Fernando Jorge Coutinho Monteiro

Region-based Spatial and Temporal Image Segmentation

Universidade do Porto

Faculdade de Engenharia

FEUP

Departamento de Engenharia Electrotécnica e de Computadores

Faculdade de Engenharia da Universidade do Porto

Julho de 2007

Fernando Jorge Coutinho Monteiro

Region-based Spatial and Temporal Image Segmentation

Universidade do Porto

Faculdade de Engenharia

FEUP

Tese submetida à
Faculdade de Engenharia da Universidade do Porto
para a obtenção do grau de
Doutor em Engenharia Electrotécnica e de Computadores

Dissertação realizada sob a supervisão do
Professor Doutor Aurélio Joaquim de Castro Campilho
Departamento de Engenharia Electrotécnica e de Computadores
Faculdade de Engenharia da Universidade do Porto
Julho de 2007

"Pedras no caminho?
Guardo-as todas,
um dia vou construir um castelo."

Autor desconhecido

À minha família e em especial à memória da minha mãe.

Resumo

Este trabalho estuda métodos baseados em regiões para a segmentação de imagens e de sequências de vídeo. Apresentam-se metodologias precisas para a segmentação de imagem e demonstra-se como é que podem ser integradas em algoritmos para a resolução de alguns dos problemas associados à segmentação do movimento. A representação baseada em regiões oferece uma forma de realizar um primeiro nível de abstracção e de reduzir o número de elementos a processar relativamente à representação clássica *pixel* a *pixel*.

A segmentação do movimento é uma técnica fundamental para a análise e compreensão de sequências de imagens reais. A segmentação do movimento "descreve" a sequência através de um conjunto de regiões compostas por pontos que apresentam um movimento coerente entre si. Esta descrição é essencial para a identificação dos objectos presentes na cena de modo a permitir uma manipulação eficaz de sequências de vídeo.

Nesta tese é apresentada uma técnica híbrida baseada na combinação de informação espacial e de informação do movimento para a segmentação dos objectos presentes numa sequência de imagens de acordo com o seu movimento. O problema é formulado como um caso de partição de um grafo onde cada nó corresponde a uma pequena região composta por pontos que apresentam o mesmo movimento. Esta é uma representação flexível de alto-nível que individualiza os objectos com movimento próprio. Partindo de uma sobre-segmentação da imagem, os objectos são formados pelo agrupamento de regiões vizinhas com base na sua similaridade espacial e temporal, tendo em atenção a informação espacial e de movimento, com ênfase na segunda. A segmentação final é obtida recorrendo a um método espectral para partição de grafos.

A fase inicial para a segmentação de objectos de acordo com o seu movimento visa a redução do ruído da imagem sem destruir a estrutura topológica dos objectos, através

de um filtro anisotrópico bilateral. Uma partição inicial em pequenas regiões uniformes é obtida através da transformada de *watershed*. O vector de movimento associado a cada região é determinado por um algoritmo variacional de cálculo de fluxo óptico. De seguida, é construído um grafo de regiões dinâmicas pela combinação normalizada de medidas de similaridade entre regiões onde são considerados, a intensidade média de cada região, a amplitude do gradiente entre regiões e a informação do movimento associado à região. A medida de similaridade de movimento entre regiões é baseado no sistema de visão humano. Finalmente, é aplicado um método espectral para obter a partição do grafo e consequente identificação de cada região de acordo com o seu movimento.

O método de segmentação do movimento é baseado num de segmentação de imagens estáticas também concebido e desenvolvido pelo autor da dissertação. Trata-se também de uma metodologia baseada na utilização de pequenas regiões que assenta na construção de um grafo de similaridades entre regiões tendo por base a informação da intensidade e da amplitude do gradiente entre regiões. Esta técnica produz segmentações mais simples e mais compactas e comparativamente vantajosa relativamente a outras técnicas. De modo a avaliar os resultados da segmentação é proposta uma nova métrica que tem em atenção o modo como os humanos visualizam os resultados.

A combinação de informação estática e do movimento numa técnica baseada em regiões permite obter resultados de segmentação visualmente significativos. São apresentados resultados experimentais do desempenho da técnica proposta tanto para a segmentação do movimento em sequências de imagens, com e sem movimento da câmara, bem como para a segmentação de imagens estáticas, sendo, neste caso, efectuada uma comparação com os resultados obtidos por outras técnicas.

Palavras chave: Segmentação de imagem, estimativa do movimento, segmentação do movimento, avaliação da segmentação, transformada de *watershed*.

Abstract

This work discusses region-based representations for image and video sequence segmentation. It presents effective image segmentation techniques and demonstrates how these techniques may be integrated into algorithms that solve some of the motion segmentation problems. The region-based representation offers a way to perform a first level of abstraction and to reduce the number of elements to process with respect to the classical pixel-based representation.

Motion segmentation is a fundamental technique for the analysis and the understanding of image sequences of real scenes. Motion segmentation 'describes' the sequence as sets of pixels moving coherently across one sequence with associated motions. This description is essential to the identification of the objects in the scene and to a more efficient manipulation of video sequences.

This thesis presents a hybrid framework based on the combination of spatial and motion information for the segmentation of moving objects in image sequences accordingly with their motion. We formulate the problem as graph labelling over a region moving graph where nodes correspond coherently to moving atomic regions. This is a flexible high-level representation which individualizes moving independent objects. Starting from an over-segmentation of the image, the objects are formed by merging neighbouring regions together based on their mutual spatial and temporal similarity, taking spatial and motion information into account with the emphasis being on the second. Final segmentation is obtained by a spectral-based graph cuts approach.

The initial phase for the moving object segmentation aims to reduce image noise without destroying the topological structure of the objects by anisotropic bilateral filtering. An initial spatial partition into a set of homogeneous regions is obtained by the watershed transform. Motion vector of each region is estimated by a variational approach. Next a region moving graph is constructed by a combination of normalized

similarity between regions where mean intensity of the regions, gradient magnitude between regions, and motion information of the regions are considered. The motion similarity measure among regions is based on human perceptual characteristics. Finally, a spectral-based graph cut approach clusters and labels each moving region.

The motion segmentation approach is based on a static image segmentation method proposed by the author of this dissertation. The main idea is to use atomic regions to guide a segmentation using the intensity and the gradient information through a similarity graph-based approach. This method produces simpler segmentations, less over-segmented and compares favourably with the state-of-the-art methods. To evaluate the segmentation results a new evaluation metric is proposed, which takes into attention the way humans perceive visual information.

By incorporating spatial and motion information simultaneously in a region-based framework, we can visually obtain meaningful segmentation results. Experimental results of the proposed technique performance are given for different image sequences with or without camera motion and for still images. In the last case a comparison with the state-of-the-art approaches is made.

Keywords: image segmentation, motion estimation, motion segmentation, segmentation evaluation, watershed transform.

Résumé

Ce travail étudie des méthodes basées sur des régions pour la segmentation d'images et de séquences de vidéo. On présente des méthodologies précises pour la segmentation d'image et on démontre comment elles peuvent être intégrées dans des algorithmes pour la résolution de certains problèmes associés à la segmentation du mouvement. La représentation basée sur des régions offre une forme de réaliser un premier niveau d'abstraction et de réduire le nombre d'éléments à traiter en comparaison avec la représentation classique *pixel* par *pixel*.

La segmentation du mouvement est une technique fondamentale pour l'analyse et la compréhension de séquences d'images réelles. La segmentation du mouvement "décrit" la séquence à travers d'un ensemble de régions composées de points qui présentent un mouvement cohérent entre eux. Cette description est essentielle pour l'identification des objets présents dans la scène afin de permettre une manipulation efficace de séquences de vidéo.

Dans cette thèse on présente une technique hybride basée sur la combinaison d'informations spatiales et du mouvement pour la segmentation des objets présents dans une séquence d'images conformément à son mouvement. Le problème est formulé comme un cas de partition d'un graphe où chaque nœud correspond à une petite région composée par des points qui présentent le même mouvement. Celle-ci est une représentation flexible de haut niveau qui individualise les objets avec mouvement propre. En partant d'une sur-segmentation de l'image, les objets sont formés par le regroupement de régions voisines basé sur leurs similitude spatiale et temporel, tenant en compte les informations spatiales et surtout du mouvement. La segmentation finale est obtenue en faisant appel à une méthode spectrale pour partition de graphes.

La phase initiale pour la segmentation d'objets conformément à son mouvement vise la réduction du bruit de l'image sans détruire la structure topologique des objets,

à travers un filtre anisotrope bilatéral. Une séparation initiale de petites régions uniformes est obtenue à travers la transformée de *watershed*. Le vecteur de mouvement associé à chaque région est déterminé par un algorithme de calcul de flux optique basé sur le système de vision humain. Après, on construit un graphe de régions dynamiques utilisant la combinaison normalisée de mesures de similitude entre des régions où sont considérés l'intensité moyenne de chaque région, l'amplitude du gradient entre régions et les informations du mouvement associé à la région. Finalement, on applique une méthode spectrale pour obtenir la séparation du graphe et la conséquente identification de chaque région conformément à son mouvement.

La méthode de segmentation du mouvement est basée sur une méthode de segmentation d'images statiques aussi conçu et développé par l'auteur de cette thèse. Il s'agit aussi d'une méthodologie basée sur l'utilisation de petites régions, préalablement obtenues, basées sur la construction d'un graphe de similitudes entre régions tenant en compte les informations de l'intensité et de l'amplitude du gradient entre des régions. Cette technique produit des segmentations plus simples et plus compactes et comparativement avantageuses à l'égard d'autres techniques. Afin d'évaluer les résultats de la segmentation on propose une nouvelle métrique qui tient en compte la façon de visualiser les résultats par les être humains.

La combinaison d'informations statiques et du mouvement dans une technique basée sur des régions permet d'obtenir des résultats de segmentation visuellement significatifs. On présente des résultats expérimentaux sur la performance de la technique proposée dans le cas de la segmentation du mouvement dans des séquences d'images, avec et sans mouvement de la chambre, ainsi que pour le cas de la segmentation d'images statiques, étant, dans ce cas aussi, effectué une comparaison avec les résultats obtenus par autres techniques.

Mots-clés: segmentation de l'image, estimation du mouvement, segmentation du mouvement, évaluation de la segmentation, transformée *watershed*.

Agradecimentos

Este documento é fruto da investigação desenvolvida durante os últimos quatro anos. Quero por isso agradecer àquelas pessoas e instituições que contribuiram, directa ou indirectamente, na realização deste trabalho:

Ao Professor Aurélio Campilho, pela sua orientação, disponibilidade e confiança que tem depositado em mim. Ao longo destes últimos anos, foram muitas as conversas que mantivemos e não posso deixar de mostrar o meu apreço pela possibilidade que sempre me foi dada de expor e confrontar, sem quaisquer limitações, as minhas ideias. À pessoa que sempre esteve disposta a ouvir e, nos momentos certos, soube mostrar que eu estava errado, o meu sincero agradecimento.

À Inês pelo amor e pela compreensão demonstrada, tendo muitas vezes feito o papel de mãe e de pai do João Pedro e do José Miguel, tentado sempre proporcionar-me o tempo necessário para trabalhar; e pela ajuda preciosa na correcção ortográfica desta dissertação.

Ao Carlos Balsa pela tradução do *Résumé*.

Agradeço também ao Instituto Politécnico de Bragança, particularmente à Escola Superior de Tecnologia e de Gestão, pelos apoios concedidos. Ao Fundo Social Europeu pela concessão da bolsa PRODEP de doutoramento 5.3/N/199.023/03.

Finalmente, quero agradecer também aos meus amigos e à minha família por todo o apoio recebido durante estes anos.

Acção de Doutoramento co-financiada pelo Fundo Social Europeu.

Ministério da Educação
prodep III
PROGRAMA DE DESENVOLVIMENTO EDUCATIVO PARA PORTUGAL

União Europeia

Fundo Social Europeu

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

Among all the human perceptual mechanisms, vision is undoubtedly the most important. The effortlessly way that we often look, interpret and ultimately act upon what we see belies the complexity of visual perception. The comparatively young science of vision research is aimed at the understanding of the general issue of seeing. The automation of the task by the use of image capture equipment in place of our eyes, computers and algorithms in place of the not yet understood visual system, constitutes what is termed computer vision. The Human Visual System (HVS) is an important model for any work in vision because it is, clearly, both efficient and general purpose, which are also the goals of any computer vision system.

Human often take for granted the solution of apparently simple computer vision problems like the segmentation and the recognition of objects, or the detection and the interpretation of motion. We solve these tasks so automatically that it can be surprising how difficult it is to instruct a computer to solve the same tasks, given just a series of two-dimensional arrays of pixel values.

When humans look at a scene, the visual system is able to decompose and identify objects in a complex scene in one instant. It is, essentially, the process of subdividing an image into basic parts and extracting these parts of interest which are the objects. In a conventional sense, image segmentation is the partitioning of an image into coherent regions, in a manner consistent with human perception of the content, where parts within a region are similar according to some uniformity property and dissimilar between neighbouring regions. The development of MPEG-4 and MPEG-7 standards which allow the object-based image coding and content-based image description and retrieval, reinforced the interest in image segmentation algorithms.

Image segmentation and perceptual grouping have traditionally relied on different image cues. Segmentation is often based mostly on pixel appearance, being it by brightness, colour or some measure of texture similarity (though the issue of cue integration for segmentation has received a reasonable amount of attention, see [Malik 01]); whereas perceptual grouping usually relies on the information provided by image edges and on grouping principles that exploit the regularities among edges that belong to object contours.

The information provided by image segmentation and perceptual grouping is also complementary. Segmentation results indicate what regions in the image look homogeneous under a chosen similarity measure, without considering boundary regularity; while grouping results indicate which edges in the image form regular groups that are likely to correspond to salient boundaries. It is reasonable to expect that combining the results produced by segmentation and grouping should lead to a better segmentation. Motion information may be used to link adjacent but visually dissimilar regions or to divide surfaces not easily separable by static criteria alone. Often, ambiguous object boundaries in a single image frame are easily resolved when dynamic effects are evaluated based on a sequence of frames.

For image segmentation, evaluation and, where possible, validation against other methods are crucial. In some cases we have been able to compare our results against state-of-the-art techniques from other researchers. Still, in most cases the ground truth will remain concealed such that evaluation must be conducted with due care and attention, even if the so-called 'gold standards' are available.

Motion segmentation is another important research field with many commercial applications including surveillance, navigation, robotics, and image coding and compression. As a result, the field has received a great deal of attention and there are a wide variety of motion segmentation techniques which are often specialised for particular problems. The relative performance of these techniques, in terms of both accuracy and of computational requirements, is often found to be data dependent and no single technique is known to outperform all others for all applications under all conditions.

Motion segmentation is usually defined as grouping of pixels of similar intensity that are associated with smooth and uniform motion information. However, this is a problem that is loosely defined and ambiguous in certain ways. Though the definition of motion segmentation says that regions with coherent motion are to be grouped, the

resulting segments may not correspond to meaningful object regions in the image. To alleviate this issue the motion segmentation problem is placed at two levels namely low level and high level. Low level motion segmentation tries to group pixels with homogeneous motion vectors without taking no other information apart from intensity or image gradient. High level motion segmentation divides the image into regions that exhibit coherent motion and it also uses other image cues to produce image segments that correspond to projections of real objects.

This thesis intends to present efficient and effective image segmentation techniques and to demonstrate how these techniques may be integrated into algorithms that solve motion segmentation problems. Region based representations offer a way to perform a first level of abstraction and reduce the number of elements to process with respect to the classical pixel based segmentation. Morphological watershed transform and spectral-based graph cut methods will play a central role.

We can think of a video as a sequence of images so the basic unit on which the video segmentation algorithms operate is actually an image or a frame. The difference is that video segmentation must consider a larger feature space because they have moving objects. Informally we can say that video segmentation is essentially a segmentation problem, similar to the image segmentation problem with pixel motion being an important dimension of the feature space.

In image segmentation, the pixels of an image need to be partitioned into regions corresponding to the different intensity patterns existent in the image. In motion segmentation, the pixels of a pair (or a set of images) need to be partitioned into regions based on a coherent motion criterion. A moving scene is thereby recorded by a single camera and the initial task is to find a dense field of displacement vectors that transform one frame into a subsequent one. The most popular motion estimation method is the optical flow approach. Horn and Schunck [Horn 81] defined optical flow as follows: *The optical flow is a velocity field in the image that transforms one image into the next image in a sequence.* As such it is not uniquely determined.

Motion estimation and segmentation are important sources of information for many applications in multimedia and video analysis. Motion estimation is concerned with the estimate of the motion parameters of a moving object while motion segmentation attempts to identify the boundary of these objects. Both of these problems are directly related and a number of methods have been presented. The tasks of motion estimation

and segmentation are highly ill-posed[1].

It has been acknowledge by many authors that it is very difficult to determine the motion of pixels in areas of smooth intensity and that the motion in these areas must invariably be found by extrapolating from nearby features. These smooth areas of the image can be determined prior to any motion analysis by performing an initial segmentation based purely on intensity (or other spatial cues) to combine these smooth areas into individual *atomic regions*. The motion of these regions, rather than pixels, is then determined and these regions clustered together according to their motions.

In this work we propose a hybrid spatial and temporal technique that tries to overcome those problems by the combination of the spatial information with the motion information. Based on the assumption that motion discontinuities go along with discontinuities in the intensity image, we take benefit from the spatial segmentation information in three ways. First, the motion values inside each segment are constrained to follow the same motion model, which allows the assignment of smooth flow values in regions of poor texture. Secondly, we believe that motion boundaries can be accurately identified by the use of static cues, such as the partition of the reference frame into regions of homogeneous intensity. Thirdly, occluded regions can be assigned to meaningful flow values that are propagated using the segmentation information.

By its very nature, the problem of defining the objects composing a moving scene is an ill-posed problem. There is a strong interdependence between the estimation of the spatial support of an object and of its motion characteristics. On one hand, estimation of the motion information of the object depends on the region of support of the object. Therefore, an accurate segmentation of the object is needed in order to estimate the motion accurately. On the other hand, a moving object is characterized by coherent motion characteristics over its entire region of support (assuming that only rigid motion is permitted). Thus, an accurate estimation of the motion is required in order to obtain an accurate segmentation of the object. Furthermore, accurate object definition involves not only motion information, but also spatial characteristics. In particular, the spatial information provides important cues about object boundaries. However, the best strategy for combining these two types of information remains an open issue.

[1] A problem is called *well-posed* (in the sense of Hadamard), if it has a unique solution that depends continuously on the data. If one of these conditions is violated, it is called *ill-posed*.

1.1 Motivation

Motion segmentation is useful since in many real world examples the moving objects are precisely the interesting objects. For example when crossing the road it is the moving cars that are of primary importance; stationary cars are uninteresting background despite the fact that both moving and stationary cars are the same physical objects. Indeed, in many applications knowing that "something" is moving in a particular way is much more important than knowing semantically what it is.

The segmentation of images based on spatial or temporal (motion) information are key problems in computer vision. Motion information allows to distinguish stationary from moving objects and thus to detect and avoid obstacles. This makes it particularly useful for tasks where vehicles have to be guided safely through an unknown environment. Another field of application that is more related to image processing than to computer vision is the compression of video sequences where the basic idea is to decompose a sequence of images into a small set of key frames and encode the differences to the remaining frames as flow fields. Extending this idea to an even more compact representation based on object shapes and single displacement vectors describing their motion, one obtains the specification of the current MPEG-7 compression standard [Chang 01].

The goal of this thesis is to provide segmentation methods that are robust, fast and flexible enough to meet the requirements of the majority of the natural image analysis settings. Further, the methods are intended to serve as a basis for motion segmentation schemes.

The best known to assign segment labels to each pixel in an image is the normalized cuts algorithm developed by Shi and Malik [Shi 00]. This algorithm creates a weighted graph in which each pixel is connected to every other and the weights represent the similarity between them. A cut of the graph is a set of links whose removal divides the pixels into two groups. A minimum cut is the cut whose total links weights are the smallest, which is biased towards separating small regions from the remainder of the image. Normalized cuts corrects this bias by dividing the cut value by associativity factors that penalize small partitions.

Many methods have been proposed to perform the task of image segmentation with the cooperative methods among the most promising ones (see Chapter 2). This class of

approaches is based on the combination, integration or iteration between methods. It is known that the resulting segmented image from a watershed approach while accurate tends to over-segmenting the original image. In this research a region merging method using a graph based technique will be applied as a post image processing to overcome such problem. By applying these two methods in a combined manner, it is expected that a better image segmentation will be obtained.

Our idea is motivated by the observation that graph-cut algorithms have some drawbacks due to the use of pixel-based graphs. We think that combining watershed pre-segmentation with normalized cut approaches can lead to a faster and better segmentation. Moreover, using accurate uniform regions as the basis to any segmentation algorithm has to increase computational speed and allows to obtain smoother results on segmentation.

Recently, region-based algorithms have become popular in the motion and image segmentation community. Although quite different from each other, all methods of this category take benefit of the segmentation information to increase their robustness in traditionally challenging areas of motion segmentation. This is well reflected by the good experimental results of those techniques.

We identify the advantages of region-based motion segmentation as follows:

- Probably the most obvious advantage is that region-based motion segmentation techniques constrain the flow field inside a region to follow a single model. In other words, smoothness within a segment is explicitly enforced. This is advantageous, since it allows the assignment of smooth flow field values in regions of poor texture.
- Often, flow field boundaries can often be more accurately identified by the use of static cues. Each object (or region) has also a compact boundary.
- The robustness in areas affected by occlusion is improved. In theory, matching might even succeed for a segment that is partially occluded, since it is still possible to match the segment's non-occluded pixels. However, this does not mean that occlusions can be ignored. Note that since a single flow field model is assigned to the complete segment, those parts that are also affected by occlusion are automatically filled.
- The number of segments is usually significantly smaller than the number of pixels. This gives rise to potentially much faster motion segmentation algorithms.

Nevertheless, using the region-based assumption also involves some disadvantages:

- The most severe problem associated with region-based approaches is that the segmentation assumption is, in general, not guaranteed to hold true. More precisely, the success of such methods depends on the ability of the segmentation algorithm to accurately delineate the objects outlines. It is therefore safer to apply over-segmentation.
- The flow field model can be inappropriate to represent the "real" displacement of a segment. This is, of course, rather a problem of using a model and not specifically bound to the segmentation aspect. However, choosing an appropriate model is a difficult task by itself. While simple models may oversimplify the real displacement, complex models may over fit the data and show undesired effects due to image noise.

1.2 Contributions

The main emphasis in this thesis is in the presentation of a hybrid framework that produces accurate segmentation results in still images and in motion segmentation. To achieve those purposes some contributions are made during this thesis:

- The development of a new evaluation metric for image segmentation where additions from different errors are weighted accordingly to their visual relevance.
- The presentation of an improved watershed approach, the definition of a new structure for a region-based similarity graph and the application of multiclass normalized cuts approach to group atomic regions which produces accurate image segmentation.
- The definition of a similarity measure which overcomes some of the common problems associated with normalized cuts approach such as the partition of homogeneous regions.
- The incorporation of spatial and motion information simultaneously in a region-based framework to segment an image sequence. This method effectively allows the partition of the frames into multiple areas according to their different motions.
- The integration of the recently proposed motion estimation scheme developed by Brox et al. [Brox 04] in the region-based motion segmentation framework.

1.3 Thesis overview

This thesis is implicitly divided in two parts: the first part deals with theoretical and practical approaches towards image segmentation and that provides a suitable basis for the chapter of motion segmentation. The second part of this thesis focuses on the segmentation of moving objects. Thus, the remainder of this thesis is organized as follows.

In Chapter 2 a review of the commonly used image segmentation methods is given, with emphasis on the existing cooperative methods. The advantages and the disadvantages that exist within each method are described.

Chapter 3 introduces a segmentation evaluation measure which takes into account the way humans perceive visual information.

Chapter 4 presents the major contribution of this thesis - the use of atomic regions as nuclear features for image segmentation. An investigation on image segmentation approaches which produce an over-segmentation result will be given with the suggestion of a combined framework between watershed transform and spectral-based graph cut method for image segmentation. The resulting atomic regions are then encoded in a region-based graph where nodes correspond to regions. Afterwards, a multiclass spectral-based graph-cut method is used to cluster these regions in segments.

Chapter 5 takes the spatial atomic regions and a variational motion estimation method and combines them into a complete algorithm producing a reliable motion segmentation framework. The chapter begins with a review for motion estimation and segmentation. Afterwards, optical flow and its associated problems are discussed, with the description on the variational optical flow method. Finally, the complete framework for motion segmentation is presented.

Chapter 6 presents the experimental results of the proposed approaches to image segmentation and to motion segmentation. It includes a comparison of the proposed image segmentation method with the state-of-the-art image segmentation methods.

Finally, Chapter 7 presents a summary of the techniques developed in this work and draws conclusions from them. We then highlight some of the weaknesses of the algorithms and indicate some of the possible directions for further research.

Appendix A contains an extension of the experimental results.

CHAPTER 2

Survey on recent image segmentation methods

This chapter[1] *reviews some of the recent contributions in the area of image segmentation with emphasis on the cooperative segmentation methods. It also presents a new categorization of image segmentation algorithms.*

2.1 Introduction

There are many methodologies to approach the image segmentation problem that are traditionally organized into two main categories: 1) the region-based, and 2) boundary-based approaches. Other categorizations are possible as the ones we will survey in this chapter. In these approaches similarity or dissimilarity concepts are involved for measuring the homogeneity within a region or for evaluating the location of the boundaries. Each of the approaches presents its own advantages and drawbacks, they can be used isolated or combined in any convenient manner to explore the complementary properties of each method, or they can be unsupervised without any user intervention or interactive as often required by medical imaging applications [Olabarriaga 01].

Many issues still remain opened in the image segmentation problem, as the many different approaches, the different applications areas where image segmentation is mandatory and the evaluation of the performance of an image segmentation algorithm. We will also look at this problem from a different level, trying to identify those contributions where the integration, fusion, combination, cooperation or interaction are the

[1] The following survey on image segmentation is based largely on [Campilho 07].

major keywords for approaching the segmentation issue. This means that we will also review the methods based on the use of different and complementary methodologies that anyhow explore the advantages and disadvantages of a particular method in order to improve the overall segmentation performance.

We just give a brief overview of two earlier surveys, the [Haralick 85] paper and [Pal 93]. In [Haralick 85] the authors describe the main ideas of the image segmentation methods that are grouped into five major classes: (1) measurement space guided spatial clustering (further divided into thresholding and measurement space clustering); (2) region growing (divided into: single linkage, hybrid linkage, and centroid linkage schemes); (3) hybrid linkage combination techniques; (4) spatial clustering, and (5) split-and-merge. This typology reflects the approach to image segmentation as a clustering process, and the interaction between the grouping within the spatial domain (the segmentation itself) and the grouping in the measurement space (the clustering process). In Pal and Pal [Pal 93] the authors reviewed some image segmentation methods (distributed by 178 papers) by covering fuzzy and non-fuzzy techniques including colour image segmentation and neural network based approaches. The authors compare some of the methods and also provide some comments on quantitative evaluation of segmentation results.

Specialized surveys in a specific image segmentation topic can be found in [Davis 75] for edge detection, [Zucker 76] for region-based segmentation methods, [Sahoo 88, Sezgin 04] for thresholding techniques, [Reed 93] for texture and feature extraction methods, [Hoffman 87, Hoover 96] for range images, [Cheng 01, Lucchese 01] for colour images, and [Archip 02] reviews the use of neural networks for image processing in general and image segmentation in particular.

There has been a remarkable growth in the number of algorithms that segment colour images in the last decade [Cheng 01, Lucchese 01] and references on them. Most of the times, these are extensions of techniques originally devised for grey-level images. Thus, colour image segmentation algorithms exploit the well established background laid down in grey-level segmentation field. In other cases, they are ad hoc techniques specialized on the particular nature of colour information and on the physics driving the interaction between light and coloured materials.

Related surveys of interest in close fields of image segmentation can be found in the following papers: [Duncan 00] for medical image analysis and [Zitova 03] for image

registration methods. Other important surveys or reviews can be found in [Jain 99] for data clustering, [Jain 00] for statistical pattern recognition, [Antani 02] for the use of pattern recognition methods for abstraction, indexing and retrieval of images and video, [Shum 03] for image data compression and [Petersen 02] for image processing with neural networks.

The cooperation is useful when some sort of complementary properties are explored among the individual methods. For instance, it is common to combine edges with region-based approaches, as the first method presents good localization characteristics but it is sensitive to noise usually resulting in several edge gaps, while the region-based methods have poor accuracy on boundaries, although producing natural closed contours, and they are more insensitive to noise. Or, to overcome the over-segmentation result from a watershed approach we need the use of other post-processing methods. The human-computer cooperation is important when we need to accurately define the regions in a demanding image segmentation task or mandatory when we deal with crucial identification of regions in a medical image analysis segmentation problem.

In this formal context, the easiest form of cooperation appears at feature level as it is possible to conceive several levels of cooperation among the decision making processes using different sets of features. Other forms occur on the different ways of partitioning an image. There are different methods of partitioning that can cooperate. All of these forms of cooperation will be surveyed in a later section.

In our study, according to the work domain of each algorithm, we broadly classify the segmentation methods into three categories, namely image domain, feature domain, and methods that use a combination of these (cooperative methods). Feature domain is further divided into two main classes: thresholding and clustering methods. Image domain is split into boundary-based and region-based methods. According to the used framework, cooperative methods are classified as sequential, parallel, hybrid and interactive. Based on the above discussions, we adopt the classification of image segmentation as shown in Figure 2.1.

The desirable characteristics that a good image segmentation should exhibit were clearly stated in [Haralick 85]: "*Regions of an image segmentation should be uniform and homogeneous with respect to some characteristics such as grey tone or texture. Region interiors should be simple and without many small holes. Adjacent regions of a segmentation should have significantly different values with respect to the characteristic*

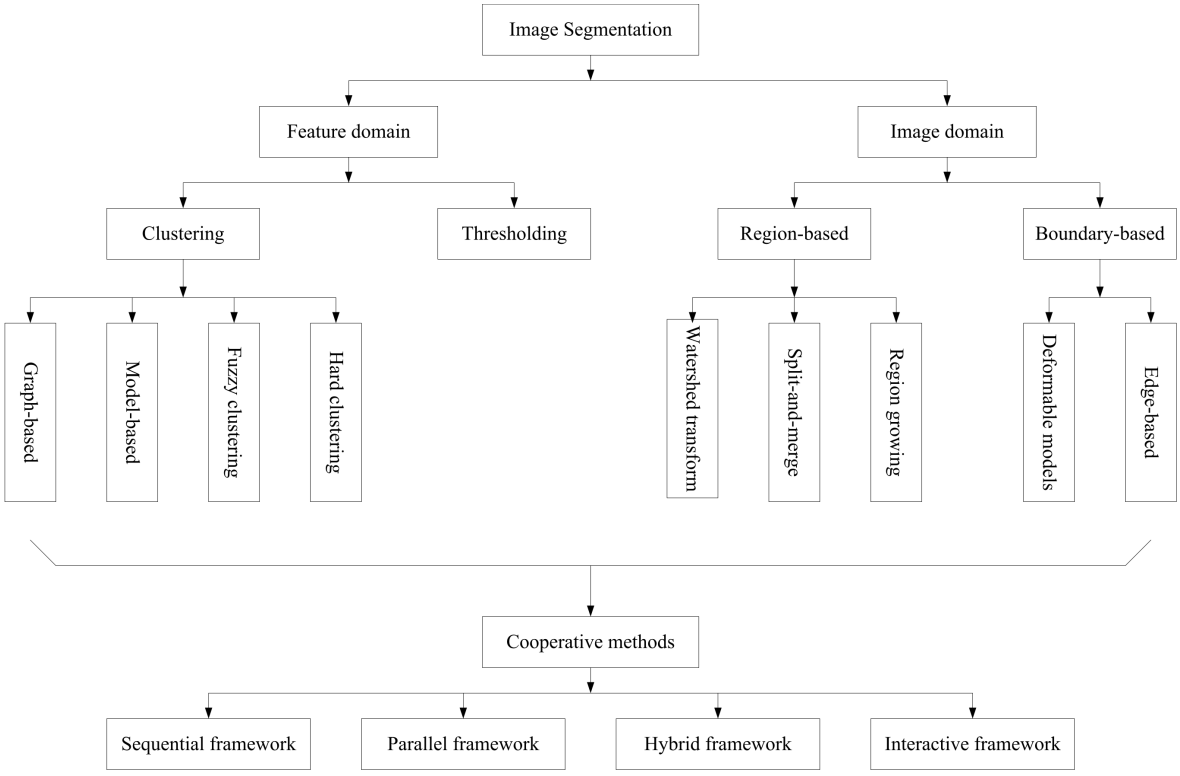


Figure 2.1: An overview of image segmentation approaches.

on which they are uniform. Boundaries of each segment should be simple, not ragged, and must be spatially accurate".

A more precise definition of segmentation accounting for the principal requirements listed above is given in [Pal 93] in the following way: "*Segmentation is a process of partitioning the image into some non-intersecting regions such that each region is homogeneous and the union of two adjacent regions is not homogeneous*".

Formally the segmentation process is the partition of an image I into k disjoint homogeneous regions (the segments) $R_1, R_2, ..., R_k$, obeying the following conditions:

1. $I = \bigcup_i R_i \quad \text{for } i = 1, 2, ..., k$
2. $R_i \cap R_j = \emptyset \quad \text{for } i \neq j$
3. $P(R_i) = TRUE \quad \text{for all } i$
4. $P(R_i \cup R_j) = FALSE, \quad \text{for } i \neq j \text{ and } R_i, R_j \text{ are adjacent}$

where the logical predicate $P(R)$ is the homogeneity property function of region R. This homogeneity function characterizes the uniformity of the region in terms of colour,

texture, shape or other features that enable the discrimination of a segment from the other segments. The consequence of the first condition is the complete spatial coverage of the image by all the detected non-overlapping regions. The non-overlapping is guaranteed by condition 2 which ensures that a pixel can be assigned to only one group. The pixel homogeneity within a region is implicit in condition 3, whilst condition 4 is an indication that two neighbouring regions must be different (in terms of the measured property).

Adjacency relationships between regions are not really taken into account in this definition, at the exception of the fourth condition which specifies that two adjacent regions cannot be similar. In order to compensate this lack, some authors suggest to use region adjacency graphs [Sanfeliu 02, Makrogiannis 05] or region similarity graphs [Monteiro 07].

As a result of the segmentation process we have a labelled image, corresponding at each region R_i $(i = 1, 2, ..., k)$ a label L_m $(m = 1, 2, ..., M)$. In general the number of regions, k, is equal to the number of labels, M, but they can also be different in some cases, with the restriction of neighbouring regions that must have different labels.

2.2 Image domain

In the literature of segmentation of grey-level images, many techniques have been suggested that try to satisfy both feature-space homogeneity and spatial compactness at the same time [Pal 93]. These approaches consider the connectivity of individual image pixels and then assign them to regions. According to the strategy preferred for spatial grouping, these algorithms are usually divided into boundary-based and region-based techniques.

The main advantages of the boundary-based methods for image segmentation rely on the accuracy of the location of the boundaries. Though as they are usually based on intensity gradient operators they are highly sensitive to noise and to small variations of the edges and they may produce incomplete and open edges with many gaps which will demand more powerful and time-consuming edge-linking tools. In many situations, as the analysis of outdoor scenes, the regions borders cannot be based on intensity or colour features only. Other texture features may be needed and eventually consider the combination of different cues, to completely describe scenes with a reasonable

complexity. Other advanced methods involving optimization methodologies, try to integrate several dimensions of the segmentation problem in order to obtain closed boundaries. However, they usually depend on the initialization and may be locked in a local minimum.

Region growing works well only if the initial seeds are representative of the regions of interest. The choice of the homogeneity and stopping criteria is crucial to the success of these methods and depends on the nature of the input image. These problems are overcome in the watershed algorithm which uses only an edge map as input and hence can be used to segment a variety of images. The algorithm produces the segmentation result without any user intervention. It is suitable for distributed implementation and it can produce significant system optimization.

2.2.1 Boundary-based methods

Boundary-based methods aim to segment an image from the edges of each region by locating the pixels where the intensity changes when compared to the pixels of its surroundings.

The basic approach for determining region boundaries is to detect the edges, by using an edge enhancement method, followed by thresholding the gradient magnitude. Here we consider a boundary as a contour in the image plane that corresponds to the separation between objects or surfaces in the world plane. An edge is an abrupt change in some feature in the image plane, as brightness, texture or colour. Edge detectors can be simple such as the Sobel or Roberts operators, or more complex such as the Canny approach. The output of most existing edge detectors can only provide candidates for the region boundaries, because the obtained edges are normally discontinuous or over-detected. Edge detection is usually followed by edge linking and boundary detection methods to obtain meaningful boundaries.

Edge-based

Edge detection aims to segment an image by finding the edges of each region by locating the pixels in the image where the intensity values change dramatically. These discontinuities are usually found by running a mask through the image. By using different values for the coefficients in the mask, different forms of edges could be sought

[Gonzalez 92]. It may also be necessary to perform some edge linking as the edges obtained by applying various masks to the image may not give complete boundaries.

The edge location is commonly computed from the local discontinuities in a local property as brightness [Canny 86], colour [Ruzon 01], texture [Will 00], or a combination of these local image cues [Martin 04]. In principle the edge detection operator can be applied simultaneously all over the image. One technique is high-emphasis spatial frequency filtering. Since high spatial frequencies are associated with sharp changes in intensity, one can enhance or extract edges by performing high-pass filtering using the Fourier operator.

The edge-based segmentation methods will respond to edge brightness or colour even if it does not correspond to a boundary as it happens in textured regions. Furthermore they are not able to detect boundaries between texture regions. On the other hand, texture based approaches may not detect brightness edges. These facts lead Martin et al. [Martin 04] to develop a method where all these features were combined. The approach of this paper is to look at each pixel for local discontinuities of these features at several orientations and scales, being the free parameters in each one of the features calibrated on the training data set. Malik et al. [Malik 01] also explored simultaneously brightness and texture as cues of contour, which are used as the primitives in a graph theoretical framework of normalized cuts for image segmentation.

Heath et al. [Heath 97] presented a study of five edge detection operators (Canny, Nalwa, Iverson, Bergholm, and Rothwell). The results show that significantly better performances are obtained when the algorithm parameters are adapted to each image than when one set of fixed parameters are used. The analysis of the relative performance of the algorithms resulted in a ranking of the algorithms as (Canny, Nalwa) $<$ Bergholm for fixed parameters and as (Iverson, Nalwa) $<$ (Rothwell, Bergholm, Canny) for adapted parameters. The performance increases from left to right and the parentheses group algorithms whose difference in performance was not statistically significant. The Canny algorithm had the highest performance when the parameters were adapted for each image, but the lowest performance when the parameters were fixed. They concluded that the choice of the edge detection algorithm depends on its application.

In the ideal case, the edge operator should find points lying only on the boundaries between regions. The main weaknesses of these methods are its sensitivity to image noise (as it is amplified by the gradient computation) and the generation of many

gaps between edge elements. To reduce the noise influence some authors proposed to firstly smooth the image by a low-pass filter. However, this will penalize the location properties of the edge detector. Resulting regions may not be connected hence edges need to be joined. To obtain a closed contour around the region other approaches for edge following and edge linking are needed to fill in the gaps. The Hough transform [Illingworth 88] can be used for boundary detection if the shape can be parameterized (e.g. as a line, a circle or an ellipsis).

A boundary detection scheme based on "edge flow" is proposed in [Ma 00]. This approach utilizes a predictive coding model to identify the direction of change in colour and texture at each image location at a given scale, and constructs an edge flow vector. By propagating the edge flow vectors the boundaries can be detected at image locations which encounter two opposite directions of flow in the stable state.

Deformable models

Active contours constitute a general technique of matching a deformable model onto an image by means of energy minimization. Since their introduction by Kass et al. in [Kass 88], deformable models have been used in many applications of image segmentation [Caselles 97, Davison 00, McInerney 00, Paragios 02, Han 03, Brox 06b]. Particularly, numerous algorithms based on the theory of deformable models have been proposed for the purpose of medical image segmentation [McInerney 96, Duta 98, Niessen 98, Paragios 03, Xu 04]. See [Xu 00] for a review on deformable models.

Various names such as snakes, active contours or surfaces, balloons and deformable contours or surfaces have been used in the literature to refer to deformable models [Xu 00].

Depending on the implementation there are essentially two types of deformable models: parametric deformable models [Kass 88, McInerney 95, Davison 00] and geometric deformable models [Caselles 97, Han 03]. Parametric deformable models represent curves and surfaces explicitly in their parametric forms during deformation. This representation allows direct interaction with the model and can lead to a compact representation for fast real-time implementation. Adaptation of the model topology such as splitting or merging parts during the deformation, can be difficult using parametric models. On the other hand geometric deformable models can handle topological changes naturally. These models, based on the theory of curve evolution [Sapiro 93]

and the level set method [Caselles 97], represent curves and surfaces implicitly as a level set of a higher-dimensional scalar function. They offer many advantages over parametric approaches. In addition to their straightforward implementation level sets do not require any parametrization of the evolving contour. Their parameterisations are computed only after complete deformation, thereby allowing topological adaptivity to be easily accommodated. Self-intersections, which are costly to prevent in parametric deformable models, are naturally avoided and topological changes are automated. Many fundamental properties of the active contours, such as the normal or the curvature, are also easily computed from the level set function. The ability to automatically change topology is often presented as an advantage of the level set method over explicit deformable models. Despite, in biomedical image segmentation, where the topology of the target shape is prescribed by anatomical knowledge, this behaviour is not desirable. Despite this fundamental difference, the underlying principles of both methods are very similar [Xu 00].

Kass et al. [Kass 88] introduced a global minimum energy contour called snakes or active contours. Given an initial approximation to a desired contour, a snake locates the closest minimum energy contour by iteratively minimizing an energy functional which combines internal forces to keep the active contour smooth, external forces to attract the snake to image features, and constraint forces which help to define the overall shape of the contour. A snake may be thought of as an elastic curve that, through minimization of an energy functional, deforms and adjusts its initial shape on the basis of additional image information to provide a continuous boundary [Davison 00].

The classic implementation of snakes by Kass et al. [Kass 88] allowed the problem to be reduced to a matrix form. However, this puts constraints on the energy functions. Davison et al. [Davison 00] proposed a less complicated form of the energy functions, and energy minimization is carried out by adjusting individual vertices on the snakes. This allows a larger range of energy functions, and the addition of internal energy functions like area and symmetry terms without complicating the minimization process as would be the case with the classic implementation.

The snakes approach had a large impact in the segmentation community. Yet, Cremers et al. [Cremers 07] identified several drawbacks on these approaches:

- The implementation of contour evolutions based on an explicit parameterisation requires a delicate re-parameterisation process to avoid self-intersection and

overlap of control or marker points.

- The explicit representation by default does not allow the evolving contour to undergo topological changes so that the segmentation of several objects or multiply-connected objects is not straight-forward.
- The segmentation obtained by a local optimization method is bound to depend on the initialization. The snake algorithm is known to be quite sensitive to the initialization. For many realistic images, the segmentation algorithm tends to get stuck in undesired local minimum, in particular, in the presence of noise.
- The snakes approach lacks a meaningful probabilistic interpretation. Extensions to other segmentation criteria such as colour, texture or motion are not straight-forward.

The snake method is known to solve boundary refinement problems by locating the object boundary from an initial plan. Though it should be stressed that the objective of these algorithms is generally to segment not a whole image but individual objects from an image.

Xu and Prince [Xu 98] presented a new class of external forces for active contour models that addresses some of the problems listed above. These fields, which they call gradient vector flow (GVF) fields, are dense vector fields derived from images by minimizing a certain energy functional in a variational framework. The minimization is achieved by solving a pair of decoupled linear partial differential equations that diffuses the gradient vectors of a grey-level or binary edge map computed from the image. They call the active contour that uses the GVF field as its external force a GVF snake. Particular advantages of the GVF snake over a traditional snake are its insensitivity to initialization and its ability to move into boundary concavities.

2.2.2 Region-based methods

Region-based techniques including region growing, region splitting, region merging and their combination attempt to group pixels into homogeneous regions. These techniques aim at partitioning the image domain by progressively fitting statistical models to the intensity, colour, texture or motion in each set of regions. These techniques rely on the assumption that adjacent pixels in the same region have similar visual features. In contrast to edge-based schemes, region-based methods tend to be less sensitive to

noise. Obviously, the performance of these approaches largely depends on the selected homogeneity criterion.

In the region growing approach, a seed region is first selected then expanded to include all homogeneous neighbours, and this process is repeated until all pixels in the image are labelled. In the region splitting approach the initial seed region is simply the whole image. If the seed region is not homogeneous it can be divided into four square sub-regions, which become new seed-regions. This process is repeated until all sub-regions are homogeneous. The region merging approach is often combined with region growing or region splitting to merge the similar regions for making a homogeneous region as large as possible.

Given the seeds, the seed region growing algorithm tries to find an accurate segmentation of images into regions with the property that each connected component of a region meets exactly one of the seeds. Moreover, high-level knowledge of the image components can be exploited through the choice of seeds.

Region growing

Region growing algorithms [Zucker 76, Adams 94, Sanfeliu 02, Fan 05, Grady 06] typically start from a pre-selected seed pixel, then progressively agglomerate points around it satisfying one or several homogeneity criteria such as intensity, colour or texture. These criteria can be defined according to local, regional and global relationships. The growth process stops when no more points can be added to the region. A common post-processing approach consists of a merging phase that eliminates small regions or neighbouring regions with similar attributes, generating broader regions accordingly. Fan et al. [Fan 05] presented a recent comparative study on seed region growing algorithms.

This strategy needs an initial set of seeds to work, as well as a general homogeneity criterion to join neighbouring regions. Though it is difficult to specify homogeneity because the concept of homogeneity is often vague and fuzzy and it is not translated easily into a computable criterion. Region-growing can be considered as a sequential clustering or classification process. Thus, the results may depend on the order according to which image points are processed. The main advantage offered by this kind of techniques is that regions obtained are certainly spatially connected and rather compact.

These methods are known to be sensitive to the seed choice process together with the way segment statistics are computed, which is done to guess whether two adjacent regions might join or not [Sanfeliu 02]. The application of a region growing process can lead to different types of errors [Pavlidis 90]: a) region boundaries are not close to edges; b) the boundaries are close but they are not coincident with the edges; c) there are edges not corresponding to boundaries.

Adams et al. [Adams 94] proposed the seeded region growing (SRG) where the initially defined seed pixels (by user interaction or by some pre-processing stage) control the growing process by measuring the dissimilarity between adjacent pixels. Given the set of seeds, each step of SRG tries to find an accurate segmentation into regions with the property that each connected component of a region meets exactly one of the seeds. These initial seeds are further replaced by the centroids of the generated homogeneous regions, and by incorporating the additional pixels step by step. An advantage of SRG is that the high-level knowledge of the image components can be exploited through the choice of seeds [Fan 05]. However, a poor starting estimate of region seeds or bad pixel sorting may result in an incorrect segmentation.

Hojjatoleslami and Kittler [Hojjatoleslami 98] presented a region growing approach by pixel aggregation which uses similarity and discontinuity measures. A unique feature of the proposed approach is that in each step at most one candidate pixel exhibits the required properties to join the region. They argue that this makes the direction of the growing process more predictable. The procedure offers a framework in which any suitable measurement can be applied to define a required characteristic of the segmented region.

Deng and Manjunath [Deng 01] proposed the *JSEG* algorithm, a colour quantization technique to smooth the image colours into several representative classes (J-images). The J-values measure the distances between different classes over the distances between the members within each class. For the case of an image consisting of several homogeneous regions, the colour classes are more separated from each other and the value of J is large. The scheme has the ability to segment colour textured images without attempting to estimate a specific model for a texture region. Instead, it tests for the homogeneity of a given colour-texture pattern. The basic idea of the algorithm is to separate the segmentation process into two independent stages, colour quantization and spatial segmentation. In the first stage colours in the image are quan-

tized to several representative classes that can be used to differentiate regions in the image. This quantization is performed in the colour space without considering the spatial distribution of the colours. Then, the image pixel values are replaced by their corresponding colour class labels, thus forming a class-map of the image. In the second stage, a region growing method is then performed directly on this class-map without considering the corresponding pixel colour similarity.

Grady [Grady 06] proposed a method for performing multi-label, interactive image segmentation. Given a small number of pixels with user-defined labels (seeds), the algorithm operates by assigning each unseeded pixel to the label of the seed point that a random walker starting from that pixel would be most likely to reach first, given that it is biased to avoid crossing object boundaries (i.e., intensity gradients).

Most of region growing methods have an inherent dependence on the order in which the pixels and regions are examined. This weakness implies that a desired segmented result is sensitive to the selection of the initial growing pixels. Wan and Higgins [Wan 03] defined a set of theoretical criteria for a subclass of region growing algorithms that are insensitive to the selection of the initial seeds. This class of algorithms referred to as symmetric region growing algorithms, leads to a single-pass region growing algorithm applicable to any image dimension.

Mehnert and Jackway [Mehnert 97] have confirmed that a different order of processing pixels leads to different final segmentation results. They also noticed two types of order dependencies. The first type is called inherent order dependencies, while the second is called implementation order dependencies. They also presented an algorithm that improves the seeded region-growing algorithm by making it independent of the pixel order of processing and making it more parallel. Parallel processing ensured that the pixels with the same priority were processed in the same manner simultaneously.

Region splitting and merging

These methods start with an initial inhomogeneous partition of the image and then keep splitting until reaching homogeneous partitions as proposed in a starting paper [Horowitz 76], describing the split-and-merge techniques. In this approach an image is initially subdivided into a set of disjoint regions and then merged and/or split until each region satisfies some conditions indicating that it is one segment. A data structure used to implement this procedure is the quadtree representation. In the first step, the

whole image is considered as one region. If this region does not satisfy a homogeneity criterion the region is split into four quadrants (subregions) and each quadrant is tested in the same way; this process is recursively repeated until every square region created in this way contains homogeneous pixels. After the splitting phase, there are usually many small and fragmented regions which have to be somehow connected in a merging phase. Therefore, in a next step all adjacent regions with similar attributes may be merged following other (or the same) criteria. The region adjacency graph (RAG) is the data structure commonly adopted in this phase. The process ends when no more splitting or merging is possible.

Gevers [Gevers 02] described a split-and-merge method based on Delaunay triangulation. This tessellation grid is adaptive in the sense that it is data dependent by measuring region and edge properties. A recent paper inspired in the same split-and-merge basic principle is presented in [Chung 05]. Here the authors proposed a quadrilateral-based segmentation framework, where the splitting phase is computed on a gradient image, which is followed by a merging process.

Watershed transform

Watershed transform is an important paradigm for image segmentation, and it is a main step in several hybrid image segmentation frameworks (see Section 2.4). Although watershed is usually considered as a region-based approach, De Smet et al. [De Smet 99] pointed out that the watershed transform has proven to be a powerful basic segmentation tool that can hold the attributed properties of both edge detection and region growing techniques which makes it a cooperative approach.

The main drawback of watershed transform for image segmentation is the over segmentation introduced by creating a large number of small regions. To overcome this problem pre-processing or post-processing phases are considered by several authors. The pre-processing phase has a main goal to regularize image intensities variations by image denoising, using anisotropic filters (as used in Weickert [Weickert 01] or special application oriented image heuristic enhancement steps [Adiga 01] or edge preserving noise filters [Haris 98]. It is also common the introduction of a post-processing phase after applying the watershed transform for merging the less significant regions in order to obtain larger regions with a better correspondence to objects. Other authors, as Haris et al. [Haris 98] and Adiga et al. [Adiga 01] used both a pre-processing

and post-processing steps. Nevertheless the performance of a watershed-based image segmentation method depends largely on the algorithm used to compute the gradient.

The main advantages of the watershed transform are:

- it produces coherent regions where boundaries are always guaranteed to be connected and closed. Unlike traditional edge detectors which most often form disconnected boundaries that need post-processing to produce closed regions, watershed transforms produce closed contours and give good performance at junctions and places where the object boundaries are diffuse. This means that all of the boundary pixels for a single object can be trivially extracted without complex tracking or edge-linking, thereby avoiding one of the pitfalls of many edge detection methods;
- the boundaries of the resulting regions always correspond to contours which appear in the image as obvious contours of objects. This is in contrast to split-and-merge methods where the first splitting is often a simple regular sectioning of the image leading sometimes to unstable results;
- gradient watershed regions can be used to interactively construct the image region associated with an object of interest;
- the union of all the regions form the entire image region.

One of two different algorithms are generally used to implement watershed segmentation, namely immersion and rainfalling simulation. Each of these can be used to detect the segments in the image either directly or using morphological operators. As watershed is a method largely used in this thesis, we briefly review some of these approaches as follows.

Since the early 1990s, there has been a considerable amount of scientific work on the watershed transform that was originally proposed by Beucher and Lantuéjoul [Beucher 79] as an image processing tool. An excellent and recommended overview on definitions, algorithms, and parallelisation strategies was published by Roerdink and Meijster [Roerdink 01].

A major breakthrough in the implementation of the watershed was made by Vincent and Soille [Vincent 91] with the introduction of the first queue based implementation of the watershed transform. Basically, the algorithm consists of two steps: a sorting step and a flooding step. The sorting step first computes the frequency distribution of

each image grey level. The cumulative frequency is then computed so that each pixel can be assigned to a unique cell in a sorted array. In the flooding step the catchment basins are recursively grown by using a FIFO (First In First Out) ordered queue for the computation of the geodesic influence zones. The queue based flooding is indeed quite fast but remains computationally intensive. This is due to the fact, that each update step of the catchment basins requires a full scan of the image. Since updating is performed recursively for each of the grey-levels in the image, the total number of scans can be quite large.

Two problems arise when applying the above watershed method to an image. The first problem is the occurrence of flat regions, i. e. regions of constant grey value, as discussed in numerous publications [Gauch 99, Stoev 00, Roerdink 01]. The second problem, which is partly linked to the flat region problem, is the dependency of the watershed location on both the used algorithm and the grid connectivity [Roerdink 01].

Moga and Gabbouj [Moga 97] described a parallel approach for computing the watershed transformation, based on rainfalling simulation within a grey-scale image. The first step transforms the original image into a lower complete image. In this lower complete image the pixels belonging to a non-minimum flat region are labelled with the geodesic distance to the flat region's nearest lower pixel. In doing so, a second ordering relation for the pixels in a non-minimum flat region is introduced in the resulting image. Afterwards a raindrop starts at each pixel and its path towards the line with the steepest descent is followed until a regional minimum is reached. The set of all pixels attracted on the way to a particular regional minimum defines the catchment basin for this minimum.

Stoev and Strasser [Stoev 00] presented a sequential approach where every pixel p is compared with the adjacent pixels and if possible the path of the steepest descent is followed and p is pushed on a stack S_c containing the pixels on the current path. Otherwise, if a flat region is reached, the whole flat region is processed in order to determine the nearest outdoor. If there are outdoors, the inner pixels are assigned to the appropriate outdoors. Every time a regional minimum is reached, which is either a flat region without outdoors or an isolated minimum, the pixels pushed on the stack S_c are traversed and marked with the label of the reached minimum.

Weikcert [Weickert 01] introduced a pre-processing step before applying the watersheds. It includes a regularization step using two partial differential equations based

methods (a non-linear isotropic diffusion filter and a convex quadratic variational image restoration method) followed by watershed and a simple region merging process.

Gauch [Gauch 99] avoided flat region problems by working with Gaussian smoothed floating point images. This removes all regions with uniform intensity. However, this approach has several problems: if the neighbours of an edge decrease in intensity rapidly on the left and gradually on the right the detected location of the edge will be to the right of the correct position; in a lot of smoothed images which have few intensity minima, the tops of some ridge like structures may be missed.

Grau et al. [Grau 04] identified the two common drawbacks for watershed based image segmentation, over segmentation and sensitivity to noise, together with two particular inconvenient in medical image segmentation: poor detection of significant areas of low contrast, and poor detection of thin structures. To overcome these drawbacks the authors defined an improved version of the classical watershed transform, enabling the use of prior knowledge of the objects that can be adapted depending on the application, namely using the information available from a statistical anatomical atlas registration, through the use of markers.

