Universitat Autònoma de Barcelona

Unsupervised image segmentation based on material reflectance description and saliency.

A dissertation submitted by **Eduard Vazquez Marco** at Universitat Autònoma de Barcelona to fulfil the degree of **Doctor en Informàtica**.

Bellaterra, 2011

Director: **Dr. Ramon Baldrich**
Universitat Autònoma de Barcelona
Dep. Ciències de la Computació & Centre de Visió per Computador

This document was typeset by the author using LaTeX 2_ε.

The research described in this book was carried out at the Computer Vision Center, Universitat Autònoma de Barcelona.

ISBN:

Printed by Ediciones Gráficas Rey, S.L.

'Porque los catalanes siempre hablan de lo mismo, es decir, de trabajo... No hay en la Tierra gente ms aficionada al trabajo que los catalanes. Si supieran hacer algo, serían los amos del mundo.'
Eduardo Mendoza. *Sin noticias de Gurb*

Navega por los resultados o realiza una nueva consulta en el formulario de la parte superior.
Web anònima; en clara referència a la metodologia científica.

Agraïments

Aquesta tesis no hagués estat possible sense l'intervenció de l'atzar. Èsser sempre tan il·lustre i afable com, ja per definició, d'imprevisible bonhomia. Què és el que fa que una tesis rutlli com un rellotge que funciona bé? Encert en les decisions? Potser. Serietat? Déu ens n'atorgui! Treballar dur? No està malament. Tanmateix, aquesta llista podria ser tan extensa com una tesis, o fins i tot, com la Santa i sagrada bíblia, el Quijote, o els Pilars de la Terra (el maligne els tingui en compte). Però com que la tesis no va sobre un llistat de virtuts de les que, sense cap mena de dubte, n'estan folrats tots els doctorands d'aquest centre, amb la comptada excepció de cap d'ells, em limitaré a finalitzar tal llistat amb un element que, de fet, ja ha estat apuntat: l'atzar, el gran oblidat. Ja no es tracta de que haguem arribat a bon port, es tracta de la immensa (i de moment impossible de modelar) casuística que em va dur a acabar sent professor i doctorand, quan la gent de baixa casta com un humil servidor està cridada a un destí molt menys avantatjós. Des d'aquell moment en el que en Marçal Rossinyol i l'Eduard Vàzquez van precaritzar el seu futur amb una beca de pigmeus ingressos, fins al moment en el que el grup Color i Textura, va dipositar coses en mi, han hagut de passar un grapat de coses. Així doncs, només m'agradaria emetre un exaltat crit a favor de la sort i del tarannà de les persones de bon cor, si és que això tingués algun sentit en aquest paràgraf. Juntament amb l'atzarós amic, caldria fer referència, i fer-ho per sincera i profunda deferència, que no és poca cosa, al meu director de tesis: el molt excel·lent Doctor Ramon Baldrich, a qui, val a dir, la fortuna, (i entenem-la com una medieval manera d'anomenar l'atzar), va deparar certs mals de cap derivats de la gestió duta a terme per a aquest redactor d'agraïments (i de coses en general que ara tampoc venen a tema). De totes maneres, li estic profundament agraït per la confiança dipositada en la meva persona, la qual, sense comptar amb rebut ni similar, podria haver estat com un xec de molts calers al portador que en comptes de al seu destinatari original acabés en mans d'un Sabadellenc, o, pitjor encara, d'algú de Barcelona. Juntament amb en Ramon, em veig obligat (literalment, doncs temo les represàlies més que un claustrofòbic un ascensor de l'eixample), a esmentar a la molt il·lustre Doctora Maria Vanrell. Dona, sí, però tanmateix jefa. Amb ella, més enllà de tesis, també he viscut uns anys d'interessentisísima (notis, per favor, el doble superlatiu), docència en Intel·ligència Artificial. Un camp que, més enllà de ser interessant, queda molt bé quan ho dius a les amistats i derivats.

És moment, doncs, de creuar el bassal, si es que es vol fer una de les rutes més estúpides per anar a Holanda que s'hagin vist en temps. D'aquell país provingué, tal dia com avui, el molt excepcional Doctor Joost van de Weijer, amb qui he tingut

fructificants publicacions i conversacions en magnes quantitats. En aquell llunyà país on la pluja cau i el fum emana dels alvèols dels turistes, (si es que no estan copsant amb l'esma just per a respirar l'anar i venir de dracs gegants, fades i nans petits) vaig tindre la fortuna, deia (i és que no pot deixar d'aparèixer la fortuna), de treballar amb el molt honorable Professor Theo Gevers i amb el molt prohom Doctor Marcel Lucassen, gent que em va donar un enfocament sobre la recerca que em va obrir coses. Aprofito per a demanar disculpes per l'extensió excessiva de la frase anterior.

Ara, si se'm permet, tornaré a vindre a aquell planeta anomenat Catalunya (i dic això amb el cor a la mà i una llàgrima regalimant pel meu immund rostre de galifardeu), per a esmentar a la resta de persones del meu grup , sense les quals no seriem un grup, doncs la absència de tothom, a excepció d'un mateix, o en el cas extrem, també d'un mateix, crearia un grup d'una o cap persones respectivament, i això té tant de grup com quelcom que no en té ni rastre de tal cosa (aquí demanaria un esforç al lector per a posar la seva pròpia metàfora, i és que Bolonya ens ha ensenyat que cal ser interactius i fer parçonera a la gent). De res. Així d'entrada, no podríem oblidar, deu nos guard de tal baixesa, d'esmentar a la resta dels molt superlatius doctors del grup: Xavier Otazu, Robert Benavente, Alejandro Pàrraga i Olivier Penacchio. Gràcies pels vostres comentaris enriquidors tan a nivell idiomàtic com espiritual. L'Anna Salvatella, la Susanna Àlvarez i el Francesc Tous, a qui desitjo molta sort des de les ferèstegues terres Londinenques on em trobo actualment en un sentit no literal. En Javier Vázquez (la vida ara és menys funesta) i en Fahad Shahbad, els propers en caure sota el jou de la tesis. Com no, la Sheyda Beighpour i la Naila Murray (visca l'eye-tracking!), amb qui sumem ja una pila de dones (arribarem a Europa?), Jordi Roca (visca!), Jaime Moreno, David rojas, Ekain Artola i Ahmed Mounir.

Obviarem el fet de que ara aquí he de posar una llista de gent del Centre de Visió per Computador que ha estat important, d'excepcional importància, rellevant o absolutament res, en la meva carrera predoctoral. Començaré, pel simple fet de l'hora que és i de que encara no he fet cafè, amb els companys de despatx: Marçal Rossinyol (i van dues), Javier Vázquez (la qestió és repetir), Juan Mas, (qui vol treure'm protagonisme doctorant-se abans, sense tindre en compte que només importo jo), Alicia Fornés (sempre protegint-nos les esquenes en una típica formació de Tango i Cash), Dani Rowe (estiguis on estiguis, continua així), Ferran Diego (mestre dels MRF i de la preparació d'esdeveniments), Carles Sánchez (l'altre que ha decidit malmetre el meu protagonisme) i Mohammad Rouhani (who makes our office a better place, thank you mate!), Eric Sommerlade (is nitch jia!) i, es clar, no em podria oblidar de la Carlita! Altres persones, o en el seu defecte individus, que cadascú jutgi serien, i encara són: José Antonio Álvarez (amb qui he compartir Amsterdam, patiments i opinions), Agnés Borràs (qui també ha hagut de veure la seva tesis des de la perjudicant perspectiva de la segmentació), Jordi González (qui ha inflat els pulmons de la meva ànima), Xavier Baró i Sergio Escalera (que em van fer entendre de la importància dels congressos), Pau Baiget i Ivan Huerta, que aquí hem estat des del principis i ara lluitant-ho des de la ciutat de Barcino, Parta Pratim (sí, la col·laboració és possible), a l'Aura Hernández pels seus consells i vitalitat, a la Camp Devesa per omplir l'espai amb la seva alegria, Al Dimosthenis Karatzas pel que ha representat en la meva cerca per a una recerca multidisciplinària, al Juan José Villanueva, qui va

obrir-me les portes del CVC i de la seva confiança, a l'Ana Pires, per demostrar-me la passió per la recerca més enllà de sous i de patiments, en Farshad Nourbakhsh i l'Antonio Clavelli amb qui he compartit centre i pis i tants i tants d'altres que si m'hi poso a pensar no em quedarien energies més que per a anar al bar. En tot cas, i perquè no sigui dit, imploro al lector sedent de noms que visiti tan ràpid com li sigui possible: http://www.cvc.uab.cat/personal.asp.

Acabo, i és que allò que comença bé ha d'acabar d'una manera o altre, excepte, es clar, els arguments fàcils de les superproduccions, amb la llista d'amics que no tenen relació directe amb el meu doctorat, però els que han omplert de delit i de fal·lera així com d'altres coses, la meva caduca i mortal existència i m'han fet ser qui soc i arribar on estic (desenganyem-nos). Un crit exaltat pels companys a temps més o menys parcial de la UAB: Nacho, Silvana, Kiki, Adolfo, Roger, Jordi (Font ,(que nhi ha una pila!)), Otger, Jordi (Quadra), Jaume Clon i aquell paiu a qui li vaig tirar el cafè i va haver de marxar a casa a les 11. Ara, una xisclet d'exitació pels de Terrassa que no he encara anomenat i la resta d'amics amb qui he compartit tota una vida: Anna, Santi, Mari Angels, Delgado, Navarro, Almi, Marta, Tristollo, Centollo, Totiano, Montse, Markitus, Mari i el tiu de la barba i la pandereta que rondava per la Plaça Vella.

Moltes gràcies a la Gabriella qui, tot i no estar present durant la tesis, ha estat de gran ajuda en tot el procés d'escriptura, ja a Granada com a Londres.

A la meva família, tota, que si bé és molt menuda, és una gran família a qui méstimo molt... i gràcies Su per aquests fantàstics 15 anys que, com aquests agraïments, han hagut d'arribar a la fi.

Resum

***Keywords**: Segmentació, saliency, color.*

La segmentació d'imatges té com a objectiu el partir una imatge en un conjunt de regions no sobreposades anomenades segments. Tot i la simplicitat d'aquesta definició, la segmentació d'imatges esdevé un problema molt complex. La pròpia definició de *segment* és encara poc clara. Quan demanem a un humà que segmenti una imatge, aquesta persona segmenta emprant diferents nivells d'abstracció. Alguns segments poden ser formats per una sola textura ben definida, mentre que d'altres corresponen a un objecte en l'escena que conté diverses textures i colors. Per aquesta raó, la segmentació d'imatges es divideix en *bottom-up* and *top-down*. La segmentació *bottom-up* es caracteritza per ser independent del problema específic a tractar, és a dir, que tracta propietats generals de les imatges, com les textures o la il·luminació. La segmentació *top-down*, és especifica per a un problema, cercant entitats en l'escena tals com objectes coneguts.

Aquesta tesi està centrada en la segmentació *bottom-up*. Començant amb l'anàlisi de les mancances dels mètodes actuals, proposem un mètode anomenat RAD. El nostre mètode millora les principals mancances d'aquells mètodes que usen les propietats físiques de la llum per a realitzar la segmentació. RAD és un mètode topològic que descriu la reflectància d'un material.

Després, tractem un dels principals problemes de la segmentació: adaptació no supervisada al contingut de la imatge. Per a aconseguir un mètode no supervisat utilitzem un mètode de *saliency* també presentat en aquesta tesi. Aquest mètode calcula la *saliency* de les transicions cromàtiques d'una imatge mitjançant un anàlisi estadístic de les derivades de la imatge. El mètode de *saliency* s'utilitza per a construir la nostra proposta final de segmentació: spRAD, un mètode no supervisat de segmentació.
El model de *saliency* ha estat validat mitjançant un experiment psicofísic així com computacionalment, millorant un mètode actual de *saliency*.

spRAD també millora els mètodes actuals de segmentació, com queda palès pels resultats obtinguts en una base de dades de segmentació àmpliament utilitzada.

Abstract

Keywords: *Image segmentation, saliency, color.*

Image segmentations aims to partition an image into a set of non-overlapped regions, called segments. Despite the simplicity of the definition, image segmentation raises as a very complex problem in all its stages. The definition of *segment* is still unclear. When asking to a human to perform a segmentation, this person segments at different levels of abstraction. Some segments might be a single, well-defined texture whereas some others correspond with an object in the scene which might including multiple textures and colors. For this reason, segmentation is divided in bottom-up segmentation and top-down segmentation. Bottom up-segmentation is problem independent, that is, focused on general properties of the images such as textures or illumination. Top-down segmentation is a problem-dependent approach which looks for specific entities in the scene, such as known objects.

This work is focused on bottom-up segmentation. Beginning from the analysis of the lacks of current methods, we propose an approach called RAD. Our approach overcomes the main shortcomings of those methods which use the physics of the light to perform the segmentation. RAD is a topological approach which describes a single-material reflectance.

Afterwards, we cope with one of the main problems in image segmentation: non supervised adaptability to image content. To yield a non-supervised method, we use a model of saliency yet presented in this thesis. It computes the saliency of the chromatic transitions of an image by means of a statistical analysis of the images derivatives. This method of saliency is used to build our final approach of segmentation: spRAD. This method is a non-supervised segmentation approach.

Our saliency approach has been validated with a psychophysical experiment as well as computationally, overcoming a state-of-the-art saliency method.

spRAD also outperforms state-of-the-art segmentation techniques as results obtained with a widely-used segmentation dataset show.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this chapter we introduce color image segmentation. First, we draw the main lines of this dissertation. Afterwards, we briefly introduce the problem of image segmentation and we analyze the concepts of bottom-up and top-down segmentation. Next, we present a brief analysis of image saliency as the second main field of this thesis. Subsequently, we present the motivations and objectives of this work and the structure of the chapters and sections that can be found in this dissertation.

1.1 Thesis scope

The main focus of this dissertation is bottom-up image segmentation. We propose a model to cope with the problems derived from shadows and highlights in common segmentation methods. Current segmentation approaches are focused either on the feature space (histogram) or on the image. These methods do not model a surface reflectance, therefore presenting some inconsistencies when shadows or highlights are present in the scene. Our proposal is inspired by the dichromatic reflection model [171], which proposes a simple mathematical model to describe a surface reflectance from the shadows to the highlights. Nonetheless, this proved to be a model too rigid, which fails to describe the interaction of light with a surface in real scenes. To solve this problem, we apply a creaseness operator to the image histogram followed by a ridge extraction process. The resulting ridges are a simplification of the image histogram where the dominant structures (mountains) do actually represent different material reflectances. These ridges are used to cluster the histogram. A representative color for each cluster is computed and remapped into the original image, thus performing the segmentation. The ridge extraction process has been designed to simplify a manifold, but although it is not explicitly meant to follow common illumination changes as described by the dichromatic reflection model. To improve its behaviour in the framework of segmentation, statistical information of illumination changes in real images is included.

This dissertation also presents a multi-scale image saliency method, validated by

means of a psychophysical experiment. Image saliency has been used to overcome one of the main challenges encountered in images segmentation so far, namely, non-supervised adaptability to image content. Segmentation coarseness varies depending on the image content or on the requirements of the specific application in which segmentation is to be applied. Current segmentation algorithms commonly have a set of parameters which can be tuned for adapting its segmentation coarseness to a given problem. Nonetheless, such tuning proves itself insufficient when some grade of adaptability is expected for any image, *i.e.*, in general for purpose segmentation. Different images display diverse scenarios, with representative objects at different scales, either indoor or outdoor, and so on. This makes general purpose segmentation a challenging problem. A way to automatically set the parameters of a segmentation method is therefore needed. Image saliency can be applied for this purpose. It is assumed that salient objects and regions in the scene ought to be properly segmented in a *correct* segmentation. A saliency method, validated and improved in this work, is used for the non-supervised parameter tuning. The saliency method proposed is an extension of *color boosting* proposed by van de Weijer, Gevers and Bagdanov [198].

Summarizing, in this dissertation we present a non-supervised model of image segmentation (spRAD) and a model of image saliency evaluated with both psychophysics and segmentation. Our segmentation schema is able to model a surface reflectance in a more flexible way than the dichromatic reflection model. Furthermore, the inclusion of saliency information makes spRAD a non-supervised segmentation method.

It is our believe that computer vision and psychophysics, as presented in this dissertation, are two fields which have been increasingly drawing nearer to each other.

Psychophysics and multidisciplinarity in computer vision

The link between computer vision and Psychophysics is getting stronger. Nonetheless, it is still difficult to find articles about psychophysics in computer vision, whereas is easier to find articles of computer vision in conferences and journals more related with psychophysics and perception.

We will mention several articles of psychophysics in this dissertation which are required to understand what 'segmentation' means in computer vision.

We have referred to another technique which is included in this thesis, namely, saliency. Actually, saliency is a phenomena which can be hardly understood without a solid psychophysical experimentation and fundamentals behind. Not in vain, the most well-known and accepted algorithms of saliency are strongly based on psychophysical experimentation and biological evidences.

Initially, image segmentation algorithms were shinny based on perceptual mechanisms, and were mainly mathematical theories based on learning or, directly, *ad hoc* algorithms tested and optimized on a certain set of images. Computer vision is becoming nowadays a more interdisciplinary and wider field. Machine learning, mathematics and physics, are now combined with psychology, psychophysics and neuroscience. More and more, biologically-inspired and psychophysical experiments are being introduced in computer vision algorithms. For instance, in addition to the

work about saliency and psychophysics detailed in Chapter 4, we have also performed during this thesis (although out of the scope of this dissertation, and therefore not included) an experiment with an eye-tracking to aid in damage painting restoration. That is, a multidisciplinary work which combine computer vision, psychophysics and experts on damage painting restoration.

In this thesis we follow this train of thoughts and we use a multidisciplinary approach to validate our proposals, although it is, indeed, a thesis on computer vision.

Organization of this chapter

Firstly, section 1.2 we present a discussion about image segmentation, which will be further extended with a state of the art in chapter 2. A discussion about top-down and bottom-up segmentation is presented in section 1.3, and extended with a state of the art in top-down segmentation in section 1.4. Afterwards, in section 1.5 we introduce image saliency and present a brief state of the art on this field. Subsequently, in section 1.6 we explain the main objectives and contributions of this thesis. The organization of the thesis is presented in section 1.7.

1.2 Image segmentation

Computer vision aims to build computational tools to analyze and understand images as well as actions in video sequences. Such a comprehensive understanding of scenes requires coping with the problem from many different points of view. An example of what, generically, computer vision aims to yield is depicted in Fig. 1.1. In the first image (Fig.1.1a) we can see an example of different materials/textures which computer visions aims to automatically detect. Another level of recognition is object detection and classification. In this case, the idea is to detect objects which are commonly formed by multiple and complex parts, such as the aeroplane showed in 1.1b. In these cases, it is necessary to include top-down information to guide the algorithms in finding specific objects. The best methods for object classification and detection are presented regularly in the PASCAL challenge [56]. In the PASCAL challenge the participants have to detect 20 classes[1] in real scenes both outdoor and indoor. Furthermore, this complex task can be much more difficult if we are focused on the real aim of computer vision, namely, to be able to analyze a scene as a human being would do so. An example of what we can expect in this case in presented in Figs.1.1c,d. Both images have been downloaded from the labelMe dataset [164], which is an open tool of online image labelling. In this web site there is a collection of images which can be labelled by every one who desires it. Hence, there are no restrictions as in the PASCAL dataset. For instance in the indoor scene showed in Fig. 1.1c we can see how people segmented the corridor, the walls, the stair and even the advertisements of the wall. The outdoor scene showed in Fig.1.1d shows an even more interesting scene. In this case, in addition of the objects present in the scene such as trees, cars, windows and so on, there is a person which is related with another

[1]The 20 classes are: person, bird, cat, cow, dog, horse, sheep, aeroplane, bicycle, boat, bus, car, motorbike, train, bottle, chair, dining table, potted plant, sofa, tv/monitor

level of recognition: action recognition. For instance, we can see a person who is *crossing* the street. Additionally, we can see that this person is doing it *correctly*, that is, across the zebra crossing.

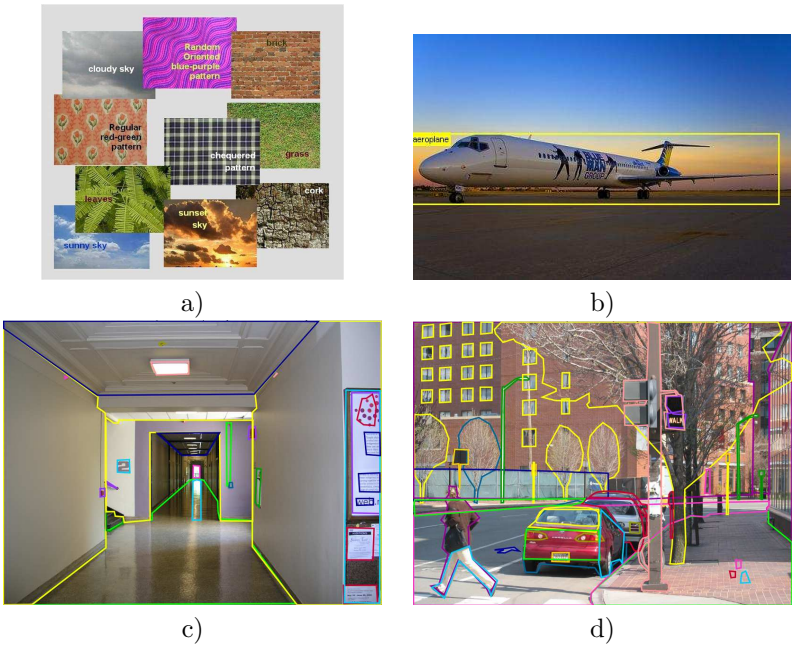


Figure 1.1: Examples of complex scene/object classification extracted from LabelMe [164] and PASCAL dataset [56].

It is clear that detecting a certain (known and well-analyzed) material seems easier than detecting if a person is having a *suspicious* behavior. Nevertheless, in both cases we need to exactly know where the material or the person is. That is, to get the *segment* of the image which contains such element. This Thesis is focused on this issue, namely, Image Segmentation [154] [182] [225] [34] [113] [126]. It can be focused in detecting parts of the image which share certain characteristics, which do not necessarily correspond with a physical object, or to detect concrete objects on the image. It is what makes the main difference between top-down and bottom-up segmentation [202] [17] [212] [86] [24] [118].

1.3 Top-down and bottom-up image segmentation

The information used in computer vision is classified in two main families: bottom-up information and top-down information. A bottom-up system consists in the piecing of little systems to give rise to bigger systems. Hence, the information used in this approaches is very detailed. By the other side, in a top-down approach, the overview of the system is firstly formulated. In computer vision, it means that bottom-up information uses cues which are rather simple and problem-independent. Thus, bottom-up

information is supposed to perform well independently of the nature of the problem. By the other hand, top-down information is commonly included in a problem dependent paradigm where, commonly, some learning is required.

The question about whether image segmentation is actually a pure bottom-up process or not has been treated in several articles, mainly those related with psychophysics. An interesting work regarding this issue was presented in 1997 by Vecera and Farah [202]. They point out that there is a paradox in this discussion: *"One could also argue a priori for bottom-up image segmentation because interactive image segmentation seems to pose a paradox: If the purpose of image segmentation is to group locations of an object in order to recognize that object, then how can object information be used to guide this process, since, presumably, the object hasnt been recognized until completion of the segmentation?"*.

The authors of this article present a brief dissertation about this issue by analyzing some previous works. Afterwards, they present four experiments from which stands out that there is indeed a top-down influence in segmentation. Nowadays, this theory is well accepted and no further discussed. It is actually easy to find a great collection of article oriented to the addition of top-down information in image segmentation [17] [212] [86] [24] [118].

