hanson78 [HR78a], pages 129–164.

[HR78c] Allen R. Hanson and Edward M. Riseman. Visions: A computer system for interpreting scenes. In Hanson78 [HR78a], pages 303–333.

[ICC87] *First International Conference on Computer Vision, (London, England, June 8–11, 1987)*, Washington, DC., 1987. IEEE Computer Society Press.

[ICC88] *Second International Conference on Computer Vision (Tampa, FL, December 5–8, 1988)*, Washington, DC., 1988. Computer Society Press.

[ICC95] *Fifth International Conference on Computer Vision (Cambridge, MA, June 20–23, 1995)*, Washington, DC., 1995. Computer Society Press.

[IUW90] *Image Understanding Workshop (Pittsburgh, Penn., Sep 11-13 1990).* Morgan Kaufmann, 1990.

[Kli93] Gudrun J. Klinker. *A Physical Approach to Color Image Understanding.* A K Peters, Wellesley, Massachusetss, 1993.

[KSK87] G.J. Klinker, S.A. Shafer, and T. Kanade. Using a color reflection model to separate highlights from object color. In ICCV87 [ICC87], pages 145–150.

[KSK88] G.J. Klinker, S.A. Shafer, and T. Kanade. Image segmentation and reflection analysis through color. In *Image Understanding Workshop (Cambridge MA, April 6-8, 1988)*, pages 838–853. Morgan Kaufmann, 1988.

[LB90] S.W. Lee and R. Bajcsy. Detection of specularity using color and multiple views. In *Second European Conference on Computer Vision (Santa Margherita Ligure, Italy, May 19-22, 1992)*, pages 99–114. Springer-Verlag, 1990.

[LEC90] H. Lee, E.J.Breneman, and C.P.Schulte. *Modeling Light Reflection for Computer Color Vision*, pages 3–10. IEEE Computer Society Press, 1990.

[Low87] D. G. Lowe. Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, 31:35–395, 1987.

[Mal86] L.T. Maloney. Evaluation of linear models of surface spectral reflectance with small number of parameters. *J. Opt. Soc. Am.*, 3:1673–1686, Octobe 1986.

[MB86] L.T. Maloney and B.A.Wandell. Color constancy: a method for recovering surface spectral reflectance. *Journal of the Optical Society of America*, 3:29–33, January 1986.

[MMK95a] J. Matas, R. Marik, and J. Kittler. The color adjacency graph representation of multi-coloured objects. Technical report, University of Surrey VSSP-TR-1/95, 1995.

[MMK95b] J. Matas, R. Marik, and J. Kittler. On representation and matching of multi-coloured objects. In ICCV95 [ICC95], pages 726–732.

[MRKI94] J. Matas, P Remagnino, J Kittler, and J Illingworth. Control of scene interpretation. In *Vision as Process*, pages 347–373. Springer-Verlag, 1994.

[NB93] Shree K. Nayar and Ruud M. Bolle. Reflectance ratio: A photometric invariant for object recognition. In *Fourth International Conference on Computer Vision (Berlin, Germany, May 11-14, 1993)*, pages 280–285, Washington, DC., 1993. IEEE Computer Society Press.

[NM80] Makoto Nagao and Takashi Matsuyama. *A Structural Analysis of Complex Aerial Photographs*. Plenum Press, New York and London, 1980.

[OKS80] Y. Ohta, T. Kanade, and T. Sakai. Color information for region segmentation. *Computer Graphics and Image Processing*, 13:222–241, 1980.

[OPR78] R. Ohlander, K. Price, and D. R. Reddy. Picture segmentation using a recursive region splitting algorithm. *Computer Graphics and Image Processing*, 8:313–333, 1978.

[RKMI95] P Remagnino, J Kittler, J Matas, and J Illingworth. Camera control for establishing the current and next-look direction in an active vision object recognition system. In *Vision as Process*, pages 403–417, 1995.

[SB90] M.J. Swain and D.H. Ballard. Indexing via color histograms. In *Third International Conference on Computer Vision (Osaka, Japan, December 4–7, 1990)*, pages 390–393, Washington, DC., 1990. IEEE Computer Society Press.

[Sha84] S.A. Shafer. Using color to separate reflection components. Technical report, University of Rochester, 1984.

[SPK95] A. Stoddart, M. Petrou, and J. Kittler. Probabilistic relaxation as an optimizer. In *British Machine Vision Conference, Birmingham*, 1995.

[SS94] M. Stricker and M. Swain. The capacity of color histogram indexing. In *CVPR'94 (IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, 1994)*, pages 704–708, Washington, DC., Jun 1994. Computer Society Press.

[Swa90] M.J. Swain. *Color Indexing*. PhD thesis, University of Rochester, 1990.

[Tom93] S. Tominaga. Analysis of color images for inhomogeneous objects with highlights and interreflections. In *ACCV 93: Asian Conference on Computer Vision, Osaka, Japan*, pages 167–170, 1993.

[Ull86] Shimon Ullman. Artificial intelligence and the brain: Computational studies of the visual system. *Annual Review of Neuroscience*, 9:1–26, 1986.

[Wix90a] Lambert E. Wixson. Exploiting world structure to efficiently search for objects. Technical report, U of Rochester, 1990.

[Wix90b] L.E. Wixson. Real-time qualitative detection of multi-colored objects for object search. In IUW90 [IUW90], pages 631–638.

[ZGB99] A. Zisserman, P. Giblin, and A. Blake. The information available to a moving observer form specularities. In C. Brown, editor, *Advances in Computer Vision (2 vols.)*, pages 167–172, Hillsdale, NJ, 1999. Erlbaum.

Chapter 2

Performance Characterisation

In model-based recognition, localisation and identification of instances of objects are achieved by matching pre-stored object description ('the model') with some quantities derived from image measurements ('the features'). These measurements, in our case the observed colour, depend not only on the properties of the imaged object, ie. its surface reflectance, but also on a number of scene parameters, eg. viewing geometry and illumination conditions. In order to limit the complexity of matching and to make experimentation feasible, researchers have had to design recognition systems assuming that they would operate in a restricted environment where some of the parameters influencing object appearance are constant and (or) known and where phenomena like occlusion cannot occur.

In a book on object recognition [Gri90], Grimson presents four criteria by which to judge methods for solving the localisation and identification problem: efficiency, correctness, robustness and scope. In Grimson's definition ([Gri90], page 401), correctness is measured by the error rate of the system [DH73] and efficiency is expressed both in terms of formal complexity and run-time performance. Robustness characterises the response of a method to noise in the sensory data; the noise being either in the form of reduction of the amount of relevant data (eg. due to occlusion), increase in irrelevant data (due to background clutter) or simply in the form of statistical measurement noise. Finally, the scope is defined as the set of circumstances under which the method meets the criteria of efficiency, correctness and robustness.

Clearly, efficiency, correctness and robustness depend critically on the amount of variability permitted in the observed world. Given the dependency, investigation of the *scope*, or its complement – the restrictions that must be imposed on environment

1. **Viewpoint** transformations. Classified as
   (a) rigid
   (b) scaling
   (c) affine
   (d) perspective
   (e) changing aspect
2. **Illumination.** Change of/due to
   (a) intensity
   (b) spectral power distribution
   (c) position of light sources
   (d) shadows and mutual illumination
3. **Other.**
   (a) non-rigid change of shape
   (b) changes in acquisition chain

Table 2.1: Factors influencing sensed object colour.

to obtain good performance – must necessarily precede any attempt to evaluate and compare the other three performance characteristics of a recognition system. Often differences in reported efficiency and correctness of colour recognition methods can be ascribed to (unreported) difference in scope (generality) of testing conditions rather then to any inherent advantages of a particular algorithm. In order to maintain consistency of evaluation of the reviewed methods, we compiled a list of common factors that influence observed colour and hence the complexity of colour matching (see table 2.1).

The parameters shown under items 'Illumination' and 'Viewpoint' are directly related to quantities that appear in equations 4.1 and 4.2 which link intrinsic object property – surface reflectance – with sensor responses. The entries in the 'Illumination' section are independent parameters of scene illumination. The character of the items of the 'viewpoint' group is different; viewpoint transformations are classified

according to the effects on the image, ranging from the simple rigid transformation caused eg. by rotation of the camera around its optical axis to a transformation with full perspective effects. In general, a change of viewing position will change the part of surface of the object that is visible. The aspect changes if the topology of surface patches is altered as a result of movement.

Besides the factors that influence object appearance, the complexity of recognition depends on the context in which it is carried out, namely on the number of models in the database, properties of unknown objects present in the scene – 'the background clutter', and the absence of expected features due to occlusion. Methods also differ in their suitability for the following categories of recognition problems [SB90]:

1. Localisation of given objects ('Where is X?')
2. Identification of a set of selected features ('What is X?')
3. Interpretation ('What is in the image? Where? ')

In the paragraphs above we discuss a number of factors that influence a recognition system. A comparison of recognition methods is a complex undertaking and we are not finished yet! So far we have considered circumstances affecting a system that is operating in *isolation*. We were concerned how to measure, at least qualitatively, the loss of generality of a system caused by adopting the 'Constrain your World' strategy. In a recent shift of perspective, the proponents of *active vision* have argued [BY93] that a number of vision related tasks are greatly simplified if the imaging sensor is controlled purposefully to facilitate a given task. Others have studied the benefits of exploiting temporal context for recognition in a *continuously* operating system [CC95].

Cooperation strategy provides another alternative to the set of environments proposed to enlarge the set of conditions in which a recognition system is capable of successful operation. It is not difficult to imagine situations where colour provides an easy way to localise an object, eg. when looking for a red football on a grassy pitch. It is equally easy to see it is of no value for sorting castings of different shape. The essence of this approach is expressed in Garvey's statement [Gar76]:

This system is based on the philosophy that there normally exist easy ways of finding objects.

At a first glance it may seem that the issues related to active, purposive, continuous and cooperative vision have no bearing on the design of a recognition method. Once techniques are developed that successfully operate in a specific niche then only a simple supervising mechanism is needed to control the incorporated recognition modules. While working on a four-year project [CC95] aimed at building a continuously operating system exploiting cooperation of multiple recognition engines [MRKI94] we realised that individual recognition engines have to possess the following two properties if the integrated system is to benefit from cooperation. First, every recognition engine must be able to focus its resources on the part of the scene where it is most competent. Secondly, a confidence estimate must accompany results to enable combination of evidence. Other desirable features include the ability to respond to changes in the system goal [KMBN95].

The last characteristic of a recognition system we will mention, the complexity of model acquisition, is not directly related to performance. But the practicality of any method is undoubtedly reduced if models are not created (semi-) automatically especially if the model acquisition process requires a significant amount of human intervention.

To summarise, we will be looking at four sets of properties characterising a colour recognition method: its scope, its capability to operate in different context, the suitability of the method for operation in a cooperative system and the complexity of the off-line preparation of models.

References

[BY93] Andrew Blake and Alan Yuille. *Active Vision.* MIT Press, Cambridge, Massachusetts, 1993.

[CC95] J Crowley and H I Christensen. *Vision as Process.* Springer-Verlag, Berlin, 1995.

[DH73] R. E. Duda and P. E. Hart. *Pattern Classification and Scene Analysis.* John Wiley, 1973.

[Gar76] T.D. Garvey. An experiment with a system for locating objects in multisensory images. In *Unknown*, pages 319–325, 1976.

[Gri90] W. Eric L. Grimson. *Object Recognition*. MIT Press, 1990.

[KMBN95] J. Kittler, J. Matas, M. Bober, and L. Nguyen. Image interpretation: Exploiting multiple cues. In *International Conference on Image Processing and Applications, Edinburgh, July 3-6*, pages 1–5, 1995.

[MRKI94] J. Matas, P Remagnino, J Kittler, and J Illingworth. Control of scene interpretation. In *Vision as Process*, pages 347–373. Springer-Verlag, 1994.

[SB90] M.J. Swain and D.H. Ballard. Indexing via color histograms. In *Third International Conference on Computer Vision (Osaka, Japan, December 4–7, 1990)*, pages 390–393, Washington, DC., 1990. IEEE Computer Society Press.

Chapter 3

Colour-based Recognition: A Survey

3.1 Introduction

This chapter provides a literature survey of several approaches to colour-based object recognition with a focus on methods that exploit colour *structure* of multicoloured objects. We start by presenting a number of algorithms designed for recognition of single-coloured objects. Methods that represent multi-coloured objects as collections of single-coloured surface patches are included in the same section (section 3.2). Such methods, using neither information about adjacency of colour patches, nor properties of colour boundaries or information about colour composition of the whole surface of the objects, differ insignificantly from the methods for recognition of single-coloured objects by considering localisation and matching of patches rather than objects.

For reasons discussed in section 1.1, only a limited amount of work has been reported on recognition of objects with complex colour structure. Following the general framework laid down in chapter 2 we review the histogram-based methods (section 3.3) and a method matching region adjacency-graphs (section 3.4) in detail. Both the colour histogram and the regions adjacency graph are related to the colour adjacency graph representation presented in chapter 4 where, in section 4.1 we discuss the links between the two representations.

We decided not to include in the overview references to work on classification of multi-spectral images, a topic that has been extensively studied in the fields of image interpretation and remote sensing [Ric93] [LK94]. The interpretation techniques de-

veloped for remote sensing applications and colour recognition methods for robotic applications share the same type of input - spatially indexed multi-dimensional measurements, but the phenomena influencing appearance of the objects of interest differ significantly. In remote sensing the measurements are influenced by properties of the atmosphere (eg. haze), a factor with negligible effect in laboratory and outdoor environments. On the other hand, remote sensing images have very little depth relative to the distance from the sensor. Moreover, the position (or at least the orientation) of the sensor with respect to objects on the surface of earth is known. The relative position to the predominant source of illumination, the Sun, and its spectral properties are also known. The parts of objects imaged from different viewpoints are identical or have a large overlap and mutual occlusion is a rare problem. As a consequence of the differences, remote sensing techniques are not directly applicable to the 'terrestrial' colour recognition problem. We refer the interested reader to [Ric93].

3.2 Recognition of Single-Coloured Objects

Early approaches to colour recognition were based on standard pattern recognition techniques. Areas corresponding to objects in the model database were detected by per-pixel classification followed by connected component analysis, possibly interleaved with image processing operations (filtering, morphology). Image formation process was not explicitly modelled. Methods varied in the choice of colour space or the type of decision surfaces used in the classifier. For instance, Garvey [Gar76] used the planes perpendicular to the axis of the hue-saturation-brightness space for decision-making [1]. Ali [AMA79] transformed *rgb* values into $r/(r+g+b)$, $g/(r+g+b)$ and $r+g+b$ [2] and used non-linear decision surfaces in the classification stage.

Methods with similar structure for processing of colour information are commonly used in applications. In [FV92], Ferri describes an algorithm designed for a "Robotic Citric Harvester". The *rgb* measurements are linearly transformed into the *yuv* space. At each pixel the u and v values, together with features computed in

[1] It is fair to note that Garvey was interested primarily in integration of recognition cues, rather than in colour-based recognition.

[2] In fact, we use an identical transformation for the CAG representation (see chapter 4, equations 4.7.

the local neighbourhood, are classified by the nearest neighbour rule. Each class is represented by a number of automatically selected prototypes, allowing effectively for an arbitrary decision surface.

A system for colour-based road following for vehicle guidance is described in [Cri93]. Predictions from previous frames are used to select regions that belong with high probability to the road and non-road classes. Prototypes for a nearest-neighbourhood classification are extracted by clustering the data from the regions. Subsequently, conditional probability of a pixel belonging to the road class is computed. The final decision on the location of the road is obtained by fitting the likelihood surface by geometric models for a straight road, bend, intersection etc. These two examples demonstrate the evolution of the classify and group paradigm to colour interpretation. In the classification stage, arbitrary decision surfaces replaced the constrained planar or smooth non-linear boundaries. In the region extraction stage the use of the likelihood function allowed to postpone the final interpretation decision until context has been taken into account.

In [MMK93] we presented a colour-based system for generation, verification and localisation of the objects centred around the idea of avoiding the 'hard' classification decision. Instead, at every pixel the aposteriori probability for each single-coloured object in the model database is computed. A mean and covariance matrix in the chromatic plane serves as a model. The algorithm then *focuses on the pixels of least ambiguity* of interpretation, ie. pixels with the highest aposteriori probability for the most likely model. Regions are defined as areas of identical most likely labels. The method is suited for cooperative operation as it focuses computational resources and generates hypotheses about the presence of objects in the regions where colour information is providing the strongest index into the model base. The same algorithm is employed for localisation of a given object; in this case the regions with maximum aposteriori probability to the object of interest are output. In [MRKI94] we extended the approach to environments with changing illumination colour. It is assumed that the resulting spectral distribution of illumination is a mixture of light from only a limited number of light sources and that chromatic models for the different illuminants are known. The modified method attempts to find regions that have unique interpretation, regardless of the current spectral distribution of illumination. Once such regions are found, the observed colour of the recognised

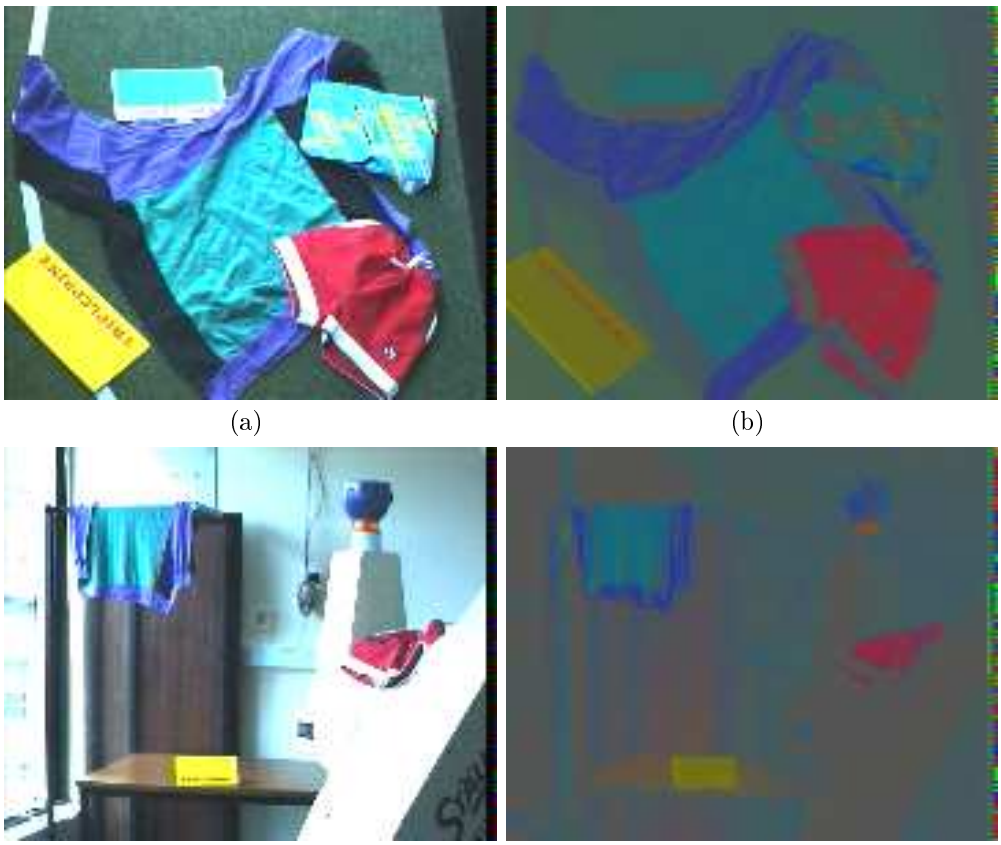
objects constrain the range of possible illumination spectral distributions.

In [GROK94], Grimson et al. describe a system which integrates cues from colour and stereo to fixate candidate regions in which objects are recognised by shape-based alignment. The processing of colour information is straightforward, as it is intended for fast figure-ground separation. Each object model contains a list of component colours. Each component is represented by a mean and a covariance matrix in the *hsv* space. Image pixels with hue and saturation within three standard deviations from the mean are selected as potentially corresponding to the object. Intensity is not used, quoting: "since intensity values can change drastically with lighting conditions". Shape of connected components of selected pixels is approximated by an ellipse. Lines inside the elliptical regions are matched to models by the method of alignment. The process is repeated in turn for every object in the model database.

Although there are significant differences in the way colour information is represented and processed in the above mentioned systems, they all share a number of common characteristics. Firstly, it is the simplicity of colour representation coupled with basic and well understood pattern recognition classification algorithms. These properties facilitating fast and reliable implementation are primarily responsible for the popularity of the above mentioned methods. Moreover, model acquisition is trivial and incorporation into a large cooperative system (eg. [GROK94]) poses little problems.

An inherent non-negligible drawback of the 'list-of-colour-patches' methods stems from their inability to capture the invariant information computable from ratios of sensor responses from the boundaries of patches with different surface reflectances, as proposed by Nayar [NB94] and Finlayson [FF94].

The effect is visualised in figure 3.1. Images 3.1a and 3.1c show colour originals used in experiments described in chapter 7. In order to achieve independence on geometry and intensity of illumination under the assumption of the simplest reflection model (ie. unichromatic model, equation 4.2), *rgb* measurements in the methods described above are projected into a two-dimensional subspace, eg. a chromatic plane or a unit sphere, losing the informative and invariant boundary ratios. As a result, a 'list-of-colour-patches' method then 'sees' the impoverished chromatic images as shown in 3.1b and 3.1d respectively.



(a) (b)

(c) (d)

Figure 3.1: Colour originals (a)(b) and corresponding purely chromatic images (c)(d).

3.3 Histogram Based Methods

All the methods presented in the previous section try to localise and match regions corresponding to objects (or parts of objects) of *single* colour. However, the technique starts to fail for even a modestly sized model base. As consequence of the spread of measurements due to effects like noise, specularities, interreflections etc., overlaps in colour space become unavoidable. In contrast, co-occurence of a number of colours, or even two colours, of a multi-coloured object in a neighbourhood significantly reduces the probability of multiple interpretations.

In his doctoral thesis [Swa90], Swain proposed two methods for recognition based on colour structure; *histogram intersection* for object identification and *histogram backprojection* for localisation. Both methods represent colour structure straightfor-

wardly, by a $16 \times 16 \times 8$ 3D histogram in the 'opponent' colour space:

$$\begin{array}{rcl} rg & = & red - green \\ by & = & 2 * blue - red - green \\ wb & = & red + green + blue \end{array} \tag{3.1}$$

Swain justifies the choice of the colour space in the following way: " (the axes) were used here simply to allow the intensity (wb) axis to be more coarsely sampled than the other two, because the intensity axis is more sensitive to lighting variation from shadows and distance from the light source". However, this in general is not true. Viewing geometry changes and effects like increase in illumination intensity or a change in the acquisition equipment (change of focal length, aperture, camera gain) on the measured rgb values are well modelled under very general assumptions[3] by a multiplication by a scalar k.

Clearly, as a consequence of the linearity of equation 3.1, the opponent values rg, by, wb will undergo an identical change - scaling. For different values of k corresponding to various conditions the measured rg, by, wb lie on a straight line. Only unsaturated colours close to white, and therefore almost parallel to the intensity wb, gain a degree of insensitivity by the reduced resolution of the wb axis. But shades of white are rarely the distinguishing features of an object, and saturated colours do not gain from the transformation.

Swain's colour histograms are therefore sensitive to changes in illumination intensity, position of light source, angle of the viewed surface with respect to the observer as well as any change in the acquisition chain - in brief, any change in the environment affecting the magnitude of measurements.

So far we have described *global* effects that can be, at least in principle, compensated for by normalisation techniques like histogram equalisation. But consider the effects of moving a light source around the object. Different parts of the surface will be illuminated by light of changing intensity that will *locally* deform the shape of the histogram.

The counts in the bins of a colour histogram are proportional to the projected area of the object's surface with colour falling in each given bin. What transformation of viewpoint (or, equivalently, object pose) are the projected areas invariant

[3]The model holds if the acquisition chain is linear. No assumptions about surface reflection besides linearity are necessary.

to? The answer is not encouraging: *translation parallel to image plane and rotation around the optical axis of a camera.* If the histogram is normalised, then bin counts represent relative projected areas. The ratio of areas is invariant under affine transformation of a plane [JA92]. We conclude that for *planar* objects the normalised colour histogram is invariant under affine transform. For objects of arbitrary shape the system is effectively two dimensional.

Other factors influencing object appearance (table 2.1) not yet discussed include change of aspect, occlusion and a non-rigid change of shape, none of which is treated in detail in Swain's work. All objects, but one, are represented by a histogram taken from a single viewpoint. This is certainly not generally sufficient. Consider a cube with pairs of opposite sides coloured blue, red and green. *Any* proportion of red, blue and green may appear in the histogram depending on the viewpoint. The non-trivial problem of selection of representative viewpoints approximating any possible histogram, clearly necessary if histograms are to be used in a truly three dimensional setting, is analysed neither in Swain's nor in any consecutive work.