2.3 Feature domain

A number of approaches to segmentation are based on finding compact clusters in some feature space [Comaniciu 02, Felzenszwalb 04]. In this technique, a vector of local properties ('features') is computed at each pixel and then mapped into the feature space. Features such as intensity, texture or motion are the commonly studied parameters. Significant features will be shared by numerous pixels, and thus form a dense region in feature space. The feature space is then clustered, and each pixel is labelled with the cluster that contains its feature vector. Clusters in feature space can then be used for image segmentation, typically by fitting a parametric model to each cluster and then labelling the pixels whose feature vectors lie in the cluster with the parameters. The common techniques include histogram thresholding, clustering and graphs.

These approaches generally assume that the image is piecewise constant because searching for pixels that are all close together in some feature space implicitly requires that the pixels are alike (e.g., similar colour). Comaniciu and Meer [Comaniciu 02] used a technique where feature space clustering first transforms the data by smoothing

it in a way that preserves boundaries between regions. This smoothing operation has the overall effect of bringing points in closer clusters together. The method then finds clusters by dilating each point with a hypersphere of some fixed radius, and finding connected components of the dilated points.

Segmentation algorithms which exclusively operate in some feature spaces return segments that are expected to be homogeneous with respect to the characteristics represented in those space. However, there is no guarantee that these segments also show spatial compactness, which is a desirable property in segmentation applications beside homogeneity. For instance, histogram thresholding accounts in no way for the spatial locations of pixels; the description they provide is global and it does not exploit the important fact that points of the same object are usually spatially close due to surface coherence. On the other hand, if pixels are clustered exclusively on the basis of their spatial relationships, the final result is likely to be with regions spatially well connected but with no guarantee that these regions will also be homogeneous in a certain feature space.

2.3.1 Thresholding methods

Thresholding techniques are based on the assumption that adjacent pixels whose value (grey level, colour value, texture) lies within a certain range belong to the same class [Fan 01]. These methods achieved reasonable performance when the input is characterized without noise and with small number of regions. This explains why these methods are mainly used in text segmentation [Solihin 99, Kim 02]. For a review of thresholding techniques readers are referred to the survey papers [Sahoo 88, Pal 93, Sezgin 04].

Histograms have been extensively used in image analysis mainly for two reasons: they provide a compact representation of large amounts of data, and it is often possible to infer global properties of the data from the behaviour of their histogram [Delon 07]. The histogram of intensities of an image made of different regions shall exhibit several peaks, each one ideally corresponding to a different region. Finding suitable threshold values that could find valleys between peaks in the histogram and produce a segmentation of the grey level image into objects and background is the core of the thresholding operation.

The traditional thresholding approach is basically a one-stage thresholding ap-

proach where an image is separated into two classes of pixels: the object pixels and the background pixels. Global thresholding techniques attempt to find a single threshold value that best separates the two classes of pixels in an image. In local or adaptive thresholding the threshold values are determined locally, e.g. pixel by pixel or region by region. Then, a specified region can have 'single threshold' that is changed from region to region according to the threshold candidate selection for the given area.

Among the algorithms proposed for histogram segmentation we can distinguish between parametric and non-parametric approaches. In the first ones [Papamarkos 94, Wang 04a] the histogram is considered to be a probability density function of a Gaussian and the segmentation problem is reformulated as a parameter estimation followed by pixel classification. If the number of objects is known optimization algorithms can estimate efficiently the parameters of these distributions. The main drawback of these approaches is that histograms obtained from real images cannot always be modelled as mixtures of Gaussians, for example luminance histograms of natural images.

Non-parametric approaches do not use any assumption on the underlying data density and they divide the histogram into several segments by minimizing some energy criterion. Among them we have methods that analyse the histogram of the whole image [Cheriet 98, Solihin 99, Kim 02], and methods based on the histogram of edge pixels [Wang 03a].

An early review of thresholding methods was reported in the highly cited paper of [Sahoo 88]. Sahoo et al. surveyed segmentation algorithms based on thresholding and attempted to evaluate the performance of some thresholding techniques using uniformity and shape measures. They categorized global thresholding techniques into two classes: point-dependent techniques (grey-level histogram based) and region-dependent techniques (modified histogram or co-occurrence based). Discussion on probabilistic relaxation and several methods of multi-thresholding techniques was also given.

More recently, Sezgin and Sakur paper [Sezgin 04] presented an exhaustive survey of forty (40) image thresholding methods both global and local. They conduct a quantitative performance evaluation and conclude that local methods perform better. Nevertheless this evaluation took into consideration only text document images that were degraded with noise and blur.

Cheriet et al. [Cheriet 98] presented a general recursive approach for image segmentation by extending Otsu's method [Otsu 79]. This approach has been implemented

in the area of document images. This approach segments the brightest homogeneous object from a given image at each recursion, leaving the darkest homogeneous object. Li et al. [Li 97] suggested that the use of two dimensional histograms of an image is more useful to find thresholds for segmentation rather than just using grey level information in one dimension. In 2D histograms, the information on pixels as well as the local grey level average of their neighbourhood is used.

Kim et al. [Kim 02] proposed a locally adaptive thresholding algorithm where a text document image is regarded as a 3D terrain and its local property is characterized by a water flow model [Beucher 79]. The water flow model locally detects the valleys corresponding to regions that are lower than neighbouring regions. The deep valleys are filled with dropped water whereas the smooth plain regions keep up dry. The final step in this method concerns the application of a global thresholding on a difference image between the original terrain and the water-filled terrain. A shortcoming of this method is the selection of two critical parameters, namely, the amount of rainfall and the mask size which is done on an experimental basis. Besides, the final binarization results are obtained by applying a global thresholding method to the amount of filled water. Thus, objects in a poor contrast background are often removed as the corresponding valleys are only filled with a little water.

Other authors proposed thresholding techniques which select threshold from histogram of edge pixels. In [Wang 84] edge pixels are first classified on the basis of their neighbourhood as being relatively dark or relatively light. Then two grey level histograms are obtained respectively for these two sets of edge pixels. The threshold is selected as one of the highest peaks of the two histograms. By recursively using the procedure, multiple thresholds can be obtained. In [Wang 03a] for each given object, its threshold is deduced from the histogram of the discrete sampling points of boundary.

The poor performance of histogram thresholding based methods in real images can be attributed to the fact that, generally, the profiles of the histograms are rather jagged giving rise to spurious peaks that complicate the selection of suitable threshold values. This is due to objects with non-uniform colour, intensity gradients caused by illumination or variations in surface reflectance, texture, noise, and backgrounds that are not uniformly coloured; to overcome this problem, some smoothing filters are usually adopted. Moreover, it is often the case that even if suitable thresholds can be found, the resulting segmentation is inaccurate because of overlap in grey-level

intensities between different elements of the image, which leads to disconnected regions with the same label. In complex images it also becomes difficult to separate different peaks in the histogram and to determine how many thresholds are required. Another weakness of thresholding segmentation methods is that they neglect all of the spatial information of the image and do not cope well with noise or blurring at boundaries [Adams 94].

2.3.2 Clustering methods

Clustering techniques appeared earlier in the literature and were used in numerous applications [Jain 99]. Following the selection of image features usually based on intensity, colour or texture, clustering operates on the feature space in order to capture the global characteristics of the image. Ignoring spatial information and using a specific distance measure, the feature samples are handled as vectors and the objective is to group them into compact but well-separated clusters. After the clustering process is completed the data samples are mapped onto the image plane typically by fitting a parametric model to each cluster and then labelling the pixels according to each parametric model to produce the final regions [Makrogiannis 05].

Turi [Turi 01] classified clustering algorithms as hierarchical or partitional. Hierarchical techniques involve the clusters themselves being classified into groups, where the process is repeated at different levels [Shi 00, Boykov 01b, Barbu 05]. Partitional techniques form clusters by optimizing a clustering criterion, where the classes are mutually exclusive, thus forming a partition of the data [Pham 02, Chen 04, Cai 07].

A characteristic of the hierarchical clustering techniques is that once a sample is assigned to a particular cluster it cannot be changed. Therefore if the sample is incorrectly assigned to a particular cluster at an early stage there is no way to correct the error. This is where the partitional clustering techniques such as hard or fuzzy clustering have an advantage over the hierarchical clustering techniques, as partitional techniques allow a data point to be reassigned to a different cluster if it improves the clustering.

Partitional clustering techniques present, however, some disadvantages: if the same data is input in a different order it may produce different clusters. Pixels from non-adjacent regions of the image can be grouped together, if there is an overlap in their

feature space values which produces several noisy areas and incomplete region borders in the segmentation results.

The partitional form of clustering where a class label is assigned to each data value identifying its class is referred to by some authors as hard clustering [Jain 99]. In recent years fuzzy clustering approaches have been developed where a fractional degree of membership for each cluster is assigned to each data value [Udupa 96].

For the case of natural images, the data-clustering problem is quite complex and the literature of clustering algorithms is very rich. [Jain 99, Turi 01] presented excellent reviews on clustering methods. The method known as K-means and its fuzzy counterpart fuzzy C-means are some of the most common techniques in the segmentation field. Based on the assumptions that the number of clusters is known a priori and the cluster shape is approximately spherical, these algorithms converge to the final cluster centres. The main difference between hard K-means and fuzzy C-means is that fuzzy partition allows the pixels to partially belong to different clusters.

Hard clustering

Currently K-means is among the most popular clustering algorithms due to its simplicity and efficiency in unsupervised classification. It starts with a random initial partition and keeps reassigning the features to clusters based on the similarity between the feature and the cluster centres until a convergence criterion is met. A major problem with this algorithm is that it is sensitive to the selection of the initial partition and may converge to a local minimum of the criterion function value if the initial partition is not properly chosen.

In [Pappas 92], Pappas indicated two problems with K-means algorithm which are: use of no spatial constraints and it assumes that each cluster is characterized by a constant intensity. In order to overcome these problems Pappas introduced a generalization of the K-means clustering algorithm and applied this procedure on grey-level images. This approach aims to separate the pixels in the image into clusters based not only on their intensity but also on their relative spatial location. This algorithm considers the segmentation of grey-level images as a maximum a posteriori probability (MAP) estimation problem.

The advantages of K-means are that it is a very simple method and it is based on intuition about the nature of a cluster, so the intra-cluster error should be as small as

possible [Turi 01]. K-means clustering has although some weaknesses: the number of clusters must be known a priori; if the same data is inputted in a different order it may produce different clusters; it is sensitive to initial conditions. We never know which feature contributes more to the grouping process since it assumes that each attribute has the same weight; weakness of arithmetic mean is not robust to outliers. Very far data from the centroid may pull the centroid away from the real one. The final clusters have circular shape because K-means is based on centroid distances.

Work by Turi [Turi 01] described a method of automatic determination of the optimal number of clusters in K-means clustering. It proposes a validity measure using the ratio of intra-cluster and inter-cluster measures incorporated with a Gaussian multiplier. The optimal number of clusters is found by minimizing the validity measure.

The mean-shift algorithm is a non-parametric statistical method that finds peaks (local maxima) of the histogram without estimating the underlying density function. It has been used for the first time by Fukunaga and Hostetler in [Fukunaga 75] with the goal of proposing an intuitive estimation of the gradient probability density of a set of points; later it has been used extensively for image segmentation [Comaniciu 02].

This method is designed to locate the centroids of clusters with high local density in the feature space. To satisfy this objective, mean-shift uses a simple mechanism by shifting iteratively every pixel to the mean of its neighbouring pixels. A segmentation of an image I into a set of k disjoint regions where each region R_i is described by its contour Γ_i and its model parameters Θ_i, $R_i = (\Gamma_i, \Theta_i) : i = 1, ..., k$, with the latter involving the estimation of a mean vector and a covariance matrix $\Theta_i = \{\mu_i, \Sigma_i\}$.

The algorithm starts with a set of initial guesses for cluster centres, and then repeats the following two steps iteratively: a) Compute a weighted mean of the points within a small window centred at the current centroid location, using weights based on the distance between each point and the current centroid; b) Update the centroid location to be the newly estimated weighted mean (by this operation the centroid location is shifted to the mean of the local distribution). Each data point becomes associated with a point of convergence which represents the local mode of the density in the d-dimensional space. Convergence points sufficiently close in the joint domain are fused to obtain the homogeneous regions in the image. This procedure is repeated until a convergence condition is satisfied.

The mean-shift algorithm produces segmentations that correspond well to human

perception. However, this algorithm is quite sensitive to its parameters and it tends to detect too many peaks in histograms coming from real noisy data which results in evident over-segmentation. Some criterion is, therefore, needed to decide which peaks from the detected ones correspond to true modes.

Fuzzy clustering

In the last years there has been considerable interest in the use of fuzzy segmentation methods, which are able to retain more information from the original image than hard segmentation methods. Fuzzy clustering theory was first introduced by Zadeh [Zadeh 65] to generalise the conventional cluster theory. Based on the definition of a fuzzy event [Zadeh 65] grey level image can be seen as a fuzzy event modelled by a probability space.

Fuzzy C-means (FCM) is one of the most well-known methodologies in clustering analysis [Bezdek 93, Udupa 96]. The reason for its success is due to the introduction of fuzziness for the belongingness of each image pixels. Unlike hard clustering methods like K-means which force pixels to belong exclusively to one class during their operation and in their output, FCM methods allow pixels to belong to multiple classes with varying degrees of membership. The degree is decided by a membership function which depends on how compatible the member is to the properties of the cluster. The FCM algorithm classifies the image by grouping similar data points in the feature space into clusters. This clustering is achieved by iteratively minimizing a cost function that is dependent on the distance of the pixels to the cluster centres in the feature domain.

In most situations FCM uses the common Euclidean distance which supposes that each feature has equal importance in FCM. This assumption seriously affects the performance of FCM since in most real world problems features are not considered to be equally important. In [Wang 04b], Wang et al. proposed a new robust metric, which is distinguished from the Euclidean distance, to improve the robustness of FCM. The feature-weight learning FCM technique [Yeung 02, Wang 04b] assigns various weights to different features to improve the performance of clustering. The spatial function can be estimated at each iteration and incorporated into the membership function which makes the new FCM technique less sensitive to noise. Another drawbacks of FCM include its computational complexity and the fact that it not consider spatial information in image context, which makes it very sensitive to noise and other imag-

ing artefacts. Recently, many researchers have incorporated local spatial information into the original FCM algorithm to improve the performance of image segmentation [Pham 02, Chen 04, Cai 07].

Pham [Pham 02] modified the FCM function by including a spatial penalty on the membership functions. The penalty term leads to an iterative algorithm, which is very similar to the original FCM and allows the estimation of spatially smooth membership functions. Ahmed et al. [Ahmed 02] proposed the FCM_S algorithm to compensate for the intensity inhomogeneity and to allow the labelling of a pixel to be influenced by the labels in its immediate neighbourhood.

In order to reduce the computational load of FCM_S Chen and Zhang [Chen 04] proposed two variants, FCM_S1 and FCM_S2, which simplified the neighbourhood term of the objective function. The essence of FCM_S1 is to make both the original image and the corresponding mean-filtered image have the same prototypes or segmentation result with aiming to guarantee the grey homogeneity. However, this variant is unsuitable for the images corrupted by impulse noise such as salt and pepper noise. In order to overcome that problem Chen and Zhang proposed the FCM_S2 in which the median filtered image replaces the mean filtered one.

As pointed out by Cai et al. [Cai 07] these approaches still have the following disadvantages: 1) although the introduction of local spatial information to the corresponding objective functions enhances their robustness to noise to some extent, they still lack enough robustness to noise and outliers, especially in absence of prior knowledge of the noise; 2) in their objective functions, there is a crucial parameter α used to control the effect of the neighbours term and to balance between robustness to noise and effectiveness of preserving the details of the image, and generally its selection has to be made by experience; and 3) the time of segmenting an image is heavily dependent on the image size.

Szilagyi et al. [Szilagyi 03] proposed the enhanced FCM (EnFCM) method to accelerate the image segmentation process. In this approach a linearly-weighted sum image is in advance formed from both original image and its local neighbour average grey image, and then clustering of the summed image is performed on the basis of the grey level histogram instead of pixels in the image. Consequently, the time complexity of EnFCM is drastically reduced.

To speed up even more the segmentation process, Cai et al. in their recent paper

[Cai 07] proposed the Fast Generalized Fuzzy C-means (FGFCM) algorithm for fast and robust image segmentation. They replace the parameter α, that is shared by EnFCM, FCM_S and its two variants, by a locality factor S_{ij} where the i-th pixel is the centre of the local window (for example, 3×3) and j-th pixels are the set of the neighbours falling into a window around the i-th pixel. This factor incorporates simultaneously both the local spatial relationship and the local grey-level relationship and its value varies from pixel to pixel for the image within the local window, i.e., spatially and grey dependent. Thus, S_{ij} can be adaptively determined by local spatial and grey-level information rather than artificially or empirically selected like α. In the second step the fast segmentation method [Szilagyi 03] is performed on the grey-level histogram of the generated image.

Krishnapuram and Keller [Krishnapuram 93] proposed a *possibilistic* clustering algorithm in which the membership values for a given feature pixel across all clusters was not constrained to add to one. Barni et al. [Barni 96] have shown on several series of examples that the classical possibilistic C-means algorithm gives rise to identical clusters. Such a problem is essentially due to the missing of an inter-class distance. Khrisnapuram and Keller [Krishnapuram 96] have proposed to consider an iterative version of the algorithm. If a class is found, pixels of cluster data having values greater than an appropriate cut are removed from the image partition. Processing is iterated again until the achievement of inconsistent clusters. However, it caused clustering being stuck in one or two clusters.

Zhang and Chen [D. Zhang 04] proposed a spatially constrained kernelized FCM (SKFCM) which uses a different penalty term containing spatial neighbourhood information in the objective function and simultaneously the similarity measurement in the FCM was replaced by a kernel-induced distance.

Model clustering

A feature vector is labelled with a probability distribution over clusters instead of a single cluster. A number of techniques for doing spatially coherent clustering have been developed in a Bayesian framework. Marroquin et al. [Marroquin 03] referred to such methods as segmentation/model estimation methods.

Statistical approaches, especially parametric ones, labels pixels according to probability values, which are determined based on the intensity distribution of the image.

With a suitable assumption about the distribution, statistical approaches attempt to solve the problem of estimating the associated class label, given only the intensity for each pixel [Zhang 01b]. This formulation of the segmentation problem leads naturally to a hierarchical model [Barker 98].

Markov Random Fields (MRF) have been and are increasingly being used to model a prior belief about the continuity of image features such as region labels, textures, edges, or motion. An MRF can be used to model the discrete label field containing the individual pixel classification. The methodology of using MRF models to the problem of segmentation has emerged later and has created a lot of interest [Won 92, Panjwani 95, Barker 98, Sarkar 00]. The MRF forms a probabilistic model for a set of variables that interact on a lattice structure. The distribution for a single variable at a particular site is conditioned on the configuration of a predefined neighbourhood surrounding that site. This effectively defines the Markov property of the process: the process is Markov not in the causal or even the bilateral sense, but with respect to this particular neighbourhood structure [Barker 98].

Difficulties associated with MRFs are the proper selection of the parameters controlling the strength of spatial interactions and they require computationally intensive algorithms [Held 97]. These methods work well in supervised mode, wherein the number of regions and their associated parameters are known or can be estimated beforehand. A solution to this problem consists in iterating an estimation/segmentation cycle [Won 92]. Given a candidate number of regions and an initial random set of region parameters, a first segmentation is computed. Region parameters are then recomputed using the current segmentation. This cycle is repeated, with different candidate region numbers, several times until convergence. The number that optimizes a model fitting criterion is retained as the true number of regions [Won 92].

To perform semi-unsupervised segmentation, where the number of classes is assumed to be known a priori, a method of concurrently estimating the underlying image and any associated model parameters is required. Alternatively, the problem may be viewed as one of parameter estimation from incomplete data. The Expectation-Maximization (EM) algorithm was first proposed by Dempster et al. [Dempster 77] as an iterative maximal-likelihood procedure for parameter estimation from missing or incomplete data.

The EM clustering provides a framework for incorporating our knowledge about a

domain. K-means and the hierarchical algorithms make fairly rigid assumptions about the data. For example, clusters in K-means are assumed to be spheres. EM clustering offers more flexibility. The clustering model can be adapted to what we know about the underlying distribution of the data. The methodology has been extensively applied to the problem of image segmentation [Belongie 98, Zhang 01b, Carson 02, Robles-Kelly 02]. The EM algorithm is an iterative process where each iteration consists of two steps. The first of these (E-step) finds an expression for the expected value of the log likelihood over the hidden data, given the previous parameter estimate. The second step (M-step) maximises this expectation over the parameter space.

Note that like thresholding and clustering algorithms, EM does not directly incorporate spatial modelling and it can therefore be sensitive to noise and intensity inhomogeneities. Recently, a diffused expectation–maximization (DEM) algorithm has been proposed for grey-level images [Boccignone 04], in which a diffusion step provides spatial constraint satisfaction.

Minimum Description Length (MDL) principle suggests that the optimal model is one which minimizes the sum of the coding length of the data given the model and the coding length of the model itself, that is, the best fitted model is the one that produces the shortest code length of the data. These two lengths formally correspond to likelihood and prior probability in the Bayesian framework, respectively. Therefore, minimizing description length is equivalent to maximizing a posterior probability. MDL has been effectively applied to image segmentation by a number of authors [Pateux 00, Galland 03]. The advantage of applying MDL to merge regions is that decisions are made adaptively by taking into account local region statistics.

Hierarchical clustering (Graph-based)

Hierarchical clustering techniques are based on the use of a proximity matrix indicating the similarity between every pair of data points to be clustered [Turi 01]. The final result is a "*dendogram representing the nested grouping of patterns and similarity levels at which grouping change*" [Jain 99]. One of the drawbacks of hierarchical algorithms is the time complexity. The memory space complexity is also a problem due to the similarity matrix needing to be stored.

An interesting category of hierarchical clustering algorithms is originated from graph theory. These methods generally present interesting results and a complete

analysis and a comparison of the different methods of graph cuts are proposed in [Soundararajan 03].

Graph cut algorithms use the Gestalt principles of perceptual grouping to form the image regions. These algorithms try to divide the initial graph into subgraphs that correspond to image regions. Though several partitioning techniques exist they all use the same underlying representation of the image: a graph $G = (V, E, W)$ with vertices (nodes) $v \in V$ corresponding to image elements (which may be pixels, feature descriptors, atomic regions, or others), links[2] $e \in E \subseteq V \times V$ and an associated weighted matrix W. The link between two vertices v_i and v_j, is denoted by e_{ij}. The weight of a link $w_{i,j}$ is proportional to the similarity between the nodes v_i and v_j and it is usually referred to as the affinity between elements i and j in the image.

A graph theory based on image segmentation consists of two main steps: 1) the graph creation and 2) the graph partitioning. These algorithms are usually applied on the pixel-based graph, where the nodes correspond to the pixels and the links to their connections. The weights associated to an edge express the (dis)similarity of the pair of nodes it connects. The similarity value can use any number of image cues including grey level intensity, colour, texture, and other image statistics. It is also common to add a distance term that ensures that the graph is sparse by linking together only those nodes that correspond to elements in the image that are near each other. Once the graph is built, the segmentation process consists on determining which subsets of nodes and links correspond to homogeneous regions in the image. The key principle here is that nodes that belong to the same region or cluster should be joined by links with large weights if a similarity measure is used, while nodes that are joined by weak links are likely to belong to different regions.

A popular criterion for such partitions is based on extremal cuts through the graph. In computer vision, the idea of segmenting images by way of optimally partitioning a graph into k subgraphs so that the maximum inter-subgraph cut is minimized was introduced by Wu and Leahy [Wu 93]. The algorithm works recursively by splitting a segment in two regions A, B by a minimum cut:

$$cut\,(A, B) = \sum_{i \in A, j \in B} w_{i,j} \tag{2.1}$$

[2] *Links* are usually noted as *edges* though we decide to use *links* notation here to distinguish from the image edges.

$$MinCut\,(A, B) = \min\{cut\,(A, B)\} \tag{2.2}$$

until the whole graph is partitioned into k parts. Intuitively, the minimum cut corresponds to finding the subset of links of minimum weight that can be removed to partition the image.

Although performing well in many situations Wu and Leahy pointed out a few problems that result from the underlying principle behind min-cut. For example, since the algorithm returns the smallest cut separating the clusters, the algorithm will often return the cut that minimally separates the clusters even though they are strongly linked to the rest of the graph. The problem is that it is often cheaper to cut a few strong links than many weak ones. Finally, multiple "minima cuts" may exist in the image that are quite different from each other. Therefore, a small amount of noise (occurring even in a single pixel) could cause the segmentation to change drastically [Grady 06].

Veskler [Veksler 00] introduced a new graph node t and connect the pixels that delimit the image to t with links of appropriately chosen small weight. Given a pixel p in the image, the minimum cost contour separating p from the image can be found using the minimum cut that separates p from t. Results shown in the paper indicate that the algorithm is indeed capable of finding interesting image regions without many of the associated artefacts that occur in typical min-cut segmentation. It is important to keep in mind that the images upon which the above algorithms work are usually limited in size. This limitation is common to graph-theoretic algorithms and it is a consequence of the amount of memory required to store the graphs associated with large images and of the computational cost of partitioning such graphs.

Boykov et al. [Boykov 01b] presented an algorithm that relies on min-cut to perform energy minimization efficiently. They address the problem of assigning labels to a set of pixels so that the labelling is piecewise smooth and consistent with observed data. They define a suitable energy functional and show that given an initial labelling min-cut can be used to approximately minimize this functional with regard to two classes of operations that work respectively on single labels and label pairs.

In [Wang 01], Wang and Siskind proposed a modification to the minimum cut criterion to reduce the preference of minimum cut for small boundaries. They propose

the use of minimum mean cut, defined as

$$MeanCut\,(A,B) = \frac{cut\,(A,B)}{L} \tag{2.3}$$

where L is the length of the boundary dividing A and B. Like other min-cut based algorithms, the minimum mean cut is used recursively to produce finer segmentations. It is interesting to point out that this algorithm uses an additional step of region merging, since the minimum mean cut may lead to some spurious cuts where no image edge exists. [Wang 03b] generalized the minimum mean cut by using two edge weights to connect pairs of nodes, the first weight comes from the similarity measure and the second weight corresponds to a normalization term based on the segmentation boundary length.

Dupuis and Vasseur [Dupuis 06] developed an approach for the computation of the affinity matrix based on the combination of affinity matrices from various cues and its integration in the segmentation process. A principal components analysis (PCA) applied to the whole set of the normalized affinity matrices provided the uncorrelated relevant cues and their respective weights for the final combination. They propose to integrate the evaluation of the affinity matrix at each iteration of an agglomerative algorithm in order to take into account the dynamics of the segmentation process. Finally, they define a criterion of satisfaction based on the covariance matrix of the affinity matrices, which determines the end of the iterations.

Introduced by Felzenszwalb and Huttenlocher [Felzenszwalb 04], the so-called efficient graph-based image segmentation algorithm is another method using clustering in feature space. This method works directly on the data points in feature space, without first performing a filtering step, and uses a variation on single linkage clustering. The key to the success of this method is adaptive thresholding. To perform traditional single linkage clustering, a minimum spanning tree of the data points is first generated, from which any edges with greater length than a given threshold are removed. In the end of the process the components that remain connected become the clusters in the segmentation.

The graph cuts segmentation algorithm has been extended in two different directions in order to address issues of speed. The first type of extension to the graph cuts algorithm has focused on speed increases by coarsening the graph before ap-

plying the graph cuts algorithm. This coarsening has been accomplished in many ways: 1) by applying a standard multilevel approach and solving subsequent, smaller graph cuts problems in a fixed band to produce the final, full-resolution segmentation [Sharon 00, Yu 04] and 2) by applying some over-segmentation algorithm to the image and treating each atomic region as a "super-node" in a coarse graph to which graph cuts are applied [Callaghan 05].

Spectral analysis uses the data representation provided by the dominant eigenvalues and eigenvectors of a similarity matrix. There are many different algorithms that use the spectral properties of the affinity matrix, they differ in the number of eigenvectors/eigenvalues used, as well as in the clustering procedure, but all use the data representation provided by the dominant eigenvalues and eigenvectors of the affinity matrix. We refer the reader to [Weiss 99, Ng 02] for a review.

Perona and Freeman [Perona 98] suggested a clustering algorithm (known as the 'factorization method') based on treating as an indicator function the first *largest* eigenvector v_1 of the similarity matrix W. A threshold T is chosen, and each node i is assigned to one part if $v_{1_i} < T$ and to the other part otherwise. Perona and Freeman motivated the approach by showing that for block diagonal affinity matrices, the first eigenvector has non-zero in components corresponding to points in the dominant cluster and zeros in components corresponding to points outside the dominant cluster.

In [Weiss 99], Weiss discussed the relationships between four different spectral algorithms [Perona 98, Shi 00, Scott 90, Costeira 95], and proposed an interesting combination of the Shi and Malik algorithm [Shi 00] with Scott and Longuet-Higgins algorithm [Scott 90]. In Ng et al. [Ng 02], the normalized row vectors of the matrix formed by the first k weighted eigenvectors are used as the input to a K-means clusterer, and a perturbational analysis was used to show that the results should be stable if the data was already "nearly clustered".

Shi and Malik [Shi 00] used a quite different eigenvector for solving clustering problems. Rather than examining the first eigenvector of W they look at generalized eigenvectors. Let D be the degree matrix of W: $D_{ii} = \sum_j w_{i,j}$. Define the generalized eigenvector $\mathbf{y}$ as a solution to:

$$(D - W)\,\mathbf{y} = \lambda D\mathbf{y} \tag{2.4}$$

and define the second generalized eigenvector, denoted by $\mathbf{y}_2$, as the $\mathbf{y}$ corresponding to the second *smallest* eigenvalue λ. Shi and Malik suggested thresholding this second generalized eigenvector of W in order to partition the nodes into two parts. They motivated the approach by showing that the second generalized eigenvector is an approximate solution to a continuous version of a discrete problem in which the goal is to minimize:

$$\frac{\mathbf{y}^T (D - W) \mathbf{y}}{\mathbf{y}^T D \mathbf{y}} \tag{2.5}$$

subject to the constraint that $\mathbf{y}_i \in \{1, -b\}$ and $\mathbf{y}^T D \mathbf{1} = \mathbf{0}$, where $\mathbf{1}$ is a vector of appropriate length consisting of unit entries and b is a positive real constant.

The significance of the discrete problem is that its solution can be shown to provide the partition that minimizes the normalized cut (NCut) criterion for two regions.

$$NCut(A, B) = \frac{cut(A, B)}{links(A, V)} + \frac{cut(A, B)}{links(B, V)} \tag{2.6}$$

where $links(A, V) = \sum_{i \in A, j \in V} w(i, j)$ is the total connection from nodes in A to all nodes in the graph V and $links(B, V)$ is similarly defined.

The great advantage of the normalized cut over previous minimum cut methods is the normalization, which rescales the cut weight to remove trivial solutions associated with the removal of one or very few nodes. As Shi and Malik noted, there is no guarantee that the real solution will bear any relationship with the discrete one. Computing the normalized cut exactly for a given graph is an NP-complete problem, however, Shi and Malik showed that an approximate solution can be obtained from the eigenvector with the second largest eigenvalue.

In spectral clustering, there is research showing that using more eigenvectors and directly computing k-way partitioning is better [Yu 03]. Yu and Shi [Yu 03] studied multi-way partitions in the context of normalized cuts and spectral clustering. Meila and Shi [Meila 01] showed a connection between the eigenvectors and eigenvalues used in normalized cuts and those of a Markov matrix obtained by normalizing the affinity matrix W.

The original NCut formulation relies on the fact that the affinity matrix can be made sparse, which allows the algorithm to handle larger images than would be possible otherwise and it also allows for the use of optimized eigensolvers that work on such sparse matrices. However, this is not sufficient for large images. Belongie et

al. [Belongie 02], and Fowlkes et al. [Fowlkes 04] introduced a modification to the NCut framework that makes it possible to segment large images, or image sequences [Fowlkes 01]. The modification is based on the Nyström method for approximating the leading eigenvalues of a matrix using only a small number of randomly sampled image pixels. These random samples are used to build a smaller (non-square) affinity matrix whose leading eigenvectors can be computed at a much lower computational expense than those of the affinity matrix for the full image. These eigenvectors are then used to interpolate the complete solution to the NCut problem.

Sharon et al. [Sharon 00, Sharon 01] proposed a different approach for making the NCuts practical on large images. Their method solves a coarser NCut problem which includes region based statistics in the affinity measure, and then interpolates the solution to finer levels of detail, providing a hierarchy of segmentations for a given image.

2.4 Cooperative methods

Elementary segmentation techniques based on boundaries or regions often fail to produce accurate segmentation results on their own. To overcome this difficulty there has been a trend towards algorithms that take advantage of the complementary nature of both techniques. More elaborated image segmentation approaches based on the combination, integration or iteration between methods, especially those of edge detection and uniform region extraction have been proposed.

The cooperative schemes are useful when some sort of complementary properties are explored among the individual methods. For instance, it is common to combine edge-based with region-based approaches. As the first method presents good localization characteristics but it is sensitive to noise usually resulting in several edge gaps, the region-based methods have poor accuracy on boundaries, although producing natural closed contours and they are more insensitive to noise. By using the complementary nature of edge-based and region-based information, it is possible to reduce the problems that arise in each individual method. The trend towards integrating several techniques seems to be the best way to follow [Muñoz 03]. By having a cooperative method it is expected that it will cover a wider range of images on which the algorithm is able to work for segmentation.

Combining the outputs of image segmentation and edge detection to improve the quality of the segmented image is an old idea. Muñoz et al. [Muñoz 03] in their recent review on combining methods classified these proposals by the timing of the integration between methods as: embedded integration, when the edge information is extracted first, and this information is then used within the segmentation algorithm, which is mainly based on regions; post-processing integration, where edge and region information are extracted independently as a preliminary step, and an a posteriori fusion process tries to exploit the dual information in order to modify, or refine, the initial segmentation obtained by a single technique.

We append two new classes to this classification: the hybrid framework and the interactive framework. Thus we distinguished the cooperative methods into four different types: the sequential [Beveridge 89, Gambotto 93, Fan 01], the parallel [Chu 93, Zhu 96, Germond 00], the hybrid [Haris 98, Kermad 02, Lezoray 03, Makrogiannis 05, Callaghan 05, Duarte 06], and interactive frameworks [Mortensen 99, Olabarriaga 01, Blake 04, Rother 04, Farmer 05]. Sequential and parallel types correspond respectively to embedded and post-processing classes of Muñoz et al. classification. Hybrid framework combines methods that are themselves cooperative approaches. The interactive framework class includes the methods which, due to a high demand for accurate results, usually adopt human intervention.

2.4.1 Sequential framework

The sequential framework usually consists of using previously extracted edge information within a region segmentation algorithm. Although the method obtained in a sequential framework is more robust than its individual components, the cooperation between the modules is still rudimentary: each sub-task is performed sequentially and its result is used to feed the following task.

Figure 2.2 illustrates the sequential framework. The decision to merge in region growing algorithms is generally based only on the contrast between the current pixel and the region. The edge map integration provides an additional criterion in such decisions. The seeds are lunched in placements which are free of edges. Finding an edge pixel means that the growing process has reached the boundary of the region and therefore the pixel must not be aggregated.

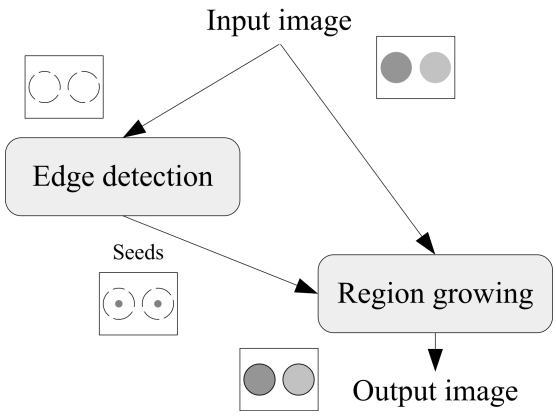


Figure 2.2: Scheme of sequential framework for image segmentation.

The work of Beveridge et al. [Beveridge 89] offered a good example of a procedure that integrates both histogram analysis and region merging. In their paper an input image is divided into sectors of fixed size and fixed location. An intensity histogram is calculated for each sector and used to produce a local segmentation. For every sector, information from its neighbours is used to detect clusters for which there may not be enough local support due to the artificially induced partition of the image. After the local segmentations are complete, the sector boundaries are removed by merging together similar regions in neighbouring sectors. The results show that this algorithm produces good segmentations in parts of the image that are reasonably homogeneous, and over-segmented regions when there is texture, significant intensity gradients, or objects with non-uniform brightness.

Gambotto [Gambotto 93] suggested using edge information to stop the region growing process. His proposal assumes that the gradient takes a high value over a large part of the region boundary. The iterative growing process is thus continued until the maximum of the average gradient computed over the region boundary is detected. Yu and Wang [Yu 99] used the edge information to determine the seeds for region growing but applied a new algorithm. A so-called difference in strength (DIS) map is first created. The pixel with the smallest DIS value among the unlabelled pixels is chosen as the seed of a region. The region grows until no more neighbouring pixels can be joined to it. Then, a new seed is chosen from the unlabelled pixels. The process continues until all pixels in the image are labelled. The major problems of cooperative techniques that are based on region growing are accuracy of the segmentation and efficiency in terms of speed of region growing around the pixels.

Fan et al. [Fan 01] developed an automatic colour image segmentation technique by integrating colour-edge extraction and seeded region growing on the YUV colour space. The colour-edges are first obtained by an isotropic colour-edge detector and the centroids between the adjacent edge regions are taken as the initial seeds for region growing. Moreover, the results of colour-edge extraction and SRG are integrated to provide more accurate segmentation of images. The disadvantage is that their seeds are over-generated.

Sclaroff and Liu [Sclaroff 01] proposed a method for deformable shape detection and recognition based on over-segmentation and region merging guided by statistical shape model and MDL principle. Luo and Khoshgoftaar [Luo 04] proposed an image segmentation algorithm by combining mean shift clustering and minimum description length (MDL) principle to complement their strengths and weaknesses. Their approach is to apply mean shift clustering to generate an initial over-segmentation and then merge regions based on MDL principle.

Pantofaru and Hebert [Pantofaru 05] presented a framework which consists of comparing the performance of mean shift [Comaniciu 02] and efficient graph-based clustering [Felzenszwalb 04], based on three important characteristics: correctness, stability with respect to parameter choice, and stability with respect to image choice. They propose a hybrid algorithm which first performs the first stage of mean shift-based segmentation, mean shift filtering, and then applies the graph-based segmentation scheme, as an attempt to create an algorithm which preserves the correctness of the mean shift-based segmentation but it is more robust with respect to parameter and image choice. They demonstrated that, although both the mean shift segmentation and hybrid segmentation algorithms can create realistic segmentations with a wide variety of parameters, the hybrid algorithm has slightly improved stability.

2.4.2 Parallel framework

After the extraction of edge and region information obtained independently the parallel framework carry out a posteriori integration. Parallel framework is based on the fusion of the results from single segmentation methods, attempting to combine the map of regions and the map of edge outputs with the aim of providing an accurate and meaningful segmentation.

Figure 2.3 shows a diagram of this parallel approach. This framework considers region-based segmentation as an approximation to segmentation which is then combined with salient edge information to achieve a more accurate representation of the boundary of the objects. Thus, edge information enables an initial result to be refined.

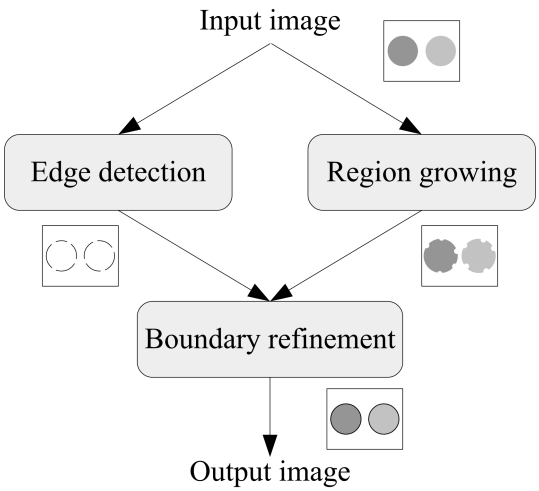


Figure 2.3: Scheme of parallel framework for image segmentation.

Chu and Aggarwal [Chu 93] presented an optimization method that integrates multiple region segmentation maps and edge maps in parallel cooperation, where an arbitrary mixing of region and edge maps are allowed.

Zhu and Yuille [Zhu 96] proposed a region competition approach to unify the active contour model, region growing, and Bayes for image segmentation. This approach is derived by minimizing a generalized Bayes criterion using the variational principle and combines aspects of active contour model and region growing. Their approach alternates boundary estimation and region estimation steps. It requires the selection of a number of seed regions for initialisation of the statistical measurements on which the region estimation is based. It would be advantageous both to minimise dependence on such initial conditions and for the region and boundary processing to be autonomous, so that where necessary one could be used independently from the other.

Germond et al. [Germond 00] proposed to mix in a cooperative framework several types of information and knowledge provided and used by complementary individual systems like a multi-agent system, a deformable model or an edge detector, where a cooperative segmentation performed by a set of region and edge agents constrained

automatically and dynamically by both, the specific grey levels in the considered image, statistical models of the brain structures and general knowledge about MRI brain scans. Interactions between the individual systems follow three modes of cooperation: integrative, augmentative and confrontational cooperation, combined during the three steps of the segmentation process namely, the specialization of the seeded region growing agents, the fusion of heterogeneous information and the retroactivity over slices.

Kermad and Chehdi [Kermad 02] presented a system that integrates the information resulting from two complementary segmentation methods: edge detection and region extraction. This permits the correction and adjustment of the control parameters of the methods used. The suggested cooperation approach introduces a mechanism which checks the coherence of the results through a comparison of the two segmentations. From over-segmentation results both methods are iterated by loosening certain constraints until they converge towards stable and coherent results. This coherence is achieved by minimising a dissimilarity measure between the edges and the boundaries of the regions.

Christoudias et al. [Christoudias 02] presented an algorithm where a region adjacency graph is created to hierarchically cluster the modes of the mean shift approach. Also, edge information from an edge detector is combined with the colour information to better guide the clustering.

Zhou et al. [Zhou 05] presented a method that combines the classical gradient vector flow (GVF) algorithm [Xu 98] with mean shift. Due to the dependence on the gradient vectors of an edge map, the classical GVF is sensitive to the shape irregularities, and hence the snake cannot be ideally located on the concave boundaries. They propose an improved representation of the internal energy force by reducing the Euclidean distance between the guessed centroid and the estimated one of the snake. The mean shift technique is used to constrain the spatial diffusion of the gradient so that it is able to handle efficiently boundary concavities.

2.4.3 Hybrid framework

Figure 2.4 gives a possible structure of a hybrid framework. This example begins by obtaining an edge map which is used in the watershed algorithm to obtain an over-segmented result. This result is then compared with the result from the dual

approach: each boundary is checked to find out if it is consistent in both results (edges and regions). When this correspondence does not exist the boundary is removed. This is achieved by using a region similarity graph where the similarity is proportional with the intervening contours between the regions. This graph is segmented by some graph cut method.

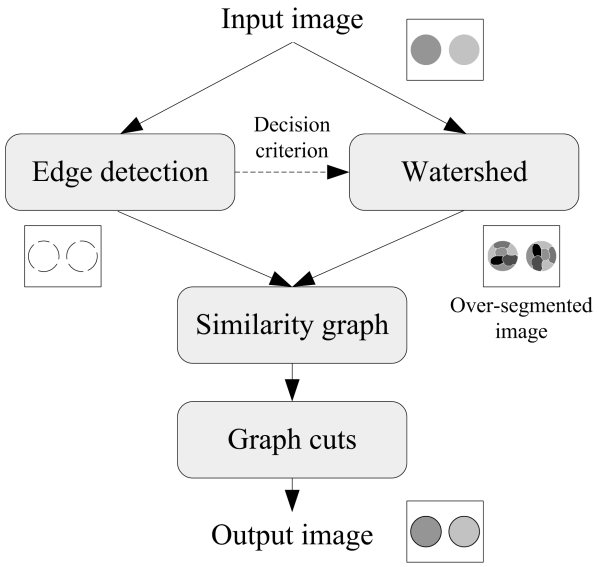


Figure 2.4: Scheme of hybrid framework for image segmentation.

Haris et al. [Haris 98] proposed a hybrid image segmentation using watersheds and fast region merging algorithm which combines edge and region-based techniques through the morphological algorithm of watersheds. This is done by applying edge preserving statistical noise filter to compute an estimate of the image gradient. The image is then partitioned into primitive regions using the watershed transform on the image gradient magnitude. The result of this is then used as an input to a bottom up region merging process. The objective cost function, the so-called region dissimilarity function, is a function of the square error of the piecewise constant approximation of the observed image, and is defined over the space of the partitions. For region merging the authors adopt a solution for fast region merging, the fast neighbour region merging by creating a simplification of the region adjacency graph (RAG). This algorithm was designed and implemented for 2D and 3D images and it produces very satisfactory results in segmentation performance and execution.

Lezoray and Cardot [Lezoray 03] combined different types of methods to obtain a segmentation of a colour image. They divided the segmentation process into three

stages: colour clustering, region merging and watershed segmentation. In the first stage 2D histograms are used to obtain a rapid and coarse clustering of the colour image. This clustering is fast, simple and unsupervised, although over-segmented. The second stage proceeds to a region merging of adjacent regions until the stabilization of a cost associated with the partitioning of the colour image. In the third stage, a segmentation refinement is based on a morphological filtering and colour watershed.

Makrogiannis et al. [Makrogiannis 05] proposed a hybrid algorithm that combines the concepts of multi-resolution fuzzy clustering and region-based graph segmentation to produce the final regions. Watershed approach is applied to produce the initial over-segmented image and a second stage, known as the merging stage, is used to form the final regions. This stage consists of the dissimilarity calculation process and the merging algorithm. The dissimilarity calculation is carried out using a multiscale generation process in the feature space. A clustering approach based on non-parametric density estimation, known as subtractive clustering, is used to determine the population and location of the most prominent cluster centres at different resolutions. The fuzzy C-means algorithm is subsequently employed to produce the membership vectors. The dissimilarity at each resolution is inferred using standard fuzzy arithmetic operations. The multiscale dissimilarity function takes into account the structure of clusters over multiple scales to evaluate the degree of dissimilarity. The result of this operation is the integration of the global cluster analysis results into the general region-based scheme.

Pan et al. [Pan 03] proposed a combination of mean shift with watershed algorithm. First, mean shift procedure is used to find the highest density regions which correspond to clusters centred on the modes (local maxima) of the underlying probability distribution in the feature space. The principal component of each significant colour is extracted by mode. Secondly, homogeneous regions corresponding to the modes are used as markers to label an image, then marker-controlled watershed transform is applied to the image segmentation. The algorithm was applied to blood cells segmentation.

O'Callaghan and Bull [Callaghan 05] proposed the combination of an initial segmentation using watershed transform with spectral methods. The morphological watershed transform is applied to a gradient image which is a result of combination of a texture gradient and modulated intensity gradient, trying to embed in a single image all perceptual gradients. For texture representation the authors use sub-band median

filters applied to the texture sub-band magnitude (the magnitude of the complex detail coefficients computed from a wavelet complex transform). This method follows an approach proposed by Hill et al. [Hill 03] which also integrates edge and texture information. This sequence of operations results in a set of homogeneous texture regions, although over segmented images. To further reduce the number of segments, the primitive regions are represented in a graph and processed using spectral clustering, using a weighted mean cut algorithm. The authors argued that weighting the cuts by the length of the boundary makes the partition independent of the number and geometric arrangements of the segments while taking into consideration the importance of the boundary lengths. For building the similarity matrix, the authors followed a non-parametric approach of Puzicha et al. [Puzicha 99] by measuring the similarity between feature distributions. In this way rather than clustering single feature points the spectral method cluster feature distributions. This morphological-spectral combination strategy leverage the over segmentation weakness of the watershed by providing to the spectral approach small texture patches that can be characterized statistically.

In [Duarte 06], Duarte et al. proposed an approach that starts from an over-segmented image which is obtained by applying a standard morphological watershed transformation on the original image. Then, this over-segmented image is described by a simplified undirected weighted graph, where each node represents one region and weighted links measure the dissimilarity between pairs of regions according to their intensities, spatial locations and original sizes. Finally, the resulting graph is iteratively partitioned in a hierarchical fashion into two subgraphs, corresponding to the two most significant components of the actual image, until a termination condition is met. They use a histogram thresholding method to automatically determine the merging termination. This graph-partitioning task is solved by a normalized cut approach using a hierarchical social meta heuristic.

Li and Zeng [Li 06] presented a strategy based on wavelet, morphology and combination method. Firstly, the wavelet transforms and morphology are used to get rid of the effect of the defocusing and get the sub-images that include the particles. Then based on the characteristics of the sub-images, edge detection and adaptive thresholding are employed adaptively. Finally, a simplified watershed algorithm for the overlapping particles is used.

2.4.4 Interactive framework

The intervention of a human operator is often needed to initialise the method, to check the accuracy of the result produced automatically, or even to correct the segmentation result manually. Interaction is usually adopted in applications with a high demand for accurate results and where the volume of images is reasonable, allowing for human manipulation. A major disadvantage of these methods is that they are only suitable for foreground-background segmentation.

All the above-mentioned algorithms are automatic. A major advantage of this type of segmentation algorithms is that they can extract boundaries from a large number of images without occupying human time and effort. However, in an unconstrained domain, for non-preconditioned images, the automatic segmentation is not always reliable. On the other hand, a simple user assistance in object recognition is often sufficient to complement deficiencies and to complete the segmentation process. There are many difficult segmentation tasks that require a detailed user assistance. This is often true in medical applications, where image segmentation is particularly difficult due to restrictions imposed by image acquisition, pathology and biological variation. To address these problems a variety of interactive segmentation methods were developed [Olabarriaga 01, Rother 04].

Figure 2.5 gives an example of an interactive framework. In this example the user draws a fat pen trail enclosing the object boundary, marked in blue. This defines the trimap with foreground/background/unclassified labels. The automatic segmentation algorithm produces a first segmentation result. Missing parts of the object can be added efficiently by user refinement: the user roughly applies a foreground brush, marked in red, and the automatic segmentation method adds the whole region.

Recently, researchers have managed to improve image cut-out by using region-based methods, e.g., intelligent paint [Barrett 02], sketch-based interaction [Tan 01], interactive graph cut image segmentation [Boykov 01a] and GrabCut [Rother 04]. Region-based methods work by allowing the user to give loose hints as to which parts of the image are foreground or background without enclosing regions or being pixel accurate. These hints usually take the form of clicking or dragging on foreground or background elements and are thus quick and easy to do. An underlying optimization algorithm extracts the actual object boundary based on the user input hints.

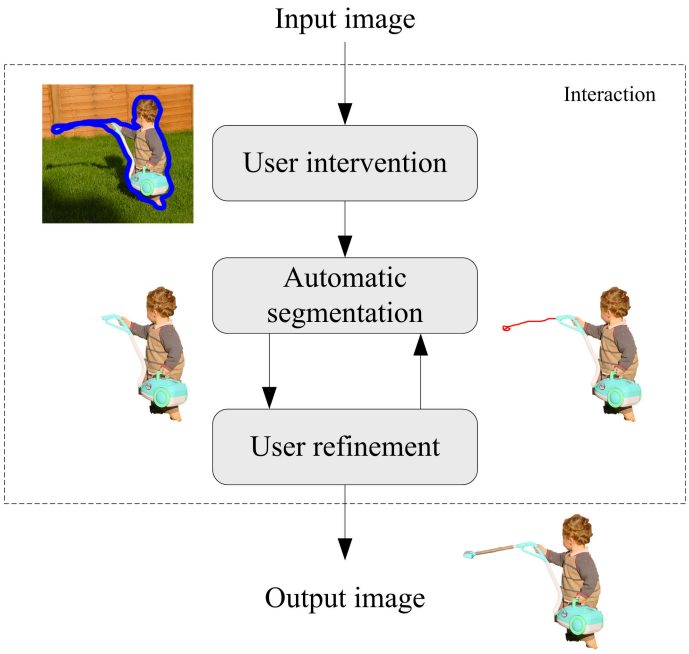


Figure 2.5: Scheme of interactive framework for image segmentation.

The interactive image segmentation algorithms [Boykov 01a, Blake 04, Rother 04] aim to separate, with minimal user interaction, a foreground object from its background so that, for practical purposes, it is available for pasting into a new context. Some studies [Ruzon 00, Wang 05] focus on inference of transparency in order to deal with mixed pixels and transparent textures such as hair. Other studies [Boykov 01a, Blake 04] concentrate on capturing the tendency for images of solid objects to be coherent, via Markov Random Field prior. For a review on interactive approaches for image segmentation see e.g. [Olabarriaga 01, Rother 04].

Rui et al. [Rui 96] proposed a segmentation algorithm based on clustering and grouping in spatial–colour–texture space. The user defines where the object of interest is and the algorithm groups regions into meaningful objects. Wang and Cohen [Wang 05] combined the segmentation and matting[3] problem together and proposed a unified optimization approach based on belief propagation [Yedidia 02]. They iteratively estimate the opacity value for every pixel in the image, based on a small sample of foreground and background pixels marked by the user.

Boykov and Jolly [Boykov 01a] proposed a method for interactive segmentation based on graph cuts. The user input is minimal, consisting of a few mouse-clicks indi-

[3]Matting approaches try to simplify the problem by photographing foreground objects against a constant coloured background, which is called *blue screen matting*.

cating some pixels which are inside the object of interest, and other are outside. An energy function based on both boundary and region information is then minimized subject to these user-imposed constraints. The global minimum is found by using graph cut techniques. With a relatively small amount of user input, the algorithm successfully segments a variety of objects from both medical and natural images. GrabCut [Rother 04] extends the graph-cut by introducing iterative segmentation scheme, that uses graph-cut for intermediate steps. The user draws a rectangle around the object of interest - this gives the first approximation of the final object/background labelling. Then, each iteration step gathers colour statistics according to the current segmentation, re-weights the image graph and applies graph-cut to compute new refined segmentation. After the iterations stop the segmentation results can be refined by specifying additional seeds, similar to the original graph-cut.

Intelligent Paint proposed by Barrett and Cheney [Barrett 02] is a region-based interactive segmentation technique based on hierarchical image segmentation by toboggan watershed [Liu 03]. The strategy it uses coordinates human-computer interaction to extract regions of interest from backgrounds using paint strokes with a mouse.

Protiere and Sapiro [Protiere 07] proposed an interactive algorithm for soft segmentation of natural images. The user first roughly scribbles (user-provided labels) different regions of interest and from them the whole image is automatically segmented. This soft segmentation is obtained via fast, linear complexity computation of weighted distances to the user-provided scribbles. The adaptive weights are obtained from a series of Gabor filters and are automatically computed according to the ability of each single filter to discriminate between the selected regions of interest.

Boundary-based methods cut out the foreground by allowing the user to surround its boundary with an evolving curve. The user traces along the object boundary and the system optimizes the curve in a piecewise manner. Examples include intelligent scissor [Barrett 98] and LiveWire [Falcão 00]. Besides being easier to manipulate rather than just selecting pixels manually with a traditional selection tool, these techniques still demand a large amount of attention from the user. There is never a perfect match between the features used by the algorithms and the foreground image. As a result, the user must control the curve carefully. If a mistake is made, the user has to "back up" the curve and try again. The user is also required to enclose the entire boundary, which can take some time for a complex high-resolution object [Li 04].

The intelligent scissors segmentation tool described in [Barrett 98] allows objects within images to be extracted quickly and accurately using simple gesture motions with a mouse. When the gestured mouse position comes in proximity to an object edge, a live-wire boundary "snaps" to, and wraps around the object of interest [Barrett 98]. It formulates boundary finding as an unconstrained graph search in which the boundary is represented as an optimal path within the graph. The live-wire tool allows the user to interactively select an optimal boundary segment by immediately displaying the minimum cost path from the current cursor position to a previously specified "seed" point in the image.