Indeed there is no such a paradox. Whereas we can easily find a fairly big number of authors who either implicitly or explicitly suggest that bottom-up mechanisms can not be influenced or biased by top-down processes, we know that it can be hardly the case. At least, not in practice. Actually, this misunderstanding causes some conflicts in current terminology. A good example of it is the definition of *saliency*. In existing literature, we can see articles talking about saliency as a pure problem-independent, bottom-up mechanisms, then leaving concepts such as guidance or attention as a pure top-down processes. Nevertheless, it is well known that actually attention *affects* the bottom-up saliency map. Phenomena that has been modelled in several approaches as in [36]. Hence, we know for interesting works as the one presented by Carrasco, Ling an Read [27], that indeed *attention alters appearance*, that is, the commonly considered pure bottom-up information. In this sense, Knudsen has recently suggested in a great survey about attention [107], four elements which involves attention, namely, working memory, top-down sensitivity control, competitive selection, and *automatic bottom-up filtering for salient stimuli*. The specific influences of pure bottom-up features in attention and saliency are analyzed by Itti in [91].

All this confusion can be cleared up by knowing the fact that in perception there is not the so-called *zero-moment*. That is, that our brain can not be turned on all of a sudden having no information at all inside. High order cortical areas are influencing all the time the basic bottom-up operations coming from the retina. Hence, visual processing therefore involves countercurrent streams of information [75]. Furthermore, this influence of high-order was initially found just at high-levels of the visual path. Nonetheless, nowadays we know that such influence also occurs in the earliest stages. It is even possible to go one step further in this direction. There is not just an interaction of top-down information because our brain has been non-stop working and it bias somehow bottom-up information as arriving. There is also the fact that our brain *predicts* what we can found in a specific environment or task [112]. Therefore

knowing *in advance* which kind of features or stimuli we are expected to be looking for before find them. When such a prediction is not present in our brain, it may turn in problems such as dyslexia and schizophrenia. For a further explanations about this interesting issue we encourage to read [75] and for an interesting analysis of top-down cues commonly used in a computational framework, [195].

1.4 Image segmentation: current status

In this section we describe the current status in image segmentation by briefly describing the main techniques of top-down and bottom-up image segmentation.

1.4.1 Bottom-up segmentation

We classify bottom-up image segmentation methods as:

1. Image-based segmentation.

   Exploit the spatial information contained in the image, named *spatial coherence* [95].

   - Deformable models.
   - Graph-based approaches.
   - Region growing and edge detection.
   - Split and merge.
   - Topological methods.
   - Other methods.

2. Feature-based segmentation.

   These methods are focused on the photometric information of an image represented on its histogram [5] [221].

   - Histogram thresholding.
   - Clustering Techniques.
     Hard Clustering.
     Fuzzy Clustering.

3. Physics Based methods.

   These methods use the knowledge about the physical formation of the scene (light, surfaces reflectance), to perform the segmentation.

   - Segmentation based on the dichromatic reflection model.
   - Spatial transformations.

4. Hybrid methods.

   Hybrid techniques combine methods of the previous categories.

Due to its importance in this dissertation, we present a comprehensive state of the art in bottom-up image segmentation in chapter 2, where each of these categories is detailed.

What we are concerned in this section is about the current status of these methods.

The most meaningful recent advances in segmentation belong to image-based and feature-based methods. Therefore, methods most commonly used in current articles of segmentation belong to these two categories. Concretely, the most widely used method of feature-based segmentation is the Mean Shift [39] whereas the Efficient Graph-based Segmentation method [58] is the most used among all the image-based methods.

Regarding physics-based techniques, they have been basically stacked in last years and no meaningful advances have been presented. These methods are mainly extensions of the *dichromatic reflection model* of Shafer [171]. Whereas this model is a good theoretical framework, its application on real images has demonstrated to yield poor results. The main reason is that it is a too rigid model and its not capable of adapting to the artifacts of the images, mainly caused by acquisition conditions, clipping and image compression.

Each of the categories described above (feature, image and physics based) have its own potential advantages. A concerning of this dissertation is to find a way to exploit each of these advantages by proposing a hybrid segmentation method as a combination of all of them. Thus, in this dissertation we present a bottom-up segmentation method which is inspired in the dichromatic reflection model. It is performed an analysis on the feature space to find shapes similar to the ones described by this model. It is done with a ridge-based analysis of the histogram. The method proposed, called RAD, outperforms state of the art segmentation techniques as detailed in chapter 3. The method proposed is therefore a feature-based segmentation technique. This method is further complemented with an statistical analysis to include physical knowledge, making a hybrid (feature plus physics) method. This method is called pRAD and is also presented in chapter 3. Finally, we include image spatial coherence using a saliency model presented in chapter 4. By means of this model, we yield a full hybrid (feature plus physics plus image) segmentation method which outperforms existing bottom-up methods.

Due to the importance of segmentation in this dissertation we present a separated state of the art in image segmentation in Chapter 2.

1.4.2 Top-down segmentation and object recognition

As pointed out before, the influence of top-down interaction in visual tasks such as object recognition is accepted. In opposition with the bottom-up features, top-dow interactions are learned and strongly depend on the subject's experience. These influences, thus, change the strategies of search in a way that differs, either gently or strongly, among subjects. In other words, the internal representations of the world, acquired by experience, affect our brains strategy for analyzing visual scenes [75].

Another important issue, if we expect to follow a coherent strategy for top-down

segmentation, is the complexity of the top-down representation of the objects and scenes in our brain. The Gestalt psychologists stated that "There are entities where the behavior of the whole cannot be derived from its individual elements nor from the way these elements fit together; rather the opposite is true: the properties of any of the parts are determined by the intrinsic structural laws of the whole".

Currently, there are evidences that along the visual path, from the primal visual cortex to the high-order visual areas, neurons become selective to progressively more complex stimuli [75]. This complexity can therefore reach a whole object, as the Gestalt school holds. But it can reach even more complex information, namely, the influence of context. In these cases, we use the information that certain objects tend to appear in conjunction with some other objects in certain situations [59], a process called *prediction* [112]. It means that this top-down influence can appear at different stages and with different complexity. An interesting study of the earliest stages where top-down influence can appear can be read in [8]. For instance, there is the basic conception of attention, as understood in the framework of saliency. Basically, it is a kind of top-down influence which guides our attention by modifying the precognitive, bottom-up, saliency maps. We treat this issue below in section 1.5. But attention can be focused in entire objects [51] [167] and even predicting the apparition of objects in specific known environments and situations.

Current top-down segmentation and object recognition methods either use these evidences in different ways or perform an *ad-hoc* combination of features which have been seen to be useful for certain classes. Reached this point we have to set a difference between object recognition and image segmentation. The border between both is narrow and smooth, nonetheless, it is enlightening the difference made in the PASCAL challenge [55], where object detection is separated from object segmentation. For object detection it is just necessary to mention wether a certain object is present or not in an image, without the necessity to specify its borders. By the other side, object segmentation is directly related with the presented dissertation, where the object have to be effectively segmented in the image. A clear description as well as the relation between both concepts (recognition and segmentation) can be read in [195]. The relation is simple to explain. For object recognition, a learning procedure is performed to find those features which better describe a certain class (*i.e.* horse, face, car, bike, and so on). For object segmentation the same information can be used, once learned, to decide which of the segments facilitated by a bottom-up segmentation method, corresponds with a known object. It stands out the fact that, without a good bottom-up information, it would be fairly difficult (if not impossible) to correctly segment an object. These features can be combined or extracted in different ways to detect or segment an object. Some well-known examples are SIFT [124] or the textons [177]. For a detailed explanation of the features and techniques involved in bottom-up segmentation, we refer to Chapter 2.

In current literature we can find a division between those top-down methods which require a previously labelled set of images for the learning procedure and those which do not need it.

Top-down segmentation methods requiring labelled data

The method proposed by Borenstein and Ullman in [17] is an example of a method which requires previously labelled data for the learning procedure. They propose to learn those segments more representative of a given class by using two sets of images: the set C which contains examples of class images, and the set NC which do not. Then, they divide each image into a large set of rectangular sub-images which sizes vary from $\frac{1}{50}$ to $\frac{1}{7}$ of the image size. They compute optimal fragments based on the Neyman-Pearson decision theory, namely, optimal fragments are defined as those having maximal frequency within the class. Then, they build a figure- ground, labelled image. In this image each pixel is mark as figure (class) or ground. Finally, they use this large set of segments to detect whether an specific class is present or not in an image and to delineate the borders of the class, that is, to perform the segmentation.

Another interesting shape-based model is presented in [117]. They learn an Implicit Shape Model (ISM). It does not try to define an explicit model for all possible shapes a class object may take, but instead define *allowed* shapes implicitly in terms of which local appearances are consistent with each other. It is in conjunction with those approaches which allow the definition of new novel objects by a combination of class prototypes as in [97]. A further improvement of this method has been recently presented by Gorelick and Basri in [76]. This model represents shape using two types of local descriptors. One encodes the boundaries of the shape. The other is a regional descriptor, which includes a dense local orientation field derived from the shape by solving a Poisson equation on the shape. It is also added to this representation a color histogram with each shape. To solve partial matches they use a modification of the voting scheme proposed in [117]. The main advantage of [76] is that it allows more objects occlusions than [117]. An example is showed in Fig.1.2 , where we can see how the approach of Gorelick and Basri in [76] (Fig.1.2 4th column), solves better the occlusions than the method in [117].

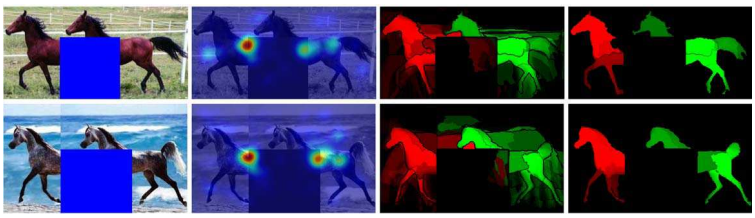


Figure 1.2: An example of the improvement achieved by Gorelick and Basri [76] (4th column) with respect to the approach presented in [117] (3rd column). From [76].

Another interesting family of methods are those which use Conditional Random Fields. An example can be found in [86], which is an extension of a previous work called multiscale CRF [85]. In this article a CRF in used to try to give a context to

an image which has been previously segmented in superpixels.

Another approach based on CRF has been presented by Levin and Weiss [118]. In this case CRF is used to consider at the same time bottom-up and top-down features (bottom-up are again, superpixels of the image). This methodology yields good results with less segments of the original image. Both methods [86] and [118] allow occlusions. A recent approach based in CRF is the OBJCUT, presented by Kumar, Torr and Zisserman in [111]. This method has as a main advantage in relation with other CRF-based methods that it allows more inter-class variability.

Top-down segmentation methods without labelled data

The most common way of learning object categories is by means of tools like probabilistic Latent Semantic Analysis [88] and Latent Dirichlet Allocation [16] based on a bag of words approach, which, when applied to images is called *visual words* [41] and SIFT -like feature region description [179] [124].

The main problem with visual bag of words is that all the words that describe an image are placed into a single histogram, therefore loosing all the spatial information. It means that the fact that those words which describe a class are present in a image does not necessarily imply that this class is present in the image. Further, with a simple bag of words approach it is difficult, for the same reason, to segment an object. Commonly, the borders of the objects are smoothed and irregular and hardly correspond with the correct borders. This problem has been treated in some rather recent works. An example is the approach proposed by Russell *et al.* in [163]. The authors state that a correct segment (on which correctly draws the object boundary) will be described by coherent groups (topics), whereas segments overlapping object boundaries will need to be explained by a mixture of several groups (topics).

In this approach, for a given image a set of subsegmentations (candidate segmentations) are computed using a bottom-up segmentation algorithm. Then, images are represented using a SIFT descriptor and quantized in about 2000 words using k-means algorithm. Once the visual words are computed for an image, each image segment is represented by a histogram of visual words contained within the segment (the bag of words model). Afterwards Latent Dirichlet Allocation is applied to find image topics. Then, each segmented its sorted by its similarity within the learned visual words.

Another approach to overcome the main shortcomings of the visual words is presented by Cao and Fei-Fei [24]. The authors propose what they call spatially coherent latent topic model (Spatial-LTM). Spatial-LTM represents an image containing objects in a hierarchical way by oversegmented image regions of homogeneous appearances and the salient image patches within the regions. The main difference with the approach proposed in [163] is that Cao and Fei-Fei generate the topic distribution at the region level instead of the word level as in [163]. In other words, Cao and Fei-Fei use the candidate segments to build a hierarchical representation of the object, instead of treating each segment as a word as in [163], what makes the methods less sensitive to segments quality than in [163].

But there are also other techniques for top-down segmentation without labelled data. An interesting example is presented by Winn and Jojic in [212]. This method called LOCUS is a shape based model. Concretely, LOCUS define the class informa-

tion with the broad global shape, the edge model defining the typical edge locations (with Canny) and (optionally) in a mild prior on the appearance features (color or texture). Even when the main advantage of this method compared with other shape based approaches is that it allows a considerable variability in the shape of an object, it is still required that a given object have to have a similar orientation to be detected.

Another interesting approach is an information based technique for the learning as presented by Ullman in [195]. The learning procedure is based in the detection of those features which are more informative for a given class. Furthermore, it is suggested that single, simple, features such as certain color of a pixel or a corner is not enough discriminative and several neighboring features have to be combined in what the author calls *fragments*. These segments are build if they are more informative (discriminative) for a given class. Then, for each class a hierarchical classification of fragments is extracted. This idea of a hierarchical representation actually corresponds with the HVS as previously commented. A graphical example is showed in Fig.1.3.

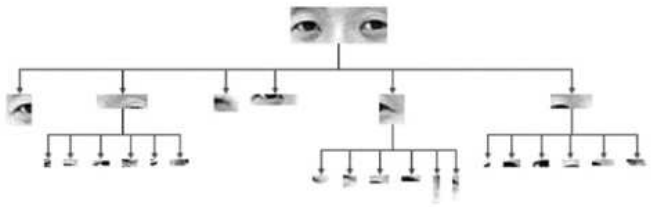


Figure 1.3: Hierarchical classification of a class. From its more complex fragments until the simple features. From [195].

Another method which do not require a training set of labelled images can be found in [191] where a tree structure is proposed to catch the relatives frequencies of the features which describe a class.

1.5 Image Saliency

The other main field which we will treat in this Thesis is 'Image Saliency'. As mentioned before, the definition of saliency still raises some controversy. Basically, the problem is if saliency is whether a pure bottom-up mechanism or it involves top-down mechanisms. A deep analysis of this question was presented by Itti and Koch in [93]. Again, as happens with image segmentation, the problem of whether we can conceive 'saliency' as a bottom-up process or not, just drives to confusion. The basic pure bottom-up map can be modified by task-dependent top-down information. This theory, which some articles still discuss, have strong evidences which can be hardly refuted. An interesting experiment about it was done by Carrasco, Ling and Read in 2004 [27]. From this work it stands out that top-down alters the bottom-

up maps. Concretely, they found that attention (task-dependent top-down) boosts stimulus contrast. By the other side, there is the phenomenon called top-down inhibition [36], which is a task-dependent way to do not consider certain information which is meaningful for the bottom-up saliency. More evidences of the inhibition of saliency due to attentional and task-dependent information was presented by Cutzu and Tsotsos in [42]. In this article the authors present evidences about the existence of a suppressive annulus around the attended item.

Reading existing literature from the fields of psychophysics, perception, computational neuroscience and computer vision, it stands out that probably the line to be drawn is between *saliency* and *attention*. For some researchers, it is clear that saliency is pure bottom-up, whereas attention is completely task-dependent, therefore, top-down. It might be a rather acceptable difference between what is task-dependent and completely bottom-up, if we accept that one can affect the other. Nevertheless, it is a generical definition which nobody would accept without some reticences. For instance, in a publication included in this dissertation it was necessary to mention the existence of pure bottom-up saliency, as understood in psychophysics and top-down saliency, as commonly treated in computer vision [199]. This thin border between these two concepts often appears to be trespassed as in [91] where it is mentioned the concept of bottom-up up visual attention or in [98] where what is called attention, is actually computed by means of a saliency map (based in orientation or based in orientation and color).

In this work, for the sake of clarity, we will refer to *attention* as a task-dependent process, whereas *saliency* will refer to a bottom-up mechanism. This separation can be found in several articles as in [54], where the authors claim in the abstract from experimental evidences that '*the results suggest rapid feature analysis mediating detection, followed by attention-demanding binding for identification and localization*'. Nonetheless, we do not attempt to be strict in this sense, since, as mentioned, some authors might disagree with such an strict separation. A graphical example of what a bottom-up saliency map is, is shown in Fig.1.4. The left image is the original image. The central image is a is a computational pure bottom-up saliency map. The image on the right is a human saliency map obtained by means of an eye-tracker, a common technique used to evaluate saliency.

Regarding what is considered attention, namely, a top-down task, it is defined in [98] as: *'Visual attention is the ability of a vision system, be it biological or artificial, to rapidly detect potentially relevant parts of a visual scene, on which higher level vision tasks, such as object recognition, can focus'*. Attention needs indeed to have a map of features, which will be the saliency map, to decide, among all these features, which of them are relevant for an specific task. There is a considerably amount of experiments related with this. For instance the one presented in [48], which follows a classical paradigm of experimentation.

An outline of the experiments of attention presented in [48] is shown in Fig.1.5. In Fig.1.5a, it is shown an experiment where the subjects should report the number of black letters appearing in the stimuli which is flashed briefly enough to avoid eye movements. In this case, the experiment is designed to find evidences of the limited capacity of processing information. It was found that the probability of reporting the target letter N is much lower with two accompanying targets (the panel on the

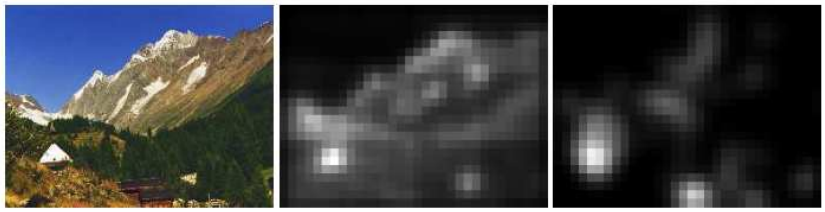
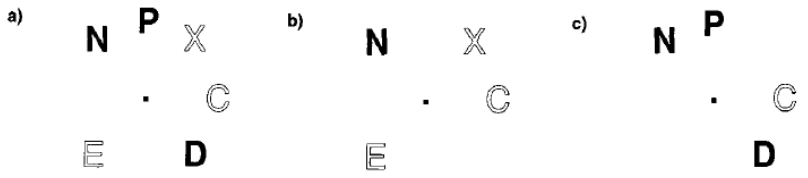


Figure 1.4: Example of bottom-up image saliency obtained from [153]. Left image: original image. Center image: Computational map of bottom-up saliency. Right image: Human saliency map obtained with an Eye-Tracker.

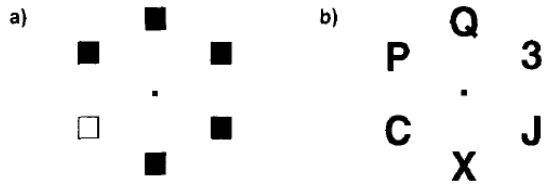
left of Fig.1.5a) than with none (central panel in Fig.1.5a). Another experiment is outlined in Fig.1.5b. In this case it was thought to find a difference between pure bottom-up discrimination and top-down. Effectively, when the target differs from the rest in simple low-level features such as in the example showed in Fig.1.5b left, it is much more faster to detect than those cases when the target differs in more complex feature (Fig.1.5b right). This kind of experiments where a target (expected to be salient) appears among a set of distracting targets, are used because they represent in a simplified way a real situation. In a real cluttered scene, we would expect to have a target combined with several distracting (not-interesting) targets.

Following a similar procedure it has been determined several low-level features which are involved in visual attention, such as contrast [48], color [98], motion [165], rarity based on information theory [132] or much more recently, the adaptive idea of *surprise* as detailed in [92]. Other interesting studies aim to figure out the influence of time in saliency, as in [187]. Comprehensive and interesting surveys on the features involved was presented by Wolfe and Horowitz in 2004 [216] and in 2007 by Knudsen [107]. Wolfe and Horowitz define 5 categories of features depending on its confidence [216]:

- **Undoubted attributes**: Color, Motion, Orientation and size (including length and spatial frequency).
- **Probable attributes**: Luminance onset (flicker), Luminance polarity, Vernier offset, Stereoscopic depth and tilt, Pictorial depth cues, shape, line terminator, closure, topological status and curvature.
- **Possible attributes**: Lightning direction (shading), glossiness (luster), expansion, number and aspect ratio.
- **Doubtful cases**: Novelty, letter identity (over-learned sets in general and alphanumeric category.



a)



b)

Figure 1.5: Outline of the experiments of attention presented in [48]. a) Subjects should report the number of black letters appearing in the stimuli which is flashed briefly enough to avoid eye movements. b) Subjects should report the mismatching element.

- **Probable non-attributes**: Intersection, optic flow, color change, three-dimensional volumes (such as geons), faces (familiar, upright, angry and so on, your name, semantic category (*e.g. animal, scary).*

note that some of these features are bottom-up (as color) whereas some others are top-down (as faces).

As stated before, we consider saliency as a bottom-up mechanisms which can be further biased by top-down information. Evidences of the formation of this bottom-up map in the V1, are presented in [120][227] and [109]. The influence of bottom-up features in visual attention are studied in [53] and [91]. The former presenting an experiment in a dataset of labelled images (called LabelMe [164]). The experiment is designed to answer the question '*How do we decide which objects in a visual scene are more interesting?*'. They sate that whereas an initial intuition would be to answer the question with high-level features, it is indeed a high influence of low-level features. An analysis of some features which are commonly involved in the formation of the bottom-up saliency map was presented by Parkhurst and Niebur in [158]. Basically, current models of saliency use color, contrast and orientation as main features [94] [215] [213] [144] [128] [121]. Unlike that these features are nowadays accepted, it was a large controversy with the role of color in visual saliency. Some important works pointed that color was actually not salient. For instance in the work of 1994 titled

'*Abrupt luminance change pops out; abrupt color change does not*' [188], or also in 2001 [194] among others as [74]. It was basically due to an error in the experimentation, as pointed out by Snowden in [181].

Another question about the formation of the saliency map, was whether the features involved are computed separately or in a parallel way. The results presented by Nothdurft in [147] and by Krummenacher, Muller and Heller, [110] show that there is a parallel computation of these features. A more recent work [109] sates that not all the combinations of features are equally useful. In this work it is found that contrast+motion and contrast+color increase the saliency of a feature, whereas color+motion does not. The method of saliency presented in this Thesis computes color and orientation in a parallel way as these works state.

In the next section we briefly describe the main models of visual saliency by dividing them into bottom-up and bottom-up plus top-down models.

1.5.1 Models of saliency

It has been reported many different models of saliency based on different features and points of view. In this section, we basically enumerate the main models of image saliency by splitting them in pure bottom-up models and top-down guided models.

Bottom-up saliency

Doubtlessly, the most relevant model of bottom-up saliency is the one introduced by Itti and Koch in [94]. This model has a public available version implemented in Matlab [207]. Further it has been also extended to dynamic scenes by means of neural networks as detailed in [45].

Itti and Koch model is based on the basic model of Koch and Ullman's [108]. An schema of the model is depicted in Fig.1.6. It is shown a kind of structure followed by several saliency methods. First, a set of features in the image are computed. The set of features can vary depending the method. Afterwards, there is a competition among all feature maps in order to find the most salient locations for each individual feature, yielding the conspicuity maps. Afterwards, all these maps are combined in a single representation using a Winner-Take-All network, which leads to the finding of the most salient location.