Although a histogram, being a global description of an object, is sensitive to occlusion, the degradation of match is graceful as shown by Swain. The last factor of table 2.1 not yet mentioned, change of spectral distribution of illuminations, is proposed to be eliminated by using histograms of images processed by a colour constancy algorithm. Taking into account recent advances in the area [FF94, FB95] such approach seems sensible.

In the framework developed by Swain, *identification* of objects is achieved by histogram intersection. In this method the quality of a match between image histogram I and model histogram M, each containing n bins, is defined as

$$\frac{\sum_{j=1}^{n} min(I_j, M_j)}{\sum_{j=1}^{n} M_j} \tag{3.2}$$

In the case of *normalised* histograms, histogram intersection has several desirable properties. It is equivalent to a city block metric in the n-dimensional space. More importantly, it provides an estimate of the *Kolmogorov variational distance* [Vaj70], a fundamental measure of a difference of two distributions used in pattern recognition [Kit75].

From the discussion above we see that colour histograms are influenced by almost any change in the viewing or illumination conditions and changes in the acquisition

chain; nor does the histogram intersection match features insensitive to the effects. It is therefore not surprising that the set of images used for recognition contains scenes with only a *single* object on black background, viewed as in the model images directly from above and from the same distance. The transformation between the model and the test image is almost Euclidean. A single experiment with a slight 3D rotation (of about 15 degrees) was performed on a Snoopy doll.

The method for localisation, *histogram backprojection*, suffers from similar limitations. It starts by computing the ratio histogram:

$$R_i = \frac{M_i}{I_i} \tag{3.3}$$

The probability of presence of object M at a given pixel is estimated by the bin count of the ratio histogram that the pixel colour indexes into. A spatial average over a circular neighbourhood is used as the final estimate of the probability of model presence. The method requires a priori knowledge of the size of the neighbourhood for averaging. Moreover, it cannot work for extended non-compact objects as they would not record a high average probability in any *circular neighbourhood*.

The histogram backprojection can be viewed as a simplified version of Bayesian estimation. If quantity I_i is interpreted as an estimate of probability $P(i)$ of an occurence of colour i in an image and M_i is interpreted as an estimate of probability $P(i|object)$ of an occurence of colour i given that the pixel corresponds to the object,

$$M_i \approx P(i|object) \tag{3.4}$$

$$I_i \approx P(i) \tag{3.5}$$

then histogram backprojection ratio R_i becomes proportional to the aposteriori probability of colour i belonging to object, if equal a priori probability for every object is assumed. Invoking Bayes rule and substituting from equations 3.4 and 3.5 we obtain:

$$P(object|i) = \frac{P(i|object)P(object)}{P(i)} \tag{3.6}$$

$$\approx P(object)R_i \tag{3.7}$$

The factor $P(object)$ has no effect, because the histogram backprojection only selects point that maximise R_i and the selection process is not affected by scaling. We have

now touched upon another problem - histogram backprojection selects most likely locations for a given object, but no rule is given for the detection of false positives (ie. what level of $P(object|i)$ is significant). Usually, the probability of a given colour i is established through the rule of total probability:

$$P(i) = \sum_j P(i|object_j)P(object_j) + P(i|background)P(background) \qquad (3.8)$$

In this case the strength with which a given colour indicates the presence of a particular object depends on the frequency of the colour in the model database and in the environment[4]. In equation 3.3 other models are not taken into account, unless they appear in the processed image. In his thesis, Swain presents only a single image on which the histogram backprojection was tested - it is a collage of all the images used to construct the model base. For this particular image, the image histogram I_i is identical to the $P(i)$ computed as shown in equation 3.8!

Other recently reported experiments using histogram intersection and backprojection [Str92] [SS94] [EM93] were conducted on 2D data.

Stricker and Swain [SS94] studied the indexing capacity of a histogram. The probability of a false match was experimentally tested and the selection of an optimal threshold discussed. The experiments were conducted on images drawn from a large database (approximately five hundred images) maintained by the Smithsonian Institute. The work only compares histograms of unrelated images, not histograms of scenes and objects taken from different viewpoints or under different illumination conditions.

Ennesser and Medioni [EM93] modified the histogram intersection algorithm by inserting a weighting function into the sum of equation 3.2. The function is chosen to amplify the influence of colours specific to the model. In Swain's original method, model histogram is matched with a global colour histogram computed over the whole image. In Ennesser's approach, a dense map of local histograms is computed and matched with the model histogram. The method therefore simultaneously identifies and localises an instance of the model. A rescaling scheme is proposed which makes the algorithm applicable even in the absence of prior knowledge about the scale (size) of the localised object. The method seems to be aimed at 2D applications and the presented experiments are carried out completely in 2D.

[4] estimation of $P(i|background)$ being the main difficulty of applying the rule

In [Wix90] Wixson states that the dependence on scale of histogram intersection made the method ill-suited for *object search*. With this purpose in mind, Wixson proposed a different matching scheme based on colour histograms . Firstly, Wixson uses only 2D histograms (resolution 16x16) obtained from Swain's 3D histograms by summing along the *wb* axis. Then a *signature* of a histogram was computed from local maxima in the 2D histogram with counts above a threshold. The signature is the set of ratios of bin counts obtained by considering all pairs of local maxima. The goodness of match is expressed in terms of a comparatively complex cost function which attains a high value when similar ratios of local maxima with similar 2D colour indices are found. The scheme relies both on geometric and colour information and the reported results suggest it is more robust than Swain's original methods. The ratios stored in the signature of Wixson's method are only invariant under affine transformation of a plane and the chosen 2D space representation of colour is sensitive to the same effects mentioned in connection with Swain's method - change of intensity of illumination, different aperture setting and change of relative position to the light source[5]. It seems that the improved suitability of the method for operation in truly 3D environments is due partially to the fact that the method does not rely on absolute bin counts. Additional stability may stem from the structure of the matching algorithm; the local maxima in the colour histogram are certainly more stable then the associated counts.

In [HS94], Healey describes a system that uses illumination invariant descriptors of a colour histogram to index into a database. It is shown that if surface reflectances are approximated by three basis functions then an *arbitrary* change in illumination intensity and spectral power distribution leads to a linear transformation of measurements. Image histogram is then described by affine invariant functions of moments, computed by a method developed by Taubin and Cooper ([JA92], pages 297 – 375).

The method has several limitations. As mentioned in the paper, histograms are computed from the whole image, and therefore the object must be *segmented* and *unoccluded*. Other problems arise from the global nature of the descriptors based on moments - any *local* change in relative intensity affects only a part of measurements and is not described by an affine transformation of the colour space. Changes in

[5]This problem is not inherent to the method and the dependency could have been avoided by using a different 2D representation of colour, eg. hue and saturation or chromaticity values.

scene geometry and viewpoint are not considered. The test database contained nine objects.

In [FF94] Funt and Finlayson propose a method where, instead of direct *rgb* measurements, ratios of responses at neighbouring pixels are histogrammed. A theoretical analysis as well as experimental results show that the ratio histogram is much less sensitive to changes of spectral properties of illumination.

On the whole, the histogram based methods have not yet been shown to operate successfully in truly three-dimensional and unconstrained environments. The issue of the dependency of a colour histogram on the relative pose of an object and the related problem of establishing the number of histograms required to cover all viewpoints remains open. It is our opinion that this does not reflect the full potential of histogram based methods, but is rather an artifact of either the application for which the described method was intended or constraints on implementation, eg. in the case where real-time operation took precedence over generality of the algorithm.

3.4 Region Adjacency Graph

In [Fat92][SM92], Fathima presents a colour recognition method based on the matching of attributed region adjacency graphs. Each object is modelled by graph M_G,

$$M_G = \{V_m, E_m, C_m, R_m, S_m, B_{rm}, B_{sm}\}$$

where V_m is the set of regions in the model, E_m set of adjacencies between colour regions, C_m is the colour of region $u \epsilon V_m$, $R_m(u, v)$ is the relative size of region v with respect to region u. $S_m(u)$ is the size of region u, B_{rm} is a bound on the relative size of regions given by R_m and B_{sm} is a bound on the absolute size of regions given by S_m. The colour region information in an image is represented similarly by the image region adjacency graph:

$$I_G = \{V_i, E_i, C_i, R_i, S_i\}$$

If a given object is present in the scene, a subgraph of I_G will belong to a subgraph of M_G. A branch and bound version of the interpretation tree search [Gri90] is employed to search for a subgraph maximising a complex cost function [6].

[6] For details of the cost function see [SM92], page 120.

Although the region adjacency contains significant information about the colour structure of an object, its attributes are not invariant to a general change of viewpoint (or pose). Absolute size S is invariant only under 2D rotation and translation, relative sizes R are invariant under affine transform, but only if the two regions lie on a flat surface. The limitation might not be serious, because the stored bounds on the relative sizes implicitly hold information about planarity of different parts of object, staying stable for adjacent regions on the same planar patch and having a large range otherwise. From this point it is interesting to note that in the only recognition example shown in [SM92], the pose of the recognised object is almost identical in both the test and model acquisition images.

The practicality of the method is limited by the computational complexity of graph matching and the size of region adjacency graphs even for scenes of moderate complexity (see 6.2 and in scenes with objects with a number of coloured regions, eg. the proceedings shown in 1.3. A further complication of the problem stems from the need to store and match RAGs for a (often large) number of aspects, a problem not treated in [SM92].

References

[AMA79] M. Ali, W.N. Martin, and J.K. Aggarwal. Color-based computer analysis of aerial photographs. *Computer Graphics and Image Processing*, 9:282–283, 1979.

[Cri93] Jill D. Crisman. Color region tracking for vehicle guidance. In *Active Vision*, pages 107–120, Cambridge, Massachusetts, 1993. MIT Press.

[CVP94] *CVPR'94 (IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, 1994)*, Washington, DC., Jun 1994. Computer Society Press.

[EM93] F. Ennesser and G. Medioni. Finding waldo, or focus of attention using local color information. In *CVPR'93 (IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York City, NY, June 15–17, 1993)*, pages 711–712, Washington, DC., Jun 1993. Computer Society Press.

[Fat92] S.T. Fathima. Data and model-driven selection using color regions. In *Image Understanding Workshop (San Diego, CA, Jan 26-29, 1992)*, pages 705–716. Morgan Kaufmann, 1992.

[FB95] W.T. Freeman and D.H. Brainard. Bayesian decision theory, the maximum local mass estimate and color constancy. In *Fifth International Conference on Computer Vision (Cambridge, MA, June 20–23, 1995)*, pages 210–217, Washington, DC., 1995. Computer Society Press.

[FF94] B. V. Funt and G. D. Finlayson. Color constant color indexing. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17:522–529, May 1994.

[FV92] F. Ferri and E. Vidal. Colour image segmentation and labeling through multiedit-condensing. *Pattern Recognition Letters*, 13:561–568, August 92.

[Gar76] T.D. Garvey. An experiment with a system for locating objects in multisensory images. In *Unknown*, pages 319–325, 1976.

[Gri90] W. Eric L. Grimson. *Object Recognition.* MIT Press, 1990.

[GROK94] W.E.L. Grimson, A. Lakshmi Ratan, P.A. O'Donnell, and G. Klanderman. An active visual attention system to "play where's waldo". In *DARPA 1994 Workshop*, 1994.

[HS94] G. Healey and D. Slater. Using illumination invariant color histogram descriptors for recognition. In CVPR94 [CVP94].

[JA92] J.L.Mundy and A.Zisserman. *Geometric Invariance in Computer Vision.* MIT Press, 1992.

[Kit75] J. Kittler. Mathematical methods of feature selection in pattern recognition. *International Journal of Man-Machine Systems*, 7:609–637, 1975.

[LK94] T. M. Lillesand and R. L. Kiefer. *Remote Sensing and Image interpretation.* John Wiley, 1994.

[MMK93] J. Matas, R. Marik, and J. Kittler. Generation, verification and localisation of object hypotheses based on colour. In J. Illingworth, editor, *British Machine Vision Conference*, pages 539–548. BMVA Press, 1993.

[MRKI94] J. Matas, P Remagnino, J Kittler, and J Illingworth. Control of scene interpretation. In *Vision as Process*, pages 347–373. Springer-Verlag, 1994.

[NB94] Shree K. Nayar and Ruud M. Bolle. Reflectance ratios: An extension of land's retinex theory. In E.S. Gelsema and L.N. Kanal, editors, *Patttern Recognition in practice IV: Multiple Paradigms, Comparative Studies and Hybrid Systems*, pages 141–152. Elsevier, 1994.

[Ric93] John A. Richards. *Remote Sensing Digital Image Analysis.* Springer, 1993.

[SM92] Tanveer Fathima Syeda-Mahmood. Data and model-driven selection using color. In *Second European Conference on Computer Vision (Santa Margherita Ligure, Italy, May 19-22, 1992)*, pages i15–123. Springer-Verlag, 1992.

[SS94] M. Stricker and M. Swain. The capacity of color histogram indexing. In CVPR94 [CVP94], pages 704–708.

[Str92] Stricker92. Color and geometry as cues for indexing. Technical report, University of Chicago, 1992.

[Swa90] M.J. Swain. *Color Indexing.* PhD thesis, University of Rochester, 1990.

[Vaj70] I. Vajda. Note on discrimination information and variation. *IEEE Trans. Information Theory*, 16:771–, 1970.

[Wix90] L.E. Wixson. Real-time qualitative detection of multi-colored objects for object search. In *Image Understanding Workshop (Pittsburgh, Penn., Sep 11-13 1990)*, pages 631–638. Morgan Kaufmann, 1990.

Chapter 4

The Colour Adjacency Graph

4.1 Introduction

In this chapter we present in detail the *colour adjacency graph* (CAG) representation for multicolour objects, informally introduced in section 1.2. We start by discussing the approximation to surface reflectance adopted in the thesis - the unichromatic reflection model. We then show that the reflection model coupled with a general assumption about properties of noise predicts that observations corresponding to surface patches with identical reflectance form a unimodal cluster in the chromatic plane. In the CAG, each colour component - the set of points forming a unimodal cluster - is represented by a node, attributed by the chromatic coordinates of the *mode* (local maximum) of the unimodal cluster. The main purpose of section 4.2 is to establish that the attributes of a CAG node are, under the assumed reflection and noise models, independent of viewing geometry and illumination intensity. Edges of the CAG represent adjacency (in the image domain) of colour components. The analysis showing that the edges attribute, the reflectance ratio, is a photometric invariant can be found in [NB94].

Having established the approximate independence of the CAG representation on viewing and illumination geometry and intensity of illumination, we turn our attention in sections 4.3.1 and 4.3.2 and in appendix A to implementational issues. In section 4.3.1 and in appendix A we describe two algorithms that implement detection of unimodal clusters, the first step of building the colour adjacency graph. The graph-theoretical clustering (GT CLUSTERING) [KNF76], detailed in section 4.3.1, is an efficient partitioning scheme akin to a morphological watershed operation

carried out in the feature space. The computational complexity of the process is linear in the number of clustered measurements. The method operates entirely in the feature space (ie. the chromatic space in our particular case); its speed is attractive, but ignoring the spatial structure means that the algorithm cannot correctly separate clusters that overlap in the feature space.

The problem is alleviated by the simultaneous feature and spatial domain clustering (FSD CLUSTERING) described in appendix A. The applications of FSD CLUSTERING to a number of image processing problems, including range data segmentation, optical flow segmentation, line detection and grey-level segmentation have been reported in [MK95]. In spite of the advantages of FSD CLUSTERING over GT CLUSTERING stemming from its use of spatial information, most of the experimental results reported in the thesis have been obtained with GT CLUSTERING for the following two reasons. Firstly, the implementation of FSD CLUSTERING had not been finished at the time when most of the experiments were conducted. Secondly, objects with distinctive colours were used in experiments; in this case the simple feature space segmentation was sufficiently accurate and its efficiency advantage allowed for faster experimentation. The rather involved description of FSD CLUSTERING was moved to the appendix to avoid interrupting the presentation of the implementation of *CAG*. It is not necessary for the exposition; from the point of view of the *CAG* representation, outputs of the GT CLUSTERING and FSD CLUSTERING are identical.

Section 4.3.2 discusses computation of the reflectance ratio. We first review the original method proposed by Nayar and Bolle. Next, we introduce an efficient method with linear complexity in the number of pixels, based on a distance transform [Bor84] that computes ratios and collects histograms of ratios on colour component boundaries in three sweeps through the image. Adjacency of colour components is then defined in a robust manner by a threshold function on the number of boundary points. Finally, the histogram of ratios is analysed by a one-dimensional form of GT CLUSTERING, allowing for more than one edges to be inserted between a pair of nodes in the *CAG*.

In section 4.4, the difference between an *image*, *single view* and *object CAG* is discussed and the process of obtaining the aspect-independent *object CAG* is presented. Having introduced the *CAG* in full detail, we comment on its relationship

to other commonly used representations of colour structure of objects, the colour histogram and the region adjacency graph.

We conclude the chapter with section 4.5 where we discuss two loosely related topics: possible extensions of the CAG representation and the dependence of the CAG representation on phenomena that are difficult to model analytically and require experimentation, eg. self occlusion, error in the segmentation process and change of aspect.

4.2 Reflection Model

In [Hea89b] Healey showed that reflectance S of an opaque material can be accurately approximated as:

$$S(g,\lambda) = \begin{cases} m_i(g)C_i(\lambda) & \text{metals} \\ m_b(g)C_b(\lambda) + m_i(g)C_i(\lambda) & \text{dielectrics} \end{cases} \tag{4.1}$$

where g and λ indicate dependence on photometric angles and wavelength respectively. $C(\lambda)$ and $m(g)$ are, in general, arbitrary functions. Subscripts i and b denote terms associated with interface (specular) and body reflection. Following [Hea89a] we simplify the second part of equation (4.1) (equivalent to the dichromatic reflection model [Sha84]) by assuming that effects of interface reflection for (inhomogeneous) dielectrics are negligible over most pixels in the image. Moreover, high intensity of specular points often saturates the sensor which significantly reduces possibilities of colour analysis. The effect of a large specularity on the CAG representation is shown in figs. 1.1d and 1.3d.

Introducing the simplification to (4.1) we obtain

$$S(g,\lambda) = m(g)C(\lambda) \tag{4.2}$$

Sensor response of a standard imaging device is well modelled by a spectral integration process:

$$p_k^X = \int_{\lambda_1}^{\lambda_2} \rho_k(\lambda)S(g,\lambda)E(\lambda)d\lambda, \qquad k = 1,\cdots,n \tag{4.3}$$

$$= m(g)\int_{\lambda_1}^{\lambda_2} \rho_k(\lambda)C(\lambda)E(\lambda)d\lambda \tag{4.4}$$

where p_k^X is the response of the $k\text{-}th$ sensor at location X of the sensor array, $E(\lambda)$ is the spectral power distribution (SPD) of the light incident on the surface patch that is projected on pixel X, $\rho_k(\lambda)$ is the responsivity function of the $k\text{-}th$ sensor and n is the number of sensors at X. In our experiments we used standard *rgb* equipment; we will therefore assume $n = 3$ and $\bar{p} = (p_1, p_2, p_3) = (r, g, b)$, where for brevity we drop the explicit notation of the output dependence on sensor location X. From (4.4) it follows that measured values corresponding to a single material illuminated by a source with SPD $E(\lambda)$ of a strength controlled by a constant (ie. $kE(\lambda)$) will lie on a line passing through the origin.

We elected to represent the linear cluster by its intersection with the chromatic plane, ie. the plane passing through $(0, 0, 1)$, $(0, 1, 0)$, $(0, 0, 1)$ in the *rgb* space. Computing the intersection and substituting from eq. (4.4) we obtain:

$$ch_r = p_1/(p_1 + p_2 + p_3) = r/(r + b + g) \tag{4.5}$$

$$ch_g = p_2/(p_1 + p_2 + p_3) = g/(r + b + g) \tag{4.6}$$

$$ch_i = \frac{\int_{\lambda_2}^{\lambda_2} \rho_i(\lambda)C(\lambda)E(\lambda)d\lambda}{\sum_{k=1}^{3} \int_{\lambda_1}^{\lambda_2} \rho_k(\lambda)C(\lambda)E(\lambda)d\lambda} \tag{4.7}$$

It is clear from eq.(4.7) that for the chosen reflection model chromatic values ch_i are independent of viewing geometry and illumination intensity. Sensor noise, variations in material composition, specularities and interreflections will cause the observed values to deviate from the line. We treat the deviations as random noise. We do not model the error distribution of the deviations, but only assume that smaller errors are more likely than larger ones. Under this assumption, the noisy linear cluster corresponding to a material of a certain reflectance will form a unimodal cluster in the 2-dimensional chromatic space and the detection of such clusters can be rephrased as the problem of unsupervised clustering of unimodal distributions [Kit76].

In chapters 4, 5 and 6 we do not address the problem of changing illumination SPD. We avoid the issue with a standard disclaimer, ie. 'Input image is assumed to be processed by a colour constancy algorithm (eg. [MB86, Fin94, Fin95])'. In practice, we performed all matching experiments on *CAG*s from images illuminated with light of identical SPD. We believe that problems related to changes in illumination do not reduce the usefulness of the *CAG* representation. On the contrary

- the CAG is built in an illumination independent way and may even provide useful input (eg. correspondence) to colour constancy algorithms. This need not be necessary though. In a recent paper [Fin95], Finlayson presents a colour constancy algorithm that does not require correspondences and, in a significant generalisation, operates outside the restrictive Mondrian world. In fact, the algorithm compensates the influence of illumination SPD on *chromaticity*, ie. exactly on the representation of colour used in the CAG!

4.3 Building the Colour Adjacency Graph

4.3.1 CAG Nodes. Graph-theoretical clustering

In section 4.2 we showed that, under the assumed reflection and noise model, chromatic values ch_r, ch_g of pixels imaging the same material form a unimodal cluster in the chromatic plane. We call this set of pixels a colour component. Each node of the CAG represents a *colour component*. The problem of unimodal cluster separation has received significant attention in the pattern recognition community [KNF76][KB90][Kit76]. We adopted the Graph-theoretical clustering method of Koontz and Fukunaga. The method can be outlined as follows (for details and comparison with other approaches see [Fuk90]).

Algorithm 1: **Graph-theoretical clustering (in chromatic plane)**

1. Compute chromaticity histogram. In each bin, maintain a list of pixels voting for the bin.
2. For each bin, find the bin with maximal count in a given neighbourhood. Store a link to this bin.
3. Such links form a forest, with a root of each tree in a local maximum. The set of pixels voting for bins in a single tree form a unimodal cluster.

The graph-theoretical clustering algorithm requires two parameters: the size of a histogram bin and the size of the neighbourhood considered when searching for

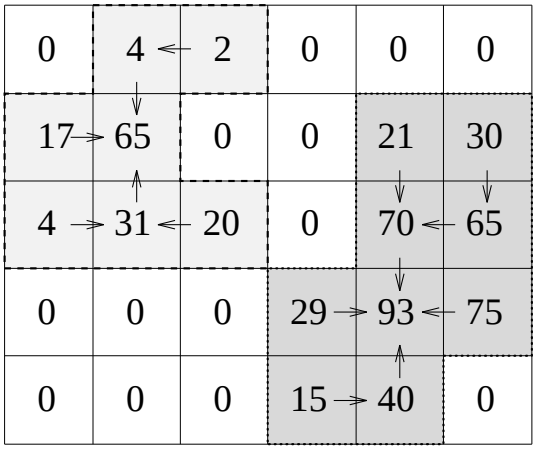


Figure 4.1: Graph-theoretical clustering in 2D space. Arrows symbolise links to bins with the largest count in local neighbourhood. Different shading highlights the two clusters of the example. Although the example used 4-neighbourhood, in this particular case (and generally with well separated clusters) the result is fairly insensitive to the size of searched neighbourhood. In this example, identical clustering is obtained with 8 and 12 neighbourhoods.

maximal bins in local neighbourhood. Methods for automatic selection of these parameters are given in [KB90].

The result of the graph-theoretical clustering for a simple example is shown in figure 4.1. The arrows symbolise links to maximal bins in the local neighbourhood (in this case 4-neighbourhood). Shaded areas correspond to clusters detected by the method.

In our implementation we compute the links to maximal neighbourhood bins in a single sweep. In the same pass we insert indices of all local maxima in a list. We sort the list according to the bin count. Next, bins belonging to the cluster associated with the local maximum currently at the top of the list are obtained by traversing the tree rooted at the local maximum. Note that this implementation first outputs the cluster with a mode in the global maximum; other salient clusters follow. The algorithm therefore need not (and typically does not) completely segment the feature space. The process can be terminated if eg. a cluster with interesting properties was found or a sufficiently large part of pixels have been analysed. The process can also be easily tuned to serve a top-down process by eg. considering maxima only from a selected part of the chromatic space, corresponding to a specific object.