Mortensen and Barret [Mortensen 99] proposed a region-based intelligent scissors approach which uses toboggan watershed for image over-segmentation and then treats homogeneous regions as graph nodes. After applying the toboggan segmentation, each connected region is assigned with a different label. Next, a weighted graph is constructed by tracing the boundary of each region successively. Once the weighted graph is constructed, the remaining algorithm is the same as the pixel-based approach. However, when compared with the pixel-based approach, the number of graph nodes created by the region-based approach is lessened and hence the computational cost is greatly reduced.

Suetake et al. [Suetake 07] argued that the intelligent scissors is too sensitive to a noise and texture patterns in an image since it utilizes the gradient information concerning the pixel intensities. They propose a new intelligent scissors based on the concept of the separability in order to improve the object boundary extraction performance. Rother et al. [Rother 04] evaluated the performance of some of the described methods and have clearly shown that methods based on graph cuts allow achieving better segmentation results with less user effort required when compared with the other methods.

A generic approach for feature selection that is related with the interactive framework uses the classification method as a subroutine, rather than as a postprocessor. Farmer and Jain [Farmer 05] proposed a closed-loop framework called wrapper-based segmentation that not only adapts the parameters of the segmentation algorithm, but also actually direct the segmentation based on the underlying shape of the object of interest. Figure 2.6 shows the closed-loop wrapper-based segmentation framework presented in [Farmer 05].

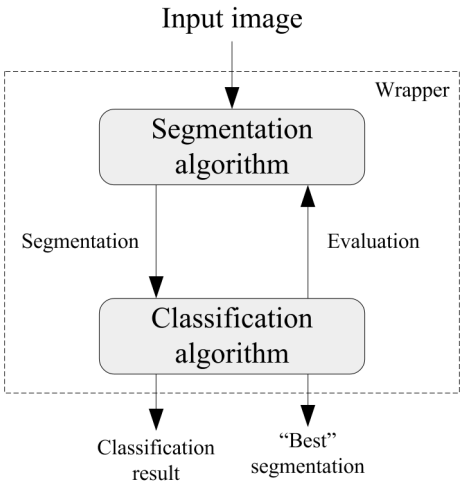


Figure 2.6: Wrapper-based image segmentation.

They initially perform low-level segmentation to label the image as a set of non-overlapping blobs. Then they use the wrapper framework to select the blobs that comprise the final segmentation based on the classification performance of the wrapper. The selection process involves grouping the set of homogeneous regions in the image that together comprise the object of interest. The blob combination with the highest probability of correct classification, based on their classification against a set of training images, for a given class is considered the most likely combination.

2.5 Summary

In this chapter we have reviewed a lot of image segmentation proposals. Special emphasis has been placed on the strategy used to carry out the cooperative process which integrate edge and region information and identified the various strategies and methods used to fuse such information. A classification of cooperative segmentation techniques has been proposed and we have described several algorithms, pointing out their specific features.

Based on all the techniques discussed in this chapter, it is clear that image segmentation procedure is a complex issue. Another conclusion is that image segmentation is application dependant and some parameters have to be refined accordingly to the type of image. The large amount of methods is an indication that the "final solution" is still far to come.

Actually, it is not feasible to determine the best approach to segmentation. There are several reasons for this, being the two most important factors (1) the lack of a generally accepted and clear methodology for evaluating segmentation algorithms [Zhang 96], and (2) the difficulty in implementing other people's algorithms due to the lack of necessary details [McCane 97]. Obviously, unless a given segmentation algorithm is specifically implemented and tried out on the same set of images, it is very difficult to evaluate from the published results how well it will work for those images. Thus, we would like to emphasize the need for the image segmentation community to create a central repository of algorithm implementations, data and evaluation measures so that researchers can quickly and effectively compare their algorithms with well established methods. We will address this evaluation issue on the next chapter.

CHAPTER 3

Image segmentation evaluation

This chapter proposes a new approach for evaluation of segmentation based on regions that takes into account not only the accuracy of the boundary localization of the created segments but also the under- and over-segmentation effects, regardless to the number of regions in each partition. In addition it takes into account the way humans perceive visual information. This new metric can be applied both to provide a ranking among different segmentation algorithms automatically and to find an optimal set of input parameters of a given algorithm.

3.1 Introduction

[1]The practical application of an image segmentation algorithm requires that we understand how its performance varies in different operating conditions. Evaluating algorithms let researchers know the strengths and weaknesses of a particular approach and identifies aspects of a problem where further research is needed. Haralick [Haralick 94] underlines the necessity of the evaluation of computer vision algorithms if the field is to produce methods of practical use to engineers.

In spite of significant advances in image segmentation techniques, evaluation of these methods thus far has been largely subjective. Typically the effectiveness of a new algorithm is demonstrated only by the presentation of a few segmented images that are evaluated by some method, or it is otherwise left to subjective evaluation by the reader.

[1]The work included in this chapter was presented at the International Conference on Image Analysis and Recognition (ICIAR2006) [Monteiro 06].

The readers frequently do not know whether the results have been opportunistically selected or they are typical examples, and how well the demonstrated performance extrapolates to larger sets of images.

Evaluating the output of segmentation algorithms is still problematic. The work of Martin et al. [Martin 01] presents a significant advance in this direction by providing segmentation results that can be used as a baseline for comparing the output of different methods, as well as suitable error metrics to quantify the performance of the algorithms in terms of the quality of their segmentations. However, at this time to our knowledge only the normalized cuts algorithm has been evaluated in this way, and the results of this evaluation cannot be interpreted in a meaningful way in the absence of comparative results for other segmentation methods. In fact there are very few comparative studies on the methods used for evaluation [Zhang 96].

The selection of an appropriated method for the segmentation of a particular image is a difficult issue, as there is no universally accepted figure(s) of merit to evaluate the performance of an image segmentation result. We still need to rely in the experience, knowledge and intuition of the person in charge of conceiving the image segmentation algorithm in the selection phase, together with the semantic information about the type of images to be segmented and the qualitative assessment of the final user.

Typically researchers show their segmentation results on a few images and point out why the results 'look good'. We never know from such studies if the results are good or typical examples. Since none of the proposed segmentation algorithms are generally applicable to all images, and different algorithms are not equally suitable for a particular application, there is the need to make comparisons so that the better ones can be selected. The majority of studies proposing and comparing segmentation methods evaluate the results only with one evaluation method. However, results vary significantly among different evaluators, because each evaluator may have distinct standards for measuring the quality of the segmentation.

The main difficulty in evaluating segmentation algorithms stems from the ill-defined nature of the problem being addressed. Zhang, in his survey [Zhang 96], proposes this definition of image segmentation: *'[Image segmentation] consists of subdividing an image into its constituent parts and extracting these parts of interest (objects).'*

Without explicit knowledge of what one would like the output of the algorithm to be, it is hard to say whether one algorithm is better than another. Many researchers

prefer to rely on quality human judgement of results for evaluation. Borra and Sarkar [Borra 97] argued that segmentation performance can be evaluated only in the context of a task such as object recognition. Pal and Pal [Pal 93] say that 'a human being is the best judge to evaluate the output of any segmentation algorithm'. McCane [McCane 97] proposes an evaluation method based entirely on the application for which the algorithm was designed. If a segmentation method leads to a better performance on a task, then that segmentation method is better for that task, regardless of what a human expert thinks about the quality of the segmentation.

In some sense boundary detection and region segmentation are two dual problems and their performance evaluation appears to be a similar task. One may convert a segmented region map to an equivalent boundary map by marking the region boundaries only and then applying any boundary detection evaluation method. However, a simple example as shown in Figure 3.1, reveals a fundamental difference: although in terms of the boundaries the two segmentation results only differ marginally, their discrepancy in terms of regions is substantially larger. In the present work although we made a review on boundary based evaluation, our first concern is with region segmentation evaluation.

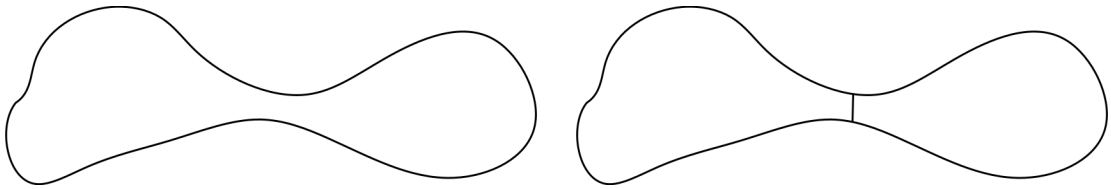


Figure 3.1: Two segmentation results.

Some researchers argue that segmentation algorithms should be evaluated in the context of a particular task such as object recognition [Borra 97], that is different algorithms should be compared in terms of the potential benefit they provide for a particular higher-level task. Other researchers (see for example [Martin 01]) propose that segmentation algorithms should be evaluated as stand-alone modules by comparing their output to 'ground truth' which is usually a segmentation produced by human observers.

This latter view is more suitable for our purposes so, for the remainder of the chapter, experimental results are considered in the light of what a human observer would see

in a given image. This leads us to two essential problems: 1) Different human observers will produce different segmentations of the same image, and 2) Human observers use high level knowledge, and solve high level vision problems such as recognition and perceptual completion while segmenting the image. Research by Martin et al. [Martin 01] indicates that human segmentations do not vary randomly, instead they show regularities that can be exploited to design and evaluate segmentation algorithms. They also suggest ways in which the use of higher level knowledge by human observers can be accounted for, thus allowing for the direct comparison of segmentations produced by human observers and segmentation algorithms.

A potential problem for a measure of consistency between segmentations is that there is no unique segmentation of an image. One approach is to ask human subjects to segment the images by hand. If a reasonable consensus emerges, the hand segmentations can be treated as ground truth, and compared to the outputs of segmentation schemes. Martin et al. [Martin 01] take this approach. They present a database containing hand segmented images from the Corel database [Martin 01]. They define an error measure which quantifies the consistency between segmentations of differing granularities and find that different human segmentations of the same image are highly consistent. According to Martin et al. [Martin 01], two subjects may segment an image differently for any of several reasons:

- **Perception**. If two subjects perceive the same scene in two different ways, then they may see different objects and produce different segmentations.
- **Attention**. Subjects may pay attention to different parts of the scene to different degrees, and may therefore over-segment the objects of focus, and under-segment the other objects.
- **Refinement**. Two subjects may segment an image identically in all regards, except that one subject may divide objects into smaller pieces than the other subject did.

The two last effects produce variations between segmentations but not inconsistencies, then the error should be smaller. This implies that we need to define segmentation consistency measures that do not penalize such differences. If two segmentations arise from different perceptual organizations of the scene then it is fair to declare the segmentations inconsistent. One desirable property of a good measure is to accommodate

refinement only in regions that human segmenters find ambiguous and to penalize differences in refinement elsewhere.

An alternative approach is to allow human subjects to evaluate directly the output of segmentation algorithm using psychovisual tests and judge which of segmentations is more meaningful to them. Shaffrey et al. [Shaffrey 02] proposed an evaluation procedure that subjects human observers to a psychovisual test comparing directly the output of different segmentation algorithms and judge which pair of segmentations is more meaningful to them. Heath et al. [Heath 97] evaluated the output of different edge detectors on a subjective quantitative scale using the criterion of ease of recognizability of objects (for human observers) in the edge images. Chalana and Kim [Chalana 97] use multiple expert observers to agree on ground truth in the context of medical imagery, while Hoover et al. [Hoover 96] do so in computer vision through carefully created ground truth to test range finding algorithms.

Only a few evaluation methods actually explore the segments obtained from the segmentation process. Some measures are best suited to evaluate edge detection [Sahoo 88], working directly on the binary image of the regions' boundaries [Huang 95]. Although we can always treat segmentation as a boundary map, the problem is in the simplified use of the edge map, as simply counting the misclassified pixels, on an edge/non-edge basis. Pixels on different sides of an edge are different in the sense that they belong to different regions - that is why it may be more reasonable to use the segmentation partition itself.

Evaluation of image segmentation differs considerably from the binary foreground background segmentation evaluation problem examined in [Goumeidane 03, Huang 95], in that the correctness of the two class boundary localization is not the only quantity to be measured. This derives from the presence of an arbitrary number of regions in both the reference segmentation and the segmentation to be evaluated.

3.2 Problem formulation

An evaluation metric is desired to take into account the following effects:

- **Over-segmentation**. A region of the reference is represented by two or more regions in the examined segmentation.

- **Under-segmentation**. Two or more regions of the reference are represented by a single region in the examined segmentation.
- **Inaccurate boundary localization**. Ground truth is usually produced by humans that segment at different granularities.
- **Different number of segments**. We need to be able to compare two segmentations when they have a different number of segments.

Under-segmentation is considered to be as a much more serious problem as it is easier to recover true segments through a merging process after over-segmentation rather than trying to split an heterogeneous region. One desirable property of a good evaluation measure is to accommodate refinement only in regions that human segmenters could find ambiguous and to penalize differences in refinements elsewhere. In addition to being tolerant to refinement, any evaluation measure should also be robust to noise along region boundaries and tolerant to different number of segments in each partition.

Segmentation evaluation can be judged according to the amount of mis-segmented pixels estimated by a direct comparison between reference and resulted segmentation mask. Pixels can be classified into four sets: well-classified pixels (true positives, T_p), incorrectly detected pixels (false positives, F_p), correctly undetected pixels (true negatives, T_n), and incorrectly undetected pixels (false negatives, F_n). True negative pixels are ignored in some evaluation measures, e.g. Precision-Recall curves.

Let S and R be two segmentations of the same image, where $S = \{s_1, s_2, ..., s_k\}$ is the segmentation mask to be evaluated, containing k regions, and $R = \{r_1, r_2, ..., r_q\}$ is the reference mask, containing q regions. The pixel classification sets can be expressed as:

$$T_p = S \cap R \qquad F_p = S \cap \overline{R} \qquad F_n = \overline{S} \cap R \qquad T_n = \overline{S \cup R} \tag{3.1}$$

where $\overline{R}$ and $\overline{S}$ denotes the complement of R and S respectively. We assume that an image is composed of objects that when aggregated form all the image. So, if a pixel is classified as true for one object it is classified as false for other object. Figure 3.2 shows the classification of pixels according to the comparison between the reference object and the segmented object.

These possible measures can be arranged in a *confusion matrix* [Stehman 97]. This matrix contains information about actual and segmented regions done by a segmen-

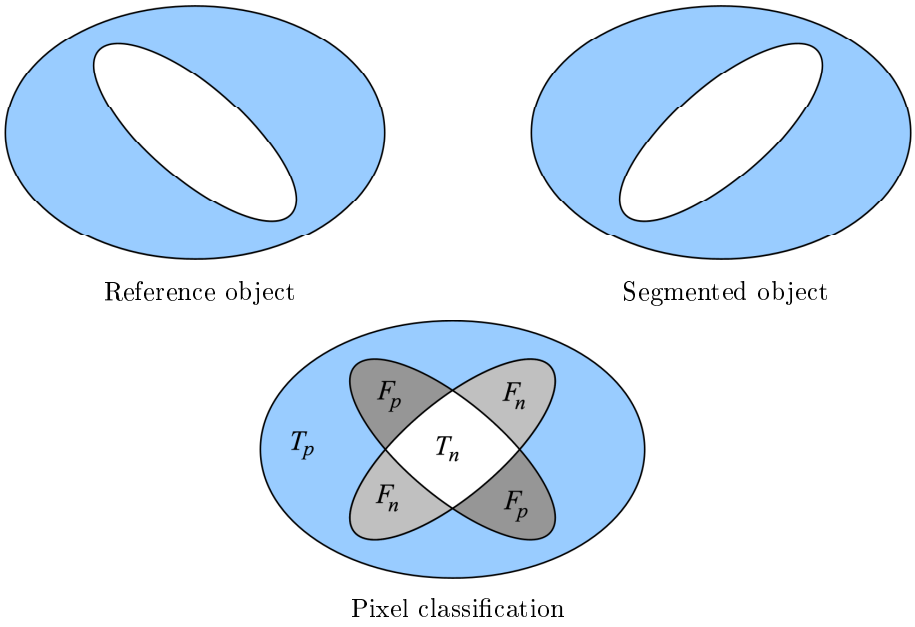


Figure 3.2: Pixel classification in the segmentation evaluation process.

tation system. The diagonal elements represent correctly classified pixels while the cross-diagonal elements represent misclassified pixels. Figure 3.3 shows the confusion matrix for a two region segmentation algorithm.

		Segmented	
		YES	NO
Reference	YES	T_p	F_n
	NO	F_p	T_n

Figure 3.3: Confusion matrix in a two region segmentation problem.

3.3 Related work

A review on evaluation of image segmentation is presented by Zhang in [Zhang 96], who classifies the methods into three categories: *analytical*, where performance is judged not on the output of the segmentation method but on the basis of their properties, principles, complexity, requirements and so forth, without reference to a concrete implementation of the algorithm or test data. While in domains such as edge detection this may be useful, in general the lack of a general theory of image segmentation limits

these methods; *empirical goodness methods*, which compute some kind of 'goodness' metric such as uniformity within regions [Borsotti 98, Huang 95], contrast between regions [Levine 85], or shape of segmented regions [Sahoo 88]. For edge detection, human intuition based measures have been introduced by Heath et al. [Heath 97] that propose an edge detection assessment based on the bootstrap resampling technique; and finally, *empirical discrepancy methods*, which evaluate segmentation algorithms by comparing the resulting segmented image against a manually-segmented reference image, which is often referred to as ground truth, and computes error measures.

As stated by Zhang [Zhang 96], the major difficulty in applying analytical methods is the lack of general theory for image segmentation. The analytical methods may only be useful for simple algorithms or straightforward segmentation problems, where the researchers have to be confident in the models on which these algorithms are based.

Empirical goodness methods, also known as unsupervised evaluation methods quantitatively evaluate the results of segmentation algorithms according to some human characterization about the properties of the *ideal* segmentation. These methods have the advantage that they do not require manually segmented images to be supplied as ground truth data. The great disadvantage is that these metrics are heuristic and may exhibit strong bias towards a particular algorithm. For example the intra-region and the inter-region grey-level uniformity metric will assume that a well-segmented image region should have low variance of grey-level. This will cause that any segmentation algorithm which forms regions of uniform texture to be evaluated poorly. Although these evaluation methods can be very useful in some applications [Palmer 96, Borsotti 98], their results do not necessarily coincide with the human perception of the goodness of segmentation. For this reason, when a reference image is available or can be generated, empirical discrepancy methods are preferred.

Empirical discrepancy methods, which compare segmentation output with ground truth segmentation of the test data and quantify the levels of agreement and/or disagreement, have the benefit that the direct comparison between a segmented image and a reference image is believed to provide a finer resolution of evaluation, and as such, they are the most commonly used methods of segmentation evaluation. A detailed survey on different discrepancy errors can be found in [Ortiz 06].

Zhang [Zhang 96] has proposed a discrepancy evaluation based on misclassified pixels. Yasnoff et al. [Yasnoff 77], in one of the earliest attempts, have shown that

measuring the discrepancy based only on the number of misclassified pixels does not consider the pixel position error. Their solution is based on the number of misclassified pixels and their distance to the nearest correctly segmented pixels, where each pixel has an associated correct class, and takes measures of classification error from the pixelwise class confusion matrix. Two error measures, the misclassification percentage and pixel distance error are used. However, they only applied it to foreground/background segmentation.

Other discrepancy measures calculate the distances between wrong segmented pixels and the nearest correctly segmented pixels [Odet 02], thus introducing a spatial component to the measure, or are based on differences between feature values measured from regions of the correctly segmented and output images. Huang and Dom [Huang 95] introduced the concept of distance distribution signatures. In [Odet 02] the use of binary edge masks and scalable discrepancy measures are proposed. Although it was adapted to segmentation region maps in [Goumeidane 03], that was only done with background/foreground segmentations.

Another concept sometimes used in evaluation is the receiver operating characteristic (ROC) curve that comes from psychophysics and signal detection theory and has received an important amount of attention within the vision community [Bowyer 01, Brown 06, Fawcett 06]. A ROC curve is a plot of false positive rate against true positive rate as some parameter is varied. The confusion matrix can be used to construct a point in ROC space. ROC curves are commonly used by the medical community, who found them useful in bringing out the *sensitivity* (true positive rate) versus *specificity* (1 − false positive rate), and in recent years have been increasingly adopted in the evaluation of medical imaging techniques [Skudlarski 99, Sorenson 05, Mendonça 06]. The major drawback of ROC curves is that they are only suitable for binary segmentation problems, such as edge detection. An exception of the two-class classification problems is the work of Rees et al. [Rees 02] which addressed multi-class classification evaluation by means of ROC analysis. An extensive literature research on the use of ROC curves can be found in Kelly Zou's bibliography of ROC literature [Zou 05].

In their recent work, Davis and Goadrich [Davis 06] demonstrate that for a given dataset of positive and negative examples, there is a one-to-one correspondence between a curve in ROC space and a curve in Precision-Recall space, such that the curves contain exactly the same confusion matrices, if there is at least one true positive pixel.

3.4 Previous evaluation measures

In this section we present some of the best known measures used in image segmentation evaluation. According to the evaluation approach we divide these measures in region-based and boundary-based.

3.4.1 Region-based evaluation

The region-based scheme evaluates the segmentation accuracy in the number of regions, the locations and the sizes. Let the segmentation be S and the corresponding ground truth be R. Both S and R are functions on the image plane with labels as their function values. A region-based evaluation between two segmented images can be defined as the total amount of differences between corresponding regions. Of course only regions that are likely the same in both segmentations should be taken into account.

Hamming distance

Huang and Dom [Huang 95] introduced the concept of directional Hamming distance between two segmentations, S and R, denoted by $d_H(S \Rightarrow R)$. They began by establishing the correspondence between region $i = \{1, 2, ..., k\}$ of S with region $j = \{1, 2, ..., q\}$ of R such that $s_i \cap r_j$ is maximized. The directional Hamming distance from S to R is defined as:

$$d_H(S \Rightarrow R) = \sum\nolimits_{r_i \in R} \sum\nolimits_{s_t \neq s_j, s_t \cap r_i \neq \emptyset} |r_i \cap s_t| \tag{3.2}$$

where $|\cdot|$ denote the size of a set. Therefore, $d_H(S \Rightarrow R)$ is the total area under the intersections between all $r_i \in R$ and their non-maximal intersected regions from S. A region-based evaluation measure based on normalized Hamming distance is defined as

$$D_H = 1 - \frac{d_H(S \Rightarrow R) + d_H(R \Rightarrow S)}{2 \times |S|} \tag{3.3}$$

where $|S|$ is the image size and $D_H \in [0, 1]$. The smaller the degree of mismatch the closer the D_H is to one.

Moreover, they define two types of errors in region segmentation: missing rate (e_R^m) and false alarm rate (e_R^f). The former indicates the percentage of the points in R being

mistakenly segmented into the regions in S which are non-maximal with respect to the corresponding region in R; while the latter describes the percentage of points in S falling into the regions in R which are non-maximal intersected with the region under consideration. We therefore have

$$e_R^m = \frac{d_H(S \Rightarrow R)}{|S|} \qquad \text{and} \qquad e_R^f = \frac{d_H(R \Rightarrow S)}{|S|} \tag{3.4}$$

These measures have been used to compare several segmentation algorithms by integration of region and boundary information [Freixenet 02].

Local Consistency Error

To compensate for the difference in granularity while comparing segmentations, many measures allow label refinement uniformly through the image. Martin, in his thesis [Martin 02] proposed an error measure to quantify the consistency between image segmentations of differing granularities - *Local Consistency Error* (LCE) that allows labelling refinement between segmentation and ground truth.

Let $r(S, p_i)$ be the set of pixels corresponding to the region in segmentation S that contains the pixel p_i. Then, the local refinement error associated with p_i is

$$E(S, R, p_i) = \frac{|r(S, p_i) \setminus r(R, p_i)|}{|r(S, p_i)|} \tag{3.5}$$

where $\setminus$ denotes set difference. Finally, the overall performance measure is defined as

$$LCE(S, R, p_i) = \frac{1}{N} \sum_i \min\{E(S, R, p_i), E(R, S, p_i)\} \tag{3.6}$$

where $E(S, R, p)$ measures the degree to which two segmentations agree at pixel p, and N is the size of region where pixel p belongs. Note that LCE is an error measure, with a score 0 meaning no error and a score 1 meaning maximum error.

Due to its tolerance of refinement, this measure is not sensible to over- and under-segmentation and may be therefore not applicable in some evaluation situations. Thus, it is only meaningful if the two segmentations have similar number of segments. As observed by Martin [Martin 02], there are two segmentations that give zero error for LCE - one pixel per segment, and one segment for the whole image.

Bidirectional Consistency Error

To overcome the problem of degenerate segmentations, Martin proposed an adaptation of the LCE formula that penalizes dissimilarity between segmentations proportional to the degree of region overlap. If we replace the pixelwise minimum with a maximum we get a measure that does not tolerate refinement at all. The *Bidirectional Consistency Error* (BCE) is defined as:

$$BCE\left(S,R,p_i\right)=\frac{1}{N}\sum_i \max\left\{E\left(S,R,p_i\right),E\left(R,S,p_i\right)\right\} \tag{3.7}$$

Partition distance measure

Cardoso and Corte-Real [Cardoso 05] proposed a discrepancy measure - *partition distance* (d_{sym}) defined as: "given two partitions P and Q of S, the partition distance is the minimum number of elements that must be deleted from S, so that the two induced partitions (P and Q restricted to the remaining elements) are identical". $d_{sym}\left(Q,P\right)=0$ means that no points need to be removed from S to make the partitions equal, i.e., when $Q=P$.

In addition to d_{sym} measure, they proposed an *asymmetric partition distance* defined as: "given two partitions R and Q defined in a set S of N elements, the asymmetric partition distance is the minimum number of elements that must be deleted from S, so that the induced partition Q is finer than the induced partition R".

3.4.2 Boundary-based evaluation

Boundary-based approach evaluates segmentation in terms of both localization and shape accuracy of extracted regions boundaries.

Distance Distribution Signatures

Huang and Dom in [Huang 95] presented a boundary performance evaluation scheme based on the distance between distribution signatures that represent boundary points of two segmentation masks.

Let B_S represent the boundary point set derived from the segmentation S and B_R the set of boundary pixels of the ground truth R. A distance distribution signature from the set B_S to the set B_R of boundary points, denoted $d_B\left(B_s,B_R\right)$, is a discrete

function whose distribution characterizes the discrepancy, measure in distance, from B_S to B_R. The distance from x in set B_S to B_R is defined as the minimum absolute distance from all the points in B_R:

$$d(x, B_R) = \min\{d_E(x, y)\}, \forall y \in B_R \tag{3.8}$$

where d_E denotes the Euclidean distance between points x and y.

The discrepancy between B_S and B_R is described by the shape of the signature, which is commonly measured by its mean and standard deviation. As a rule, $d_B(B_s, B_R)$ with a near-zero mean and a small standard deviation indicates high similarity between segmentation masks. Since the Huang and Dom [Huang 95] paper do not normalize these measures, we cannot determine between two different results which segmentation is the most desirable.

In order to normalize the evaluation measure between 0 and 1, we propose a modification to the distance distribution signature of Huang and Dom. Thus, we introduce a c value that sets an upper limit for the error. For $d(x, B_R) = \min\{d_E(x, y), c\}$, the two boundary distances could be combined in a function similar to the one presented in Equation (3.3):

$$D_B = 1 - \frac{d_B(B_S, B_R) + d_B(B_R, B_S)}{c \times (|R| + |S|)} \tag{3.9}$$

where $|R|$ and $|S|$ are the number of boundary points in reference mask and segmented mask, respectively.

Precision-Recall measures

Martin in his thesis [Martin 02], propose the use of *precision* and *recall* measures to characterize the agreement between the oriented boundary elements (termed *edgels*) of the region boundaries of two segmentations. Thus, given two segmentations, S and R, where S is the result of segmentation and R is the ground truth, precision is proportional to the fraction of *edgels* from S that matches with the ground truth R, and recall is proportional to the fraction of *edgels* from R for which a suitable match was found in S. Precision and recall measures are defined as follows:

$$Precision = \frac{T_p}{T_p + F_p} \qquad Recall = \frac{T_p}{T_p + F_n} \tag{3.10}$$

To compute precision and recall we must determine which true positive pixels are correctly detected, and which detections are false. We could simply consider coincident boundary pixels as true positive and declare all others pixels to be either false positive or false negative. However, this approach would not tolerate any localization error, and would be a poor indicator of performance since the ground truth data contains boundary localization errors as a result of handmade segmentation. In Martin's work, precision and recall are computed using a bipartite matching formulation that matches *edgels* using their location and orientation. He uses Andrew Goldberg's Cost Scaling Assignment package [Goldberg 95] to solve the assignment problem that allows to compare two boundary maps while both permitting localization error and avoiding over-counting. In cases where segmentation classifies pixels as on-boundary or off-boundary, we can correspond boundary pixels instead of *edgels*, and omit the orientation penalty from the *edgels* weight.

In probabilistic terms, precision is the probability that the result is valid, and recall is the probability that the ground truth data was detected. A low recall value is typically the result of under-segmentation and indicates failure to capture salient image structure. Precision is low when there is significant over-segmentation, or when a large number of boundary pixels have greater localization errors than some threshold.

Precision and recall measures have been used in the information retrieval systems for a long time [Raghavan 89]. These measures are also used in the medical community where they go under the names of *specificity* and *sensitivity*, respectively. The interpretation of the precision and recall for evaluation of segmentation are a little different from the evaluation of retrieval systems. In retrieval, the aim is to get a high precision for all values of recall. However in image segmentation, the aim is to get both high precision and high recall. The two statistics may be distilled into a single figure of merit:

$$F = \frac{PR}{\alpha R + (1 - \alpha) P} \tag{3.11}$$

where α determines the relative importance of each term. Following [Martin 02], α is selected as 0.5, expressing no preference for either.

The main advantage of using precision and recall for the evaluation of segmentation results is that we can compare not only the segmentations produced by different algorithms, but also the results produced by the same algorithm using different input

parameters. However, since these measures are not tolerant to refinement, it is possible for two segmentations that are perfect mutual refinements of each other to have very low precision and recall scores.

Earth Mover's Distance

Using the concept of Earth Mover's Distance (EMD) to measure perceptual similarity between images was first explored by Peleg et al. [Peleg 89] for the purpose of measuring distance between two grey-scale images. More recently EMD has been used for image retrieval [Rubner 00].

EMD evaluates dissimilarity between two distributions or *signatures* in some feature space where a distance measure between single features is given. The EMD between two distributions is given by the minimal sum of costs incurred to move all the individual points between the signatures.

Let $P = \{(p_1, w_{p_1}), ..., (p_m, w_{p_m})\}$ be the first signature with m pixels, where p_i is the pixel representative and w_{p_i} is the weight of the pixel; the second signature with n pixels is represented by $Q = \{(q_1, w_{q_1}), ..., (q_n, w_{q_n})\}$; and $D = [d_{ij}]$ the distance matrix where d_{ij} is the distance between two contour points' image coordinates p_i and q_j. The flow f_{ij} is the amount of weight moved from p_i to q_j. The EMD is defined as the work normalized by the total flow f_{ij}, that minimizes the overall cost:

$$EMD(P,Q) = \frac{\sum_i \sum_j f_{ij} d_{ij}}{\sum_i \sum_j f_{ij}} \tag{3.12}$$

As pointed by Rubner et al. [Rubner 00], if two weighted point sets have unequal total weights, EMD is not a true metric. It is desirable for robust matching to allow point sets with varying total weights and cardinalities. In order to embed two sets of contour features with different total weights, we simulate equal weights by adding the appropriate number of points, to the lower weight set, with a penalty of maximal distance. Since normalizing signatures, with the same total weight do not affect their EMD, we made $\sum_{i,j} f_{ij} = 1$. Equation (3.12) becomes,

$$EMD(P,Q) = \sum_i \sum_j f_{ij} d_{ij} \tag{3.13}$$

subject to the following constraints: $f_{ij} \geq 0$, $\sum_j f_{ij} = w_{p_i}$ and $\sum_i f_{ij} = w_{q_j}$.

As a measure of distance for the EMD ground distance we use

$$d_{ij} = 1 - e^{-\frac{\|p_i - q_j\|}{\alpha}} \tag{3.14}$$

where $\|p_i - q_j\|$ is the Euclidean distance between p_i and q_j and α is used in order to accept some deformation resulted from manual segmentation of ground truth. The exponential map limits the effect of large distances, which otherwise dominate the result.

3.5 Weighted evaluation measure

In the context of image segmentation, the reference mask is generally produced by humans. There is an agreement that interpretations of images by human subjects differ in granularity of label assignments, but they are consistent if refinements of segments are admissible [Martin 02]. One desirable property of a good evaluation measure is to accommodate refinement only in regions that human segmenters could find ambiguous and to penalize differences in refinements elsewhere. In addition to being tolerant to refinement, any evaluation measure should also be robust to noise along region boundaries and tolerant to different number of segments in each partition.

For the purpose of evaluating image segmentation results, a correspondence between the examined segmentation mask, S, and the reference mask, R, has initially been established, indicating which region of S better represents each reference region. This is performed by associating each region r_i of mask R with a different region s_j of mask S on the basis of region overlapping, i.e. s_j is chosen so that $r_i \cap s_j$ is maximized. The set of pixels assigned to s_j but not belonging to r_i are false positives, F_p, that can be expressed as $F_p = s_j \cap \overline{r}_i$, where $\overline{r}_i$ denotes the complement of r_i. The pixels belonging to r_i but not assigned to s_j are false negatives, F_n, and can be expressed as $F_n = \overline{s}_j \cap r_i$.

The minimum required overlap between r_i and s_j is 50% of the reference region. Pixels belonging to regions where this ratio is not achieved are considered as false pixels. These measure quantify the errors due to under and over segmentation. Clearly, more visually significant regions that were missed in the segmented mask are assigned a significantly higher error.

The normalized sum of false detections is an objective discrepancy measure that quantifies the deviation of the results of segmentation from the ground truth and can be expressed as:

$$\varepsilon_F = \frac{F_p + F_n}{2N} \tag{3.15}$$

where N is the set of all pixels in the image. The value of ε_F is proportional to the total amount of errors and indicates the accuracy of region boundaries localization. The quality of the segmentation is inversely proportional to the amount of deviation between the two masks.

In applications where the final evaluator of quality is the human being, it is fundamental to consider human perception to deal with the fact that different kind of errors are not visually significant to the same degree. To build a spatial accuracy measure with high perceptive meaning, we have to use the following assumptions:

- The visual relevance of a wrong pixel increase with its distance from the border of the reference mask.
- As we move away from the border, false negative pixels achieve always greater relevance, since they mean that a bigger part of the object is being missed.
- With false positives the situation is slightly different. Although they also increase their relevance at far locations, that increment tends to stabilize with the distance from the reference border.

To accommodate human perception, the different error contributions are weighted according to their visual relevance. Gelasca et al. [Gelasca 04] present a psychophysical experiment to assess the different perceptual importance of errors. They conclude that a false positive pixel contributes differently to the quality than a false negative. False negatives are more significant, and the larger the distance the larger the error.

We define two weighted functions w_p and w_n to deal with that fact where w_p is associated with false positive pixels and w_n is associated with false negative pixels. Let d_p be the distance of a false positive pixel from the boundary of the reference region, and d_n be the distance of a false negative pixel.

$$w_p = \frac{\alpha_p \log(1 + d_p)}{D} \tag{3.16}$$

$$w_n = \frac{\alpha_n d_n}{D} \tag{3.17}$$

These functions are normalized by the image diagonal distance D. The weighted function for each false pixel is also represented in Figure 3.4.

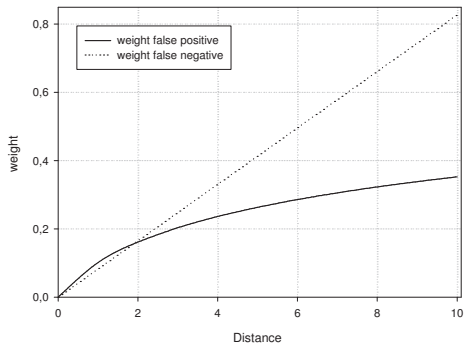


Figure 3.4: Weight functions for false negative and false positive pixels.

The weights for false negative pixels increase linearly and are larger than those for false positive pixels at the same distance from the reference region border. As we move away from the border of an object, missing parts are more important than added background, e.g., in medical imaging, it may be enough that the segmented region overlaps with the true region, so the tumour can be located. But if there are missing parts of the tumour the segmentation results will be poor.

To obtain a measure between $[0, 1]$, we normalize the total amount of weight by the image size. The discrepancy measure of weighted distance, ε_w, becomes:

$$\varepsilon_w = \frac{1}{N}\left(\sum_{f_n} w_n + \sum_{f_p} w_p\right) \tag{3.18}$$

where f_n and f_p represent the false pixels. We define a new measure of similarity as $s_w = 1 - \varepsilon_w$. The value of $s_w = 1$ indicates a perfect match between the segmentation and the reference mask.

3.6 Analysis on evaluation methods

We conducted two experiments to validate the measure proposed in this work. The first with results obtained from manual segmentations and the second with synthetically generated segmentations.

To achieve comparative results about different evaluation methods, two strategies can be followed: the first one consists in applying the evaluation methods to segmented

images obtained from different segmentation approaches. The second one consists in simulating results of segmentation processes. To exempt the influence of segmentation algorithms, the latter has been adopted and a set of images obtained from manual segmentation available at the Berkeley Segmentation Database [Martin 01] was used. As the ground truth is not unique, we used as ground truth the manual segmentation with the best F-measure against all the others. Figure 3.5 shows the segmentation results used in this comparative study where result (i) is also used to set up the weighted parameters of false pixels.

A good evaluation measure has to give large similarity values for results (a)-(e) and has to strongly penalize other results ((f)-(i)). Figure 3.6 shows the comparison results between the proposed method and the methods presented in Section 3.4.1, for the images in Figure 3.5, expressed in terms of region-based evaluation.

Due to its tolerance to refinement, LCE gives low error (high similarity) scores, even when the segmentation result is very different from the ground truth (images (f)-(i)). Measure D_H has a similar behaviour. BCE and d_{sym} give good evaluations for images ((f)-(i)). However, since these measures are not tolerant to refinement, the results are poor for results ((a)-(e)).

The results obtained from images ((a)-(e)) show that the proposed measure is tolerant to refinement, in accordance with the way human perceive visual information. Since our measure weights the segmentation errors according to their distance to the correct segmentation it strongly penalizes segmentation errors of images ((f)-(i)).

Results of boundary-based evaluation on the same set of segmentation results are reported in Figure 3.7. On comparing the results of the boundary-based measures, it is made evident that they are well correlated. EMD tolerates well some amount of deformations that normally happens in the manual segmentation process. However, when the number of pixels in ground truth differs a lot from the number of pixels in the segmented image, EMD gives poor results. Despite its success, the EMD method still needs to be refined to address the limitation in the complexity of algorithm that require to be further reduced. The D_B measure gives similar results with F-measure, but it is even more intolerant to refinement.

Table 3.1 presents the evaluation results obtained from a set of trivial synthetically generated segmentations presented in Figure 3.8, where we make constant the number of false detections in each segmentation.

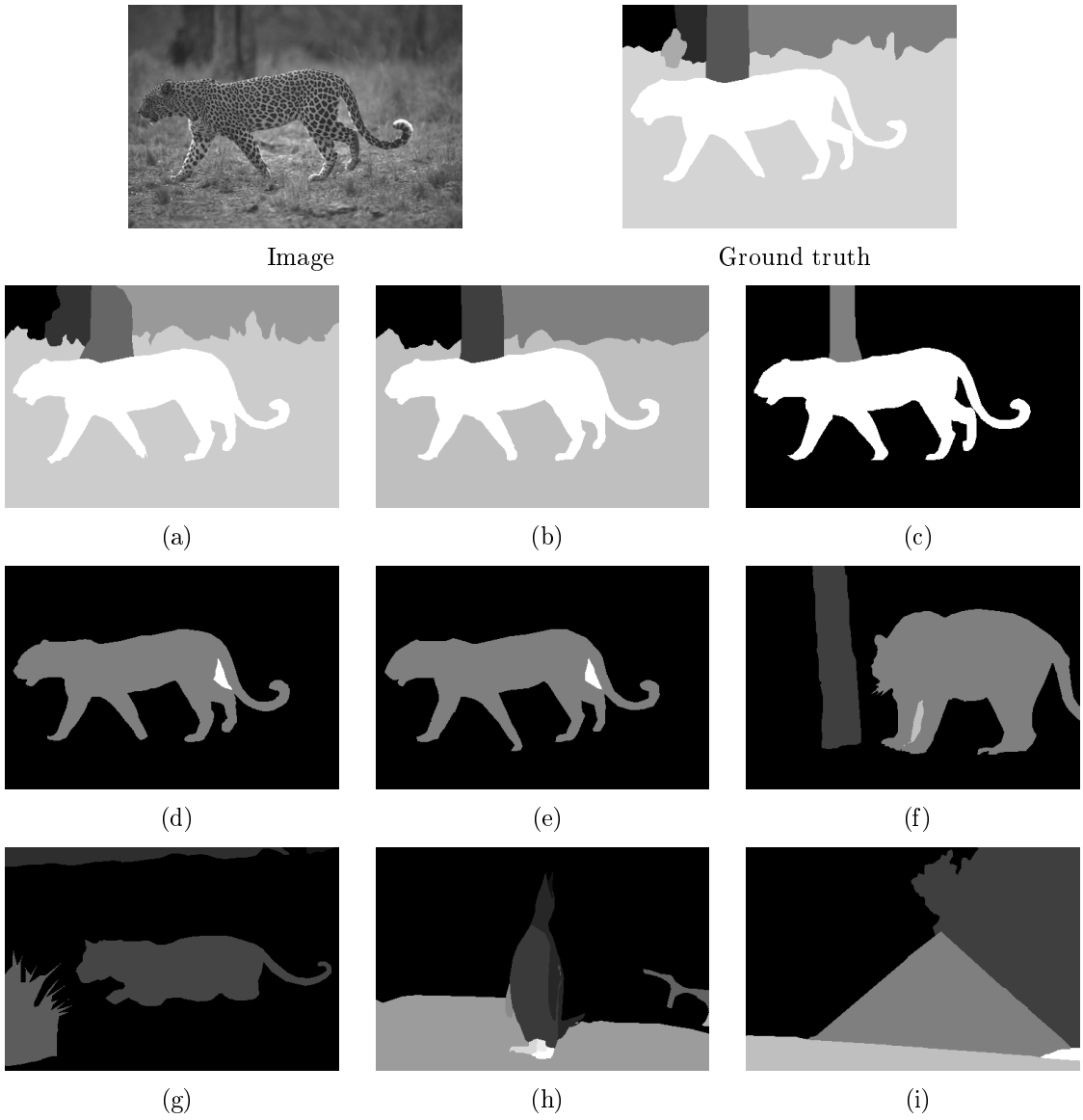


Figure 3.5: The first row shows original image and the segmentation ground truth. From (a) to (e) we have different manual segmentations of the same image. Images from (f) to (i) are segmentation results of other images.

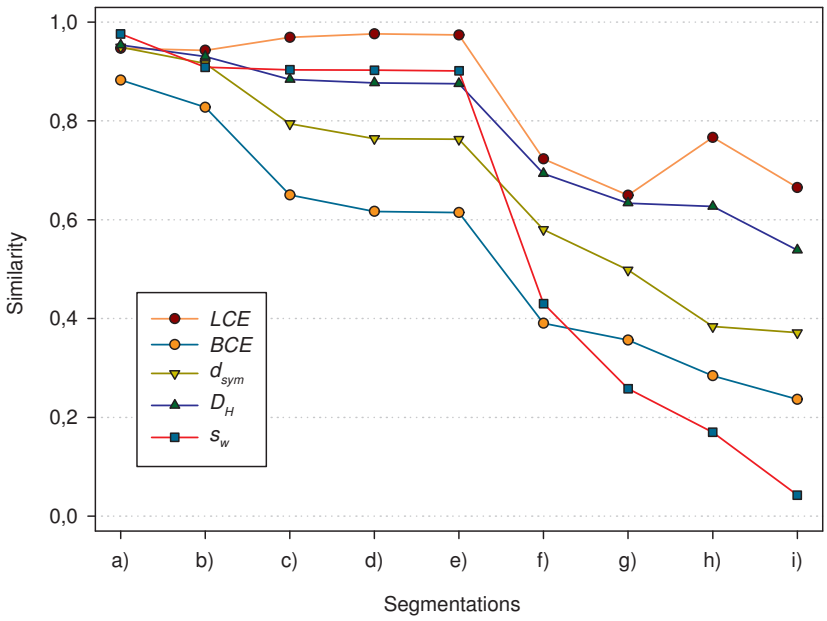


Figure 3.6: Evaluation of segmentation, in terms of similarity, from a set of evaluation schemes based on regions.

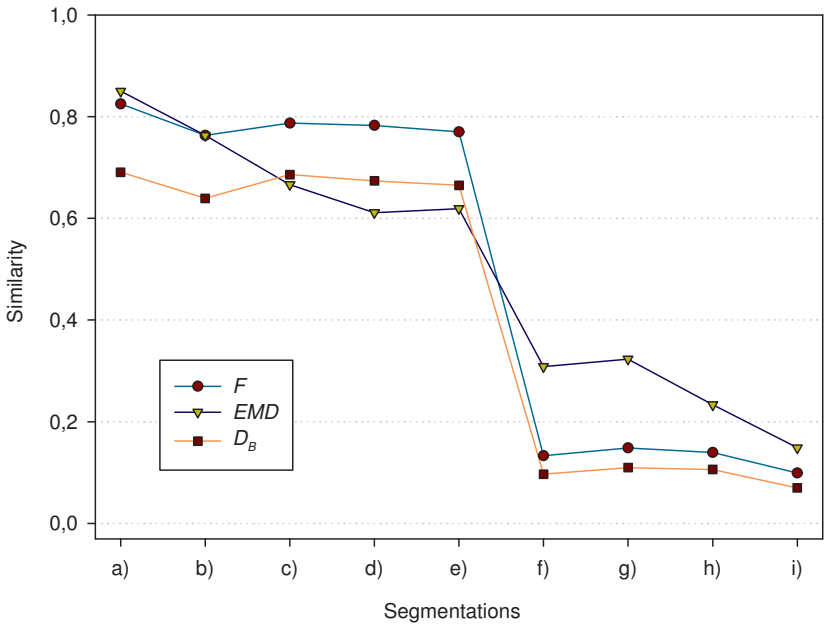


Figure 3.7: Evaluation of segmentation, in terms of similarity, from a set of evaluation schemes based on boundaries.

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

Figure 3.8: Synthetically generated set of segmentations, where (a) is the reference.

Since LCE, BCE, d_{sym} and D_H, are just proportional to the total amount of false detections, different position of those pixels do not affect the similarity. This makes those methods unreliable for applications where the results will be presented to humans. Note that s_w produces results that agree with the visual relevance of errors.

Table 3.1: Numerical evaluation of segmentations from Figure 3.8.

images	LCE	BCE	$\mathbf{d_{sym}}$	$\mathbf{D_H}$	$\mathbf{s_w}$
(b)	0.99380	0.98088	0.99349	0.99349	0.99741
(c)	0.99380	0.98088	0.99349	0.99349	0.99612
(d)	0.99380	0.98088	0.99349	0.99349	0.99159

3.7 Summary

In this chapter we introduce a new approach for segmentation evaluation based on regions that takes into account, using a single metric, not only the accuracy of the boundary localization but also the under-segmentation and over-segmentation effects according to the ambiguity of the regions, regardless to the number of segments in each partition. The proposed metric is based on examining the spatial accuracy of segmentation results using a manually generated reference mask. Its output is a weighted sum of misclassified pixels, effectively indicating how well the examined segmentation mask corresponds to the reference one. We introduce a modification to the distance signature of Huang and Dom, the D_B measure; and apply the concept of Earth Mover's Distance to the evaluation of image segmentation. We experimentally demonstrated the efficiency of the new measure against well known methods. This metric can be applied both to automatically provide a ranking among different segmentation algorithms and to find an optimal set of input parameters of a given algorithm. This measure will be used in the evaluation of image segmentation experimental results in Chapter 6.

CHAPTER 4

Hybrid spatial segmentation: the model

This chapter presents a new framework to spatial image segmentation. The main idea is to use atomic regions to guide a segmentation using the intensity and gradient information through a spectral graph-cut approach. This method produces simpler segmentations less over-segmented and it is compared favourably with state-of-the-art methods (See also Chapter 6).

4.1 Introduction

Image segmentation is one of the largest domain in image analysis, and aims at identifying regions, the so-called segments that have a specific *meaning* within images. Another definition of image segmentation is the identification of regions that are *uniform* with respect to some parameter, such as image intensity, texture or motion. While the latter definition is often used for technical reasons, the former definition should be preferred from an application point of view. Although the effort made in the computer vision community there is no algorithm that is known to be optimum in image segmentation. Different images require different methods, different applications demand new approaches. Much research is being done to discover new methods building up on previous ideas.

Since the Gestalt movement in psychology [Wertheimer 38], it has been known that perceptual grouping plays a powerful role in human visual perception. The main goal of this chapter is to develop an algorithm for efficient segmentation of a grey level

image that a) identifies perceptually homogeneous regions in the images, b) makes minimal assumptions about the scene, and c) avoids merging of multiple objects into single segments and vice-versa. The presentation of an improved rainfalling watershed approach, the definition of a new structure for region based graph, the presentation of a new similarity function, and the application of multiclass normalized cuts to group atomic regions are the main contributions of this chapter.

Spectral segmentation is a promising approach to perceptual grouping or image segmentation that takes into account global image properties as well as local spatial relationships. It treats image segmentation as a graph partitioning problem. A common characteristic of these techniques is the idea of clustering/separating pixels or other image elements using the dominant eigenvectors of a matrix derived from the pairwise pixel similarities, as measure by one or more cues. It thus segments an image from a global point of view. The advantage of having a global objective function is that hard decisions are made only when information from the whole image is taken into account at the same time [Malik 01].

These methods use the eigenvectors and the eigenvalues of a matrix representation of a graph to partition an image into disjoint regions. A salient region in the image is the one for which the similarity across its border is small, whereas the similarity within the region is large. A well known spectral graph analysis method is normalized cut algorithm [Shi 00] that minimizes a discriminative energy function defined in terms of the graph link weights. The normalized cut algorithm is a graph partitioning algorithm that has previously been used successfully for image segmentation. It has originally applied to pixels by considering each pixel in the image as a node in the graph. One important issue of this approach is the size of the corresponding similarity matrix. If the graph node set contains all the pixels of an image, the size of the similarity matrix is equal to the squared number of pixels, and therefore generally too large to fit into computer memory completely.

The energy function modelled by the normalized cut is capable of generating clean results, even though the intensity regions can sometimes be broken into a small number of pieces. As a recent paper [Carson 02] notes: "*large, uniform background areas in the image are sometimes arbitrarily split into two pieces due to the use of position as a feature. On the whole, however, including position yields better segmentation results than excluding it.*"

Since the use of positional information as a feature is known to be problematic [Carson 02], several authors have explored alternatives. One possibility is to perform a fairly atomic segmentation at the very beginning, and then compute feature vectors from these regions rather than from pixels. Thus to reduce the size of the graph, nodes can be used to represent disjoint atomic regions covering the image instead of single pixels. The output of the preliminary segmentation step is a set of spatially coherent clusters, which could then be used to compute the affinity matrix for the spectral-based segmentation algorithm. It can also be used directly for segmentation by a merging process.

Our WNCUT[1] approach overcomes the problem of over-segmentation in the preliminary segmentation stage by using the spectral methods to intelligently re-assemble the sub-set of atomic regions into the final segmented object based on a similarity function among the regions. Actually our approach prefers the objects to be over-segmented into a number of smaller regions to ensure that a minimal amount of background is connected to any of the object regions.

In order to apply WNCUT, first we must represent the micro-regions in graph terms. Suppose that the image under consideration is partitioned into a set of k disjoint regions denoted by $R = \{R_1, ..., R_k\}$. Then R can be represented by a set of k nodes in an undirected graph, called the Region Similarity Graph (RSG). An evident computational advantage is obtained describing the image by a set of regions instead of pixel in the RSG structure: it enables a faster region merging in images with higher spatial resolution.

4.2 Overview of the proposed method

The proposed methodology has four major stages. First, we smooth image noise, as a pre-processing stage, using an anisotropic filter. Next, we create an over-segmented image based on the initial magnitude gradient image. In the third stage, the over-segmented image will be the input for the image RSG construction. Finally, we apply a multiclass normalized cut approach on the RSG. A block diagram of the proposed method is depicted in Figure 4.1.

This framework integrates edge-based and region-based segmentation with spectral-

[1] From *Watershed Normalized Cut.*

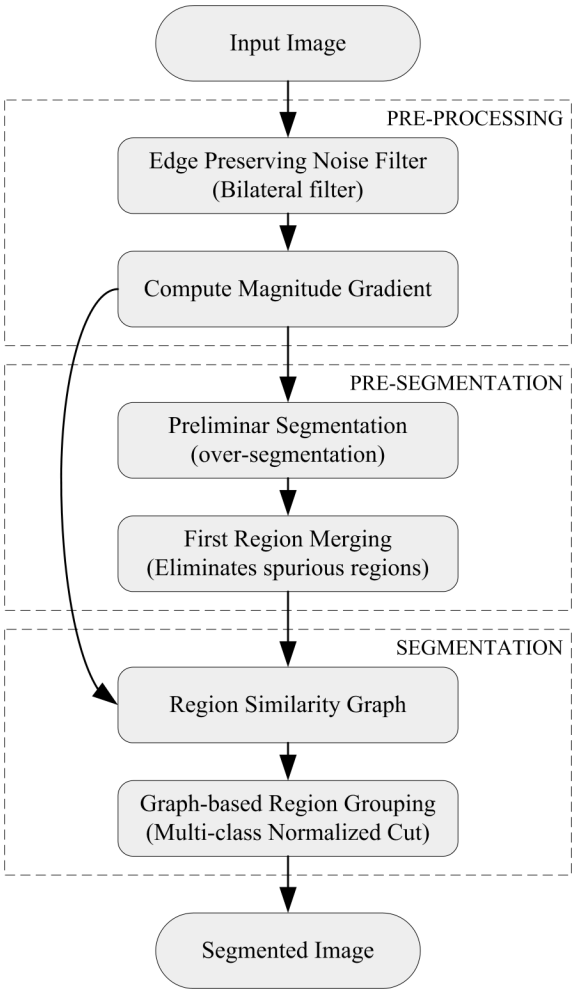
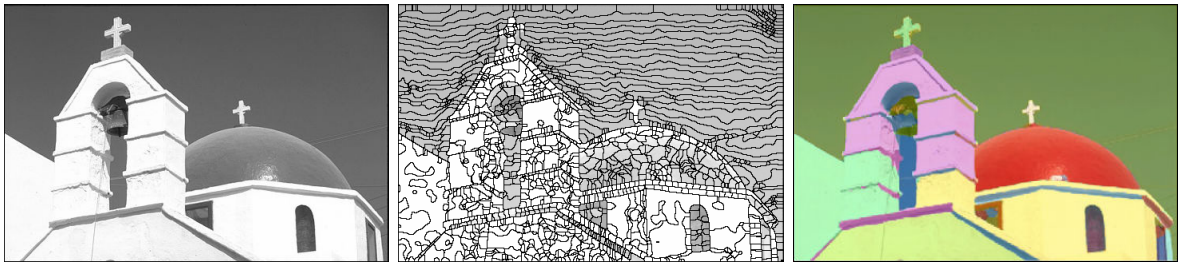


Figure 4.1: Block diagram of the proposed method.

based clustering as follows:

1. Reduce image noise using the bilateral filter;
2. Compute gradient magnitude and remove the weakest edges by gradient minima suppression (*pre-flooding*);
3. Make initial partitioning using the gradient information;
4. Make a simple post-processing to remove single tiny regions by merging them with neighbouring regions. These regions are considered to be spurious;
5. Calculate the statistics of all atomic regions;
6. Initialize the region similarity graph where each node corresponds to an atomic region;
7. Use a spectral-based approach in order to obtain the final segmentation.