An schema of the model of Itti and Koch is presented in Fig.1.7. In this case the bottom-up features computed are color, orientation and intensity. These features are generated at nine spatial scales. Further, these features are computed in a center surround-schema due to biological reasons [43] [26]. It is shown that typical visual neurons are most sensitive in a small region of the visual space (the center), while stimuli presented in a broader, weaker antagonistic region concentric with the center (the surround) inhibit the neuronal response. After the center-surround calculus, maps at multiple scales are combined in a single one following the classical schema of a pyramid of Gaussians [78]. More specifically, the center is a pixel at scale $c \in \{2, 3, 4\}$, being the surrounding the corresponding pixel at scale $s = c + \delta$ ($\delta \in \{3, 4\}$. Finally, the combination of maps is obtained by interpolation to the finer scale and point-by-point substraction. Finally, by means of a Winner-Take-All neural network [193] they

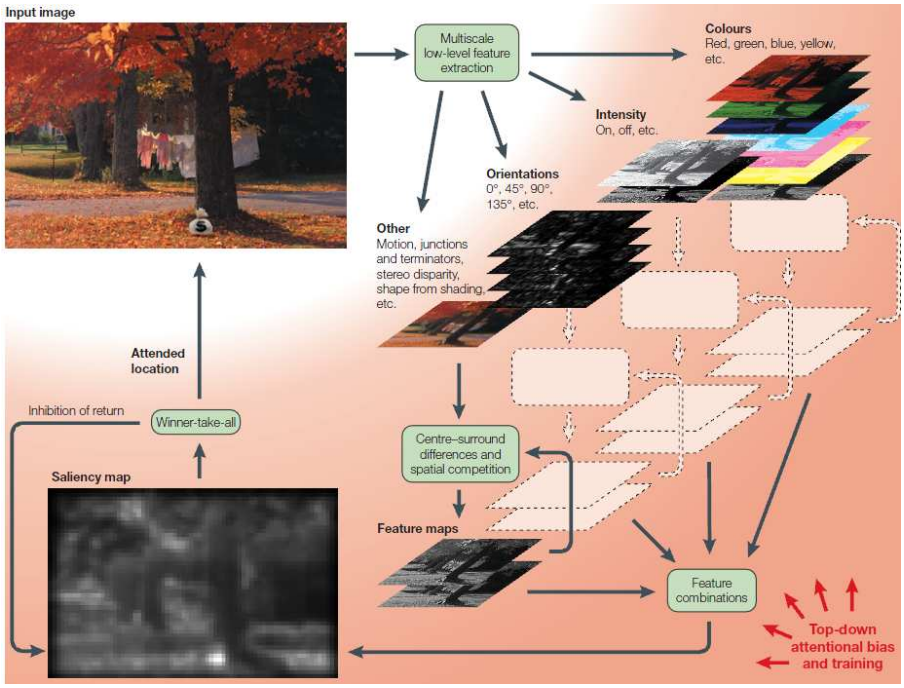


Figure 1.6: Saliency model of Koch and Ullman [108], extracted from [93]. The model is based in a number of features such as color or orientation represented in parallel. Afterwards a combination of the features in performed using a winner-take-all process which yields the most conspicuous location.

combine all the maps into a single map which will be used to guide attention in a bottom-up way. The model of Itti and Koch can be considered nowadays as a reference methodology and is probably the most used saliency method. Some subsequent bottom-up saliency methods use the methodology proposed by Koch and Ullman in 1985 [108] but modifying certain aspects such as the features computed, the previous steps or the post-processing.

An interesting approach which was focused in a further correspondence between the model and the Human Visual System (HSV), was proposed by Le Meur *et al.* in [115]. The authors point out that the main drawbacks of the Itti and Koch model are:

1. Several normalization steps are applied before and after the fusion step.

2. Each channel is normalized independently to a common scale in order to be independent of the feature extraction mechanisms.

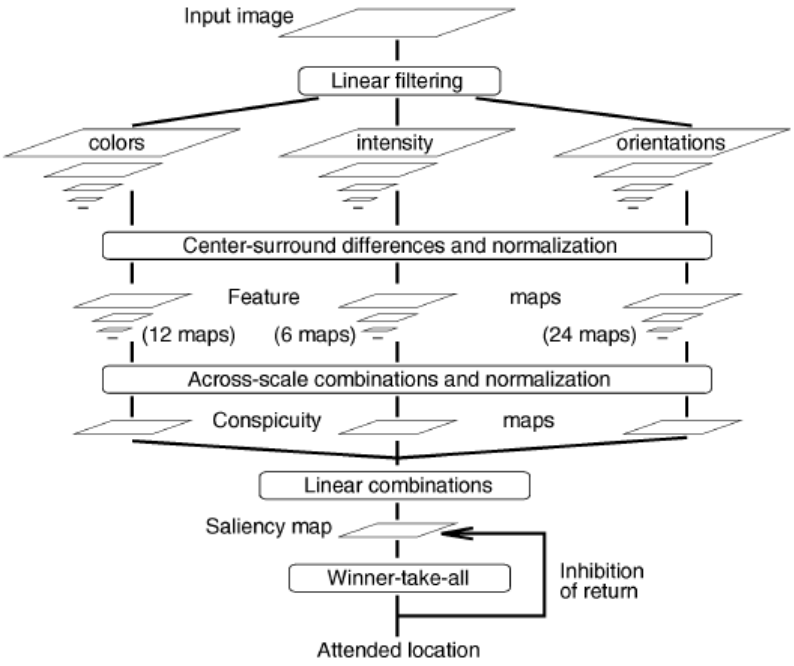


Figure 1.7: Saliency model of Itti and Koch [94], extracted from [94].

3. There are strong links between the visual sensitivity and the viewing distance. However, this has been overlooked.

Le Meur *et al.* propose a computational framework for visual saliency which copes with numerous properties of the HSV. They propose an schema divided in three aspects of the HSV: visibility process, perception and perceptual grouping. An schema of the method is showed in Fig.1.8.

The first part, namely, the ***visibility process*** is made to perform a coherent normalization to solve the first drawback of the Itti and Koch model. It simulates the limited sensitivity of the HVS. Firstly, it changes the chromatic representation of the image to the Krauskop color space. Afterwards they apply a contrast sensitivity function which also is a mechanism of the HSV. A visibility threshold is computed which determines at which scales a component is visible. The early visual features computed are based on spatial frequencies and orientations. Finally, they compute a masking which is a way to change the visibility threshold depending on the context.

The second part of the model, that is, the ***perception*** aims to determine the achromatic components required to compute the saliency map. First, it is reinforced the saliency of an achromatic structure if there is a high chromatic contrast in this

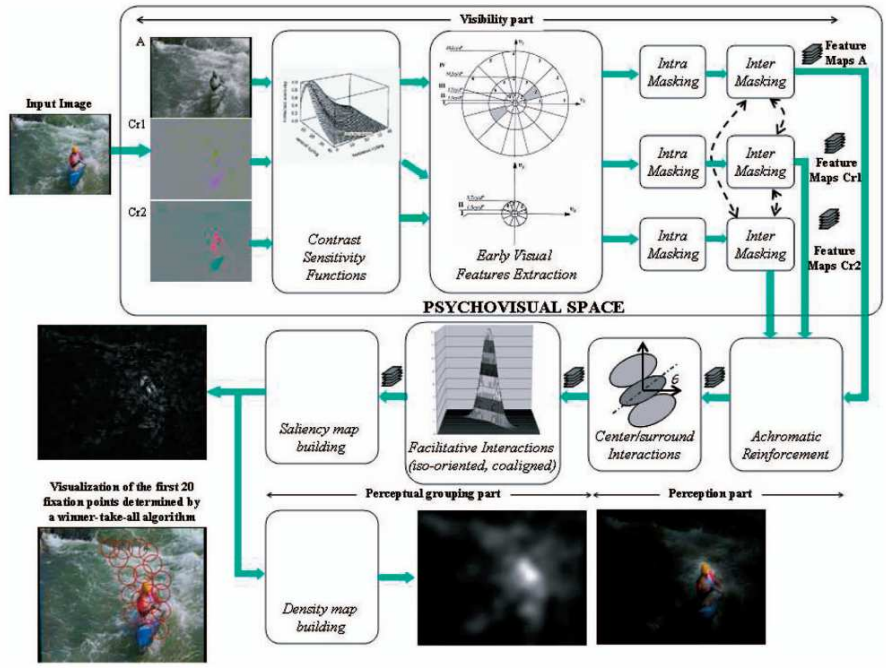


Figure 1.8: Bottom-up saliency model as suggested by Le Meur *et al.* in [115] from which it has been extracted. A three-level schema is proposed which is inspired in the HVS. The three parts are: visibility process, perception and perceptual grouping.

area in the original image. Afterwards a center-surround suppressive interaction is applied.

Finally the last part, called ***perceptual grouping***, refers to the capability of the HSV to group several features to build a high-order structure. To simulate such a capability the authors use a contour grouping technique.

Finally, the saliency map is computed by directly summing the output of the different achromatic channels.

Another model which is based in the Itti and Koch saliency model was presented by Liu *et al.* in 2007 [121]. The authors pose the problem of saliency as a problem of image segmentation[2]. In the same article a public available dataset consisting in 20.000 images for a quantitative evaluation of saliency is also presented. The method is based on a supervised learning using a Conditional Random Field which correctly combine the low-level features proposed which are a multicontrast schema based on a

[2]The relation between saliency and segmentation is one of the main conclusions of this Thesis. Chapter 6 is focused on this issue, where it can be seen the relevance of saliency to evaluate image segmentation.

Pyramid of Gaussian [78]. A center-surround histogram is performed to detect those areas/objects which have a high contrast in its surrounding. As a global feature, the authors propose to use a *color spatial-distribution*. An object is expected to be more salient if it has a color which is not wider distributed in the image.

Another model of bottom-up saliency was introduced by Ma and Zhang in [128]. The method is based on local contrast analysis. The authors said that in image processing, techniques share a sort of common properties, namely, color, texture and shape. The authors claim that these three main properties have a common principle behind: the contrast. They compute contrast by means of a Gaussin difference. Therefore, being computed at a single scale given by the size of the Gaussian. They compute all images at the same scale by resizing them, turning in a pure *ad hoc* step hard to justify. Afterwards, they convert the image to CIE Luv space and compute the contrast image. With this procedure they generate the saliency map. Afterwards, a fuzzy-growing procedure is performed in order to find the saliency areas. In this article, the authors also define three levels of attention analysis: attended view, attended area and attended points.

Besides features such as color, contrast and orientation, there is also a common feature in saliency, which is actually used the method of Liu *et al.* [121], that is, rarity or information. It is expected that a feature which barely appears in the image to be salient. In other words, as more rare the feature is, more its saliency. This characteristic is exploited in many different ways, as the theory of surprise, which quantifies how data affects a natural or artificial observer, by measuring the difference between posterior and prior beliefs of the observer. The work presented by Itti and Baldi in [92] describe a Bayesian definition of surprise. The same idea of rarity is also used by Kadir, Zisserman and Brady in [100], which is an extension of a previous work of Kadir and Brady [99]. They compute the rarity or surprise of features such as contrast or color with the Shannon's entropy formulation at multiple scales. For each scale they compute the PDF of the features and detect the most representative scales, that is, those which are more informative (more surprise). Another method which uses Shannon's theory is presented by Bruce and Tsotsos in [22]. In this case, the computation is done by means of a Neural circuit which is claimed to have similarities with the circuitry existent in the primate visual cortex. Finally, a graph-based schema to compute the rarity of a feature was proposed in [83]

Another work based on rarity is presented in this Thesis. Our proposal is based on Color Boosting, introduced in [198]. Our approach was presented in [199] and will be deeply analyzed in this Thesis. Our proposal detects the most rare chromatic transitions (color plus contrast) in a multiscale framework.

Finally, some other approaches of saliency are aimed to exploit other characteristics such as motion, as the method presented by Guironnet *et al.* in [80].

In the next section we briefly analyze top-down saliency.

Top-down saliency

It has been commented that there is some misunderstanding with the concepts of saliency and attention. Basically, the line which splits both concepts in some works is whether there are top-down features and mechanisms involved or not. In this section we briefly describe some approaches which propose different ways to include top-down information in image saliency and image attention.

Probably, the most important method of top-down saliency or attention was the work entitled *Guided Search: An Alternative to the Feature Integration Model for Visual Search* presented by Wolfe in 1989 [215]. The theory was revised 5 years later by the same author in what he called Guided Search 2.0 [213]. First, the called *feature maps* are computed (pure bottom-up feature maps). They can guide attention in a bottom-up way, if a feature is relevant enough on its environment. Nevertheless, they will not guide attention to a desired item if the low-level features of that item are not unusual. Here, is when the top-down mechanisms appear, which are achieved by means of direct user interaction, when looking for an specific target.

We have already introduced a method which, can be considered as top-down in the second step. The method is the one introduced by Liu *et al.* in [121]. They use a learning process which uses high-level information, that is, human-labeled images. Concretely, they present a dataset with 20.000 images. 15.000 of them were labelled by 3 subjects, whereas 5.000 images considered to have a high consistency were labelled by 6 additional subjects. This information was used for a learning procedure based on a CRF which yields the best way to combine bottom-up saliency maps.

An interesting approach of top-down attention was presented by Lee, Buxton and Feng in 2005 [116]. This approach proposes a dynamic way to include top-down information to bottom-up saliency. A schema of the method is depicted in Fig.1.9. First, they compute a set of low-level feature to generate the bottom-up map. The feature computed are color, aspect ratio, symmetry and ellipse. These features are combined in a single overall bottom-up map. Afterwards, a top-down map is computed. In the article they use just color as a high-level feature whereas, as the authors explain, the model is not limited to this single feature. They combine both maps by means of what they call *interactive spiking neural network*, which finds the consistency between both maps.

A model of the influence of the task in attention was proposed by Navalpakkam and Itti in [144]. They propose a schema where the same features are shared between top-down maps and bottom-up ones. First, there is the initial phase, called *eyes closed*. In this moment, the system receive a keyword which will guide the high-level search. Those features which correspond with the entity or object to be found are computed and it is translated in the task-relevant map which could, for instance, give a great relevance to the center of the scene. In a second phase called *computing*, the system receives the image and the bottom-up features are computed. These low-level map can be biased by the information contained in the task-relevance map. In the third phase, *attending*, an object recognition module determines the correspondence of the most salient location with the keyword introduced. In the last phase, *updating*, the task-relevance map is updated with the new information. Thus, it is a hierarchical

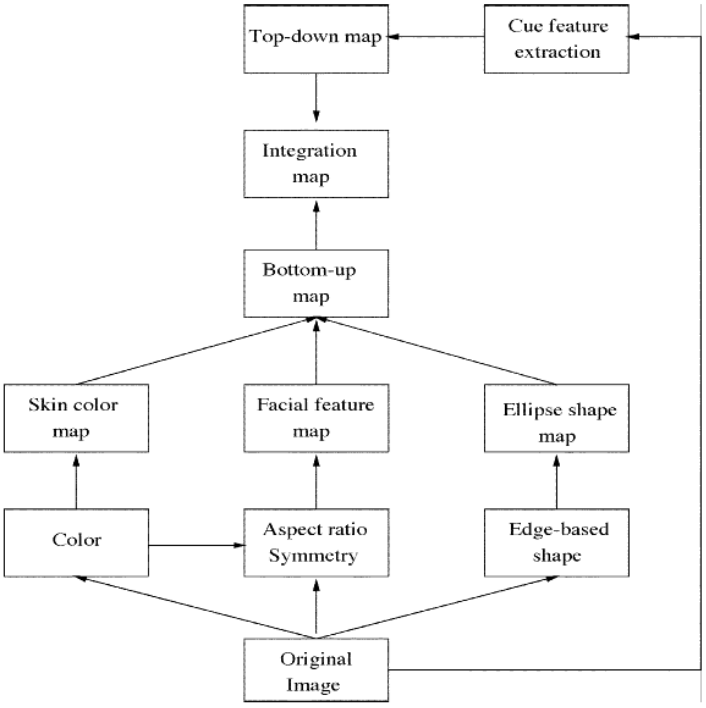


Figure 1.9: Schema of the model of top-down attention proposed in [116] when applied to face recognition. The main parts are the bottom-up map and the top-down map, which are combined with an interactive spiking neural network to generate the integration map. Extracted from [116].

schema of attention.

Finally, we stand out the fact that the inclusion of high-level information to visual saliency has facilitated the apparition of some interesting applications such as object recognition [176], or image retrieval [168] [114].

1.5.2 Image saliency evaluation

The evaluation of models of image saliency has been improving with the popularization of eye-tracking devices, which can be currently acquired for a fairly affordable prices. With an eye-tracker it is possible to generate a saliency map from a a human subject. An example of a human-based saliency map is showed in Fig.1.4. A straight comparison can be carried having a human map and a computational map.

Following this methodology we can find several works to evaluate saliency models. But eye-tracking can be used not just to validate an specific model as in [22][90], but to determine the influence of specific features in bottom-up saliency. Some interesting examples are the works presented by Parkhurst and Niebur[158], by Tatler, Baddeley and Gilchrist [187] or by Ouerhani *et al.* in [153]. A Behavioral analysis of some methods implemented for a robot and using eye-tracking was presented by Shic and Scassellati in [174].

When no eye-tracker is available it is possible to evaluate the feasibility of an specific feature by performing some psychophysical experiment as in our approach of image saliency [199] or as the experiment already referred in the introduction of saliency [48]. Another ways to evaluate saliency can be comparing models in a certain application such as robots [174], object recognition [176], or image retrieval [168] [114].

Finally, it has been published in 2007 a dataset consisting in 20000 images for image saliency evaluation, the dataset is further commented in chapter 4.

1.6 Objectives and contributions of this thesis

1.6.1 Main objectives of this work

Overcoming the Dichromatic reflection model

As explained in section 1.4, physics based segmentation methods have experienced no meaningful advances in last years, even though they are a good theoretical framework for bottom-up image segmentation. One objective, and further, one of the main motivations of this work is to overcome the main shortcomings of the dichromatic reflection model [171]. This is a parametric model of the behavior of the light when reflected in a specific body and how it can be detected in the image histogram. It has good theoretical foundations but some limitations when applying on real images. The main problem is that it is a parametric model which turns into a too rigid methodology to find the expected structures as representative of an object. Basically, these structures are expected to follow two main orientations, whereas in real images, due to acquisition conditions, clipping, compression and some other effects, they have unexpected shapes. To overcome this problem, we were aimed to find a non-parametric operator to find these structures which correspond with material reflectances. Thus, we took the main idea of the dichromatic model, that is, that a material reflectance in a real scene is present in the histogram as an elongated object which is found to form a ridge-like structure. Ridges are extracted by means of the MLSEC-ST creaseness operator [123][122], which will be explained in detail in chapter 3. This operator jointly with a ridge extraction procedure was able to extract the structures representative of a material reflectance in a more flexible fashion than with the dichromatic reflection model.

Hybrid segmentation method

Bottom-up segmentation methods are classified in three main categories [180][34][126]:

- Feature-based: when performing the segmentation in the histogram domain.

- Image-based: when using the *spatial coherence* of the image, that is, relationships between pixels where its spatial position in the image is considered.
- Physics-based: are those which use physical properties of the light to perform the segmentation.

The method initially presented in this work to overcome the main shortcomings of the dichromatic reflection model is a feature-based segmentation method inspired in a physics based one. Nevertheless, our initial proposal does not consider the properties and strengths that might come from image based approaches. Ideally, we want to define a model of segmentation which might consider all the main contributions of each segmentation category, yet avoiding as far as possible its weaknesses.

Non-supervised evaluation

A segmentation method has to adapt to image content. Some images present meaningful objects at different scales, some images are indoor, some other outdoor, and so on. Each image requires a level of coarseness in the segmentation, what implies to fit the segmentation parameters to each image. In can be done either interactively or automatically. Needless to say that facilitating a method to do it automatically is the best option. Hence, an important objective of this thesis is to propose a method to automatically evaluate a segmentation in order to decide among a set of segmentations which adapts better to a given image or problem.

Main contributions of this work

Regarding the first objective, namely, to find a methodology to overcome the main shortcomings of the classical dichromatic reflection model, we have successfully proposed an alternative which has demonstrated to outperform state-of-the-art segmentation methods. The method presented is called RAD.

The second objective, that is, to propose a hybrid segmentation method, has been also performed by means of an statistical approach and a learning procedure to force the ridges (feature-space) to follow those directions expected for the dichromatic model (physics-based). Additionally, we have included the spatial coherence of the image (image-based) with a multiscale approach based on saliency. The resulting method, spRAD, outperforms the segmentation method initially proposed (RAD).

Non-supervised evaluation has been achieved by means of image saliency. A new method of image saliency has been proposed. This method outperforms the Itti [94] saliency model. The non-supervised evaluation performed using this new model of saliency also outperforms sate-of-the-art evaluation methods.

As a results of the evaluation we have proposed a general schema to evaluate image segmentation which can be also used to combine a set of segmentations of the same image to compose a combined segmentation image which is demonstrated to clearly outperform state-of-the-art segmentation techniques, including RAD and its further improvements.

1.7 Organization

- In chapter 2 we draw the state of the art in image segmentation. We classify segmentation methods in four main categories, namely, feature-based, image-base and physics-based and hybrid methods. Additionally, we briefly describe the main techniques and error measures of segmentation evaluation.
- In chapter 3 we present a segmentation method called RAD. It is focused on segmenting a single material reflectance (including shadows and highlights) by means of a topological analysis of the color histogram. RAD is able to overcome the main shortcomings of the dichromatic reflection model.
- In chapter 4 we detail a saliency method based on the information of the image chromatic derivatives. This saliency measure forms the basis of the our final proposal for image segmentation.
- In chapter 5 an extension of RAD in presented. Firstly, we add physics based statistics to our model. Afterwards, we include our saliency measure to form the final proposal of image segmentation called spRAD.
- In chapter 6 we further use our saliency measure for unsupervised segmentation evaluation.
- In chapter 7, we draw the conclusions of the present dissertation along with a discussion on image segmentation.
- In appendix A, we describe the most commonly used color spaces.

Chapter 2

Survey and State of the art in Image Segmentation

In this section we draw the state of the art in image segmentation. We classify segmentation methods in three main categories, namely, feature-based, image-base and physics -based segmentation methods. We also describe the main advantages and shortcoming of these different approaches and we also describe the main methods of each category. Finally, we briefly describe a category of methods, called hybrid methods, which combine techniques of the three main categories. In addition to the segmentation algorithms, in this chapter we briefly describe the main techniques and error measures of segmentation evaluation which is, as the segmentation itself, still an open issue.

2.1 State of the art in image segmentation

Due to its relevance as a preprocessing step in several computer vision applications, image segmentation has been widely studied and, consequently, there exist several different methods covering a broad spectrum of points of view. The main surveys on Image Segmentation can be found in [154] [182] [225] [34] [113] [126].

An initial classification of segmentation techniques was presented by Skarbek and Koschan in [180]. This work draws the basis of the current classifications of segmentation methods. Concretely, the authors propose a classification as follows:

1. Pixel based segmentation.
    - Histogram thresholding.
    - Clustering in colour space.
    - Fuzzy clustering in colour space.
2. Area based segmentation.

- Region growing.
- Split and merge.

3. Edge based segmentation.
   - Local techniques.
   - Global techniques.
4. Physics based segmentation.
   - Inhomogeneous dielectrics.
   - General approaches.

This initial division in four main categories has been modified as in [34] and [126]. Hence, current authors classify first techniques (pixel-based), as *feature space analysis* methods, since they perform the segmentation in the histogram space. Furthermore, *area-based* and *edge-based* methods work directly in the image space. For this reason, this methods are currently classified as *image-based* segmentation methods. Hence, we can classify segmentation methods as follows:

1. Image-based segmentation.

   Exploit the spatial information contained in the image, named *spatial coherence* [95].
   - Deformable models.
   - Graph-based approaches.
   - Region growing and edge detection.
   - Split and merge.
   - Topological methods.
   - Other methods.
2. Feature-based segmentation.

   These methods are focused on the photometric information of an image represented on its histogram [5] [221].
   - Histogram thresholding.
   - Clustering Techniques.
     Hard Clustering.
     Fuzzy Clustering.
3. Physics-based methods.

   These methods use the knowledge about the physical formation of the scene (light, surfaces reflectance), to perform the segmentation.
   - Segmentation based on the dichromatic reflection model.

- Spatial transformations.

4. Hybrid methods.

   Hybrid techniques combine methods of the previous categories.