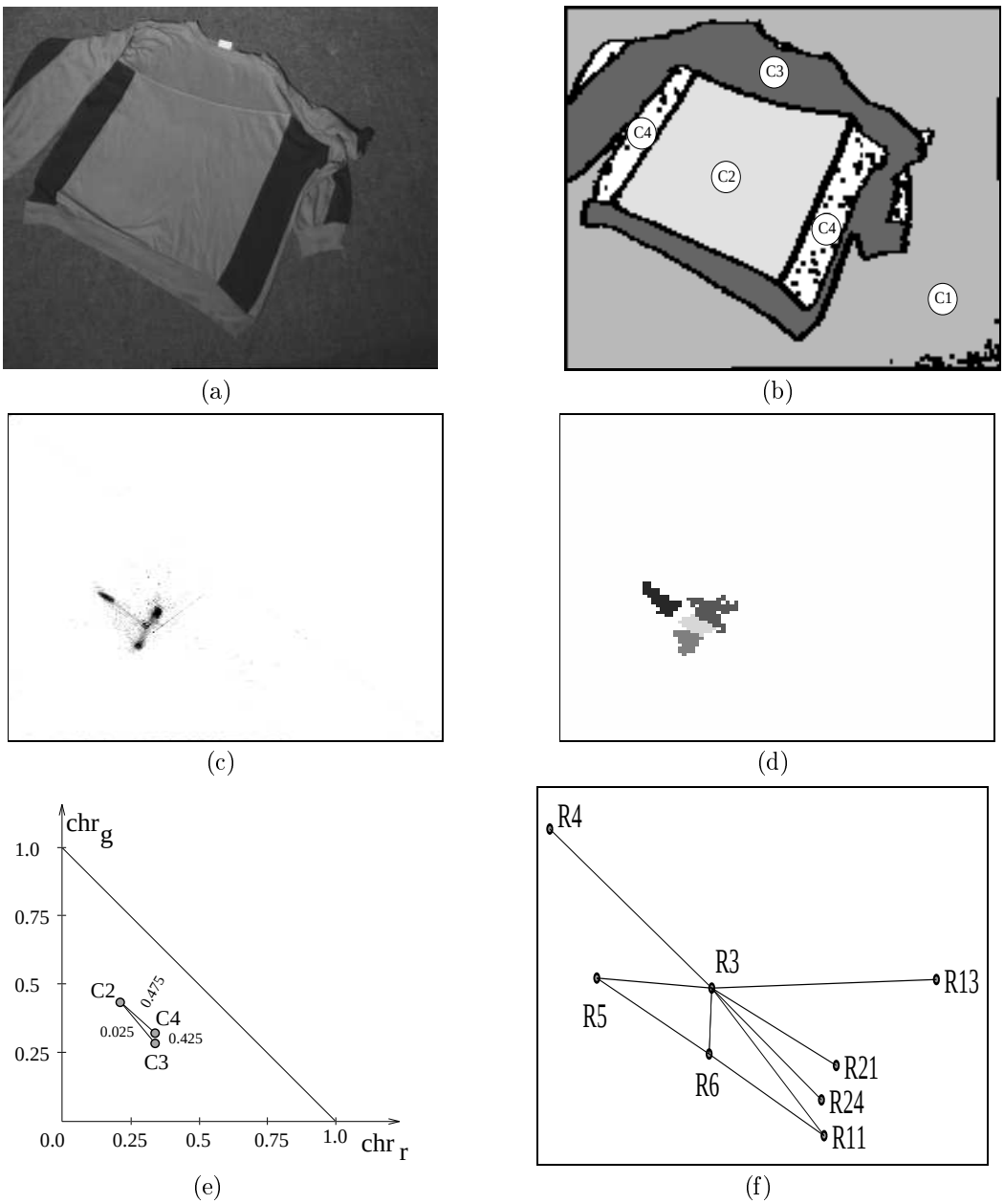


Figure 4.2: Building the *CAG* - nodes. Acquisition of the 'pyjamaTop' model. (a) Original image. (b) Segmentation induced by the partitioning of the feature space. Colour components are shown after boundary pixel removal, with pixels labelled according to cluster membership. (c) Chromaticity histogram. Darker regions represent higher counts. Chromacity coordinates as in (e). (d) Partioning of the chromatic space by GT CLUSTERING. (e) *CAG*. The largest component, C1, is known to belong to background and therefore not included in the object *CAG*. (f) *RAG*.

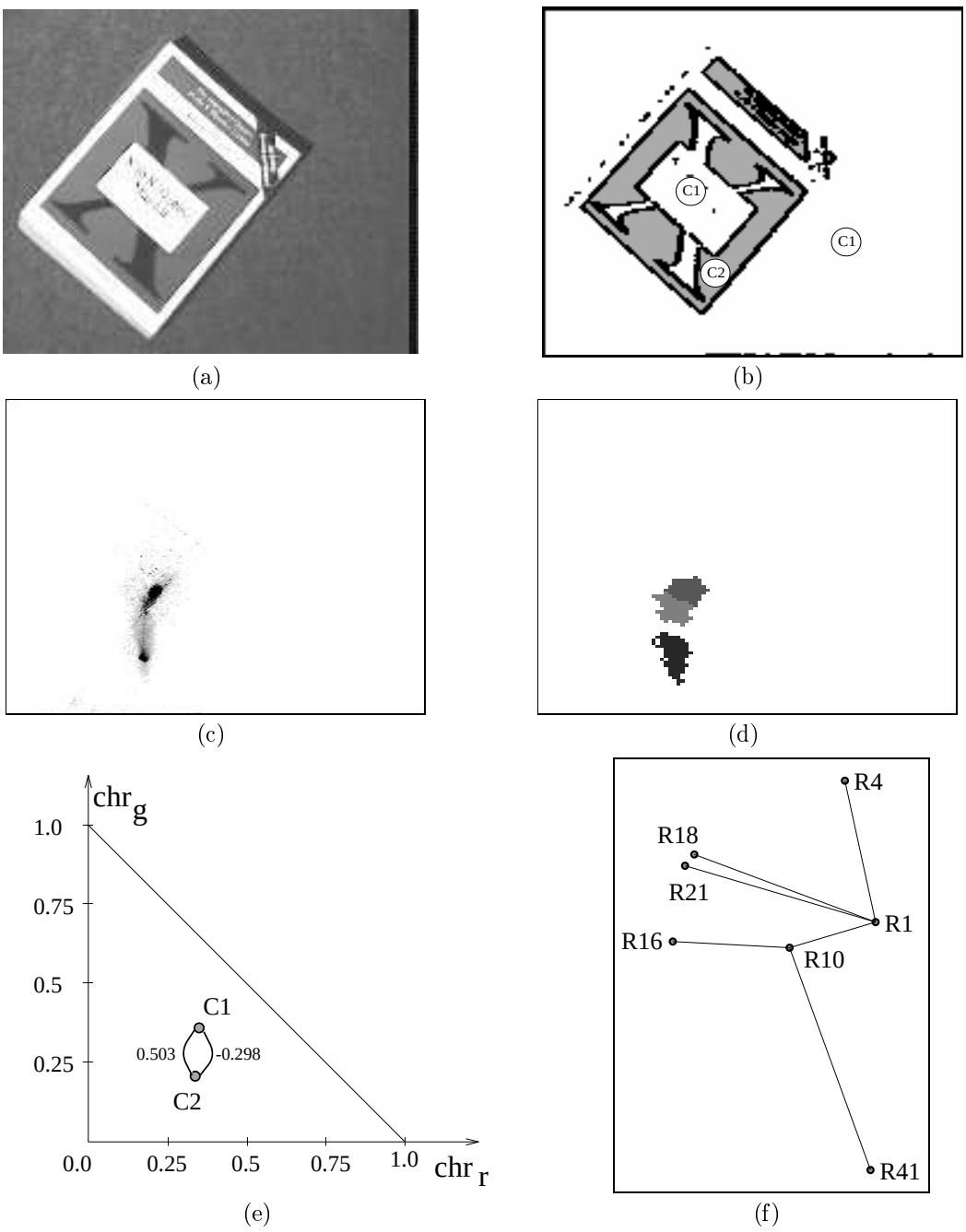


Figure 4.3: Building the *CAG* - nodes. Acquisition of the 'manual3' model. (a) Original image. (b) Segmentation induced by the partitioning of the feature space. Colour components are shown after boundary pixel removal, with pixels labelled according to cluster membership. (c) Chromaticity histogram. Darker regions represent higher counts. Chromacity coordinates as in (e). (d) Partioning of the chromatic space by GT CLUSTERING. (e) *CAG*. (f) *RAG*.

Colour Label	Chromaticity		Cluster Maximum	Cluster Size
	r	g		
1	0.385	0.375	10895	53245
2	0.205	0.445	8173	17810
3	0.335	0.285	6375	20067
4	0.355	0.315	2555	8703
5	0.265	0.395	106	403
6	0.295	0.365	61	351

Table 4.1: Results of the graph-theoretical clustering on the 'pyjamaTop' image (fig. 4.2a). The clusters are detected in the order defined by the bin count of the maximal bin (peak size). Chromaticity associated with the maximal bin (second and third columns) attributes the corresponding *CAG* node. The total number of pixels in each cluster is shown in the last column. After boundary pixel removal, no pixels are left in clusters 5 and 6. Knowing that cluster 1 originates from the background, only clusters 2,3 and 4 are inserted in the object *CAG* (see fig. 4.2e)

Table 4.1 summarises performance of the graph-theoretical clustering for the image shown in fig. 4.2(a). The chromaticity of the maximal bin in each cluster is used as attribute of the node. In a post-processing step, pixels lying on colour component boundaries are removed by erosion. In general, boundary pixel colour is a mixture of colour inside and outside the colour component and is therefore unreliable.

The intermediate results and the output of the process are visualised in figures 4.2 and 4.3. The input images are shown in subfigures 4.2a and 4.3a. Figures 4.2c and 4.3c depict the histogram of chromaticities that is passed as input to the GT CLUSTERING process. The partitioning of the chromatic space into unimodal clusters is shown in figures 4.2d and 4.3d. The partitioning-induced segmentation into colour components is shown, after the erosion described below, in 4.2b and 4.3b respectively. The *CAG* embedded in the chromatic plane is depicted in 4.2e and 4.3e. To highlight the difference in complexity of the *CAG* and *RAG* representations, the *RAG* for the two images are also shown (subfigures (f)). In both examples the computation necessary to obtain nodes of the colour adjacency graph took less than 0.5 second on SUN SPARC 10.

4.3.2 Edges. Reflectance Ratio

Edges of the *CAG* represent adjacency of colour components. Nayar and Bolle [NB93] show both experimentally and theoretically that, under general reflection conditions, intensity ratios on smooth surfaces are photometrically invariant and equivalent to 'reflectance ratios'. To avoid the mathematical inconveniences caused by the fact that an intensity ratio can attain values in the range $\langle 0, \infty)$ we adopt the convention of [NB93] and define the ratio as:

$$r_I = \frac{I_1 - I_2}{I_1 + I_2} \tag{4.8}$$

which maps the ratio conveniently into the interval $\langle -1, 1\rangle$, with the sign expressing the 'brighter than' relationship.

The implementation of the ratio computation must take into account two effects. Nayar's proof of photometric invariance of the reflectance ratio hinges on the assumption that boundary pixels adjacent in the image lie on surface patches with identical geometry with respect to the viewer and the sources of illumination. This assumption suggests that the ratio should be computed at neighbouring pixels. On the other hand, no imaging system is perfect and boundary pixels are affected by blur, finite resolution of the sensor and limited bandwidth of the acquisition chain.

In the original algorithm of Nayar and Bolle, a compromise between the two opposing requirements was achieved as follows ([NB93], page 1533). The algorithm operates on a segmented image with labelled regions; some pixels are labelled as belonging to no valid region and are not considered. The reflectance ratio is computed for every boundary pixel inside a valid region by invoking formula 4.8 on the intensity of the boundary pixel and a pixel outside the region in the same row or column at distance d, depending on the direction of the boundary. Distance d is a user-supplied parameter, usually between two and five. The reflectance ratio of a *region* is defined as an arithmetic mean of the ratio computed along a region boundary. The main advantages of this algorithm are the simplicity of implementation and speed. Unfortunately it has significant limitations. First of all, it is assumed that the reflectance ratio is constant along the entire boundary or, in other words, that a region is completely surrounded by a *single* region with constant reflectance. This is not true for *any* of the regions shown in figs. 4.2 and 4.3 . Moreover, The arithmetic mean is not a robust statistic and may be influenced by outliers, eg.

in the situation shown in fig. 4.5d where the left peak in the ratio histogram is asymmetric. Finally - the pair of intensity values selected for ratio computation is not selected symmetrically with respect to the region boundary; in fact, one of the values is taken from a possibly contaminated value exactly on the boundary while the other sample is taken d pixels from the edge.

In our implementation we proceed as follows. The image of colour component labels is eroded. The erosion removes pixels on the boundaries affected by blurring. A non-square 5×3 structural element was used, reflecting the different horizontal and vertical bandwidths in the acquisition chain (figs. 4.2b, 4.3b). At this point we would like to find, for each pixel on the boundary of the eroded labelled image, the nearest pixel belonging to a different colour component[1]. We abandoned the direct implementation of the idea after initial tests; the search took order of minutes on SUN SPARC 10 (depending on the maximum distance allowed). Instead an efficient approximation by distance transform was adopted [Bor84]. In two sweeps through the image, at every pixel not belonging to the eroded colour components, a distance from the nearest pixel belonging to a colour component is computed. In the same two sweeps, the intensity of the nearest colour component pixels is propagated to the appropriate pixels in the eroded zones. The ratio of intensity values is computed at every point that is equally distant from two or more colour components and then inserted into a list of ratios, maintained for every pair of colour components. An edge is inserted into the CAG between the nodes representing colour components if the size of their boundary list is larger than a small threshold (default set to five).

We analyse the intensity ratios in the list of boundary pixels using the 1D version of the graph-theoretical clustering mode. If a single mode is found, the ratio corresponding to the maximal bin attributes the edge. If more modes are found, than an edge is inserted for every mode[2]. As described, the computation of CAG edges and the reflectance ratios required less than ten seconds[3] on SUN SPARC 10, eg. 4.6 seconds for the 'pyjamaTop' model example.

[1]Note that this pixel is also on a boundary, so the ratio is computed on a pair of values on boundaries of eroded colour components. Unfortunately we also have to note that the 'nearest from' relation is not (in our context) symmetrical.

[2]Strictly speaking the CAG is a multi-graph rather then a simple graph[RW90], as more than one edge may connect a pair of nodes. Throughout the text we do not make the distinction as it has neither implication on our implementation of the graph structure nor on the matching strategy.

[3]The exact time depends on the number of boundary points in the image.

The computation of CAG edges is visualised in figures 4.4 and 4.5. The eroded labelled image is shown in the top left corner (figs. 4.4a and 4.5a). The distance map on the eroded areas is depicted in the top right corner (4.4b and 4.5b). The graph in the bottom left corner shows the set of ratios between a pair of components. In the 'manual3' example this information looks highly ambiguous, but the peaks of the ratio histogram (shown in the bottom right corner) are sharp and easily detected by the 1D variant of GT CLUSTERING. If the smoothness constraint on the boundary is violated, then different ratios will be observed for the same pair of colour components and inserted in the CAG. Matching of such colour components will not be affected by the ratios, as any possible ratio will be represented in the model.

The benefit of using intensity ratios is clearly seen in the example shown in fig. 4.3. A purely chromatic (or hue and saturation) approach would describe the object colour structure with four numbers, namely the chromaticity (hue-saturation) of gray (the black and white part) and of blue. In our approach, the same object is described by six numbers, which greatly reduces a random mismatch.

4.4 The Image, Single View and Object CAG

In the recognition experiments described in section 6 a CAG description of a scene is matched against CAGs representing objects. The form of the object, single view and image (scene) CAG is identical:

$$CAG = (N, E, \psi, rat, chr)$$

where N is the set of nodes, each representing a colour component, E is the set of edges, each representing adjacency of colour components *with a certain reflectance ratio*, ψ is the mapping from E to an *ordered*[4] pair of nodes of N. Function $chr : N \rightarrow \mathcal{R}$ assigns a chromaticity to every node and function $rat : E \rightarrow \mathcal{R} \times \mathcal{R}$ a reflectance ratio to every edge.

The three types of the CAG differ in the set of nodes, or more precisely in what the colour components included in the set represent. In the image (scene)

[4]The reflectance ratio of an edge changes sign if order of the corresponding nodes changes; it is therefore necessary to treat the CAG as a directed graph; adjacency per se, being a symmetric relation, would be represented by an undirected graph.

CAG all colour components of the image are included. In the model acquisition stage, image *CAG* is transformed into a single view object *CAG* by removing nodes corresponding to background, obtaining a *CAG* where the set of nodes represent the set of object colours visible from a single view. It is quite common that not all colours on the surface of an object are visible from a single view. To obtain the full *object CAG* a number of single view *CAG*s are combined by a *CAG*-union operation (described below). So, starting from a completely view-dependent image centred representation we obtain an object centred description where the set of nodes of the *CAG* corresponds to the totality of colours on object surface.

Both conversions, ie. from image *CAG* to single view *CAG* and then to object *CAG* are performed automatically during model acquisition. We use two methods of automatic removal of the background each based on different assumptions. When we had control over conditions in which models were acquired, we placed objects on a cloth with simple uniform known chromaticity. The node corresponding to the background together with all incident edges were consequently removed from the *CAG*. In environments with complex background structure we used colour differencing [YKM94] to segment the region of the image covered by the "introduced" model and then clustered only data from the segmented region.

If more than one view is needed to obtain a full object *CAG*, a *CAG*-union operation must be performed. The union operation was implemented as *CAG* matching (chapter 5) of the current single-view *CAG* with the object *CAG* of the given model. If no new colour components appear in the latest single-view *CAG*, the single-view *CAG* is a subgraph of the object *CAG*[5]. If unmatched nodes and edges remain in the single-view *CAG*, they are inserted into the object *CAG* .

4.4.1 Link to other representations

In chapter 3 we reviewed three types of representations of colour structure of multi-coloured objects: the list of colour components, the region adjacency graph and the colour histogram. The colour adjacency representation is useful in refining the hierarchy of colour representations between the simple list of colours void of representation of colour topology and the adjacency of colour regions, which is typically

[5]The match is not in general one-to-one, because the latest single-view *CAG* need not contain all previously seen object colours or colour edges.

significantly more complex as well as less stable than adjacency of colour components.

Despite its name, the bin counts in the colour histogram are related to geometry and relative projected areas and therefore metric shape. Colour information is coded in the *indices*. By storing the indices of the chromatic histogram, the CAG uses the stable information about colour structure, discarding the geometric structure. This is not to say that the structure contained in the histogram is not useful for recognition. We see a hierarchy, from the most basic to the complicated:

$$set\ of\ colours \subset colour\ adjacency \subset region\ adjacency$$

Histogram methods stand apart and cannot be directly included in the hierarchy above, as they exploit geometric rather than topological information. If objects differ in the most fundamental way (from the point of view of colour recognition) and the sets of their surface colours are different then discriminating is easy without a recourse to the CAG representations. Similarly, it the CAGs of two object are different, comparison of the (in general more complex) RAGs is not necessary. But why not use the most powerful structural information? Because the strength comes at a price, in term of loss of invariance (to viewpoint, illumination, aspect, etc.).

The problems and the relationship of the CAG, colour histogram and RAG representations are nicely highlighted in a problem suggested by a question in the annotation of a bibliographical entry referring to our work on the CAG found in the electronic database at the University of South California [K.P]: " (Do) French and UK flags have the same CAG?". The straightforward answer would be: "Yes, the structure of the two CAGs is identical, only some attributes may differ, especially the intensity ratio between white and blue". The implication is that it is difficult to distinguish the two flags using the CAG representation, where the other colour-based approaches would succeed. This may well be the case, but we do not see this as an example pointing to a fundamental problem of the CAG approach. Firstly, the red, white and blue regions may be (ambiguously) interpreted by CAG matching as either French or British (Dutch etc.) flags and RAG matching used for disambiguation. Such process may be more efficient than direct RAG matching. As drawn in textbooks, the RAG of the French flag has three nodes and is identical to its CAG. The RAG of the British flag is a graph with seventeen nodes; its CAG

has three nodes. It is interesting to note that most people would draw the UK flag as having one red connected region in the center (the two crosses), eight white and eight blue regions. In fact, the flag has five red, four white and eight blue regions. The former topology is usually perceived if the flag is seen at low resolution. A further complication stems from the fact that the RAGs of the two flags change when the flags fly in the air, when they are partially occluded or folded. In such circumstances the RAG changes significantly and it is easy to occlude the British flag so that its RAG is identical to that of an unoccluded, spread French flag.

The colour histogram methods are in trouble too. In the 'fully spread' position both flags have very similar histograms. Moreover, the colour histogram cannot easily capture the difference in the brightness of the two blue colours and yet remain insensitive to illumination intensity. Under occlusion or non-rigid transformation of shape the projected relative areas and hence the bin counts for the red, blue and white may change arbitrarily. It is also interesting to note that logos (eg. British Airways, Air France) linked to the national flag often preserve the CAGs, but do not preserve topology of regions (the RAG) nor relative areas of colour patches (the histogram).

4.5 Properties: Final Remarks.

In section 4.2 we studied the stability of the *attributes* of the CAG representation to changes of viewing and illumination geometry and the intensity of illumination. So far we have not touched upon the significant issue of robustness in the presence of occlusion, non-rigid change of shape and segmentation errors. Each of these problems is difficult to treat analytically; what is the set of possible occlusions? Is it possible to model the unconstrained change of shape of an object like a shirt? How should segmentation errors be modelled?

We do not intend to try to resolve these issues, rather we point out that *none* of these problems is avoided in the experiments presented in the thesis. For heavy occlusion, see figs. 1.1f, 5.1b or 6.4; for non-rigid transformation of shape see the series of experiments shown in fig. 6.3. The effects of imperfect segmentation are visible even in the controlled environment used for model acquisition in the examples shown in this chapter. Looking at fig. 4.2 we see that the holes in colour component C4

(subfigure b) do not influence the CAG (no new colour component is introduced, see subfigure e); the same is not true for the region adjacency representation (subfigure f).

We also observed a high degree of insensitivity to self-occlusion and change of aspect. Disregarding singular viewpoints, non of the objects in our database (fig. 6.1) needed more than two single-view CAGs to obtain a full object CAG. Again, it is difficult to make general statements but we observed this property on many common objects - books, toys, instruments, animals etc. In general, self-occlusion may only distort the edge part of the CAG information by introducing accidental new adjacencies or by skewing the reflectance ratio of an existing edge. Although such accidental edges clearly appeared in our experiments (eg. fig. 6.3), they matched some of the edges existing on the model and did not adversely affect the matching result.

4.5.1 CAG extensions

Other attributes of the CAG nodes, besides the chromatic coordinates of the mode of the corresponding unimodal cluster, can be derived from the cluster histogram. Efficiency is not of concern as the histogram is available in a convenient form after the clustering process. In our experiments we elected not to insert the additional information contained in the histogram into the CAG representation. The main reason for the decision was that the shape of the cluster is dependent on phenomena like specularity and interreflections, treated in our approach as noise. Although estimates of noise associated with measurements are necessary for certain matching techniques (see eg. [SPK95]) we felt that they are not part of the object description. Moreover, the level of noise depends on the scene; including information of this type runs against one of our central objectives of defining a representation independent of environment.

The same is not true in the case of CAG edges. The reflectance ratio is invariant only at points on the colour component boundaries that do not coincide with surface orientation discontinuities. A viewpoint-dependent change in the value of the local maximum (maxima) in the histogram of reflectance ratios should be interpreted as indication of the presence of surface orientation discontinuity rather than noise. The observation is equally valid for points near occluding boundaries of objects,

where pairs of pixels adjacent in the image are, in general, on surfaces with different orientation and widely separated in the scene. Inclusion of a measure of stability of a reflection ratio attribute of the *CAG* edge and its implication on *CAG* matching performance is a highly relevant issue and will be investigated in our future research.

References

[Bor84] G. Borgefors. Distance transformations in arbitrary dimensions. *Computer Vision, Graphics and Image Processing*, pages 321–345, 1984.

[Fin94] G. Finlayson. Color constancy and changing illumination. In *SPIE, Feb. 94, San Jose CA*, 1994.

[Fin95] G. D. Finlayson. Color constancy in diagonal chromaticity space. In *Fifth International Conference on Computer Vision (Cambridge, MA, June 20–23, 1995)*, pages 218–223, Washington, DC., 1995. Computer Society Press.