We illustrate the algorithm by an example shown in Figure 4.2. An input image is decomposed into a number of atomic regions to reduce the graph size in a pre-segmentation stage as in Figure 4.2.(b). Each atomic region has nearly constant intensity and it is represented by a node in the graph G. Two vertices are connected if their atomic regions are adjacent (i.e. share the same boundary). Figure 4.2.(c) shows the result produced by our algorithm where each closed region is assigned a colour.



(a) Input image (b) Atomic regions (c) Segmentation

Figure 4.2: Example of image segmentation. (a) Input image. (b) Atomic regions. Each atomic region is a node in the graph G. (c) Segmentation (labelling) result.

4.3 Noise reduction and gradient computation

Images taken with digital cameras will pick up noise from a variety of sources. As the watershed algorithm is very sensitive to noise it is desirable to apply noise reduction filter in the pre-processing step. Several methods have been proposed in the literature to reduce the spurious boundaries created due to noise and produce a meaningful watershed segmentation. Ogor [Ogor 95] proposed morphological opening and closing. Gauch [Gauch 99] used Gaussian blurring. Hernandez and Barner [Hernandez 00] suggested median filtering. However, some of these filters tend to blur image edges while they suppress noise which is undesirable for the watershed algorithm.

4.3.1 Bilateral filter

To prevent this effect we use the non-linear bilateral filter [Tomasi 98]. The bilateral filter was first introduced by Smith and Brady under the name "SUSAN" [Smith 97] as a non-linear filter that combines domain and range filtering. It was rediscovered later by Tomasi and Manduchi [Tomasi 98] who called it the 'bilateral filter' which is now the most commonly used name.

The basic idea underlying bilateral filtering is to replace the intensity of a pixel (*nucleus*) by taking a weighted average of the pixels within a neighbourhood (in a circle) with the weights depending on both the spatial and intensity difference between the central pixel and its neighbours. In smooth regions, pixel values in a small neighbourhood are similar to each other and the bilateral filter acts essentially as a standard domain filter, averaging away the small, weakly correlated differences between pixel values caused by noise. Bilateral filter preserves image structure by only smoothing over those neighbours which form part of the "same region" as the central pixel.

Expressed formally, given an input signal $f(x)$, using a continuous representation notation as in [Tomasi 98], the output signal $h(x)$ is obtained by:

$$h(x) = \frac{\int_{\Omega} f(\xi)\, c(\xi, x)\, s(f(\xi), f(x))\, d\xi}{\int_{\Omega} c(\xi, x)\, s(f(\xi), f(x))\, d\xi} \tag{4.1}$$

where $c(\xi, x)$ measures the spatial closeness between the centre pixel x and a nearby point ξ; the photometric similarity is given by $s(f(\xi), f(x))$, and Ω represents the filter support.

Considering a grey level image I, the result of the bilateral filter I^{bf} is defined as:

$$I^{bf}(p_0) = \frac{\sum_{p \neq p_0} I(p) \cdot c(p, p_0) \cdot s(I(p), I(p_0))}{\sum_{p \neq p_0} c(p, p_0) \cdot s(I(p), I(p_0))} \tag{4.2}$$

where the so-called nucleus $p_0 := (u_0, v_0)$ is the pixel which is going to be filtered and $p := (u, v)$ is a pixel which belongs to the convolution mask around the nucleus.

The decreasing weight functions c and s, which represent *closeness* (in the spatial domain) and *similarity* (in the range domain) respectively, are Gaussian distributions of the form:

$$c(p, p_0) = \exp\left(-\frac{(p - p_0)^2}{2\sigma_s^2}\right) \tag{4.3}$$

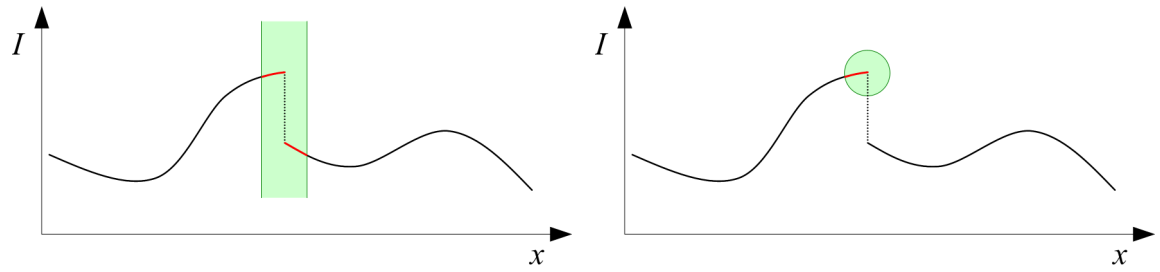
$$s(I(p), I(p_0)) = \exp\left(-\frac{(I(p) - I(p_0))^2}{2\sigma_r^2}\right) \tag{4.4}$$

The parameter σ_s is the standard deviation of the spatial component of the blurring function and σ_r is the standard deviation of the intensity component. The non-linearity

of the filter comes from the division by the two Gaussian distributions and from the dependency on the pixel intensities through the spatial component.

We can control the spatial support of the filter and thus the level of blurring by varying σ_s. By varying σ_r we can control how much an adjacent pixel is down weighted because of the intensity difference. If the grey level difference between two regions is larger than σ_r, the algorithm computes averages of pixels belonging to the same region as the reference pixel. Thus, the algorithm does not blur the edges which is its main scope. In our experiments we apply the bilateral filter implementation of Smith and Brady [Smith 97] with $\sigma_r = 30$ and $\sigma_s = 4$.

Figure 4.3 shows the comparison between the usual unilateral filter (e.g. the mean filter) and the bilateral filter for an 1D signal. Since the spatial support of the bilateral filter is a circle with radius σ_s the bilateral filter preserves discontinuities where the unilateral filter uses both object and background intensities in the smoothing process, as showed by the red lines of Figure 4.3.



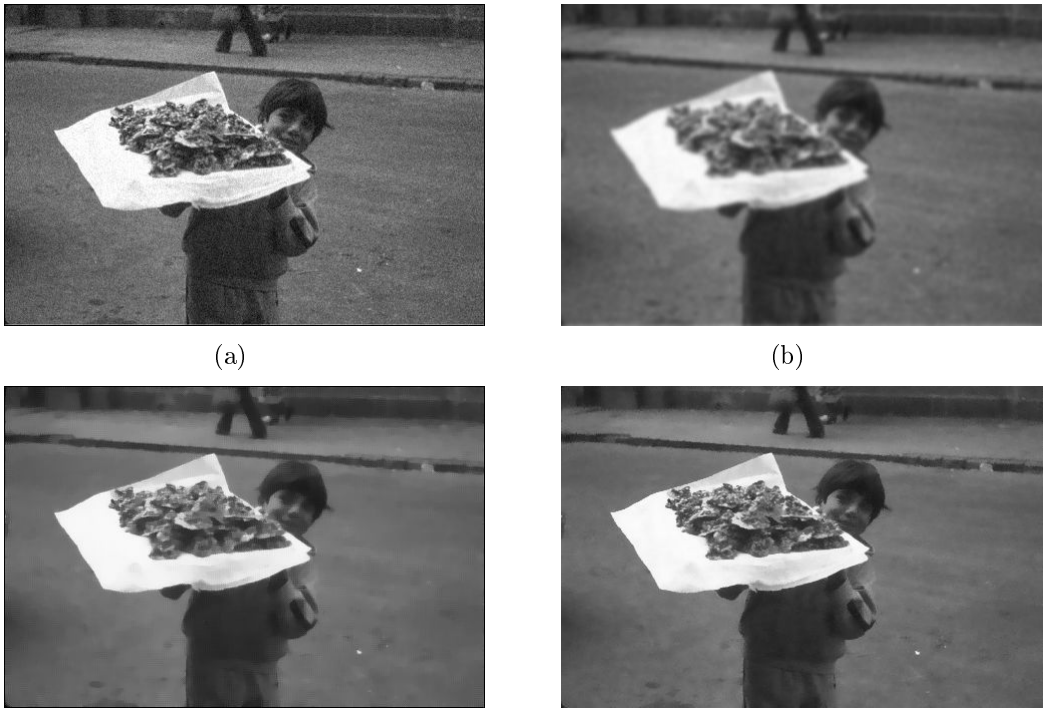
(a) (b)

Figure 4.3: Unilateral *versus* bilateral filter. (a) Unilateral filter. (b) Bilateral filter.

It is well known that median filters preserve the location of edges while eliminating structures such as impulses, which can correspond to undesirable local intensity minima or maxima. In cases where the central pixel is uncorrelated with the whole neighbourhood, and hence it is treated as pulse noise, the denominator of Equation (4.2) is zero. This is dealt by replacing the intensity of the pixel intensity with the median of its closest neighbours.

Figure 4.4 shows the result of smoothing an image with Gaussian smoothing, anisotropic diffusion [Perona 90] and bilateral filter, respectively.

Since bilateral filtering does not involve the solution of partial differential equations it is a good non-iterative alternative to anisotropic diffusion proposed by Perona and



(a) (b)

(c) (d)

Figure 4.4: Noise reduction filters. (a) Image with added Gaussian noise with $\sigma = 10$. (b) After Gaussian filter with $\sigma = 2$. (c) After anisotropic diffusion filter with 100 iterations. (d) After bilateral filter with $\sigma_r = 30$ and $\sigma_s = 4$.

Malik [Perona 90]. Despite the difference in implementation both methods are designed to smooth the image while edges are preserved.

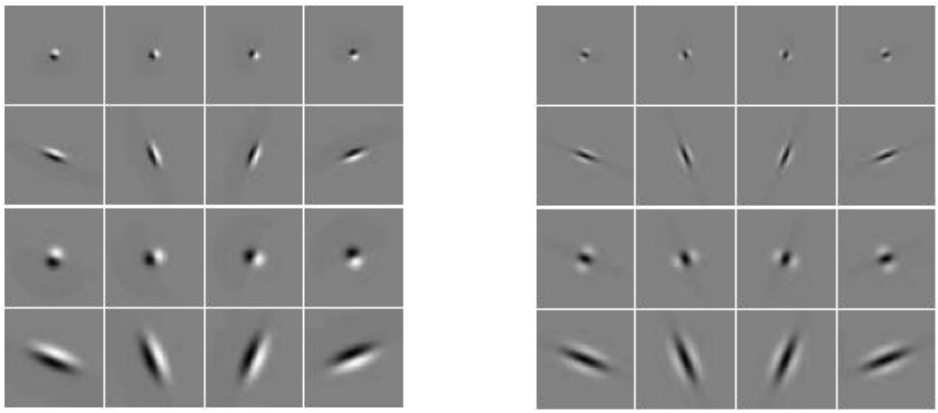
4.3.2 Gradient computation

The gradient computation step is crucial as it is used in two different sections of the proposed algorithm: in the preliminary segmentation and in the construction of the region similarity graph.

Provided that the original noise level is not high or the noise has been effectively reduced in the first stage, then any of the known gradient operators, namely classical Sobel, Prewitt or morphological operators may perform well. However, if the original noise level is high or the noise has not been effectively reduced in the first stage, the use of small scale Gaussian derivative filters may further reduce noise.

Images are first convolved with Gaussian oriented filter pairs to extract the magnitude of orientation energy (OE) of edge responses as used by Malik et al. in [Malik 01]. The filters shown in Figure 4.5 are tuned to detect edges of different shapes, parame-

terised by $\rho = \{\rho_o, \rho_s, \rho_e\}$, where ρ_o, ρ_s and ρ_e refer to orientation, scale and elongation respectively.



$F_o(\rho)$ $F_e(\rho)$

Figure 4.5: Linear filters of 4 orientations, 2 elongations and 2 scales, in both odd and even phases that form quadrature pairs.

Given image I, the orientation energy approach can be used to detect and localize the composite edges, and it is defined as:

$$OE(\rho) = (I * F_e(\rho))^2 + (I * F_o(\rho))^2 \tag{4.5}$$

where $F_e(\rho)$ and $F_o(\rho)$ represent a quadrature pair of even and odd-symmetric filters which differ in their spatial phases. The even-phase filters are the second-order derivative and the corresponding odd-symmetric filters are their Hilbert transforms which correspond to the first-order derivative, both smoothed with Gaussian functions specified by ρ.

At each pixel i, we can define the dominant orientation energy $(OE_i(\rho)^*)$ and the parameter (ρ_i^*) as the maximum energy across scale, orientation and elongation:

$$OE_i(\rho)^* = \max OE(\rho) \qquad \rho_i^* = \arg\max OE(\rho) \tag{4.6}$$

Orientation energy $OE(\rho)$ has a maximum response for contours of shape ρ, whereas the zero-crossing of filter $F_e(\rho)$ locate the positions of the edges. The value OE^* is kept at the location of i only if it is greater than or equal to the neighbouring values. Otherwise it is replaced with a value of zero.

4.4 Over-segmentation as pre-processing

An ideal over-segmentation should be easy and fast to obtain, and should not contain too many segmented regions and it should have its region boundaries as a superset of the true image region boundaries. In this section we present a pre-processing stage that groups pixels into "atomic regions". The motivations of this preliminary grouping stage resemble the perceptual grouping task: (1) abandoning pixels as the basic image elements, we instead use small image regions of coherent structure to define the corresponding graph representation. In fact, since the real world does not consist of pixels, it can be argued that this is even a more natural image representation than pixels as those are merely a consequence of the digital image discretization; and (2) the number of pixels in natural images is high even at moderate resolutions. By treating regions as the elementary unit for image processing, we can reduce the computational complexity without a corresponding loss of accuracy.

This section presents two strategies for the pre-segmentation stage: chunk graphs and rainfalling watershed. Alternatively, the atomic regions could be computed using other methods, such as normalized cuts [Ren 03], graph cuts [Felzenszwalb 04], edge detection followed by edge tracing and contour closing [Barbu 05] or by an over-segmented version of the mean-shift approach [Luo 04].

4.4.1 Chunk graph

The objective is to partitioning the image into a number of disjoint regions so that each region has consistent intensity. In this section we propose a graph coarsening approach based on a chunk graph defined below. This refinement or coarsening could be thought of as a hierarchical structure on the image where graph computation is performed at different levels of granularity with the connected pixels from the lower level collapsing into nodes in the higher level. In addition to significantly reducing the number of nodes in the graph, this coarsening creates small aggregates of pixels which have similar intensities, adapted to the image at hand.

A chunk graph $G' = (V', E')$ of a graph G is defined as follows: Each node of G' represents a chunk, which is a subset of G; each chunk corresponds to a set of homogeneous pixels; chunks on G' are disjoint and their union is G.

A graph is then constructed to present the spatial relationship of the pixels. The

graph G is initially set to represent the 8-neighbour of each pixel in the image. Since we want to find sets of homogeneous nodes the processing order of the nodes is not important. The edges corresponding to connections between homogeneous nodes are removed. The resulting graph G' contains nodes that represent homogeneous atomic regions in the image. Therefore, we transform graph $G = (V, E)$ into a new graph $G' = (V', E')$, where $E' \subseteq E$. Graph G' is composed by a set of subgraphs (*chunks*) that follow the normalized cut criterion in their construction. This means that edges between two nodes in the same chunk should have relatively high similarity weights, and edges between nodes in different chunks should have lower similarity weights. Figure 4.6 shows an example of a two level chunk graph.

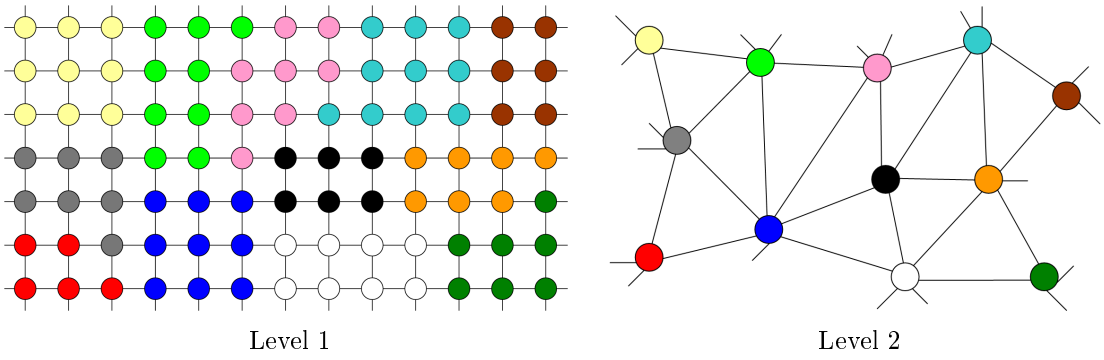


Figure 4.6: Graph chunk sampling. Computation is performed at different levels of granularity where the connected pixels from the lower level collapse into nodes in the higher level.

In the following discussion, we denote nodes of graph G' using v_i and v_j, and use e_{ij} to represent the edge connecting nodes v_i and v_j. An edge e_{ij} is labelled according to the absolute difference of the mean intensities of nodes v_i and v_j. A merge, $M(i, j)$, is a graph transformation operation that merges the nodes v_i and v_j. The procedure of node merging is actually to integrate two or more chunks into a bigger one. It is also called an *edge contraction* as the edge e_{ij} is removed. The graph G is transformed in a new graph G' that has node v_i and all other nodes of G except node v_j. The graph links weights between the atomic regions are defined in terms of the smallest matching cost for fitting both atomic regions by the same intensity.

By the above definition, a merge always reduces the total number of regions. This merge process is guaranteed to converge. A decision function, called the *merge criterion* determines whether two nodes should be merged. Basically, this merge criterion measures the strength of the boundary between two regions by comparing two quan-

tities: one based on measuring the dissimilarity between elements along the boundary of the two components and the other based on the measure of the dissimilarity among neighbouring elements within each of the two components. We define two measures

$$In_w(A) = \max_{e_{ij} \in N_8(A,E)} w_{ij} \tag{4.7}$$

$$Out_w(A,B) = \min_{v_i \in A, v_j \in B, (v_i,v_j) \in E} w_{ij} \tag{4.8}$$

where A and B are regions, $In_w(A)$ is the internal variation within the region, $N_8(A,E)$ are the 8-neighbours of A, and $Out_w(A,B)$ is the external variation between the regions. We merge together two regions[2] when the external variation between them is small regard to their respective internal variations

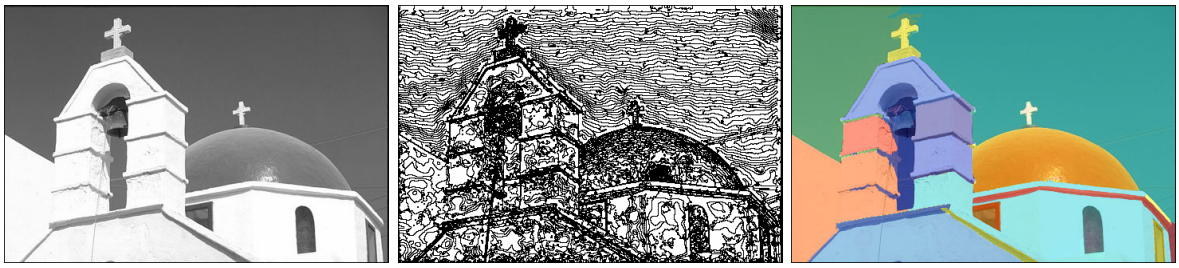
$$Out_w(A,B) \leq MIn_w(A,B) \tag{4.9}$$

with

$$MIn_w(A,B) = \min\left(In_w(A) + \tau(A), In_w(B) + \tau(B)\right) \tag{4.10}$$

where the threshold value $\tau(A) = \alpha/|A|$ determines how large the external variation can be with regards to the internal variation to still be considered similar, α is some constant parameter, and $|A|$ is the size of A.

Neighbouring pixels whose properties are similar enough are joined. A pixel is not chained until all the pixel pairs which are more similar are chained. This ensures that each pixel is always joined to its best fit neighbour. We illustrate the algorithm by an example on image segmentation shown in Figure 4.7.



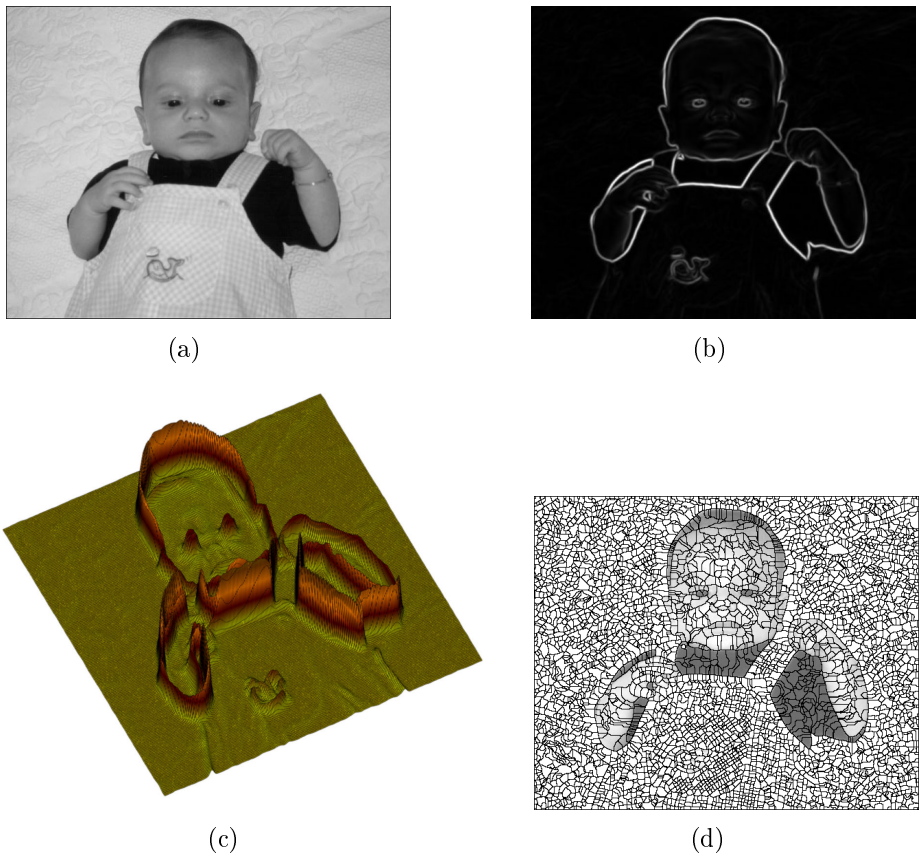
(a) Input image (b) 9461 chunks (c) Segmentation

Figure 4.7: Example of image segmentation. (a) Input image. (b) Atomic regions produced by the chunk graph. (c) Segmentation result.

[2]A region could be formed only by a single pixel.

4.4.2 The watershed transform

Watershed transform is a classical and effective method for image segmentation in grey scale mathematical morphology. For images the idea of the watershed construction is quite simple. An *activity image* is considered as a topographic relief, as shown in Figure 4.8, where for every pixel in position (x, y), its activity level plays the role of the z-coordinate in the landscape. Local maxima of the activity image can be thought of as mountain tops, and minima can be considered as valleys.



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

Figure 4.8: Image as a topographic relief. (a) Intensity image, (b) gradient and (c) its topographic representation. (d) Watershed segmentation result.

A drop of water placed anywhere on this surface will follow the path of steepest descent until it reaches a minimum. This idea helps to establish an equivalent relationship among pixels that trace to the same minimum and it is used to group pixels in the image under different catchment basins. Thus, the algorithm works by finding the minima of the surface, which correspond to the catchment basins and tries to group every other pixel under one of these basins, producing a segmented output. Since

most structures contain several catchment basins, generally watershed segmentation produces a large number of regions even for simple images.

A general topographic interpretation of a two-dimensional function is depicted in Figure 4.9. The most important notions in this context are the ones of minima, catchment basins (or simply basins), and watersheds that are separating basins from each other. Using this terminology, the watershed approach transforms an image into a disjoint set of basins plus a set of watersheds.

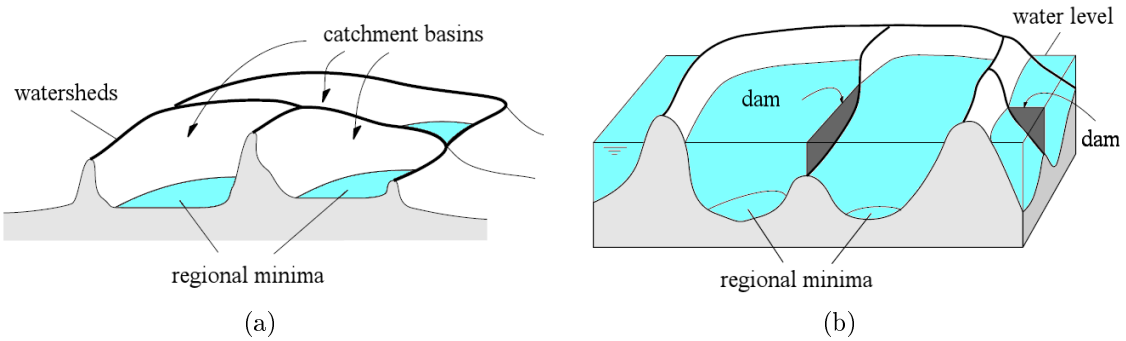


Figure 4.9: (a) Minima, catchment basins, and watersheds on the topographic representation of a gradient image. (b) Building dams at the places where the water coming from two different minima would merge (adapted from [Vincent 91]).

The watershed approach has been applied in many image segmentation problems and it is known to yield robustness in extracting meaningful regions and contours [Roerdink 01]. The watershed transform approach to image segmentation combines region growing and edge detection techniques: it groups the image pixels around the regional minima of the image and the boundaries of adjacent regions follow the crest lines dividing the influence zones of the minima. This transform is a powerful technique to partition an image into many regions while retaining edge information and it produces a complete division of the image in separated regions even if the contrast is poor, thus avoiding the need for any kind of contour joining.

Several algorithms have been proposed for the computation of watershed transform applied to images [Vincent 91, Beucher 93, Moga 97, De Smet 99]. Yet, the application of watershed algorithms to an image is often disappointing: like many other methods, the watershed algorithm is sensitive to noise and local texture, and often the image is over-segmented into a large number of tiny regions due to the large number of minima within an image or its gradient. However, unlike other methods, which typically produce incorrect or displaced boundaries in the presence of noise, the watershed al-

gorithm usually produces extra boundaries. This is referred to as over-segmentation, which means that apart from the real boundaries, the algorithm also produces spurious boundaries due to noise. Even though small changes in the edge map values can re-route the flow of water producing different watersheds. This problem can be removed by pre-processing the image to reduce noise and using a good post-merging scheme. This can make the watershed algorithm robust and if combined with the right merging scheme it is a good choice for automatic and semi-automatic segmentation problems.

One of two different algorithms are generally used to implement watershed segmentation, namely immersion and rainfalling simulation. Each of these can be used to detect the segments in the image either directly or using morphological operators. We briefly review these approaches as follows.

Immersion watershed

In the flooding or immersion approach [Vincent 91], single pixel holes are pierced at each regional minimum of the activity image which is regarded as topographic landscape. When sinking the whole surface slowly into a lake water leaks through the holes, rising uniformly and globally across the image, and proceeds to fill each catchment basin. Then, in order to avoid water coming from different holes merge, virtual dams are built at places where the water coming from two different minima would merge (cf. Figure 4.10). When the image surface is completely flooded the virtual dams or watershed lines separate the catchment basins from one another and correspond to the boundaries of the regions.

Figure 4.10 illustrates the immersion simulation approach. Figure 4.10.a) shows a 1D function with five minima. Water rises in and fills the corresponding catchment basins, as in Figures 4.10.b)-c). When water in basins b_3 and b_4 begin to merge a dam is built to prevent this overflow of water. Similarly, the other watershed lines are constructed. The final result containing five segments is shown in Figure 4.10.d).

Rainfalling watershed

The original concept behind the watershed transform was rainfalling on a terrain and flowing down paths of steepest descent to local minima [Beucher 79]. If a drop of water were to fall on any point of the altitude surface, according to the law of gravitation, it would flow down to a lower altitude, along the steepest slope path, until it reaches a

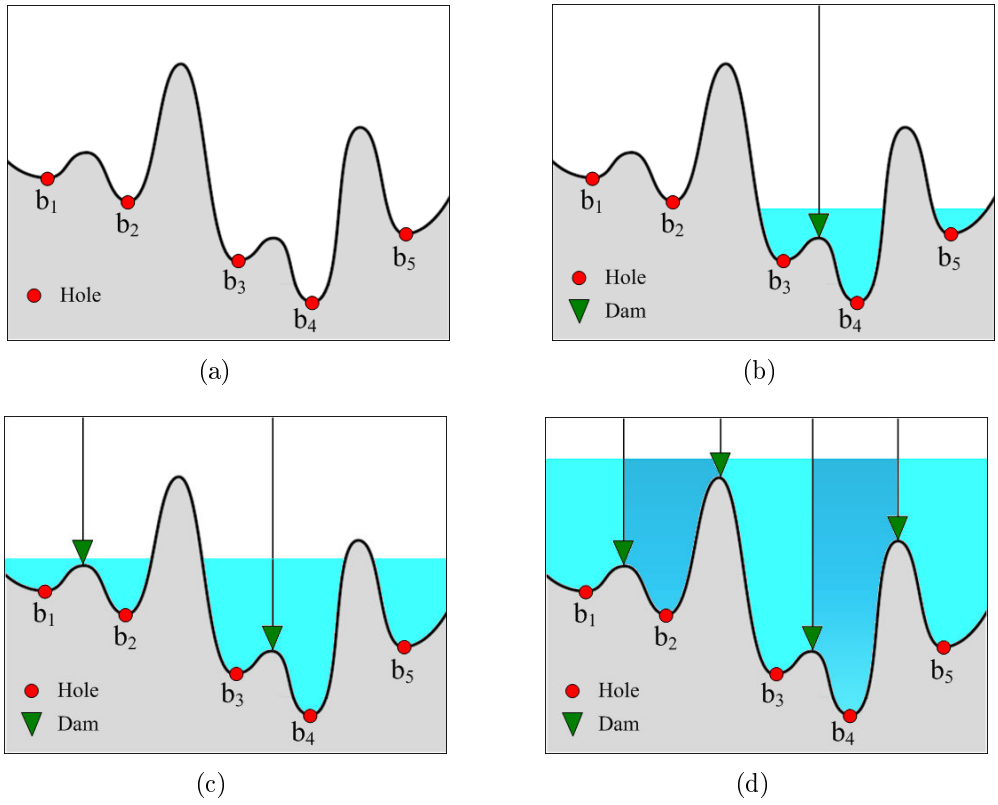
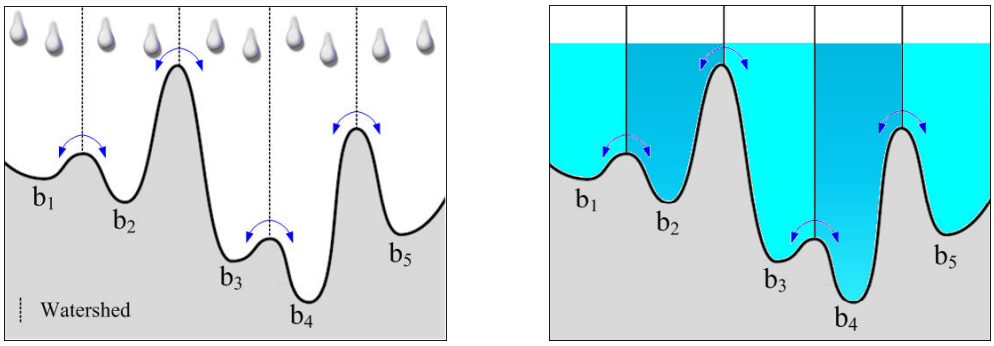


Figure 4.10: Illustration of immersion watershed transform on a continuous 1D function interpreted as a landscape. The landscape is sequentially flooded from bottom to top. a) Holes are pierced at each regional minimum. b) At certain flooding height there are two regions with one dam between basin b_3 and basin b_4. c) At intermediate flooding height there are three regions with two dams. d) Final segmentation with five segments.

point or region of minimum altitude. The accumulation of water in the neighbourhood of a minimum is called catchment basin. The whole set of points of the surface whose steepest slope paths reach a given minimum constitutes the catchment basin associated with this minimum, and all points that drain into a common catchment basin are part of the same watershed, in other words, watersheds are the borders between catchment basins. Thus, raindrops falling on both sides of a watershed line flow into different catchment basins. An illustration of a complete flooding process on a one-dimensional function is given by Figure 4.11 where five catchment basins are defined by the rainfalling simulation.

In the case of the rainfalling approach, every pixel can be traced to a minimum independent of the tracing of other pixels while in the immersion approach most pixels get their labels from a previously labelled neighbour.

The optimized implementation of the rainfalling method is two or three times faster



(a) (b)

Figure 4.11: Illustration of rainfalling watershed transform on a continuous 1D function interpreted as a landscape. a) Rainfalling process defines four top levels or dams. b) Final segmentation with the same five catchment basins as immersion watershed approach.

than the immersion method [De Smet 00]. Moreover, the rainfalling watershed treats the floating point type so that there is no round-off error in the implementation. Therefore, rainfalling-based watershed is more accurate than immersion-based.

4.5 Rainfalling watershed implementation

We propose a new implementation to the rainfalling watershed simulation in order to overcome some of the problems associated with watershed transform. To describe our implementation, we first define terms that are required to understand the working of the algorithm. We then discuss in detail our implementation of watershed segmentation by rainfalling simulation.

Let us consider a gradient image f whose domain is denoted as $D_f \subset \mathbb{R}^2$. Let $N_8(p)$ denote the neighbours of a pixel p in a $8-$*connectivity* grid.

Definition 1 (Regional minimum) *A pixel $p \in D_f$ is called a regional minimum if $\nexists q \in N_8(p)$ so that $f(q) < f(p)$.*

A regional minimum is a connected set of one or more pixels of similar value surrounded by pixels of higher value. In other words, a pixel belongs to a regional minimum if there is no descending path leading from it to another pixel with strictly lower value.

Definition 2 (Activity slope) *A pixel p is on an activity slope if $\forall p \in D_f$, $\exists q \in N_8(p)$ so that $f(q) < f(p)$.*

Definition 3 (Flat region) *A pixel p lies on a flat region with altitude h if $\exists q \in N_8(p)$ so that $h = f(q) = f(p)$*

A flat region is a smooth connected-component region of uniform gradient values from which it is impossible to reach a location of different altitude without having to descend or climb. A flat region can be classified into three types namely maximum plateau, plateau, and minimum flat region.

Definition 4 (Inner pixel) *A pixel p is an* inner pixel *of a flat region if $\forall q \in N_8(p)$ so that $f(q) = f(p)$.*

Definition 5 (Border pixel) *A pixel p is called a* border pixel $p \in B$ *of a flat region if p is on the flat region and it is not an inner pixel.*

Definition 6 (Indoor pixel) *A pixel p is an* indoor pixel *of a flat region if p is on the flat region and $\exists q \in N_8(p)$ so that $f(q) > f(p)$.*

Definition 7 (Outdoor pixel) *A pixel p is an* outdoor pixel *of a flat region if p is on the flat region and $\exists q \in N_8(p)$ so that $f(q) < f(p)$.*

Definition 8 (Maximum plateau region) *A flat region is called a* maximum plateau region *in D_f if $\forall q \in B$, q is an outdoor.*

Definition 9 (Plateau region) *A flat region is called a* plateau region *in D_f if $\exists p, q \in B$, so that p is an outdoor and q is an indoor.*

Definition 10 (Minimum flat region) *A flat region is called a* minimum flat region *in D_f if $\forall q \in B$, so that q is an indoor.*

Definition 11 (Catchment basin) *A pixel belongs to a catchment basin for a given regional minimum (RM) if one of the following three conditions are fulfilled:*

1. *The pixel is on a slope line which is connected to the RM or to an indoor pixel of the minimum flat region of RM.*
2. *The pixel is on the same flat region as the RM.*
3. *The pixel is on a activity slope line which is connected to one of the pixels fulfilling condition 2.*

The catchment basin of a regional minimum ρ_k is defined as the set of pixels that are topographically closed to ρ_k than to any other minimum.

Definition 12 (Watershed) *The boundaries between basins form the watersheds.*

Unlike standard watershed algorithms, the aim of the approach described in this section is to provide a strategy for watershed segmentation which does not require a pre-processing step in order to either sort all pixels of the input image [Vincent 91], to pre-compute the local minima from where the basins are flooded [Meyer 94], or to introduce a metric for plateau pixels [Moga 97].

4.5.1 Plateau regions analysis

Two problems arise when applying the watershed transform to an image. The first problem is the occurrence of plateau regions, i. e. regions of constant activity value as discussed in numerous publications [Gauch 99, Stoev 00, Roerdink 01]. The second problem, which is partly linked to the plateau region problem, is the dependency of the watershed location on both the used algorithm and the grid connectivity [Roerdink 01].

A pixel is said to be part of a plateau region if its value is equal to the value of at least one of its 8-neighbouring pixels in the activity image and its value is over the pre-flooding threshold. In our work a plateau region belongs to a unique catchment basin and a catchment basin has at most only one plateau region, as we will see below.

Conventionally, motion on a plateau surrounded by lower altitudes is oriented toward the closest downward outdoor of the plateau [Moga 97]. However, physical meaning of flat regions in intensity images [Vincent 91, Moga 97] is not the same as in gradient magnitude images [Gauch 99, Stoev 00]. Flat regions in intensity images correspond to uniform intensity regions of the image, while in gradient magnitude images, flat regions correspond to uniform variations of image intensity (ramps). Therefore,

the approach we use to analyse flat regions in rainfalling simulation has a different interpretation relying on the activity image used (intensity image or gradient image). To our knowledge this is the first time that this specificity is handled.

Moga and Gabbouj [Moga 97] described a parallel implementation for computing watershed transform based on rainfalling simulation. To deal with plateau regions, they transform the original image into a "lower complete image", i.e. an image where the only pixels without neighbours of lower altitude are the pixels of minima. In this lower image the pixels belonging to a non-minimum plateau are labelled with the geodesic distance to the plateau's nearest outdoor. Afterwards a raindrop starts at each pixel and its path toward the line with the steepest descent is followed until a regional minimum is reached.

Stoev and Strasser [Stoev 00] presented a sequential approach where every pixel p is compared with the adjacent pixels and if possible the path of steepest descent is followed and p is pushed on a stack S_c containing the pixels on the current path. Otherwise, if a flat region is reached, the whole flat region is processed in order to determine the nearest outdoor. If there are outdoors, the inner pixels are assigned to the appropriate outdoors and the path continues. They do not make any distinction between plateau regions and minimum flat regions, so it does not detect ramps in intensity images.

Gauch [Gauch 99] avoided flat region problems by working with Gaussian smoothed floating point images. This removes all regions with uniform intensity. However, this approach has several problems: if the neighbours of an edge decrease in intensity rapidly on the left and gradually on the right, the detected location of the edge will be to the right of the correct position; in very smoothed images which have few intensity minima the tops of same ridge-like structures may be missed.

A characteristic of some rainfalling approaches [Gauch 99, Hernandez 00] is the predominance of edges along a 45° angle. This is due to the fact that they do not scale the neighbouring pixels in diagonal directions on the computation of steepest descent which produces higher values on those directions. It increases the tendency to follow 4-connected directions.

Classical rainfalling method pours water onto the terrain surface of the entire image many times [Gauch 99, Kim 02], thus requiring a long processing time to obtain a satisfactory segmented image. Moreover, if the water falls on a wide and flat surface,

the flow route to the lowest position becomes longer, and the processing time increases in proportion to the length of the flow route. Therefore, to solve such problems, plain regions corresponding to flat regions need to be excluded from the rainfall process.

In the next section we propose an improved approach that can increase the speed and overcome the main shortcoming of rainfalling watershed segmentation method - the flat regions. Our activity image is the magnitude gradient of an image which simplifies the detection of uniform intensity regions as they are represented by zeros on the gradient magnitude. The only plateaus are result of ramps in the image intensity which occur less times than uniform intensity regions.

The proposed method performs rainfall only within the regions of interest (ROI) in which a pixel shows variation in gradient magnitude (see Figure 4.12). The set of neighbour pixels with constant gradient magnitude, i.e. within a flat region, are desert regions where rain rarely falls or, to be more precise, where only a raindrop falls.

4.5.2 Water flow tracing

The regional minima are the points which define the bottoms of watersheds, so the goal here is to identify the drainage directions for each pixel in the image. By following the image gradient downhill from each point in the image, the set of points which drain to each *regional minimum* can be identified.

We smooth the input image with an anisotropic filter described below and convert it to a floating point image gradient to predict the direction of drainage in the image. This simplifies the process of identifying minima points and reduce the over-segmentation problem. The use of floating point gradient is quite important as it avoids the problem of quantize the activity image which would lead to a loss of information and accuracy.

The watershed approaches usually require a pre-computation of the input image in order to detect the minima pixels (lower complete image in [Moga 97]). Since the plateau computing can be performed only when it is reached, in our algorithm we avoid the pre-computation step by sequentially scan the input image only once. For each not yet labelled pixel, the gradient descent labelling can be implemented efficiently in a single pass through the image. Figure 4.12 presents an example of the search process to find the regional minimum in the 3D terrain surface of an image. The yellow textured region represents the desert region, while the other region represents the ROI.

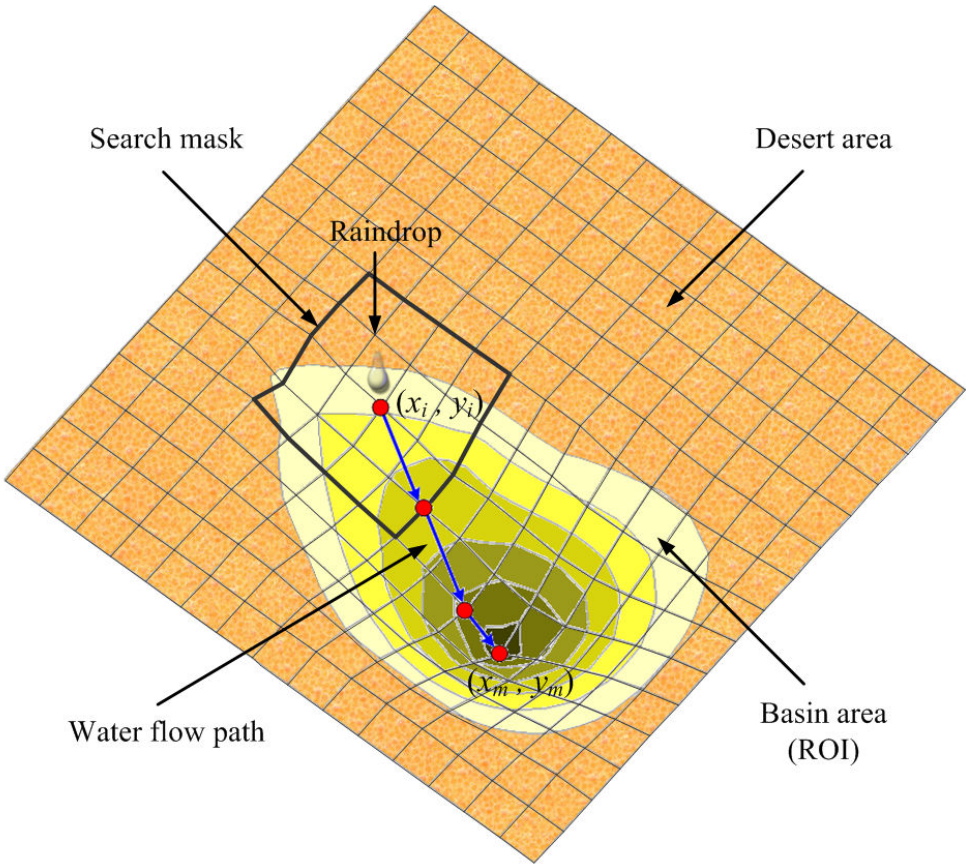


Figure 4.12: Example of water flow procedure using search mask (For a better illustration of the flow procedure, the search mask in the figure is 5×5).

Since we use a 3×3 search mask to compute the downhill search of the rainfalling watershed, to handle the image borders we build a one-pixel wide wall around the activity image and set the height to a value higher than the maximum value of the gradient image. This step is used to prevent water from leaking out of the surface.

A drop of water falls at (x_i, y_i) within the ROI, excluding the desert areas. The downhill or gradient descent direction of a pixel is then computed by examining its connected neighbours. Each pixel p is compared with its 8-neighbours and if it is on a steepest descent line to some pixel q, the value of p in the label is set to point to q. This search process is then repeated until the centre position of the mask has the lowest height. Hence, every time a regional minimum (x_m, y_m) is reached, the path setted in the direction of the predecessor q is traversed backwards and the pixels are labelled with the regional minimum's Id. Here, restricting the rainfall to ROIs reduces both the target region to be processed and the length of the flow route, thereby increasing the speed of the segmentation method based on the water flow model.

nwp	*np*	*nep*
wp	***cp***	*ep*
swp	*sp*	*sep*

Figure 4.13: The 3×3 search mask used in water flow trace (steepest descent).

In this step, the rainfalling concept is carried out by calculating the steepest descent direction for each pixel p. The directions are limited to the pixels neighbouring the central pixel ***cp*** of a 3×3 search mask, as shown in Figure 4.13, according to the following formula:

$$\begin{aligned}\text{steepest descent} = \min \Big\{ &(nwp - cp) \Big/ \sqrt{2}, (np - cp), (nep - cp) \Big/ \sqrt{2}, \\ &(wp - cp), (ep - cp), \\ &(swp - cp) \Big/ \sqrt{2}, (sp - cp), (sep - cp) \Big/ \sqrt{2} \Big\}\end{aligned}$$

At this time we present a new approach to handle the problem of plateau regions. If we assume that the pixel p, which has not yet been processed, is the next pixel on the path, five cases illustrated in Figure 4.14 can happen:

Case 1: p has no adjacent pixel with lower altitude, hence p is an isolated regional minimum;

Case 2: p has only one adjacent pixel q with lowest altitude. This is the regular case, where the algorithm follows the steepest descent path;

Case 3: p has adjacent pixels with the same altitude which means that p is an indoor pixel;

Case 4: p has at least one adjacent pixel with the same altitude and at least one lower pixel q which means that p is an outdoor pixel;

Case 5: p has more than one adjacent pixel with lowest altitude where q_1 and q_2 are non-adjacent pixels. In this case the algorithm cannot determine which of the adjacent pixels is the one the raindrop should flow to.

When **case 1** occurs, a regional minimum is reached and a new Id is assigned to the pixels on the path. In **case 2**, the pixel p is assigned to the path and if the lowest neighbour q is not marked yet, it is considered as the next processed pixel: $p \leftarrow q$. If

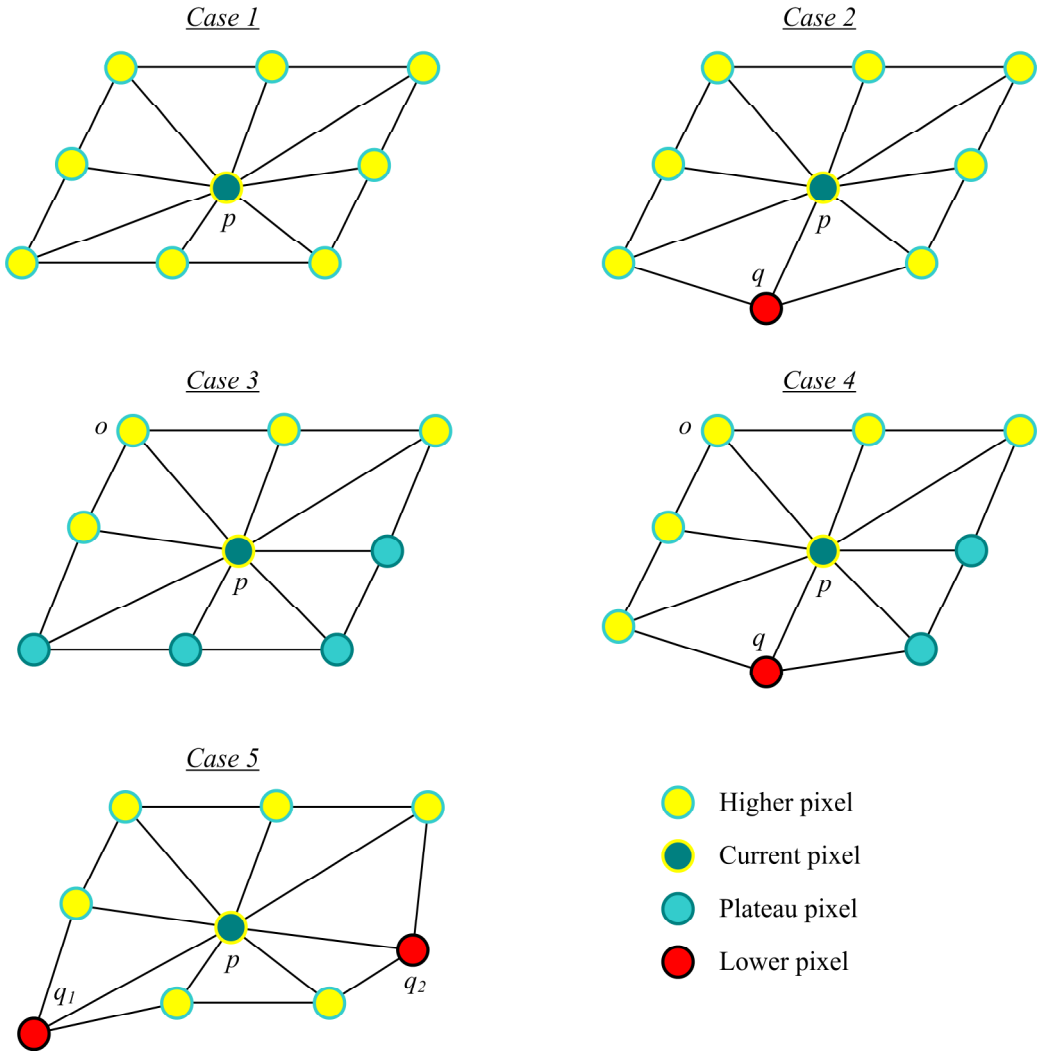


Figure 4.14: The five cases which can occur when the steepest descent path is followed.

q is already marked and it is an indoor pixel to a plateau region, the current path is terminated and its pixels are labelled with the label of p; if q is not an indoor pixel, the current path is also terminated and its pixels are labelled with the label of q. In **case 3**, if p is an indoor pixel to a plateau region the current path comes to an end and its pixels are labelled with the label of the pixel that precedes p in the path. Then the reached plateau has to be processed, since the steepest path cannot be unequivocally determined within plateaus. Thus, when a plateau is reached we label every pixel on the same plateau with the same label. We hold the location of indoor pixels to be used in cases 2 and 5. If p is an indoor pixel to a minimum flat region[3], the path is labelled with the label of p. The same label is assigned to all the pixels in the flat region. In

[3]A pixel is on a minimum flat region if its value is lower than the pre-flooding threshold.

case 4, the path is terminated and labelled with label of o; the plateau is labelled with p label and a new drop is put in q pixel which begins a new path. **Case 5** occurs when the pixel p is adjacent to m non-adjacent pixels $q_i, i = 1, .., m$ with the same altitude. In this case the algorithm cannot unequivocally decide which pixel should be processed next. In [Moga 97], the authors consider the first detected pixel with the lowest altitude as the next pixel to be processed which could produce erroneous results. In our approach all adjacent lowest pixels are traversed as if they were hit by a raindrop. After processing all q_i, the pixel with the lowest and nearest minimum is chosen to be the next processed one $p \leftarrow q_j$ and the path computation continues.

Since this approach is directed towards image segmentation, we put emphasis on the decomposition of an image into labelled regions or, in terms of the watershed transform, into catchment basins, whereas the extraction of watershed lines is not considered as an output of the algorithm. Our watershed produces a segmentation with zero-width watershed lines. This means that we assign to each pixel the label of the catchment (minimum) it belongs to so that the set of basins tessellates the image plane. Once all pixels in the image have been associated with their respective minima, the output image will contain the watershed regions of the image. We can simply locate the watershed lines by bounding the output image detecting changes in watershed region numbers.

4.6 Multiclass normalized cut

Although the pre-processing step serves to reduce the number of regions in the output of the watershed algorithm, it does not resolve the problem of over-segmentation. From our observation and testing it reveals that even when the small gradients are set to zeros, it could still cause over-segmentation. Generally there are two methods to reduce this over-segmentation. One is to use the markers [Grau 04, Levner 07] before the initial segmentation to extract the desired regional minimal to flood them. Although the markers work well for many types of images (especially medical images) their selection requires either explicitly prior knowledge of the image structure or careful user intervention. The other is to use some criteria to merge the regions produced by the initial segmentation. In our algorithm we use the latter method to produce the final segmentation. Thus we propose a spectral-based multiclass normalized cut approach to produce a meaningful segmentation.

Traditionally graph-based methods map an image onto a graph where nodes are composed of pixels and links are composed of connections between nodes. Each node has a weight based on some features and each link has a weight generally defined by the weight difference of the nodes it connects. The algorithm will group nodes or will cut the graph into connected regions [Shi 00] by link weight (reflecting similarity of pairs of nodes). It can be used without any supervision, and it does not require a learning phase. Graph-based segmentation takes into account global image properties as well as local spatial relationships and results in a region map that is ready for further processing, e.g. region labelling.

These methods have been applied in clustering and particularly in image segmentation. It is largely recognized that segmentation can be considered as a graph-partitioning problem; there are several approaches in the literature to solve this problem, including the spanning trees [Kwok 97], graph cuts [Shi 00], and the binary partition tree [Salembier 00].

There are different ways to measure the quality of a segmentation but in general we want the elements in a region to be similar and the elements in different regions to be dissimilar. This means that links between two nodes in the same region should have relatively low weights, and links between vertices in different regions should have higher weights. The normalized cut criterion balances the weight of the cut with the weights of the resulting regions.

The core computational technique of the normalized cut algorithm is a generalized eigenvalue problem. Although it is an elegant way to optimize the normalized cut criterion, the computational complexity of an eigenvalue decomposition is very high. In the original description of the normalized cut algorithm for image segmentation, one node corresponds to one pixel, so the number of nodes in the graph equals the number of pixels in the image.

Spectral methods use the eigenvectors and eigenvalues of a matrix derived from the pairwise similarities of pixels. The problem of image segmentation based on pairwise similarities can be formulated as a graph partitioning problem in the following way: consider the weighted undirected graph $G = (V, E, W)$ where each node $v_i \in V$ corresponds to a locally extracted image features, e.g. pixels and the links in E connect pairs of nodes. A weight $w_{i,j} \in \mathbb{R}_0^+$ is associated with each link based on some property of the pixels that it connects (e.g., the difference in intensity, colour, motion, location

or some other local attribute). Let $\Gamma = \{V_i\}_{i=1}^{k}$ be a multiclass disjoint partition of V such as $V = \cup_{i=1}^{k} V_i$ and $V_i \cap V_j = \emptyset, i \neq j$. Image segmentation is reduced to the problem of partitioning the set V into disjoint non-empty sets of nodes $(V_1, .., V_k)$, such similarity among nodes in V_i is high and similarity across V_i and V_j is low. The solution in measuring the goodness of the image partitioning is the minimization of the normalized cut as a generalized eigenvalue problem.

In order to reduce the number of nodes in the graph we replace the individual pixels by micro segments in a pre-processing stage. Image is decomposed into a number of atomic regions where each one is a vertex in the graph RSG. However, it is very important that the atomic regions will already yield a meaningful segmentation, i.e. the atomic regions must be homogeneous and the edges contained in the image must correspond to segment boundaries. Watershed segmentation is a classical and effective method for image segmentation in grey scale mathematical morphology that delivers these requirements. This method, in a wide perspective, has been applied successfully into some fields like remote sensing images processing of satellite and radar [Chen 04], biomedical applications [Grau 04] and computer vision [Kim 03].