This classification put in concordance the different categorizations which can be found in existing literature. For instance, in [113] we find a division in three categories: ***Stochastic Techniques*** instead of *Feature Space methods* ***Structural Techniques*** instead of *Image Based methods* and ***Hybrid Approaches***. Therefore, in [113] *Physics Based methods* are not described and a new division of *Image Based techniques* is done. It does not imply a misclassification, but the fact that segmentation techniques differ depending the framework. Just as [154] makes a survey of gray-scale segmentation and [113] do it on medical image methods, a survey of segmentation methods in the motion framework can be found in [226] and is also treated, briefly, in [40]. Whereas medical image segmentation methods are quite related with this state-of-the-art, motion segmentation has its own techniques and will not be described in this chapter.

Another example of classification is the work presented on [34], where the author divide segmentation techniques in six different points: *Histogram Trhesholding* (which includes clustering methods), *Region based* (region growing, region splitting and merge), *Edge Detection*,*Fuzzy Techniques*, *Physics based* and *Neural Network approaches*. Other reviews of colour segmentation are [126], [62], [5] or [40]; a survey on intelligent interactive segmentation methods, e.g., oriented to user intervention, can be read on [149].

2.1.1 Image-based segmentation

Image-based segmentation methods are those which use image's *spatial coherence* to perform the segmentation, namely, the relationship existing between pixels in the image domain. These methods mainly aim to detect the borders of the objects/surfaces in the scene. Several different methods have been proposed, which we classify in six subcategories.

Deformable models

Deformable models are one of the first techniques specifically proposed for image segmentation. These have been studied and improved among the last years. [138]. Deformable models are further split in parametric active contours and geometric active contours. One of the early methods of parametric active contours with acceptable results was the *snakes* proposed in 1988 by Kass, Witkin and Terzopulos [102]. A snake is an energy-minimizing spline guided by external constraint forces and influenced by image forces that pull it toward features such as lines and edges. Snakes are active contour models: they lock onto nearby edges. Deformable models include the already mentioned snakes as well as all those methods which deal with active contours. An extensive survey of them can be found in [13]. Since the original and simple method proposed in [102] was too affected by local irregularities, initialization, and local minima in the energy minimization function, subsequent methods have been

focused on the addition of new cues to guide the active contours towards a better convergence to meaningful borders. Well known techniques are the geodesic active contours, as proposed Caselles, Kimel and Sapiro in [28] or the introduction of a new external force derived form the image gradient called *gradient vector flow* as detailed in [219]. Recently, the combination of active contours with other techniques have been proposed to solve the problem of local minima in the energy minimization function as proposed by Bresson *et al.* in [21]. Another approach to avoid both local minima and self-crossing contours was suggested by Xu, Ahuja and Bansal in [220], by means of a combination of a graph cut based methodology to iteratively guide the contour deformation.

Graph-based approaches

Graph based approaches for image segmentation have been widely studied and several variations of them have been proposed. Graph-based methods include techniques such as the *intelligent scissors* algorithm [143], based on Dijkstra's shortest path algorithm, detailed in [6]. These methods treat the image as a graph and the user places several marks along the desired object boundary. Then, Dijkstra's shortest path algorithm is used to find a minimum length path connecting all marks and this path is returned as the object boundary. The main drawback of this approach is that it is affected by noise in the images. Other well-known approach is the normalized cuts algorithm [173]. Graph-cuts algorithms [217] are further analysed in [20]. The shortcomings of these methods are their difficulty to segment elongated objects and that they tend end up at local-minima. These problems are treated by Vicente, Kolmogorov and Rother in [205] and with the proposal of *ratio cut* as explained in [210]. Finally, another technique of graph-based segmentation which deserves special attention is the random walk [77]. Graph cuts and random walk approaches have been combined in an interesting approach by Sinop and Grady in [178].

In general, the main drawback with the graph-based segmentation algorithms is that they tend to be excessively time consuming. The *efficient graph based* segmentation algorithm proposed by Felzenszwalb and Huttenlocher is an example that a graph-based segmentation algorithm [58] can be fast yielding good results.

Region growing and edge detection

Whereas edge detection techniques could be classified as a category *per se* we have chosen to include them together with region growing algorithms for two reasons. First, the rest of the categories commonly look for the borders of the regions and are, therefore, edge-detection methods. Second, because gradient information-based edge detection is commonly a criterion of stability for the region growing procedure [159] [40] or determinant to define what a region is [61]. Another example was propose in [62] where the authors use a homogeneity criterion consisting of the weighted sum of the contrast between the region and the pixel, and the value of the modulus of the gradient of the pixel.

Region growing methods have in common that they begin with the positioning of a seed and afterwards, a criterion of growing from these seeds is established in order to converge to the homogeneous regions in the image [4]. This family of methods

propose different criteria for stability as explained in [126]. These techniques include Markov Random Fields or Neural Networks [73] [52] among others [126]. Region growing techniques strongly depend on the selection of the growing mechanism and on the initialization of the seeds, which might causes fairly different results form the same image [62] [175].

Split and Merge

Typical split and merge techniques [33] consist of two basic steps. First, the whole image is considered as one region. If this region does not satisfy a homogeneity criterion the region is split into more regions (the number depends on the method) which are tested in the same way. A common structure used for the split process is the quadtree representation [19]. Afterwards, in the merging step all adjacent regions with similar attributes may be merged following other (or the same) criteria. A common methodology used in the merging procedure is the region adjacency graph [19] [84]. Therefore, in these methods the main characteristic is the criterion used for both split and merge steps. In this sense, for instance, Dai and Maeda [44] propose a pyramidal method which uses statistical geometrical features as texture descriptors. Another common criterion is the color homogeneity as in [101]. Other techniques used are by means of MRFs or Neural Networks [126][62].

Topological methods

We include among topological methods those approaches which consider the image a landscape using its intensity values. Among these methods, the most well-known, and widely used is the watershed algorithm, which aims to find the catchment basins of a landscape. Vincent and Soille [206] define a catchment basin associated with a local minimum M as the set of pixels p of the landscape such that a water drop falling at p flows down along the relief, following a certain descending path called the downstream of p, and eventually reaches M. The original method proposed in [206] is based on the gradients of the image and it is too affected by local irregularities. A new version of the algorithm was presented in [69].The proposed algorithm, is not based on the gradient vectors of a landscape but on the idea of *immersion* which is more stable and reduces over-segmentation. Basically, the flooding process begins on the local minima and, iteratively, the landscape sinks on the water. Those points where the water coming from different local minima join, compose the watershed lines. To avoid potential problems with irregularities [123], a more proper marks instead of the local minima have to be found as for instance the ridges used in [200]. More segmentation techniques based on watershed can be found in [170], [29] and [82], where a learning procedure based on the borders drawn in the Berkeley benchmark [135]. Finally, a recent evaluation of watershed-based algorithms can be found in [29].

The second group of algorithms inside the category of topological methods is those which treat the borders of the borders as ridges and, therefore, perform a ridge-extraction algorithm. A good example of these methods can be found in [183]. One of the best methods of ridge extraction was the MLSEC-ST algorithm introduced by Lopez *et al.* in [123]. The same author presented an extensive study of ridge extraction algorithms in [122].

Other methods

There is a set of methods which can be hardly classified as a single category. The most clear among all of them are the ones which combine two or more of the previous categories of image-based segmentation methods. For instance, a method which uses a the watershed algorithm to merge the regions of the image [84] or a method which combines edge detection and nCuts [131]. Finally, we include to this category Neural Network approaches [52] and Markov Random Fields [18] [14] [103] approaches.

2.1.2 Feature-based segmentation

Feature-based segmentation methods are those which perform the segmentation in the image's histogram. Feature-based methods can be further split in three main categories: histogram thresholding, clustering and fuzzy clustering.

Histogram thresholding

Histogram thresholding techniques assume that there exist a threshold value that isolates all pixels representative of an object in a scene. Early methods in segmentation use this idea, as summarized in [182] and [154]. Nonetheless, this idea has been exploited in many different ways. For instance, with the idea of the probability density functions [81]. A compilation of such techniques can be found in [169]. The main shortcoming of these methods is that they are affected by local minima and irregularities. It has been treated by means of a segmentation in the CIE Luv space plus a k-means-based postprocessing [127] or with the fuzzy thresholding, as proposed in [190]. More recently, the histogram thresholding has been integrated with the Parzen window technique to yield a more adaptive thresholding [209]. Finally, as in the previous categories, Neural Network approaches are also used for histogram thresholding [47].

Hard Clustering techniques

Hard clustering techniques divide the histogram space in a set of well-defined clusters or regions. Probably the most used and well-know technique of hard clustering is the k-means algorithm. Its main limitations are that the number of clusters has to be known before the segmentation. This issue is treated in [162] and for the ISODATA algorithm, as explained in [186]. After the k-means algorithm and its variation, the other widely-used algorithm of hard clustering is the Mean shift [64], proposed as a segmentation method in [39] [37]. This method is based on probability density functions. Basically, it looks for the modes of the landscape and afterwards an iterative process is proposed to find its basis of attraction, that is, the clusters of the histogram obtained from the local maxima. As happened with the image-based techniques, here we also find some interesting approaches based on topological features. For instance, a watershed-based algorithm in the CIE Luv space [170] or with a watershed plus ridge-extraction algorithm called RAD, as proposed in [200].Another common methodology of hard clustering is composed by the spectral clustering methods, which are based on the Karhunen-Loeve transformation [49] [145]. A comprehensive survey of spec-

tral clustering methods can be found in [204]. Some other well-known techniques of pattern recognition can be used, as for instance, a k nearest neighbors based segmentation as explained in [60]. These approaches require to determine well-discriminative features.

Fuzzy Clustering techniques

In these approaches, a membership functions have to be obtained in order to perform a fuzzy classification of the feature space [185]. For instance, the fuzzy version of k-means belongs to this category. This method is called fuzzy c-means and it was proposed by Bezdek and Ehrlich in [11]. In this fuzzy version of k-means, the grade of belonging to a class is weighted by the distance of each point to the center of the class. These segmentation methods have been evolved to several variations. For example, a version which considers not just distance but orientation [166] or, more recently, the proposal of the possibilistic c-means, which propose a normalization among all distances [155]. As happened with the k-means, the initialization of the method can affect the results. This issue is analyzed and treated in [105]. An extension of fuzzy c-means which has turned in a segmentation method *per se* is the Gath-Geva algorithm, originally proposed in [68]. It combines a basic fuzzy c-means algorithm with the fuzzy maximum likelihood estimation, based on density criteria. It is proposed as a method much more adaptive to different cluster's shapes than the fuzzy c-means algorithm, which is one of the main drawbacks of both k-means and fuzzy c-means. This method has been further extended in different ways, for instance, by the addition of an expectation maximization methodology, as suggested in [3]. Another big family of fuzzy-clustering techniques are the mixture models [1] [129] which are a way to look for areas of high density. The fuzzy membership is also treated by means of neural networks, as in [203] or [119]. Finally, a fairly recent analysis of clustering techniques can be found in [5].

2.1.3 Physics-based segmentation

Physics-based segmentation methods are those which model the physical behaviour of the light in the scene. They are further classified in two main categories.

Segmentation based on the Dichromatic Reflection Model

The main contribution to these techniques was done by S.A. Shafer in 1985 with the introduction of the dichromatic reflection model (DCM) [171]. DCM, has been the basis of several segmentation techniques [7] [106], which limitations regarding different materials (metals and inhomogeneous dielectrics) geometry and non Lambertian surfaces have been also treated [136] [137] [151]. An interesting approach where a fuzzy reasoning has been included in the DCM model can be found in [223].Furthermore, physical formation of the scene has been also the inspiration of some other approaches, including our approach, pRAD [200]. Thus, DCM explains under a theoretical point of view the sort of shapes that a single surface can form in the histogram due to illumination interactions. The fact that these shapes do not correspond with the common feature-based clustering techniques such as Mean Shift [64] [39] which can not give the

elongated shapes described by the DCM. Some other proposals to find these structures are, for instance, with an statistical approach based on b-splines fitting in the HSV [104], or by means of a generalized Hough transform method, gradient descent method, and eigenvectors method as suggested in [146].

Spatial transformations

In addition to these approaches we include within physics-based approaches those models of color spaces proposed to cope with shadows and highlights. The first good proposal for this aim, was the Ohta space [148] proposed in 1980 which is a linear transformation of the RGB space that has been used in several approaches for images segmentation. Other interesting proposals for color spaces robust to, or that deal with, shadows and highlights, comprises an eigen color representation [2], an illuminant independent log-opponent representation [10] or an specific model to deal with color distortion [150].

2.1.4 Hybrid segmentation

There is a set of segmentation methods which can not be classified in one of the above detailed categories since they are a combination of two or more of them. Typical hybrid methods are those which add image spatial constraints to some feature-based segmentation techniques such as k-means [157] or more recently with fuzzy c-means [38] [23]. The JSEG segmentation method [46] is a two-step schema following a similar idea. First, a clustering of the color space is performed. Afterwards, a criterion of *good* segmentation is applied using the spatial coherence of the image. Another schema proposes that a good segmentation region should be formed by strongly connected pixels with homogeneous colors [130]. A criteria to combine color homogeneity with the texture in the image space by means of Gabbor filters is proposed in [32]. A similar idea, with the addition of a Markov Random Field model is proposed in [103]. An approach to combine image edges and color features by means of a Bhattacharyya-based Gradient Flow approach is proposed in [141]. Finally, in this thesis we also present a hybrid segmentation method, which combines RAD (feature-based) segmentation with image's spatial coherence in a context of saliency.

2.1.5 Discussion

Bottom-up segmentation methods handle the problem in different ways. More robustness is expected from those methods that combine the main strengths of each category while minimizing its weaknesses.

The main strengths of each method are:

- Image-based approaches: since these methods exploits the image spatial coherence, the borders of the segments tends to better coincide with an objects borders. Spatial coherence is lost in the other segmentation categories.
- Feature-based approaches: methods belonging to this category segment in a more consistent way than image-based approaches the colors appearing on the image, which are expected to represent objects or parts of the objects.

- Physics-based approaches: these methods follow a similar aim than feature-based, although even more robustness to chromatic changes from shadows to highlight is to be expected.

For the other side, the weaknesses of these categories are:

- Image-based approaches: abrupt illumination changes resulting from shadows and highlights as well as certain textures might affect these approaches. The main effect is oversegmentation.
- Feature-based approaches: colors found in the histogram space do not always correspond with physical objects or surfaces in the image.
- Physics-based approaches: they share the main drawback with physics based approaches.

Hybrid approaches are expected to combine the strengths of each category. A good example is Mean shift, which look for certain structures in the histogram space yet including the image space,

The method presented in this dissertation has been build to include the main strengths of each of the three categories which, as results hold, improve the performance of our approach.

2.2 State of the art in segmentation evaluation

Image segmentation evaluation is still a challenging issue. Some authors argue that it can be evaluated only in the context of the task in which the segmentation is done. Moreover, it is also considered an ill-posed problem since there is no consensus on which is the best segmentation of an image as explained in [172]. After a set of experiments with 14 subjects, the authors conclude that each subject performs a different segmentation. Therefore, to answer the question about what a *correct* segmentation is, turns in a difficult issue to handle. From this, we can conclude that a simply hand-made ground truth does not allow comparing several methods, further than in a specific context where the objects to be segmented are clearly identified.

Consequently, the question is if we can find a performance evaluation method of general purpose. In this section we analyse the different proposals about image segmentation evaluation by focusing in three main subjects, namely, ground-truth generation for supervised evaluation [222], error measures on supervised evaluation [225] [226] and unsupervised segmentation evaluation [224].

2.2.1 Ground-truth generation for supervised segmentation evaluation

If a hand-made benchmark for evaluation is not valid for a general purpose, how we can validate a method? This is the first question we need to solve. Despite its relevance for the segmentation field, this issue has not been treated deeply enough and just some articles about it can be found. Similarly, few useful ground-truths are

nowadays available for the scientific community. The question rising is, can we yield a real and useful general purpose segmentation benchmark or we rather have to give a specific solution to any framework?

As we have told before, it feasible to generate a hand-made ground truth to validate a method in a concrete context such as human skin segmentation or in some industrial applications. In these cases, we know what a good segmentation is. But, what happens when we want to know if a segmentation method is good enough to be applied, *a priori*, in any context? The works presented in [172], [135] and [70] suggest a possible solution.

These kind of psychophysical works, (also named psychovisual [172]) start with the idea that even the human subjects used in experiments do not coincide in the segmentations done, as shown in Figs. 2.1 and 2.2. Nonetheless, either in [172] and [135] the authors conclude that even though the segmentations are not the same, all of them share some characteristics and can be combined in some way to find a ground truth of general purpose. In [135] the authors argue that a given segmentation of an image is, in fact, a refinement of another one or vice versa. Following this reasoning, the segmentation showed in figure 2.1c is a refinement of 2.1b, and 2.1d is a refinement of 2.1c. Evidently, all these segmentation have to be considered as a correct segmentation since all of them have been performed by a human subject. This dataset, called *the Berkeley segmentation dataset and benchmark*, proposed in [135] has become a standard in the evaluation of image segmentation. Nonetheless its main limitation is that its variability is fairly low, that is, the set of images of this dataset are all of them quite similar.

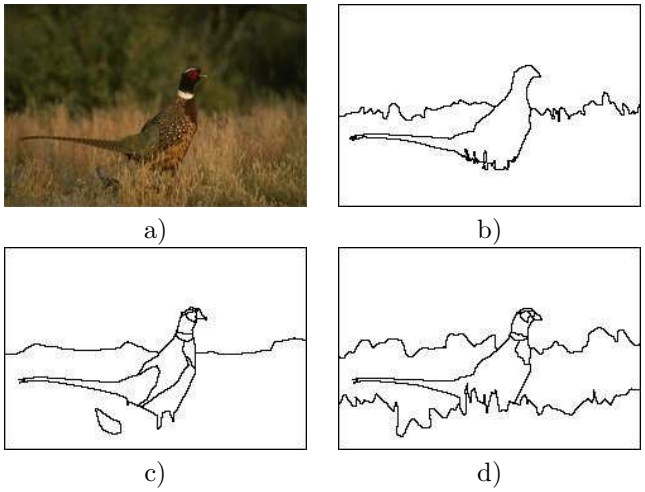


Figure 2.1: Human segmentation example of the Berkeley Dataset. (a)Original image. (b,c,d)Segmentation of three different human subjects.

In the last years the interest for object recognition techniques has resulted in the generation of new benchmarks. For instance, the PASCAL challenge which appears for the first time in 2005 [56]. This is a benchmark for object recognition focused on

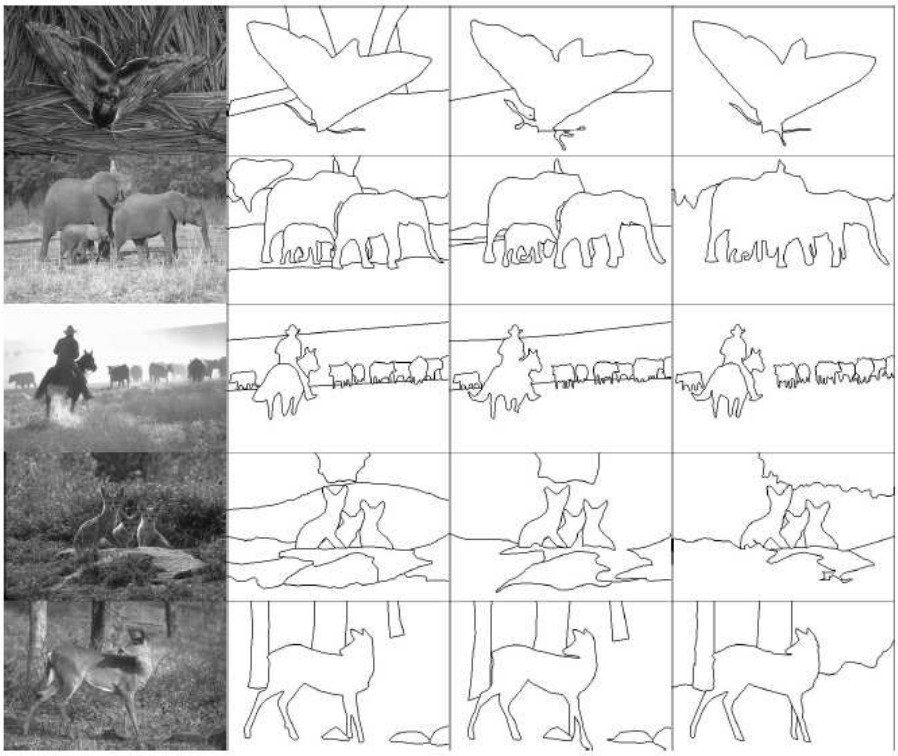


Figure 2.2: More examples belonging to the Berkeley segmentation datset and benchmark.

several classes. This dataset has been modified and extended year after year. This dataset is divided in object segmentation and class segmentation. The difference between both can be seen in Fig. 2.3. These examples belong to the PASCAL VOC2009.

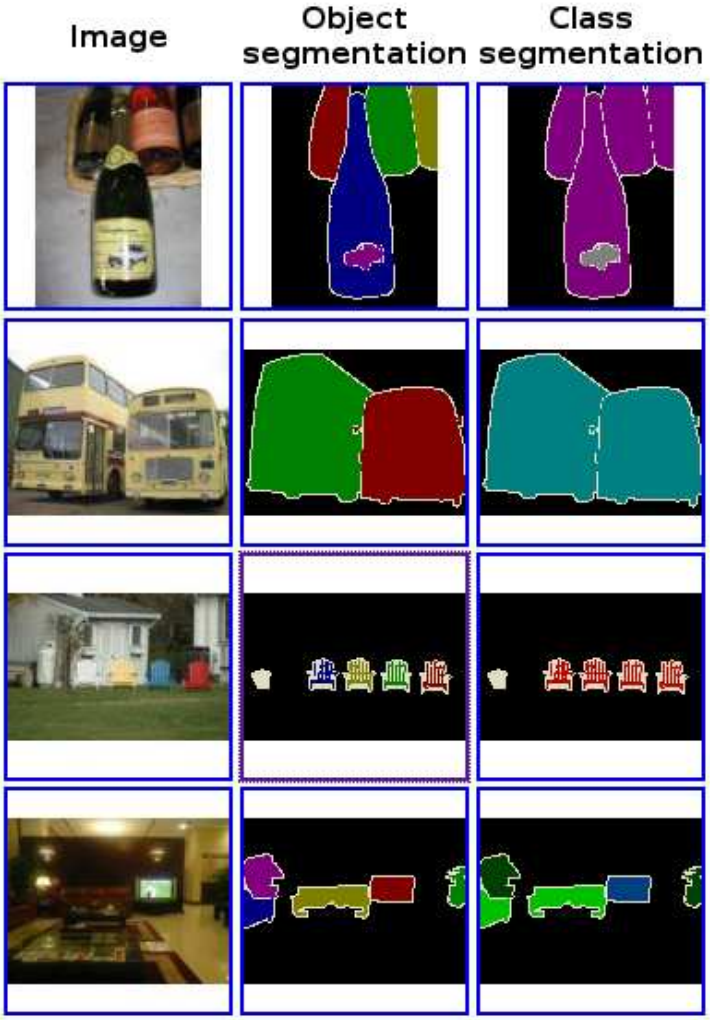


Figure 2.3: Examples of PASCAL's object-class based dataset.

The object segmentation pixel indices correspond to the first, second, third object etc. The class segmentation pixel indices correspond to classes in alphabetical order (1=aeroplane, 2=bicycle, 3=bird, 4=boat, 5=bottle, 6=bus, 7=car , 8=cat, 9=chair, 10=cow, 11=diningtable, 12=dog, 13=horse, 14=motorbike, 15=person, 16=potted plant, 17=sheep, 18=sofa, 19=train, 20=tv/monitor) The main drawback of this dataset is that it is focused on 20 classes and images are not completely valid for general purpose segmentation, but for object recognition. Nonetheless, a good segmentation is expected to segment these objects in an acceptable way.