[Fuk90] K. Fukunaga. *Introduction to Statististical Pattern Recognition*. Academic Press, 1990.

[Hea89a] G. Healey. A parallel color algorithm for segmenting images of 3-d scenes. In *Image Understanding Workshop (Palo Alto, CA, May 23-26, 1989)*, pages 1038–1042. Morgan Kaufmann, 1989.

[Hea89b] G. Healey. Using color for geometry-insensitive segmentation. *Journal of the Optical Society of America*, 6:86–103, June 1989.

[KB90] A. Khotanzad and A. Bouarfa. Image segmentation by a parallel , non-parametric histogram based clustering algorithm. *Pattern Recognition*, pages 961–973, 1990.

[Kit76] J. Kittler. A locally sensitive method for cluster analysis. *Pattern Recognition*, 8:23–33, 1976.

[KNF76] W.L.G. Koontz, P. M. Narendra, and K. Fukunaga. A graph-theoretic approach to nonparametric clustering. *IEEE Trans. Computers*, C-25:936–944, 1976.

[K.P] K.Price.
Recognition by color indexing. ¡a href="http://iris.usc.edu/Vision-Notes/bibliography/texture45.html"¿ WWW page ¡/a¿.

[MB86] L.T. Maloney and B.A.Wandell. Color constancy: a method for recovering surface spectral reflectance. *Journal of the Optical Society of America*, 3:29–33, January 1986.

[MK95] J. Matas and J. Kittler. Spatial and feature based clustering: Applications in image analysis. In *6th International Conference on Computer Analysis and Patterns , Prague, Czech Republic, September 6-8 1995*, 1995.

[NB93] Shree K. Nayar and Ruud M. Bolle. Computing reflectance ratios from an image. *Pattern Recognition*, 26:1529–1542, 1993.

[NB94] Shree K. Nayar and Ruud M. Bolle. Reflectance ratios: An extension of land's retinex theory. In E.S. Gelsema and L.N. Kanal, editors, *Patttern Recognition in practice IV: Multiple Paradigms, Comparative Studies and Hybrid Systems*, pages 141–152. Elsevier, 1994.

[RW90] Lennart Rade and Bertil Westergren. *Beta - Mathematics Handbook*. Studentlitteratur, Chartwell-Bratt, Studentlitteratur,Box 141, S-221 00 Lund; Chartwell-Bratt Ltd., Old Orchard, Bickley Road, Bromley, Kent BT1 2NE, UK, 1990.

[Sha84] S.A. Shafer. Using color to separate reflection components. Technical report, University of Rochester, 1984.

[SPK95] A. Stoddart, M. Petrou, and J. Kittler. Probabilistic relaxation as an optimizer. In *British Machine Vision Conference, Birmingham*, 1995.

[YKM94] D. Yang, J. Kittler, and G. Matas. Recognition of cylindrical objects using occluding boundaries obtained from colour based segmentation. In E. Hancock, editor, *British Machine Vision Conference*, pages 439–440, 1994.

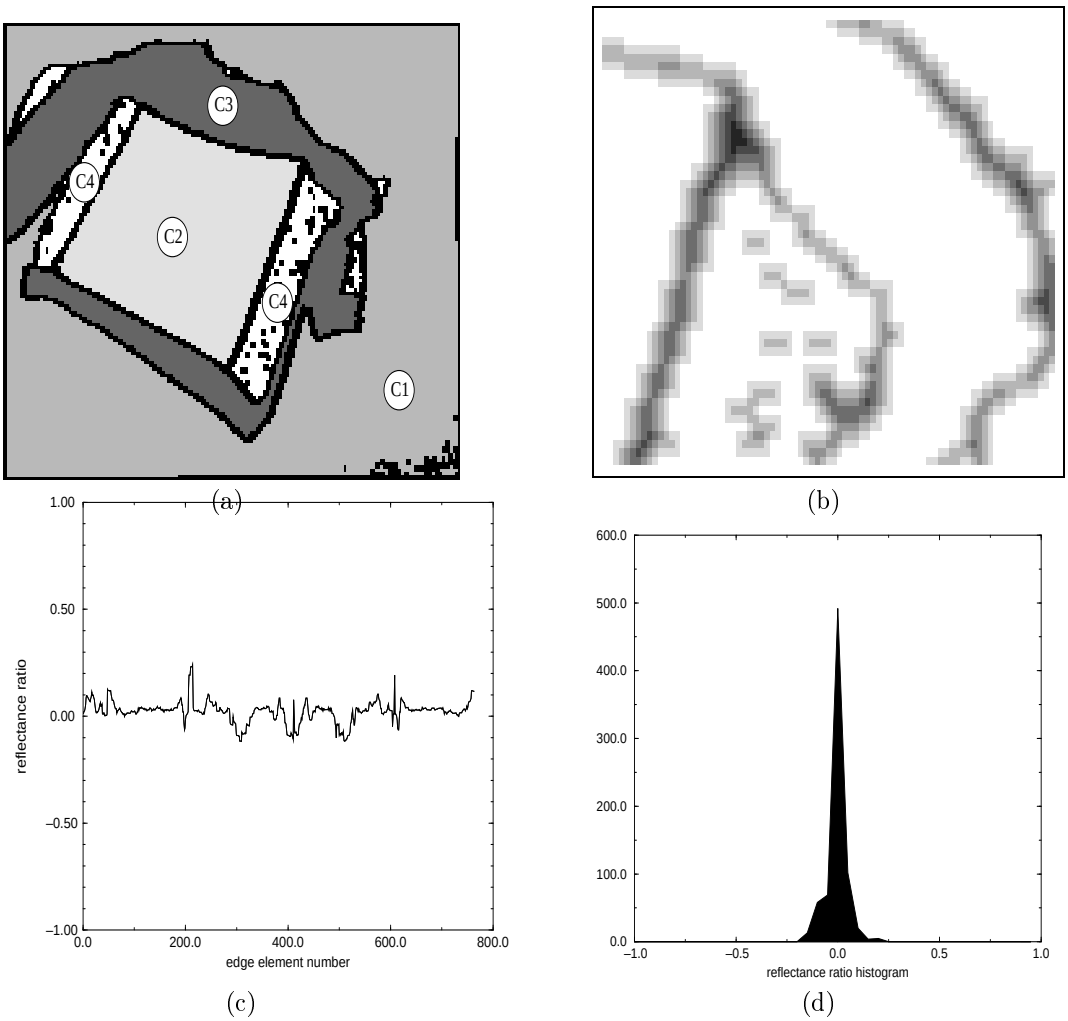


Figure 4.4: Building the CAG - edges. Acquisition of the 'pyjamaTop' model. (a) Colour components of the 'pyjamaTop' model. (b) Output of the distance transform, detail of the junction of components C2, C3 and C4. (c) Reflectance ratios along the boundary between components C2 and C3. (d) Histogram of (c). The very low contrast between components C2 and C3 in fig. 4.2a a is consistent with the reflectance ratio distribution around 0.0. The colour adjaceny graph is shown in 4.2e

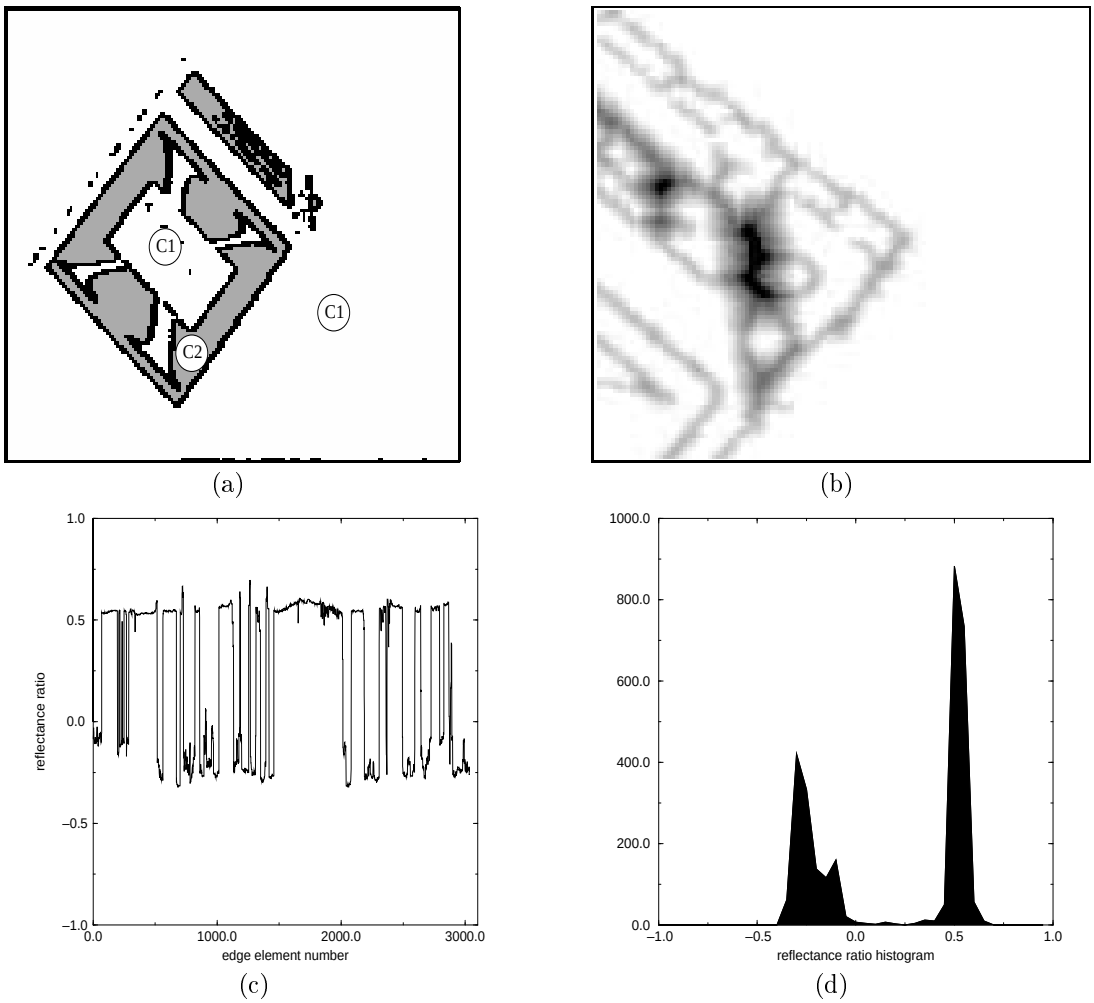


Figure 4.5: Building the *CAG* - edges. Acquisition of the 'manual3' model. (a) Colour components of the 'manual3' model. (b) Output of the distance transform, detail of the left hand corner of the manual. (c) Reflectance ratios along boundary between components C1 and C2. (d) Histogram of (c). The two reflectance ratio modes are represented by two edges between components C1 and C2 in 4.3e. Modes of reflectance ratios are detected by a 1D form of graph-theoretical clustering. The small peak close to 0.0 in (c) does not define a mode, because large values to the left are in its neighbourhood (neighbourhood diameter 0.15).

Chapter 5

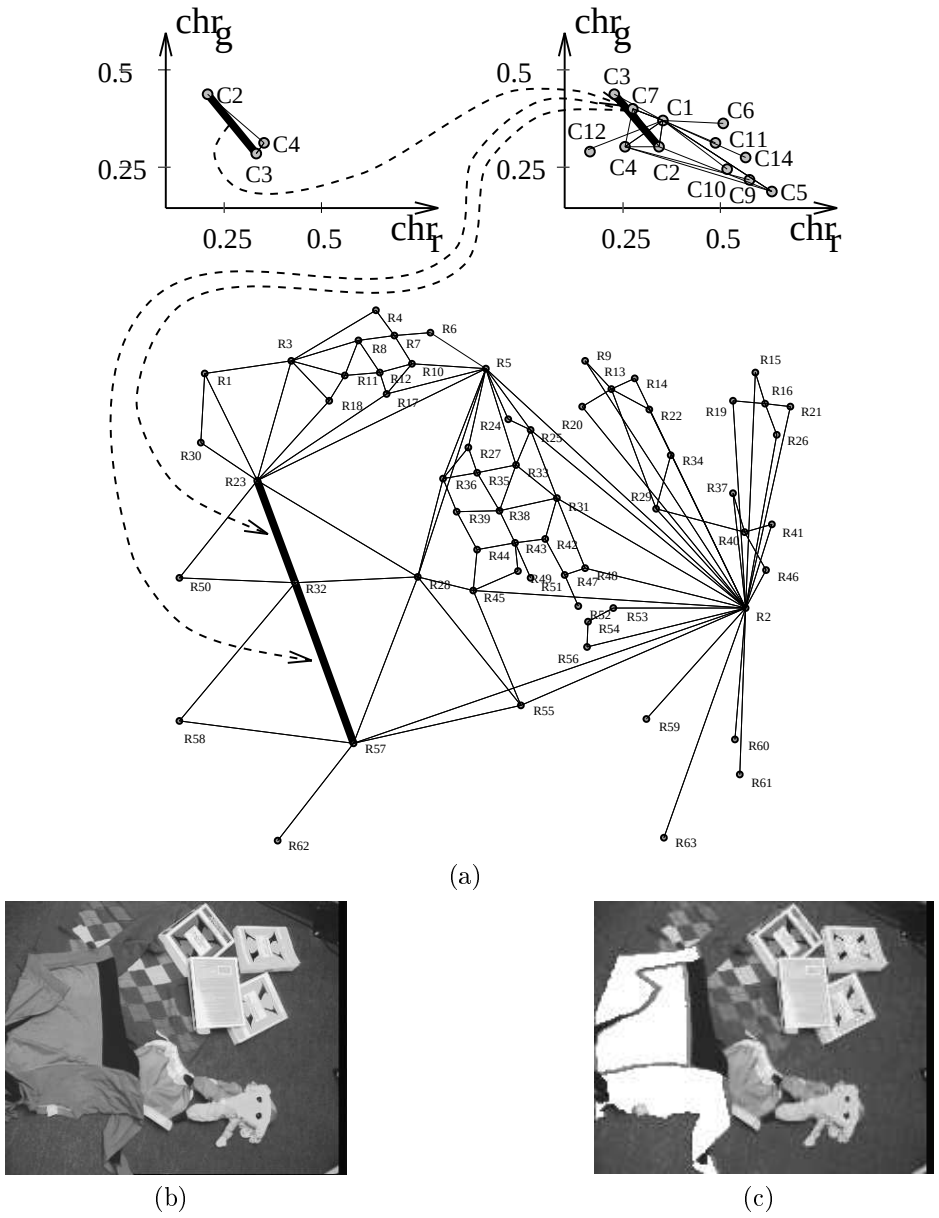
CAG Matching

5.1 Introduction

If a scene contains an object whose CAG is known, then a subgraph of the scene CAG will match a subgraph of the model CAG corresponding to the visible part of the object. Recognition and localisation of objects using CAG representation can be therefore accomplished by subgraph matching. Subgraph matching subsumes all the three categories of recognition problems mention on page 21 of chapter 2: localisation ('where'), identification ('what') and interpretation ('what and where'). More precisely, identification leads to the problem of *subgraph isomorphism* and localisation and interpretation to the problem of *bidirectional subgraph isomorphism* [BM95].

Because the work presented in the thesis focuses on the CAG representation and not on graph matching per se, we implemented an algorithm performing the most general recognition task - interpretation - via simple subgraph matching. The algorithm provides a viable alternative to the more sophisticated methods (eg. [CKP],[BM95] or references therein) because of a very low probability of mismatch between CAG edges.

The matching algorithm performing a variant of a best-first search is described in section 5.2. All of the experiments reported in chapter 6 used this algorithm. In section 5.3 we discuss possible modifications of the matching strategy for the localisation and identification task. Finally, section 5.4 describes experiments in CAG matching using a sophisticated algorithm based on probabilistic relaxation reported in [SPK95].



(a)

(b)

(c)

Figure 5.1: Matching in Experiment 3. (b) The scene. (a) The scene *CAG* is shown in the top-right corner. The scene *RAG* is depicted in the center. The top-left corner shows the 'pyjama' model *CAG*. During the recognition process edge C2,C3 of the 'pyjama' model *CAG* is matched to edge C3,C2 of the image *CAG*. The match is propagated in the *RAG* and backprojected into the image. In this particular case, three regions are labelled. By coincidence, *CAG* edges between nodes with identical indices (C2, C3) are matched. This is due to the fact that the coloured patches of the 'pyjama' belong to the largest colour components of the image after the background; their relative size is reversed, as C2,C3 is matched to C3,C2.

5.2 Simple CAG matching

Let us first schematically outline the algorithm:

Algorithm 2: **Simple CAG matching**

1. Let $\mathcal{M}$ be the set of CAG edges of all models in the database.
   Let $\mathcal{I}$ be the set of edges of the image CAG.
   Let $\mathcal{P}$ be the set of all pairs (e_M, e_I), $e_M \in \mathcal{M}$, $e_I \in \mathcal{I}$
2. For all $(e_M, e_I) \in \mathcal{P}$ compute quality of the edge match.
3. Order pairs (e_M, e_I) according to match quality.
4. **While** best match quality higher than threshold
   if edge match is consistent with previous matches
   (a) label the image CAG edge
   (b) label the image regions corresponding to the matched edge

The 'edge quality' (step 2) was approximated (ad hoc) as a weighted sum of differences in chromaticity and reflectance ratios. Two tests of consistency (step 4) are performed. A single model CAG edge can only match a single image CAG edge and vice versa. The second test checks the consistency of image interpretation. Before starting the matching process, image RAG is computed by connected component analysis of colour components. Once an image CAG edge is matched, the pair of regions mapping into the image CAG edge are labelled as belonging to the matching model CAG. Any match that would attempt to relabel a region is not accepted. An example of a pair of regions labelled as a result of a single CAG edge match is shown in fig. 5.1.

Note that although the algorithm assumes that CAG *edges* unambiguously identify the model, no such assumption is made about *nodes*. It is common that a colour component (ie. the set of pixels corresponding to an image CAG node) with a common chromaticity, eg. white-gray-black, contains regions of more than one object. This is clearly illustrated by the example of the seven X manuals, which all

have white and black patches with identical chromaticities (see fig. 6.1). Finally we remark that although the CAG can represent an object of single colour by graph with a single node, such an object cannot be matched by Algorithm 2.

5.3 Localisation and Identification

In the localisation and identification tasks, explicit constraints on location or identity of matched objects are known. In the following paragraphs we outline simple modifications of the CAG matching strategy (eg. those described in sections 5.2 and 5.4) exploiting these constraints. Such constraints are often available in a co-operative, continuously operating vision system [CC95]. Capability of a recognition engine to benefit from the constrains might significantly improve recognition efficiency [KMBN95] and allow for efficient top-down control.

Information about object location in the identification task may be exploited in a straightforward manner by computing the CAG for a given region only. This is accomplished by restricting histogram computation in GT CLUSTERING to the given region; as a consequence only unimodal clusters corresponding to the defined regions are extracted in the clustering step.

Knowledge about the object identity in the localisation task has more serious implications on the matching strategy. Analogically to the previous case, only a single model CAG of the object being localised need be matched against the scene CAG. Moreover, the information about colours on the surface of the localised object may be exploited when building image CAG by processing only clusters in the chromatic space that possibly come from a node of CAG. This strategy turns the process of building the image CAG from being purely data driven into a model-dependent one, and the unsupervised clustering problem into classification into classes with known models[1].

In general, providing the constraints on location or identity of matched objects increased the speed of computation but did not, for our test database, influence the recognition rate; for more details see chapter 6 (especially table 6.1, where the

[1] In principle the same scheme is applicable to the general recognition problem where only clusters in the proximity (in chromatic space) of one of the nodes in the model database are selected. In our experiments we found this method sensitive to variations of illumination colour and to mutual illumination.

timing results are reported. The localisation results were identical to the general recognition results reported in chapter 6.

5.4 Matching by Probabilistic Relaxation

In [SPK95], Stoddart proposed a probabilistic relaxation scheme with a modified product support rule. The method is compared with other relaxation schemes proposed by Kittler and Hancock [KH89] and Christmas, Kittler and Petrou [CKP, Chr95]. Matching of colour adjacency graphs was one of the experiments performed to establish the merit of the new rule. The *CAG* matching problem was chosen because it satisfies all the assumptions made in the formulation of the problem. Moreover, necessary estimates of prior probabilities are available. The models were taken from the publicly available database of colour objects [MMK95] and matched against modified *CAG*s of images of complex scenes. Some of the images processed by Stoddart are also analysed in experiments in this thesis, eg. the scene depicted in fig. 1.1f.

The modification of the *CAG* mentioned above comprised of estimating the probability distribution of the error of the graph attributes required by the matching algorithm. The distribution was modelled by a Gaussian with the mean and standard deviation computed from the histogram of unimodal clusters[2] In all conducted experiments, Stoddart reports successful identification of the model in the image by both his relaxation method and the method of Christmas-Petrou-Kittler. He remarks that the strong unary constraints (ie. the information in the chromaticities) made both relaxation methods converge rapidly to the correct solution (in two or three iterations). Stoddart concludes that *CAG* matching problems have too little ambiguity to allow to discriminate between the two relaxation schemes.

References

[BM95] H. Bunke and B.T. Messmer. Efficient attributed graph matching and its applications to image analysis. In C. Braccini, L. DeFloriani, and

[2]Note that in the standard *CAG* the *modes*, rather than the means, are used.

Gianni Vernazza, editors, *Image Analysis and Processing*, volume 974, pages 45–55. Springer, 1995.

[CC95] J Crowley and H I Christensen. *Vision as Process.* Springer-Verlag, Berlin, 1995.

[Chr95] W. J. Christmas. *Structural matching in computer vision using probabilistic reasoning.* PhD thesis, University of Surrey, 1995.

[CKP] W.J. Christmas, J. Kittler, and M. Petrou. Matching in computer vision using probabilistic relaxation. Accepted for publication by *IEEE Trans. on Pattern Analysis and Machine Intelligence.*

[KH89] J. Kittler and E. Hancock. Combining evidence in probabilistic relaxation. *International Journal of Pattern Recognition and Artificial Intelligence*, 3:25–51, Jan 1989.

[KMBN95] J. Kittler, J. Matas, M. Bober, and L. Nguyen. Image interpretation: Exploiting multiple cues. In *International Conference on Image Processing and Applications, Edinburgh, July 3-6*, pages 1–5, 1995.

[MMK95] J. Matas, R. Marik, and J. Kittler. On representation and matching of multi-coloured objects. In *Fifth International Conference on Computer Vision (Cambridge, MA, June 20–23, 1995)*, pages 726–732, Washington, DC., 1995. Computer Society Press.

[SPK95] A. Stoddart, M. Petrou, and J. Kittler. Probabilistic relaxation as an optimizer. In *British Machine Vision Conference, Birmingham*, 1995.

Chapter 6

Recognition Experiments

6.1 Introduction

In this chapter we present a set of experiments that illustrates the robustness of the CAG recognition method under change of viewpoint, scale and illumination intensity and in the presence of occlusion and non-rigid transformation of shape, ie. most of the environmental changes affecting object appearance listed in table 2.1. We start with a short section (6.2) discussing the design of the model database and the properties of objects included in the database. Images of the model objects and the test images were acquired in laboratory conditions described in section 6.3. The section also discusses specifications of the acquisition equipment used in the experiments. The core of the chapter, section 6.4, describes a sequence of recognition experiments followed by an analysis of the recognition results. In section 6.5 we discuss the efficiency of the CAG matching, comparing theoretical predictions with measured execution times. We summarise the experimental results in section 6.6.

6.2 Design of the model database

Matching experiments were performed with a model database containing eleven objects – seven different manuals, a manikin, a pyjama, a sweater, and an ICPR proceedings (see figs.6.1, 5.1, 6.3, 6.4 and 6.6).

The object models, ie. object CAGs, were built automatically from the acquired images as described in section 4.3. The complexity of the CAGs varied from a two

node graph for the ICPR proceedings to a four node graph for models 'manikin' and 'sweater'. The objects were chosen to represent different rigidity and appearance properties. The 'pyjama' and the 'sweater' are non-rigid objects that often undergo a dramatic change in shape with heavy self-occlusion. On the other hand, the 'manikin', although technically non-rigid, has a single stable shape and is shown in all our experiments in this position. The rest of the objects, all books, represent object with planar, specular surfaces. The set of X manuals was chosen for the experiments for three particular reasons. Firstly, it is not possible to distinguish the objects from visual clues not-related to colour[1]. Secondly the X manuals fall in the category of objects whose CAG is a multigraph[2]. Finally - the volumes differ only in a single colour and the respective colours often have very similar chromaticities (eg in the case of light and dark blue volumes, see fig. 6.1). Under such conditions, the reflectance ratios provide the information necessary for unambiguous identification of the manuals. Such conditions are valuable for assessing the role of the reflectance ratio in the overall recognition process.