Shi and Malik [Shi 00] introduced the normalized cut segmentation criterion for bipartitioning segmentation. Let V_A, V_B be two disjoint sets of the graph $V_A \cap V_B = \emptyset$. We define $links\,(V_A, V_B)$ to be the total weighted connections from V_A to V_B:

$$links\,(V_A, V_B) = \sum_{i \in V_A, j \in V_B} w_{i,j} \tag{4.11}$$

The intuition behind the normalized cut criterion is that not only we want a partition with small link cut but we also want the subgraphs formed between the matched nodes to be as dense as possible. This latter requirement is partially satisfied by introducing the normalizing denominators in the NCut equation. The normalized cut criterion for a bipartition of the graph is then defined as follows:

$$Ncut\,(A, B) = \frac{links\,(A, B)}{links\,(A, V)} + \frac{links\,(A, B)}{links\,(B, V)} \tag{4.12}$$

By minimizing this criterion we simultaneously minimize the similarity across partitions and maximize the similarity within partitions. This formulation allows us to decompose the problem into a sum of individual terms and formulate a dynamic pro-

gramming solution to the multiclass normalized cut ($kNCut$). So, the $NCut$ problem is naturally extended to a $kNCut$, finding a partition Γ that minimizes the function

$$kNCut\left(\Gamma\right) = \frac{links\left(V_1, \overline{V_1}\right)}{links\left(V_1, V\right)} + \frac{links\left(V_2, \overline{V_2}\right)}{links\left(V_2, V\right)} + ... + \frac{links\left(V_k, \overline{V_k}\right)}{links\left(V_k, V\right)} \tag{4.13}$$

where $\overline{V_i}$ represents the complement of V_i and $links\left(V_A, V_B\right) = \sum_{i \in V_A, j \in V_B} w_{i,j}$.

For a fixed k partitioning of the nodes of G, reorder the rows and columns of W accordingly so that

$$W = \begin{bmatrix} W_{11} & W_{12} & ... & W_{1k} \\ W_{21} & W_{22} & ... & W_{2k} \\ ... & ... & ... & ... \\ W_{k1} & W_{k2} & ... & W_{kk} \end{bmatrix} \tag{4.14}$$

and the rows of W correspond to the nodes in V_i. Let $D = diag\left(D_1, ..., D_k\right)$ be the $n \times n$ diagonal matrix so that D_i is given by the sum of the weights of all links on node i: $D_i = \sum_{j=1}^{k} W_{ij}$. It is easy to verify that

$$links\left(V_i, \overline{V_i}\right) = D_i - W_{ii} \qquad \text{and} \qquad links\left(V_i, V\right) = D_i \tag{4.15}$$

Barnes [Barnes 82] formulated the multiclass partitioning problem in terms of an indicator matrix. A multiclass partition of the nodes of G is represented by an $n \times k$ indicator matrix $\mathbf{X} = [\mathbf{x_1}, ..., \mathbf{x_k}]$ where $X\left(i, l\right) = 1$ if $i \in V_l$ and 0 otherwise. Since a node is assigned to one and only one partition there is an exclusion constraint between columns of X: $XI_k = I_n$. It follows that

$$links\left(V_i, \overline{V_i}\right) = \mathbf{x_i^T}\left(\mathbf{D} - \mathbf{W}\right)\mathbf{x_i} \qquad \text{and} \qquad \mathbf{links}\left(\mathbf{V_i}, \mathbf{V}\right) = \mathbf{x_i^T D x_i} \tag{4.16}$$

Therefore,

$$\begin{aligned} kNCut\left(\Gamma\right) &= \frac{\mathbf{x_1^T}\left(\mathbf{D} - \mathbf{W}\right)\mathbf{x_1}}{\mathbf{x_1^T D x_1}} + ... + \frac{\mathbf{x_k^T}\left(\mathbf{D} - \mathbf{W}\right)\mathbf{x_k}}{\mathbf{x_k^T D x_k}} \\ &= k - \left(\frac{\mathbf{x_1^T W x_1}}{\mathbf{x_1^T D x_1}} + ... + \frac{\mathbf{x_k^T W x_k}}{\mathbf{x_k^T D x_k}}\right) \end{aligned} \tag{4.17}$$

subject to $X^T DX = I_k$.

The solution for the generalized Rayleigh quotients that compose Equation (4.17) is the set of eigenvectors X associated with the set of the smallest eigenvalues $\Phi = \{0 = \nu_1 \leq ... \leq \nu_k\}$ of the system

$$(D - W)\, X = \Phi D X \tag{4.18}$$

However, this problem is NP-hard [Shi 00, Meila 01] and therefore generally intractable. If we ignore the fact that the elements of $\mathbf{x_i}$ are either zero or one, and allow them to take continuous values, by using the method of Lagrange multipliers as shown in [Chan 94], Equation (4.18) can be expressed by the standard eigenvalue problem. Let $\mathbf{y_i} = \mathbf{D^{1/2}x_i}$ and $Y = [\mathbf{y_1}, \mathbf{y_2}, ..., \mathbf{y_k}]$.

$$\widetilde{W} Y = Y \Lambda \tag{4.19}$$

subject to $Y^T Y = I_k$, where $\widetilde{W} = D^{-1/2} W D^{-1/2}$ is the normalized graph Laplacian matrix[4], with $\Lambda = \{1 = \lambda_1 \geq ... \geq \lambda_k\}$ where $\lambda_i = 1 - \nu_i$.

If Y is formed with any k eigenvectors of $\widetilde{W}$ then $\widetilde{W} Y = Y \Lambda$ where Λ is the $k \times k$ diagonal matrix formed with the eigenvalues corresponding to the k eigenvectors in Y. These k eigenvectors must be distinct to satisfy $Y^T Y = I_k$. This means that

$$Y^T \widetilde{W} Y = Y^T Y \Lambda = I_k \Lambda = \Lambda \tag{4.20}$$

and the *trace* of $Y^T \widetilde{W} Y$ is the sum of the eigenvalues corresponding to the k eigenvectors in Y. It follows that this sum is maximized by selecting the eigenvectors corresponding to the k largest eigenvalues of $\widetilde{W}$. So, Equation (4.17) becomes equivalent to

$$kNCut\,(\Gamma) = k - trace\left(Y^T \widetilde{W} Y\right) = k - \sum_{i=1}^{k} \lambda_i \tag{4.21}$$

Theorem 1 (Fan's Theorem [Fan 49]) *Let the eigenvalues λ_i of a symmetric matrix Q be so arranged that $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$. For any positive integer $k \leq n$, the sums $\sum_{i=1}^{k} \lambda_i$ and $\sum_{i=1}^{k} \lambda_{n+1-i}$ are respectively the maximum and minimum of $\sum_{j=1}^{k} y_j^T Q y_j$ when k orthonormal vectors $y_j (1 \leq j \leq k)$ vary in the space.*

[4] Although the Laplacian matrix is usually represented by $I - \widetilde{W}$, replacing $\widetilde{W}$ with $I - \widetilde{W}$ only changes the eigenvalues (from λ to $1 - \lambda$) and not the eigenvectors.

It follows from Fan's Theorem that the maximum on the right hand side of Equation (4.21) is achieved when Y is taken to by any orthonormal basis for the subspace spanned by the eigenvectors corresponding to the k largest eigenvalues of $\widetilde{W}$. From this we reach the following relaxed optimization problem

$$\min_{X^T D X = I_k} kNCut\left(\Gamma\right) = k - \max_{Y^T Y = I_k} trace\left(Y^T \widetilde{W} Y\right) \tag{4.22}$$

By putting together Fan's theorem with Equation (4.22) we establish a lower bound $l\left(k\right)$ on $kNCut\left(\Gamma\right)$ as

$$\min_{\Gamma} NCut_k\left(\Gamma\right) \geq k - \sum_{i=1}^{k} \lambda_i \tag{4.23}$$

where $\lambda_1, ..., \lambda_k$ are the k largest eigenvalues of $\widetilde{W}$. (For a proof see [Meila 01].)

For $k = 2$ the bound becomes $l\left(2\right) = 2 - \left(1 + \lambda_2\right) = 1 - \lambda_2 = \nu_2$ that is the second smallest eigenvalue of the generalized eigensystem of Equation (4.18). This is consistent with the bi-partitioning method proposed by Shi and Malik [Shi 00].

The core computational technique of the normalized cut algorithm is the eigenvalue problem Equation (4.27). It requires the solution to a large sparse system of symmetric equations. The LANCZOS algorithm [Scott 87] provides an excellent method for approximating the eigenvectors corresponding to the smallest or the largest eigenvalues of a sparse matrix with a time complexity of $O\left(n^{3/2}k\right)$ where n is the dimension of the matrix and k the number of eigenvectors.

4.6.1 Multiclass NCut in a random walk view

The Markov chain describing the sequence of nodes visited by a random walker is called a random walk on a weighted graph. We associate a random variable, s_t, representing the state of the Markov chain to every node in a step t; If the random walker is in state i at time t, we say $s_t = i$.

We define a random walk by the following single-step transition probability $p_{i,j}$ that represents the probability of jumping from a node i to a node j in one step, given that we are in node i, which is proportional to the weight $w_{i,j}$ of the link connecting nodes i and j: $p_{i,j} = Pr\left[s_{t+1} = j | s_t = i\right] = w_{i,j}/d_i$, where d_i is the degree of node i, given by the sum of links connecting node i to all the nodes.

The kNCut criterion can also be understood in the Markov random walk framework. Let $V_A, V_B \in V$. We define $P_{V_A,V_B} = \Pr[V_A \rightarrow V_B | V_A]$ as the probability of the random walk going from set V_A to set V_B in one step if the current state is in V_A.

$$P_{V_A,V_B} = \frac{\sum_{i \in V_A, j \in V_B} w_{i,j}}{\sum_{j \in V} w_{i,j}} = \frac{links(V_A, V_B)}{links(V_A, V)} \tag{4.24}$$

From this and from Equation (4.17) we express Equation (4.13) as:

$$kNCut(\Gamma) = k - \sum_{i=1}^{k} P_{V_k V_k} \tag{4.25}$$

The stochastic transition matrix P is obtained by normalizing the similarity matrix in order to the rows sums be all 1 (the degree matrix of P is the identity matrix).

$$P = D^{-1}W \tag{4.26}$$

The NCut is strongly related to the concept of low conductivity sets in the Markov random walk [Meila 01]. Minimizing the NCut for the bipartition V_A, V_B means that the probabilities of evading set V_A, once the walk is in it and of evading V_B are both minimized.

The relationship between the Laplacian matrix $\widetilde{W}$ and the Markov random walk transition matrix P was presented by Meila and Shi [Meila 01]. Equation (4.19) can be transformed into a standard eigenvalue problem of,

$$PZ = \Lambda Z \tag{4.27}$$

where the eigenvectors of P are related with the eigenvectors of $\widetilde{W}$ by $Z = D^{-1/2}Y$.

Since D is diagonal this means that the i-th row of Y is the same as the i-th row of Z scaled by $D_i^{1/2}$. So after the rows of Z are normalized to length 1, the optimal solution obtained from Z is identical to the solution obtained from Y.

$Z = [z_1, ..., z_k]$ is an $n \times k$ matrix formed by stacking the k largest eigenvectors of the eigensystem from Equation (4.27) in columns. The continuous solution $\widetilde{X}$ is obtained from Z by renormalizing each of Z's rows to have a unit norm.

$$\widetilde{X} = Z\left(Z^T Z\right)^{-1/2} \tag{4.28}$$

Recovering a discrete solution X from the continuous solution $\widetilde{X}$ is however a complex task. To overcome this problem, a majority of the theoretical work on spectral methods have dealt with successive bi-partitioning generating 2^k partitions [Shi 00].

4.6.2 Discrete partition

Due to the orthogonal invariance of the eigenvectors [Ng 02] any continuous solution can be replaced by a discrete solution $X = \widetilde{X}R$ for any orthogonal matrix $R \in \mathbb{R}^{k \times k}$. We can obtain this optimal discrete solution using the classical perturbation theory for matrix eigenvalue problems. In this work we follow a similar approach to the one presented by Yu and Shi in [Yu 03].

To discretize Z into X, we first normalize the rows of Z into $\widetilde{X}$ and then search for the rotation R that brings $\widetilde{X}$ the closest possible to a binary indicator vector X. The optimum discrete solution can be found iteratively. Given a continuous solution, we solve for its closest discrete partitioning solution; given a discrete solution, we solve for its closest continuous optimum. After convergence, X corresponds to a partitioning that is nearly globally optimal.

An optimal partition X should satisfy the following conditions:

$$\begin{aligned} &minimize\ \phi(X,R) = \left\| X - \widetilde{X}R \right\|^2 \\ &\text{subject to } X \in \{0,1\}^{n \times k},\ XI_k = I_n \\ &R^T R = I_k \end{aligned} \tag{4.29}$$

This can be solved by an iterative optimization process:

- Given R, we want to minimize $\phi(X) = \left\| X - \widetilde{X}R \right\|^2$. The optimal solution is given by non-maximum suppression:

$$X(i,m) = istrue\left(m = \arg\max\left[\widetilde{X}(i,k)\right]\right),\ i \in V \tag{4.30}$$

We let the first cluster centroid to be given by the row of the continuous solution $\widetilde{X}$ corresponding to the row of Z with the maximum sum, and then repeatedly choose as the next centroid the row of $\widetilde{X}$ that is closest to being 90^0 from all the centroids already picked.

- Given X, we want to minimize $\phi(R) = \left\| X - \widetilde{X}R \right\|^2$. The solution is given by singular value decomposition (SVD) diagonalization:

$$\begin{aligned} U \cdot \Omega \cdot V &= X^T \widetilde{X} \\ R &= VU^T \end{aligned} \tag{4.31}$$

  where U and V are $k \times k$ orthonormal matrices, $U^T U = V^T V = I_k$, and Ω is a $k \times k$ matrix that contains the singular values of $X^T \widetilde{X}$ in decreasing order on its diagonal and it is equal to zero elsewhere.

  Since $\phi(R) = 2(n - trace(\Omega))$, the larger $trace(\Omega)$ is the closer X is to $\widetilde{X}R$.

Such iterations monotonously decrease the distance between the continuous optimum and the discrete solution.

Figure 4.15 shows a comparison between continuous and discretized eigenvectors. Although there is correct information in the continuous solution, it could be very hard to split the pixels into segments.

4.7 Region similarity graph

Spectral-based methods use the eigenvectors and eigenvalues of a matrix derived from the pairwise similarities of features (pixels or regions). This effect is achieved by constructing a fully connected graph.

Based on the graph construction, there are two main groups of methods for image segmentation: region-based methods where each node represents a set of connected pixels, and pixel-based methods where each node corresponds to a pixel of the image. Region-based methods are usually modelled by a region adjacency graph (RAG). However, in the merging process these methods take into account only local information. Pixel-based methods construct an undirected weighted graph, taking each pixel as a node and connecting each pair of pixels with a weighted link. This reflects the likelihood that these two pixels belong to the same object. In these methods segmentation criteria are based on global similarity measures. In general, these methods are based on the partition of the graph by optimizing some cut value instead of merging the most similar adjacent regions.

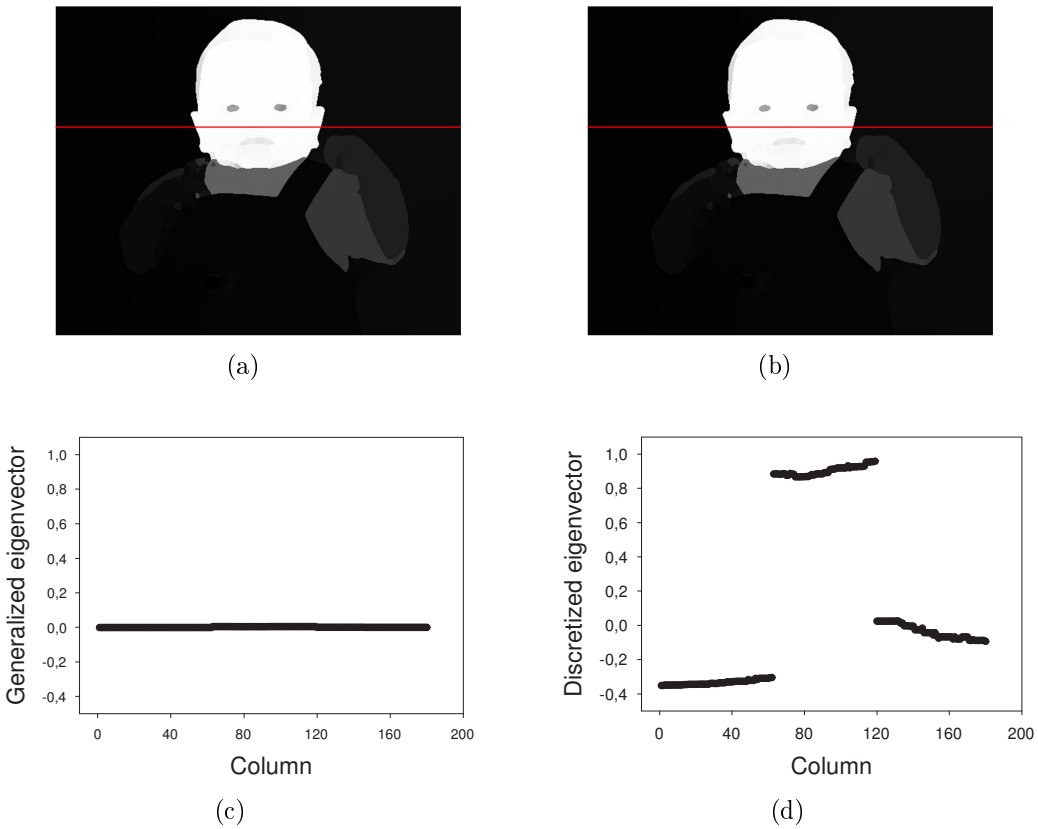


Figure 4.15: Continuous vs. discrete eigenvectors: (a) A generalized continuous eigenvector. (b) The discrete solution of the same eigenvector. (c)-(d) Graphic representation of values from the red rows in the images.

Considering all pairwise pixel relations in an image may be too computational expensive. Unlike other famous clustering methods [Shi 00, Yu 03] which use all pixels to construct the graph, our method is based on selecting links from a region similarity graph where each node corresponds to an atomic region. We represent the over-segmented image by a weighted undirected graph $G = (V, E, W)$, called region similarity graph (RSG). The RSG is similar to the region adjacency graph (RAG) [Haris 98, Hernandez 00] but it allows the existence of links between pairs of non-adjacent regions.

The proposed RSG structure takes advantages of both, region and pixel-based representations. The set of nodes V corresponds to the over-segmented regions where nodes are represented by the centroid of each micro-region. The set of links E represent relationships between pairs of regions, and the link weights W represent similarity measures between pair of regions and they are defined taking into account the intensity difference between regions and the maximum amount of gradient in the line connecting

the regions centroids (intervening contours). Figure 4.16 shows a synthetic image and its corresponding RAG and RSG.

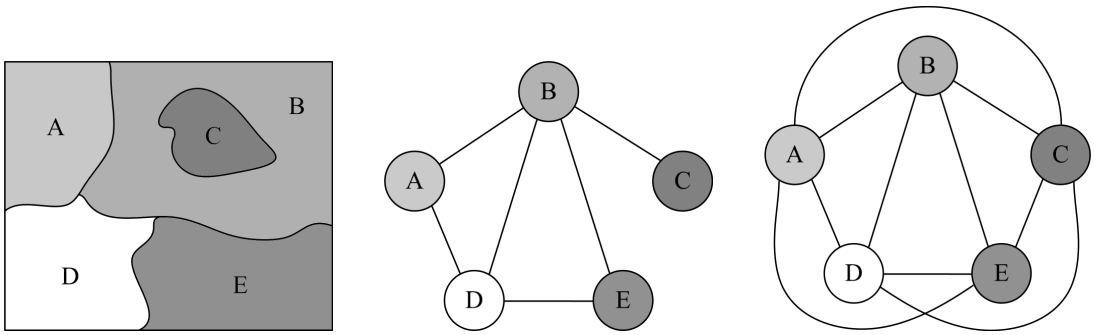


Figure 4.16: (a) Original image. (b) Corresponding RAG. (c) RSG with links between non-adjacent regions.

Some characteristics of the RSG model that yield to some relevant advantages with regard to the RAG model are:

- It is defined once and it does not need any dynamic updating when merging regions. Merging two regions in a RAG structure requires a considerable amount of processing to update RAG to reflect changes generated by the merging. It requires identity updating for every pixel in the merged region, as well as every region adjacent to those two regions.
- The segmentation, formulated as a *not necessarily* adjacent graph partition problem, leads to the fact that extracted objects are not necessarily connected.

4.7.1 Pairwise spatial similarity

The quality of a segmentation based on a RSG depends fundamentally on the link weights (similarity) that are provided as input. The weights should be large for nodes that belong to the same group and small otherwise. Using the micro-regions obtained in the pre-segmentation step as graph nodes, the corresponding weight function $W \in \mathbb{R}_0^+$ is defined assigning each link with the similarity between two nodes.

This weighted graph depends on external parameters that are related to the definition of similarity (which is task dependent) and to the transformation from perceptual similarity to link weight. Exponential decreasing function is supported psychophysically. It has been argued by Shepard [Shepard 87] that there is a robust psychological law that relates the distance between a pair of items in psychological space and the

probability that they will be perceived as similar. Specifically, this probability is a negative exponential function of the distance between the pair of items.

In the RSG model nodes are represented by the centroid of each region as a result of the initial over-segmentation. Links together with their associated weights are defined using the spatial similarity between nodes, their connectivity and the strength of intervening contours [Malik 01] between region centroids. The resulting graph is a structure where region nodes represent complete image regions.

For each pair of nodes, node similarity is inversely correlated with the maximum contour energy encountered along the path connecting the centroids of the regions. If there are strong links along a line connecting two centroids, these atomic regions probably belong to different segments and should be labelled as dissimilar. So, edge information can be integrated by reducing the pairwise similarity of such centroids. Let i and j be two atomic regions:

$$w_{ic}(i,j) = \exp\left[-\frac{\max_{t\in line(i,j)} \left\|OE^*\left(\overline{x}_i, \overline{x}_j\right)\right\|^2}{\sigma_{ic}^2}\right] \tag{4.32}$$

where $line(i,j)$ is the straight line between centroids $\overline{x}_i$ and $\overline{x}_j$.

The intensity distance between nodes contributes for the link weight according to the following function:

$$w_I(i,j) = \exp\left(-\frac{\left(I_{\overline{x}_i} - I_{\overline{x}_j}\right)^2}{\sigma_I^2}\right) \tag{4.33}$$

These cues are combined in a final link weight similarity function, with the values σ_{ic} and σ_I selected in order to maximize the dynamic range of W:

$$W(i,j) = w_{ic}(i,j) \cdot w_I(i,j) \tag{4.34}$$

In almost all the graph-based approaches proposed in the literature the spatial distance cue is also used to compute the similarity between graph nodes. However, during our experiments, we note that such cue is responsible for the partition of image homogeneous areas - an issue commonly associated to normalized cut algorithm. It is demonstrated by the common sense that if we consider two atomic regions belonging to the same homogeneous area but distant from each other if we decrease the similarity between nodes with spatial distance the probability that normalized cut will not merge

the two regions will increase. Thus, we decided not to use centroid spatial distance as a similarity cue. To this decision we take in attention the fact that intervening contours are equivalent to spatial distance without suffering from the same problems.

4.7.2 Implementation details of the RSG

For a computational consideration it is important to sort and label all the regions created by the watershed segmentation. In the following some implementation details are given about the construction of the RSG. For each region r_i, spatial location $\overline{x}_i$ is computed as centroids of their pixels. If the region is convex, the centroid is inside of it but if the region is concave, the centroid changes to the corresponding location of the nearest boundary pixel of that region. Two dynamic data structures are used through which it is very convenient to add or remove regions: 1) A label map in which each pixel value corresponds to the label of the segment that this pixel belongs to; 2) An array of segments where each segment is represented by a linked-list of pixels which correspond to the pixels that belong to the segment. This list includes the location and the grey-level of each pixel.

This dual representation of a partitioned image allows for a very efficient implementation. The label map grants us immediate access to the label of every pixel in the image. The array of lists gives us immediate access to the set of pixels that belong to each segment. Using this representation two different segments can be merged into one by iterating through the corresponding linked-lists and updating the label map. Even more, we can easily obtain the centroid and the mean value of each segment.

To compute the similarity matrix the current approach uses only image brightness and magnitude gradient. Additional features such as texture, could be added to the similarity criterion. This may slow the construction of the RSG but the rest of the algorithm will proceed with no change.

4.8 Hybrid segmentation framework

The algorithm described in this chapter can be well classified into the category of hybrid techniques (see section 4 of chapter 2), since it combines the edge-based, region-based, and the morphological techniques together through the spectral-based approach. Rather than considering our method as another segmentation algorithm, we propose

that our hybrid technique can be considered as an image segmentation framework within which existing image segmentation algorithms that produce over-segmentation may be used in the preliminary segmentation step.

To improve efficiency we introduce a graph cut formulation which is built on a pre-computed image over-segmentation instead of image pixels. In this framework graph G is not a necessarily adjacency graph with nodes being a set of atomic regions. We propose a powerful image segmentation algorithm by combining watershed transform and the multiclass spectral method to complement their strengths and weaknesses.

In most images there are usually large regions of pixels that belong to the same salient region and have only small interior intensity variations and they are thus easily identified. To combine these pixels into one region and to reduce the spatial resolution without losing important information we have decided to use a gradient watershed algorithm that provides over segmented but homogeneous regions with well located region boundaries. Since watershed segmentation provides a good set of object boundaries, this approximation produces reasonable results and improves the speed significantly.

The normalized cut and watershed approaches have complementary strengths:

- In the output of the watershed approach we have a reduced complexity representation. The dimension of the graph is far smaller when assigning nodes to atomic regions than to pixels, reducing the cluster computation.
- We have complete freedom in the choice of similarity function. This means that region interior as well as gradient information can be used. In particular atomic regions allow the comparison of distributions of feature vectors rather than single points as with the pixel based algorithms.
- Further, while the watershed depends fundamentally on local measurements of similarity (via the gradient function) region affinities can be calculated over the whole image, if desired, leading to a more global view of the similarity structure.

The combination of watershed and spectral methods solves the weaknesses of each method by using the watershed to provide small prototype regions from which similarity matrix could be obtained. Rather than clustering single feature points we will cluster micro-segments, confident that the underlying primitive segments are reliable. Our approach actually prefers the objects to be over-segmented into a number of smaller regions to ensure that a minimal amount of background is connected to any of the object

regions. The new criterion takes joint advantage of the two methods aiming at combining the best qualities of both segmentation approaches, giving a final segmentation that is more visually appropriated.

Preliminary segmentation by watershed transform is capable of producing atomic regions with complete and accurate boundaries, which can be considered as a good starting point for region merging. We present a new approach for locally applying a floating point based rainfalling simulation in a single image scan. In the second stage these atomic regions are used to construct a graph representation of the image, which is processed by a discrete multiclass normalized cut algorithm ($kNCut$). This combined framework results in a considerable speed-up of the entire algorithm.

A critical issue in watershed techniques is known to be over-segmentation i.e. the tendency to produce too many basins [Haris 98]. Several methods have been proposed in the literature to reduce the spurious boundaries created due to noise and produce a meaningful segmentation. Ogor [Ogor 95] proposes morphological opening and closing. Gauch [Gauch 99] uses Gaussian blurring. Hernandez and Barner [Hernandez 00] suggest median filtering while De Smet et al. [De Smet 99] apply non-linear filtering by anisotropic diffusion.

In this work we provide three methods to overcome this problem. First, bilateral anisotropic filtering [Tomasi 98] can be applied to remove noise from the image. Secondly, some of the weakest edges are removed by a gradient minima suppression process known as *pre-flooding*. This concept uses a measure of depth of a certain basin. Prior to the transform, each catchment basin is flooded up to a certain height above its bottom, i.e. the lowest gradient magnitude and it can be thought as a flooding of the topographic image at a certain level (flooding level). This process will create a number of lakes grouping all the pixels that lie below the flooding level (see Figure 4.17). This step is useful in reducing the influence of noise and partly eliminates over-segmentation.

The third one, handles to control over-segmentation eliminating spurious tiny regions associated with uniform regions through a merging step. This eliminates tiny regions which have similar adjacent regions, while maintaining the accuracy of the partition. This stage is required to reduce the computational complexity in the graph partitioning. Another advantage of these steps is to prevent large homogeneous (flat) regions from being split in the graph-based segmentation (a common problem with balanced graph cut methods).

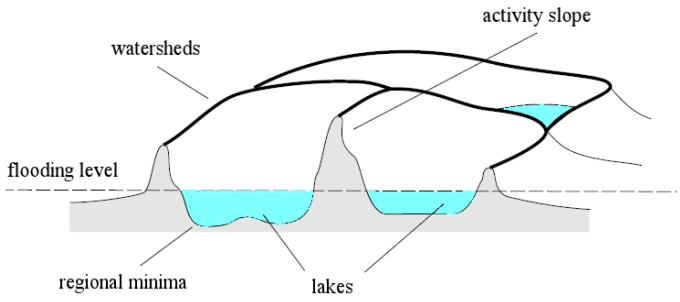


Figure 4.17: Pre-flooding process. Lakes are formed by merging neighbouring pixels below the flooding threshold.

Our approach to solve image segmentation as a graph partitioning problem is related to O'Callaghan and Bull [Callaghan 05] and De Bock et al. [De Bock 05] work. However, there are important differences between their works and ours: although De Bock et al. use a rainfalling watershed, it does not handle the problem of flat regions. Thus, when a raindrop falls in such kind of regions it forms a single region. It results in a larger number of atomic regions with dimension 1; O'Callaghan and Bull use an immersion-based watershed to compute initial segmentation; In the merging process De Bock et al. perform a bipartition normalized cut similar to the one presented in [Shi 00] and O'Callaghan and Bull use a weighted mean cut function for graph partitioning. It is also important to note that both schemes use a simple region adjacency graph structure to compute region similarity.

4.9 Summary

In this chapter we have proposed a new global image segmentation algorithm which combines edge- and region-based information with spectral techniques through the morphological algorithm of watersheds. A non-linear smoothing (bilateral filter) is used to reduce over-segmentation in the watershed algorithm while preserving the location of the image boundaries. The purpose of the pre-processing step is to reduce the spatial resolution without losing important image information. An initial partitioning of the image into primitive regions is set by applying a rainfalling watershed simulation on the image gradient magnitude. This step presents a new approach to overcome the problems with flat regions. This initial partition is the input to a computationally efficient region segmentation process (multiclass normalized cut algorithm) that produces

the final segmentation. The latter process uses a region similarity graph representation of the image regions.

To prevent large homogeneous regions from being split (a common problem of balanced graph based methods) we computed an over-segmentation of the image using the watershed technique. Clearly, large homogeneous regions are not partitioned into separate regions, unless there is a small amount of linking pixels between parts of the same region.

Using small atomic regions instead of pixels leads to a more natural image representation - the pixels are merely the result of the digital image discretization process and they do not occur in the real world. Besides producing smoother segmentations than pixel-based partitioning methods, it also reduces the computational cost in several orders of magnitude.

Any region-based segmentation algorithm which produces an over-segmented image can be used to extract the micro regions that will be combined based on the similarity function. So, our framework can easily integrate these algorithms and overcome their problems of over-segmentation in order to produce a better segmentation.

CHAPTER 5

Region-based motion segmentation: the model

This chapter describes an approach for integrating motion estimation and region clustering techniques with the purpose of obtaining precise multiple motion segmentations. Motivated by the good results obtained with the algorithm proposed in Chapter 4 we propose a hybrid approach where motion segmentation is achieved within a region-based clustering approach taken the initially result of a spatial pre-segmentation and extended to include motion information. Motion vectors are first estimated with a multiscale variational method applied directly over the input images and then refined by incorporating segmentation results into a region-based warping scheme. The complete algorithm facilitates obtaining spatially continuous segmentation maps which are closely related to actual object boundaries.

5.1 Introduction

Motion segmentation is basically defined as grouping pixels that are associated with a smooth and uniform motion profile. The segmentation of an image sequence based on motion is a problem that is loosely defined and ambiguous in certain ways. Though the definition says that regions with coherent motion are to be grouped, the resulting segments may not conform to meaningful object regions in the image.

The analysis of image motion and the processing of image sequences using motion information is becoming more and more important as video systems are finding an

increasing number of applications in the areas of coding, entertainment, robot vision, education, personal communications and multimedia. In video surveillance, segmentation can help to detect special events or to track objects over time. To reliably classify regions of an image sequence by their motion information is an important part of many computer vision systems. In video surveillance it is important to be able to detect which are the regions with movement. If one has detected a foreground object, further operations can be done on that object, such as recognition, identification or tracking. In robotics it is important to know which foreground objects are in order to properly interact with them. In video conferencing one wants to decide which objects are foreground and which ones are background to be able to encode the parts separately in order to save bandwidth, as the background needs to be transmitted only once.

Recent applications such as content-based image/video retrieval, like MPEG-7 [Chang 01], and image/video composition, require that the segmented objects are semantically meaningful. Indeed, the multimedia standard MPEG-4 [MPEG4 99] specifies that a video is composed of meaningful video objects. In order to obtain a content-based representation, an image sequence must be segmented into an appropriate set of semantically shaped objects or video object planes. Although the human visual system can easily distinguish semantic video objects, automatic video segmentation is one of the most challenging issues in the field of image processing.

Motion segmentation is closely related to two other problems, motion detection and motion estimation. Motion detection is a special case of motion segmentation with only two segments corresponding to moving *versus* stationary image regions (in the case of a stationary camera) or global *versus* local motion regions (in the case of a moving camera) [Dufaux 95]. In these cases, the motion profile of a pixel represents only the probability that a pixel is moving or not. When using stationary cameras, background subtraction is a particularly popular method to segment foreground and background. The idea behind background subtraction is to compare the current image with a reference image of the background, and from there decide what is background and what is not by looking for change at each pixel.

There is a strong interdependence between the definition of the spatial support of a region and of its motion estimation. On one hand, estimation of the motion information of the region depends on the region of support. Therefore, a careful segmentation of the regions is needed in order to estimate the motion accurately. On the other hand, a

moving region is characterized by coherent motion characteristics over its entire surface (assuming that only rigid motion is permitted). Therefore, an accurate estimation of the motion is required in order to obtain an accurate segmentation of the region.

All the motion estimation approaches assume that there is point correspondence between two consecutive frames which induces dense motion vector field of an image. No matter what method is used, at some stage we need a mechanism to assign each point to one of the recovered motions. This mechanism must take into account the smoothness of the world, i.e., the intuitive notion that the points belonging to the same motion are also spatially clustered in the image. This fact has been widely acknowledged in the literature on 2D motion segmentation [Shi 98, Cremers 05].

The estimation of an accurate motion field plays an important role in motion segmentation. However, general motion estimation algorithms often generate an inaccurate motion field mainly at the boundaries of moving objects, due to reasons such as noise, aperture problem, or occlusion. Therefore, segmentation based on motion alone results in segments with inaccurate boundaries.

In this chapter, a hybrid framework is proposed to integrate differential optical flow approach and region-based spatial segmentation approach to obtain for the accurate object motion. Our method adopts the variational optical flow approach of Brox et al. [Brox 04] in conjunction with several proposed techniques to convert the dense optical flow field to region-based motion field, with the suppression of noise and outliers.

Motion information will be initially represented through a dense motion vector field, i.e., it estimates which one best relate the position of each pixel in successive image frames. For the task at hand we adopt a high accuracy optical flow estimation based on a coarse-to-fine warping strategy [Brox 04] which can provide dense optical flow information. This method accelerates convergence by allowing global motion features to be detected immediately, but it also improves the accuracy of flow estimation because it provides better approximation of image gradients via warping. This technique is implemented within a multiresolution framework, allowing estimation of a wide range of displacements.

Handling spatial and temporal information in a unified approach is appealing as it could solve some of the well known problems in grouping schemes based on motion information alone [Wang 94, Weiss 97]. Brightness cues can help to segment untextured regions for which the motion cues are ambiguous and contour cues can impose sharp

boundaries where optical flow algorithms tend to extend along background regions. Graph based segmentation is an effective approach for cutting (separating) sets of nodes on a graph producing segmentation. As such, its extension to integrate motion information is just a matter of adding a proper similarity measure between nodes in the graph.

5.2 Previous work in motion segmentation

There is large literature on methods for segmenting from motion (see [Zhang 01a] for a comprehensive review on motion segmentation). The majority of the proposed approaches rely on the partition of each frame into solely two regions: one object and the background which could be too restrictive in some applications, e.g. coding.

A common class of methods for segmentation from motion is based on matching features points, such as corners or interest points. Since these systems process only a relatively sparse set of feature points, they are used to detect and track moving objects in a scene, rather than segmenting them with high resolution. Instead of matching feature points, some systems match small image blocks. Others, focusing on the simultaneous solution of motion estimation and segmentation assume a fixed number of regions and they are still more concerned with motion estimation for compression [Chang 97].

We can divide motion segmentation methods into the following three categories:

- Optical flow based segmentation.
- Simultaneous or sequential recovery of motion and segmentation.
- Fusion of motion estimation and static segmentation.

In the first approach, a dense optical flow field is recovered first and then segmentation is performed by fitting a model (often affine) to the computed flow field [Mémin 98]. Geometry of the scene can be used to combine this approach with a region growing approach. Reliable estimation of optical flow is difficult and separating the two processes causes errors to propagate from the first stage to the segmentation. The second approach attempts to solve the problems of the first one by doing simultaneous or sequential motion recovery and segmentation. In these techniques the segmentation is often formulated by using a Markov Random Field (MRF), which is a way of

incorporating spatial correlation into the segmentation process. The third approach aims to improve segmentation performance by using static segmentation based on the intensities of a single image to provide cues for the dynamic segmentation [Dufaux 95]. Gelgon and Bouthemy [Gelgon 95] used a region-level graph labelling approach to combine the static and dynamic segmentations. Since the support area for estimating the motion is chosen based on the static segmentation, biases in the motion estimation are likely to mislead the segmentation algorithm.

Other approaches to motion segmentation have been developed including the statistical model fitting algorithm of Bab-Hadiashar and Suter [B.-Hadiashar 98] and motion based segmentation techniques which do not use the dense motion estimation approaches are just outlined. For instance, Torr [Torr 95] proposed using the fundamental matrix for motion segmentation purposes. These algorithms are feature based and used a sparse set of features to identify the objects. Therefore, the number of data is relatively small.

Multibody factorization algorithms [Costeira 95] provide an elegant framework for segmentation based on the 3D motion of the object. These methods get as input a matrix that contains the location of a number of points in many frames and they use algebraic factorization techniques to calculate the segmentation of the points into objects, as well as the 3D structure and motion of each object. A major advantage of these approaches is that they explicitly use the full temporal trajectory of every point, therefore they are capable of segmenting objects whose motions cannot be distinguished using only two frames. Despite recent progress in multibody factorization algorithms, their performance is still far from satisfactory. In many sequences, for which the correct segmentation is easily apparent from a single frame, current algorithms that use only motion information often fail to reach this segmentation.

Most motion segmentation techniques handle the optical flow or just the image difference, as a precomputed feature that is provided to a standard segmentation method. In contrast to those methods, some more recent approaches propose to solve the problems of optical flow estimation and segmentation simultaneously [Mémin 02, Cremers 05, Brox 06a]. Cremers and Soatto introduced in [Cremers 05] the level set based motion competition technique. The optical flow is estimated separately for each region by a parametric model and the region contour is evolved directly by means of the fitting error of the optical flow. This idea has been adopted in [Brox 06a] where the

parametric model has been replaced by the better performing non-parametric optical flow model from [Brox 04]. A fundamental problem with the simultaneous segmentation and velocity estimation approach is that we typically need a segmentation in order to compute the motion model parameters and we need motion models in order to partition the image into regions.

When there is camera motion in video, segmenting or clustering motion is usually done by separating the objects (foreground) from the background. The use of normalized cuts for motion segmentation was introduced in [Shi 98], in which graph cutting techniques are used to obtain a motion related set of patches in the image sequence. The relationship between patches is defined on the basis of their motion similarity as well as their spatial and temporal proximity in the image sequence. The method is pixel-based, therefore it imposes a high computational overhead and thus, restricted to very small image sizes in order to minimize the graph cutting complexity. As a result it does not attempt to provide accurate shape recovery. Shi and Malik propose an approach to this problem which uses a sparse, approximate version of the similarity matrix in which each unit is connected only to a few of its nearby neighbours in space and time and all other connections are assumed to be zero.

The MPEG-4 video coding schemes use a block-based approach to motion estimation. The image is arbitrarily divided up into small blocks. For each block, a translational motion is estimated by making a search in the next frame for the most similar block. These systems are preferably used in the context of low-bit-rate video coding. This method again results in a rather crude segmentation with a resolution given by the block-size. However, the purpose of video coding is, in any case, compression rather than best representing the motion of the underlying object. Using regions instead of blocks provides more accuracy since block-wise motion does not fulfil real motion in the real world.

One of the earliest works on combining multiple features for segmentation is reported by Thompson [Thompson 80]. The image is segmented based on intensity and motion, by finding 4-connected regions that have similar intensity and optical flow values. The regions are then merged together using a variety of heuristics. Black [Black 92] presented an approach of combining intensity and motion for segmentation of image sequences based on Markov Random Fields (MRF). He uses three energy terms: intensity, boundary and motion. Tekalp et al. [Tekalp 98] presented a system

in which both colour and motion segmentation is done separately, followed by clustering the colour segments together that belong to the same motion segment. This assumes that the colour segments are more detailed, but nevertheless accurate, than the motion segments and they only need to be grouped together for correct segmentation.

In several approaches intensity is involved at pixel level through a spatial segmentation stage providing a set of regions that are handled by a region-based motion scheme. In [Ayer 95], a spatial segmentation stage is followed by a motion-based region-merging phase where regions are grouped by iterating estimation of the dominant motion and grouping of regions that conform to that motion. Tsaig and Averbuch [Tsaig 02] proposed a framework for automatic segmentation of moving objects with MRF model. They partitioned each frame into homogeneous regions by using watershed algorithm and constructed a region adjacency graph. They modelled MRFs on the graph and used the motion information to classify regions as foreground or background. By treating the region as an elementary unit for the MRF model, they efficiently reduced the computational complexity usually associated to MRFs. Although the method produce good results it was only applied to foreground-background motion segmentation.

Zeng and Gao [Zeng 04] followed the same framework with a solution to the occlusion problem. Occlusion has been an obstacle to estimate accurate motion vector. They detected occlusion region by forward and backward motion validation scheme and removed the potential misclassification of the uncovered background regions. In addition, region growing technique is used to improve the segmentation results.

Other methods involve, in contrast, motion-based intermediate regions or layers. The idea of segmenting an image into layers was introduced by Wang and Adelson [Wang 94] followed by Darrell and Pentland [Darrell 95]. In the paper of Wang and Adelson, affine model is fitted to blocks of optical flow, followed by a K-means clustering in motion parameter space. Motion segments are clustered in the layer extraction step of the algorithm to derive a set of layers that represent the dominant image motion. The affine model of each layer is refined based on its spatial extent. In the layer assignment step, a global cost function is optimized in order to improve the assignment of segments to layers. The algorithm, then, iterates the layer extraction and assignment steps until the costs would not be improved for a fixed number of iterations and returns the solution of lowest costs. The results presented are convincing, though the edges of segments are not very accurate, most likely due to the errors in the computation of

optical flow at occlusion boundaries. Darrell and Pentland use a robust estimation method to iteratively estimate the number of layers and the pixel assignments to each layer. They show examples with range images and with optical flow.

Smith et al. [Smith 01, Smith 04] have developed a Bayesian framework for segmentation of video sequence into ordered motion layers. Their approach is focused on the relationship between the edges in successive image frames.

Fowlkes et al. [Fowlkes 01] proposed a method for combining both static image cues and motion information considering all images in a video sequence as a space-time volume and attempt to partition this volume into regions that are coherent with respect to the various grouping cues. This approach is based on a technique for the numerical solution of eigenfunction problems known as the Nyström method. It exploits the fact that the number of coherent groups in an image sequence is considerably smaller than the number of units of volume. It does so by extrapolating the complete grouping solution using the solution to a much smaller problem based on a few random samples drawn from the image sequence.

5.3 Motion estimation

Motion segmentation schemes must also estimate, at some point in the process, the motion information in the scene. This section gives an overview of motion estimation process and the different approaches available.

Motion perception is an important cognitive element of the visual interpretation of our 3D world. In an ideal case, the movement of an object in 3D space corresponds to a 2D motion in an image sequence. These projected motions can be represented by a motion vector field in the image plane. The estimation of motion from image sequences has a long tradition in computer vision where accurate techniques for estimating the velocity field (optical flow field) are indispensable components. All work on image sequences begins by trying to find out how the image changes with time, analysing how different elements in the frame move.

Horn and Schunck [Horn 81] defined the optical flow as a velocity field in the image sequence which transforms one image into the next. In other words, the motion vector field is defined as the set of motion vectors that are used to denote the relative displacement of the image intensity values in a time-varying image sequence.

The estimation of optical flow relies on the assumption that objects in an image sequence may change position but their appearance remains the same (or nearly the same). Classically this is represented by the grey-level constancy assumption or the optical flow constraint [Horn 81, Lucas 81]. However, this assumption by itself is not sufficient for optical flow estimation. Horn and Schunck [Horn 81] add a smoothness assumption to regularize the flow, and Lucas and Kanade [Lucas 81] assume constant motion in small windows. Higher accuracy can be achieved using coarse-to-fine and/or warping methods [Black 96, Brox 04, Bruhn 05b]. These methods accelerate convergence by allowing global motion features to be detected immediately, but they also improve the accuracy of flow estimation because they provide a better approximation of the image gradients via warping [Brox 04].

From the scope of the used technique, motion estimation can be categorized into the following classes: non-parametric block-based [MPEG4 99], parametric motion model-based [Ayer 95, Torr 95, Black 96, Weiss 97], and gradient-based approaches [Horn 81, Lucas 81, Brox 04, Bruhn 05b]. All of these approaches assume that there is point correspondence between two consecutive frames which induces dense motion vector field of an image.

Block-based motion matching has been adopted in the international standards for digital video coding algorithms such as H.264 and MPEG-4. They operate by matching specific "features" (e.g., small blocks) from one frame to the next one. The matching criterion is usually a normalized correlation measure, typically by analysing the correlation in the feature neighbourhood. Block matching assumes that the motion field is piecewise translation. The current frame is broken up into blocks of equal size and for each block in the frame, the best match in the reference frame is computed within a certain neighbourhood.

Because of its simplicity, fast computation and relative robustness in visual effect, it is one of the most commonly used motion estimation methods even used as an intermediate stage in some pixel-based approaches. The weakness of the non-parametric block-based method is its inability to describe rotations and deformations, and the possibility of obtaining motion vectors that completely differ from the "true" motion. Additionally, a block-based scheme only provides a coarse motion field which is insufficient for motion segmentation.

Parametric estimation techniques (known also as *feature-based* methods) assume

that the motion in the scene (optical flow) can be described as a geometric transformation, i.e. affine or perspective transformation. Thus, rather than estimating the flow field, these techniques directly estimate the parameters of the motion model. In most cases, however, the motion between successive frames cannot be described as a single geometric transformation, due to presence of independently moving objects thus the scene is usually decomposed into several regions, each exhibiting a coherent motion, to where the motion parameters are then estimated.

The focus of this thesis is on gradient-based or differential methods (known also as *pixel-based* methods), in which the most recent progress has been made. These methods have the advantage that they do not have to find feature point correspondence. The motion vector field, or the so-called optical flow in gradient-based approaches, is estimated from the derivatives of image intensity over space and time and they are based on the assumption of data conservation (intensity and gradient). Due to the widely known aperture problem, additional assumptions are required to infer a particular 2D image velocity.

5.4 Optical flow

Optical flow is defined as the 2-D vector field that matches a pixel in one image to the warped pixel in the other image. In other words, optical flow estimation tries to assign to each pixel of the current frame a two-component velocity vector indicating the position of the same pixel in the reference frame.

Given two successive images of a sequence $I(x, y, t)$ and $I(x, y, t+1)$ we seek at each pixel $\mathbf{x} := (x, y, t)^{\mathrm{T}}$ the optical flow vector $\mathbf{v}(\mathbf{x}) := (v_x, v_y, 1)^{\mathrm{T}}$ that describes the motion of the pixel at $\mathbf{x}$ to its new location $(x + v_x, y + v_y, t + 1)$ in the next frame.

Estimating optical flow involves the solution of a correspondence problem. That is, what pixel in one frame corresponds to what pixel in the other frame. In order to find these correspondences one needs to define some property or quantity that it is not affected by the displacement. Many differential methods for optical flow are based on the assumption that the image intensity remains unchanged along motion trajectories (brightness constancy constraint) [Lucas 81]:

$$I(x, y, t) = I(x + v_x, y + v_y, t + 1) \tag{5.1}$$

The brightness constancy assumption requires that the grey value of a pixel does not change as it undergoes motion. It is customary to accommodate for this sensitivity to noise by pre-blurring the image or equivalently by using weighted windows around each pixel. In the following, we will assume that the intensity of a moving point remains constant throughout time. Expanding the total differential into partial derivatives gives a relation between the spatial image gradient and the homogeneous velocity vector, known as *optical flow constraint*:

$$I_x \cdot v_x + I_y \cdot v_y + I_t = 0 \tag{5.2}$$

as it has been formulated in the classical algorithms of [Horn 81, Lucas 81]. I_* denote partial derivatives where I_x and I_y are the spatial derivatives of image brightness, and I_t is the difference between the image sequences. It must be noted that this linearisation is only valid under the assumption that the image changes linearly along the displacement which, in general, is not the case especially for large displacements.

Obviously, this single equation is not sufficient to uniquely compute the two unknowns v_x and v_y. This issue is commonly referred to as *aperture problem*. For non-vanishing image gradients it is only possible to determine the flow component perpendicular to the image gradient. It is also clear that Equation (5.2) is only well defined in areas of the image with high gradient and then it is the results from these areas that must then be spread into the other areas of the image. In motion estimation this is typically resolved either by smoothing or by parameterising the motion.

Besides prior information on the flow magnitude, the work of Weiss and Adelson [Weiss 97] suggests that humans also use prior information about the smoothness of optical flow. In a non-rigid motion, although each pixel of an image can move freely, the motion is assumed to be locally coherent. The optical flow field undergoes two forces, one that matches the warped image with the original image and the other that keeps the optical flow field smooth.

Consequently, a second assumption is needed that is capable to provide a unique solution of the flow vector. There are two popular possibilities: local and global methods. The first one was proposed by Lucas and Kanade [Lucas 81] and assumes that the optical flow can be described by a parametric model in a local neighbourhood, which is in the simplest case the model of constant flow. This allows to locally compute

the optical flow for each pixel ignoring the situation outside the local neighbourhood. The other class of techniques is based on the work of Horn and Schunck [Horn 81] and assumes the optical flow field to be smooth. This induces a dependency of the flow vector at a pixel on the flow at all other pixels. Recently, some combined approaches have been proposed which tried to overcome the intrinsic problems to each of the two methods [Bruhn 05b].

5.4.1 Relevant literature

There are several motion estimation algorithms known in the literature. A complete survey describing the basic ideas behind the most important algorithms was presented in [Beauchemin 95], whereas the authors of [Barron 94] compare quantitatively the performance of various optical flow techniques.

Two seminal variational methods were proposed by Horn and Schunck [Horn 81] and by Lucas and Kanade [Lucas 81]. The Horn and Schunck optical flow algorithm [Horn 81] uses a global regularisation between a data term consisting of the motion constraint equation and a smoothness term constraining the velocity to vary smoothly everywhere. Lucas and Kanade [Lucas 81] assumed the velocity is constant in local neighbourhoods and formulate a least squares calculation of the velocity for each neighbourhood. Both of these methods are based on a least-squares criterion for the optical flow constraint, and some global or local smoothness assumption on the estimated flow field. In practice, flow fields are generally not smooth. The boundaries of moving objects will correspond to discontinuities in the motion field. At these discontinuities, the smoothness assumption is strongly violated. Yet, one cannot simply drop the regularisation term, since the problem of motion estimation is highly ill-posed. Ideally, one would like to enforce a regularity of the estimated motion field only in the areas corresponding to the different moving objects, allowing for discontinuities across the boundaries of objects. Yet this requires knowledge of the correct segmentation.

Many researchers have addressed this coupling of segmentation and motion estimation. Rather than first estimating local motion and subsequently segmenting or clustering regions with respect to the estimated motion [Wang 94] some researchers have proposed to model motion discontinuities implicitly by non-quadratic robust estimators [Nagel 86, Black 96, Mémin 98]. Others tackled the problem of segmenting motion by treating the problems of motion estimation in disjoint sets and optimization of the mo-

tion boundaries separately [Odobez 98, Paragios 00, Farnebäck 01]. Some approaches are based on Markov random field (MRF) formulations and optimization schemes such as stochastic relaxation by Gibbs sampling [Konrad 92], deterministic relaxation [Bouthemy 93], graph cuts [Shi 98], energy minimization via graph cuts [Boykov 01b] or expectation-maximization (EM) [Weiss 97]. As pointed out in [Weiss 97], exact solutions to the EM algorithm are computationally expensive and therefore suboptimal approximations are employed.

Ju et al. [Ju 96] proposed a "Skin and Bones" model to compute optical flow using an affine flow model with a smoothness constraint on the flow parameters to ensure continuity of motion between patches. They formulate the problem as an objective function with a data term that enforces the affine flow models within a patch and a prior term that enforces spatial smoothness between the estimated affine motions and those of neighbouring patches. Black and Anandan [Black 96] exploited locally adaptive parametric motion models to drive the optical flow estimation. Lai et al. [Lai 05] proposed a gradient-based regularisation method that includes a contour-based motion constraint equation that enforced only at zero-crossing. Farnebäck algorithm [Farnebäck 01] has three distinct components: estimation of spatio-temporal tensors, estimation of parametric motion models and simultaneous segmentation of the motion field. Mémin and Pérez [Mémin 98, Mémin 02] proposed a robust energy-based model for the incremental estimation of optical flow in a hierarchical piece-wise parametric minimization of an energy functional in regular or adaptive meshes at each hierarchical level from the coarsest to the finest levels. To increase precision as well as robustness against noise Bruhn et al. [Bruhn 05b] proposed a method that combines local and global methods, in particular, those of Horn-Schunck and Lucas-Kanade which forms the combined local-global (CLG) method. The data term in the Horn-Schunck regularisation is now replaced by the least squares Lucas-Kanade constraint.

Brox et al. [Brox 04] proposed a variational method that combines a brightness constancy assumption, a gradient constancy assumption and a discontinuity-preserving spatio-temporal smoothness constraint. In order to allow for large displacements, this technique implements a coarse-to-fine warping strategy. The results obtained with this method are among the best of all methods for optical flow estimation. Recently, Papenberg et al. [Papenberg 06] added a few additional constraints to this algorithm and got even better results.

5.4.2 Variational methods

Differential methods, and in particular variational methods based on the early approach of Horn and Schunck [Horn 81] are among the best performing techniques for computing the optical flow [Brox 04, Bruhn 05a, Papenberg 06]. Such methods determine the desired displacement field as the minimiser of a suitable energy functional, where variations[1] from model assumptions are penalised. In general, this energy functional consists of two terms: a data term that imposes temporal constancy on certain image features, e.g. on the grey value of objects, and a smoothness term that regularises the often non-unique (local) solution of the data term by an additional smoothness constraint. While the data term represents the assumption that certain image features do not change over time and thus allow for a retrieval of corresponding objects in subsequent frames, the smoothness term stands for the assumption that neighbouring pixels most probably belong to the same object and thus undergo a similar type of motion. Due to the smoothness constraint which propagates information from textured areas to nearby non-textured areas the resulting flow field is dense i.e. there is an optical flow estimate (vector) available for each pixel in the image.