Another good example of a large-scale general purpose dataset is the LabelMe [164], which is a collaborative web-based open annotation tool. In this dataset it can be found a benchmark consisting on thousands of images where 9 objects have been labelled. Images for training are partially labelled, whereas images for test are completely labelled, in opposition with the PASCAL dataset. It is also facilitated a Matlab toolbox to handle the information contained in the images. The database consisted of 111490 polygons, with 44059 polygons annotated using the online tool and 67431 polygons annotated offline. There were 11845 static pictures and 18524 sequence frames with at least one object labeled. Right now, there are 181318 images of which has been annotated a total of 56830. An example of an image is showed in Fig.2.4

Figure 2.4: Example of LabelMe's dataset.

The main drawback with this dataset is that there is no control among the segmentation of the images and fairly incorrect segmentation can be found. Nonetheless, it is possible to select images from this dataset to generate a benchmark which could be used instead of the Berkeley one.

An interesting methodology to generate a large-scale general purpose dataset for image segmentation is presented in [222]. As in the case of LabelMe, the dataset is continuously growing. At the moment of the publication of [222] it was 636,748 images and video frames. In this dataset a schema composed on three parts for image

annotation is proposed.

- Scene Level: Global geometry information, scene category (indoor/outdoor), events and activities
- Object Level: Hierarchical decomposition, object segmentation, sketching and semantic annotation
- Low-middle Level: Contours types (object boundary, surface norm change or albedo change), Amodal completions, Layered representation, etc

They propose a hierarchical classification of the image which goes from the scene level to the details of each object, as showed in Fig.2.5

Another example of dataset are the Caltech-256 presented in [79]. Finally, an interesting extension of the Berkeley dataset and benchmark is presented in [9], where the addition of high-level semantics is proposed.

All the datasets mentioned before, are focused on the segmentation of certain classes of objects or the whole image. Another point of view is presented in [70]. In this paper, the authors are focused in the most salient object of the image. The authors argue that the most salient object in a scene is always segmented by each human subject. Hence, we can expect that a good segmentation method, independently of its refinement, have to segment correctly the most salient object, which should be enough to evaluate a segmentation method. The main drawback with this dataset is that images are presented in a poor resolution and are not suitable for current segmentations methods which allow a better refinement. An example of the images proposed in this dataset based on saliency are showed in Fig.2.6.

2.2.2 Error measures for supervised segmentation evaluation

Whereas it might be a rather easy task to evaluate segmentation when a ground-truth is present, it becomes a difficult task. The main problem is that there is no simple correspondence between regions in the ground-truth and regions in the segmented image [152]. Hence, error measures are focused on the quantization of this overlapping and non-overlapping, in order to facilitate a quantitative measure of a segmentation method's performance.

On existing literature we can find three comprehensive surveys on segmentation evaluation. Two of them, presented by Y.J. Zhang in 1996 [225] and 2001 [226] are about supervised evaluation. The other, presented by H. Zhang in 2008 is about unsupervised evaluation [224] and will be commented in the next section. Finally, another interesting survey can be found in [152].

The schema depicted in Fig.2.7 is followed by [225] and [226].

As showed in this schema, supervised evaluation methods are divided in analytical methods, goodness methods and discrepancy methods. The author defines analytical methods as those which compute a set of characteristics of each method. For instance the amount of *a priori* knowledge that can be incorporated to a method or wether a method can be parallelised. Some of these methods are discussed in [225]. Although, its utility in image segmentation evaluation can be, at least, arguable. Fairly much

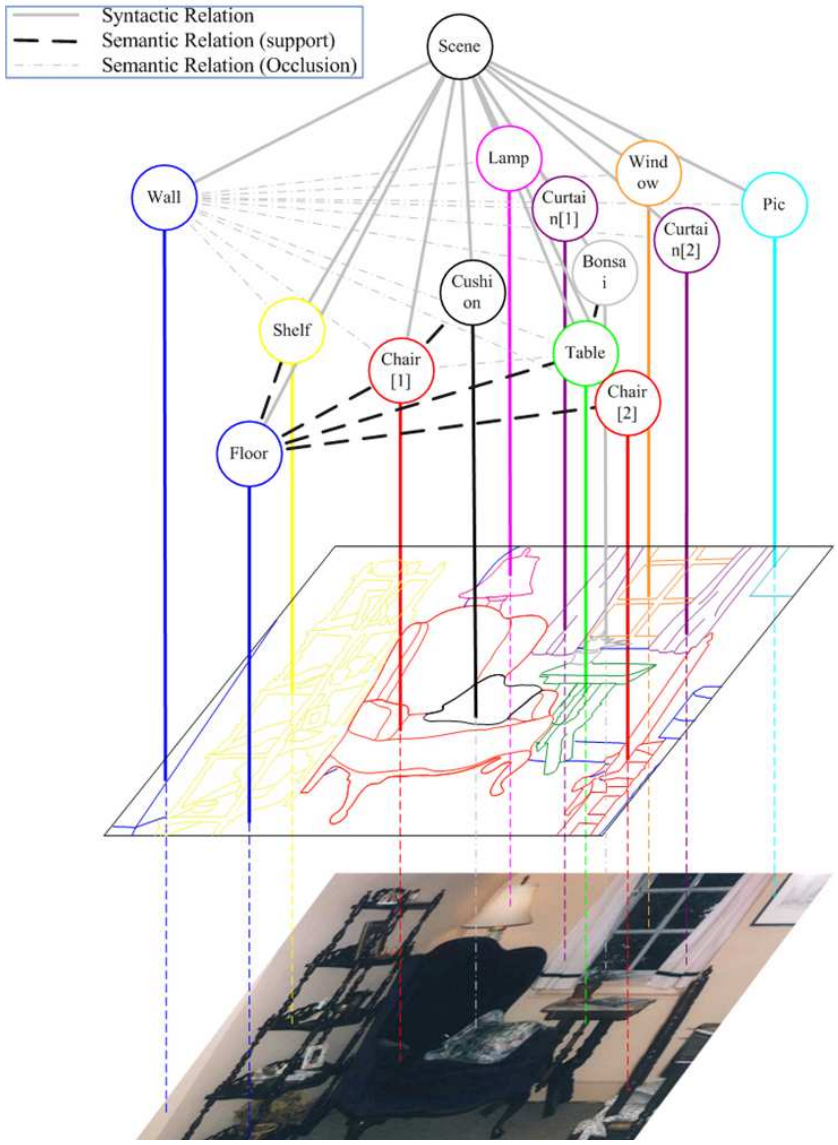


Figure 2.5: Example of Yao's approach.

useful are goodness and discrepancy methods, the ones that are called *empirical* in [225] and [226]. In this review we change the terminology *empirical measures* for *error measures* since they provide a correspondence among two images based on either its similarity or discrepancy. A survey on error measure can be found in [196] and an interesting comparison of four of the ones commented here can be found in [221].

The main error measures are the edge-based methods, the global constancy error, probabilistic rand index, variation of information and other generic and combined

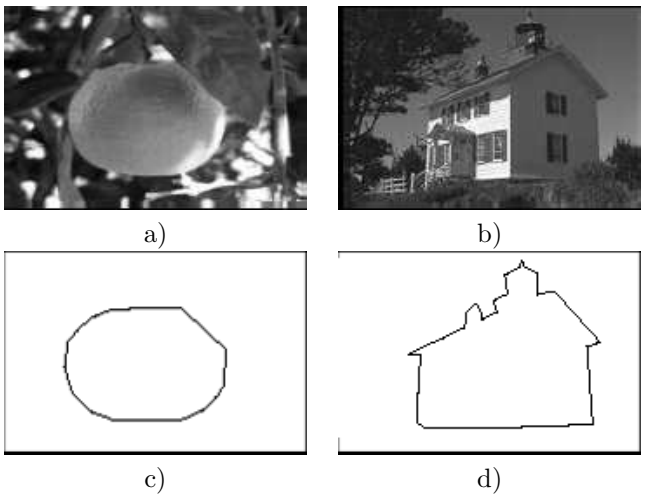


Figure 2.6: Salient object. (a,b)Original images. (c)Most Salient object of a). (d) Most salient object of b).

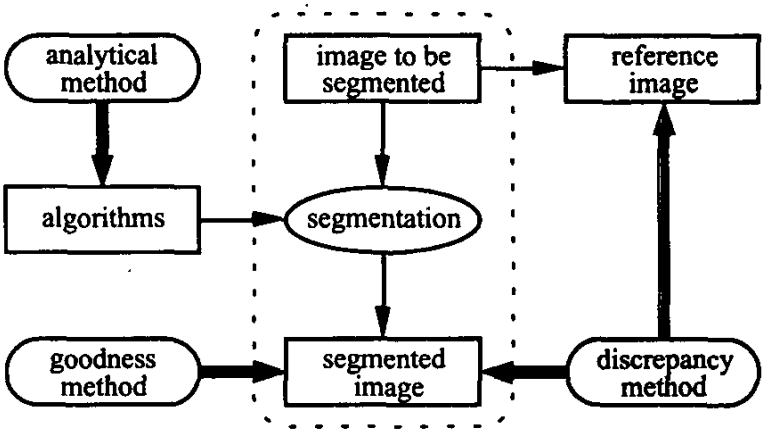


Figure 2.7: Classification of supervised evaluation methods.

measures.

Edge-based error measures

These measures rank a segmentation by considering the borders of the segments. Measures of similarity and discrepancy can be found in this category.

These error measures have a low-tolerance to refinement, but are useful to compare segmentations with a similar number of segments.

Boundary Displacement Error

This method, introduced in [89], evaluates the precision of the extracted region boundaries [62].

Let B be the estimated boundary and G^B the ground-truth boundary. The method uses two distance distribution signatures from the estimated to the ground truth borders, denoted by D_G^B and viceversa, denoted by D_B^G. For two sets of boundary points B_1 and B_2, $D_{B_1}^{B_2}$ is a discrete function whose distribution characterizes the discrepancy, measured in distance, from B_1 to B_2. The authors define this distance as the minimum absolute Euclidian distance. $D_{B_1}^{B_2}$ can be established from the distance histogram from individual $x \in B_1$ to B_2, which may be estimated through a distance transformation with respect to B_2.

The shape of $D_{B_1}^{B_2}$ defines the degree of similarity between B_1 and B_2. By means of the calculus of the standard deviation and the mean of $D_{B_1}^{B_2}$, it is possible to reflect the shape of the signature. Thus, a standard deviation near to zero means good approximation (without outliers). The same happens with the mean.

Uncertain Image Classification This method introduced by Martin Laanaya and Arnold-Bos in [134] takes care of both the well-classified and the bad classified border pixels. The singularity of this error measure is that it do not expect a *perfect* matching between segmented image and ground-truth, but is tolerant to some distance errors. The distance between a ground-truth pixel and a segmented image pixels is weighted buy a Gaussian. Hence pixels *acceptably* close are well ranked even when they do not perfectly coincide. Afterwards a measure of bad-classified pixels (those which are too far away forming a ground-thrust's pixels) is proposed. Finally, these two measures are considered in order to classify a segmentation method. The authors argue that combining a measure of goodness and badness for the classification is more robust than just considering one of them.

Global Constancy Error

This method, presented in Martin *et al.* [135], takes care of the refinement between different segmentations. Thus, for a given pixel p_i, consider the segments (sets of connected pixels), S_1 and S_2 that contain this pixel. If one segment is a proper subset of the other, then p_i lies in an area of refinement and the error measure should be zero. If there is no subset relationship, then S_1 and S_2 overlap in an inconsistent manner and the error is higher than zero.

Let $\setminus$ be the difference between two segments, $\|x\|$ the cardinality of the set x and $R(S_n, p_i)$ the set of pixels in the segmentation corresponding to a segment S_n containing pixel p_i. Then, the local refinement error is defined as:

$$E(S_1, S_2, p_i) = \frac{\|R(S_1, p_i) \setminus R(S_2, p_i)\|}{\|R(S_1, p_i)\|} \tag{2.1}$$

Finally, since this error measure is not symmetric, the authors define the Global Constancy Error (GCE) and the Local Constancy Error (LCE). GCE forces all local refinements to be in the same direction whereas LCE allows refinements in different directions. If n is the number of pixels:

$$GCE(S_1, S_2) = \frac{1}{n} min \left\{ \sum_i E(S_1, S_2, p_i), \sum_i E(S_2, S_1, p_i) \right\} \tag{2.2}$$

$$LCE(S_1, S_2) = \frac{1}{n} min \left\{ E(S_1, S_2, p_i), E(S_2, S_1, p_i) \right\} \tag{2.3}$$

As $LCE \leq GCE$, the authors propose to use GCE as error measure. The difference between LCE and GCE is graphically explained in figure 2.8. GCE allows refinements from (b) to (c) or (d) whereas LCE also accept refinement from both (c) to (d) and (d) to (c).

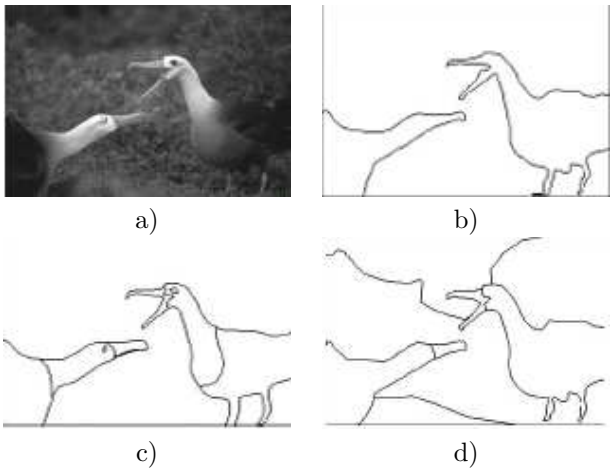


Figure 2.8: Example of refinements accepted. GCE: from b) to c) and d). LCE: also from c) to d) and d) to c). Consequently, GCE is tougher than LCE.

The main drawback of this method is that the segmentations to be compared are expected to have a very similar number of segments.

Variation of information

This technique, introduced by Meila *et al.* in [139] is a clustering comparison method based on the information theory.

For a discrete random variable taking K values, uniquely associated to the clustering C, we define $H(C)$ the entropy of the cluster C as follows:

$$H(C) = -\sum_{k=1}^{K} P(k) \log P(k) \tag{2.4}$$

where $P(k)$ is the probability of a given point to be in the cluster C. The entropy is always positive and takes value zero when there is just one cluster, i.e., when there is no uncertainty. This entropy is measured in bits. An uncertainty of one bit corresponds to a clustering with $K = 2$ and $P = (1) = P(2) = 0.5$.

In the calculus of the variation of information is also needed the *mutual information* (MI), which is a measure of the information that one cluster has about another. Let $P(k)$, $k = 1...K$ and $P'(k')$, $k' = 1...K'$ be the random variables associated with clusterings C and C'. Then we define $I(C, C')$ the mutual information as:

$$I(C, C') = \sum_{k=1}^{K} \sum_{k'=1}^{K'} P(k, k') \log \frac{P(k, k')}{P(k)P'(k')} \tag{2.5}$$

Where $P(k, k')$ is the *joint probability distribution function* defined. Denote n the total number of data points:

$$P(k, k') = \frac{|C_k \bigcap C'_{k'}|}{n} \tag{2.6}$$

if the probability that a point belongs to C_k in C and to $C'_{k'}$ in C'. The MI of two independent random variable is 0 and the MI value increase insofar the uncertainty increase, but in any case exceeding the total uncertainty. Thus,

$$I(C, C') \leq min(H(C), H(C')) \tag{2.7}$$

Finally, the Variation of information between two clusters, $VI(C, C')$, is defined as follows:

$$VI(C, C') = (H(C) - I(C, C')) + (H(C') - I(C, C')) \tag{2.8}$$

A graphical example is depicted in 2.9.

A further analysis of this error measure can be found in [140]. In the same article, a brief analysis of the most relevant techniques for clustering comparison can be also found. a more extensive analysis of clustering-based comparison techniques as well as a proposal of a new one can be found in [96].

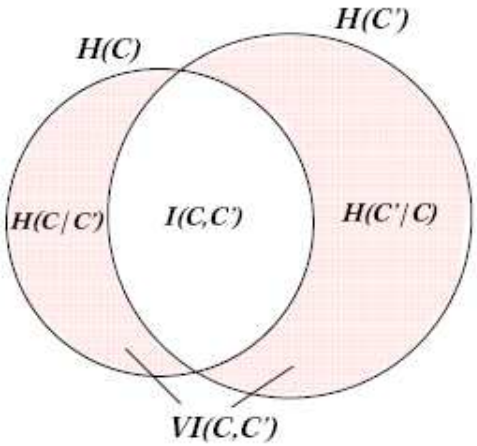


Figure 2.9: Graphical example of VI.

Probabilistic Rand Index

This measure is a variation of the Rand Index (RI) [161] by including the ability to allow some kind of refinements in the segmentation [196].

Let $S_1, S_2, ..., S_k$ be a set of ground truth segmentation images and S_{test} a segmented image to be compared with the ground truth , we will consider a segmentation S_{test} as 'good' if its labels correspond to the pairwise labels in S_i. That is, for any pair of pixels x_i, x_j, and their labels l_i^{test}, l_j^{test} if these labels correspond to l_i^{Sk}, l_j^{Sk}, the labels on the ground truth for the same pixels, it will be considered a good segmentation. In [161] it is proposed the RI to compute that as follows:

$$R(S_{test}, S_k) = \frac{1}{\binom{N}{2}} \sum_{\substack{i,j \\ i \neq j}} [I(l_i^{test} = l_j^{test} \wedge l_i^k = l_j^k) + I(l_i^{test} = l_j^{test} \wedge l_i^k = l_j^k] \quad (2.9)$$

where I is the identity function and the denominator is the number of possible unique pairs among N data points.

In [196] it is demonstrated that RI does not allow the possibility of refinement in the segmentation. To avoid that, the authors propose the Probabilistic Rand Index (PRI), which combines the desirable statistical properties of the RI with the ability to accommodate refinements appropriately. The idea is to calculate, given the manually segmented images $S_1, ..., S_K$, the empirical probability of the label relationship of a pixel pair x_i and x_j:

$$p_{ij} = \frac{1}{K} \sum_{k=1}^{K} I(l_i^k = l_j^k) \quad (2.10)$$

Then, the PR index is defined as follows:

$$R(S_{test}, S_k) = \frac{1}{\binom{N}{2}} \sum_{\substack{i,j \\ i \neq j}} [I(l_i^{test} = l_j^{test})p_{ij} + I(l_i^{test} = l_j^{test})(1 - p_{ij})] \tag{2.11}$$

the measure takes value 0 when there are no similarities and 1 when S_{test} and S_K are exactly the same. A graphical example with a synthetic image is depicted in figure 2.10.

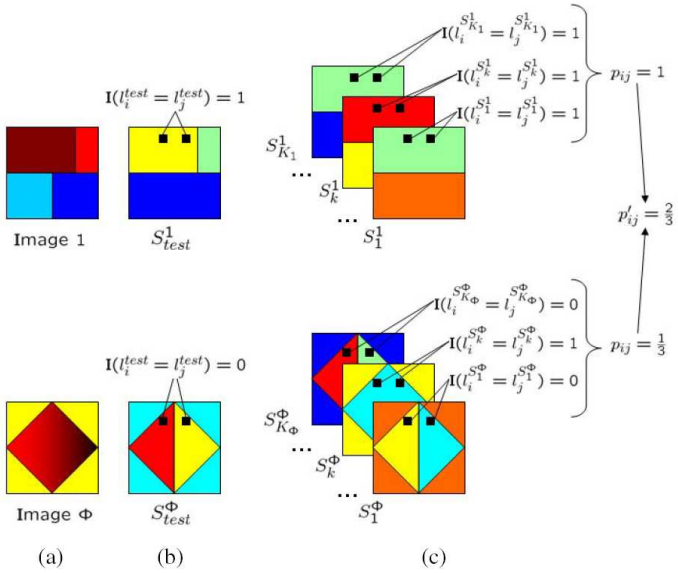


Figure 2.10: Extracted from [196]. Each row Φ has (a) an associated input Image, (b) a candidate segmentation S^{Φ}_{test} test and (c) a set of K_{Φ} available manual segmentations $S^{\Phi}_{K_{\Phi}}$.

The main drawback with PRI is that it is not possible to know if a certain score is good or bad. We can compare the score of two segmentation but we can not know if the difference is meaningful or not [156]. This is solved with the introduction of the Normalized PRI, proposed in [197].

$$NormalizedIndex = \frac{Index - ExpectedIndex}{MaximumIndex - ExpectedIndex} \tag{2.12}$$

The expected value of the normalized index is 0, so we know exactly when a segmentation is better than average or worse than average.

Generic and combined measures

Some other methods propose a combination of some properties related with segmentation evaluation.

Other clustering based measures It is the case of the work presented in [96]. In this article a set of measures of clustering comparison are analyzed and formulated. Basically, the same categories of clustering comparison as the ones described in [140] are described. This categories are:

1. Distance of clusterings by counting pairs: It counts the four different relationship existing between a pixel and two clusters one belonging to the reference image and one belonging to the segmented one

2. Distance of clusterings by set matching: This second class of comparison criteria is based on matching the clusters of two clusterings.

3. Information-theoretic distance of clusterings:Techniques basically based on Information theory and entropy as the variation of Information index already explained.

This generic measures are further analyzed and discussed in [96] and [140].

Evaluation based on overlapping matrix Another interesting review can be found in [152]. Here the authors stand-out that the main drawback of segmentation evaluation can be handled by considering three different measures:

1. The percentage of Correctly Grouped pixels: aims at accounting for those pixels which, belonging to a reference region R_i, are put together in a single output region R_j.

2. The percentage of under-segmentation: represents the amount of pixels of the image which have been assigned to regions R_j which cover several reference regions R_i.

3. The percentage of over-segmentation: accounts for pixels of output regions R_j which split a reference region R_i.

An extensive set of ways to compute these percentages is explained in the same article. Finally, the authors argue that a combination of such measures can be used to efficiently rank a set of segmentations.

Symmetric and asymmetric discrepancy measures Another interesting approach is the one suggested by Cardoso and Corte-Real in [25]. In this case, the authors present a combination of the following discrepancy measures:

1. Generic discrepancy measure: given by the normalized partition distance between the reference segmentation and the segmentation under study

2. Asymmetric measure for applications where over segmentation is not an issue

3. Asymmetric measure for applications where under segmentation is not an issue

4. Mutual partition distance: where mutual refinements can be tolerated.

2.2.3 Non-supervised segmentation evaluation

All previous error measures can be applied just when there is a ground-truth available. Nevertheless, to have a ground-truth suitable for any problem is just unfeasible. A ground-truth should be representative of any kind of problem and extensive enough to validate a segmentation method. As commented in section 2.2.1, the amount of ground-truth currently available are not well controlled or limited in many aspects. Due to that, the interest in non-supervised methods for segmentation evaluation has been increasing. Two surveys of these techniques can be found in [30] and [224].

Non-supervised evaluation is useful for several aspects:

- Rank a set of segmentations as with the supervised measures.
- Automatic selection of parameters.
- Combination (also called fusion) of different segmentation results of the same image to generate a final segmentation which theoretically takes the strength of previous segmentation.

Further details on these techniques can be found in [224]. Finally, we will analyze and compare unsupervised methods in chapter 6.

Chapter 3

Ridge-based Analysis of a Distribution (RAD)

The segmentation of a single material reflectance is a challenging problem due to the considerable variation in image measurements caused by the geometry of the object, shadows, and specularities. The combination of these effects has been modelled by the dichromatic reflection model. However, the application of the model to real-world images is limited due to unknown acquisition parameters and compression artifacts. In this chapter, we present a robust model for the shape of a single material reflectance in histogram-space. The model is a Ridge based Analysis of a Distribution (RAD). It is based on a multilocal creaseness analysis of the histogram, which results in a set of ridges representing the material reflectances. The segmentation method derived from these ridges is robust to both shadow, shading and specularities.
Qualitative results illustrate the ability of our method to obtain excellent results in the presence of shadow and highlight edges. Quantitative results obtained on the Berkeley data set show that our method outperforms state-of-the-art segmentation methods at low computational cost.

3.1 Introduction

Image segmentation is a computer vision process which aims to partition an image into a set of non-overlapped regions, called segments. A robust and efficient segmentation is required as a preprocessing step in several computer vision tasks such as object recognition or tracking. In real images changes due to illumination, shadow, shading and highlights provoke image measurements to vary significantly. These effects, are one of the main difficulties that have to be solved to yield a correct segmentation.