6.3 Acquisition hardware. Experimental set-up.

The test images and images for the model database were acquired under artificial illumination in a laboratory. The scene was lit by a pair of Reflecta 8002 studio lamps [ref94] fitted with tungsten-halogen lamps with colour temperature 3400K.

The acquisition chain comprised a single-CCD colour camera JVC TK1070E [jvc93a] and the Videopix grabber board [vid91]. The hardware proved not to be optimal for the recognition experiments. In its specification, a setting of the JVC TK1070E camera with linear response (ie. with disconnected gamma correction) is described. Unfortunately, extensive measurements showed that significant non-linearity remains in the camera response even for this setting. Having studied the service manual [jvc93a] and circuit diagram [jvc93b] of the camera we concluded that the problem was due to the number of transformation between the Y-chroma PAL TV standard representation and the RGB representation of the colour signal.

The videopix grabber [vid91] requires the input signal to be in the Y-chroma (S-

[1] Of course, unless the recognition system is 'literate'.

[2] Due to the boundary of the black and white 'X' region.

VHS) format[3]. The chromatic part of the input is subsampled four times (so called 4:1:1 digitisation). The subsampling introduces strong anisotropy; the maximum bandwidth of the colour signal is reduced four times in the horizontal direction. As a consequence only subsampled images were used in the experiments, either 360×288 or 180×144. In the latter case a decreased size of the erosion mask in the horizontal direction was used because of the reduced smearing on colour boundaries[4].

6.4 Experiments

The sequence of images used in **Experiment 1** is shown in Figure 6.2. In addition to a subset of known objects, each frame contains objects whose model is not in the database. Thus figs. 6.2a-f contain Navid's head, trousers and the background (brown table, wall, etc.); Navid's stripy shirt, which is not in the database either, is a part of scenes 6.2d and 6.2f.

The first two rows of Figure 6.3 show recognised instances of the 'sweater'. In principle, the result could have been obtained by trying to match a single model CAG to the image CAG, as described in section 5.3. Instead, as in all matching experiments reported in this section, all the models were matched and results for a particular model were visualised. The pixels of the detected object are annotated in white. The boundary pixels between two adjacent colours are shown in black to allow the reader to see the outcome of clustering (segmentation). We can see from Figure 6.2 that the object of interest undergoes dramatic transformations due to changing viewpoint and due to its nonrigidity. Thus in (c) Navid's torso is rotated by 90 degrees as compared to the pose in (a). As the surface covered by the sweater is nonplanar, the areas of the colour patches on the sweater are distorted nonlinearly which would detrimentally affect both the histogram and RAG representations. In Figures 6.2d and (e) the appearance of the sweater is changed dramatically as a result of its nonrigid nature. In (d) the sweater is thrown over an unknown object while in (e) the sweater is thrown over the pyjama. Figure 6.3b just illustrates the presence of occlusion and in Figure 6.2f the sweater is not present. We can see in Figures 6.3g-k that the sweater is always correctly detected and segmented out. No

[3] Or composite PAL signal, but this option is of limited interest because of the low bandwidth of the chromatic component of the composite signal.

[4] All experiments reported in the thesis were carried out on 360×288 images.

false alarms are reported for Figure 6.3l.

The last row of Figure 6.3 presents the results of the query "What else is in the image and where?" Nothing is detected in Figure 6.2d as shown in 6.3m whereas the pyjama is found and segmented out in 6.2e as shown in 6.3n. In Figure 6.3f we found both the pyjama and the manikin. Note that the result contains a segmentation error. The trousers which are black have the same colour as one of the constituent chromatic components of the pyjama. As these two regions happen to be adjacent to each other, the CAG is unable to discriminate between the black patch of the pyjama and the black trousers adjacent to other patches of the pyjama.

The subsequent **Experiments 2, 3, and 4** were designed to illustrate a similar behaviour of the proposed methodology on different sets of images, shown in Figure 6.4, 6.5 and 6.6 respectively. The same object model database was used in the experimentation. In each case the queried objects were detected and segmented out of the image correctly. The results demonstrate the robustness of the advocated method to shape and illumination changes, occlusion, and a widely changing viewpoint, including scale change.

Overall, all thirteen instances of objects from the model database appearing in the six images 6.2a-f, all five objects appearing in image 6.4a, six out of seven objects in image 6.5a and three out of four objects in image 6.6 are correctly recognised. We consider an object to be recognised if *at least one* of its CAG edges is correctly matched[5]. The overall recognition rate is 93%, with one false positive (fig. 6.3p).

The two false negatives were recorded in experiments 3 and 4. In figure 6.6a the X manual at the top edge of the image occluded by the telephone handset and the TV remote control is not detected. This is due to the fact that the purple part on the spine of the manual is so narrow that it is completely eroded away during the construction of CAG edges. The same situation is encountered in the case of the white rim of the only X manual in figure 6.5 whose back is facing the viewer. Consequently the two objects are represented only by a single node in the image CAG. As mentioned in section 5.2, a single node cannot be matched by the simple matching algorithm (alg. 2).

The single false positive recorded in experiments 1 through 4 as well as false

[5]Note that in several cases, eg. in images depicting the manikin (figs. 6.3f and 6.4), not *all* CAG edges on the object are matched.

Experiment	CAG nodes	CAG edges	matching	localisation
1a	0.92	4.07	0.28	0.10
1b	0.90	4.15	0.22	0.13
1c	0.97	4.11	0.21	0.08
1d	1.00	4.09	0.22	0.13
1e	0.95	4.05	0.20	0.12
1f	0.98	4.21	0.27	0.11
2	1.00	3.95	0.19	
3	0.96	3.88	0.16	
4	1.03	3.91	0.21	

Table 6.1: Speed of CAG matching on a SPARC 10 (in seconds). Image size 360×288

positives recorded in other experiments [MMK95] all appeared under the following circumstance. An object that is not in the database containing a colour similar to one of the colours of a database object occludes the database object. Thus, in figure 6.3p, Navid's black trousers occlude the 'pyjama', an object containing a black patch.

6.5 Computational Complexity and Timing.

The processing in the matching experiments described above breaks down into two parts - the computation of the scene CAG for the given scene and the subsequent matching of the scene CAG with all the object CAGs in the model database. As described in detail in chapter 4, the CAG computation proceeds in two steps. First, the CAG nodes are obtained by GT CLUSTERING (see section 4.3.1, algorithm 1). Secondly, CAG edges with the corresponding reflectance ratios are computed as described in section 4.3.2. We measured the three components of the total recognition time for experiments 1 through 4. The results are summarised in table 6.1. The time necessary to perform GT CLUSTERING to obtain CAG nodes is shown in the second column, the time to compute CAG edges in the third column and the matching time in the fourth column.

Let us first turn our attention to the time needed to build the image CAG . Looking at the three steps of algorithm 1 it is easy to deduce that the time complexity of the GT CLUSTERING, and hence the CAG node computation, is proportional to

Experiment	CAG nodes	CAG edges	RAG nodes	RAG edges
1a	10	27	75	335
1b	8	27	63	363
1c	9	24	59	215
1d	7	30	69	273
1e	9	30	64	293
1f	7	32	48	194
2	10	23	53	134
3	10	22	38	86
4	12	29	39	158

Table 6.2: Number of nodes and edges of the CAG and RAG respectively

the number of processed pixels. This is confirmed by the second column of table 6.1. The *CAG edge* computation is dominated by morphological operations; the speed of the computation is therefore again roughly proportional to the number of processed pixels. Having collected corresponding pairs of boundary points, reflectance ratio histograms are analysed by a 1D version of GT CLUSTERING. The time required to complete this step depends on the number of boundary points and is therefore data-dependent. We attribute the fluctuations in the execution times given in the third column of table 6.1 to this dependency.

The CAG matching in the experiments described above was carried out by algorithm 2. Looking at the pseudocode of algorithm 2 we see that from the point of view of asymptotic complexity the matching time is dominated by step 3 - the complexity of sorting all pairs of CAG edges, which is $O(N^2 logN)$. The complexity of step 4 is $O(N^2)$, ie. it is proportional to the number of edge pairs (tentative matches) in the list generated in step 2. The $O(N^2)$ complexity of step 4 is not obvious from the description given in algorithm 2 and is due to the implementation that performs steps 4a, 4b and the consistency test in constant time.

However, the execution times shown in the third column of table 6.1 do not correlate strongly with the number of CAG edges given in the third column of table 6.2. We believe that this is a consequence of the comparatively modest sizes of the matching problems, too small to model the asymptotic behaviour. In practice we observed that for the matching problems most of the execution time is spent on auxiliary tasks like input and output, memory allocation and initialisation of data

structures.

The role of the overheads is clear from the data in the last column of table 6.1. The execution times were measured when a model database containing only a single object - the 'sweater' was passed to the matcher. The matcher was therefore effectively performing localisation of the 'sweater'. The number of CAG edge pairs considered for matching in this case is equal to the number of the scene CAG edges given in the third column of table 6.2. With less than ten potential matches considered in all the experiments 1a through 1f, approximately 0.1 seconds are spent in the above-mentioned auxiliary routines.

As a matter of interest we include the sizes of RAGs in the experimental images. With the exception of experiment 3, the number of RAG and CAG edges differ by an order of magnitude. Given eg. the complexity of the simple matching algorithm 2 the difference would translate into two to three orders of magnitude of the time required for matching.

6.6 Conclusions

In the presented experiments, the method for colour recognition via CAG matching performed well, achieving high recognition rate while keeping false positives to the minimum. It could be argued that the 93% recognition rate would be more impressive if observed over a much larger number of experiments. We believe that it is trivial to generate literally hundreds of results in experiments that are conducted in essentially *identical* conditions. For example, we could have asked Navid in experiment 1a (fig. 6.3a) to move five centimeters forward and collect the data, and repeat the process twenty times.

A large set of experiments without any diversity is of lesser value than a small number of experiments carried out under changing conditions. From this point of view, of the factors influencing object colour appearance listed in table 2.1, only spectral power distribution of the illumination was maintained constant in the experiments presented in section 6.4. Experiments with light with changing SPD are described in chapter 7. Alternatively, as discussed in the last paragraph of section 4.2, it is reasonable to assume that most of the effects of illumination colour change can be compensated for in a preprocessing step by a colour constancy algorithm.

Therefore the experiments can be considered to test the proposed approach under the full set of environmental changes listed in table 2.1[6]

From the point of view of computational complexity and efficiency we consider the speed of *CAG* construction and matching acceptable even if it is not real-time (ie. close to frame rate). The method is fast enough to allow for easy testing and development. Compared with the other recognition engines developed for the VAP at the University of Surrey [KMBN95] interpretations system it is one of the fastest.

References

[jvc93a] *JVC TK-1070E Service Manual.* Victor Company of Japan, Ltd., May 1993.

[jvc93b] *JVC TK-1070E Service Manual Supplement.* Victor Company of Japan, Ltd., July 1993.

[KMBN95] J. Kittler, J. Matas, M. Bober, and L. Nguyen. Image interpretation: Exploiting multiple cues. In *International Conference on Image Processing and Applications, Edinburgh, July 3-6*, pages 1–5, 1995.

[MMK95] J. Matas, R. Marik, and J. Kittler. The color adjacency graph representation of multi-coloured objects. Technical report, University of Surrey VSSP-TR-1/95, 1995.

[ref94] *Reflecta 8002 studio Operationg Instructions*, Berlichingenstrasse 9, D-8540 Schwabach, Germany, 1994. Reflecta Gmbh.

[vid91] *Using VideoPix*, 2550 Garcia Avenue, Mountain View, California 94043-1100, USA, 1991. Sun Microsystems.

[6] Insensitivity of *CAG* to the change of aspect and a comparison with respect to this property with the *RAG* representation has already been discussed in the thesis when describing results shown in figs. 1.1 and 1.3.

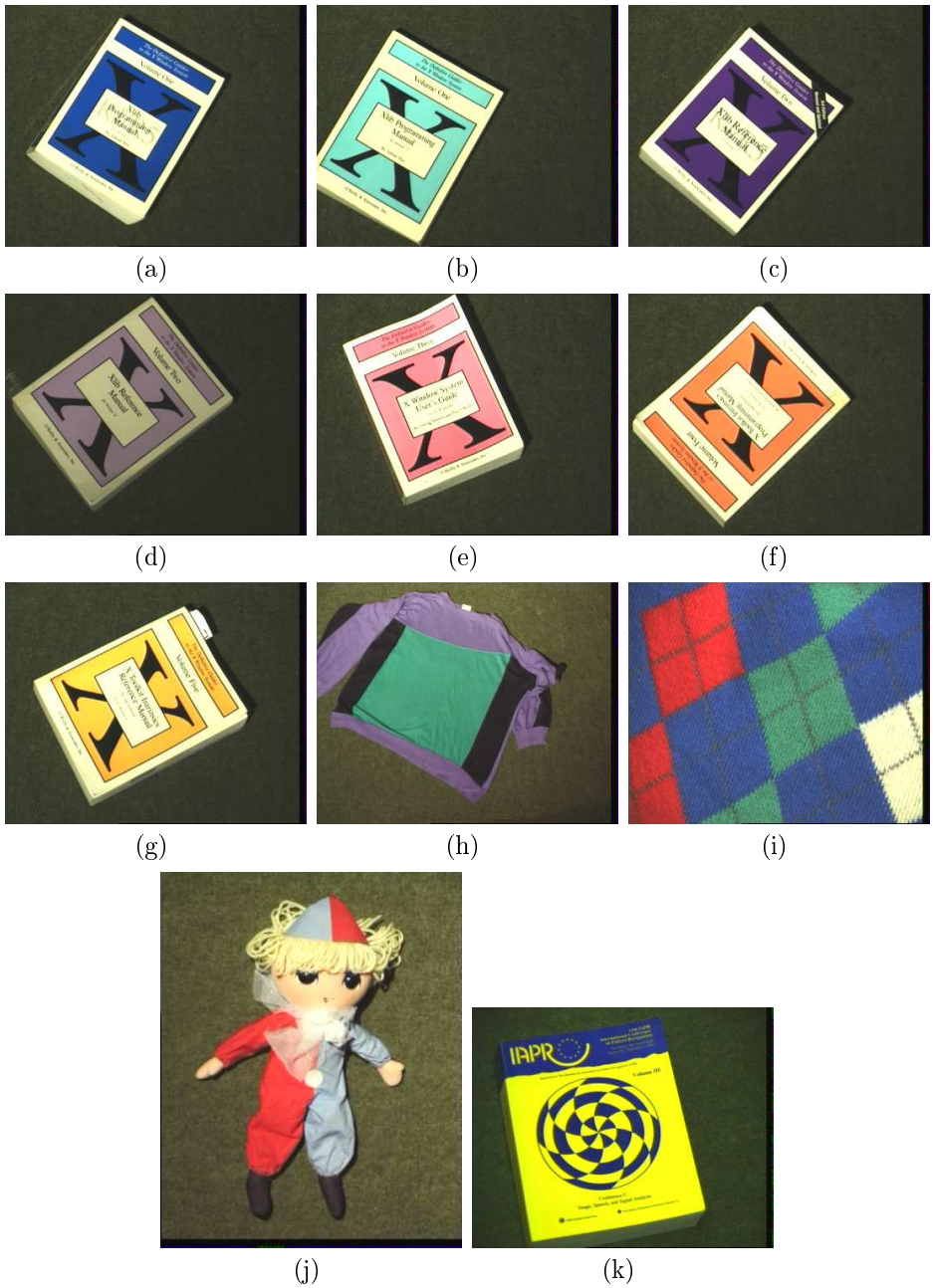
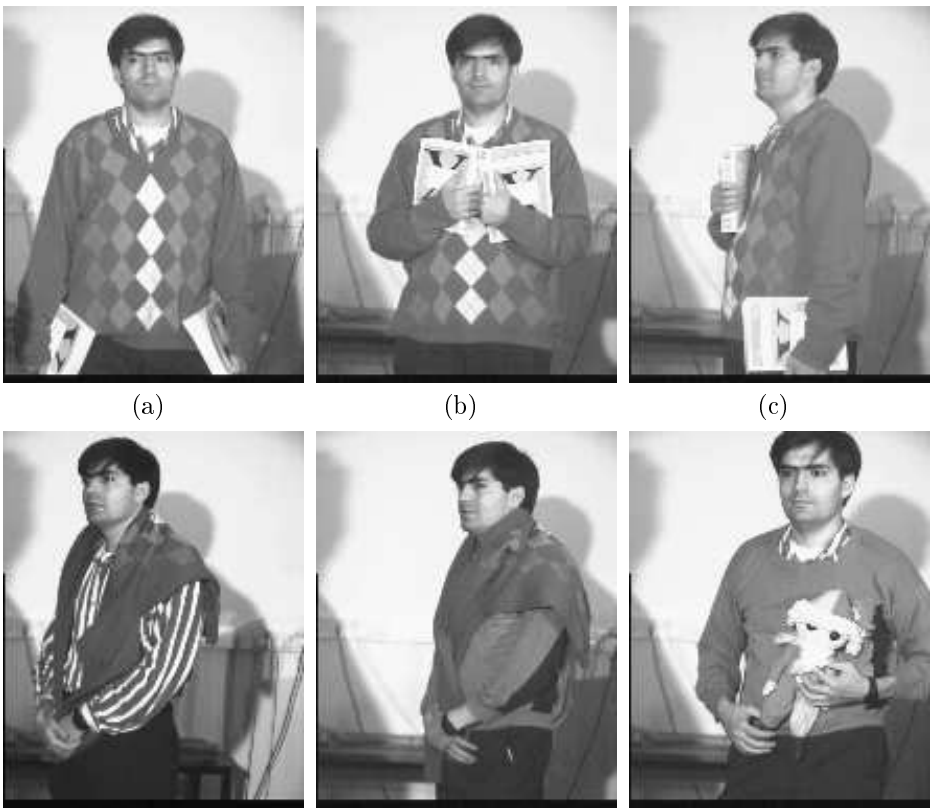


Figure 6.1: The model database: (a-g) objects 'manual1' to 'manual7'. (h) 'pyjama' ('pyjamaTop') (i) 'sweater' (j) 'manikin' (k) 'ICPR proceedings'.

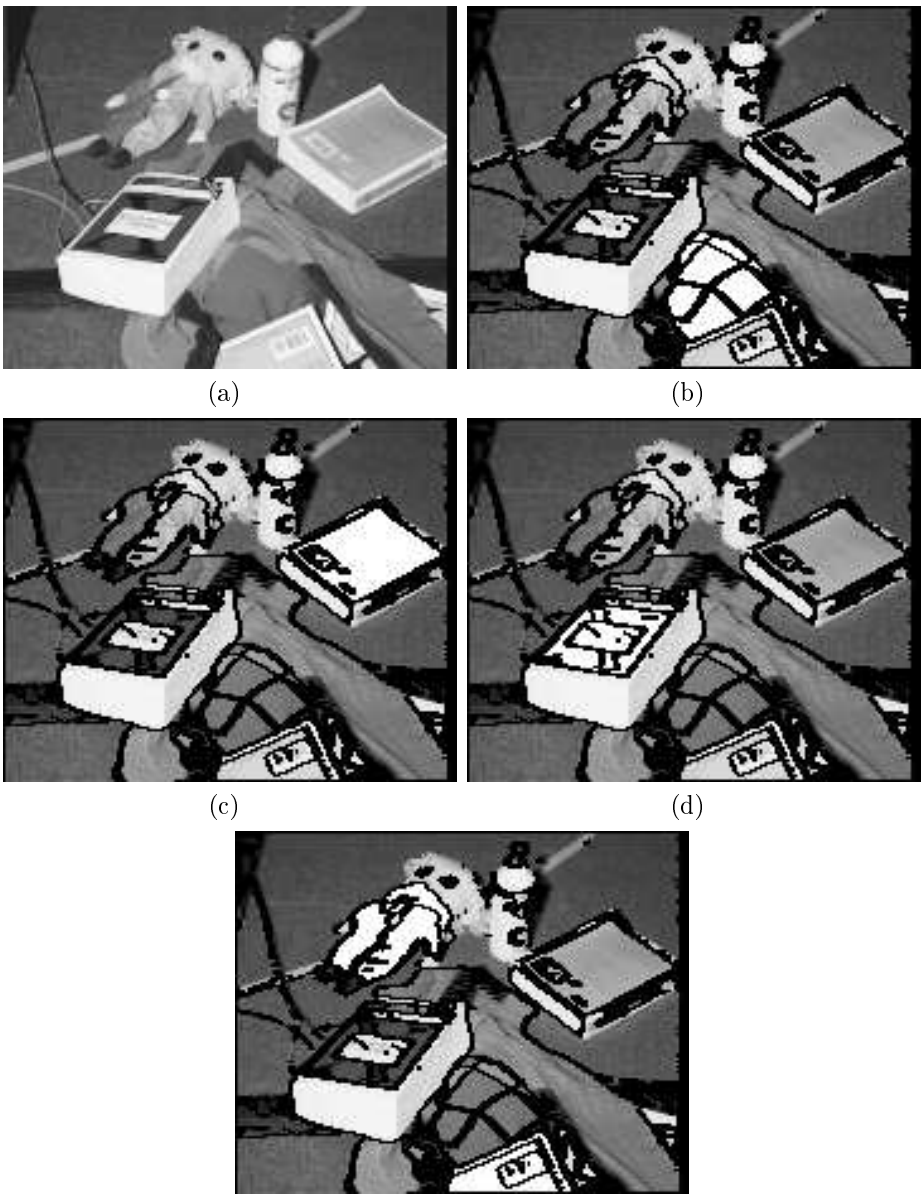


(a) (b) (c)

(d) (e) (f)

Figure 6.2: Experiment 1. (a-f) Test sequence.

Figure 6.3: Experiment 1. (g-l) Regions matched to model 'sweater' (see fig. 6.1(i)) are shown in white. Clustering results are visualised by setting to black the areas corresponding to pixels removed on the boundaries of colour components (see sect. 4.3.2). Successful recognition and localisation of the 'sweater' in images (a-e) is shown in (g-k). No false positive is reported in (l). Subimage (m) shows the answer to the query 'Any other known objects in (d)?'. No other objects are present, therefore nothing is highlighted. (n) Answer to the same query for (e). The 'pyjama' object is found and highlighted. (o-p)Answer to the same query, image (f). Both the 'manikin' and 'pyjama' objects are found. Black trousers are incorrectly recognised as part of the 'pyjama' (see text).



(a) (b) (c) (d) (e)

Figure 6.4: Experiment 2. (a) The scene. (b-e) Recognised objects: 'sweater' (b), 'manual6' (c), 'manual3' (d), 'manikin' (e). Neither the bottle nor any of the other objects in the scene whose models are not in the model database are matched, ie. there are no false positives.

(a) (b) (c)

Figure 6.5: Experiment 3 (see also 5.1). (a) Original. (b) Recognised object 'pyjama'. (c) Recognised object 'sweater'. Note that not all colour patches belonging to the objects are matched.

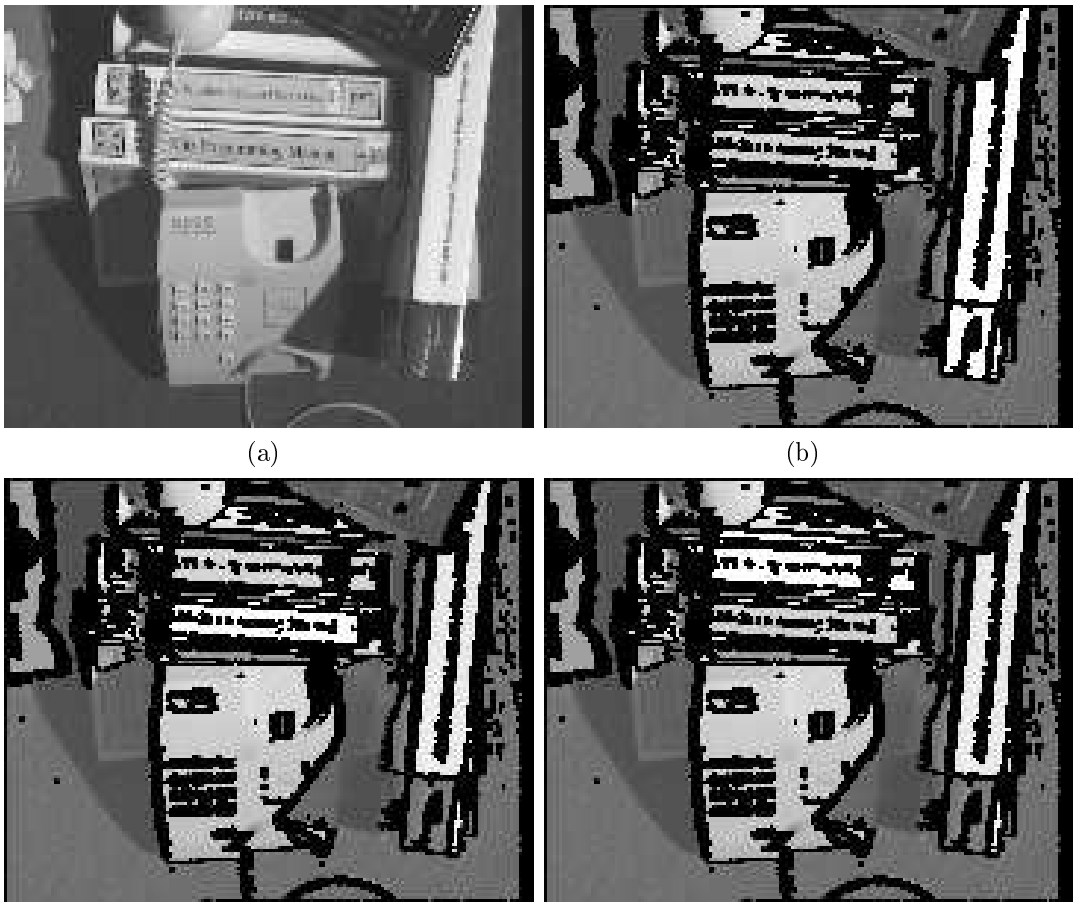


Figure 6.6: Experiment 4. (a) The scene. (b-d) Recognised objects: 'ICPR proceedings' (b), 'manual2' (c), 'manual5' (d).