A variational approach formulates some model assumptions $A_1, ..., A_m$ in terms of an energy functional [Brox 05]:

$$E\left(e_1\left(\mathbf{x}\right), ..., e_n\left(\mathbf{x}\right)\right) = \int_\Omega \left(A_1, ..., A_m\right) \mathrm{d}\mathbf{x} \tag{5.3}$$

and tries to find those functions $e_1, ..., e_n$ that minimize the energy, possibly by respecting additional constraints.

It is necessary to quantify the model assumptions by the so-called penaliser terms. Each penaliser induces a high energy for those cases where the model assumption is not fulfilled and a low energy otherwise. The theory of the calculus of variations provides a way how to minimize the energy functional. It leads to the so-called Euler-Lagrange equations, which have to be satisfied in a minimum. The Euler-Lagrange equations are partial differential equations. For sufficiently simple energy functionals, these Euler-Lagrange equations lead to a linear system of equations, which can be solved by well-founded and optimized numerical methods.

[1] This is where the term *variational method* comes from.

The combined variational approach differs from usual variational approaches by the use of a gradient constancy assumption. This assumption provides the method with the capability to yield good estimation results even in the presence of small local or global variations of illumination. Besides this, the combination of non-linearised constancy assumptions and a coarse-to-fine strategy yields a numerical scheme that provides a well founded theory for the very successful warping methods.

Given two successive images of a sequence $I(x, y, t)$ and $I(x, y, t+1)$, we aim to obtain the optical flow vector[2] $\mathbf{v} := (v_x, v_y)$ which gives the relative displacement between the pixels of the two images.

Pixels in areas of homogeneous intensity are ambiguous as they can appear similar under several different motions (optical flow constraint). Pixels in areas of high intensity gradient are also troublesome as slight errors in the motion estimate can yield pixel of a very different intensity, even under the correct motion.

Constancy assumptions on data

Estimating motion requires a solution to what pixel in one frame corresponds to what pixel in the other frame. In order to find these correspondences we need to define some assumptions that are not affected by the displacement.

- **Brightness constancy assumption**

  The common assumption is that the grey value of the pixel does not change as it undergoes motion:

  $$I(x, y, t) = I(x + v_x, y + v_y, t + 1) \tag{5.4}$$

  A first order Taylor series expansion leads this assumption to the well-known optical flow constraint of Equation (5.2).

  However, this constancy assumption cannot only deal with image sequences with either local or global change in illumination. In this case other assumptions that are invariant against brightness changes must be applied. Invariance can be ensured by considering spatial derivatives.

[2] In this thesis we represent the optical flow vector $\mathbf{v}(\mathbf{x}) := (v_x, v_y, 1)^{\mathrm{T}}$ by $\mathbf{v} := (v_x, v_y)$.

- **Gradient constancy assumption**

  A global change in illumination both shifts and/or scales the grey values of an image sequence [Papenberg 06]. Shifting the grey values will not affect the gradient. Although scaling the grey values changes the length of the gradient vector it does not affect its direction. Thus, we assume that the spatial gradients of an image sequence can be considered as constant during motion:

$$\nabla I\left(x, y, t\right) = \nabla I\left(x + v_x, y + v_y, t + 1\right) \tag{5.5}$$

  where $\nabla = (\partial x, \partial y)$ denotes spatial gradient. Although the gradient can slightly change due to changes in the grey value too, it is much less dependent on the illumination than on the brightness assumption.

Finding the flow field by minimizing the data term alone is an ill-posed problem since the optimum solution, especially in homogeneous areas, might be attained by many dissimilar displacement fields [Amiaz 07]. This is the aperture problem: the motion of a homogeneous contour is locally ambiguous. In order to solve this problem some regularisation is required. The most suitable regularisation assumption is *piecewise smoothness* [Brox 04], that arises in the common case of a scene that consists of semi-rigid objects.

The data term $E_D\left(v_x, v_y\right)$ incorporates the brightness constancy assumption, as well as the gradient constancy assumption. While the first data term models the assumption that the grey-level of objects is constant and does not change over time, the second one accommodates for slight changes in the illumination. This is achieved by assuming constancy of the spatial image gradient:

$$E_D\left(v_x, v_y\right) = \int_\Omega \psi\left(\left|I\left(\mathbf{x} + \mathbf{v}\right) - I\left(\mathbf{x}\right)\right|^2 + \gamma\left|\nabla I\left(\mathbf{x} + \mathbf{v}\right) - \nabla I\left(\mathbf{x}\right)\right|^2\right) \mathrm{d}\mathbf{x} \tag{5.6}$$

where Ω is the region of interest (the image) over which the minimization is done. The parameter γ relates the weight of the two constancy assumptions, and $\psi\left(s^2\right) = \sqrt{s^2 + \varepsilon^2}$ is a non-quadratic (convex) penaliser applied to both the data and the smoothness term which represents a smooth approximation of the L_1 norm, $L_1\left(s\right) = |s|$. Using the L_1 norm rather than the common L_2 norm reduces the influence of outliers and makes estimation robust. Due to the small positive constant ε, $\psi\left(s^2\right)$ is still convex

which offers advantages in the minimization process. The incorporation of the constant ε makes the approximation differentiable at $s = 0$; the value of ε sets the level of approximation which we choose to be 0.001.

Applying a non-quadratic function to the data term addresses problems at the boundaries of the image sequence, where occlusions occur and therefore outliers in the data compromise the correct estimation of the flow field.

Smoothness assumption

The smoothness assumption [Horn 81, Weiss 97, Brox 04] is motivated by the observation that it is reasonable to introduce a certain dependency between neighbouring pixels in order to deal with outliers caused by noise, occlusions or other local violations of the constancy assumption. This assumption states that disparity varies smoothly almost everywhere (except at depth boundaries). That means we can expect that the optical flow map is piecewise smooth and it follows some spatial coherency. This is achieved by penalising the total variation of the flow field. Smoothness is assumed by almost every correspondence algorithm. This assumption fails if there are thin fine-structured shapes (e.g. branches of a tree, hairs) in the scene.

Horn and Schunck proposed in their model the following smoothness (homogeneous) term [Horn 81]:

$$E_{S_{HS}}(v_x, v_y) = \int_\Omega |\nabla v_x|^2 + |\nabla v_y|^2 \mathrm{d}\mathbf{x} \tag{5.7}$$

However, such a smoothness assumption does not respect discontinuities in the flow field. In order to be able to capture also locally non-smooth motion it is necessary to allow outliers in the smoothness assumption. This can be achieved by the non-quadratic penaliser ψ also used in the data term. Thus, the smoothness term $E_S(v_x, v_y)$ becomes:

$$E_S(v_x, v_y) = \int_\Omega \psi\left(|\nabla v_x|^2 + |\nabla v_y|^2\right) \mathrm{d}\mathbf{x} \tag{5.8}$$

The smoothness term gives a penalty to adjacent segments which have different motion parameters.

Xiao et al. [Xiao 06] proposed an adaptive bilateral filter to regularize the flow computation which is able to achieve the smoothly varied optical flow field with highly desirable motion discontinuities. This approach combines information from regions

with similar flow and similar intensities taking into account occlusions. The method produces very similar results with the Brox et al. approach [Brox 04].

Energy functional

Applying non-quadratic penaliser functions to both the data and the smoothness term and also integrating the gradient constancy assumption, results in the optical flow model described by the following energy functional:

$$E\left(v_x, v_y\right) = E_D\left(v_x, v_y\right) + \alpha E_S\left(v_x, v_y\right) \tag{5.9}$$

where α is some positive regularisation parameter which balances the data term E_d with the smoothness term E_s: Larger values for α result in a stronger penalisation of large flow gradients and lead to smoother flow fields.

The minimization of $E\left(v_x, v_y\right)$ is an iterative process, with external and internal iterations [Amiaz 07]. The external iterations are with respect to scale. The internal iterations are used to linearise the Euler–Lagrange equations and solve the resulting linear set of equations [Brox 04]. Linearisation via fixed-point iterations is used both in the external and internal loops. The linear equations are solved using successive over relaxation. We employ the technique proposed by Brox et al. [Brox 04] which is currently one of the most accurate optical flow estimation method available. The reader is referred to Thomas Brox's PhD thesis [Brox 05] for a solution to minimize this functional.

5.4.3 Multiscale approach

In the case of displacements that are larger than one pixel per frame, the cost function in a variational formulation must be expected to be multi-modal and the minimization algorithm could easily be trapped in a local minimum [Brox 04]. A good approximation for smoothing the energy functional is to smooth the underlying images. As the smoothing of the images removes small details that are responsible for local minima, we can expect that the energy functional containing the smoothed images has considerably less local minima.

Instead of costly smoothing operations on the originally sized images it is also possible to downsample the images in a pyramid framework. The multiscale coarse-to-

fine approach is used by most actual algorithms for optical flow estimation in order to support large motion and to improve accuracy [Brox 04, Bruhn 05b, Amiaz 07]. This removes small details the same way as a smoothing operation on the original image. Additionally, it leads to a much more efficient multiscale implementation. Thus, this procedure is chosen here. Figure 5.1 shows the multiscale warping scheme used in the optical flow estimation.

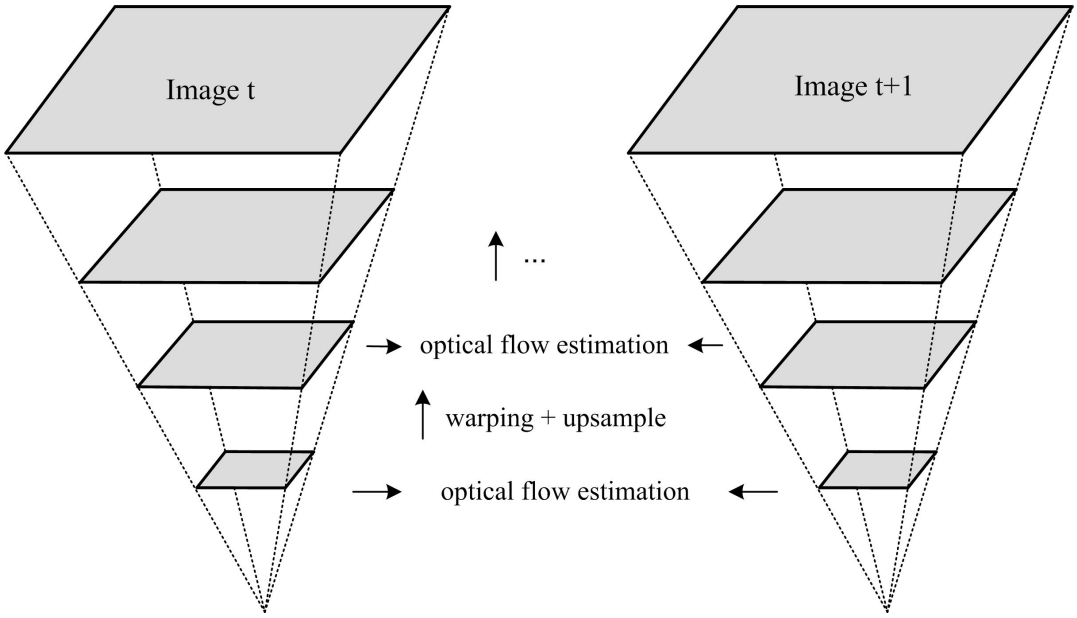


Figure 5.1: Coarse-to-fine optical flow estimation.

This approach relies on estimating the flow in a full pyramid of images, starting with the smallest possible image at coarsest scale and the upper levels are warped representations of the images based on the flow estimated at preceding scales. In the context of large displacements, the problem is compensated by the already computed motion from all coarser levels before the resolution is refined. What remains to be solved at each resolution level is the motion increment $d(v_x, v_y)$ for the difference problem. Such procedure allows to keep the displacements at each resolution level small, so that linearised constancy assumptions remain reasonable approximations. This ensures that the small motion assumption of Equation (5.2) remains valid.

Warping denotes the distortion of the image which is required for the compensation of the already computed motion. In general, it was argued that it makes sense to embed optical flow approaches for small displacements into a coarse-to-fine framework, since large displacements become smaller at coarser levels and thus allow for an accurate

estimation with linearised model assumptions. Each level in the pyramid can cause the initialization at a finer scale to be too close to a local minimum just appearing at that scale. Brox et al. [Brox 04] suggested to reduce this risk by making smaller steps. They proposed a downsampling factor $\eta \in (0, 1)$ between successive resolution levels in the pyramid, typically[3] $\eta \in [0.80, 0.95]$ which allows smooth flow projections between adjacent image levels in the pyramid. Though this high factor increases the computational cost it allows highly accurate optical flow computations.

5.4.4 Motion estimation analysis

The used optical flow estimation method has several positive properties that are important to our motion segmentation task:

- Due to non-linearised constancy assumptions the method can deal with larger displacements than most other techniques. This ensures a good estimation quality even when the object changes its location rapidly.
- It provides dense and smooth flow fields with subpixel accuracy due to the multiscale approach.
- The method is robust with respect to noise as shown in [Brox 04].
- By the introduction of the gradient constancy assumption it is fairly robust with regard to illumination changes that appear in most real-world image sequences.

For a qualitative evaluation and to a better visualization of the computed flow fields, we used a colour RGB representation shown in Figure 5.2. While the colour itself indicates the direction of the displacements, the brightness expresses their magnitude. Figure 5.3 shows how the individual model assumptions influence the quality of the computed optical flow. We used a real-world sequence (the *Dancing* sequence), where a person dances in front of the camera. Before we applied the different numerical schemes we preprocessed the sequence by convolution with a Gaussian kernel of standard deviation $\sigma = 1.0$. Starting from the classical local constraints approach (with no regularisation)

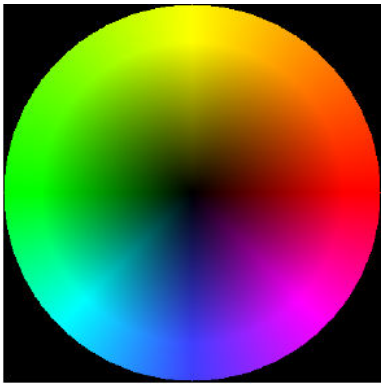


Figure 5.2: Flow colour code.

[3] This reduction factor is larger than the commonly used 0.5.

of Lucas and Kanade [Lucas 81], each extension of the optical flow model implies a significant improvement in the result.

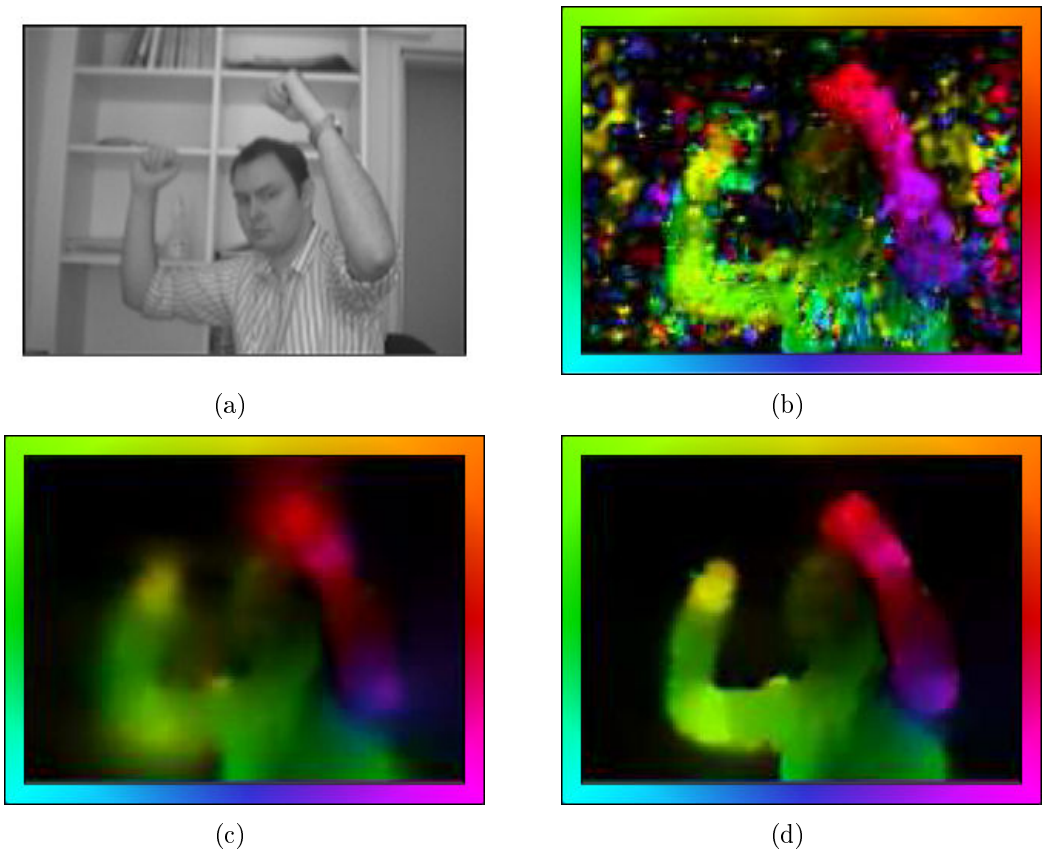


Figure 5.3: (a) One frame of *Dancing* sequence. (b) Computed flow field using only local constraints [Lucas 81]. (c) Computed flow field using homogeneous propagation of [Horn 81]. (d) Computed flow field using a non-quadratic regularisation term [Brox 04].

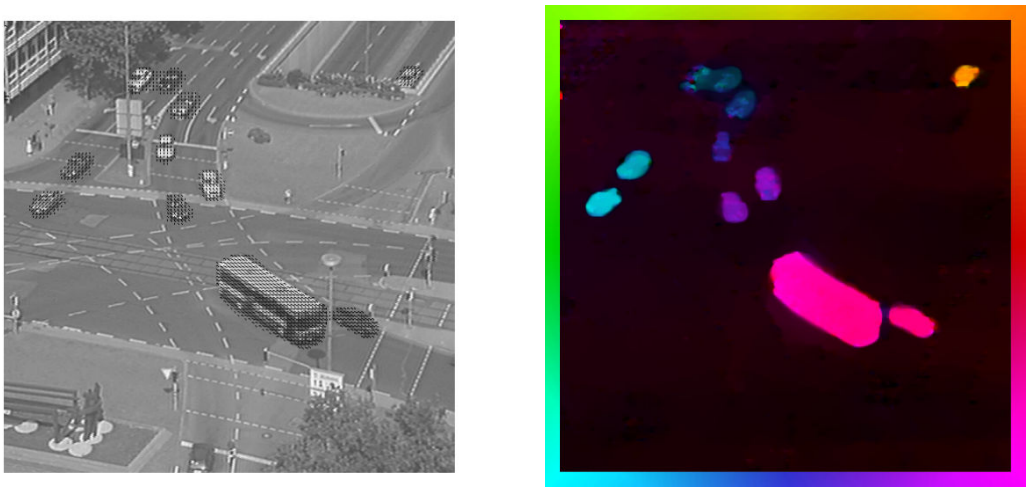
In a first step the introduction of the homogeneous propagation term of Horn and Schunk allows the model to have spatial coherency in the flow map by propagating the flow to homogeneous regions. However, this smoothness constraint does not respect discontinuities in the flow field producing over-smoothing on the flow. In the second step the incorporation of a non-quadratic smoothness term allows the model to capture the motion discontinuities more accurately. The non-quadratic regularisation term allows the propagation of information without crossing image and flow discontinuities.

In order to get a visual impression of the quality of the estimation[4] the *Ettlinger Tor* traffic sequence[5] is used. Figure 5.4 shows both the computed flow field between

[4]We used the implementation of Brox et al.'s algorithm which was available to us by courtesy of Thomas Brox. We would like to thank him for providing optical flow software.

[5]Available at *http://i2iwww.ira.uka.de/image_sequences/*.

frame 5 and 6 and its magnitude and orientation plot. As proposed in Barron et al. [Barron 94] we pre-processed each image sequence by convolution with a Gaussian kernel of standard deviation referred to as parameter σ.



(a) (b)

Figure 5.4: (a) Computed flow field between frame 5 and frame 6 of the *Ettlinger Tor* traffic sequence. (b) Magnitude and orientation of the flow field with $\sigma = 0.6$, $\alpha = 40$ and $\gamma = 20$.

Although the sequence suffers from interlacing artefacts the optical flow estimation algorithm gives very realistic results where the flow boundaries are relatively sharp. This is a direct consequence of using non-quadratic smoothing functions.

5.5 Building the region-based motion graph

Studies in motion analysis have shown that motion-based segmentation would benefit from including not only motion but also the intensity cue, particularly to retrieve region boundaries accurately [Dufaux 95, Weiss 96, Galun 05]. Hence, the knowledge of the spatial partition can improve the reliability of the motion-based segmentation.

We would like to identify prominent groups that follow the same motion structure. In order to do so, it is necessary to compute a measure of affinity between each region. Taking our cues from the Gestalt school, we consider brightness similarity, intervening contours and common fate. These sources of information should measure the likelihood that two regions R_i and R_j represent different parts of the same moving object. Such scheme requires the construction of a structure exploiting the motion information which represents the relationships among partitions and between successive image partitions.

This section focuses on this stage consisting in the introduction of a region-based motion graph representation. To this end a region-based contextual information has to be formalized and exploited. Figure 5.5 gives an overview of the scheme to construct the region-based motion graph.

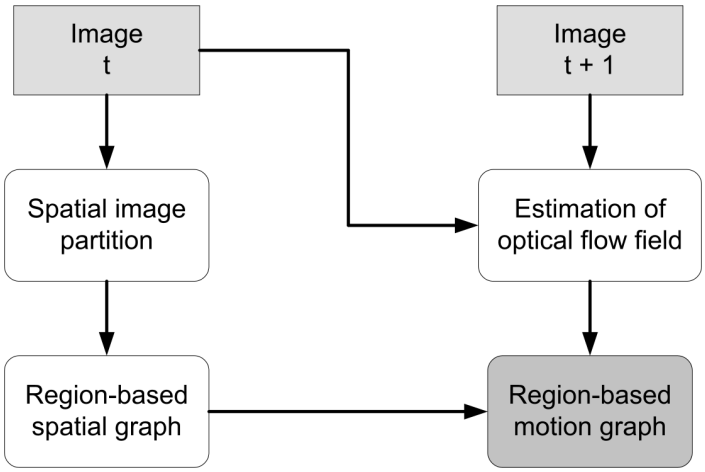


Figure 5.5: Diagram of the region-based motion graph construction.

A spatial partition of the first frame of the image sequence is first required by some over-segmentation process (e.g. watershed). A region-based spatial graph is then derived from the spatial image partition (Section 4.5). A 2D motion model is estimated within each region, and the optimal motion label configuration is sought for using an energy minimization approach, so that region undergoing similar (respective different) motion are given the same (respective different) labels.

We aim at assigning a motion vector to every node in the graph, with a view to partitioning this graph into node subsets, corresponding to groupings of regions of coherent motion. The predefined regions should be so that all pixels within a spatial atomic region were assigned the same motion label. It is generally true that motion boundaries coincide with intensity segment boundaries but not vice versa; i.e., intensity segments are almost always a subset of motion segments. Therefore, we can first perform an intensity segmentation to obtain a set of candidate motion segments. Then, those segments which have the same motion can be merged to obtain the final motion segmentation map.

Given the initial spatial partition $R_i, i = 1, ..., q$, containing q micro regions, a regular graph is derived from its topology. We denote it by Θ, the nodes V_i of which correspond to the regions R_i of the spatial partition. Let links $E_{i,j}$ join in Γ the nodes

associated with regions i and j, in the spatial partition, with a weight $W(i,j)$ given by spatial and motion similarity measures between the regions.

$$\Theta = \{\{V_1, ..., V_q\}, \{E(1,1), ..., E(q,q)\}, \{W(1,1), ..., W(q,q)\}\} \tag{5.10}$$

We attach three features to each region: the centroid location, the mean intensity, and the optical flow vector estimated between subsequent pairs of images. For the motion information characteristic segment R_i is assumed as uniquely assigned a segmentation label L_{R_i}. Each atomic region has a single motion vector that illustrates its motion, estimated using the technique described below in Section 5.5.1.

The definition of the region similarity which involves not only motion information but also spatial characteristics is a challenging issue. In particular, the spatial information provides important hints about object boundaries. All the available information should be put to work in order to robustly define the objects present in the scene.

We propose a region similarity measure that exploits both spatial similarity $w_s(i,j)$ and motion similarity $w_m(i,j)$:

$$W(i,j) = \varphi \cdot w_m(i,j) + (1-\varphi) \cdot w_s(i,j) \tag{5.11}$$

where φ is a regularisation term that reflects the importance of each measure. Spatial similarity measure is obtained using the technique described in Section 4.7, and motion similarity measure is described below in Section 5.5.2. At this phase the role of w_s is only to be a refinement measure. Therefore, in our experiments φ was set to 0.95.

5.5.1 Region motion vector

The proposed method applies spatial pre-segmentation to the first image. Using atomic regions implicitly resolves the problems identified earlier which requires smoothing of the optical flow field since the spatial (static) segmentation process will group together neighbouring pixels of similar intensity, so that all the pixels in a area of smooth intensity grouped in the same region will be labelled with the same motion. We thereby presume two basic assumptions: i) it is assumed that all pixels inside a region of homogeneous intensity follow the same motion model, and ii) motion discontinuities coincide with the boundaries of those regions. To ensure that our assumptions are met, we apply a strong over-segmentation method to the image.

Our first goal is to associate a unique optical flow vector to each atomic region. While the atomic region motion vector is computed from the optical flows, it is necessary to consider the real situation that some of the optical flows might have been contaminated with noises, causing the computation of the region motion vector deviate from its genuine motion vector. For each optical flow, its contribution to the deviation depends both on its magnitude and on its direction. Thus, another goal is to detect and exclude those optical flows which tend to cause large errors to the computation of the region motion vector. We achieve these goals by obtaining the dominant motion of the atomic regions region from the mode of each optical flow component in the region.

5.5.2 Motion similarity measure

For region-based motion segmentation, we assign a unique motion vector to each region. To reflect human perceptual characteristics for motion similarity measure, we adopt the distance metric proposed by Yoshida [Yoshida 02]. The idea here is to represent a motion vector $\mathbf{v} = (v_x, v_y)$ in a (U_x, U_y) plane (Figure 5.6) with radius ρ and the argument θ given by:

$$\rho(\mathbf{v}) = \log\left(1 + \beta\left(v_x^2 + v_y^2\right)^{1/2}\right) \tag{5.12}$$

$$\theta(\mathbf{v}) = \tan^{-1}\left(v_y/v_x\right) \tag{5.13}$$

The parameter β is a positive parameter included to reflect the variation in the similarity judgement of motion from person to person.

The motion information of each region are computed in reference to different points - the centroids of the regions. We define a motion distance $d_m(i,j)$ expressing the degree of similarity between the motion fields of two regions R_i and R_j in reference to the centroid of R_i. From Figure 5.6, $d_m(i,j)$ can be expressed as:

$$d_m(i,j) = \sqrt{(\Delta^2 U_x + \Delta^2 U_y)}$$

$$\Delta U_x = \rho_i \cos\theta_i - \rho_j \cos\theta_j \tag{5.14}$$

$$\Delta U_y = \rho_i \sin\theta_i - \rho_j \sin\theta_j$$

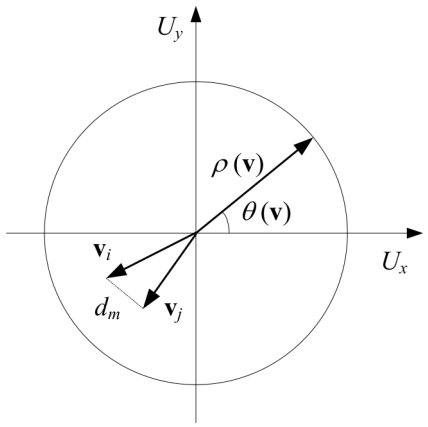


Figure 5.6: Representation of motion vectors in the (U_x, U_y) plane.

where ρ_i, ρ_j, θ_i and θ_j are calculated by Equations (5.12) and (5.13). In fact, this motion distance expresses how well the motion model of region R_j can also fit the motion of region R_i.

As the distance measures have their own range it is desirable to normalize their values. The parameter σ_m in Equation (5.15) is used to normalize the distance measure to a range $[0, 1]$.

$$w_m(i,j) = \exp\left(-d_m(i,j)^2/\sigma_m^2\right) \tag{5.15}$$

5.6 Motion segmentation algorithm

In this section, we aim to integrate spatial segmentation and motion information for high quality motion segmentation. If it is true that for synthetic sequences flow field values can be computed exactly, that is not the typical scenario, where flow field is estimated from a sequence of images. Then, our approach should be robust against inaccuracies in the motion information.

Starting from a pre-segmentation of the reference frame, the proposed technique determines the motion objects constituting the scene at hand. To that end, the over-segmented regions are merged according to their mutual spatial and temporal similarity. By treating regions as the elementary unit for image processing, we can reduce the computational complexity without a corresponding loss of accuracy. The information about spatial and temporal similarity between regions is represented by a region-based motion graph. A spectral-based clustering algorithm is used to detect clusters of similar motion regions and to achieve the motion segmentation.

We assume that a region of uniform motion (rigid motion) will be composed of one or more atomic regions each of which possessing uniform intensity. Consequently, the motion boundaries will be a subset of the intensity boundaries determined at this stage. We refer to this assumption as *segmentation assumption*. Our choice of this assumption is supported by the following fact: the atomic regions resulting from the spatial pre-segmentation are usually small enough to justify the assumption of piecewise constant intensity and motion.

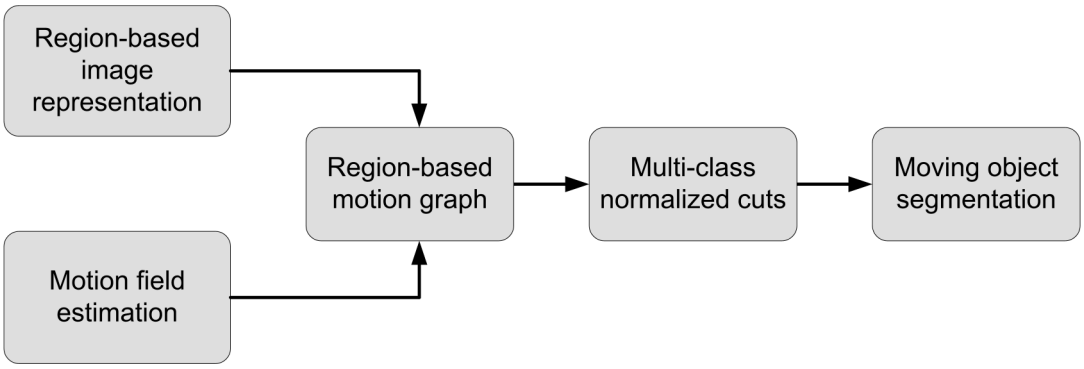


Figure 5.7: Block diagram of the proposed hybrid motion segmentation method.

The procedure of the motion segmentation algorithm is presented in the diagram of Figure 5.7 and illustrated in Figure 5.8. It can be summarized as follows:

Step 1: Spatial pre-segmentation: images of sequence are partitioned into homogeneous atomic regions based on their brightness properties using the segmentation algorithms introduced in Section 4.5

Step 2: Motion estimation: estimate the dense optical flow field with the variational scheme described in Section 5.4.2.

Step 3: Dominant motion extraction: extract the highly reliable optical flows for each atomic region. It selects from the dense flow field the dominant motion vector according to the directions and magnitudes of the optical flows. This step eliminates the influence of noise and outliers.

Step 5: Region-based motion graph: build the region-based motion graph where the nodes correspond to regions.

Step 6: Graph partitioning: multiclass spectral based graph partitioning using the normalized cut approach described in Section 4.6.

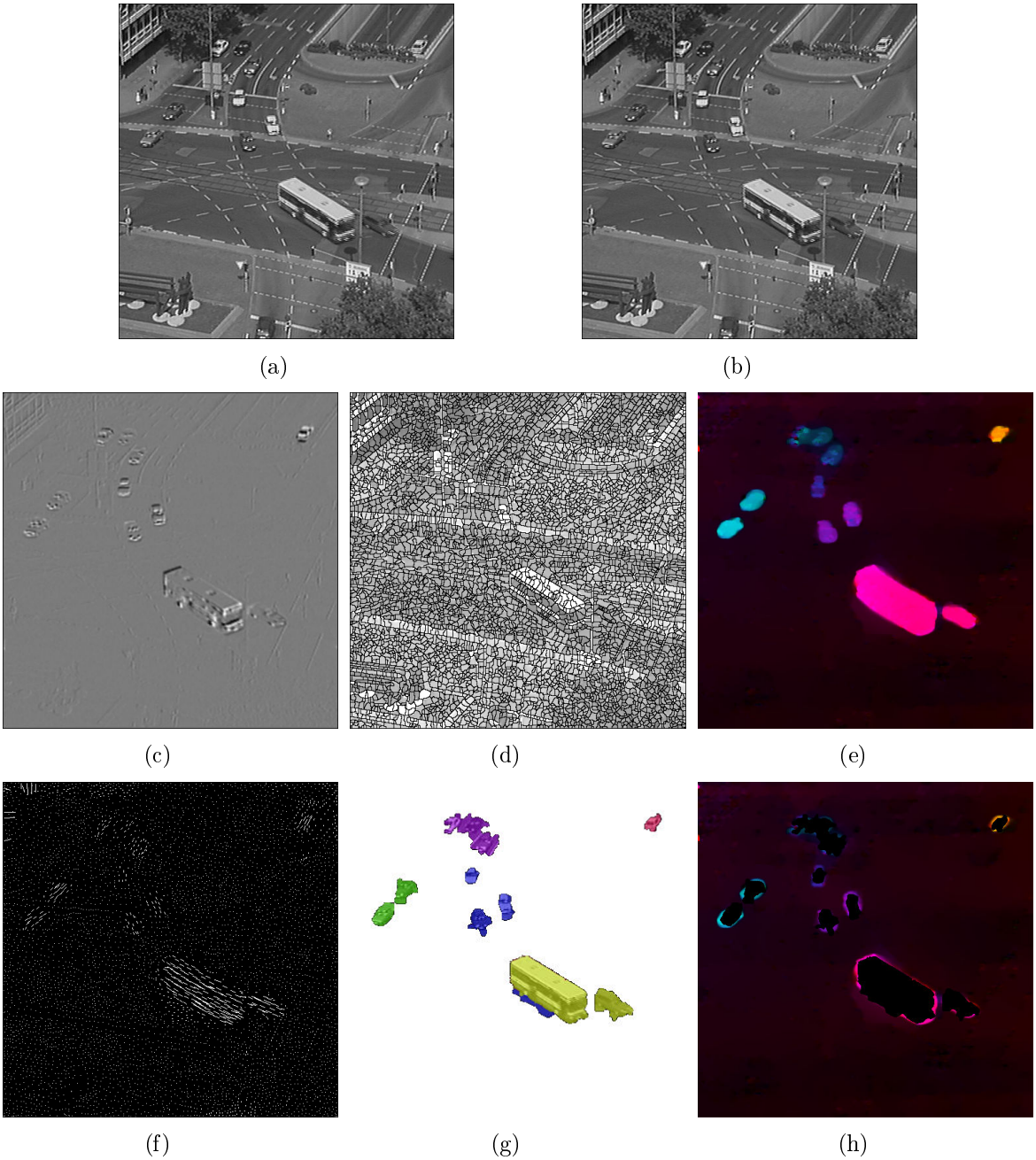


Figure 5.8: Illustration of the proposed motion segmentation algorithm. (a)-(b) Frame 5 and 6 of the Ettlinger Tor sequence (grey-scale). (c) Absolute difference between the frames. (d) Atomic regions. (e) Computed dense optical flow. (f) Region-based vector field scaled by a factor of 2. (g) Motion segmentation. (h) "Difference" between (e) and (g).

The input is represented by two consecutive frames of the Ettlinger Tor sequence (frames 5 and 6). The sequence consists of 50 frames of size 512×512 and depicts a variety of moving cars (up to 6 pixels per frame). Thereby five groups of cars can be formed according to their velocity and direction: 1) a bus and a car in the foreground are moving fast to the right; 2) in the middle area three cars are moving in a similar direction of group 1 but slower; 3) two cars on the left are moving to the left; 4) in the upper middle area three cars are moving slowly to the left; 5) on the upper right area a car is moving up.

In the first step, an initial segmentation of the frames is achieved with watershed-based segmentation. The result is a fine partition of the image into regions with intensity homogeneity where region sizes are kept small (in this case we suppress the pre-flooding step). Motion estimation between the frames is obtained with the variational method described in Section 5.4.2 and depicted in Figure 5.8.e) according to colour code proposed in Figure 5.2. In the following, a dominant motion vector is associated with each region produced in step 1. Figure 5.8.f) shows a representation of the resultant flow vectors scaled by a factor of 2. Finally, Figure 5.8.g) presents the result of the motion segmentation where different kind of motions are represented by different colours[6] in accordance with the five groups upper referenced.

It is important to understand why the area under the bus was labelled as belonging to group 2 and not to group 1. This area has been originated in the motion estimation process as a consequence of the brightness similarity between the bottom of the bus and the ground. In other words, since the smoothness term expands the optical flow along areas of homogeneous intensity it has also expanded the bus motion to the ground. However, the optical flow of the ground has a lower magnitude which makes it more similar to the motion of the cars in group 2 than to the motion of the bus. This shows the accuracy of the motion segmentation algorithm.

As it was expected the result from the motion segmentation is very similar with the motion estimation result. Figure 5.8.h) shows the refinement produced by the region-based motion segmentation. It is possible to see that it removes the "halo" originated by the smoothness term used in the motion estimation process allowing to obtain a more accurate segmentation. Even more, the segmentation effectively separates the groups of cars according to their type of motion.

[6]These colours have nothing to do with the colours in Figure 5.2.

5.7 Summary

A method for multiple motion segmentation was presented, relying on a combined region-based segmentation scheme. A region-based motion graph was built on the partition obtained in a spatial pre-segmentation stage. The derivation of a motion-based partition of the images was achieved through a graph labelling process in a spectral-based clustering approach. To achieve this aim an appropriate similarity function (energy function) was defined. Links weights now denote a similarity measure in terms of both spatial (intensity and gradient) and temporal (flow fields) features. To compute the flow field we use a high accuracy optical flow method based on a variational approach. The region-based graph-labelling principle provides advantages over classical merging methods which by operating a graph reduction imply irreversibility of merging. Moreover, spectral-based approach avoids critical dependency in the order in which regions are merged. The proposed approach successfully reduces computational cost, while enforcing spatial continuity of the segmentation map without invoking costly Markov random field models.

The algorithm takes advantage of spatial information to overcome inherent problems of conventional optical flow algorithms, which are the handling of untextured regions and the estimation of correct flow vectors near motion discontinuities. The assignment of motion to regions allows the elimination of optical flow errors originated by noise.

To partitioning each image into a set of homogeneous regions, we used the watershed transform implementation proposed in Chapter 4. By treating regions as an elementary unit for further processing, we reduced the computational complexities without a corresponding loss of accuracy. Each frame is converted into a region-based motion graph and the graph is partitioned into perceptually significant groups by means of the normalized cuts algorithm. The weights on links of the region-based motion graph are defined by the motion similarity which is computed by using a perceptual measure. By simultaneously making use of both static cues and dynamic cues we are able to find coherent groups within a variety of video sequences. The experiments presented in Chapter 6 show that the proposed method provides satisfactory results in motion segmentation from image sequences.

CHAPTER 6

Image and motion segmentation: experimental results

In order to test the performance of the proposed image segmentation framework we use a number of images from the Berkeley dataset. The results are evaluated and compared with those obtained with the state-of-the-art methods described in [Deng 01, Comaniciu 02, Cour 05]. Additionally, the results from the described motion segmentation algorithm are tested using several benchmark test sequences and therefore allowing a comparison with other algorithms. Due to the lack of motion segmentation ground truth we only show visual results for our algorithm.

6.1 Hybrid spatial segmentation: results

For spatial segmentation we mainly used images from the Berkeley Segmentation Dataset [Martin 01]. This database comprises a ground truth of 300 hand-segmented images by a minimum of 5 subjects, to compare the segmentation outputs. We identify each image with the identification number presented in [Martin 01]. To expand the field of application of our algorithm some other images are also used, including medical images. The results are shown in Appendix A. Due to the absence of ground truth to such images we present only the qualitative results of the segmentations.

Although some optimisation could be made, in our experiments we use the same threshold values for every images. Thus, in the gradient magnitude computation we use $\rho_o = 8$, $\rho_s = 1$ and $\rho_e = 3$. The smoothing bilateral filter was applied with $\sigma_r = 30$ and

$\sigma_s = 4$. The flooding level is 0.0125 times the gradient magnitude standard deviation. The standard deviation of the similarity measures proposed in Equations (4.32) and (4.33) are $\sigma_{ic} = 0.02$ and $\sigma_I = 0.02$ times the maximum intensity value of the image.

6.1.1 Evaluation

The evaluation measure proposed in Chapter 3 requires a calibration image to set up the weighted functions w_p and w_n as defined in Equations (3.16) and (3.17). We use the calibration image represented in Figure 6.1 to which correspond the threshold values $\alpha_p = 80$ and $\alpha_n = 20$.

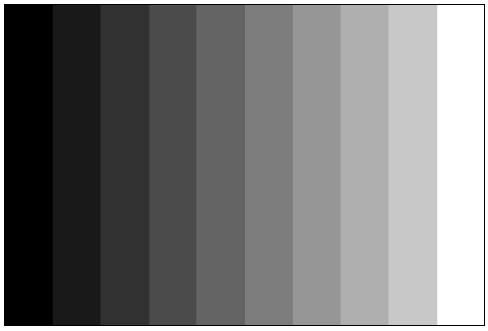


Figure 6.1: Calibration image used to set up the parameters of s_w.

Figure 6.2 depicts the experimental results on image segmentation of a set of natural scene images taken from the Berkeley Dataset. Left column shows the original image with the corresponding Berkeley identification number. Right column presents the segmentation results where each segment is labelled with a different colour. To show the accuracy of the segmentation results the labelled segments are superimposed on the original image. The number of segments is putted under each segmented image.

One problem usually associated with normalized cuts approach is the partition of homogeneous regions. Due to the suppression of spatial distance in similarity measure and to the use of the flooding level in the computation of watershed atomic regions this problem is greatly reduced in our approach.

Table 6.1 and Table 6.2 show the segmentation evaluation in terms of weighted measure s_w and F-measure from a set of randomly chosen images from Berkeley dataset. The bottom row shows the evaluation results obtained when considering the calibration image as being the reference image.

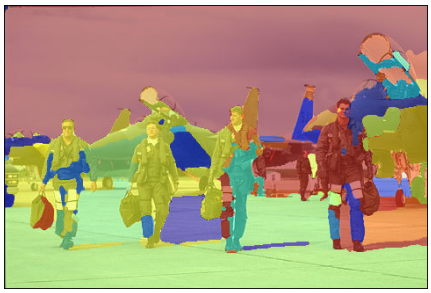
3096

5 segs

24063

10 segs

245051

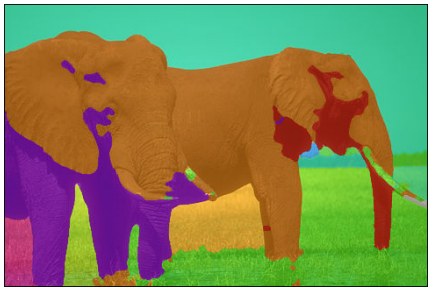


40 segs

286092

30 segs

296059



12 segs

Figure 6.2: Experimental segmentation results over images from the Berkeley dataset.

Table 6.1: Evaluation of the images in Figure 6.2 in terms of weighted measure s_w and F-measure.

Measure	**3096**	**24063**	**245051**	**286092**	**296059**
s_w	0.99	0.82	0.74	0.78	0.68
F	0.84	0.80	0.67	0.71	0.72
$s_{w_{cal}}$	0.01	0.00	0.33	0.19	0.00

Although in complex images such as images 245051 and 286092, the segmentations are not yet the ideal ones, they exhibit promising results.

Table 6.2: Evaluation of the images in Figure 6.3 in terms of weighted measure s_w and F-measure.

Measure	**37073**	**41004**	**42049**	**65019**	**90076**	**118035**	**143090**	**241004**
s_w	0.57	0.77	0.90	0.67	0.94	0.79	0.79	0.80
F	0.65	0.75	0.89	0.80	0.85	0.74	0.71	0.81
$s_{w_{cal}}$	0.00	0.12	0.11	0.15	0.00	0.09	0.02	0.02

Comparison with other segmentation methods

We have compared our method (WNCUT) with three state-of-the-art segmentation algorithms: (i) mean shift (EDISON) [Comaniciu 02], (ii) a multiscale graph based segmentation method (MNCUT) [Cour 05], and (iii) JSEG [Deng 01]. For this comparison we use the set of natural images shown in Figure 6.3. To provide a numerical evaluation measure and thus allow comparisons, the experiments for the evaluation were conducted on the manual segmentations of the Berkeley Segmentation Dataset [Martin 01]. The task is cast as a boundary detection problem, with results presented in terms of Precision (P) and Recall (R) measures.

The algorithm provides a binary boundary map which is scored against each one of the hand-segmented results of Berkeley Dataset, producing a (R, P, F) value. The final score is given by the average of those comparisons.

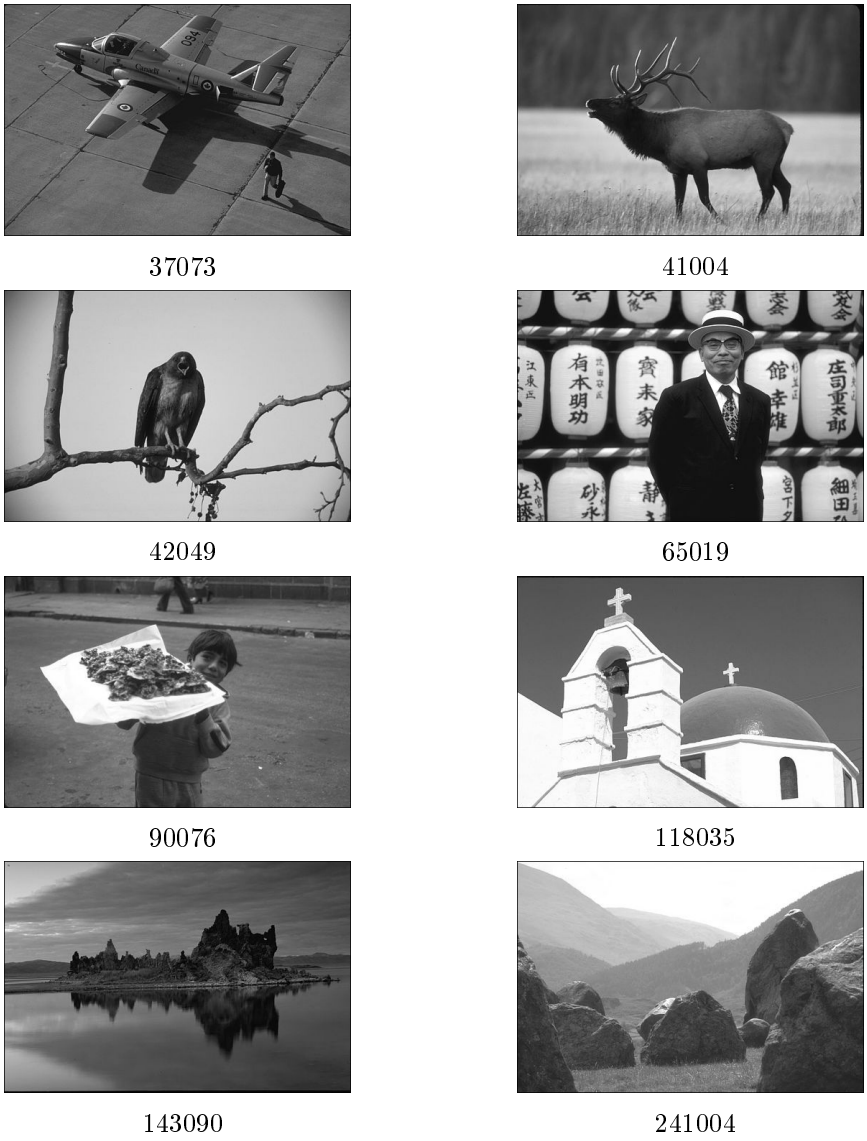


Figure 6.3: Set of tested images taken from the Berkeley dataset. Each image is identified with the Id number used in the dataset.

Mean shift methods [Fukunaga 75, Comaniciu 02] have gained popularity for image segmentation due to their lack of reliance on a priori knowledge of the number of expected segments. Mean shift is an iterative procedure to find clusters in the joint spatial and colour spaces. Given an image, the algorithm is initialized with a large number of hypothesized cluster centres randomly chosen from the data. Then each cluster centre is moved to the mean of the data lying inside the multi-dimensional ellipsoid centred on the cluster centre. The vector defined by the old and the new cluster is called the *mean shift vector*. The mean shift vector is computed iteratively until the cluster centres do not change their positions. Note that during this process

some clusters may get merged.

As described in [Comaniciu 02], the mean shift based segmentation algorithm takes as input parameters a feature bandwidth h_r, a spatial bandwidth h_s and a minimum region (in pixels) M. It uses the adaptive specification of the two bandwidths according to the data statistics in the image and colour domains to define a kernel in the joint spatial-range domain to filter image pixels and a clustering method to retrieve segmented regions. The two bandwidth parameters are critical in controlling the scale of the segmentation result. Too large values result in loss of important details, or under-segmentation; while too small values result in meaningless boundaries and excessive number of regions, or over-segmentation. In this comparison we tested the images with a set of values for each parameter, $h_s = \{7, 11, 15\}$, $h_r = \{7, 11, 15\}$ and $M = \{200, 300, 400\}$. These values were empirically found, after carrying out several tests with different images. The parameters were adjusted to each image in order to obtain the highest F-measure.

Christoudias et al. [Christoudias 02] presented an algorithm using mean shift segmentation that addresses directly to the image clustering. In this approach, a region adjacency graph is created to hierarchically cluster the modes. Also, edge information from an edge detector is combined with the colour information to better guide the clustering. This is the method used in the publicly available EDISON system, also described in [Christoudias 02]. The EDISON system is the implementation we use here as the mean shift segmentation system.

Deng and Manjunath [Deng 01] proposed the JSEG method for multiscale segmentation of colour and texture, based on colour quantization and region growing. Their algorithm also consists of two stages: colour quantization and spatial segmentation. Colour quantization maps each pixel into a class label, which is used in the second stage to minimize a homogeneity measure of colour-texture patterns. Spatial segmentation is based on seeded region growing and region merging. JSEG segmentation algorithm takes as input parameters a colour quantization threshold q_r, the number of scales n_s and a region merge threshold m. We leave for automatic determination of q_r and n_s by the original software. For each image we change the region merge threshold in a range of $0.0 - 0.8$ and as in EDISON approach found the segmentation result with the highest F-measure.

We think that it is also important to contrast our method with another successful

graph partitioning algorithm. In [Cour 05], Cour et al. presented a multiscale spectral image segmentation algorithm (MNCUT) which works on multiple scales of the image in parallel, without iteration, to capture both coarse and fine level details.

The quantitative evaluation results are summarized in Figure 6.4 for the set of tested images. To a better visualisation of the comparative results we decided to represent these results in a graphic figure. A table with the values of F-measure of Figure 6.4 is presented in Appendix A. Taking into consideration that the methods can produce results with different number of regions, we have taken as a region count reference number the average number of regions from the human segmentations available for each image. To understand the level of variability in the segmentation results, the errors among the results from the manual segmentation were also computed.

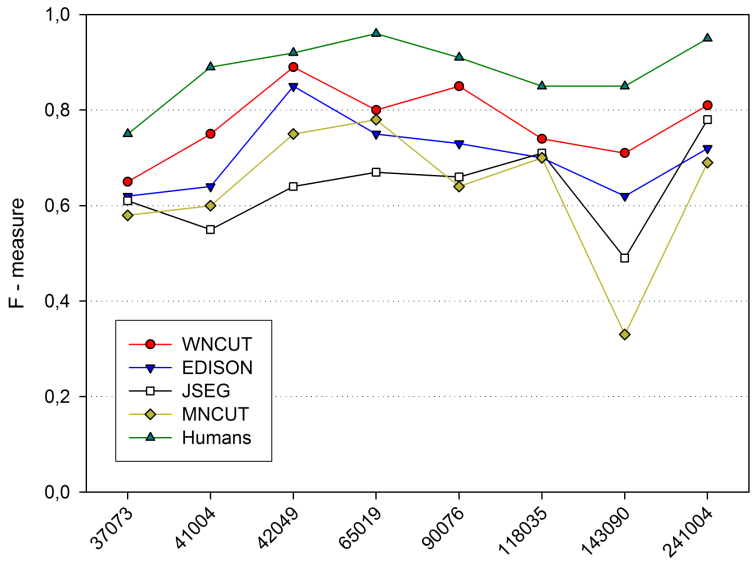
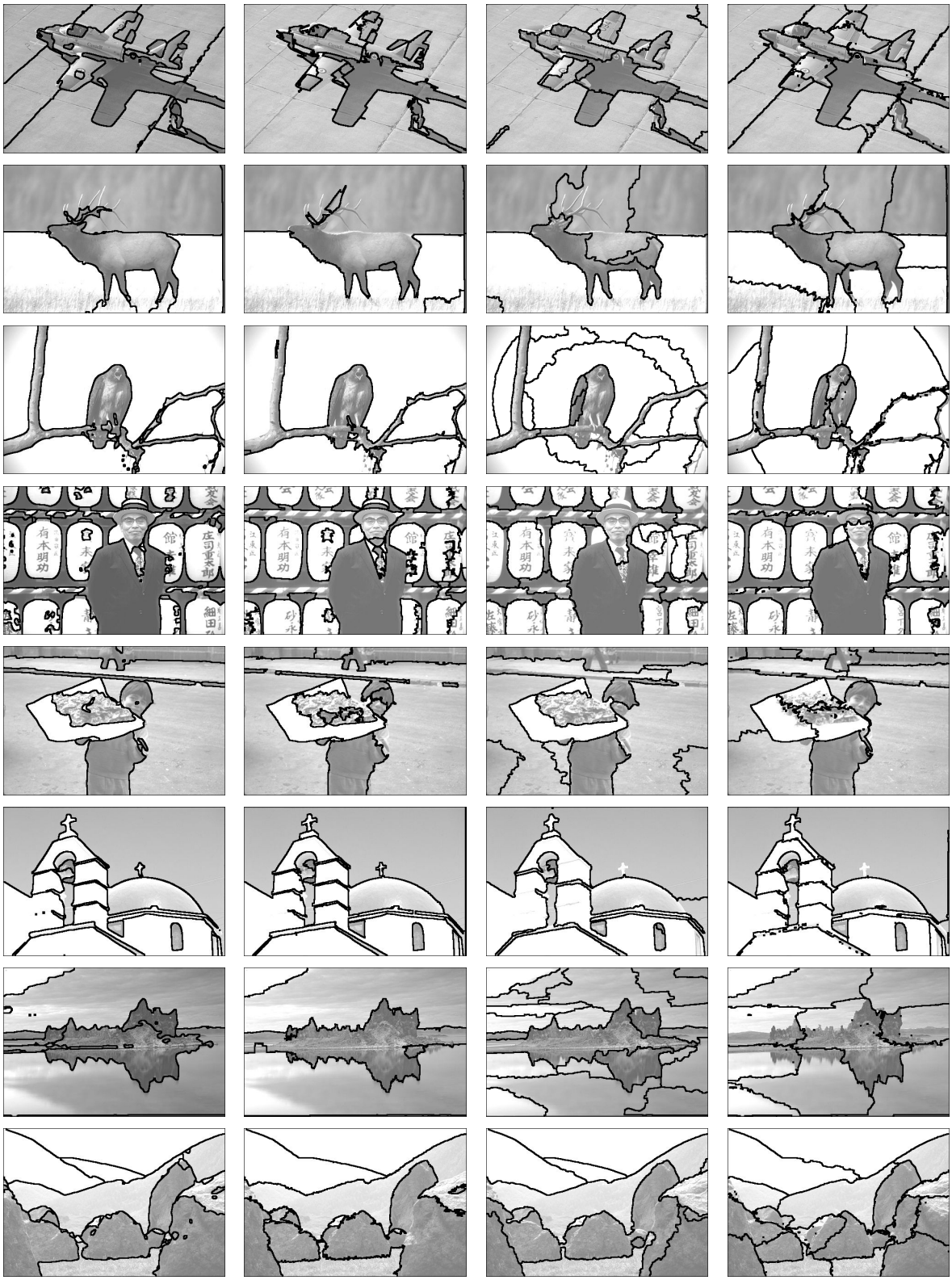


Figure 6.4: Results of F-measure evaluation for the comparison between methods.