Image segments caused by a single material reflectance form complex shapes in histogram-space, due to shading effects and specularities. The fact that these physical effects lead to undesired image segments is also confirmed by Martin *et al.* in [135]. He points out the existence of strong edges caused by such physical effects which are not considered in human segmentations, but which tend to be detected by cur-

rent segmentation methods. Previous work on image segmentation robust to shading effects and specularities is based on the reflection model of the light. These methods, called physics-based, predominantly based on the dichromatic reflection model (DCM) [171] [106] are aimed to explain the behavior of the light in a scene. Thus, from a theoretical point of view, these models are able to explain the formation of shadows and specularities. These methods are based on several assumptions which severely limit their applicability. The main problem is the presence of artifacts introduced by acquisition conditions, clipped highlights or image compression. A second set of segmentation methods are feature-based [34] [126]. These methods are not based on prior assumptions of the underlying physics and are therefore more flexible to mentioned problems. However, ignorance of the physical process often leads to incongruences in the presence of shadows and highlights.

Here, we aim to combine the strengths of physics and feature-based methods. The presented method is based on the observation that the distribution of single material reflectance can be robustly represented by a single connected ridge in histogram space. The method is named Ridge-based Analysis of Distributions (RAD). The detection of these ridges is based on a creaseness analysis of the histogram. This technique connects the shadows in the dark parts of the object, to the brighter regions, and further up to the highlights (see Fig. 3.1). Furthermore, the ridges are capable to correctly connect single material textures, such as grass or sand. The advantage over previous physics-based methods is that our method does not assume a parametric shape, and is therefore robust for non-linear acquisition and image compression.

We propose two further extensions to the basic method. Firstly, to suppress those ridges in the less probable orientations and favor those ridges in the probable ones, we extend the method to exploit the image statistics of ridge orientations. This extension is called physics-based RAD (pRAD). Secondly, ridges on the histogram can just cope with those segments derived from single materials. Segments formed by more than a material will be represented by different ridges in the histogram. These textures tend to be present at certain scales, but display weak contrast at other scales. This fact is exploited by the multi-contrast representation of the image, in which texture contrast is suppressed. This method is called spatial RAD (sRAD).

This chapter is organized as follows: in section 3.2 we explain the related work in image segmentation. Afterwards, in section 3.3 we explain the theoretical basis and motivations of our approach. Subsequently in sections 3.4 and 3.5 we explain RAD. A comparison with Mean shift and a performance evaluation of our approach is done in section 3.6. Finally, conclusions of the current work are given in section 3.7.

3.2 Related work

There exist several different methods covering a broad spectrum of points of view. The work presented by Skarbek and Koschan [180], draws the basis of the current classifications of segmentation methods. Some other comprehensive surveys of colour segmentation techniques are presented in [34] and [126], where a similar schema is followed. From these works segmentation methods are divided in four main categories: image-based, feature-based, physics-based and hybrid approaches. Feature-based ap-

proaches are focused on the photometric information of an image represented on its histogram [5] [221]. Image-based approaches exploit the spatial information contained in the image, named *spatial coherence*. Physics-based methods use the knowledge about the physical formation of the scene (light, surfaces reflectance), to perform the segmentation. Finally, hybrid techniques combine methods of the previous categories. As stated before, this chapter introduces a method that performs an analysis of the histogram (feature-based method, RAD) exploiting the statistics of the ridges (physics-based, pRAD) and adding as a final step the spatial coherence of the image (image based, sRAD) . Therefore, the segmentation method presented belongs to the category of hybrid methods.

In regard image-based methods, these include region and boundary information [62] [40] graph-based approaches as nCuts [173] or the efficient graph-based image segmentation [58], region growing algorithms [159] [175] or segmentation based on watershed [82] [29] and in general topological approaches [183]. These basic techniques are either mixed [84] or complete by means of markov random fields [18] [103] or neural network approaches [52].

Feature-based methods can be further split in three main categories: histogram thresholding, clustering and fuzzy clustering. Histogram thresholding techniques assume that there exist a threshold value that isolates all pixels representative of an object in a scene. This basic concept is exploited in several ways as explained in [169]. Clustering techniques, perform a partition of the feature space under different criteria as described in [5]. Such criteria include distance measures as k-means or ISODATA [186], probabilistic/statistical approaches, such as Mean Shift [64], or the spectral analysis of the data [204], based on the Karhunen-Loeve transformation. Fuzzy clustering includes methods such as fuzzy c-means [166] [105], Gath-Geva clustering [68], or mixture models [1] [129] which are a way to look for areas of high density. From all of them, the most related work with RAD is Mean shift. Both, Mean Shift and RAD, use topological information (modes and gradients for Mean Shift and structural tensor, creaseness and ridges for RAD) to perform the classification of colors in the histogram space.

Physics techniques model the behavior of the light in the scene. The main contribution to these techniques was done by S.A. Shafer in 1985 with the introduction of the dichromatic reflection model (DCM) [171]. DCM, has been the basis of several segmentation techniques [7] [106], which limitations regarding different materials (metals and inhomogeneous dielectrics) geometry and non Lambertian surfaces has been also treated [137] [151]. Furthermore, physical formation of the scene has been also the inspiration of some other approaches, including pRAD. Thus, DCM explains under a theoretical point of view the sort of shapes that a single surface can form in the histogram due to illumination interactions. The fact that these shapes do not correspond with the common feature-based clustering techniques such as Mean Shift [64] [39], the most related feature-based technique with RAD is the other observation that forms the basis of our proposal. Mean Shift joints different modes and its basis of attraction with a method that can not give the elongated shapes described by the DCM. Instead of this, pRAD performs an analysis of the histogram space focused in the extraction of elongated shapes that can easily follow the directions of the DCM (if present in the histogram) but without its main restrictions. Some other proposals to

find these structures are, for instance, with an statistical approach based on b-splines fitting in the HSV [104], or by means of a generalized Hough transform method, gradient descent method, and eigenvectors method as suggested in [146].

In addition to these approaches we include within physics-based approaches those models of color spaces proposed to cope with shadows and highlights. The first good proposal for this aim, was the Ohta space [148] proposed in 1980 which is a linear transformation of the RGB space that has been used in several approaches for images segmentation. Other interesting proposals for color spaces robust to, or that deal with, shadows and highlights, comprises an eigen color representation [2], an illuminant independent log-opponent representation [10] or an specific model to deal with color distortion [150].

Finally hybrid approaches combine techniques of the three previous categories. For instance, by adding image spatial constraints (spatial coherence) to a clustering technique such as k-means [157] or more recently with fuzzy c-means [38]. The JSEG segmentation method [46] is a two-step schema following a similar idea. First, a clustering of the color space is performed. Afterwards, a criterion of *good* segmentation is applied using the spatial coherence of the image. Another schema proposes that a good segmentation region should be formed by strongly connected pixels with homogeneous colors [130].

In this work, we use the spatial coherence of the image to build a multiscale information-based chromatic contrast map, called *multicontrast image*. This map will guide a procedure to combine a set of sub-segmentations computed from a single image at different feature-space scales. Hence, we use the multicontrast image, to determine the goodness of a segmentation (or a segment) [87] [70]. This extension of the method is called sRAD.

3.3 Our approach: Theoretical Foundations

Our approach to colour image segmentation is based on the insight that the distributions formed by a single-colored object have a physically determined shape in colour histogram-space. We model an image as being generated by a set of segments, each of which corresponds with a material reflectance (MR) described by a distribution in histogram-space. Each MR is related to a semantic object in the image. For example, in Figure 1 we distinguish between four different MRs, namely: red for the pepper, green and brown for the branch and black for the background.

A MR generates many image values due to geometrical and photometric variations. Our main aim is to find a good representation of the topologies which MR's are likely to form in histogram space. For this purpose, consider the distribution of a single MR as described by the dichromatic reflection model [171]:

$$\mathbf{f}(\mathbf{x}) = m^b(\mathbf{x})\,\mathbf{c}^b + m^i(\mathbf{x})\,\mathbf{c}^i \tag{3.1}$$

in which $\mathbf{f} = \{R, G, B\}$, $\mathbf{c}^b$ is the body reflectance, $\mathbf{c}^i$ the surface reflectance, m^b and m^i are geometry dependent scalars representing the magnitude of body and surface reflectance. Bold notation is used to indicate vectors. For one MR we expect

both $\mathbf{c}^b$ and $\mathbf{c}^i$ to be almost constant, whereas $m^b(\mathbf{x})$ and $m^i(\mathbf{x})$ are expected to vary significantly. Hence, as for this definition, a MR, is formed by a single body reflectance $\mathbf{c}^b$ and a surface reflectance $\mathbf{c}^i$.

The two parts of the dichromatic reflection model are clearly visible in the histogram of Figure 3.1b. Firstly, due to the shading variations the distribution of the red pepper traces an elongated shape in histogram-space. Secondly, the surface reflectance forms a branch which points in the direction of the reflected illuminant. In conclusion, the distribution of a single MR forms a ridge-like structure in histogram space.

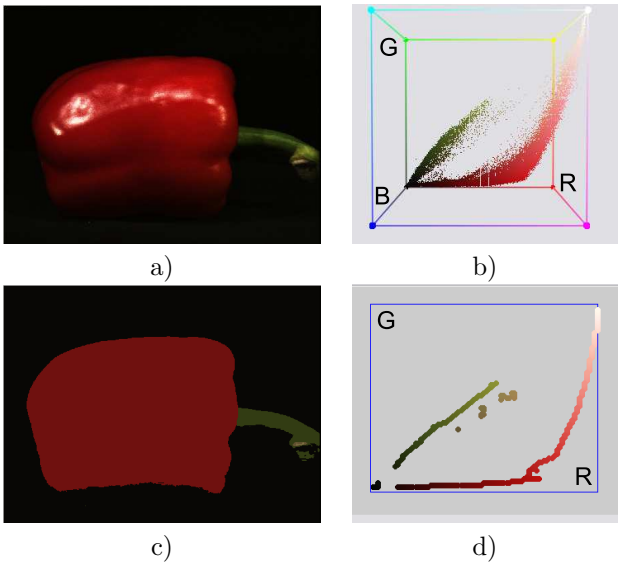


Figure 3.1: (a) An image from [72] and (b) its histogram. The effects of shading and highlights are clearly visible in the red colours of the histogram. (c) Segmented images using RAD. (d) Ridges found with RAD. Note that the three branches of the red pepper are correctly connected in a single ridge.

To illustrate the difficulty of extracting the distributions of MRs consider Figure 3.2c, which contains a patch of the horse image. The 2D Red-Green histogram of the patch is depicted in Figure 3.2d to see the number of occurrences of each chromatic combination. This is done for explanation purposes. In this 2D histogram it can be clearly seen that the density of the geometric term $m_b(\mathbf{x})$ varies significantly, and the distribution is broken in two parts. However, we have an important clue that the two distributions belong to the same MR: the orientation of the two distribution is similar, which means they have a similar $\mathbf{c}^b$. We exploit this feature in the ridge extraction algorithm by connecting neighboring distributions with similar orientation.

In literature several methods have explicitly used the dichromatic reflection model to obtain image segmentation, e.g. [106]. A drawback of such methods is however that for many images Eq. 3.3 does only approximately model the data. This can be caused by many reasons, such as non-linear acquisition systems, clipped highlights,

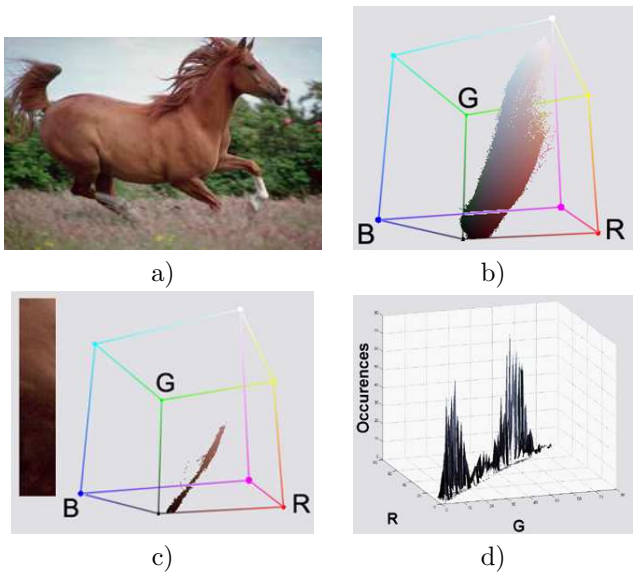


Figure 3.2: (a) An image and (b) its 3D RGB histogram. (c) A patch of a) and its RGB histogram. (d) 2D histogram of c) to illustrate the discontinuities appearing on a MR.

and image compression. We use Eq. 3.3 only to conclude that objects described by this equation will trace connected ridges in histogram space. This makes the method more robust to deviations from the dichromatic model.

3.4 A Ridge based Distribution Analysis method (RAD)

In this section we present a fast algorithm to extract MRs from histogram space. The proposed method is divided in two main steps. First, we propose a method to extract ridges as a representative of a MR. Afterwards a flooding process is performed to find the MRs from its ridges.

3.4.1 First step: Ridge Extraction

To extract a MR descriptor we need to find those points containing the most meaningful information of a MR, *i.e.*, its ridge. We propose to apply a multilocal creaseness algorithm to find the best ridge point candidates. This operator avoids to split up ridges due to irregularities on the distribution, mainly caused by the discrete nature of the data. Afterwards, we apply a ridge extraction algorithm to find the descriptor.

Multilocal Creaseness: finding candidates and enhancing connectivity

In order to deal with this commonly heavily jagged MR (see Fig. 3.2d) , we propose to apply the MLSEC-ST operator introduced by Lopez *et al.* in [123] to enhance ridge points. This method is used due to its good performance compared with other ridge detection methods [123] on irregular and noisy landscapes.

The Structure Tensor (ST) computes the dominant gradient orientation in a neighbourhood of size proportional to σ_d. Basically, this calculus enhances those situations where either a big attraction or repulsion exists in the gradient direction vectors. Thus, it assigns the higher values when a ridge or valley occurs. Given a distribution $\Omega(\mathbf{x})$, (the histogram in the current context), and a symmetric neighbourhood of size σ_i centered at point $\mathbf{x}$, namely, $N(\mathbf{x}, \sigma_i)$ the ST field S is defined as:

$$S(\mathbf{x}, \sigma) = N(\mathbf{x}, \sigma_{\mathbf{i}}) * (\nabla \mathbf{\Omega}(\mathbf{x}, \sigma_{\mathbf{d}}) \cdot \nabla \mathbf{\Omega^t}(\mathbf{x}, \sigma_{\mathbf{d}})) \tag{3.2}$$

where $\sigma = \{\sigma_{\mathbf{i}}, \sigma_{\mathbf{d}}\}$, and the calculus of the gradient vector field $\nabla\Omega(\mathbf{x}, \sigma_d)$ has been done with a Gaussian Kernel with standard deviation σ_d.

If $w(\mathbf{x}, \sigma)$ is the eigenvector corresponding to the largest eigenvalue of $S(\mathbf{x}, \sigma)$, then, the dominant gradient orientation $\overline{w}(\mathbf{x}, \sigma)$ in a neighbourhood of size proportional to σ_i centered at $\mathbf{x}$ is:

$$\overline{w}(\mathbf{x}, \sigma) = \mathbf{sign}(\mathbf{w^t}(\mathbf{x}, \sigma) \cdot \nabla^{\mathbf{t}} \mathbf{\Omega}(\mathbf{x}, \sigma_{\mathbf{d}}))\mathbf{w}(\mathbf{x}, \sigma) \tag{3.3}$$

The creaseness measure of $\Omega(\mathbf{x})$ for a given point $\mathbf{x}$, named $k(\mathbf{x}, \sigma)$, is computed with the divergence between the dominant gradient orientation and the normal vectors, namely n_k, on the r-connected neighbourhood of size proportional to σ_i. That is:

$$k(\mathbf{x}, \sigma) = -\mathbf{Div}(\overline{w}(\mathbf{x}, \sigma)) = -\frac{\mathbf{d}}{\mathbf{r}} \sum_{\mathbf{k=1}}^{\mathbf{r}} \overline{w}_{\mathbf{k}}^{\mathbf{t}}(\mathbf{x}, \sigma) \cdot \mathbf{n_k} \tag{3.4}$$

where d is the dimension of $\Omega(\mathbf{x})$. The creaseness representation of $\Omega(\mathbf{x})$ will be referred hereafter as Ω^σ.

As an example, Figure 3.3a shows the opponent colour 2D histogram of 3.3g. Its creaseness values are showed in 3.3b. There are three enhanced areas which corresponds with the three MRs of the original image. They appear as three mountains in 3.3b, clearly separated by two valleys. Note that higher creaseness values have a larger probability to become a ridge point.

Ridge Detection

In the previous section we have detected a set of candidate ridge points. In this section we discard superfluous points. As a result only those points necessary to maintain the connectivity of a MR remain. These points form the ridges of Ω^σ.

We classify ridge points in three categories. First, Transitional Ridge Points (TRP): when there is a local maximum in a single direction. Second, Saddle Points (SP): when there is a local maximum in one direction and a local minimum in another one. Third, Local Maximum Points (LMP). Formally, let $\Omega(x, y)$ be a continuous 2D surface and $\nabla\Omega(x, y)$ be the gradient vector of the function $\Omega(x, y)$. We define ω_1 and

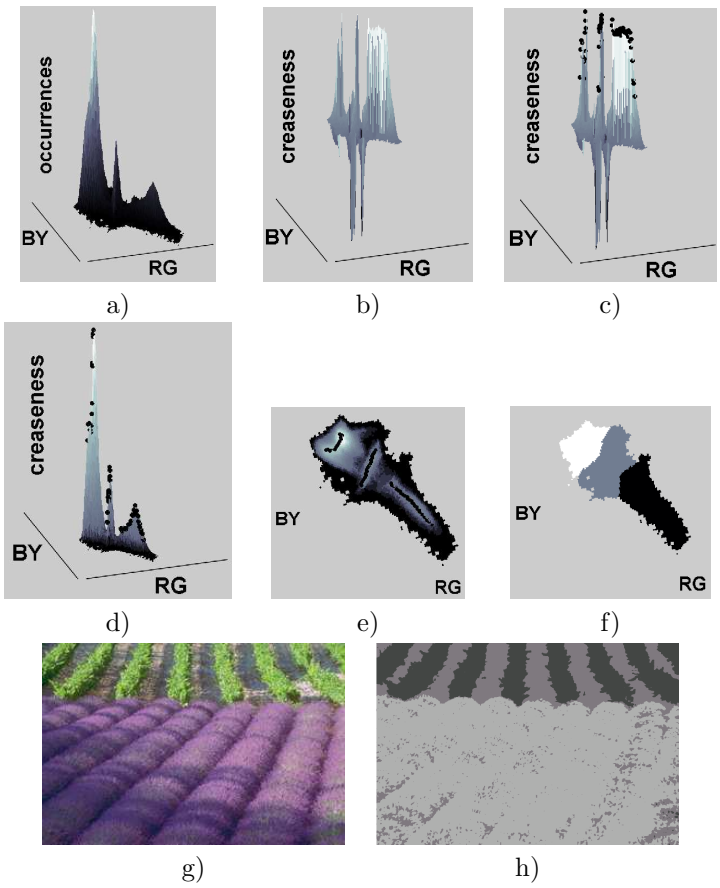


Figure 3.3: A graphical example of the whole process. (a) Opponent Red-Green and Blue-Yellow histogram $\Omega(\mathbf{x})$ of g). (b) Creaseness representation of a). (c) Ridges found in b). (d)Ridges fitted on original distribution. (e) Top-view of d). (f)MRs of a). (g) Original image. (h)Segmented image.

ω_2 as the unit eigenvectors of the Hessian matrix and λ_1 and λ_2 its corresponding eigenvalues with $\mid \lambda_1 \mid \leq \mid \lambda_2 \mid$. Then, for the 2D case:

$$\begin{aligned} LMP(\Omega(x,y)) =, \\ \{(x,y) | (\|\nabla\Omega(x,y)\| = 0), \lambda_1 < 0, \lambda_2 < 0\} \end{aligned} \tag{3.5}$$

$$\begin{aligned} TRP(f(x,y)) =, \\ \{(x,y) | \|\nabla\Omega(x,y)\| \neq 0, \lambda_1 < 0, \nabla\Omega(x,y) \cdot \omega_1 = 0, \\ \|\nabla\Omega(x,y)\| \neq 0, \lambda_2 < 0, \nabla\Omega(x,y) \cdot \omega_2 = 0, \\ \|\nabla\Omega(x,y)\| = 0, \lambda_1 < 0, \lambda_2 = 0\} \end{aligned} \tag{3.6}$$

$$SP(f(x,y)) = \{(x,y) | \|\nabla\Omega(x,y)\| = 0, \lambda_1 \cdot \lambda_2 < 0\} \tag{3.7}$$

This definition can be extended for an arbitrary dimension using the combinatorial of the eigenvalues. Hereafter we will refer to these three categories as *ridge points* (RP). Thus, $RP(\Omega(x,y)) = LMP \bigcup TRP \bigcup SP$. A further classification of ridges and its singularities can be found in [208] and [12].

A common way to detect RP is to find zero-crossing in the gradient of a landscape for a given gradient direction. Thus, we need to compute all gradient directions and detect changes following the schema proposed in [12]. In our case, we propose a way to extract a ridge without the need to calculate the gradient values for all points in the landscape. We begin on a local maxima of the landscape and follow the ridge by adding the higher neighbours of the current point, if there is a zero-crossing on it, until it reaches a flat region. This method can be easily applied to an arbitrary dimension. Formally, let $neigh(\mathbf{x}, \Omega^\sigma)$ be the set of neighbours of a point $\mathbf{x} \in \Omega^\sigma$, and $Cneigh(\mathbf{x}, \mathbf{y}, \Omega^\sigma)$ be the set of common neighbours between point $\mathbf{x} \in \Omega^\sigma$ and $\mathbf{y} \in \Omega^\sigma$. We also define a function $\mu(\mathbf{x}, \Omega^\sigma)$ as follows:

$$\mu(\mathbf{x}, \Omega^\sigma) = \sharp\{\mathbf{y} \in neigh(\mathbf{x}) \mid \Omega^\sigma(\mathbf{y}) \geq \Omega^\sigma(\mathbf{x})\} \tag{3.8}$$

Therefore, $\mu(\mathbf{x}, \Omega^\sigma) = 0$ means that $\mathbf{x}$ is a local maximum. Finally, we define μ' as:

$$\mu'(\mathbf{x}, \mathbf{y}, \Omega^\sigma) = \sharp\{\mathbf{z} \in Cneigh(\mathbf{x}, \mathbf{y}) \mid \Omega^\sigma(\mathbf{z}) \geq \Omega^\sigma(\mathbf{y})\} \tag{3.9}$$

$\mu(\mathbf{x}, \mathbf{y}, \Omega^\sigma) = 0$ means that y is a local maxima in the common neighbours between x and y. To extract ridges we propose an iterative process beginning on local maxima, that is

$$RP_0(\Omega^\sigma) = \mathbf{x} \in \Omega^\sigma \| \mu(\mathbf{x}, \Omega^\sigma) = 0 \tag{3.10}$$

Then, we just have to follow ridges starting on $RP_0(\Omega^\sigma)$ until its ending.

$$RP_{\mathbf{z}}(\Omega^\sigma) = RP_{\mathbf{z}-1}(C) \bigcup \{\mathbf{n} \in neigh(\mathbf{l}) \mid \mathbf{l} \in RP_{\mathbf{z}-1}(\Omega^\sigma), \mu'(\mathbf{l}, \mathbf{n}) = 0\} \tag{3.11}$$

Fig. 3.3c depicts the RP found on Ω^σ with black dots. Figs. 3.3d,e show a 3D view and a 2D projection view respectively of how these RPs fit in the original distribution as a representative of the three MRs. Finally, from the set of RPs of a distribution we can perform the calculus of each MR. A second example is shown in Figure 3.1. The complicated colour distribution of the pepper, caused by shading and highlight effects, is correctly connected in a single ridge.