Chapter 7

Attention-driven Illumination Invariant Recognition

7.1 Introduction

So far in the thesis we have adopted a traditional approach to recognition: we developed a recognition strategy based solely on colour information. A single static image was processed at a time and the quality of recognition was judged by the number of instancies of objects with models in the database that were correctly matched. However, it is easy to envisage situations where *cooperating* recognition systems exploiting different properties of objects outperform any of the individual systems operating in *isolation*[1].

In this chapter we describe a colour-based attention system successfully tested in a continuously operating, integrated cooperative system [CC95]. In such a system the ability of a recognition engine to *exhaustively* search the image for all instances of known objects is often less important than the capability to focus computational resources on areas of competence, where reliable matching is accomplished at high speed. In the particular case of the colour recognition system, its attention mechanism attempts to avoid analysing the whole image in the spirit of traditional methods; instead it tries to find and select for further processing areas where distinctive colour provides least ambiguous information. Using this approach, standard recognition tasks can be accomplished without performing a time-consuming analysis in parts of the image where pixel colour analysis is complex, eg. where mutual illumi-

[1]For references and more detailed discussion see chapter 2, page 21.

nation effects or specularity must be taken into account. Hopefully, these regions will have been efficiently interpreted by a different recognition engine, eg. shape-based or texture-based, before the attention of colour recognition system decides to focus on these regions. Besides improved efficiency, the focusing strategy facilitates cooperation. In a system where multiple agents running in parallel cooperate to recognise objects, it is essential to output reliable results as soon as possible to enable other interpretation agents to exploit the information. The attention mechanism, the overall structure of the colour recognition system and the region growing method that expands interest points into object hypotheses is described in section 7.3.

From the point of view of colour information processing, the main contribution of the work presented in this chapter lies in the suitability of the proposed approach for environments with *spectrally varying illumination*. The assumptions of the proposed method are orthogonal to those made by standard colour constancy methods (eg. [MB86]). We restrict the set of permitted 'canonical' illuminants to a small number and require that all objects in the model database are imaged in the model-building process under all 'canonical' illuminants. On the other hand, unlike the general colour constancy algorithms (eg. [For88] [Mal86][TO90][FDF93]), we neither impose any restrictions on the geometry of the scene nor require the whole scene to be lit by illumination with constant spectral power distribution.

In experiments described in this chapter objects are represented by a list of colour patches[2]. The illumination-independent representation can be integrated in the CAG approach by treating the 'canononical chromaticities' (described bellow) as complex attributes of the CAG nodes. A description of each patch for a number of 'canonical' illuminants ('canonical chromaticity') is stored in a database. Automatic model acquisition is generally desirable for any model based system. Considering the number of objects (> 50), patches and canonical illuminants ($4-6$) it is clearly necessary. We used the GT CLUSTERING, followed by a simple database maintenance step, to accomplish the task.

The rest of the chapter is structured as follows. In section 7.2 we first introduce formulas modeling the dependency of observed colour on the following factors - spectral reflectance of the viewed object, spectral distribution, intensity and rel-

[2]Limitations of the 'list-of-colour-patches' methods are discussed in section 3.2.

ative position (photometric angles) of illumination sources. We show, under the assumption of a limited number of known illumination sources, that the effects of changing illumination and geometry can be predicted allowing recognition in environments with spectrally variable illumination (in both time and space); we are not aware of a system with similar capabilities. Moreover, we do not impose any restriction on spectral reflectance of objects (eg. low dimensionality, low pass function of wavelength).

The attention mechanism, the overall structure of the recognition system and the region growing method that expands interest points into object hypotheses is described in section 7.3. Experiments on two test images are presented in section 7.4. The contribution of the work discussed in the chapter is summarised in section 7.5.

7.2 Surface reflectance, geometry and illumination

Sensor response of a standard imaging device is well modelled by a spectral integration process:

$$p_k^X = \int_{\lambda_1}^{\lambda_2} \rho_k(\lambda) L(\lambda, .) d\lambda, \qquad k = 1, \cdots, n \tag{7.1}$$

where p_k^X is the response of the $k-th$ sensor at location X of the sensor array, $L(\lambda, .)$ is the light emitted from the surface patch that is projected on pixel X, and $\rho_k(\lambda)$ is the responsivity function of the $k-th$ sensor. The n-dimensional vector $\underline{p}^X$ will be referred to as pixel value, pixel colour or object colour, assuming that light from a surface patch belonging to a single object falls on pixel X (the results of this section apply to any pixel; we therefore simplify expressions by dropping X from p_k^X). Besides the optical properties of the patch, the spectral power distribution (SPD) of $L(\lambda, .)$ depends on the the relative position of the patch, illumination source(s) and the viewpoint (defined by photometric angles) as well as on the spectral power distribution of the illumination source(s). Clearly, any system making use of pixel colour for recognition must separate the dependence of $\underline{p}$ on object material from the effects due to changes in illumination or geometry.

The relationship between illumination SPD and pixel colour has been studied by researchers interested in colour constancy [For88] [Mal86][TO90]. Unfortunately,

the theory of colour constancy is developed mainly in the context of the Mondrian world, ie. a world consisting of a single planar surface composed of a number of matte (Lambertian) patches. Light striking a Mondrian world is assumed to be spectrally unchanging and of uniform intensity [FDF93]. Under such conditions light reflected from a patch is independent of viewing geometry and can be expressed as

$$L(\lambda) = S(\lambda)E(\lambda) \tag{7.2}$$

$S(\lambda)$ is a surface reflectance function of the patch and $E(\lambda)$ is the (global) illumination SPD. Furthermore, surface reflectances and illuminant SPDs are approximated by a weighted sum of basis functions $S_j(\lambda)$ and $E_i(\lambda)$ respectively:

$$S(\lambda) = \sum_{j=1}^{d_S} S_j(\lambda)\sigma_j \qquad E(\lambda) = \sum_{i=1}^{d_E} E_i(\lambda)\epsilon_i \tag{7.3}$$

Substituting equations (7.3) and (7.2) into equation (7.1) we obtain:

$$\underline{p}_k = \sum_{j=1}^{d_S} \sigma_j \sum_{i=1}^{d_E} \epsilon_i \int_{\lambda_1}^{\lambda_2} S_j(\lambda)E_i(\lambda)\rho_k(\lambda)d\lambda \tag{7.4}$$

where the expression inside the integral depends only on the the sensor responsivity $\rho_k(\lambda)$ and the choice of basis functions for illumination and reflectance. Equation (7.4) lies at the heart of most colour constancy algorithms; variations exist in the assumed number of sensors at each pixel and the dimensionality of the illuminant and reflectance spaces (d_E and d_S respectively). From eq. (7.4) it can be seen that if a representation of a spectral reflectance in terms of the vector of mixing weights $\underline{\sigma}$ is known, then object colour can be computed for any illuminant described by $\underline{\epsilon}$.

Unfortunately, comparatively little work has been carried out to establish the applicability of the low dimensionality assumptions. Surface reflectance were studied by Maloney [MB86] who analysed 337 spectral reflectances of natural formations collected by Krinov. Maloney concludes that five to seven basis functions provide an almost perfect fit. In our opinion Maloney's results are difficult to interpret. On the one hand, the quality of the fit of the first three basis function seems sufficient for computer vision applications. On the other hand it is unlikely that the same basis functions are applicable to a larger set of natural and man-made objects. In contrast, SPDs of a number of artificial illuminants are known. Furthermore, three

basis functions providing practically a perfect fit to all phases of daylight have been found [WS82].

The effects of geometry on SPD of reflected light have been extensively studied [HB87],[KSK87]. The dichromatic reflection model of [Sha84] is generally regarded to be accurate for a large class of materials [Tom91]. The dichromatic model states that reflected light L consists of two independent components: light reflected on the interface and light due to sub-surface (body) reflection. Furthermore it is assumed that the SPD of neither of the two components depends on geometry. Therefore:

$$L(\lambda, g) = m_i(g)L'_i(\lambda) + m_b(g)L'_b(\lambda) \tag{7.5}$$

where g denotes the geometry (ie. the photometric angles), $m_i(g)$ and $m_b(g)$ are scaling factors and $L'_i(\lambda)$ and $L'_b(\lambda)$ are the relative spectral distributions of light reflected by interface and body reflection respectively. The quantities $L'_i(\lambda)$, $L'_b(\lambda)$ depend only on the surface reflectance and relative illuminant SPD. Besides geometry, the scaling factors model absolute changes of illumination intensity. The dichromatic model does not specify *how* quantities L' depend on illumination and spectral reflection, therefore its application always requires the assumption of spectrally unchanging illumination SPD.

For a colour based recognition system operating in the real world the assumptions imposed by colour constancy algorithms are too restrictive. We adopt a weaker set of assumptions, with a single exception of modelling surface reflection by a monochromatic reflection model:

$$L(\lambda, g) = m(g)L'(\lambda) \tag{7.6}$$

In our opinion, the simplification is justifiable for a number of reasons. In case of metals the model is equivalent to the dichromatic reflection model. For dielectrics we neglect the specular component. Moreover, specularities almost always cover only a fractional part of an image. Very often the high intensity of specular points saturates the sensor making colour analysis meaningless. And finally: the attention mechanism (see 7.3) will skip over specularities as their colour is inherently more ambiguous then areas of diffuse reflection.

To predict the effects of changing illumination we substitute for L' from eq. (7.2) into eq. (7.5).

$$L(\lambda) = m(g)S(\lambda)E'(\lambda); \tag{7.7}$$

where $E'(\lambda)$ is the relative SPD of illumination. Assuming low dimensionality of illuminant SPD but a number of sources $j = 1 \ldots N_s$ (with different SPDs) we obtain after substituting in (7.1):

$$p_k = \sum_{j=1}^{N_s} m_j(g) \sum_{i=1}^{d_E} \epsilon_{ij} \int_{\lambda_1}^{\lambda_2} S(\lambda)E_i(\lambda)\rho_k(\lambda)d\lambda \tag{7.8}$$

where $m_j(g)$ is a scaling factor covering the effects of change in the illumination intensity and geometry of the j-th source, ϵ_{ij} defines the j-th source SPD in terms of the basis function $E_i(\lambda)$. Replacing the integral (which is independent of geometry and illumination illumination) we obtain:

$$p_k = \sum_{j=1}^{N_s} m_j(g) \sum_{i=1}^{d_E} \epsilon_{ij} p_k^{E_i} \tag{7.9}$$

where $p_k^{E_i}$ is the colour of the object with spectral reflectance $S(\lambda)$ as seen under illuminant E_i. Because both $m_j(g)$ and ϵ_{ij} are non-negative, pixel colour $\underline{p}$ will lie inside a convex polyhedron - a pyramid with an apex at $\underline{0}$ and edges coincident with vectors $p_k^{E_i}$. In the case of artificial illumination we *choose* illuminant SPD as the basis function E_i. Values $p_k^{E_i}$ are obtained in a straightforward manner as they are identical to object colour under the given artificial illumination. The three $p_k^{E_i}$ corresponding to pixel colour under daylight are approximated by points on the convex envelope of $\underline{p}$'s from images taken under different daylight conditions.

Equation (7.9) shows that the absolute value of p_k arbitrarily changes with $m_j(g)$ and ϵ_{ij} and therefore carries no information about the object. Therefore we project $\underline{p}$ onto the chromatic plane (projection on a unit sphere or parameterisation by hue and saturation would achieve the same objective). The pyramid is projected on a convex polygon whose vertices correspond to canonical illuminants. Such polygons serve as our model of object colours - pixel colour of a surface patch with *arbitrary* surface reflectance will lie inside a convex polygon (regardless of geometry) if illuminated by a light source well approximated by the set of basis functions we adopted.

With the presented approach we are able to recognise object colour under any mixture of canonical illuminants. Moreover, we do not have to assume that the

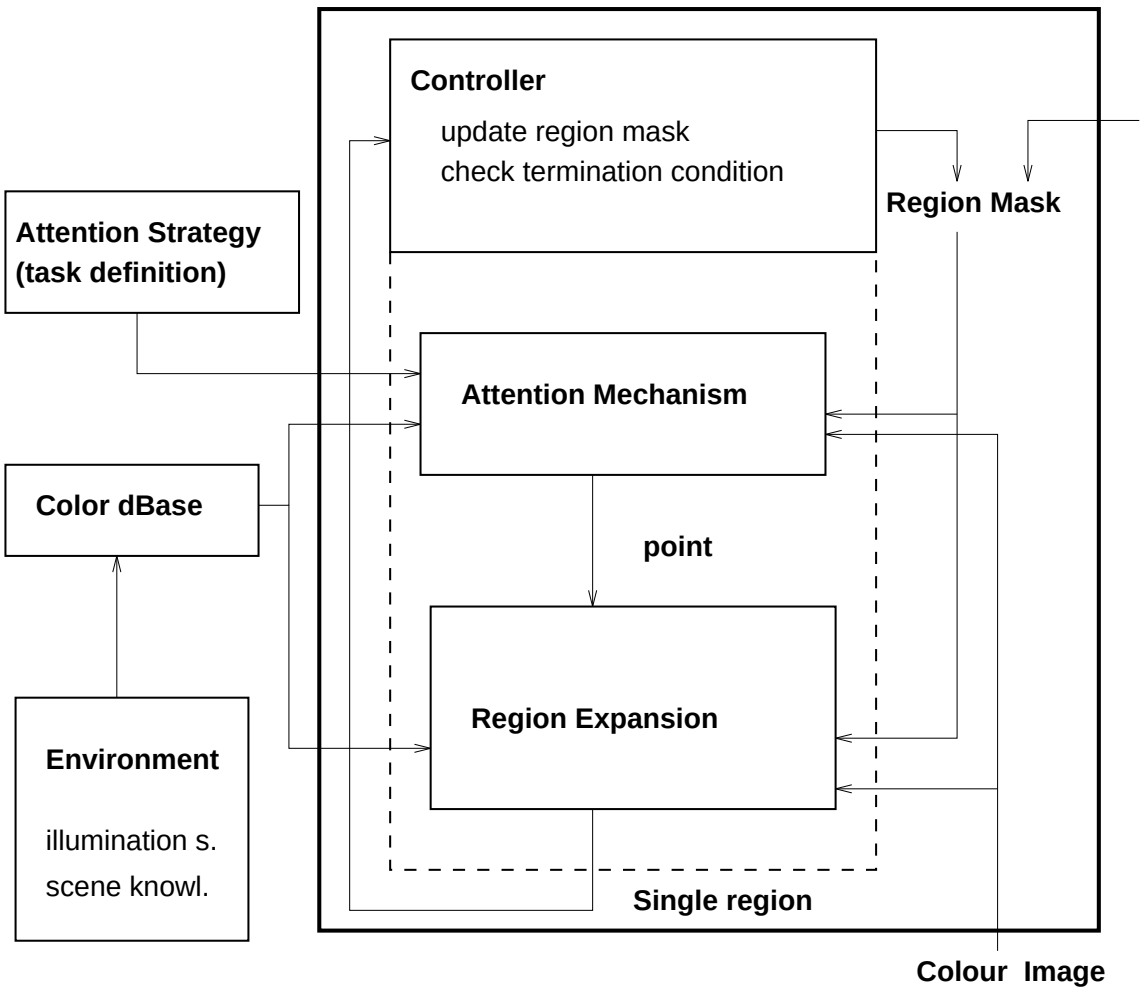


Figure 7.1: System structure

whole scene is illuminated by a single source (imagine a common indoor environment with windows and lit lamps). Information about illumination source can be easily incorporated into the scheme; only vectors $\underline{p}$ corresponding to the illuminants known to be present in a given environment are taken into account when computing the convex hull.

7.3 Overall structure

The overall structure of the colour recognition system is depicted in figure 7.1. We will first focus on the most complex part of the system - the attention mechanism.

The pixel location on which the module focuses its attention depends, besides the input data, on the definition of the recognition task, the contents of colour database, model of the environment and the current state of the region mask. The region mask serves two purposes. Pixels already processed are marked in the mask and are ignored in further processing. The mask can be set up prior to processing to control in a natural way the image area on which colour processing is performed.

The model of environment allows the system to exploit specific information about the current scene and illumination. The environment defines 1. a set of objects then can appear in the scene and 2. a set of permissible illuminants. The colour database manager builds a chromatic model for every patch of every specified object. Each patch is characterised by a convex polygon in the chromatic plane (defined in section 7.2) with vertices corresponding to patch colour under canonical illumination. At present the system can cope with scenes with up to five illuminants for which patch chromaticities are stored in the colour database (tungsten filament lamp, fluorescent lamp and three for daylight). So far we tested the approach in scenes which were lit by natural light and one artificial illuminant.

The recognition process adopted in our system represents a significant departure from our earlier work [MMK93] where Bayesian decision rule was invoked at every pixel in the image. Here prior to recognition the colour database is transformed into a look-up table. The table representing a discretised chromatic plane stores at each cell a list of patch labels that can assume the chromaticity. The conditional probability $P(\textit{patch-label}|\textit{chroma})$ is computed for each chromatic cell and the labels sorted according to it. Briefly the process can be described as follows. A set of points where $P(\textit{chroma}|\textit{patch-label})$ is non-zero is obtained by rasterising the chromatic polygon of the patch. The rasterised polygon is convolved with a smoothing filter to model effects of noise. The shape of the smoothing kernel depends on the shape of the distribution detected in the colour histogram during model acquisition. Values in the raster are taken as approximations of $P(\textit{chroma}|\textit{patch-label})$. The conditional probability $P(\textit{patch-label}|\textit{chroma})$ is finally computed from probabilities $P(\textit{chroma}|\textit{patch-label}_i)$ of all patches (with non-zero $P(\textit{chroma}|\textit{patch-label}_i)$) and the probability $P(\textit{chroma}|\textit{background})$.

The chromatic look-up table enables the attention mechanism to implement efficiently recognition strategies best described as $Where$ (is object X) and $What$ (is in

the scene). In the case of $Where$ the attention selects a pixel with chromaticity that maximises $P(patch\text{-}label_X|chroma)$. Note that although a $list$ of labels is stored at every chromatic cell only the first item of a suitably sorted list need be accessed.

Once the point of interest is selected a standard region growing algorithm [Mar92] is employed to detect a region. The region, together with a list of labels with probabilities, is passed to the controller. The controller updates the region mask and decides whether to terminate processing or run another iteration of the attention - region expansion loop.

The colour recognition system is implemented as a set of AVS modules [AVS93, AVS92a, AVS92b]. The core of the system, shown inside the bold box in figure 7.1, is implemented as a single AVS module. In the integrated VAP system [CC95], the 'Attention Strategy' module as well as the image and region mask inputs were provided by AVS modules developed at the University of Linkoping. The principal advantage of the AVS system is the simplicity with which modules can be linked to form complicated networks. Thus, in preliminary experiments carried out at the University of Linkoping [Wes95], the colour recognition module was first run inside a network where colour images were generated by a virtual environment modeller. A trivial change of the network was needed to carry out the experiments on real images described in sections 7.4.

7.4 Experiments

We demonstrate performance of the colour recognition system on images shown in figures 7.2(a) and 7.3(a). The colour database contained 51 models depicted in fig. 7.4. The environment model specified that daylight and fluorescent light could be present in the scene. The termination strategy was set as 'find n-best regions'.

The first experiment (fig. 7.2) shows performance of the recognition system under most favourable conditions reminiscent of the Mondrian world. The objects are placed close to each other on a single plane; it is therefore likely that they are illuminated by light of unchanging spectral distribution. The sequence of points of attention is shown in fig. 7.2(d). The points, together with regions into which they were expanded, are depicted in fig. 7.2(c). The sequence starts on the red trunks. Next another part on the inside of the trunks is picked (this part is separated from

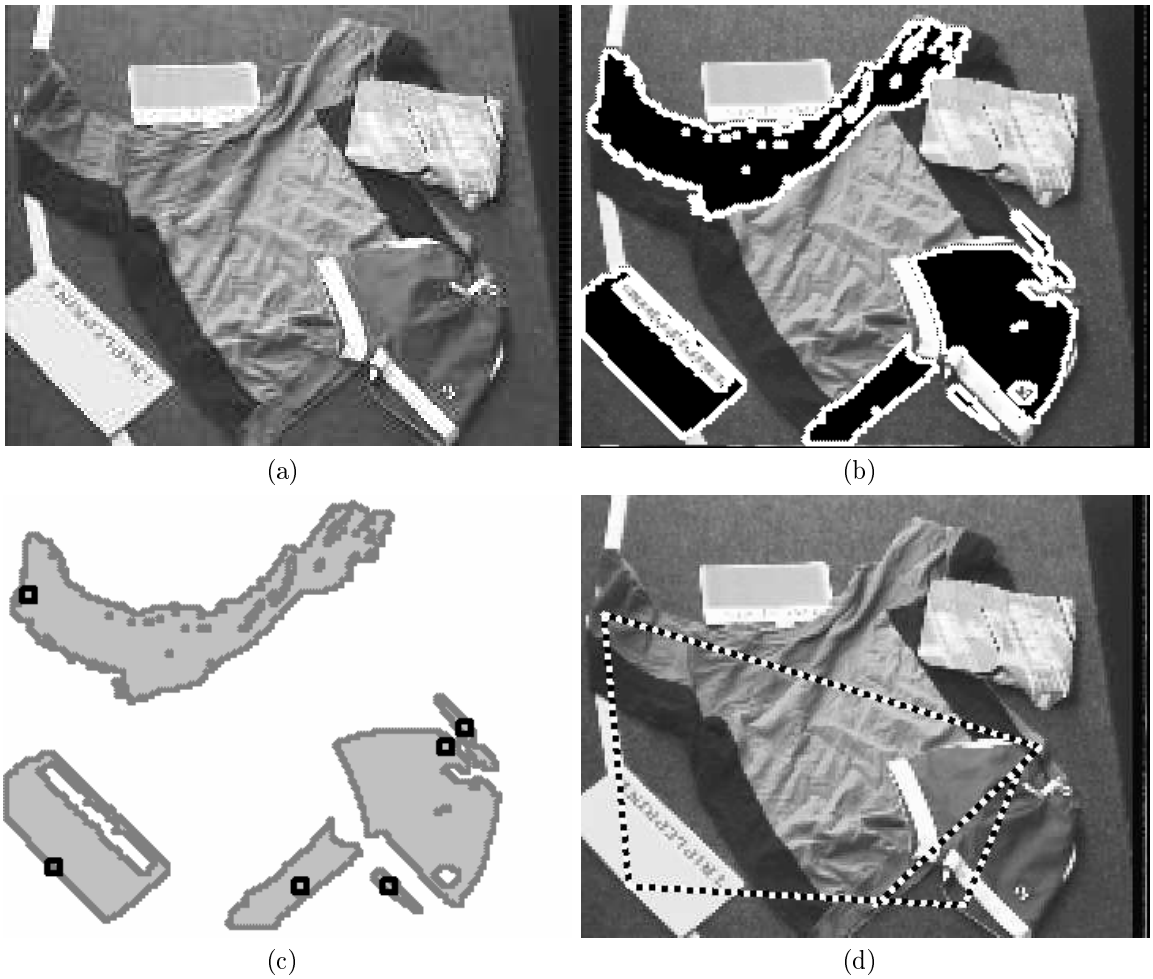


Figure 7.2: Experiment 1. (a) Image of the scene. (b) Regions expanded from the six first attention points superimposed on the image. (c) Regions with attention points. (d) Movement of the point of attention. The sequence starts on the trunks in the left part of the image.