The resulting segmentation after the application of the examined algorithms is shown in Figure 6.5. Since the F-measure is a boundary-based measure the segmentation results are presented as boundaries over the original images. The proposed approach produces segmentations of high quality. For all images in Figure 6.5 the set of segments is reasonably compact. The proposed method produce better results than the other methods for every images.

This new approach overcomes some limitations usually associated with spectral clustering approaches. As we can see from the segmentation result of image 118035,



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

Figure 6.5: Segmentation results: (a) proposed method (WNCUT), (b) Mean shift (EDISON) [Comaniciu 02], (c) JSEG [Deng 01], and (d) the multiscale segmentation MNCUT [Cour 05].

larger homogeneous regions are not partitioned into separated regions.

Compared with the other methods, the proposed approach has overall less over-segmentation and a very good boundary location. It produces an overall score of $F = 0.77$, against $F = 0.72$ for EDISON, and $F = 0.66$ for JSEG and MNCUT. Note that due to the variability of segmentations among humans, the overall score of manual segmentations is $F = 0.88$.

Although EDISON and JSEG produce results with high value of precision, the correspondent recall value is in general low. For example, with $h_s = 11$, $h_r = 4$ and $M = 100$, EDISON evaluation for image 41004 gives $R = 0.79$, $P = 0.27$ and $F = 0.40$. This is due to the over-segmentation produced by these methods.

According to these results, we can conclude that our method generally provides results with a F-measure better than other state-of-the-art methods.

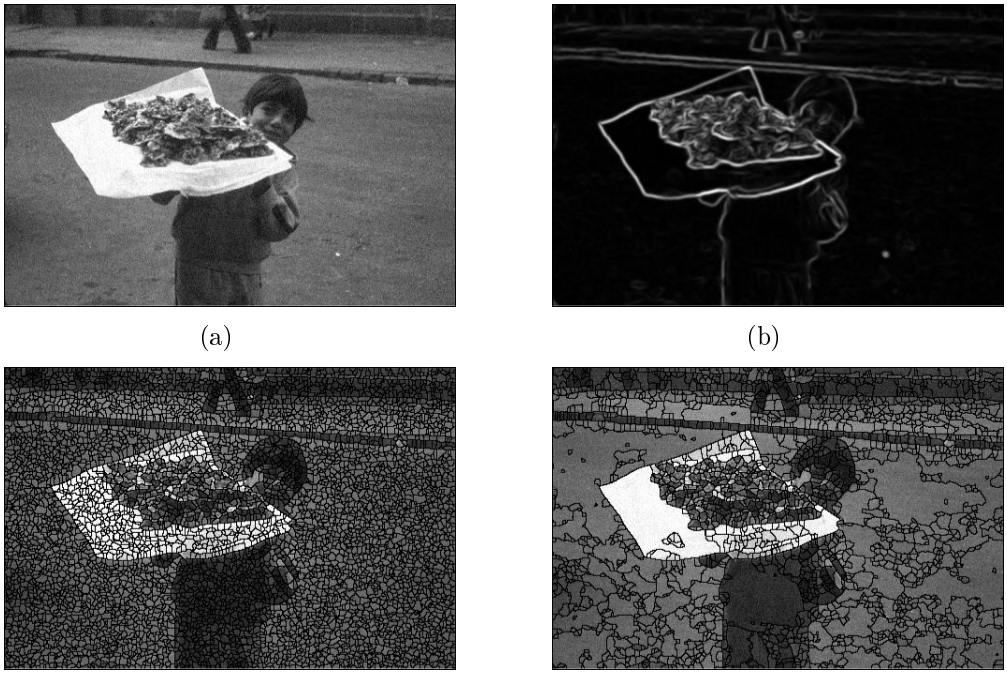
6.1.2 Robustness to noise

Larger over segmentation at the first stage will result in a graph that increase the computational cost, since the eigensystem complexity depends on the number of atomic regions being clustered. The dominant parameter controlling this stage is the flooding level threshold applied to the gradient image which we empirically set to 0.025 times the mean image gradient. This factor determines the degree of over segmentation and thus the number of nodes of the graph (Figure 6.6).

The flooding level can be a function of local image characteristics, such as gradient magnitude, intensity or variance. Such function may additionally depend on one or more parameters. Figure 6.6 compares the watershed segmentation computed without and with this modification.

To analyse the behaviour of the algorithm in presence of noise, the images were corrupted with four levels of Gaussian additive noise with standard deviations $\sigma =$ 5, 10, 20, 30. All the tests were done without changing the parameter values of the methods. The effect of the pre-processing step in reducing the noise, with a reduction on the number of irrelevant regions in the output of the watershed algorithm, can be observed in Table 6.3 and in Figure 6.7.

Our method turned out to be extremely robust to artificially added Gaussian noise. We may notice that segmentation results are not very affected till $\sigma = 20$, and it



(a) (b)

(c) (d)

Figure 6.6: Effects of pre-processing in watershed transform. (a) Original image with added Gaussian noise with $\sigma = 10$ (154 401 pixels). (b) Gradient magnitude image. (c) Regions in the "raw" watershed (6 104 segs). (d) Regions in the pre-processed image (2 223 segs).

produces a good segmentation even for added Gaussian noise with an amplitude of $\sigma = 30$. This amount of noise is greater than would be expected in a normal real image.

6.2 Motion segmentation: results

The motion segmentation algorithm described in Chapter 5 was tested using several benchmark test sequences: *Tennis*, *Salesman* and *Flower Garden with Car*. These three are among the sequences widely used by authors for testing video segmentation and coding applications.

It is difficult to access, in quantitative terms, the accuracy of a real world motion segmentation. Some authors have presented "ground truth" data to some sequences [Chung 07]. However, these reference images are not extracted in a motion-based process. They are obtained using some iterative image segmentation method like the ones presented in Chapter 2. Therefore, the results presented here are only qualitative.

Figure 6.8 shows the segmentation result with the *Tennis* sequence. In this part

Table 6.3: Results of quantitative evaluation in terms of F-measure for original image and for added Gaussian noise with $\sigma = 5, 10, 20, 30$.

σ	**37073**	**41004**	**42049**	**65019**	**90076**	**118035**	**143090**	**241004**
0	0.65	0.75	0.89	0.80	0.85	0.74	0.71	0.81
5	0.64	0.72	0.87	0.71	0.83	0.73	0.68	0.80
10	0.63	0.71	0.81	0.69	0.82	0.72	0.67	0.77
20	0.60	0.68	0.77	0.68	0.81	0.72	0.64	0.75
30	0.53	0.64	0.71	0.67	0.78	0.71	0.58	0.64

of the sequence, the player bounces the ball on his bat as he prepares to serve. The upper arm is almost stationary, and the lower arm naturally obeys a motion part-way between that of the upper arm and the bat, so an uncertain labelling is somewhat justified. The motion of the ball is, of course, a genuine fourth independent motion. The ball's displacement between frames is quite large - about 20 pixels.

This example illustrates an important dilemma in motion segmentation. Looking only at the actual motions the forearm is essentially pivoting at the elbow so that there is large motion at the bat and smaller motions on the arm, whilst the motion of the upper part of the arm is so small that it could very plausibly be classified as the same as the background (Figure 6.8.f)). This is a general problem where motions in an image (typically due to rotations) become indistinguishable from the motions of nearby regions. In this case there is always going to be some ambiguity about where the division between the motion classes should be when considered solely on the basis of the motion.

Figure 6.8.g) shows the resulting segmentation from the Tennis sequence where most of the arm is correctly classified. One exception is the bottom of the ball, which is incorrectly classified in the region in which the flow field is propagated to the adjacent three atomic regions under the ball. This is essentially due to the large motion of the ball (Figure 6.8.c)) which causes occlusions that affect the accuracy of motion estimation. Even more, the region under the ball has diffuse brightness that affects also the spatial similarity.

Figure 6.9 shows the segmentation result with the *Salesman* sequence. Here we

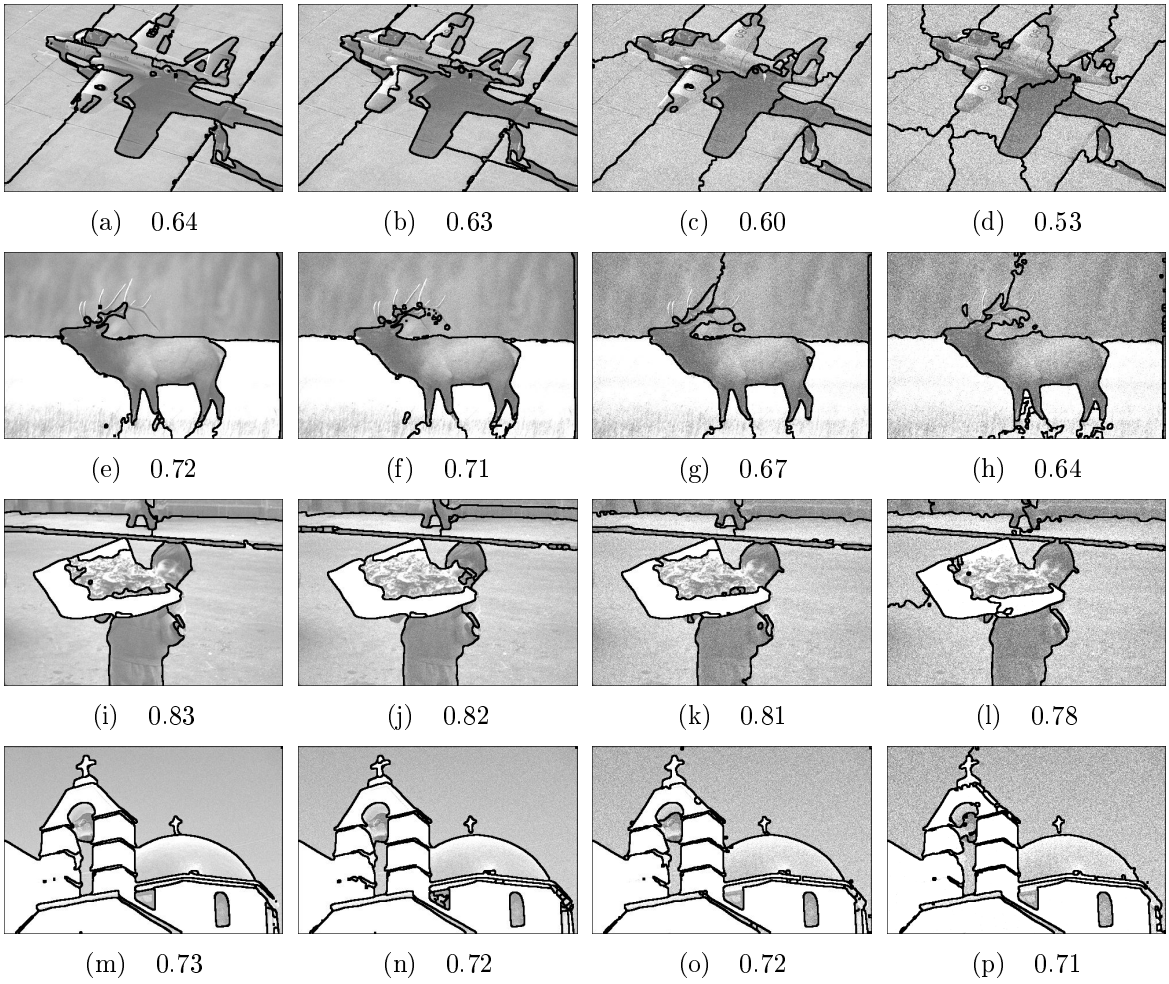


Figure 6.7: Performance of the proposed approach on noisy images. Results with added Gaussian noise with σ, from left to right, equal to 5, 10, 20, 30. The values below the images are the F-measures.

observe multiple local motions of the arm (due to movement of the shirt).

The Salesman sequence does not possess any global motion, but the motion of the non-rigid object (salesman) is significant in this sequence, especially in respect to the arm movements. It can be seen in Figure 6.9.g) that our proposed algorithm yields satisfactory multiple motion segmentation. Regions such as the arm of the Salesman and his hand, which moves with motion involving rotation, are correctly segmented. Also the shirt, that is divided in two by the arm, is correctly merged.

Figure 6.11 shows the segmentation result with the *Flower Garden with Car* sequence. This example is part of the well-known Flower Garden sequence. The sequence was shot by a camera placed on a driving car, and the image motion is related to distance from the camera. Thus the tree, which is closest to the camera moves

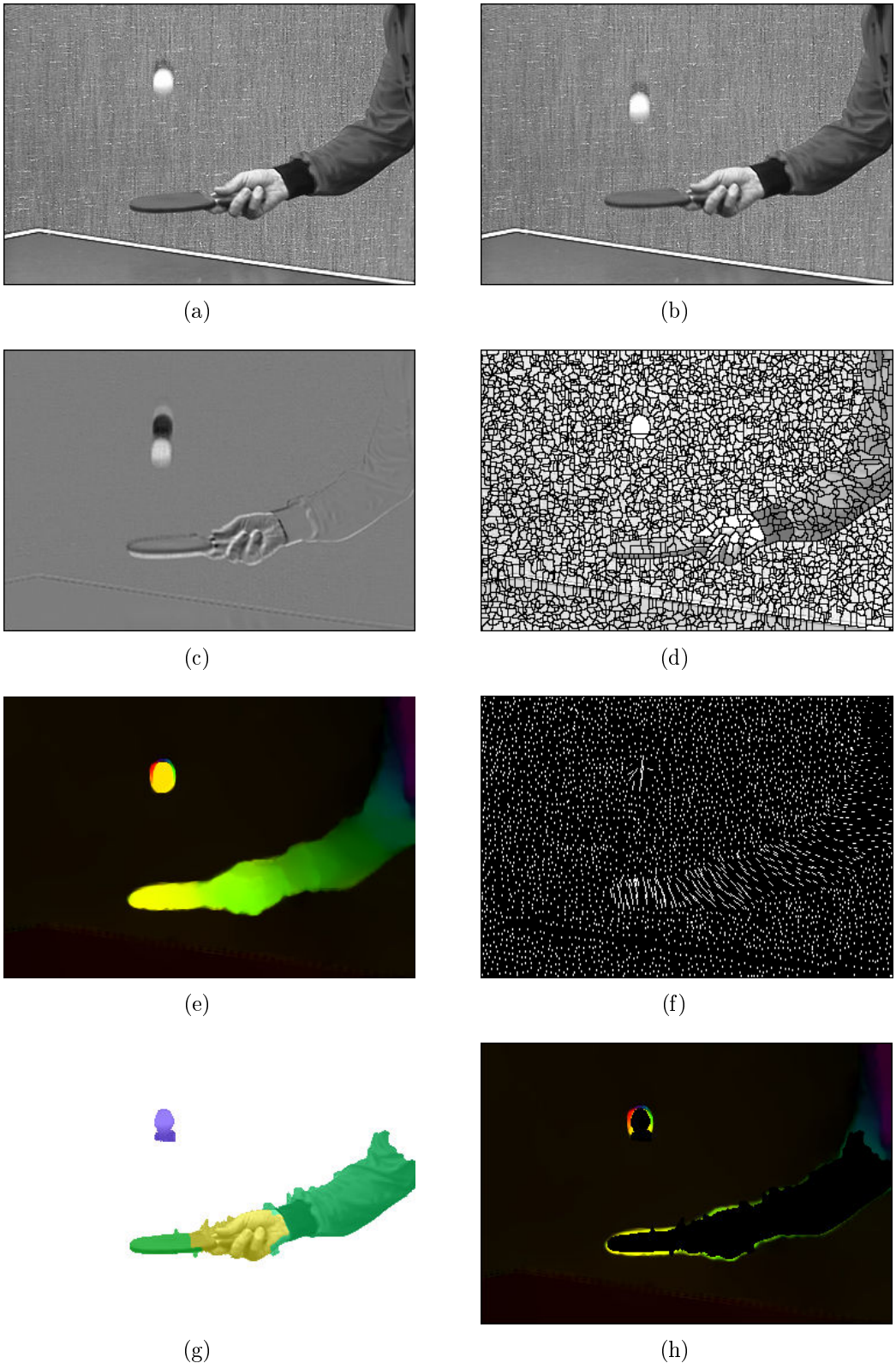
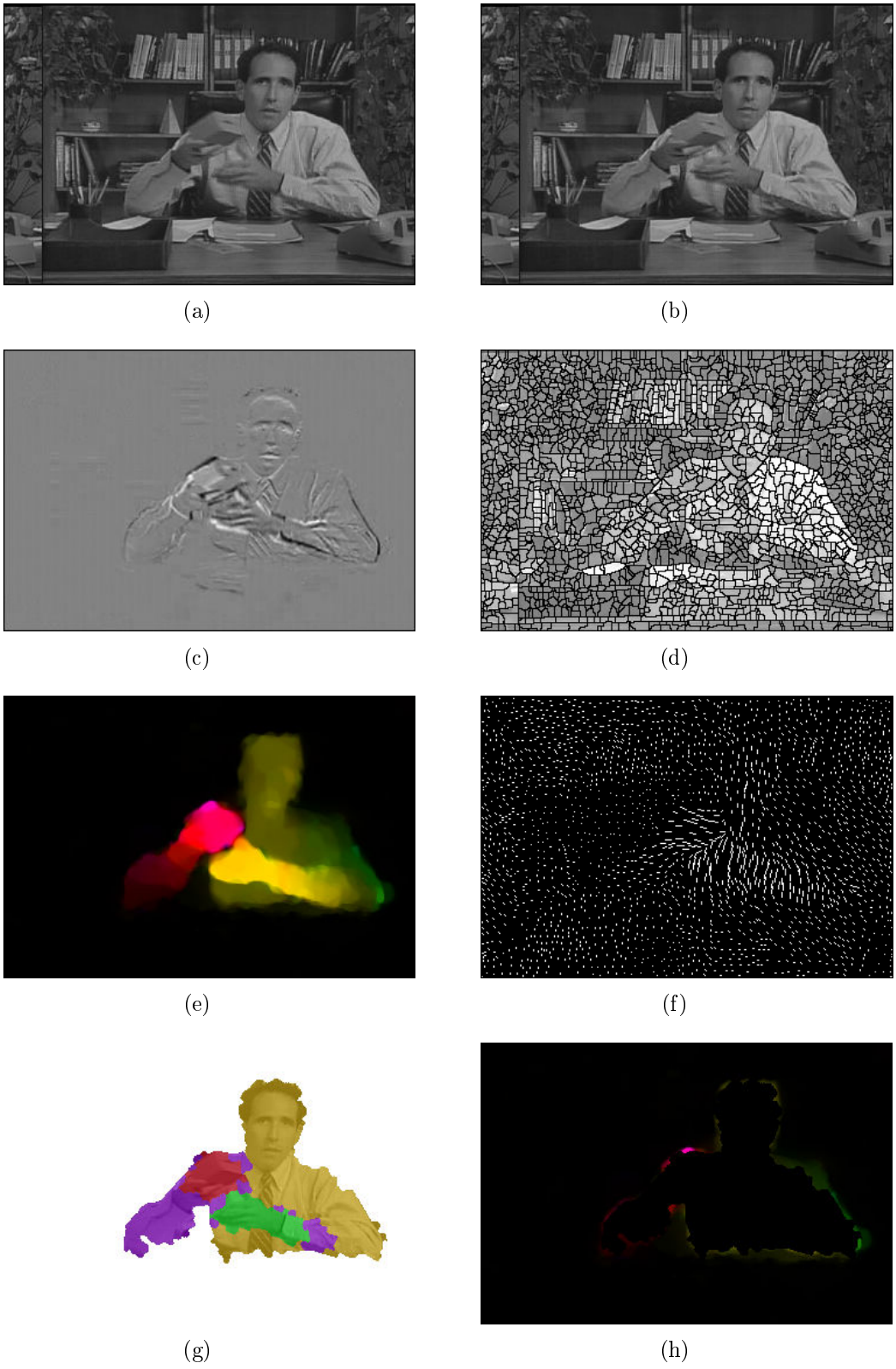


Figure 6.8: Tennis sequence. (a)-(b) Frames 8 and 9 (grey-scale). (c) Absolute difference between the frames. (d) Atomic regions. (e) Computed dense optical flow. (f) Region-based vector field scaled by a factor of 2. (g) Motion segmentation. (h) "Difference" between (e) and (g).



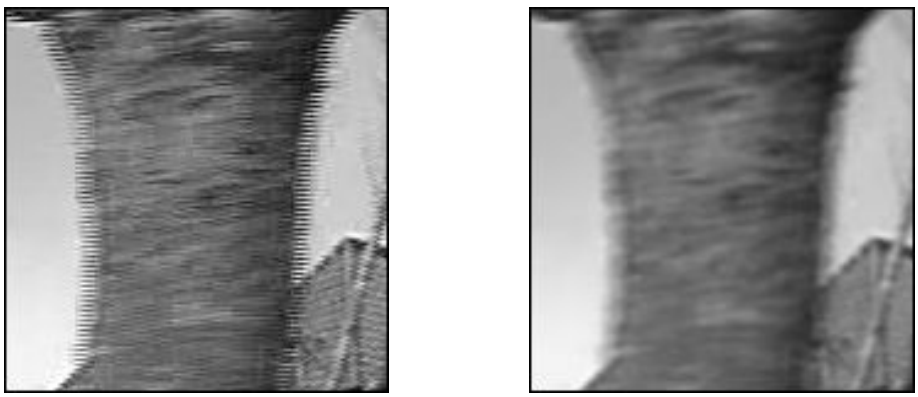
(a) (b) (c) (d) (e) (f) (g) (h)

Figure 6.9: Salesman sequence. (a)-(b) Frames 14 and 15 (grey-scale). (c) Absolute difference between the frames. (d) Atomic regions. (e) Computed dense optical flow. (f) Region-based vector field scaled by a factor of 2. (g) Motion segmentation. (h) "Difference" between (e) and (g).

fastest.

In this experiment a moving car was included in the scene. The inter-frame difference detects motion at every image pixels. Flower Garden sequence contains many depth discontinuities, not only at the boundaries of the tree but also in the background. In this sequence, the camera captures a flower garden with a tree in the centre. Also, the flower bed gradually slopes toward the horizon showing the sky and far objects. Semantically, this sequence has five layers: the tree, the car, the flower bed, the house and the sky.

We should note that this sequence has been recorded in interlacing mode and thus requires the handling of typical interlacing artefacts. These stripe artefacts that result from an alternating update of even and odd lines are typical for real-world applications. These could be reduced during the convolution with the Gaussian kernel. Figure 6.10 shows the effect of interlacing artefacts reduction.

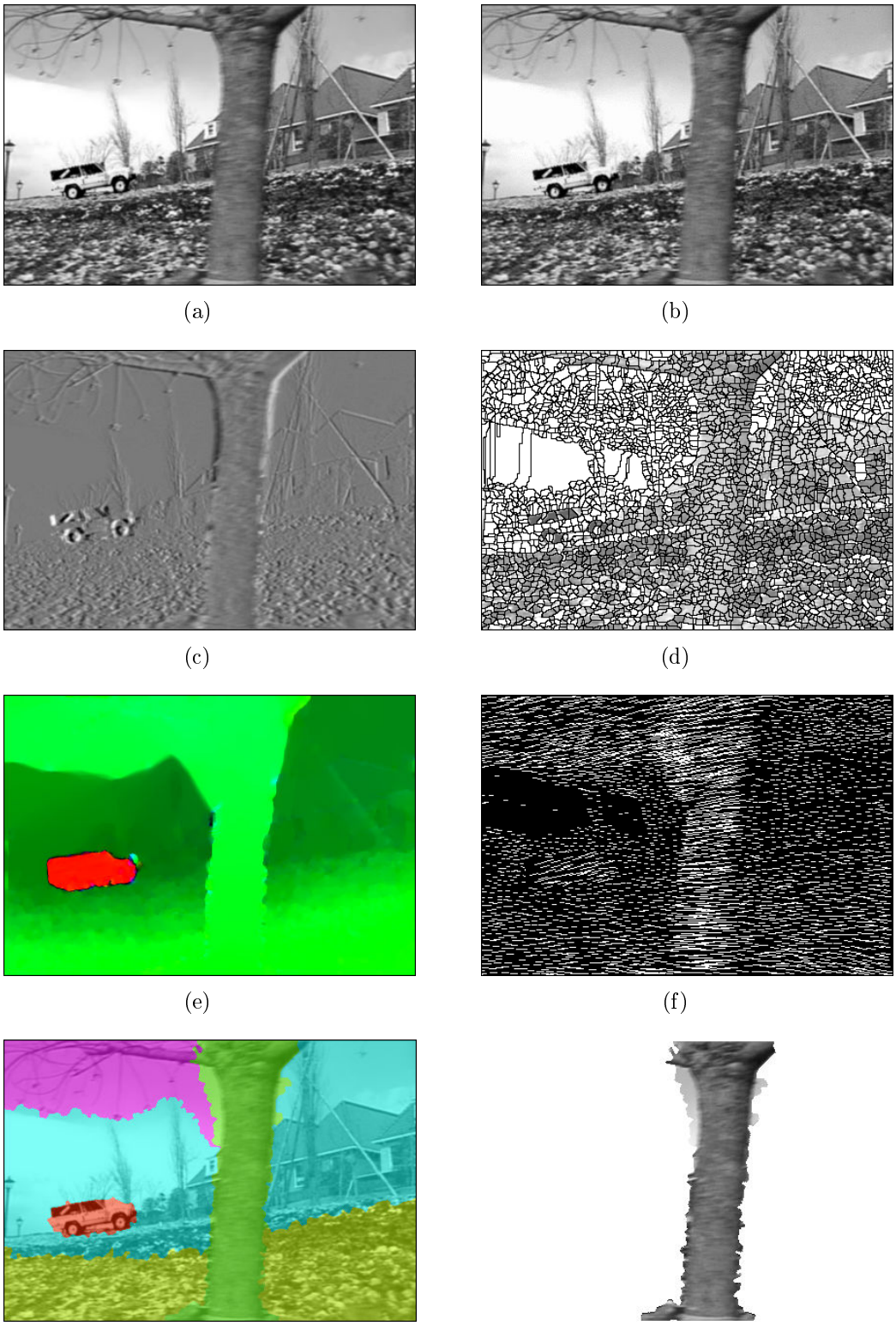


(a) (b)

Figure 6.10: Interlacing artefacts. (a) Detail from frame 5 of the Flower Garden sequence. (b) image convolved with Gaussian kernel with $\sigma = 1.0$.

Although the tree divides the flower bed the algorithm merges the two parts in one only segment. This happens also in the house layer. Note that in the area that contains the tree's branches, only one segment is chosen since the sky area has no brightness variation. Figure 6.11.e) shows the estimated optical flow with different colours represent different directions. From this figure it looks like as the bottom of the flower bed, the tree and the sky have the same motion information. However, the segmentation algorithm making use of the intensity information, correctly divides these parts.

Figure 6.11.h) shows the resulting tree segment. The region-based approach extracts



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

Figure 6.11: Flower Garden with Car sequence. (a)-(b) Frames 5 and 6 (grey-scale). (c) Absolute difference between the frames. (d) Atomic regions. (e) Computed dense optical flow. (f) Region-based vector field scaled by a factor of 2. (g) Motion segmentation. (h) Tree segment.

the tree's edges accurately along major part of the trunk, even in similar textured area of the flower bed, but less well in other areas. The fine detail of the small branches cannot be well represented by image regions, and these are segmented poorly.

6.3 Comparative results

As demonstrated by the results shown in this chapter, motion segmentation is a difficult task. It is also difficult to assess, in quantitative terms, the accuracy of a segmentation. It is therefore instructive to compare the results generated by this region-based system with work published by other authors over recent years; this gives an indication of the relative success of the region-based approach. Again, with no accepted quantitative measure of segmentation performance, a qualitative comparison is made between results.

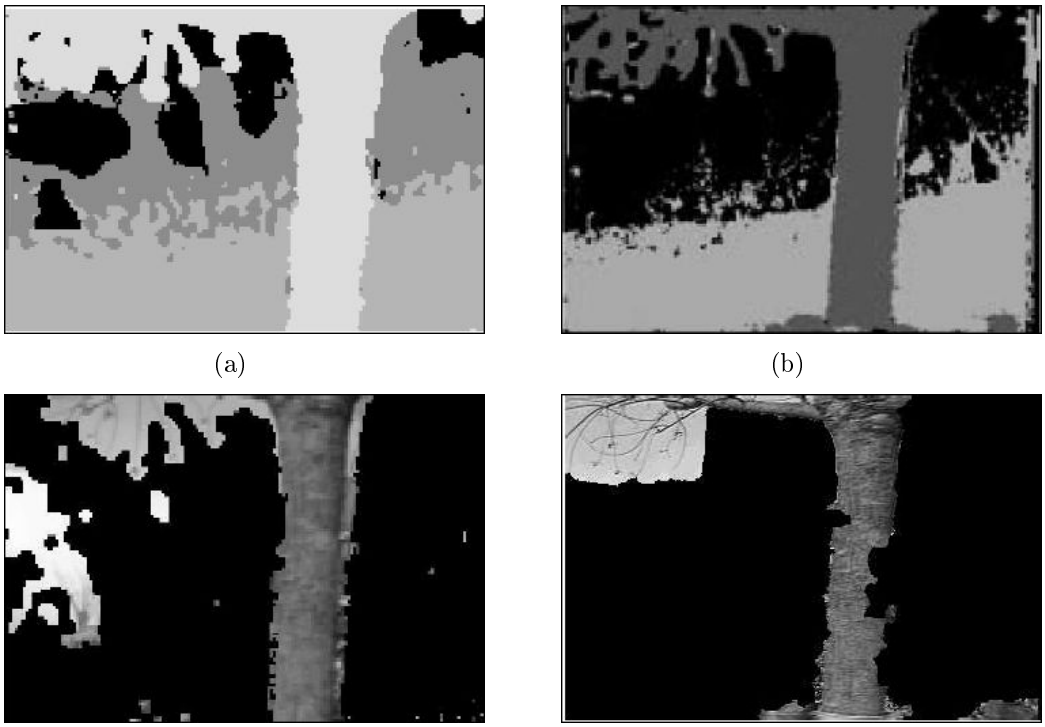
This section presents a comparison with a number of authors who have analysed the Flower Garden sequence. In this comparison we analyse the accuracy of the resulting tree segment. The results are extracted from the published papers. Although each author displays their results differently it is not difficult to compare them.

Wang and Adelson [Wang 94] presented results from this sequence in their paper introducing the layered representation. Comparisons with Ayer and Sawhney [Ayer 95] and Weiss and Adelson [Weiss 96] are also presented in Figure 6.12. Both of these authors' results show some outlying pixels or regions which are absent in our approach, which gives the system presented in this dissertation a more pleasing appearance. Figure 6.12.d) shows the result of the edge-based motion segmentation scheme from Smith [Smith 01].

The segmentation of the tree in the Wang and Adelson estimate it to be too wide, while the edge-based approach misses a few sections. Ayer and Sawhney's is a better outline, but there is more noise in the background. Although the tree segment of Weiss and Adelson is similar with our result, it is not so "clean".

6.4 Summary

This chapter has evaluated analytically and empirically the segmentation methods proposed in Chapter 4 and Chapter 5. We have experimentally shown that the pro-



(a) (b)

(c) (d)

Figure 6.12: Comparative results with the Flower Garden sequence. Results presented by (a) Wang and Adelson in [Wang 94], (b) Ayer and Sawhney in [Ayer 95], (c) Weiss and Adelson in [Weiss 96] and (d) Smith in [Smith 01].

posed approaches provide an effective region-based segmentation method for achieving high quality segmentation. It has been shown that good segmentation results can be achieved when using a combined approach between morphological and graph-based methods. We compared this new approach against other state-of-the-art segmentation techniques [Deng 01, Comaniciu 02, Cour 05]. Qualitative results for real-world sequences demonstrate the capacity of our approach to segment objects based on spatial and motion cues. A comparison with some of the best known motion segmentation methods is also made for the Flower Garden sequence.

CHAPTER 7

Conclusion

This thesis is focused on the problems of image and motion segmentation using two region-based methods.

One of the key ideas presented in this thesis is the simplification of the entry graph for the normalized cut (NCut) algorithm. A pre-segmentation process allows the construction of a region-based graph which makes the Ncut algorithm tractable to large images. This graph has a smaller size than the pixel-based graph, but still with meaningful data. The initial segmentation is not a simple "pre-processing" step such as making some assumptions on the sparsity of certain matrices [Shi 00], or using bottom-up region merging to reduce input size. By using the watershed transform we provide a ready-made matrix of relevant data as input to the NCut algorithm. We demonstrate the reliability of our algorithm with qualitative and quantitative experimental data.

Major reasons for the success of our algorithm over other similar methods are: the use of edge preserving smoothing filter; the use of intervening contours in the similarity measure; the exclusion of the spatial distance in the pairwise similarity measure; the region-based similarity graph; and the multiclass spectral-based approach. Even more, the use of watershed based regions instead of single pixels as graph nodes largely decreases the computational cost.

This region-based method also enforces spatial smoothness of the resulting motion segmentation map without using costly Markov random field models. We observe that we can tolerate over-segmentation in the spatial region formation step, since these regions will be merged later using motion vector and intensity matching. In contrast with the classical motion segmentation methods that segment sequences only as foreground/background objects, our method effectively separates the moving areas according to their motion. Experimental results demonstrate the robustness of the proposed

method, which can also be viewed as integration of motion and intensity segmentation.

Our basic assumptions for motion segmentation approach are that motion information varies smoothly inside a region of homogeneous intensity, while flow field discontinuities are located at the borders of those regions. The purpose of applying this segmentation assumption is to improve the performance of our algorithm in untextured regions and in the proximity of flow field boundaries.

There are two important advantages to estimating the velocity over a whole region rather than pixel by pixel. The first advantage is that the effects of noise and inaccuracies in the velocity vector estimation typically are reduced significantly. The second advantage is that even if the aperture problem is presented in some part of the region, information obtained from other parts can help to fill in the missing velocity component. A disadvantage with velocity estimation over a whole region is that it is assumed that the true velocity field is at least reasonably consistent with the chosen motion model. A problem here is that even if we know, e.g. from the geometry of the scene, that the velocity field should be patch-wise affine, we still need to obtain regions not covering patches with different motion parameters. There are many possible solutions to this problem, including grey level segmentation and the ideal case of a priori knowledge of suitable regions.

7.1 Contributions

There have been three main themes pursued through out this thesis. The first two are image segmentation and correspondingly evaluation, and the third is motion segmentation. This section summarizes the contributions of this work.

Our contribution in Chapter 2 is a review of the recent contributions in the area of image segmentation with emphasis on the cooperative segmentation methods. We also proposed a new categorization of image segmentation algorithms.

In Chapter 3, we introduce a new evaluation metric for image segmentation. Most of the currently used evaluation metrics measure in one way or another the quantity of false and positive pixels in the segmentation result making no perceptual differentiation among them. Our region-based measure takes into account not only the accuracy of the segments boundary localization regardless to the number of regions in each partition. From the comparison of the proposed metric with some of the best known evaluation

measures in the literature we have shown that our method is tolerant to refinement and at the same time strongly penalizes segmentation errors. This comply with the way humans perceive visual information.

In Chapter 4, we develop a new hybrid segmentation technique for still images which combines edge and region-based information with spectral techniques through the morphological algorithm of watersheds. A non-linear smoothing (bilateral filter) is used to reduce over-segmentation in the watershed algorithm while preserving the location of the image boundaries. The purpose of the pre-processing step is to reduce the spatial resolution without losing important image information. An initial partitioning of the image into primitive regions is set by applying a rainfalling watershed simulation on the image gradient magnitude. This step presents a new approach to overcome the problems with flat regions. This initial partition is the input to a computationally efficient region segmentation process (multiclass normalized cut algorithm) that produces the final segmentation. The method's accuracy and robustness were demonstrated through a series of experiments involving several real images. Our experimental results were also compared with other published results, and the comparison indicated that the proposed method produced results that fall into the most accurate category.

The third problem that we address in this thesis is the estimation and segmentation of motion. In Chapter 5, we apply the proposed framework to motion segmentation. Motion estimation is obtained with the variational method proposed by Brox et al. [Brox 04]. This method relies on a piecewise smooth assumption using a gradient constancy regularisation which yields robustness against illumination changes between the corresponding images. We also develop the theory linking the motion labelling of pixels with that of motion labelling of regions. The major advantages of this region-based motion segmentation algorithm are twofold. First, it is likely to reduce the effect of leverage pixels by encouraging flow field maps to have spatially coherent support. Optical flow vectors inside a region are constrained to follow a unique dominant vector. This allows the assignment of smooth optical flow field in regions of poor texture. Secondly, optical flow discontinuities are enforced to coincide with region borders. This is advantageous, since we believe that motion segmentation boundaries can be more accurately identified by the use of static cues than using motion information only.

The performance of this method was demonstrated in Chapter 6 through a series of experiments involving several of the most currently used image sequences.

7.2 Open topics and future research

The work presented in this thesis provides a new effective framework for image and motion segmentation which has been illustrated on various experiments. The approaches presented open several extension opportunities and a number of areas of interesting future work that are still allowed to go through for further exploration.

The motion segmentation assumption is not guaranteed to hold truth. This is a limitation of our approach and our current solution is to apply a stronger over-segmentation. However, since this does not completely overcome this problem, our algorithm could take benefit for example from an operation that allows splitting segments. It would be interesting to develop a special purpose intensity segmentation method as well that avoids producing regions which overlap a depth discontinuity.

Our image segmentation evaluation measure needs a calibration image to set up the thresholds. Further investigation on the choice of universal thresholds is needed. The segmentation algorithms' parameters are also chosen empirically. In a more advanced implementation parameter estimation could be automated (e.g. based on the expected level of image noise or optical flow field variation).

Image segmentation and motion estimation are considered to be separate problems. In further research we are planning to set up an image segmentation system that exploits temporal relationships and a motion estimation system that exploits region-based image segmentation. These should improve the quality of image segmentation as well as of motion estimation.

An explicit treatment of the occlusions and, more specifically, of occlusions in the previous frame could be beneficial. This implies the identification of segments that have just appeared in the scene and the relaxation of the assumption of the temporal continuity of the label map in such cases.

The algorithm presented here computes a motion segmentation map between any two frames of a sequence. It is also possible to extend it to temporally integrate these maps to obtain more stable motion boundaries across successive frames.

In order to improve the quality of results, we intend to apply the algorithms to specific areas, e.g. Medical Imaging where some preliminary experiments proved to achieve good results (see Appendix A).

APPENDIX A

Additional experimental results

This chapter presents additional experimental results of the region-based image segmentation algorithm described in Chapter 4.

A.1 Additional quantitative results

Table A.1 shows the quantitative evaluation results of the comparison of *WNCUT* method with the state-of-the-art methods. The same results are presented in a graphic representation in Figure 6.4.

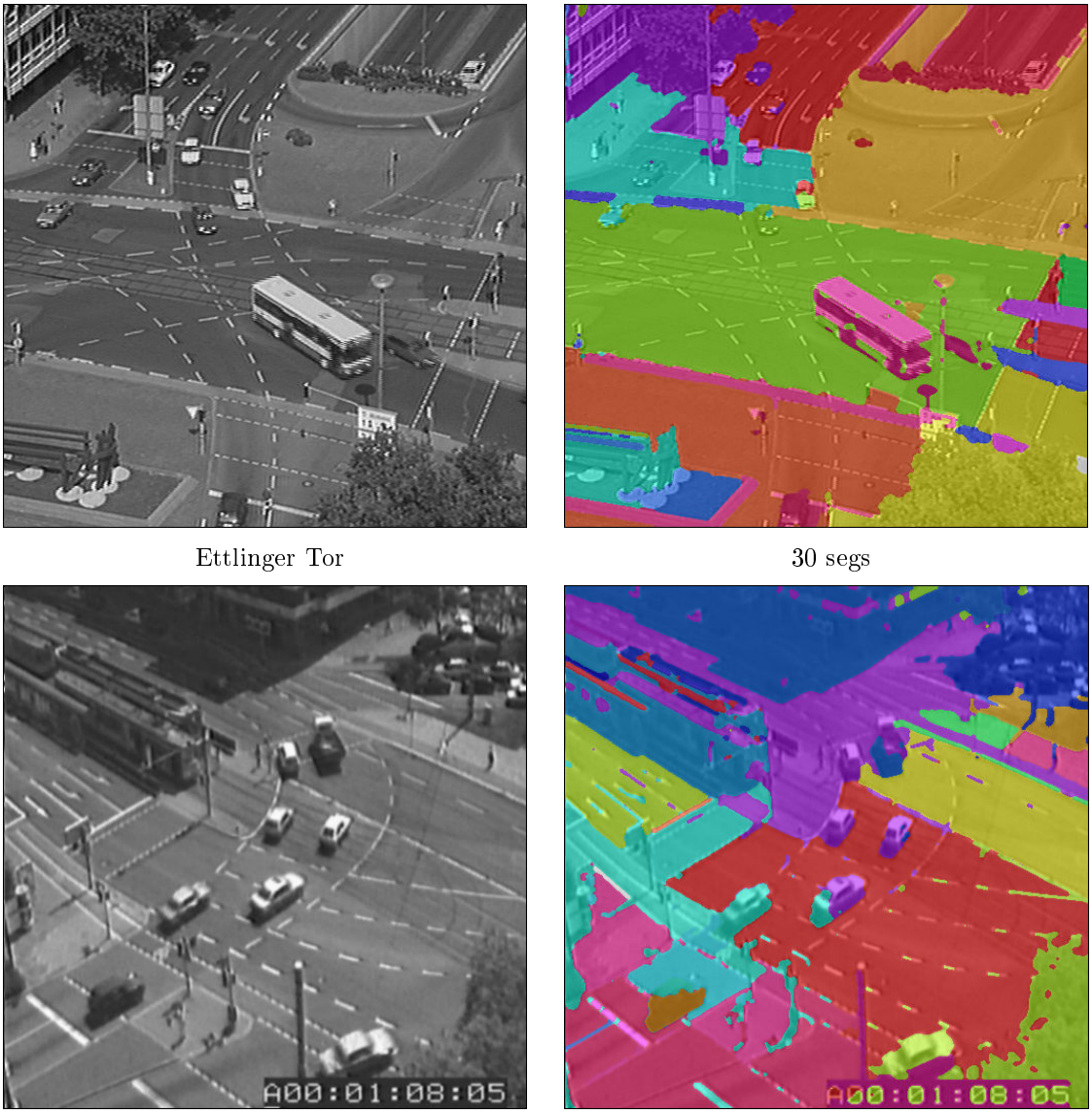
Table A.1: Results of quantitative evaluation in terms of F-measure for the comparison between the proposed method (WNCUT), Mean shift (EDISON), JSEG and the multiscale segmentation MNCUT. The last row shows the evaluation among hand-segmented results.

Method	**37073**	**41004**	**42049**	**65019**	**90076**	**118035**	**143090**	**241004**
WNCUT	0.65	0.75	0.89	0.80	0.85	0.74	0.71	0.81
EDISON	0.62	0.64	0.85	0.75	0.73	0.70	0.62	0.72
JSEG	0.61	0.55	0.64	0.67	0.66	0.71	0.49	0.78
MNCUT	0.58	0.60	0.75	0.78	0.64	0.70	0.33	0.69
Humans	0.75	0.89	0.92	0.96	0.91	0.85	0.85	0.95

A.2 Additional qualitative results

To a better visualisation of the results they are superimposed on the original images. As in the experiments of Chapter 6 the parameters were set to $\sigma_{ic} = 0.02$ and $\sigma_I = 0.02$.

Figure A.1 presents the results of the segmentation over complex real images. More results, not so complex, are shown in Figure A.2 and in Figure A.3.



Ettlinger Tor 30 segs

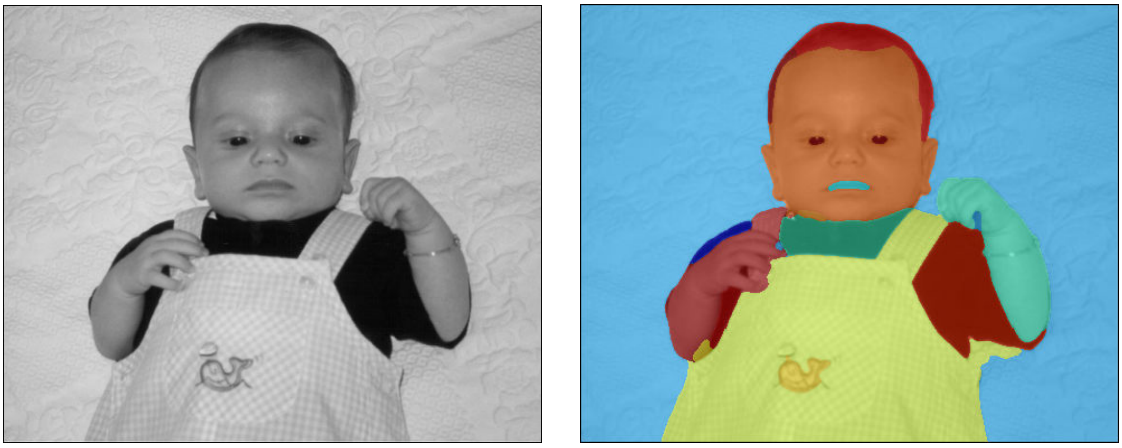
Street 30 segs

Figure A.1: Experimental segmentation results over complex real images.

Figure A.4 shows the segmentation results over medical images. It is perceptible, for example, in results of images (c), (d) and (e), the accuracy of the method as it follow the correct lung boundaries even if they are very complex.

Claire 6 segs

Peter 6 segs



JP 15 segs

Figure A.2: Experimental segmentation results over images showing humans.

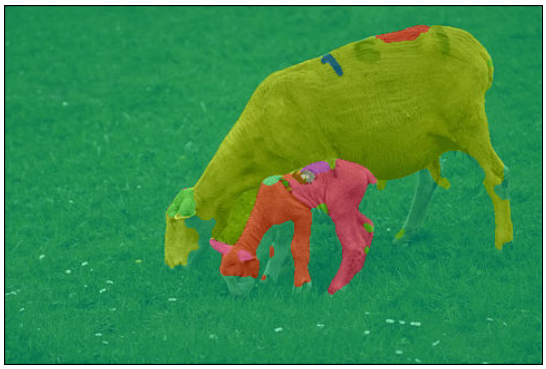
42044

8 segs

172032

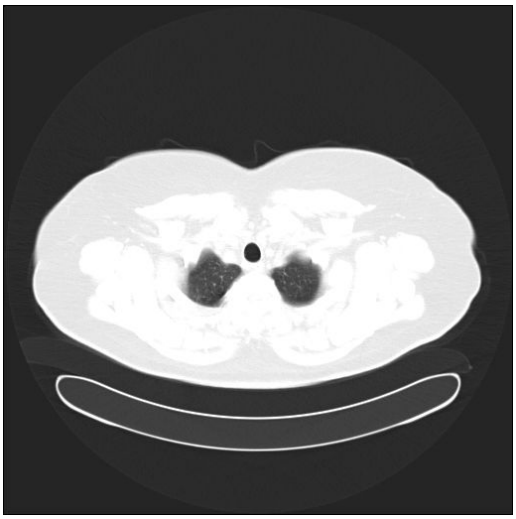
20 segs

207056

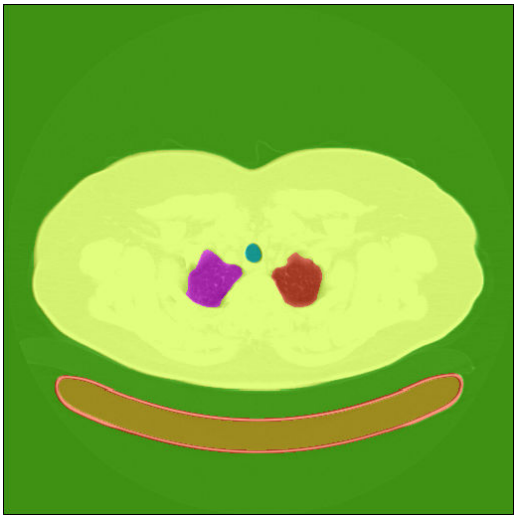


12 segs

Figure A.3: Experimental segmentation results over real images.

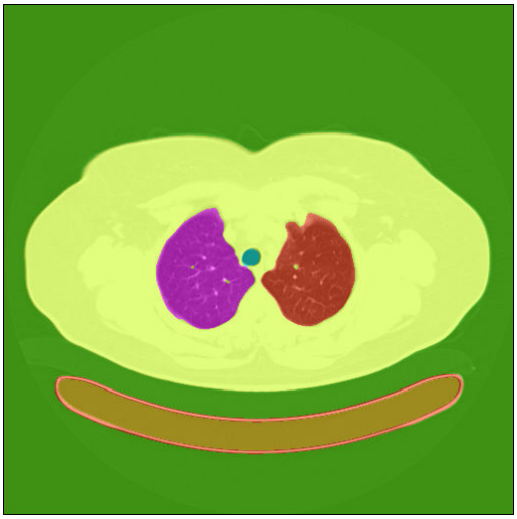


(a)



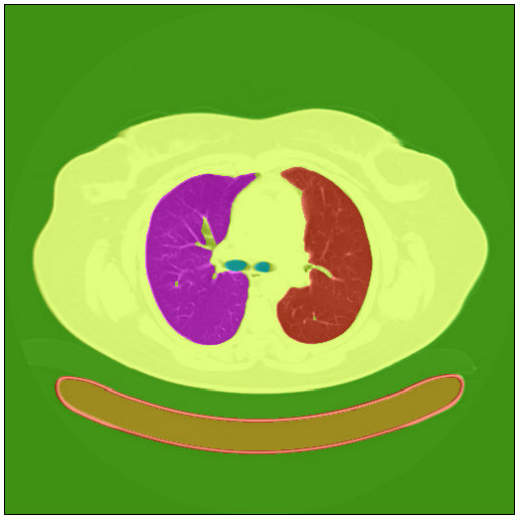
7 segs

(b)

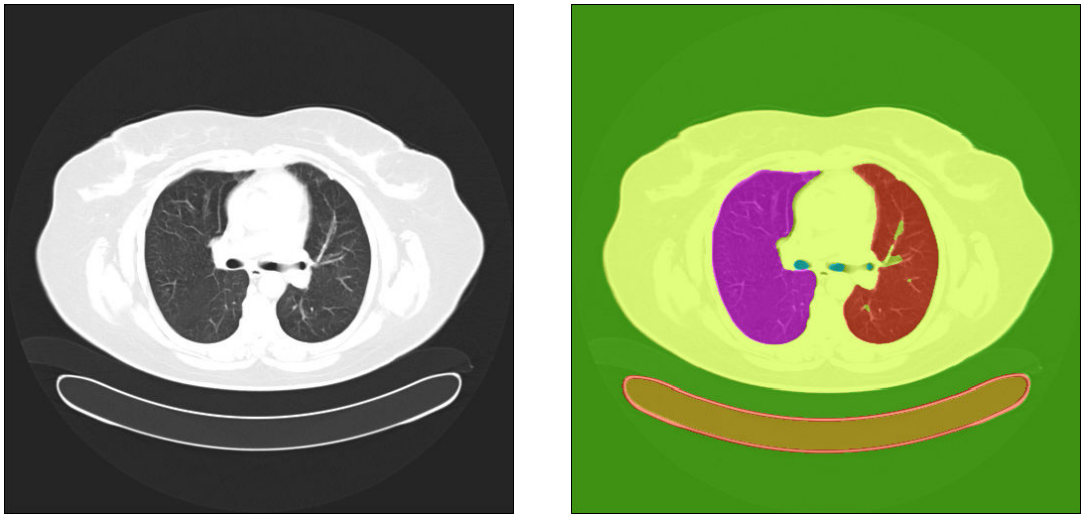


7 segs

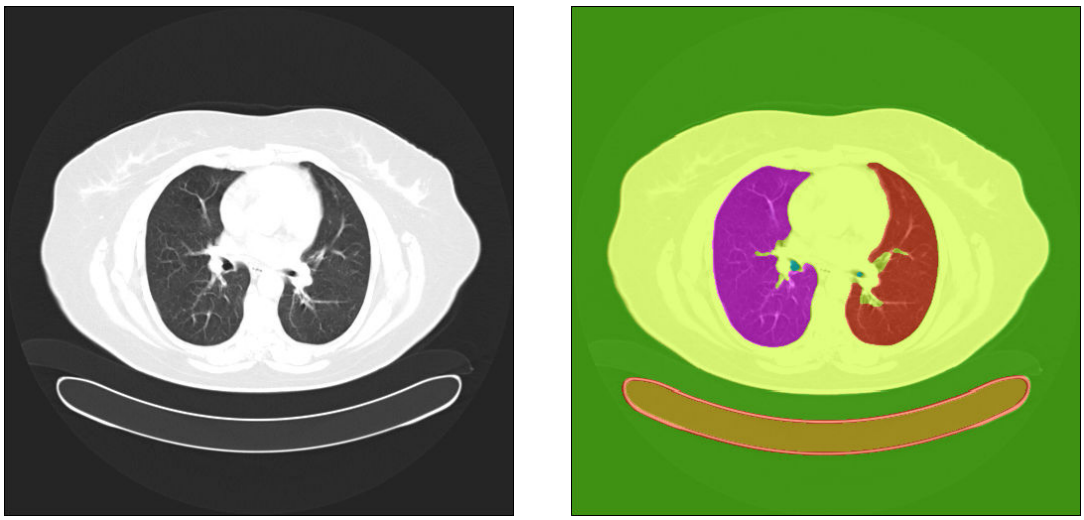
(c)



7 segs



(d) 7 segs



(e) 7 segs

(f) 7 segs

Figure A.4: Experimental segmentation results over medical images with $k = 7$.

References

[Adams 94] R. Adams & L. Bischof. *Seeded region growing: a new approach.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16, no. 6, pp. 641–647, 1994.

[Adiga 01] P.S. Adiga & B.B. Chaudhuri. *An efficient method based on watershed and rule-based merging for segmentation of 3D histo-pathological images.* Pattern Recognition, vol. 34, no. 7, pp. 1449–1458, 2001.

[Ahmed 02] M.N. Ahmed, S.M. Yamany, N. Mohamed, A.A. Farag & T. Moriarty. *A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data.* IEEE Transactions on Medical Imaging, vol. 21, no. 3, pp. 193–199, 2002.

[Amiaz 07] T. Amiaz, E. Lubetzky & N. Kiryati. *Coarse to over-fine optical flow estimation.* Pattern Recognition, vol. 40, no. 9, pp. 2496–2503, 2007.

[Antani 02] S. Antani, R. Kasturi & R. Jain. *A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video.* Pattern Recognition, vol. 35, no. 4, pp. 945–965, 2002.

[Archip 02] N. Archip, P.J. Erard, M. Egmont-Petersen, J.M. Haefliger & J.F. Germond. *A knowledge-based approach to automatic detection of the spinal cord in CT images.* IEEE Transactions on Medical Imaging, vol. 21, no. 12, pp. 1504–1516, 2002.

[Ayer 95] S. Ayer & H. S. Sawhney. *Layered representation of motion video using robust maximum-likelihood estimation of mixture models and MDL encoding.* In Proc. IEEE Int. Conference on Computer Vision, pp. 777–784, June 1995.

[B.-Hadiashar 98] A. Bab-Hadiashar & D. Suter. *Robust optic flow computation.* Int. Journal of Computer Vision, vol. 29, no. 1, pp. 59–77, 1998.