3.4.2 Second step: MR Calculus from its RPs

In this final step we find the MR belonging to each ridge found. From topological point of view, it implies finding the portion of landscape represented by each ridge.

These portions of landscape are named *catchments basins*. Vincent and Soille [206] define a catchment basin associated with a local minimum M as the set of pixels p of Ω^σ such that a water drop falling at p flows down along the relief, following a certain descending path called the downstream of p, and eventually reaches M. In our case, M are the set of RPs found and then, MRs are found using the algorithm proposed in [206] applied on the inverse Ω^σ distribution. The proposed algorithm, is not based on the gradient vectors of a landscape [69] but on the idea of *immersion* which is more stable and reduces over-segmentation. Basically, the flooding process begins on the local minima and, iteratively, the landscape sinks on the water. Those points where the water coming from different local minima join, compose the watershed lines. To avoid potential problems with irregularities [123], we force the flooding process to begin at the same time in all MR descriptors, on the smoothed $\Omega(\mathbf{x})$ distribution with a Gaussian kernel of standard deviation σ_d (already computed on the ST calculus). Then, we define RAD as the operator returning the set of MRs of Ω^σ using RPs as marks:

$$RAD(\Omega(x)) = W(\Omega^\sigma, RP(\Omega^\sigma)) \tag{3.12}$$

Following this procedure, Fig. 3.3f depicts the 2D projection of the MRs found on 3.3a.

3.5 Colour image segmentation using RAD

Once RAD has been applied we need to assign a representative colour to each MR found. Thus, let $MR_n = \{\mathbf{x_1}, ..., \mathbf{x_r}\}$ be the nth MR of $\Omega(\mathbf{x})$, and $\Omega(\mathbf{x_i})$ the function returning the number of occurrences of $\mathbf{x_i}$ in Ω. Then, the dominant colour of MR_n, namely, $DC(MR_n)$ will be the mass center of $\Omega(MR_n)$:

$$DC(MR_n) = \frac{\sum_{i=1}^{r} \mathbf{x_i} \cdot \mathbf{\Omega}(\mathbf{x_i})}{\sum_{i=1}^{r} \Omega(\mathbf{x_i})} \tag{3.13}$$

The segmented image will have as many colours as the number MRs found. Figure, 3.3h shows the segmentation obtained with RAD from 3.3g. This segmentation has been performed on the opponent colour histogram. Although RAD can be applied to any chromatic representation of an image such as CIE, RGB, Ohta spaces or 2-dimensional ones such as Opponent or normalized RGB.

3.6 Results and performance evaluation

In the experiments we qualitatively and quantitatively evaluate the proposed segmentation method. Firstly, RAD is compared with Mean Shift (MS) [64], [39]. MS has been chosen because it is widely used, has a public available version, the EDISON one [37] and it has demonstrated its good performance [156]. Additionally, Mean Shift is a feature space analysis technique, as well as RAD, and yields a segmentation in a rather reasonable time, in opposition to other set of methods such as the Graph-Based

approaches [71]. Secondly, our method is compared on the Berkeley data set against a set of state-of-the-art segmentation methods.

The MS method [39], consists of finding the modes of the underlying probability function of a distribution. The method finds the Mean Shift vectors in the histogram of an image that point to the direction of higher density. All values of the histogram attracted by one mode compound the basis of attraction of it. In a second step, the modes which are near of a given threshold are joined. Finally, all modes joined an its basis of attraction will compose a dominant colour of the image. Mean Shift has two basic parameters to adapt the segmentation to an specific problem, namely, h_s, which controls a smoothing process, and h_r related with the size of the kernel used to determine the modes and its basis of attraction. To test the method, we have selected the set parameters $(h_s, h_r) = \{(7,3),(7,15),(7,19),(7,23),(13,7)(13,19),(17,23)\}$ given in [156] and [221]. The average times for this set of parameters, expressed in seconds, are 3.17, 4.15, 3.99, 4.07, 9.72, 9.69, 13.96 respectively. Nevertheless, these parameters do not cover the complete spectrum of possibilities of the MS. Here we want to compare RAD and MS from a soft oversegmentation to a soft undersegmentation. Hence, in order to reach an undersegmentation with MS, we add the following parameter settings $(h_s, h_r) = \{(20,25),(25,30),(30,35)\}$. For these settings, the average times are 18.05, 24.95 and 33.09 respectively.

The parameters used for RAD based segmentation are (σ_d,σ_i)={ (0.8,0.05) , (0.8,0.5) , (0.8,1) , (0.8,1.5) , (1.5,0.05) , (1.5,0.5) , (1.5,1.5) , (2.5,0.05) , (2.5,0.5) , (2.5,1.5) }. These parameters vary from a soft oversegmentation to an undersegmentation, and have been selected experimentally. The average times for RAD are 6.04, 5.99, 6.11, 6.36, 6.11, 5.75, 6.44, 5.86, 5.74 and 6.35. These average times, point out the fact that RAD is not dependent of the parameters used. In conclusion, whereas the execution time of Mean Shift increases significantly with increasing spatial scale, the execution time of RAD remains constant from an oversegmentation to an undersegmentation.

The experiments has been performed on the publicly available Berkeley image segmentation dataset and benchmark [135]. We use the Global Constancy Error (GCE) as an error measure. This measure was also proposed in [135] and takes care of the refinement between different segmentations. For a given pixel p_i, consider the segments (sets of connected pixels), S_1 from the benchmark and S_2 from the segmented image that contain this pixel. If one segment is a proper subset of the other, then p_i lies in an area of refinement and the error measure should be zero. If there is no subset relationship, then S_1 and S_2 overlap in an inconsistent manner and the error is higher than zero, (up to one in the worst possible case). MS segmentation has been done on the CIE Luv space since this is the space used in [156] and [221]. RAD based segmentation has been done on the RGB colour space for two reasons. First, the Berkeley image dataset does not have calibrated images and, consequently, we can not assure a good transformation from sRGB to CIE Luv. Second, because the size of L, u and v, is not the same and the method will require six parameters, instead of two, that is, $\overrightarrow{\sigma_L}$, $\overrightarrow{\sigma_u}$ and $\overrightarrow{\sigma_v}$. Nonetheless, for the sake of clarity, we also present some results of RAD on CIE Luv to directly compare results with MS. Figure 3.4 depicts a set of examples for RAD on RGB. From left to right: original image, RAD for (σ_d,σ_i)={ (0.8,0.05) , (1.5,0.05) , (2.5,0.05) , (2.5,1.5) } and human segmentation.

Table 3.1: Global Constancy Error for several state-of the-art methods: seed [142], fow [61], MS, and nCuts [173]. Values taken from [142] and [221].

	human	RAD	seed	fow	MS	nCuts
GCE	0.080	0.1996	0.209	0.214	0.2598	0.336

Figure 3.5 shows some results for the mean shift segmentation, corresponding to $(h_s, h_r) = \{(7, 15), (13, 19), (17, 23), (20, 25), (25, 30), (30, 35)\}$.

These results point out the main advantage of RAD in favor of MS, namely, the capability of RAD to capture the DS of a histogram, whereas MS is ignorant to the physical processes underlying the structure of the DSs as Abd-Almageed and S. Davis explain in [1]. Graphically, the set of images depicted in the first row of Figure 3.5, shows this behavior in a practical case. In the last column, MS joins rocks with the mountain, and the mountain with the sky, but is not able to find one unique structure for a rock or for the mountain, whereas RAD, as shown in Figure 3.4, is able to do.

A danger of RAD is that for some parameter settings it is prone to undersegmenting. Consequently it finds only one dominant colour for the whole image. This happens in some cases for $(\sigma_d,\sigma_i)=\{(2.5,1),(2.5,1.5)\}$, as Figure 3.6 illustrates. In the first example, the aircraft has a bluish colour similar to the sky, as well as the fish and its environment in the second example.

Additional examples related to the presence of physical effects, such as shadows, shading and highlights are shown in Figure 3.7. The good performance of RAD in these conditions can be clearly observed for the skin of the people, the elephants and buffalos, as well as for the clothes of the people.

The histogram of the mean GCE values versus the percentage of images for each GCE value are shown in Figures 3.8a,b for RAD on RGB and MS respectively. As more bars are accumulated on the left, the better is the method. Figures 3.8c,d show the standard deviation along the maximum and the minimum GCE values (red lines) for each of the 10 sets of parameters for RAD on RGB and MS. Note that the behaviour of both methods in this sense is almost the same. A low and similar standard deviation along all parameters means that the method presents an stable behaviour. Figure, 3.8e depicts the mean GCE index for each image ordered by increasing index for MS (green), RAD on RGB (black) and RAD on Luv (red). This plot shows, not only the good performance of RAD, but that RAD has a similar behavior on RGB and CIE Luv spaces, even with the aforementioned potential problems on Luv. Figure 3.8f plots the GCE index differences for each image between RAD on RGB and MS. Values lower than zero indicate the number of images where RAD performs better than MS. The same but for RAD on Luv versus MS, and RAD on RGB versus RAD on Luv is depicted on Figure 3.8g,h.

Additionally, table 3.1 shows GCE values for several state-of-the-art methods. These values are taken from [142] and [221]. These experiments have been performed using the train set of 200 images. For both the RAD and MS we present the results obtained with the best parameter settings. For our method the best results were obtained with $(\sigma_d,\sigma_i)=\{(2.5,0.05)\}$.

As can be seen our method obtains the best results. Furthermore, it should be noted that the method is substantially faster than the *seed* and the *nCuts* [173]

method. In addition, the results obtained with the MS need an additional step. Namely, a final combination step, which requires a new threshold value, is used to fuse adjacent segments in the segmented image if their chromatic difference is lower than the threshold (without pre- an postprocessing MS obtains a score of 0.2972). For our RAD method we do not apply any pre- or postprocessing steps.

3.7 Conclusions

In this chapter we have described a new feature space segmentation method that extracts the Ridges formed by a dominant colour on an image histogram. This method is robust against discontinuities appearing in image histograms due to compression and acquisition conditions. Furthermore, those strong discontinuities, related with the physical illumination effects are correctly treated due to the topological treatment of the histogram. As a consequence, the presented method yields better results than Mean shift on a widely used image dataset and error measure. Additionally, even with neither preprocessing nor postprocessing steps, RAD has a better performance than the state-of-the-art methods. It points out that the chromatic information is an important cue on human segmentation. Additionally, the elapsed time for RAD is not affected by its parameters. Due to that it becomes a faster method than Mean Shift and the other state-of-the-art methods.

The main shortcoming of RAD is its tendency to oversegmentation depending parameters used. In the next chapter we detail a saliency measure which is used to detect oversegmentation in a non-supervised manner.

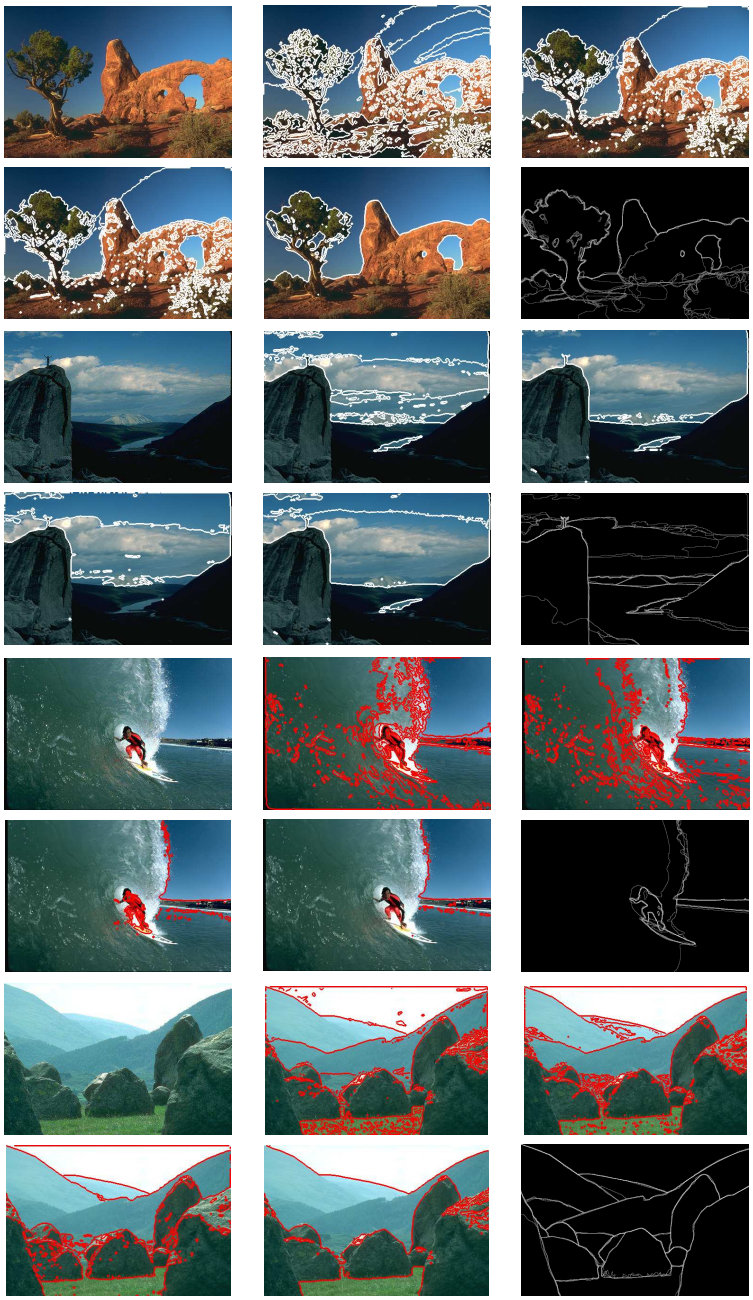


Figure 3.4: Examples of segmentation. For each image: original image, 4 segmentations with RAD on RGB with parameters (σ_d,σ_i)={(0.8,0.05),(1.5,0.05),(2.5,0.05),(2.5,1.5)} and last image corresponding with human segmentation.

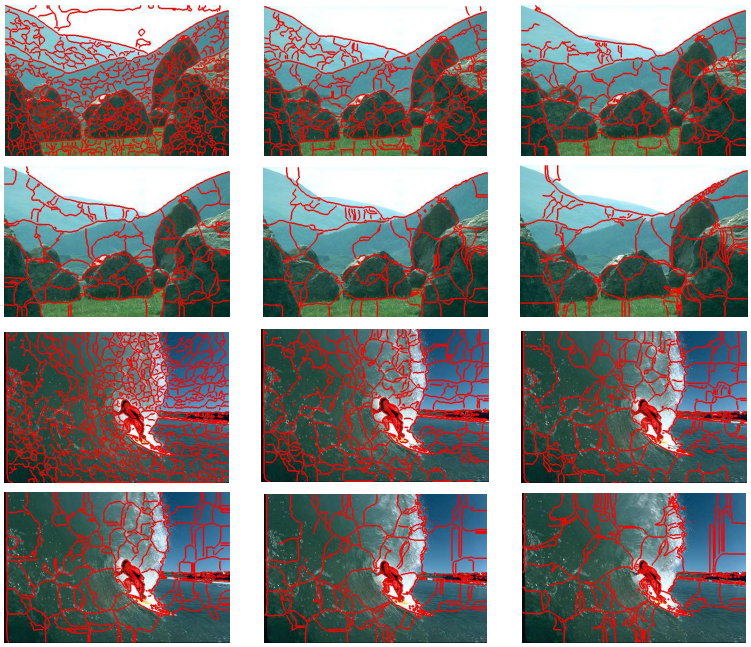


Figure 3.5: MS segmentation examples for different parameters. For each image: original image 5 segmentations showed for parameters: $(h_s, h_r) = \{(7,15),(13,19),(17,23),(20,25),(25,30)\}$.

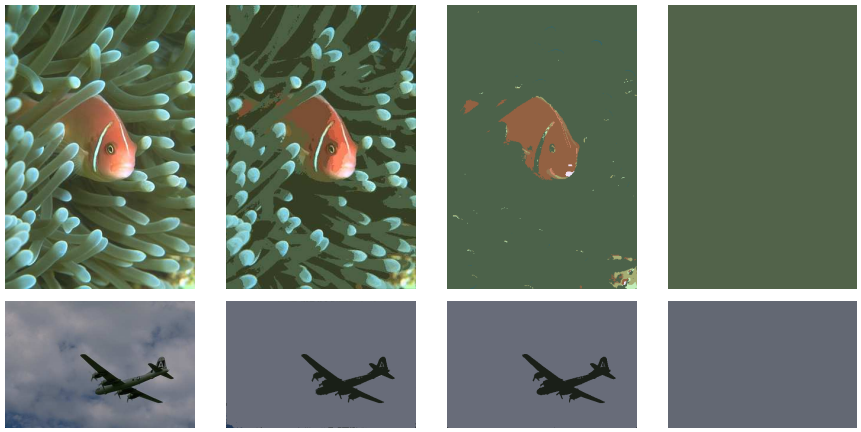


Figure 3.6: Examples of oversegmentation. For each image: original image. and segmentation with RAD with (σ_d,σ_i)=$\{(0.8,0.05),(2.5,0.05),(2.5,1.5)\}$.

Figure 3.7: Examples of segmentation in presence of shadows and highlights.

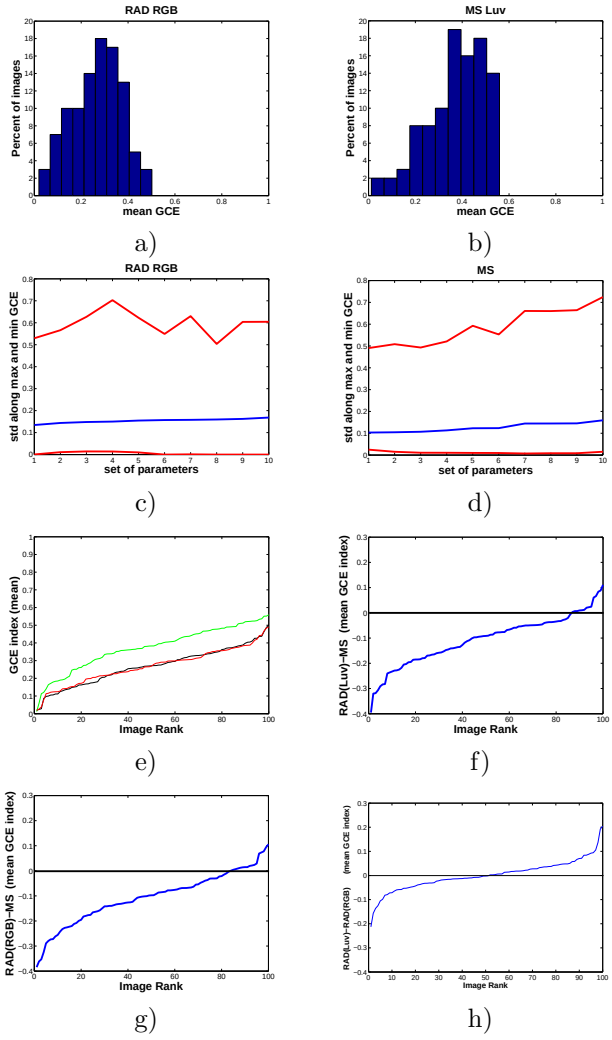


Figure 3.8: (a,b)Mean GCE values for each set of parameters. (c,d) Standard deviation of GCE along maximum and minimum values for each set of parameters. (e)Mean GCE values for each image sorted form lower to higher. (f)Values higher than zero: images where MS performs better RAD. (g,h)The same as f) but for MS and RAD Luv and for RAD RGB versus RAD Luv.

Chapter 4

Saliency of Color Image Derivatives

A shortcoming with RAD, as explained in Chapter 3, is undersegmenation. Saliency-driven methods have been proposed to detect oversegmentation and undersegmentation in an unsupervised manner. In this chapter we propose a saliency method which is used to turn RAD in an unsupervised segmentation technique by adding image spatial coherence.
Chromaticity and contrast play an important role in bottom-up saliency. Therefore, a computational model for saliency of color derivatives is proposed. The model is derived by applying Shannon's information theory to color derivative distributions. The computational model is compared to a human saliency measure which is computed from an image dataset consisting of manually labelled salient objects.
The experimental results show that the proposed method provides accurate performance to compute visual saliency with a Hit rate up to 95.2% on a large scale image dataset. Further, the psychophysical experimental results demonstrate that the proposed method performs significantly better at predicting human saliency than state-of-the-art models.

4.1 Introduction

Human visual attention is for an important part driven bottom-up by the saliency of image details. An image detail appears salient when one or more of its low-level features (e.g. size, shape, luminance, color, texture, binocular disparity, or motion) exceeds the overall feature variation of the background. Saliency determines the capability of an image detail to attract visual attention (and thus guide eye movements) in a bottom-up way [189] [109]. Current models of human visual search and detection suggest that this preattentive stage indicates potentially interesting image details, whereupon the focus of attention is sequentially shifted to each of these regions and the serial stage is deployed to analyze them in detail [192].

Several information theoretical approaches have been proposed to compute visual saliency from local image features [100] [63][132]. These methods are based on the assumption that feature saliency is inversely related to feature occurrence (i.e. rare

features are more informative and therefore more salient than features that occur more frequently). It is indeed plausible that interesting image details correspond to locations of maximal self information, a measure closely related to local feature contrast [22] [66]. Using this notion, recent models of human visual fixation behavior assume that saliency driven free viewing corresponds to maximizing information sampling [65][224]. These models have successfully been deployed to model human fixation behavior, pop-out, dynamic saliency, saliency asymmetries, and to solve classic computer vision problems like dynamic background subtraction [65] [66][67].

Because of its importance for many practical applications, we focus on bottom-up saliency in this chapter. The parallel, preattentive, or bottom-up stage of human vision is thought to guide a serial (computationally intensive) attentive or top-down stage. Among all features that contribute to a detail's saliency, orientation and color are generally considered to be the most significant ones [98][211] [216]. Consequently, most current saliency models are based on local color and orientation contrast (e.g. [78][94][214]).

There exist evidence that the human visual system combines and processes low-level features in an early stage [109] [110]. Most popular models of visual attention compute individual saliency maps for different features like color, orientation or motion, and merge these in a later stage into a single overall saliency map, e.g. [94] (late fusion of features). Here we present an information theoretical method to compute the saliency of color edges. by combining chromaticity and contrast (early fusion of features).

In this chapter, a method is proposed which computes image saliency from the information content (the frequency of occurrence, probability, or self information) of both local chromatic and orientation derivatives. The method is based on the observation that in natural images, color transitions of equal probability (i.e. isosalient transitions) form ellipsoids in decorrelated color spaces [198]. The transformation that turns these ellipsoidal isosaliency surfaces into spherical ones (called the color saliency function), effectively replaces gradient strength with information content. After the color saliency transformation, vectors of equal length have equal information content and thus equal impact on the saliency function. We introduce three different ways to calculate the saliency transformation, using either a single image, a collection of images, or the restriction that the eigenvectors of the transformation matrix coincide with the opponent color space. Further, we investigate whether there is a correspondence between our approach and human visual perception.

The chapter is organized as follows. In the next section, we propose three computational saliency measures, and one human saliency measure. In Section 4.3, the psychophysical experiment is outlined. Finally, in Section 4.4, the results are presented and conclusions are drawn.

4.2 Saliency of Color Edges

In this section, we present two different methods to compute color edge saliency. The first one, introduced in section 4.2.1, is a computational method based on the self information of color edges. We present three versions of this method: a local version

(that estimates color edge saliency from only a single image), a global version (that uses a collection of images to compute color edge saliency), and a version in which the eigenvectors of the transformation matrix are restricted to the opponent color space. Then, in section 4.2.2, we propose a measure of color edge saliency based on (human-labeled) salient object detection data.