the main part of the trunks by a white stripe). The attention point then moves clockwise to the yellow envelope and the purple sleeve. Next a small triangular patch to the left of the trunks is detected. This patch is a part of the back side of the pyjama top and has the same colour. The process was terminated after focusing on the bottom purple part of the pyjama top (the small triangular patch). Fig. 7.2(c) shows the position of the attention points in the expanded regions. It might appear counterintuitive but there is no reason why the points should lie near the region centers as the colour near the center (of a sufficiently large object) is on

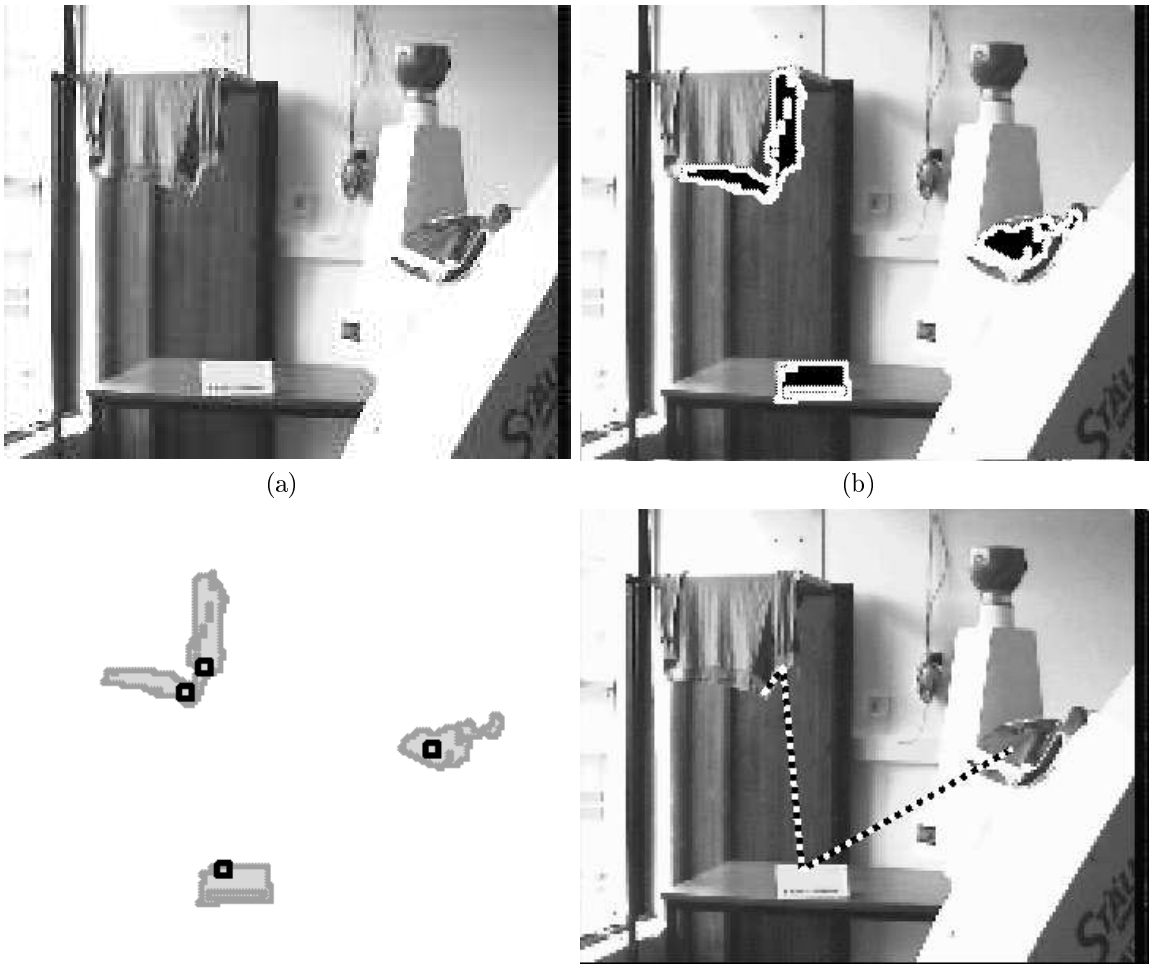


Figure 7.3: Experiment 2. (a) Image of the scene. (b) Regions expanded from the four first attention points superimposed on the image. (c) Regions with attention points. (d) Movement of the point of attention. The sequence starts on the trunks on the PUMA robot arm.

average equally likely to provide the point with least ambiguous chromaticity (in case of a tie the top-left pixel is chosen). With the exception of the group of holes on the top part of the pyjama top the expanded regions fit well the the image data. The holes correspond to almost black points on the dark side of the creases.

The second experiment was carried out using the same objects as in experiment one (fig. 7.3). However, the complexity of the analysis is increased by two factors. Firstly, the proportion of daylight to fluorescent light is higher for objects closer to

the window (on the left edge of fig. 7.3(a)). Secondly, objects are not placed on a planar surface. Results of the second experiment are presented in the same way in fig. 7.3. The point of attention shifted from right to left, ie. from the trunks to the pyjama top.

7.5 Conclusions

This chapter draws mainly on the results published in [MMK94] and [MMK95]. It describes a colour-based attention system successfully tested in a continuously operating, integrated cooperative system [CC95]. The main contributions of the approach are:

- capability to operate under illumination with spectral distribution varying in both space and time
- adoption of a recognition strategy with focus of attention

Experiments reported in [CC95] and [KMBN95] show that benefits in terms of speed and quality of output are gained by focusing processing to areas of salient colour.

References

[AVS92a] *AVS Developer's Guide*, volume 320-0013-02, 300 Fifth Ave., Waltham, MA 02154, USA, May 1992. Advanced Visual Systems Inc.

[AVS92b] *AVS User's Guide*, volume 320-0011-02, 300 Fifth Ave., Waltham, MA 02154, USA, May 1992. Advanced Visual Systems Inc.

[AVS93] *AVS Module Reference*, volume 320-0014-03, 300 Fifth Ave., Waltham, MA 02154, USA, February 1993. Advanced Visual Systems Inc.

[CC95] J Crowley and H I Christensen. *Vision as Process.* Springer-Verlag, Berlin, 1995.

[FDF93] Graham D. Finlayson, Mark S. Drew, and Brian V. Funt. Diagonal transforms suffice for color constancy. In *Fourth International Conference on Computer Vision (Berlin, Germany, May 11-14, 1993)*, Washington, DC., 1993. IEEE Computer Society Press.

Figure 7.4: Fifty selected objects from the database.

[For88] D.A. Forsyth. A novel approach to colour constancy. In ICCV88 [ICC88], pages 9–18.

[HB87] G. Healey and T.O. Binford. Local shape from specularity. In ICCV87 [ICC87], pages 151–160.

[ICC87] *First International Conference on Computer Vision, (London, England, June 8–11, 1987)*, Washington, DC., 1987. IEEE Computer Society Press.

[ICC88] *Second International Conference on Computer Vision (Tampa, FL, December 5–8, 1988)*, Washington, DC., 1988. Computer Society Press.

[KMBN95] J. Kittler, J. Matas, M. Bober, and L. Nguyen. Image interpretation: Exploiting multiple cues. In *International Conference on Image Processing and Applications, Edinburgh, July 3-6*, pages 1–5, 1995.

[KSK87] G.J. Klinker, S.A. Shafer, and T. Kanade. Using a color reflection model to separate highlights from object color. In ICCV87 [ICC87], pages 145–150.

[Mal86] L.T. Maloney. Evaluation of linear models of surface spectral reflectance with small number of parameters. *J. Opt. Soc. Am.*, 3:1673–1686, Octobe 1986.

[Mar92] R. Marik. Regions and image segmentation. Technical report, PhD. thesis, Department of Control Engineering Czech Technical University, Prague, Czechoslovakia, 1992.

[MB86] L.T. Maloney and B.A.Wandell. Color constancy: a method for recovering surface spectral reflectance. *Journal of the Optical Society of America*, 3:29–33, January 1986.

[MMK93] J. Matas, R. Marik, and J. Kittler. Generation, verification and localisation of object hypotheses based on colour. In J. Illingworth, editor, *British Machine Vision Conference*, pages 539–548. BMVA Press, 1993.

[MMK94] J. Matas, R. Marik, and J. Kittler. Illumination invariant colour recognition. In E. Hancock, editor, *British Machine Vision Conference*. BMVA Press, 1994.

[MMK95] J. Matas, R. Marik, and J. Kittler. Illumination invariant colour recognition. *IVC*, 1995. to appear.

[Sha84] S.A. Shafer. Using color to separate reflection components. Technical report, University of Rochester, 1984.

[TO90] M. Tsukada and Y. Ohta. An approach to color constancy using multiple images. In *Third International Conference on Computer Vision (Osaka, Japan, December 4–7, 1990)*, pages 385–389, Washington, DC., 1990. IEEE Computer Society Press.

[Tom91] S. Tominaga. Surface identification using the dichromatic reflection model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:658–670, June 1991.

[Wes95] Westelius95. *Focus of Attention and Gaze Control for Robot Vision*. PhD thesis, University of Linkoping, 1995.

[WS82] G. Wyszecki and W. S. Stiles. *Color Scence: Concepts and Methods, Quantitative Data and Formulea*. John Wiley, 1982.

Chapter 8

Conclusions and Future Work

8.1 Summary and Conclusions

In the first chapter of the thesis we introduced the problem of colour-based object recognition. An intuitive example was used to highlight the issues related to the problem and to informally present the approach developed in the thesis. We discussed the evolution of colour-related research within the field of computer vision, focusing in particular on colour-based recognition systems. We then stated the objectives and outlined and the structure of the thesis.

The short chapter on performance characterisation of a colour recognition system built a framework that allowed us to judge and compare the generality of the reviewed colour recognition systems and the CAG methods proposed in the thesis. We believe that the analysis of factors that influence colour recognition is of interest in its own right, especially because it is omitted in standard computer vision textbooks [CPW93, SHB93], where the treatment of colour-related issues is confined to image processing topics.

Chapter 3 provided a literature survey of several approaches to colour-based object recognition with a focus on methods that exploit colour *structure* of multi-coloured objects. We started by presenting a number of algorithms designed for recognition of single-coloured objects. Methods that represent multi-coloured objects as collections of single-coloured surface patches were then considered. We discussed the limitations of this class of methods and showed that they fail to capture invariant information computable on the boundaries of patches with different surface reflectances. We reviewed in detail histogram-based methods and a method

for matching region adjacency-graphs. We analysed the histogram backprojection method from a new point of view and showed its equivalence to a simplified form of evaluating the Bayesian aposteriori probability.

A new representation for objects with multiple colours - *the colour adjacency graph* - has been proposed in chapter 4. Each node of the CAG represents a single chromatic component of the image defined as a set of pixels forming a unimodal cluster in the chromatic scattergram. Edges encode information about adjacency of colour components and their reflectance ratio. The CAG has been shown to be related to both the histogram and region adjacency graph representations. It preserves and combines the best features of these two approaches while avoiding their drawbacks. In particular, the nodes of the CAG correspond to peaks of the chromatic histogram. However the locations of the maxima, unlike the associated bin counts, are stable under viewpoint change and are robust to occlusion. Experimental results confirmed robustness of the CAG representation predicted by the surface reflectance analysis carried out in section 4.2. In comparison with the region adjacency graph, the CAG is significantly simpler, facilitating efficient matching. Moreover, its edge attributes (two chromaticities and reflectance ratio) are very rarely ambiguous. It has been shown that the CAG can be generated very efficiently by means of theoretical graph clustering. Object recognition is achieved by subgraph matching of the image and model CAGs. Multiple instances of the same object or several distinct objects can be recognised and segmented out simultaneously.

The proposed CAG approach has been applied to a representative range of images containing rigid and nonrigid objects under varied viewing conditions, and subject to occlusion, change of scale, specular reflection distortion and 3D aspect change ie. most of the environmental changes affecting object appearance listed in table 2.1. The results of the experiments presented in chapter 6 showed that the method for colour recognition via CAG matching performs well, achieving high recognition rate while keeping false positives to the minimum. The timing information presented demonstrated that the method is computationally efficient.

In chapter 7 we described a colour-based attention system successfully tested in a continuously operating, integrated cooperative system [CC95]. In such a system the ability of a recognition engine to *exhaustively* search the image for all instances of known objects is often less important than the capability to focus computational

resources to areas of competence, where reliable matching is accomplished at high speed. As shown in experiments reported in [CC95] and [KMBN95], benefits in terms of speed and quality of output are gained by focusing processing to areas of salient, ie. least ambiguous, colour. From the point of view of colour information processing, the main contribution of the work presented in this chapter lies in the suitability of the proposed approach for environments where spectral distribution of illumination varies in both space and time.

In appendix A we proposed a novel approach to image segmentation, called *feature and spatial domain* clustering. The method is devised to group pixel data by taking into account simultaneously both their feature space similarity and spatial coherence. The FSD CLUSTERING algorithm is practically application independent. It has been successfully tested on a wide range of image segmentation problems, including grey and colour image segmentation, edge and line detection, range data and motion segmentation. In comparison with existing segmentation approaches, the method can resolve image features even if their distributions significantly overlap in the feature space. It can distinguish between noisy regions and genuine fine texture. Moreover, if required, FSD CLUSTERING can produce partial segmentation by identifying salient regions only.

8.2 Lessons from Past and Ideas for Future Research

The importance of a suitable sensor and acquisition equipment in general for colour related research was one of our hard-earned lessons. The low cost television-standard equipment commonly used in computer vision experiments is perhaps suitable for analysis of intensity images. The low spatial resolution of the chromatic information, that was added ex post to the composite TV signal, is just one of the problems. Non-linearities, arbitrary offsets and colour system transformations [Hut71] preclude, in our opinion, the use of a TV colour camera as a scientific measurement instrument. The dynamic range of commonly used grabbers (eg.[abe93, vid91]), theoretically at two hundred and fifty-six levels but practically less than two hundred levels, similarly limits experimentation. We found that images containing bright yellow, orange or green objects next to dark blue or purple objects could not be acquired without

either the former being saturated or the latter being effectively black.

In our future work we would like to investigate whether the dynamic range problem can be overcome by the method of varying camera aperture proposed by Asada and Matsuyama [AM92]. As the name of the method suggests, a colour image of high dynamic range is acquired by combining a sequence of images taken by a static camera with different aperture settings. A motorised zoom lens controlled from a desktop computer via a controller developed recently at the University of Surrey by Steve Procter provides the hardware support for the experiment.

In chapter 2 we list a number factors that influence object appearance. The set of possible appearances of an object is potentially large. But if we take into account occlusions, types of background, differences in the model database etc., we must conclude that the space of possible recognition experiments is practically limitless. In experiments presented in chapter 6 we sampled this 'recognition' space, trying to place our probes at distant corners of this abstract space. In our future research we would like to investigate methods that allow to explore the space systematically and on a finer scale. The easier approach, advocated recently by Terzopoulos [Ter95], would immerse the recognition system into a virtual world rendered by a computer graphics system. Rather than following the virtual reality path, we would prefer to carry out the technically more demanding 'real world' simulations where random scenes are created by a robot manipulator. The automatic approach to experimentation is, in our opinion, the only way to gather significant amounts of data sufficient for statistical evaluation of a recognition method. Moreover, the ground truth is known in both types of simulations, which opens the opportunity for automatic detection of matching errors. Consequently, the development of a system with feedback and/or learning capabilities may be contemplated.

The last research direction we will mention is the issue of adaptive choice of colour recognition strategy. In certain circumstances, simple single-colour-based method is capable of successfully recognising the object of interest at very high speed. On the other hand, as discussed in the 'flag' example (page 52), the information contained in the CAG is not always sufficient for discrimination of two objects. We feel that a combined strategy starting with the fastest but weakest method followed by sequence of progressively more sophisticated, but also more time consuming, fall-back methods may form a basis of a very fast (on average) recognition system with

very high performance.

References

[abe93] *A26 D1-Analogue Converter Manual*, 12 Portman Road, Reading, Berkshire RG3 1EA, England, 1993. Abekas Video Systems Ltd.

[AM92] N. Asada and T. Matsuyama. Color image analysis by varying camera aperture. In *11th IAPR International Conference on Pattern Recognition (The Hague, The Netherlands, August 30–September 3, 1992)*, volume A, pages 466–469, Washington, DC, 1992. IEEE Computer Society Press.

[CC95] J Crowley and H I Christensen. *Vision as Process.* Springer-Verlag, Berlin, 1995.

[CPW93] C. H. Chen, L. F. Pau, and P. S. P. Wang, editors. *Handbook of pattern recognition and computer vision.* World Scientific, 1993.

[Hut71] G.H. Hutson. *Colour television theory: PAL system principles and receiver cirucitry.* 1971.

[KMBN95] J. Kittler, J. Matas, M. Bober, and L. Nguyen. Image interpretation: Exploiting multiple cues. In *International Conference on Image Processing and Applications, Edinburgh, July 3-6*, pages 1–5, 1995.

[SHB93] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis and Machine Vision.* Chapman, 1993.

[Ter95] Terzopoulos95. Animat vision: Active vision in artificial animals. In *Fifth International Conference on Computer Vision (Cambridge, MA, June 20–23, 1995)*, pages 801–815, Washington, DC., 1995. Computer Society Press.

[vid91] *Using VideoPix*, 2550 Garcia Avenue, Mountain View, California 94043-1100, USA, 1991. Sun Microsystems.

Appendix A

Spatial and Feature Space Clustering

A.1 Introduction

Many low and intermediate level image analysis tasks in computer vision are essentially image segmentation problems. The aim of image segmentation is to identify, in the image, regions which can be associated with perceptually meaningful scene primitives. This can mean either a complete image partitioning or the extraction of a subset of such regions selected according to some criteria. As pixels in such regions represent the same physical properties or phenomena, they are expected to exhibit a degree of similarity, i.e. a clustering tendency, in the feature space characterising the relevant property. However, a distinctive aspect of image data is its spatial ordering. Thus in the data analysis terms, image segmentation is a process of clustering *spatially indexed data*. Consequently, the grouping of pixels into clusters must take into account not only their similarity in the feature space but also the requirement of their spatial coherence.

Classically, image segmentation problems have often been viewed as purely feature space clustering problems [JD88]. Accordingly, the vector of image features observed at each pixel is considered as a point in the feature space. Although spatial coherence is not a constraint explicitly built in, each cluster in the feature space is expected to group pixels which on the grounds of their homogeneous properties will come from coherent regions in the image. For low dimensional feature spaces the clustering can be performed efficiently from a statistical summary of the data in

terms of the histogram. By analysing the histogram one can determine the groups of pixels satisfying the homogeneity property and thus obtain the corresponding segmentation.

The global statistical analysis of the image data in the feature space has the advantage of providing a good assessment of the clustering tendency of the data. However, feature space clustering does not automatically guarantee spatial coherence of the cluster based pixel groups. If measurements overlap in the feature space, the segmentation results obtained by this approach can be disappointing, giving a very noisy appearance. This problem has been tackled in the literature in a number of different ways. One possibility is to augment the feature vector either by measurements on the neighbouring pixels, or by pixel coordinates which will encourage spatial consistency of the segmentation result. The former approach has been particularly popular among the remote sensing community [Ric93]. However this dramatically increases the dimensionality of the segmentation problem and the associated computational complexity. Alternatively, one may deal with the issue of spatial coherence by means of postprocessing. Morphological filtering techniques are one of the examples of methods specifically developed for this purpose. The main problem with the postprocessing refinement of the raw segmentation is that at this stage it is impossible to distinguish between noisy labelling and fine genuine image structures such as those characterised by texture.

On the other hand, region growing methods stress spatial coherence. Techniques vary, but typically, from an arbitrary starting point in the image an initially small region is grown to subsume neighbouring regions, provided the appended pixels satisfy a prespecified similarity measure. A number of termination methods have been devised to stop the region growing process [HS93]. By definition, the extracted regions are spatially coherent. However, the use of local statistics to compute the similarity measure often results in an excessive sensitivity to thresholds leading to oversegmentation or undersegmentation of the image data.

In this work we argue that, in order to achieve successful segmentation (either image partitioning or selection of salient regions) one should exploit simultaneously global and local statistics that can be computed from the image, together with pixel connectivity information. The global statistics of the clustered features defines a model that can be used in local statistical testing to determine the segment identity

of individual pixels. The idea of performing image segmentation by means of a parallel process of feature space analysis under spatial coherence constraints has been voiced by Ballard [Bal84]. However, his paper only identified the issues without advocating any specific solutions to address them. We have developed an effective algorithm that achieves the objectives of simultaneous feature space-image space clustering. Moreover, we show that the proposed approach is completely general and consequently applicable to a wide spectrum of problems, ranging from grey and colour image segmentation, to edge and line extraction, range data and motion segmentation. In other words, the approach defines a unified framework for a family of image segmentation problems. It can also be used either to achieve a complete segmentation of the image or select a subset of image regions according to a prespecified criterion. Currently, region saliency is used to define priority.

We discuss the relationship of the proposed method to the perceptual grouping approach based on robust statistical testing. We show that the proposed method is more resilient to contamination due to clutter and consequently it can resolve image features to much enhanced resolution. The practical significance of the proposed method is that it can distinguish between noisy regions and genuine texture. This is particularly important for correctly segmenting textured motion fields (resulting from object transparency).

The chapter is organised as follows. In the next section we introduce the novel feature and spatial domain (FSD) clustering method. In Section A.3 we illustrate the main differences between GT CLUSTERING and FSD CLUSTERING clustering on a set of test problems. In Section A.4 we describe a number of diverse experiments in image segmentation to demonstrate the properties and versatility of the proposed procedure. Finally, conclusions are drawn in Section A.5.

A.2 Algorithm

A large number of image segmentation problems involve low dimensional feature spaces and huge quantities of pixel data. In such situations it is most efficient to base the cluster analysis of the data on its statistical summary in the form of a histogram. Accordingly the clustering problem can be viewed as one of partitioning the feature space into regions over which the histogram is locally unimodal.

The problem of unimodal cluster separation has received significant attention in the pattern recognition community [KNF76][KB90]. We adopted the Graph-theoretical clustering method of Koontz and Fukunaga, described in detail in section 4.3.1 as the basis of the FSD CLUSTERING. We outline the FSD CLUSTERING algorithm below. The detailed behaviour of the algorithm is discussed in the next section in connection with simple test cases, each highlighting effects of one particular step of algorithm 3.

Algorithm 3: **Clustering in Feature and Spatial Domains**

1. Perform graph-theoretical clustering.
2. Terminate if:
   (a) No local maximum exists.
   (b) Global maximum is bellow a threshold or the required number of regions is found.

   else select the current global maximum.
3. Traverse the tree rooted at current maximum to obtain a list of bins that belong to a unimodal cluster (identical to GT-clustering)
4. Check local (ie. connected component) consistency.
   (a) Backproject all pixels voting into bins of the unimodal cluster in the image. Perform connected component analysis.
   (b) Sort the list of connected components by size.
   (c) Histogram computed on the largest connected component forms the initial model of the cluster distribution (in feature space).
   (d) Iterating through the sorted list, statistically test whether the current model of cluster distribution is consistent with the distribution computed on the next connected component. Votes from accepted components are removed from the global histogram.

5. Iteratively add pixels that are *both* in spatial and feature space neighbourhoods of the cluster.

6. Update the list of local maxima and go to step 2

The complexity of computing connected components is near linear [Sed88]. The histogram comparison is carried out by a modified method described in [PFTV88] or by histogram intersection [Swa90]. The complexity is linear in the number of non-zero bins of the compared histograms.

A.3 Test Cases

This section presents five 'test cases', ie. synthetically generated examples of simple yet common circumstances where feature space clustering alone cannot in principle result in error-free segmentation. This is due to the fact that the class distributions are chosen so that they overlap in the feature space.

For simplicity the problems are presented in the context of intensity image segmentation where dimensionality of the feature space is 1 and the data are taken from a 2D domain. The images contain rectangular patches with pixel values drawn from different Gaussian distributions. As discussed in detail bellow, the 'cases' differ in the character of the global histogram, number of distributions involved and adjacency relation of the rectangular patches. The distributions are chosen to highlight a particular limitation of the purely feature-based clustering methods like Modesp [Kit76] or the Graph-theoretical (GT) clustering [KNF76] [Fuk90]. In particular we shall compare the results of the proposed FSD method with GT clustering, but the conclusions drawn, in general, are expected to hold for a large class of feature space clustering algorithms.