[Barbu 05] A. Barbu & S-C. Zhu. *Generalizing Swendsen-Wang to sampling arbitrary posterior probabilities.* IEEE Transactions on Pattern Analysis and Machine

Intelligence, vol. 27, no. 8, pp. 1239–1253, 2005.

[Barker 98] S.A. Barker. *Image segmentation using Markov random fields models.* PhD thesis, Department of Engineering, University of Cambridge, 1998.

[Barnes 82] E. Barnes. *An algorithm for partitioning the nodes of a graph.* SIAM J. on Algorithm and Discrete Method, vol. 3, no. 4, pp. 541–550, 1982.

[Barni 96] M. Barni, V. Cappellini & A. Mecocci. *A possibilistic approach to clustering.* IEEE Fuzzy Systems, vol. 4, no. 3, pp. 393–396, 1996.

[Barrett 98] W. Barrett & E. Mortensen. *Interactive segmentation with intelligent scissors.* Graphical Models and Image Proc., vol. 60, no. 5, pp. 349–384, 1998.

[Barrett 02] W. Barrett & A.S. Cheney. *Object-based image editing.* In Proc. of International Conference on Computer Graphics and Interactive Techniques, pp. 777–784, San Antonio, Texas, USA, 2002.

[Barron 94] J.L. Barron, D.J. Fleet & S. Beauchemin. *Performance of optical flow techniques.* Int. Journal of Computer Vision, vol. 21, no. 1, pp. 43–77, 1994.

[Beauchemin 95] S.S. Beauchemin & J.L. Barron. *The computation of optical flow.* ACM Computer Surveys, vol. 27, no. 3, pp. 433–466, 1995.

[Belongie 98] S. Belongie, C. Carson, H. Greenspan & J. Malik. *Color- and texture-based image segmentation using EM and its application to content-based image retrieval.* In Proc. of IEEE International Conference on Computer Vision, pp. 675–682, Washington, DC, USA, 1998.

[Belongie 02] S. Belongie, C. Fowlkes, F. Chung & J. Malik. *Partitioning with indefinite kernels using the Nyström extension.* In Proc. European Conf. on Computer Vision, volume II, pp. 21–31, Copenhagen, Denmark, 2002.

[Beucher 79] S. Beucher & C. Lantuéjoul. *Use of watersheds in contour detection.* In International Workshop on Image Processing, Real-Time Edge and Motion Detection/Estimation, Rennes, France, September 1979.

[Beucher 93] S. Beucher & F. Meyer. *The morphological approach to segmentation: the watershed transformation.* In Int. Workshop on Image Processing, Real-Time Edge and Motion Detection/Estimation, Rennes, France, September 1993.

[Beveridge 89] J.R. Beveridge, J.S. Griffith, R.R. Kohler, A.R. Hanson & E.M. Riseman. *Segmenting images using localized histograms and region merging.* Int. Journal of Computer Vision, vol. 2, no. 3, pp. 311–352, 1989.

[Bezdek 93] J. Bezdek, L. Hall & L. Clarke. *Review of MR image segmentation techniques using pattern recognition.* Med. Physics, vol. 20, no. 4, pp. 1033–1048, 1993.

[Black 92] M.J. Black. *Combining intensity and motion for incremental segmentation and tracking over long image sequences.* In Proceedings of European

Conference on Computer Vision, volume 588 of *LNCS*, pp. 485–493, Santa Margherita Ligure, Italy, May 1992.

[Black 96] M.J. Black & P. Anandan. *The robust estimation of multiple motions: parametric and piecewise-smooth flow fields.* Computer Vision and Image Understanding, vol. 63, no. 1, pp. 75–104, 1996.

[Blake 04] A. Blake, C. Rother, M. Brown, P. Perez & P.H.S. Torr. *Interactive image segmentation using an adaptive GMMRF model.* In Proceedings of European Conference on Computer Vision, volume 588 of *LNCS*, pp. 428–442, Santa Margherita Ligure, Italy, May 2004.

[Boccignone 04] G. Boccignone, M. Ferraro & P. Napoletano. *Diffused expectation maximisation for image segmentation.* Electronics Letters, vol. 40, no. 18, pp. 1107–1108, 2004.

[Borra 97] S. Borra & S. Sarkar. *A framework for performance characterization of intermediate level grouping modules.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 11, pp. 1306–1312, 1997.

[Borsotti 98] M. Borsotti, P. Campadelli & R. Schettini. *Quantitative evaluation of color image segmentation results.* Pattern Recognition Letters, vol. 19, no. 8, pp. 741–747, 1998.

[Bouthemy 93] P. Bouthemy & E. Francois. *Motion segmentation and qualitative dynamic scene analysis from an image sequence.* International Journal of Computer Vision, vol. 10, no. 2, pp. 157–182, 1993.

[Bowyer 01] K.W. Bowyer, C. Kranenburg & S. Dougherty. *Edge detector evaluation using empirical ROC curves.* Computer Vision and Image Understanding, vol. 84, no. 1, pp. 77–103, 2001.

[Boykov 01a] Y. Boykov & M-P. Jolly. *Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images.* In Proc. IEEE Int. Conference on Computer Vision, pp. 105–112, Vancouver, Canada, July 2001.

[Boykov 01b] Y. Boykov, O. Veksler & R. Zabih. *Fast approximate energy minimization via graph cuts.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 11, pp. 1222–1239, 2001.

[Brown 06] C.D. Brown & H.T. Davis. *Receiver operating characteristics curves and related decision measures: A tutorial.* Chemometrics and Intelligent Laboratory Systems, vol. 80, no. 1, pp. 24–38, 2006.

[Brox 04] T. Brox, A. Bruhn, N. Papenberg & J. Weickert. *High accuracy optical flow estimation based on a theory for warping.* In Proceedings of European Conference on Computer Vision, volume 3024 of *LNCS*, pp. 25–36, Prague, Czech Republic, May 2004.

[Brox 05] T. Brox. *From pixels to regions: Partial differential equation in image analysis.* PhD thesis, Dept. of Mathematics and Computer Science, Saarland University, Germany, 2005.

[Brox 06a] T. Brox, A. Bruhn & J. Weickert. *Variational motion segmentation with level sets.* In Proc. of European Conference on Computer Vision, volume 3951 of *LNCS*, pp. 471–483, Graz, Austria, May 2006.

[Brox 06b] T. Brox & J. Weickert. *Level set segmentation with multiple regions.* IEEE Transactions on Image Processing, vol. 15, no. 10, pp. 3213–3218, 2006.

[Bruhn 05a] A. Bruhn & J. Weickert. *Towards ultimate motion estimation: Combining highest accuracy with real-time performance.* In Proc. of IEEE International Conference on Computer Vision, pp. 749–755, Beijing, China, October 2005.

[Bruhn 05b] A. Bruhn, J. Weickert & C. Schnörr. *Lucas/Kanade meets Horn/ Schunck: combining local and global optic flow methods.* International Journal of Computer Vision, vol. 61, no. 3, pp. 1–21, 2005.

[Cai 07] W. Cai. *Fast and robust fuzzy C-means clustering algorithms incorporating local information for image segmentation.* Pattern Recognition, vol. 40, pp. 825–838, 2007.

[Callaghan 05] R.J. O' Callaghan & D.R. Bull. *Combined morphological-spectral unsupervised image segmentation.* IEEE Transactions on Image Processing, vol. 14, no. 1, pp. 49–62, 2005.

[Campilho 07] A. Campilho, M. Kamel & F.C. Monteiro. *Image segmentation: a review of recent contributions*, 2007. in preparation.

[Canny 86] J. Canny. *A computational approach to edge detection.* IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 8, no. 6, pp. 679–698, 1986.

[Cardoso 05] J.S. Cardoso & L. Corte-Real. *Toward a generic evaluation of image segmentation.* IEEE Transactions on Image Processing, vol. 14, no. 11, pp. 1773–1782, 2005.

[Carson 02] C. Carson, S. Belongie, H. Greenspan & J. Malik. *Blobworld: image segmentation using Expectation-Maximization and its application to image querying.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 8, pp. 1026–1038, 2002.

[Caselles 97] V. Caselles, R. Kimmel & G. Sapiro. *Geodesic active contours.* International Journal of Computer Vision, vol. 22, no. 1, pp. 61–79, 1997.

[Chalana 97] V. Chalana & Y. Kim. *A methodology for evaluation of boundary detection algorithms on medical images.* IEEE Transactions on Medical Imaging, vol. 16, no. 5, pp. 642–652, 1997.

[Chan 94] Y-L. Chan & X. Li. *Adaptive image region-growing.* IEEE Transactions on

Image Processing, vol. 3, no. 6, pp. 868–872, 1994.

[Chang 97] M.M. Chang, A.M. Tekalp & M.I. Sezan. *Simultaneous motion estimation and segmentation.* IEEE Transactions on Image Processing, vol. 6, no. 9, pp. 1326–1333, 1997.

[Chang 01] S-F. Chang, T. Sikora & A. Puri. *Overview of the MPEG-7 standard.* IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 6, pp. 688–695, 2001.

[Chen 04] S. Chen & D. Zhang. *Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure.* IEEE Trans. on Systems, Man and Cybernetics - Part B, vol. 34, no. 4, pp. 1907–1916, 2004.

[Cheng 01] H. Cheng, X. Jiang, Y. Sun & J. Wang. *Color image segmentation: advances and prospects.* Pattern Recognition, vol. 34, no. 12, pp. 2259–2281, 2001.

[Cheriet 98] M. Cheriet, J.N. Said & C.Y. Suen. *A recursive thresholding technique for image segmentation.* IEEE Transactions on Image Processing, vol. 7, no. 6, pp. 918–920, 1998.

[Christoudias 02] C.M. Christoudias, B. Georgescu & P. Meer. *Synergism in low level vision.* In Proc. of IEEE International Conference on Pattern Recognition, volume 4, pp. 150–155, Quebec, Canada, August 2002.

[Chu 93] C. Chu & J.K. Aggarwal. *The integration of image segmentation maps using region and edge information.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 12, pp. 1241–1252, 1993.

[Chung 05] R. Chung, N. Yung & P. Cheung. *An efficient parameterless quadrilat-eral-based image segmentation method.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 9, pp. 1446–1458, 2005.

[Chung 07] P-C. Chung, C-L. Huang & E-L. Chen. *A region-based selective optical flow back-projection for genuine motion vector estimation.* Pattern Recognition, vol. 40, no. 3, pp. 1066–1077, 2007.

[Comaniciu 02] D. Comaniciu & P. Meer. *Mean Shift: a robust approach toward feature space analysis.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 603–619, 2002. Software available at: *http://www.caip.rutgers.edu/riul/research/code/EDISON/.*

[Costeira 95] J. Costeira & T. Kanade. *A multi-body factorization method for motion analysis.* In Proc. of IEEE International Conf. on Computer Vision, pp. 1071–1076, Cambridge, Massachusetts, USA, June 1995.

[Cour 05] T. Cour, F. Benezit & J. Shi. *Spectral segmentation with multiscale graph decomposition.* In Proc. of IEEE International Conference on Computer Vision and Pattern Recognition, volume II, pp. 1124–1131, Washington, DC, USA,

June 2005. Software available at: *http://www.seas.upenn.edu/~timothee/software/ncut_multiscale/*.

[Cremers 05] D. Cremers & S. Soatto. *Motion competition: a variational framework for piecewise parametric motion segmentation.* International Journal of Computer Vision, vol. 62, no. 3, pp. 249–265, 2005.

[Cremers 07] D. Cremers, M. Rousson & R. Deriche. *A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape.* International Journal of Computer Vision, vol. 72, no. 2, pp. 195–215, 2007.

[D. Zhang 04] S. Chen D. Zhang. *A novel kernelized fuzzy C-means algorithm with application in medical image segmentation.* Artificial Intelligence in Medicine, vol. 32, no. 1, pp. 37–50, 2004.

[Darrell 95] T. Darrell & A.P. Pentland. *Cooperative robust estimation using layers of support.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, no. 5, pp. 474–487, 1995.

[Davis 75] L.S. Davis. *A survey of edge detection techniques.* Computer Graphics and Image Processing, vol. 4, pp. 248–270, 1975.

[Davis 06] J. Davis & M. Goadrich. *The relationship between Precision-Recall and ROC curves.* In Proceedings of the 23rd international conference on Machine learning, pp. 233–240, Pittsburgh, Pennsylvania, June 2006.

[Davison 00] N.E. Davison, H. Eviatar & R.L. Somorjai. *Snakes simplified.* Pattern Recognition, vol. 33, no. 10, pp. 1651–1664, 2000.

[De Bock 05] J. De Bock, P. De Smet & W. Philips. *Image segmentation using watersheds and normalized cuts.* In Proceedings of the IS&T/SPIE Electronic Imaging 2005, volume 5675, pp. 164–173, January 2005.

[De Smet 99] P. De Smet, R. Pires, D. De Vleeschauwer & I. Bruyland. *Activity driven nonlinear diffusion for color image watershed segmentation.* Journal of Electronic Imaging, vol. 8, no. 3, pp. 270–278, 1999.

[De Smet 00] P. De Smet & R. Pires. *Implementation and analysis of an optimized rainfalling watershed algorithm.* In Proceedings of the IS&T/SPIE Electronic Imaging 2000, volume 3974, pp. 759–766, January 2000.

[Delon 07] J. Delon, A. Desolneux, J-L. Lisani & A.B. Petro. *A nonparametric approach for histogram segmentation.* IEEE Transactions on Image Processing, vol. 16, no. 1, pp. 253–261, 2007.

[Dempster 77] A. Dempster, N. Laird & D. Rubin. *Maximum likelihood from incomplete data via the EM algorithm.* Journal of Royal Statistical Society, vol. B-39, no. 1, pp. 1–38, 1977.

[Deng 01] Y. Deng & B. Manjunath. *Unsupervised segmentation of color-texture re-*

gions in images and video. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 8, pp. 800–810, 2001. Software available at: *http://vision.ece.ucsb.edu/segmentation/jseg/ software/index.htm.*

[Duarte 06] A. Duarte, A. Sánchez, F. Fernández & A. Montemayor. *Improving image segmentation quality through effective region merging using a hierarchical social metaheuristic.* Pattern Recognition Letters, vol. 27, no. 11, pp. 1239–1251, 2006.

[Dufaux 95] F. Dufaux & F. Moscheni. *Spatio-temporal segmentation based on motion and static segmentation.* In Proc. IEEE Int. Conference on Image Processing, volume 1, pp. 306–309, Washington DC, USA, Oct. 1995.

[Duncan 00] J. Duncan & N. Ayache. *Medical image analysis: progress over two decades and the challenges ahead.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 85–108, 2000.

[Dupuis 06] A. Dupuis & P. Vasseur. *Image segmentation by cue selection and integration.* Image and Vision Computing, vol. 24, no. 10, pp. 1053–1064, 2006.

[Duta 98] N. Duta & M. Sonka. *Segmentation and interpretation of MR brain images: an improved active shape model.* IEEE Transactions on Medical Imaging, vol. 17, no. 6, pp. 1049–1062, 1998.

[Falcão 00] A.X. Falcão, J.K. Udupa & F.K. Miyazawa. *An ultra-fast user-steered image segmentation paradigm: Live Wire on the fly.* IEEE Transactions on Medical Imaging, vol. 19, no. 1, pp. 55–62, 2000.

[Fan 49] K. Fan. *On a theorem of Weyl concerning eigenvalues of linear transformations.* Proc. of the National Academy of Sciences, vol. 35, no. 11, pp. 652–655, 1949.

[Fan 01] J. Fan, D.K. Yau, A. Elmagarmid & W.G Aref. *Automatic image segmentation by integrating color-edge extraction and seeded region growing.* IEEE Transactions on Image Processing, vol. 10, no. 10, pp. 1454–1466, 2001.

[Fan 05] J. Fan, G. Zeng, M. Body & M. Hacid. *Seeded region growing: an extensive and comparative study.* Pattern Recognition Letters, vol. 26, no. 8, pp. 1139–1156, 2005.

[Farmer 05] M.E. Farmer & A.K. Jain. *A wrapper-based approach to image segmentation and classification.* IEEE Transactions on Image Processing, vol. 14, no. 12, pp. 2060–2072, 2005.

[Farnebäck 01] G. Farnebäck. *Very high accuracy velocity estimation using orientation tensors, parametric motion, and segmentation of the motion field.* In Proc. of International Conference on Computer Vision, volume 1, pp. 171–177, Vancouver, Canada, July 2001.

[Fawcett 06] T. Fawcett. *An introduction to ROC analysis.* Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, 2006.

[Felzenszwalb 04] P. Felzenszwalb & D. Huttenlocher. *Efficient graph-based image segmentation.* Int. Journal of Computer Vision, vol. 59, no. 2, pp. 167–181, 2004.

[Fowlkes 01] C. Fowlkes, S. Belongie & J. Malik. *Efficient spatio-temporal grouping using the Nyström method.* In Proc. of IEEE Computer Vision and Pattern Recognition, volume I, pp. 231–238, Hawaii, Dec. 2001.

[Fowlkes 04] C. Fowlkes, S. Belongie, F. Chung & J. Malik. *Spectral grouping using Nyström method.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 2, pp. 214–225, 2004.

[Freixenet 02] J. Freixenet, X. Muñoz, D. Raba, J. Martí & X. Cufí. *Yet another survey on image segmentation: region and boundary information integration.* In Proceedings of European Conference on Computer Vision, volume III, pp. 408–422, Copenhagen, Denmark, May 2002.

[Fukunaga 75] K. Fukunaga & L.D. Hostetler. *The estimation of the gradient of a density function, with applications in pattern recognition.* IEEE Transactions on Information Theory, vol. 21, no. 1, pp. 32–40, 1975.

[Galland 03] F. Galland, N. Bertaux & P. Refregier. *Minimum description length synthetic aperture radar image segmentation.* IEEE Transactions on Image Processing, vol. 12, no. 9, pp. 995–1006, 2003.

[Galun 05] M. Galun, A. Apartsin & R. Basri. *Multiscale segmentation by combining motion and intensity cues.* In Proc. of IEEE Computer Vision and Pattern Recognition, volume I, pp. 256–263, 2005.

[Gambotto 93] J.P. Gambotto. *A new approach to combining region growing and edge detection.* Pattern Rec. Letters, vol. 14, no. 11, pp. 869–875, 1993.

[Gauch 99] J. Gauch. *Image segmentation and analysis via multiscale gradient watershed hierarchies.* IEEE Trans. on Image Processing, vol. 8, no. 1, pp. 69–79, 1999.

[Gelasca 04] E.D. Gelasca, T. Ebrahimi, M.C.Q. Farias, M. Carli & S.K. Mitra. *Towards perceptually driven segmentation evaluation metrics.* In Proc. of IEEE Computer Vision and Pattern Recognition Workshop, volume 4, page 52, 2004.

[Gelgon 95] M. Gelgon & P. Bouthemy. *A region level graph labeling approach to motion based segmentation.* In Proc. IEEE Int. Conference on Computer Vision and Pattern Recognition, pp. 514–519, Puerto Rico, 1995.

[Germond 00] L. Germond, M. Dojat, C. Taylor & C. Garbay. *A cooperative framework for segmentation of MRI brain scans.* Artificial Intelligence in Medicine, vol. 20, no. 1, pp. 77–93, 2000.

[Gevers 02] T. Gevers. *Adaptive image segmentation by combining photometric invariant*

region and edge information. IEEE Transactions on Pattern Analysis And Machine Intelligence, vol. 24, no. 6, pp. 848–852, 2002.

[Goldberg 95] A. Goldberg & R. Kennedy. *An efficient cost scaling algorithm for the assignment problem.* Mathematical Programming, vol. 71, no. 2, pp. 153–177, 1995. Software available at: *http://www.avglab.com/andrew/soft.html.*

[Gonzalez 92] R. Gonzalez & R. Woods. Digital Image Processing. Addison-Wesley, 1992.

[Goumeidane 03] A.B. Goumeidane, M. Khamadja, B. Belaroussi, H. Benoit-Cattin & C. Odet. *New discrepancy measures for segmentation evaluation.* In Proc. of IEEE Int. Conference on Image Processing, volume II, pp. 411–414, 2003.

[Grady 06] L. Grady. *Random walks for image segmentation.* IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 28, no. 11, pp. 1768–1783, 2006.

[Grau 04] V. Grau, A. Mewes, M. Alcaniz, R. Kikinis & S.K. Warfield. *Improved watershed transform for medical image segmentation using prior information.* IEEE Transactions on Medical Imaging, vol. 23, no. 4, pp. 447–458, 2004.

[Han 03] X. Han, C. Xu & J.L. Prince. *A topology preserving level set method for geometric deformable models.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 6, pp. 755–768, 2003.

[Haralick 85] R. M. Haralick & L. Shapiro. *A survey on image segmentation.* Computer Vision and Image Processing, vol. 29, pp. 100–132, 1985.

[Haralick 94] R.M. Haralick. *Performance characterization protocol in computer vision.* In ARPA Image Understanding Workshop, volume I, pp. 667–673, Monterey, California, 1994.

[Haris 98] K. Haris, S.N. Efstratiadis, N. Maglaveras & A.K. Katsaggelos. *Hybrid image segmentation using watersheds and fast region merging.* IEEE Transactions on Image Processing, vol. 7, no. 12, pp. 1684–1699, 1998.

[Heath 97] M.D. Heath, S. Sarkar, T. Sanocki & K.W. Bowyer. *A robust visual method for assessing the relative performance of edge-detection algorithms.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 12, pp. 1338–1359, 1997.

[Held 97] K. Held, E.R. Kops, B.J. Krause, W.M. Wells, R. Kikinis & H.W. Muller-Gartner. *Markov random field segmentation of brain MR images.* IEEE Transactions on Medical Imaging, vol. 16, no. 6, pp. 878–886, 1997.

[Hernandez 00] S. Hernandez & K. Barner. *Joint region merging criteria for watershed-based image segmentation.* In Proc. of IEEE Int. Conference on Image Processing, volume II, pp. 108–111, Vancouver, Canada, Sep. 2000.

[Hill 03] P.R. Hill, C.N. Canagarajah & D.R. Bull. *Image segmentation using a texture gradient based watershed transform.* IEEE Transactions on Image Processing,

vol. 12, no. 12, pp. 1618–1633, 2003.

[Hoffman 87] R. Hoffman & A. K. Jain. *Segmentation and classification of range images.* IEEE Transactions on Pattern Analysis Machine Intelligence, vol. 9, no. 5, pp. 608–620, 1987.

[Hojjatoleslami 98] S.A. Hojjatoleslami & J. Kittler. *Region growing: a new approach.* IEEE Transactions on Image Processing, vol. 7, no. 7, pp. 1079–1084, 1998.

[Hoover 96] A. Hoover, G. Jean-Baptiste, X. Jiang, P. Flynn, H. Bunke, D. Goldgof, K. Bowyer, D. Eggert, A. Fitzgibbon & R. Fisher. *An experimental comparison of range image segmentation algorithms.* IEEE Transactions on Pattern Analysis Machine Intelligence, vol. 18, no. 7, pp. 673–689, 1996.

[Horn 81] B.K.P. Horn & B.G. Schunck. *Determining optical flow.* Artificial Intelligence, vol. 17, no. 1-3, pp. 185–203, 1981.

[Horowitz 76] S. Horowitz & T. Pavlidis. *Picture segmentation by a tree traversal algorithm.* Journal of the ACM, vol. 23, no. 2, pp. 368–388, 1976.

[Huang 95] Q. Huang & B. Dom. *Quantitative methods of evaluating image segmentation.* In Proc. IEEE Int. Conf. on Image Processing, volume III, pp. 53–56, 1995.

[Illingworth 88] J. Illingworth & J. Kittler. *A survey of the Hough transform.* Computer Vision, Graphics, and Image Processing, vol. 44, no. 1, pp. 87–116, 1988.

[Jain 99] A.K. Jain, M.N. Murty & P.J. Flynn. *Data clustering: a review.* ACM Computing Surveys, vol. 31, no. 3, pp. 264–323, 1999.

[Jain 00] A.K. Jain, R.P. Duin & J. Mao. *Statistical pattern recognition: A review.* IEEE Transactions on Pattern Analysis Machine Intelligence, vol. 22, no. 1, pp. 4–37, 2000.

[Ju 96] S. Ju, M.J. Black & A.D. Jepson. *Skin and Bones: multi-layer, locally affine, optical flow and regularization with transparency.* In Proc. of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 307–314, San Francisco, USA, June 1996.

[Kass 88] M. Kass, A. Witkin & D. Terzopoulos. *Snakes: active contour models.* Int. Journal of Computer Vision, vol. 1, no. 4, pp. 321–331, 1988.

[Kermad 02] C.D. Kermad & K. Chehdi. *Automatic image segmentation system through iterative edge–region co-operation.* Image and Vision Computing, vol. 20, no. 8, pp. 541–555, 2002.

[Kim 02] I-K. Kim, D-W. Jung & R-H. Park. *Document image binarization based on topographic analysis using a water flow model.* Pattern Recognition, vol. 35, no. 1, pp. 265–277, 2002.

[Kim 03] J-B. Kim & H-J. Kim. *Multiresolution-based watersheds for efficient image*

segmentation. Pattern Recognition Letters, vol. 24, no. 3, pp. 473–488, 2003.

[Konrad 92] J. Konrad & E. Dubois. *Bayesian estimation of motion vector fields.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 9, pp. 910–927, 1992.

[Krishnapuram 93] R. Krishnapuram & J. M. Keller. *A possibilistic approach to clustering.* IEEE Transactions on Fuzzy Systems, vol. 1, pp. 98–110, 1993.

[Krishnapuram 96] R. Krishnapuram & J. M. Keller. *The possibilistic C-means algorithm: Insights and recommendations.* IEEE Transactions on Fuzzy Systems, vol. 4, pp. 385–392, 1996.

[Kwok 97] S. Kwok & A. Constantinides. *fast recursive shortest spanning tree for image segmentation and edge detection.* IEEE Transactions on Image Processing, vol. 6, no. 2, pp. 328–332, 1997.

[Lai 05] S.H. Lai & B. Vemuri. *Reliable and efficient computation of optical flow.* International Journal of Computer Vision, vol. 29, no. 2, pp. 87–105, 2005.

[Levine 85] M.D. Levine & A.M. Nazif. *Dynamic measurement of computer generated image segmentations.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 7, no. 2, pp. 155–164, 1985.

[Levner 07] I. Levner & H. Zhang. *Classification-driven watershed segmentation.* IEEE Transactions on Image Processing, vol. 16, no. 5, pp. 1437–1445, 2007.

[Lezoray 03] O. Lezoray & H. Cardot. *Hybrid color image segmentation using 2D histogram clustering and region merging.* In Proc. of IEEE Int. Conf. on Image and Signal Processing, volume 1, pp. 22–29, Agadir, Morocco, June 2003.

[Li 97] L. Li, J. Gong & W. Chen. *Gray-level image thresholding based on Fisher linear projection of twodimensional histogram.* Pattern Recognition, vol. 30, no. 5, pp. 743–749, 1997.

[Li 04] Y. Li, J. Sun, C-K. Tang & H-Y. Shum. *Lazy snapping.* ACM Transactions on Graphics, vol. 23, no. 3, pp. 303–308, 2004.

[Li 06] Y-M. Li & X-P. Zeng. *A new strategy for urinary sediment segmentation based on wavelet, morphology and combination method.* Comp. Methods and Programs in Biomedicine, vol. 84, no. 2-3, pp. 162–173, 2006.

[Liu 03] M.G. Liu, J. Jiang & C.H. Hou. *Hybrid image segmentation using watersheds and adaptive region growing.* In Proc. Int. Conference on Visual Information Engineering, pp. 282–285, Guildford, UK, July 2003.

[Lucas 81] B.D. Lucas & T. Kanade. *An iterative image registration technique with an application to stereo vision.* In Proc. of the 7th Int. Joint Conference on Artificial Intelligence, pp. 674–679, Vancouver, Canada, April 1981.

[Lucchese 01] L. Lucchese & S.K. Mitra. *Color image segmentation: A state-of-the-art survey.* Proc. of the Indian National Science Academy, vol. 67, no. 2, pp. 207–221, 2001.

[Luo 04] Q. Luo & T.M. Khoshgoftaar. *Efficient image segmentation by mean shift clustering and MDL-guided region merging.* In Proceedings of IEEE International Conference on Tools with Artificial Intelligence, pp. 337–343, 2004.

[Ma 00] W-Y. Ma & B.S. Manjunath. *EdgeFlow: a technique for boundary detection and image segmentation.* IEEE Transactions on Image Processing, vol. 9, no. 8, pp. 1375–1388, 2000.

[Makrogiannis 05] S. Makrogiannis, G. Economou, S. Fotopoulos & N.G. Bourbakis. *Segmentation of color images using multiscale clustering and graph theoretic region synthesis.* IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 35, no. 2, pp. 224–238, 2005.

[Malik 01] J. Malik, S. Belongie, T. Leung & J. Shi. *Contour and texture analysis for image segmentation.* International Journal of Computer Vision, vol. 43, no. 1, pp. 7–27, 2001.

[Marroquin 03] J. Marroquin, E. Santana & S. Botello. *Hidden Markov measure field models for image segmentation.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 11, pp. 1380–1387, 2003.

[Martin 01] D. Martin, C. Fowlkes, D. Tal & J. Malik. *A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics.* In Proc. of IEEE Int. Conference on Computer Vision, volume II, pp. 416–423, 2001. Online at: *http://www.cs.berkeley.edu/projects/vision/grouping/segbench/.*

[Martin 02] D. Martin. *An empirical approach to grouping and segmentation.* PhD thesis, University of California, Berkeley, 2002.

[Martin 04] D. Martin, C. Fowlkes & J. Malik. *Learning to detect natural image boundaries using local brightness, color, and texture cues.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 5, pp. 530–549, 2004.

[McCane 97] B. McCane. *On the evaluation of image segmentation algorithms.* In Proc. of Digital Image Computing: Techniques and Applications, pp. 455–460, 1997.

[McInerney 95] T. McInerney & D. Terzopolos. *A dynamic finite element surface model for segmentation and tracking in multidimensional medical images with application to cardiac 4D image analysis.* Computerized Medical Imaging and Graphics, vol. 19, no. 1, pp. 39–83, 1995.

[McInerney 96] T. McInerney & D. Terzopolos. *Deformable models in medical image analysis: A survey.* Medical Image Analysis, vol. 1, no. 2, pp. 91–108, 1996.

[McInerney 00] T. McInerney & D. Terzopolos. *T-snakes: topology adaptive snakes.* Medical Image Analysis, vol. 4, no. 2, pp. 73–91, 2000.

[Mehnert 97] A. Mehnert & P. Jackway. *An improved seeded region growing.* Pattern Recognition Letters, vol. 18, pp. 1065–1071, 1997.

[Meila 01] M. Meila & J. Shi. *A random walk view of spectral segmentation.* In Proc. of the International Workshop on Artificial Intelligence and Statistics, 2001.

[Mendonça 06] A.M. Mendonça & A. Campilho. *Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction.* IEEE Transactions on Medical Imaging, vol. 25, no. 9, pp. 1200–1213, 2006.

[Meyer 94] F. Meyer. *Topographic distance and watershed lines.* Signal Processing, vol. 38, no. 1, pp. 113–125, 1994.

[Mémin 98] E. Mémin & P. Pérez. *Dense estimation and object-based segmentation of the optical flow with robust techniques.* IEEE Transactions on Image Processing, vol. 7, no. 5, pp. 703–719, 1998.

[Mémin 02] E. Mémin & P. Pérez. *Hierarchical estimation and segmentation of dense motion fields.* International Journal of Computer Vision, vol. 46, no. 2, pp. 129–155, 2002.

[Moga 97] A.N. Moga & M. Gabbouj. *Parallel image component labeling with watershed transformation.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 5, pp. 441–450, 1997.

[Monteiro 06] F.C. Monteiro & A. Campilho. *Performance evaluation of image segmentation.* In Proc. International Conference on Image Analysis and Recognition, volume 4141 of *LNCS*, pp. 248–259, Póvoa de Varzim, Portugal, Sept. 2006.

[Monteiro 07] F.C. Monteiro & A. Campilho. *Combining watershed and multi-class spectral methods for image segmentation.* Image and Vision Computing, 2007. submitted.

[Mortensen 99] E.N. Mortensen & W.A. Barrett. *Toboggan-based intelligent scissors with a four parameter edge model.* In Proc. of IEEE International Conf. on Computer Vision and Pattern Recognition, volume II, pp. 452–458, Fort Collins, USA, June 1999.

[MPEG4 99] MPEG4. *MPEG-4 video verification model, Version 15.0.* ISO/IEC/JTC1/SC29/WG11 N3093, 1999.

[Muñoz 03] X. Muñoz, J. Freixenet, X. Cufí; & J. Martí. *Strategies for image segmentation combining region and boundary information.* Pattern Recognition Letters, vol. 24, no. 1-3, pp. 375–392, 2003.

[Nagel 86] H.H. Nagel & W. Enkelmann. *An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences.* IEEE

Transactions on Pattern Analysis and Machine Intelligence, vol. 8, no. 5, pp. 565–593, 1986.

[Ng 02] A. Ng, M. Jordan & Y. Weiss. *On spectral clustering: analysis and an algorithm.* In Advances in Neural Information Processing Systems 14, pp. 849–856, 2002.

[Niessen 98] W.J. Niessen, B.M. ter Haar Romeny & M.A. Viergever. *Geodesic deformable models for medical image analysis.* IEEE Transactions on Medical Imaging, vol. 17, no. 4, pp. 634–641, 1998.

[Odet 02] C. Odet, B. Belaroussi & H.B. Cattin. *Scalable discrepancy measures for segmentation evaluation.* In Proceedings of International Conference on Image Processing, volume I, pp. 785–788, 2002.

[Odobez 98] J-M. Odobez & P. Bouthemy. *Direct incremental model-based image motion segmentation for video analysis.* Signal Processing, vol. 66, pp. 143–155, 1998.

[Ogor 95] B. Ogor, V. Haese-Coat & K. Kpalma. *Cooperation of mathematical morphology and region growing for remote sensing image segmentation.* In Proceedings of SPIE: Image and Signal Processing for Remote Sensing, volume 2579, pp. 375–386, November 1995.

[Olabarriaga 01] S.D. Olabarriaga & A.M. Smeulders. *Interaction in the segmentation of medical images: a survey.* Medical Image Analysis, vol. 2, no. 5, pp. 127–142, 2001.

[Ortiz 06] A. Ortiz & G. Oliver. *On the use of the overlapping area matrix for image segmentation evaluation: A survey and new performance measures.* Pattern Recognition Letters, vol. 27, no. 16, pp. 1916–1926, 2006.

[Otsu 79] N. Otsu. *A threshold selection method from grey-level histograms.* IEEE Trans. on Systems, Man and Cybernetics, vol. 9, no. 1, pp. 62–66, 1979.

[Pal 93] N.R. Pal & S.K. Pal. *A review on image segmentation techniques.* Pattern Recognition, vol. 26, no. 9, pp. 1277–1294, 1993.

[Palmer 96] P.L. Palmer, H. Dabis & J. Kittler. *A performance measure for boundary detection algorithms.* Computer Vision and Image Understanding, vol. 63, no. 3, pp. 476–494, 1996.

[Pan 03] C. Pan, C-X. Zheng & H-J. Wang. *Robust color image segmentation based on mean shift and marker-controlled watershed algorithm.* In Proceedings of IEEE International Conference on Machine Learning and Cybernetics, volume 5, pp. 2752–2756, November 2003.

[Panjwani 95] D. Panjwani & G. Healey. *Markov random field models for unsupervised segmentation of textured color images.* IEEE Transactions on Pattern Analysis

and Machine Intelligence, vol. 17, no. 10, pp. 939–954, 1995.

[Pantofaru 05] C. Pantofaru & M. Hebert. *A Comparison of image segmentation algorithms.* Technical report CMU-RI-TR-05-40, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, September 2005.

[Papamarkos 94] N. Papamarkos & B. Gatos. *A new approach for multilevel threshold selection.* Computer Vision, Graphics, and Image Processing, vol. 56, no. 5, pp. 357–370, 1994.

[Papenberg 06] N. Papenberg, A. Bruhn, T. Brox, S. Didas & J. Weickert. *Highly accurate optic flow computation with theoretically justified warping.* Int. Journal of Computer Vision, vol. 67, no. 2, pp. 141–158, 2006.

[Pappas 92] T.N. Pappas. *An adaptive clustering algorithm for image segmentation.* IEEE Transactions on Signal Processing, vol. 40, no. 4, pp. 901–914, 1992.

[Paragios 00] N. Paragios & R. Deriche. *Geodesic active contours and level sets for the detection and tracking of moving objects.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 3, pp. 266–280, 2000.

[Paragios 02] N. Paragios & R. Deriche. *Geodesic active regions and level set methods for supervised texture segmentation.* International Journal on Computer Vision, vol. 46, no. 3, pp. 223–247, 2002.

[Paragios 03] N. Paragios. *A level set approach for shape-driven segmentation and tracking of the left ventricle.* IEEE Transactions on Medical Imaging, vol. 22, no. 6, pp. 773–776, 2003.

[Pateux 00] S. Pateux. *Spatial segmentation of color images according to the MDL formalism.* In Proc. of IEEE International Conference on Image Processing, volume 2, pp. 92–95, 2000.

[Pavlidis 90] T.Y. Pavlidis & T. Liow. *Integrating region growing and edge detection.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 3, pp. 225–233, 1990.

[Peleg 89] S. Peleg, M. Werman & H. Rom. *A unified approach to the change of resolution: Space and gray-level.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 7, pp. 739–742, 1989.

[Perona 90] P. Perona & J. Malik. *Scale-space and edge detection using anisotropic diffusion.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 7, pp. 629–638, 1990.

[Perona 98] P. Perona & W. Freeman. *A factorization approach to grouping.* In Proceedings of European Conference on Computer Vision, volume 1406 of *LNCS*, pp. 655–670, Freiburg, Germany, June 1998.

[Petersen 02] M. Egmont- Petersen, D. de Ridder & H. Handels. *Image processing with*

neural networks-a review. Pattern Recognition, vol. 35, no. 10, pp. 2279–2301, 2002.

[Pham 02] D.L. Pham. *Fuzzy clustering with spatial constraints.* In Proc. of the International Conference Image Processing, volume II, pp. 65–68, 2002.

[Protiere 07] A. Protiere & G. Sapiro. *Interactive image segmentation via adaptive weighted distances.* IEEE Transactions on Image Processing, vol. 16, no. 4, pp. 1046–1057, 2007.

[Puzicha 99] J. Puzicha, T. Hofmann & J.M. Buhmann. *Histogram clustering for unsupervised segmentation and image retrieval.* Pattern Recognition Letters, vol. 20, no. 8, pp. 899–909, 1999.

[Raghavan 89] V. Raghavan, P. Bollmann & G. Jung. *A critical investigation of recall and precision as measures of retrieval system performance.* ACM Transactions on Information Systems, vol. 7, no. 3, pp. 205–229, 1989.

[Reed 93] T. Reed & J. du Buf. *A review of recent texture segmentation and feature extraction techniques.* Computer Vision Image Understanding, vol. 57, no. 3, pp. 379–372, 1993.

[Rees 02] G.S. Rees, W.A. Wright & P. Greenway. *ROC method for the evaluation of multi-class segmentation/classification algorithms with infrared imagery.* In Proc. of the British Machine Vision Conference, pp. 537–546, Cardiff, UK, Sept. 2002.

[Ren 03] X. Ren & J. Malik. *Learning a classification model for segmentation.* In Proc. of IEEE International Conference on Computer Vision, volume 1, pp. 10–17, Nice, France, October 2003.

[Robles-Kelly 02] A. Robles-Kelly & E. Hancock. *An Expectation-maximisation framework for segmentation and grouping.* Image and Vision Computing, vol. 20, no. 9-10, pp. 725–738, 2002.

[Roerdink 01] J.B. Roerdink & A. Meijster. *The watershed transform: definitions, algorithms and parallelization strategies.* Fundamenta Informaticae, vol. 41, no. 1-2, pp. 187–228, 2001.

[Rother 04] C. Rother, V. Kolmogorov & A. Blake. *GRABCUT: interactive foreground extraction using iterated graph cuts.* ACM Transactions on Graphics, vol. 23, no. 3, pp. 309–314, 2004.

[Rubner 00] Y. Rubner, C. Tomasi & L.J. Guibas. *The Eearth Mover's Distance as a metric for image retrieval.* International Journal of Computer Vision, vol. 40, no. 2, pp. 99–121, 2000.

[Rui 96] Y. Rui, A. She & T. Huang. *Automated region segmentation using attraction-based grouping in spatial-color-texture space.* In Proc. of IEEE Int. Conf. on

Image Processing, volume 1, pp. 53–56, Lausanne, Switzerland, Sept. 1996.

[Ruzon 00] M. Ruzon & C. Tomasi. *Alpha estimation in natural images.* In Proc. of IEEE International Conference on Computer Vision and Pattern Recognition, volume I, pp. 18–25, Hilton Head Island, June 2000.

[Ruzon 01] M. Ruzon & C. Tomasi. *Edge, junction, and corner detection using color distributions.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 11, pp. 1281–1295, 2001.

[Sahoo 88] P.K. Sahoo, S. Soltani, A.K.C. Wang & Y.C. Chen. *A Survey of thresholding techniques.* Computer Vision, Graphics, and Image Processing, vol. 41, no. 2, pp. 233–260, 1988.

[Salembier 00] P. Salembier & L. Garrido. *Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval.* IEEE Transactions on Image Processing, vol. 9, no. 4, pp. 561–576, 2000.

[Sanfeliu 02] A. Sanfeliu, R. Alquezar, J. Andrade, J. Climent, F. Serratosa & J. Verges. *Graph-based representations and techniques for image processing and image analysis.* Pattern Recognition, vol. 35, no. 3, pp. 639–650, 2002.

[Sapiro 93] G. Sapiro & A. Tannenbaum. *Affine invariant scale-space.* International Journal of Computer Vision, vol. 11, no. 1, pp. 25–44, 1993.

[Sarkar 00] A. Sarkar, M.K. Biswas & K.M. Sharma. *A simple unsupervised MRF model based image segmentation approach.* IEEE Transactions on Image Processing, vol. 9, no. 5, pp. 801–812, 2000.

[Sclaroff 01] S. Sclaroff & L. Liu. *Deformable shape detection and description via model-based region grouping.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 5, pp. 475–489, 2001.

[Scott 87] D.S. Scott. *Solving sparse symmetric generalized eigenvalue without factorization.* SIAM Journal of Numerical Analysis, vol. 18, pp. 102–110, 1987. Software available at: *http://www.netlib.org/laso/snlaso.f.*

[Scott 90] G.L. Scott & H.C. Longuet-Higgins. *Feature grouping by "relocalisation" of eigenvectors of the proximity matrix.* In Proceedings of British Machine Vision Conference, pp. 103–108, 1990.

[Sezgin 04] M. Sezgin & B. Sankur. *Survey over image thresholding techniques and quantitative performance evaluation.* Journal of Electronic Imaging, vol. 13, no. 1, pp. 146–168, 2004.

[Shaffrey 02] C. Shaffrey, I. Jermyn & N. Kinsbury. *Psychovisual evaluation of image segmentation algorithms.* In Proc. of Advanced Concepts for Intelligent Vision Systems, Ghent university, Belgium, Sept. 2002.

[Sharon 00] E. Sharon, A. Brandt & R. Basri. *Fast multiscale image segmentation.* In

Proc. of IEEE International Conference on Computer Vision and Pattern Recognition, volume I, pp. 339–344, June 2000.

[Sharon 01] E. Sharon, A. Brandt & R. Basri. *Segmentation and boundary detection using multiscale intensity measurements.* In Proc. of IEEE International Conference on Computer Vision and Pattern Recognition, volume I, pp. 469–476, June 2001.

[Shepard 87] R.N. Shepard. *Towards a universal law of generalization for psychological science.* Science, vol. 237, pp. 1317–1323, 1987.

[Shi 98] J. Shi & J. Malik. *Motion segmentation and tracking using normalized cuts.* In Proc. of IEEE Int. Conference on Computer Vision, pp. 1154–1160, 1998.

[Shi 00] J. Shi & J. Malik. *Normalized cuts and image segmentation.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 888–905, 2000.

[Shum 03] H-Y. Shum, S.B. Kang & S-C. Chan. *Survey of image-based representations and compression techniques.* IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 11, pp. 1020–1037, 2003.

[Skudlarski 99] P. Skudlarski, R.T. Constable & J.C. Gore. *ROC analysis of statistical methods used in functional MRI: Individual subjects.* NeuroImage, vol. 9, no. 3, pp. 311–329, 1999.

[Smith 97] S. Smith & J. Brady. *SUSAN - A new approach to low level image processing.* International Journal of Computer Vision, vol. 23, no. 1, pp. 45–78, 1997. Software available at: *http://www.fmrib.ox.ac.uk/~steve.*

[Smith 01] P. Smith. *Edge-based motion segmentation.* PhD thesis, Department of Engineering, University of Cambridge, 2001.

[Smith 04] P. Smith, T. Drummond & R. Cipolla. *Layered motion segmentation and depth ordering by tracking edges.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 4, pp. 479–494, 2004.

[Solihin 99] Y. Solihin & C.G. Leedham. *Integral ratio: A new class of global thresholding techniques for handwriting images.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no. 8, pp. 761–768, 1999.

[Sorenson 05] J.A. Sorenson & X. Wang. *ROC methods for evaluation of fMRI techniques.* Mag. Resonance in Medicine, vol. 36, no. 5, pp. 737–744, 2005.

[Soundararajan 03] P. Soundararajan & S. Sarkar. *An in-depth study of graph partitioning measures for perceptual organization.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 6, pp. 642–660, 2003.

[Stehman 97] S.V. Stehman. *Selecting and interpreting measures of thematic classification accuracy.* Remote Sensing of Environment, vol. 62, no. 1, pp. 77–89, 1997.

[Stoev 00] S. Stoev & W. Strasser. *Extracting regions of interest applying a local watershed transformation.* In Proc. of the Annual IEEE Conference on Visualization, pp. 21–28, Salt Lake City, Utah USA, Oct. 2000.

[Suetake 07] N. Suetake, E. Uchino & K. Hirata. *Separability-based intelligent scissors for interactive image segmentation.* IEICE - Transactions on Information and Systems, vol. E90-D, no. 1, pp. 137–144, 2007.

[Szilagyi 03] L. Szilagyi, Z. Benyo, S.M. Szilagyi & H.S. Adam. *MR brain image segmentation using an enhanced fuzzy C-means algorithm.* In Proc. IEEE International Conference on Engineering in Medicine and Biology Society, volume 1, pp. 724–726, Sept. 2003.

[Tan 01] K-H. Tan & N. Ahuja. *Selecting objects with freehand sketches.* In Proc. IEEE International Conference on Computer Vision, pp. 337–344, Vancouver, Canada, July 2001.

[Tekalp 98] A.M. Tekalp, Y. Altunbasak & P.E. Eren. *Region based parametric motion segmentation using color information.* Journal of Graphical Models and Image Processing, vol. 60, no. 1, pp. 13–23, 1998.

[Thompson 80] W.B. Thompson. *Combining motion and contrast for segmentation.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 2, no. 6, pp. 543–549, 1980.

[Tomasi 98] C. Tomasi & R. Manduchi. *Bilateral filtering for gray and color images.* In Proc. of IEEE International Conference on Computer Vision, pp. 839–846, Bombay, India, January 1998.

[Torr 95] H. Torr. *Motion segmentation and outlier detection.* PhD thesis, Dept. of Engineering Science, University of Oxford, 1995.

[Tsaig 02] Y. Tsaig & A. Averbuch. *Automatic segmentation of moving objects in video sequences: a region labelling approach.* IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 7, pp. 597–612, 2002.

[Turi 01] R.H. Turi. *Clustering-based colour image segmentation.* PhD thesis, Monash University, Australia, 2001.

[Udupa 96] J. Udupa & S. Samarasekera. *Fuzzy connectedness and object definition: theory, algorithms, and applications in image segmentation.* Graphical Models and Image Processing, vol. 58, no. 3, pp. 246–261, 1996.

[Veksler 00] O. Veksler. *Image segmentation by nested cuts.* In Proc. IEEE International Conference on Computer Vision and Pattern Recognition, volume I, pp. 339–344, June 2000.

[Vincent 91] L. Vincent & P. Soille. *Watersheds in digital spaces: an efficient algorithm based on immersion simulations.* IEEE Transactions on Pattern Analysis

and Machine Intelligence, vol. 13, no. 6, pp. 583–589, 1991.

[Wan 03] S.Y. Wan & W.E. Higgins. *Symmetric region growing.* IEEE Transactions on Image Processing, vol. 12, no. 9, pp. 1007–1015, 2003.

[Wang 84] S. Wang & R. Haralick. *Automatic multi-threshold selection.* Computer Vision Graphics and Image Processing, vol. 25, no. 1, pp. 46–67, 1984.

[Wang 94] J. Wang & E. Adelson. *Representing moving images with layers.* IEEE Transactions on Image Processing, vol. 3, no. 5, pp. 625–638, 1994.

[Wang 01] S. Wang & J. Siskind. *Image segmentation with minimum mean cut.* In Proc. of IEEE International Conference on Computer Vision, volume I, pp. 517–524, Vancouver, Canada, July 2001.

[Wang 03a] L. Wang & J. Bai. *Threshold selection by clustering gray levels of boundary.* Pattern Recognition Letters, vol. 24, no. 12, pp. 1983–1999, 2003.

[Wang 03b] S. Wang & J. Siskind. *Image segmentation with ratio cut.* IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 25, no. 6, pp. 675–690, 2003.

[Wang 04a] H. Wang & D. Suter. *Robust adaptive-scale parametric model estimation for computer vision.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 11, pp. 1459–1474, 2004.

[Wang 04b] X. Wang, Y. Wang & L. Wang. *Improving fuzzy C-means clustering based on feature-weight learning.* Pattern Recognition Letters, vol. 25, no. 10, pp. 1123–1132, 2004.

[Wang 05] J. Wang & M.F. Cohen. *An iterative optimization approach for unified image segmentation and matting.* In Proc. of IEEE Int. Conference on Computer Vision, volume 2, pp. 936–943, Beijing, China, Oct. 2005.

[Weickert 01] J. Weickert. *Efficient image segmentation using partial differential equations and morphology.* Pattern Recognition, vol. 34, pp. 1813–1824, 2001.

[Weiss 96] Y. Weiss & E.H. Adelson. *A unified mixture framework for motion segmentation: incorporating spatial coherence and estimating the number of models.* In Proc. of IEEE Int. Conference on Computer Vision and Pattern Recognition, pp. 321–326, San Francisco, USA, June 1996.

[Weiss 97] Y. Weiss. *Smoothness in layers: motion segmentation using nonparametric mixture estimation.* In Proc. of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 520–527, Puerto Rico, June 1997.

[Weiss 99] Y. Weiss. *Segmentation using eigenvectors: a unifying view.* In Proc. IEEE Int. Conference on Computer Vision, pp. 975–982, Kerkyra, Greece, 1999.

[Wertheimer 38] M. Wertheimer. Laws of organization in perceptual forms (translation), pp. 71–88. A Sourcebook of Gestalt Psychology. W. B. Ellis Editor, 1938.

[Will 00] S. Will, L. Hermes, J.M. Buhmann & J. Puzicha. *On learning texture edge detectors.* In Int. Conf. on Image Processing, volume III, pp. 877–880, 2000.

[Won 92] C.S. Won & H. Derin. *Unsupervised segmentation of noisy and textured images using Markov random fields.* Computer Vision, Graphics and Image Processing, vol. 54, no. 4, pp. 308–328, 1992.

[Wu 93] Z. Wu & R. Leahy. *An optimal graph theoretic approach to data clustering: theory and its application to image segmentation.* IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 15, no. 11, pp. 1101–1113, 1993.

[Xiao 06] J. Xiao, H. Chen, H. Sawhney, C. Rao & M. Isnardi. *Bilateral filtering-based optical flow estimation with occlusion detection.* In Proc. of the European Conference on Computer Vision, volume I, pp. 211–224, 2006.

[Xu 98] C. Xu & J.L. Prince. *Snakes, shapes, and gradient vector flow.* IEEE Transactions on Image Processing, vol. 7, no. 3, pp. 359–369, 1998.

[Xu 00] C. Xu, D.L. Pham & J.L. Prince. Handbook of medical imaging - medical image processing and analysis, volume 2, pp. 129–174. SPIE Press, 2000.

[Xu 04] M. Xu, P.M. Thompson & A.W. Toga. *An adaptive level set segmentation on a triangulated mesh.* IEEE Transactions on Medical Imaging, vol. 23, no. 2, pp. 191–201, 2004.

[Yasnoff 77] W.A. Yasnoff, J.K. Mui & J.W. Bacus. *Error measures in scene segmentation.* Pattern Recognition, vol. 9, no. 4, pp. 217–231, 1977.

[Yedidia 02] J.S. Yedidia, W.T. Freeman & Y. Weiss. *Understanding belief propagation and its generalizations.* Technical report TR-2001-22, Mitsubishi Electric Research Lab., Cambridge, Massachusets, USA, Jan. 2002.

[Yeung 02] D.S. Yeung & X.Z. Wang. *Improving performance of similarity-based clustering by feature weight learning.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 4, pp. 556–561, 2002.

[Yoshida 02] T. Yoshida. *Distance metric for motion vector histograms based on human perceptual characteristics.* In Proc. of IEEE International Conference on Image Processing, volume I, pp. 904–907, 2002.

[Yu 99] Y. Yu & J. Wang. *Image segmentation based on region growing and edge detection.* In Proc. of IEEE International Conference on Systems, Man and Cybernetics, volume 6, pp. 798–803, Tokyo, Japan, Oct. 1999.

[Yu 03] S. Yu & J. Shi. *Multiclass spectral clustering.* In Proc. of IEEE International Conference on Computer Vision, pp. 313–319, Nice, France, Oct. 2003.

[Yu 04] S. Yu. *Segmentation using multiscale cues.* In Proc. of IEEE International Conference on Computer Vision and Pattern Recognition, volume 1, pp. 247–254, Washington DC, USA, June 2004.

[Zadeh 65] L. Zadeh. *Fuzzy sets.* Information and Control, vol. 8, pp. 338–353, 1965.

[Zeng 04] W. Zeng & W. Gao. *Accurate moving object segmentation by a hierarchical region labeling approach.* In Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, volume III, pp. 637–640, Montreal, Canada, May 2004.

[Zhang 96] Y.J. Zhang. *A survey on evaluation methods for image segmentation.* Pattern Recognition, vol. 29, no. 8, pp. 1335–1346, 1996.

[Zhang 01a] D. Zhang & G. Lu. *Segmentation of moving objects in image sequence: A review.* Circuits, Systems, and Signal Proc., vol. 20, no. 2, pp. 143–183, 2001.

[Zhang 01b] Y. Zhang, M. Brady & S. Smith. *Segmentation of brain MR images through a hidden Markov random field model and the Expectation-Maximization algorithm.* IEEE Trans. on Medical Imaging, vol. 20, no. 1, pp. 45–57, 2001.

[Zhou 05] H. Zhou, T. Liu, H. Hu, Y. Pang, F. Lin & J. Wu. *A hybrid framework for image segmentation.* In Proc. of IEEE Int. Conference on Acoustics, Speech, and Signal Processing, volume II, pp. 749–752, Philadelphia, USA, 2005.

[Zhu 96] S.C. Zhu & A. Yuille. *Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, no. 9, pp. 884–900, 1996.

[Zitova 03] B. Zitova & J. Flusser. *Image registration methods: A survey.* Image and Vision Computing, vol. 21, no. 11, pp. 977–1000, 2003.

[Zou 05] K.H. Zou. *Receiver operating characteristic (ROC) literature research*, 2005. Available at: *http://splweb.bwh.harvard.edu:8000/~zou/roc.html.*

[Zucker 76] S.W. Zucker. *Region growing: childhood and adolescence.* Computer Graphics and Image Processing, vol. 5, pp. 382–399, 1976.