Problem 1. Two non-adjacent regions A and B on a black background are shown in Figure A.1 (a). The two overlapping clusters shown in Figure A.2 (a) form a single unimodal cluster in feature space. The situation arises eg. when the clusters are normally distributed and $\mu_1 - \mu_2 \leq \sigma_1 + \sigma_2$ In this case, GT-clustering detects a single cluster (see Figure A.1 (b)), regardless of the fact that each region has significantly different statictical properties. This can be shown either

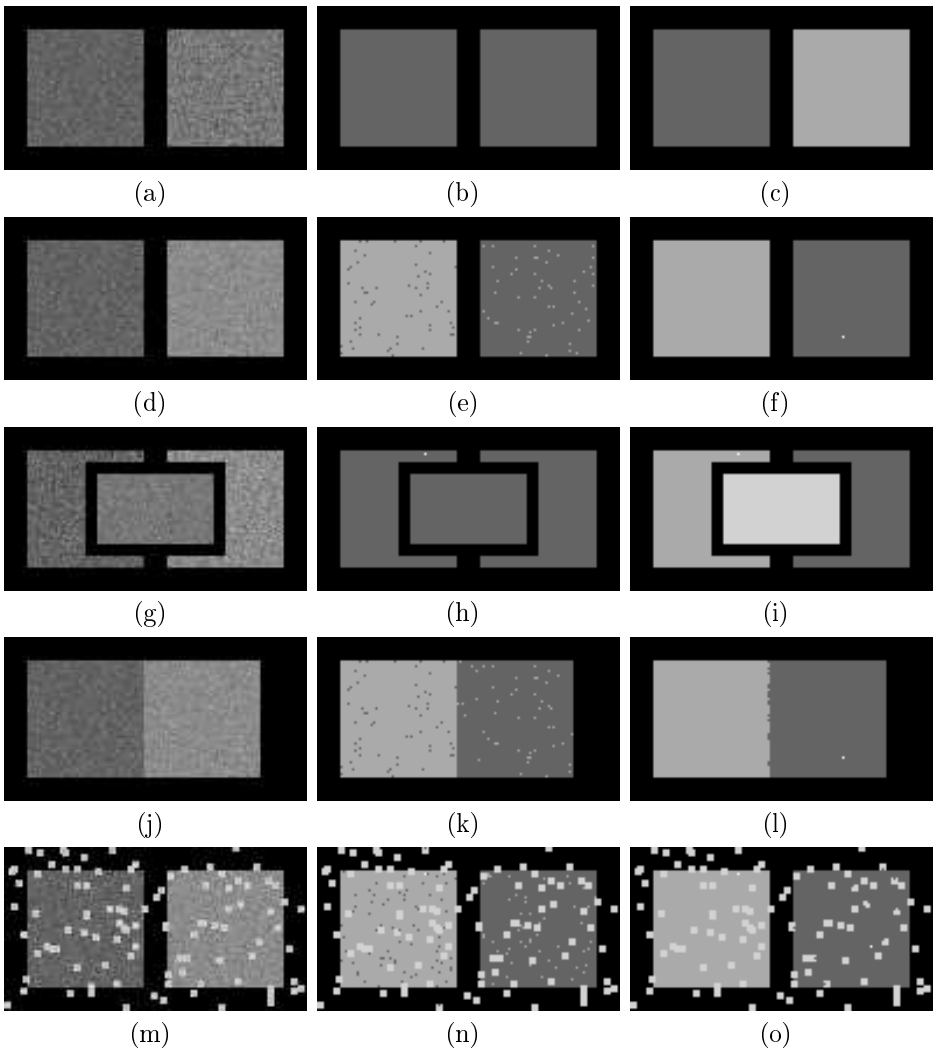


Figure A.1: Test case images.

by a Chi-square test on the feature (ie. grey-level) distributions of the regions or simply by comparing interval estimates of the means and standard deviations of the two distributions. The distributions used in the example were $N(100, 10)$ and $N(120, 15)$. In Step 4.d, FSD clustering rejects the hypothesis that the clusters from the two regions are identical and produce the correct segments (see Figure A.1 (c)). Note that in the case of the GT-clustering method, a change in the level of noise could qualitatively affect the segmentation result. The FSD method is considerably more stable. Using the spatial context, it resolves the single modal histogram into

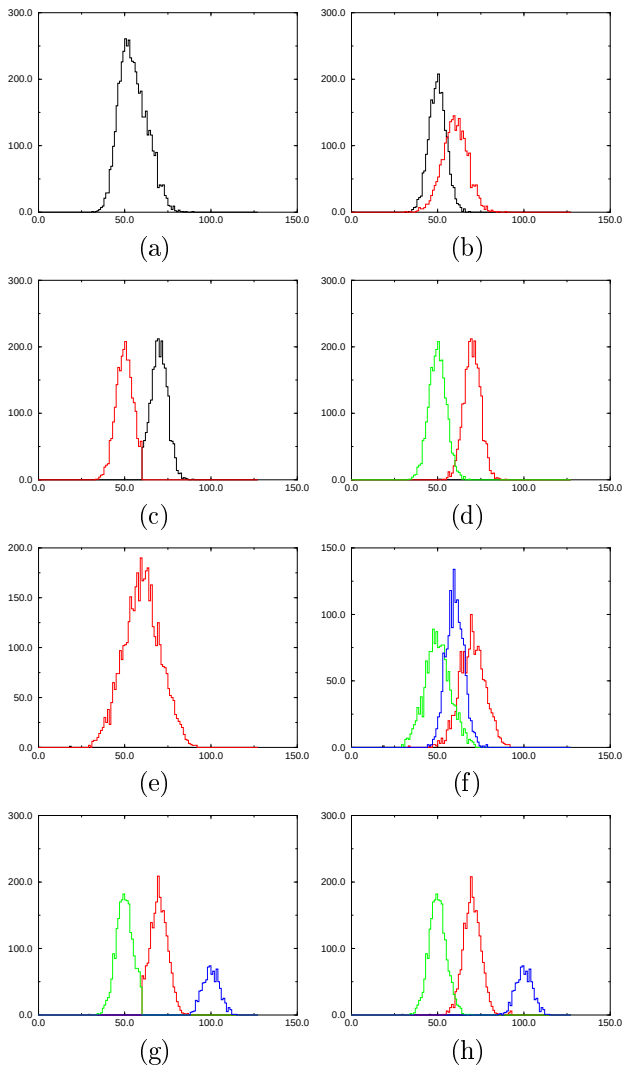


Figure A.2: Test case image histograms.

two mixture components (Figure A.2 (b)) characterising the respective statistical properties of the two regions.

Problem 2. Two non-adjacent regions A and B on a black background are shown in Figure A.1 (d). The regions form a histogram with two overlapping modes as shown in Figure A.2 (c). The distributions used were $N(100, 10)$ and $N(140, 10)$. GT-clustering segments the image data by partitioning the histogram at the valley point between the two modes. This results in pixel misclassification (oversegmentation) illustrated in Figure A.1 (e). In the case of the FSD method, in Step 4.d small

regions are rejected on the grounds that their pixels are not from different distributions. In Step 5 the small regions are then incorporated into a neigbouring larger region on the grounds of spatial adjacency to deliver an error free segmentation of the image as shown in Figure A.1 (f).

Problem 3. Figure A.1 (g) shows two adjacent regions A and B generated according to the same distributions as in Problem 2. A third region C distributed as $N(120, 10)$, not adjacent to either A or B, overlaps in the feature space heavily both with A and B. Globally, the feature histogram is unimodal, as apparent from Figure A.2 (e). As in Problem 1 the GT-clustering throws all pixel data into the same foreground cluster (see Figure A.1 (h)). The proposed method resolves the data into three separate clusters in agreement with human perception. This is depicted in Figures A.1 (i) and A.2 (f). The example shows how the GT result depends on non-local context.

Problem 4. Identical to Problem 2, but the two regions with the overlapping histograms are adjacent. It is interesting to observe the behaviour of the segmentation methods on the boundary shown in Figures A.1 (k) and (l). In particular, the position of the boundary produced by the FSD method can be slightly biased because of the inherent ambiguity of region membership of pixels which are adjacent both spatially and in the feature space. The FSD algorithm will attempt to grow a region as far as possible. The bias will reflect the order in which regions are processed.

Problem 5. Identical to Problem 2, but a number of small regions with distribution shown in Figure A.2 (g) are added to the image (see Figure A.1 (m)). This simulates the effect of textured region. Here we illustrate the case when postprocessing by morphological filtering would destroy useful image content. Without the filtering, the GT-clustering method produces a noisy segmentation as shown in Figure A.1 (n). In contrast, the FSD method treats differently the pixels with feature values lying in the overlapping tails of the cluster distributions and those exhibiting distinct feature space identity, regardless of the size of the region they form as can be seen in Figure A.1 (o).

A.4 Experiments

All the experiments were carried out using the same program where only the voting procedure (approximately five to ten lines of code) was application specific. The section of the code for the FSD clustering computation which is about three orders of magnitude larger remained unchanged.

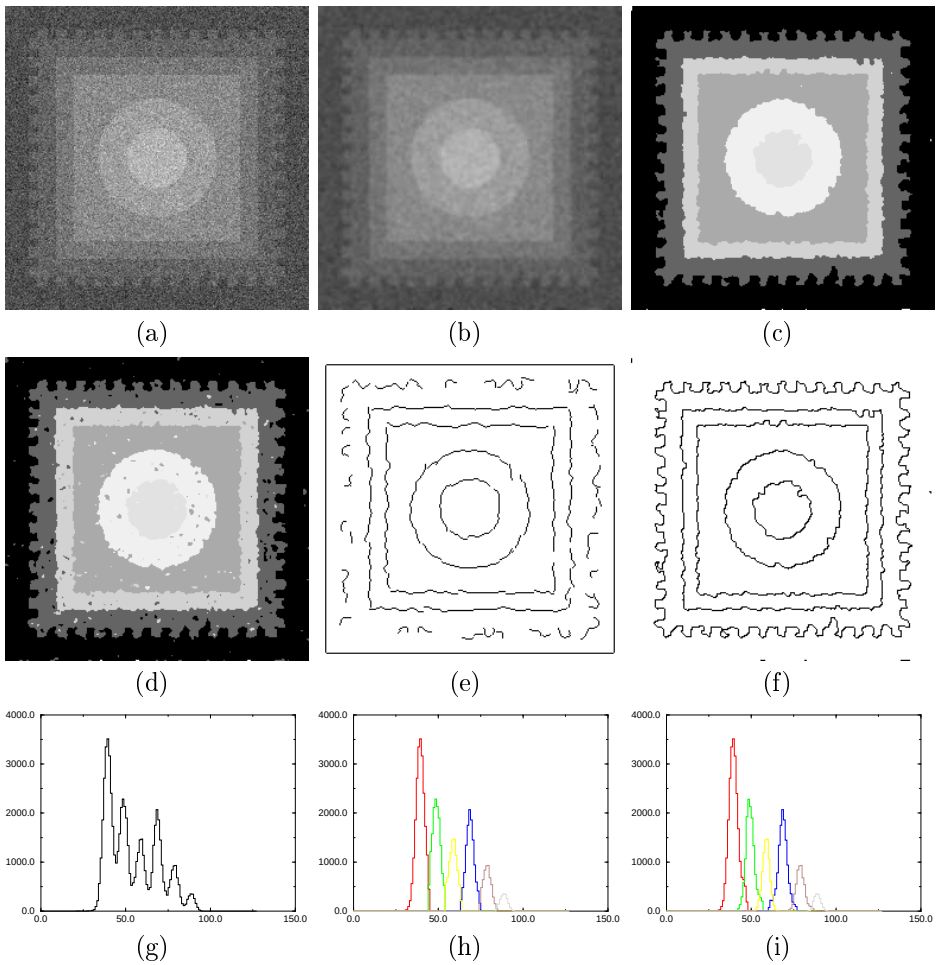
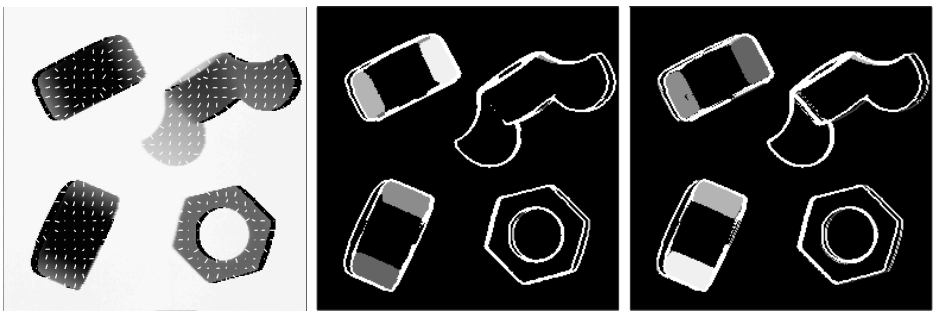


Figure A.3: Stamp (from [Pet94]).

Segmentation of a gray level image. A very noisy synthetic image of Figure A.3 (a) referred to as "stamp" which has been used as a benchmark in edge detection studies [PK91], has been used to demonstrate the properties of the FDS segmentation scheme. For comparison with the segmentation approach based on edge detection the image was first filtered with an optimal filter [PK91], [Pet94]

used in the Petrou-Kittler edge detector. The result obtained with the edge detector is shown in Figure A.3 (e). The filtered image Figure A.3 (b) is the input to the FSD and GT clustering algorithms which produce outputs shown in Figures A.3 (c) and (d) respectively. Note the noisy appearance of the GT result. On the other hand, the FSD method's biased treatment of larger regions explains why the outer circular ring is about one pixel wider than its true size. From the FSD segmentation output the image edges can be obtained by extracting the region boundaries as shown in Figure A.3 (f). When compared to the edge detector output we note the relative quality of the FSD performance. The edge map in Figure A.3 (f) is less noisy and the boundaries are closed which may help the subsequent analysis. The edge detector output produces less noisy edge chains along strait lines. However, considering that in contrast to the edge detection method the FSD approach has no inbuilt knowledge of shape, the clustering scheme works remarkably well. In the histogram terms, the segmentation result of the two methods can be seen in Figure A.3 (h) for the GT algorithm and in Figure A.3 (i) for the FSD algorithm, with Figure A.3 (g) showing the image histogram.



(a) (b) (c)

Figure A.4: Bolts. (a) Original image. (b) FSD clustering. (c) GT clustering.

Range Data The approach was tested on depth images from the NRCC database [RC88]. The NRCC database contains dense image data with depth available at every pixel location. A sample image is shown in Figure A.4 (a). We estimated the depth image derivatives $\frac{dz}{dx}$ and $\frac{dz}{dy}$ using the Sobel operator. Each pair of derivatives defines a normal which is superimposed on the depth image. Note that points of constant derivative values belong to a set of pixels lying on parallel planes. The clustering was performed in the $\frac{dz}{dx}$, $\frac{dz}{dy}$ space. The resulting segmentation for the

FSD and GT clustering methods is shown in Figures A.4 (b) and (c) respectively. At some pixel location the depth measurements are not available. At such points estimates of the derivatives are randomly spread over the feature space and consequently the pixels are not assigned to any cluster (see the white points in Figures A.4 (b) and (c).

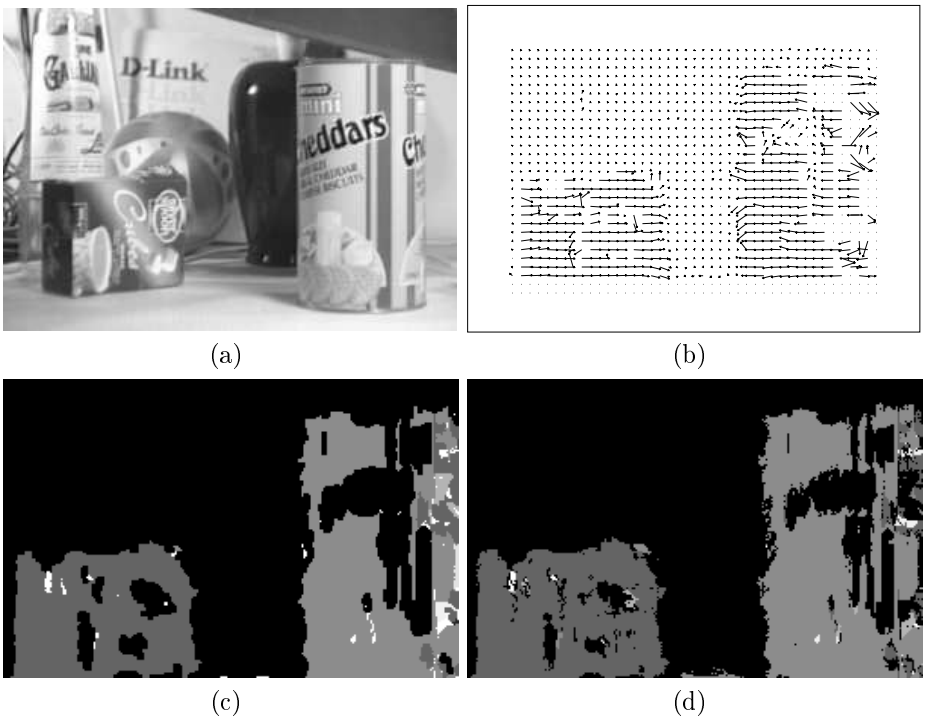
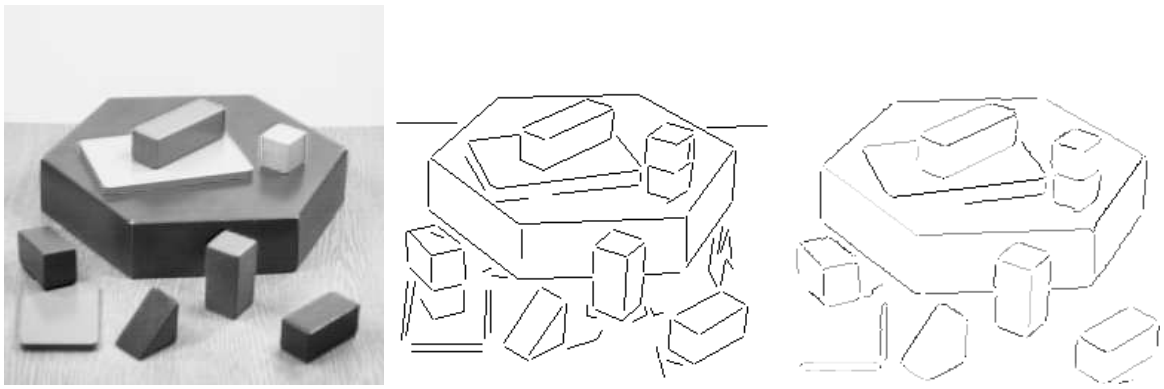


Figure A.5: Lab sequence.

Optic Flow We tested the suitability of the technique for motion segmentation on a pair of images A.5 (a) taken from the 'Lab' sequence of images analysed in the thesis of Bober [Bob95], page 53. The image contains a stationary background and two moving objects undergoing distinct translatory motion. The optic flow field in Figure A.5 (b) was computed using a robust Hough Transform technique described in [BK94], with the following parameters: block size 16x16, final resolution 0.01 pixel. The sequence was preprocessed by Gaussian filtering, sigma 3 pixels. The results of the FSD and GT clustering methods are shown in Figures A.5 (c) and

(d) respectively. It can be seen that both moving objects are segmented out. The differences are insignificant as the clusters corresponding to the two motions are not overlapping. In both cases the segmentation is noisy because the regions contain flow vectors which are distribution outliers. Note that these errors are the artifact of the optic flow computation rather than produced by the clustering procedures. It should also be pointed out that for nontranslatory motion, a more general motion model would have to be assumed to achieve feature space and spatial domain coherence to allow satisfactory segmentation.



(a) (b) (c)

Figure A.6: A polyhedral scene. (a) Original image. (b) Lines detected by Hough Transform (c) Results of clustering

Line Detection by Hough Transform In order to demonstrate a completely different application of the FSD clustering method we consider the problem of detecting straight lines in the image shown in Figure A.6 (a). Edge position and direction is mapped into the ρ - θ line parameter space. The connected component analysis (Step 3.a) in this case effectively compares the properties of the individual line segments associated with the infinite straight line defined by the mode of each cluster. Note that at this stage the lines that are grouped together in a single unimodal cluster will be resolved into a set of contributing lines with distinct parameters. In this sense the proposed approach does not have the tendency to merge several lines of slightly different parameters into a long ghost line with incorrect 'averaged' parameters. The FSD clustering output is shown in Figure A.6 (c) for comparison with the robust Hough transform output [Pri90] in Figure A.6 (b).

There are no significant differences in performance. The slight oversegmentation produced by the Hough transform method could be reduced by threshold modification. Also note that the FSD method outputs edge chains, rather than line models. Different clusters of edge pixels are denoted by different graylevels. The main aim here is to demonstrate the versatility of the FSD method.

(a) (b) (c)

Figure A.7: Colour. (a) Original image. (b) GT clustering. (c) FSD clustering.

Segmentation of Colour Images The RGB values of the image in Figure A.7 (a) were projected into the chromatic plane to achieve illumination invariant segmentation results [MMK95]. The underlying assumptions for successful segmentation are diffused reflection, and the ability to treat specularities and interreflections as noise as discussed in [MMK94]. Here the FSD scheme performs significantly better (see Figure A.7 (c)) than GT clustering (Figure A.7 (b)), specially in the dark and/or specular regions where clusters corresponding to different reflectances overlap. They are disambiguated by the FSD procedure by considering the spatial context.

A.5 Conclusions

We have proposed a novel approach to image segmentation, called *feature and spatial domain* clustering. The method is devised to group pixel data by taking into account simultaneously both their feature space similarity and spatial coherence. The FSD CLUSTERING algorithm is practically application independent. It has been successfully tested on a wide range of image segmentation problems, including grey and colour image segmentation, edge and line detection, range data and motion segmentation. In comparison with existing segmentation approaches, the

method can resolve image features even if their distributions significantly overlap in the feature space. It can distinguish between noisy regions and genuine fine texture. Moreover, if required, FSD CLUSTERING can produce partial segmentation by identifying salient regions only.

Acknowledgements

The experiments on range data were carried out with help from Andrew Stoddart and Adrian Hilton. We thank Miroslaw Bober for providing the input data for the optical flow field segmentation and to Maria Petrou for the stamp image.

References

[Bal84] D.H. Ballard. Parameter nets. *Artificial Intelligence*, pages 235–267, 1984.

[BK94] M. Bober and J. Kittler. Robust motion analysis. In *CVPR94*, pages 947–952, 1994.

[Bob95] M. Bober. *General Motion Estimation and Segmentation from Image Sequences*. PhD thesis, VSSP Group, University of Surrey, United Kingdom, 1995.

[Fuk90] K. Fukunaga. *Introduction to Statististical Pattern Recognition*. Academic Press, 1990.

[HS93] R.M. Haralick and Shapiro. *Computer and Robot Vision - Volume II*. Addison-Wesley, 1993.

[JD88] Anil K. Jain and Richard C. Dubes. *Algorithms for Clustering Data*. Prentice Hall, 1988.

[KB90] A. Khotanzad and A. Bouarfa. Image segmentation by a parallel , non-parametric histogram based clustering algorithm. *Pattern Recognition*, pages 961–973, 1990.

[Kit76] J. Kittler. A locally sensitive method for cluster analysis. *Pattern Recognition*, 8:23–33, 1976.

[KNF76] W.L.G. Koontz, P. M. Narendra, and K. Fukunaga. A graph-theoretic approach to nonparametric clustering. *IEEE Trans. Computers*, C-25:936–944, 1976.

[MMK94] J. Matas, R. Marik, and J. Kittler. Illumination invariant colour recognition. In E. Hancock, editor, *British Machine Vision Conference*. BMVA Press, 1994.

[MMK95] J. Matas, R. Marik, and J. Kittler. On representation and matching of multi-coloured objects. In *Fifth International Conference on Computer Vision (Cambridge, MA, June 20–23, 1995)*, pages 726–732, Washington, DC., 1995. Computer Society Press.

[Pet94] M. Petrou. The differentiating filter approach to edge detection. In *Advances in Electronics and Electron Physics*, volume 88, pages 297–345. Academic Press, 1994.

[PFTV88] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1988.

[PK91] M. Petrou and J. Kittler. Optimal edge detectors for ramp edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, May 1991.

[Pri90] J. Princen. Hough transform methods for curve detection and parameter estimation. Technical report, University of Surrey, 90.

[RC88] M. Rioux and L. Cournmoyer. *The NRCC Three-dimensional Image Data Files*. National Research Council Canada, Ottawa, Ontario, Canada, 1988.

[Ric93] John A. Richards. *Remote Sensing Digital Image Analysis*. Springer, 1993.

[Sed88] R. Sedgewick. *Algorithms*. Addison-W