4.2.1 Multi-contrast computational saliency

The color saliency method by Van de Weijer *et al.* [198] is inspired by the notion that a feature's saliency reflects its information content. Consider an image $\mathbf{f} = (R, G, B)^{\mathbf{t}}$. The information content, I, of an image derivative $\mathbf{f_x}$, according to information theory, is given by the logarithm of its probability p:

$$I = -log(p(\mathbf{f_x})). \tag{4.1}$$

Hence, color image derivatives which are equally frequent have equal information content. To map image derivatives to a saliency map, a function g is required for which the following holds:

$$p(\mathbf{f}_x) = p(\mathbf{f}'_x) \leftrightarrow |g(\mathbf{f}_x)| = |g(\mathbf{f}'_x)|. \tag{4.2}$$

The saliency function g transfers color image derivatives to a space where their norm is proportional to their information content.

In Fig. 4.1, the distribution of color derivatives for the COREL dataset is given. The derivatives form an ellipsoid-like distribution, of which the longest axis is along the luminance direction. This indicates that equal displacements are more informative along the color directions (perpendicular to the luminance) than in the luminance direction. The saliency transformation in [198] is restricted to a transformation based on known color spaces. Now we propose a more general transformation to compute g in that it is not fixed to a pre-defined color space.

Let the distribution of the ellipsoid to be described by the covariance matrix $\mathbf{M}$:

$$\mathbf{M} = \overline{\mathbf{f_x} \left(\mathbf{f_x}\right)^{\mathbf{t}}} = \begin{pmatrix} \overline{R_{\mathbf{x}} R_{\mathbf{x}}} & \overline{R_{\mathbf{x}} G_{\mathbf{x}}} & \overline{R_{\mathbf{x}} B_{\mathbf{x}}} \\ \overline{R_{\mathbf{x}} G_{\mathbf{x}}} & \overline{G_{\mathbf{x}} G_{\mathbf{x}}} & \overline{G_{\mathbf{x}} B_{\mathbf{x}}} \\ \overline{R_{\mathbf{x}} B_{\mathbf{x}}} & \overline{G_{\mathbf{x}} B_{\mathbf{x}}} & \overline{B_{\mathbf{x}} B_{\mathbf{x}}} \end{pmatrix} \tag{4.3}$$

where the matrix elements are computed by

$$\overline{R_{\mathbf{x}} R_{\mathbf{x}}} = \sum_{i \in S} \sum_{\mathbf{x} \in X^i} R_{\mathbf{x}}\left(\mathbf{x}\right) R_{\mathbf{x}}\left(\mathbf{x}\right) \tag{4.4}$$

where S is a set of images, and X^i is the set of pixels coordinates $\mathbf{x}$ in image i. Matrix $\mathbf{M}$ describes the derivatives energy in any direction $\hat{v}$. This energy is computed by $E(\hat{v}) = \hat{v}\mathbf{M}\hat{v}^{\mathbf{t}}$. Matrix $\mathbf{M}$ can be decomposed into eigenvector matrix $\mathbf{U}$ and eigenvalue matrix $\mathbf{\Lambda}$ according to $\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{\Lambda}\mathbf{U}^{\mathbf{t}}$. This provides us with the saliency function g:

$$\mathbf{g}\left(\mathbf{f_x}\right) = \mathbf{\Lambda}^{-1}\mathbf{U}^{\mathbf{t}}\mathbf{f_x}. \tag{4.5}$$

Substitution of Eq. 4.5 into Eq. 4.3 yields

$$\mathbf{g}\left(\mathbf{f}_{\mathbf{x}}\right)\left(\mathbf{g}\left(\mathbf{f}_{\mathbf{x}}\right)\right)^{\mathbf{t}} = \mathbf{\Lambda}^{-1}\mathbf{U}^{\mathbf{t}}\mathbf{U}\mathbf{\Lambda}\mathbf{\Lambda}\mathbf{U}^{\mathbf{t}}\mathbf{U}\mathbf{\Lambda}^{-1} = I, \tag{4.6}$$

meaning that the covariance matrix of the transformed image is equal to the identity matrix. This implies that the derivative energy in the transformed space is equal in all directions.

We consider three methods derived from information theory to compute the saliency of color edges:

- *Local color saliency:* saliency is defined by the rarity of the color derivatives in a single image. Thus, when applied to a set of images, each image is transformed by its own individual saliency matrix $\mathbf{M}_l^c$ (where c stands for computational and l for local). For its computation, S in Eq. 4.4 contains only a single image.
- *Global color saliency:* saliency is defined as the rarity of the color derivatives over a set of images. Hence, a single matrix $\mathbf{M}_g^c$ is computed based on the color derivatives of all images in a data set (S contains all images). The same saliency matrix is then applied to all images in the data set.
- *Global opponent color-space saliency [198]*: saliency is defined as the rarity of the color derivatives in a set of images, with the additional restriction that the eigenvectors of the saliency matrix coincide with the vectors which span the opponent color space

$$\mathbf{U} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}. \tag{4.7}$$

  The color saliency transformation $\mathbf{M}_o^c = \mathbf{\Lambda}^{-1}\mathbf{U}^{\mathbf{t}}$ only differs in the scaling of the axes as given by the eigenvalue matrix $\mathbf{\Lambda} = diag\left(\alpha, \beta, \gamma\right)$. Applied to a set of images, the same eigenvalue matrix is applied to all images.

An example of local and global computational saliency is given in Fig. 4.2. Based on global saliency, the edges of the red American flag are considered salient. However, for local saliency, which is computed based on only the statistics of this image, the red edges are not considered salient. Instead the brown edges of the pastry are considered more salient. This is in correspondence with human assessment of this image.

Multi-scale color saliency: The three types of information theoretical saliency maps can be computed at multiple spatial scales. Maps computed at multiple scales can be combined into a single saliency map as follows:

$$s\left(\mathbf{x}\right) = \sum_{\sigma \in \Sigma} \sum_{\mathbf{x}' \in N(\mathbf{x})} \left\| \mathbf{M}^{\sigma}\left(\mathbf{f}^{\sigma}\left(\mathbf{x}\right) - \mathbf{f}^{\sigma}\left(\mathbf{x}'\right)\right)\right\| \tag{4.8}$$

where $\mathbf{f}^{\sigma}$ denotes the Gaussian smoothed image at scale σ, and $\Sigma = \{1, 2, 4, 6, 8, 10, 12, 14\}$. $N\left(\mathbf{x}\right)$ is a 9x9 neighborhood window. $\mathbf{M}^{\sigma}$ is the transformation matrix computed from Gaussian derivatives of scale σ and can be any of the three before mentioned ones: $\mathbf{M}_l^c$, $\mathbf{M}_g^c$ or $\mathbf{M}_o^c$. Note that leaving out $\mathbf{M}$ from Eq. 4.8 results in the multi-scale contrast approach proposed by Liu et al. [121]. An example of a multi-scale color saliency map is given in Fig. 4.2. The edges of the salient pastry are considered more salient by the multi-scale color saliency map.

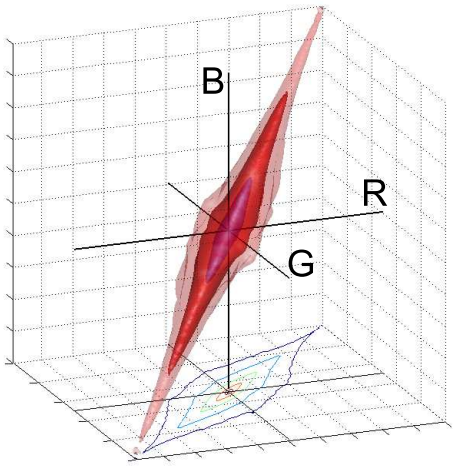


Figure 4.1: Histogram of the distribution of opponent derivatives computed for the Corel image dataset.

4.2.2 Human saliency measure

In the previous section, we proposed different versions to color saliency based on information theory, which we called computational saliency measures. Since our goal is to obtain saliency maps that closely relate to human perception, a more direct approach would be to learn the optimal transformation $\mathbf{M}^h$ (where h stands for human) of color derivatives from a labeled set of training images. The images should be labeled with salient objects in the scene. From this data, the saliency transformation $\mathbf{M}^h$ can be derived that maximally agrees with the human labeled data.

To compute $\mathbf{M}^h$, a large-scale image data set of human labeled salient objects is used [121]. Example images of this dataset are shown in Fig. 4.3. The data set contains a large number of high quality images obtained from different sources such as image forums and image search engines. Images all contain a single salient object or a distinctive foreground object. For each image, users drew a rectangle enclosing the most salient object in the image. We use the a set of images called set B in [121]. This set consists of 5000 images which were labeled by nine users. Foreground pixels are those pixels which are considered to be foreground by a majority of the users. Then, this set is divided in 10 subsets of 500 images each $(B_1, ..., B_{10})$. We use the 500 images in B_1 for training and the rest of the 4500 images for testing.

We evaluate the performance of the saliency measure with the precision index as follows. An image is divided in a foreground region f^i and a background b^i, where i is the image index. Let $f^i_{\mathbf{M}}$ be the summed saliency of the foreground for a certain saliency transformation $\mathbf{M}$. Let $b^i_{\mathbf{M}}$ denote the same for the background. Further, let $A(f^i)$ and $A(b^i)$ denote the area of the foreground and background respectively. The

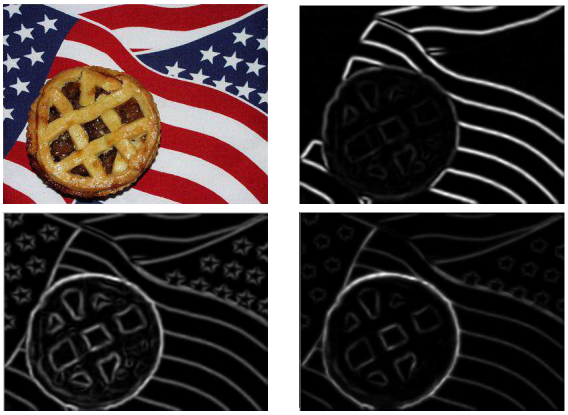


Figure 4.2: Top left: Original image. Top right: computational global saliency. Bottom left computational local saliency. Bottom right: local boosting with multi-scale contrast. The local statistics used in the local transformation suppress the colorful edges of the American flag, therefore the pastry is better detected, which is the part of the scene selected as the most salient one by the 9 users.

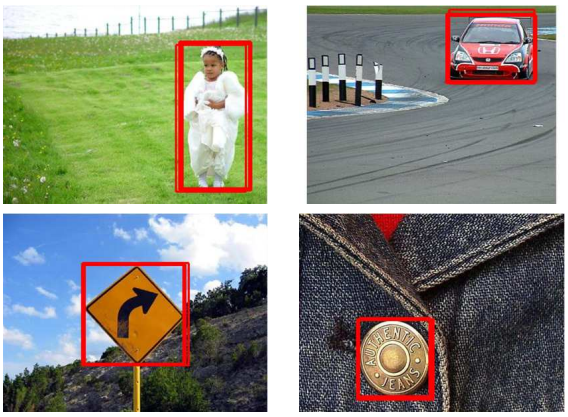


Figure 4.3: Labeled images from image set B consisting of 5000 images which were labeled by nine users obtained from [121].

confidence measure used is the Precision index P^i_Λ:

$$P^i_{\mathbf{M}} = \frac{A\left(b^i\right) f^i_{\mathbf{M}}}{A\left(f^i\right) b^i_{\mathbf{M}}}. \tag{4.9}$$

In other words, $P^i_{\mathbf{M}}$ provides the likelihood to select from the salient map the most salient object.

To reduce the set of possible transformations, the transformation is used which corresponds to the opponent color space. We define $\mathbf{M}^h_o$ as that transformation which maximizes $P^i_{\mathbf{M}}$ by varying the parameters $\Lambda = diag(\alpha, \beta, \gamma)$. Therefore, an exhaustive search is performed in the $\alpha\beta\gamma$ space and $P^i_{\mathbf{M}}$ is computed for all training set images. The best transformation $(\alpha_l, \beta_l, \gamma_l)$ is the one corresponding with the highest average precision score. Hence, it is that transformation which obtains the maximum correspondence (given the opponent transformation) to human assessments of object saliency.

4.2.3 Comparing computational and human color saliency

In this section, we compare the saliency maps obtained with the computational and human saliency measures. To this end, $(\alpha_m, \beta_m, \gamma_m)$ are computed according to the computational opponent color-space measure. Table 4.1 summarizes the results of the computational saliency and the human saliency measures in terms of the precision index.

When comparing the human saliency measure with the computational saliency measure, it can be inferred that the results obtained by the computational approach are very close to the best possible transformation, that is, the human saliency measure. In both cases, γ is a fairly small value. This is because there is a high amount of achromatic transitions in the images as opposed to chromatic ones. Hence, these transitions are less informative, as predicted by the computational saliency measure, and to obtain a good saliency map the weights of these transitions should be decreased. Thus, α and β values are larger and close to each other in both cases.

To quantitatively show the resemblance of the saliency maps computed by the computational and human measure, we have calculated the intersection of the normalized saliency maps. The averaged score over all images reaches 97.43%, whereas the overlapping between human saliency and saliency based on RGB edges (without additional transformation) is only 83.11%. A qualitative comparison between computational and human saliency is depicted in Fig. 4.4.

This indicates the relevance of using information theory as a valid saliency processing model from a computational point of view. To provide more insight in the relation between the computational and human saliency measure, we propose to validate this observation with human perception by means of a psychophysical experiment that is conducted in the next section. The goal of this experiment is to study whether information theory is a valid underlying mechanism of color saliency in human vision.

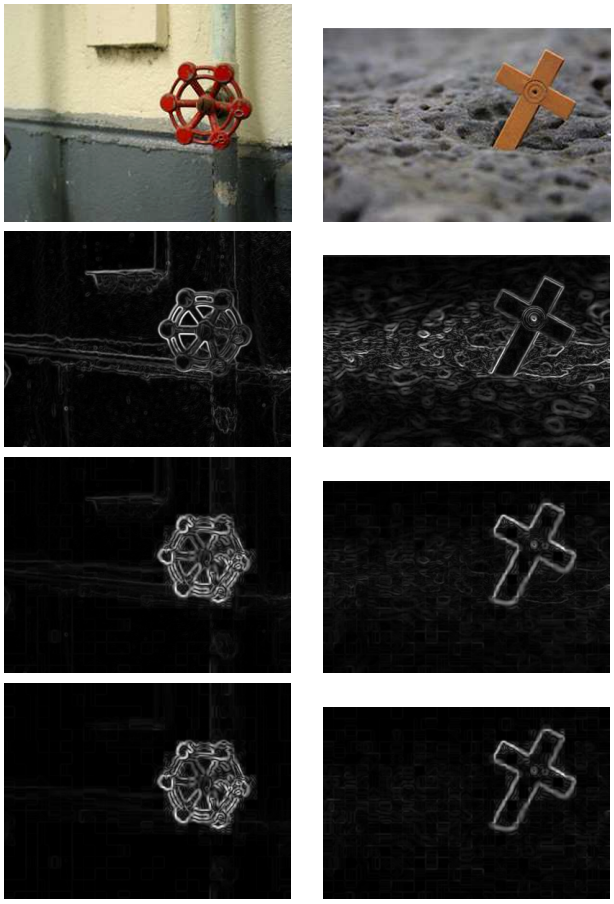


Figure 4.4: Color saliency example. First row: original image. Second row: RGB edges. Third row: computational global saliency $\mathbf{M}_o^c$ (see table 4.1). Fourth row: $\mathbf{M}_o^h$ (see table 4.1). The overlap between human and computational maps over all images reaches 97.43% whereas the overlapping with the RGB edges is 83.11%.

Measure	α	β	γ	$P^i_{\mathbf{M}}$
$\mathbf{M}^h_o$	0.65	0.34	0.01	0.49
$\mathbf{M}^c_o$	0.53	0.43	0.04	0.45
$\mathbf{M}^c_l$	image dep.	image dep	image dep	0.51

Table 4.1: Results obtained for human global saliency measure $\mathbf{M}^h_o$ $(\alpha_l, \beta_l, \gamma_l)$, computational global saliency $\mathbf{M}^c_o$ saliency $(\alpha_m, \beta_m, \gamma_m)$ and computational Local Saliency $\mathbf{M}^c_l$ with different transformation values depending the image. The fourth column shows the average precision score.

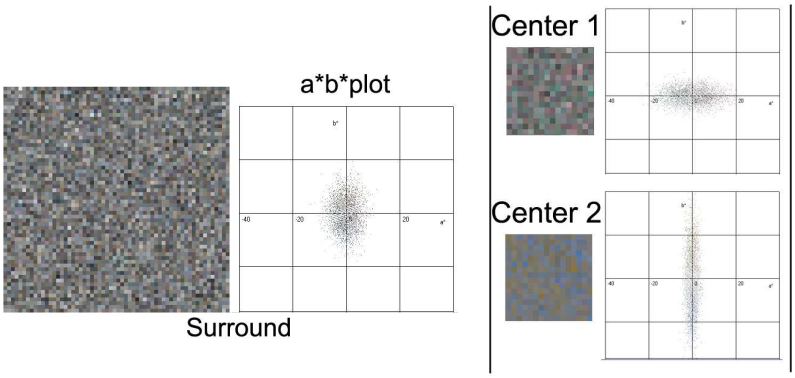
4.3 Psychophysical Evaluation of Color Edge Saliency

In this section, a psychophysical experiment is proposed to determine if the classic information theory can explain human perception in saliency. In the previous section, we have shown how computational saliency corresponds with the human saliency measure. Nonetheless, in the comparison carried in section 4.2 there is also an uncontrolled cognitive top-down mechanism in the human measure. Therefore, the goal of this section is to investigate whether in a controlled scenario, a human subject and computational saliency measures will provide the same response by avoiding any possible effect of cognitive top-down mechanisms. Hence, a set of images are designed where any possible known pattern is avoided to block the apparition of top-down mechanisms. A way to solve this is to generate images without any familiar shape for a human subject.

The next question is how to measure chromatic information in a synthetic image. To this end, the spatial distribution formed by chromatic transitions in an image is taken. As shown in Fig. 4.1, these transitions form an ellipsoid in opponent color space. Then, the principle is to generate synthetic images having the same distribution. An example of such a synthetic image forming a controlled distribution is depicted in Fig. 4.5.

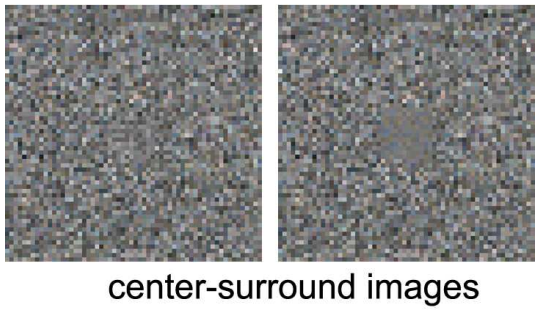
Saliency is the degree to which an item or location stands out from its surround. Therefore, we propose a center-surround experiment. Every image is composed by a background (surround) following a controlled distribution as the one in Fig. 4.5, and a central foreground (center) following another distribution. The color patterns are defined in the CIELAB color space [218]. Other color spaces might have been selected but the aim is to specify colors in terms of a perceptual space and enable comparison of the results with other studies [125].

Given a certain distribution of L*a*b* values in the surround, the saliency of the center will depend on the difference between the L*a*b* distribution of the center and that of the surround. The more the two distributions differ, the higher the saliency of the center is expected to be. The aim is to determine how strong this saliency depends on the underlying L*a*b* distributions (and associated edge transition distributions). We therefore transform these distributions in a systematic manner. So, most of the energy is contained in the L* direction, followed by b* and a*. Following the observations made in section 4.2, we should expect that the center should be the most salient. In other words, our model of computational saliency should provide corresponding answers as a human subject in the psychophysical experiment.



a)

b)



c)

Figure 4.5: a) Example of a synthetic image with specified distribution in CIELab color space, which forms the surround. b) Two different transformations of the color distribution shown in a) form two centers. c) Layout of the psychophysical experiment, showing two center-surround color patterns side-by-side. The surrounds are identical, the centers are different. Subjects had to indicate which of the two centers stood out most from the surround, i.e. was considered most salient.

4.3.1 Method

Subjects

Five men and three women (ages ranging from 22 to 29) participated in our experiment. They had normal or corrected-to-normal acuity and normal color vision as confirmed by testing on the HRR pseudoisochromatic plates (4^{th} edition). Subjects were unaware of the purpose of the experiment.

Apparatus

Stimuli were presented on a self-calibrating LCD monitor (Eizo, ColorEdge CG211) operating at 1600x1200 pixels (0.27 mm dot pitch) and 24-bit color resolution. Using a spectrophotometer (GretagMacbeth, Eye-one) the monitor was calibrated to a D65 white point of 80 cd/m^2, with gamma 2.2 for each of the three color primaries. CIE 1931 x,y chromaticities coordinates of the primaries were (x,y) = (0.638, 0.322) for red, (0.299,0.611) for green and (0.145,0.058) for blue, respectively, closely approximating the sRGB standard monitor profile [184]. Spatial uniformity of the display, measured relative to the center of the monitor, was $\Delta E^*_{ab} < 1.5$ according to the manufacturer's calibration certificates.

Stimuli and design

Fig. 4.5c shows the layout of the experiment. From Fig. 4.1, it is observed that for the Corel dataset, we have 5 times more transitions (edges) in Intensity than transitions in RG and BY. In L*a*b* space this corresponds to $\sigma_{L_{Corel}} = 54$, $\sigma_{b_{Corel}} = 27$ and $\sigma_{a_{Corel}} = 16$. We generate a controlled synthetic image which forms a distribution that corresponds with these statistics. Then, we transform this distribution (e.g. multiplying each axis by a certain value) in order to create the two central patches. As shown in Fig. 4.5, two square center-surround color patterns were shown side-by-side. Subjects have to indicate which of the two centers is most salient.

In addition to the surround corresponding with the COREL statistics, we use three more surrounds where, the axis containing the most information was a* in one of them and b* in the other instead of L*. The last surround generated was a surround with an spherical distribution to verify what happens in equal conditions of energy in all directions. Table 4.2 summarizes the values used to generate the distributions forming these three surrounds.

Each surround listed in Table 4.2 was combined with 13 different center distributions. These center distributions were obtained by applying the transformation

$$\begin{pmatrix} L' \\ a' \\ b' \end{pmatrix} = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} L \\ a \\ b \end{pmatrix} \tag{4.10}$$

One of the transformations, labeled C_1, is predicted by our computational saliency measure $\mathbf{M}_l^c$ (computational local transformation, which is here fixed to L*a*b* space)

Surround	σ_L	σ_a	σ_b
S_L	$\sigma_{L_{Corel}}$	$\sigma_{a_{Corel}}$	$\sigma_{b_{Corel}}$
S_a	$\sigma_{a_{Corel}}$	$\sigma_{L_{Corel}}$	$\sigma_{b_{Corel}}$
S_b	$\sigma_{b_{Corel}}$	$\sigma_{a_{Corel}}$	$\sigma_{L_{Corel}}$
S_{eq}	$\sigma_{L_{eq}}$	$\sigma_{a_{eq}}$	$\sigma_{b_{eq}}$

Table 4.2: Surrounds with systematic changes in the standard deviations (σ) along the L*, a* and b* axes of perceptual color space. The statistics of the first surround (S_L) comply with the energy distributions of natural images contained in the Corel image dataset. The last surround (S_{eq}) has equal amounts of energy in the three directions.

as the most salient between all possible transformations, having values α_0, β_0 and γ_0. Five more center patches (C_2 - C_6) are generated with α_0, β_0 and γ_0 interchanged. Centers C_L, C_a and C_b were created by maximizing the energy of the axis indicated by the subscript, while the energy in the remaining two axes are equal. Centers C_{La}, C_{Lb} and C_{ab} were created by maximizing the energy of two axes (indicated by the subscripts). Finally, center C_{eq} was obtained by having the same amounts of energy in all three axes.

Summarizing, for each surround (background) we generate 13 different centers (foregrounds). One of these centers is predicted from the computational saliency measure as the most salient. The question is whether the human observers also find this center to be the most salient. If so, this means that information theory is a valid underlying mechanism for saliency.

Procedure

After passing the color vision test, the subjects were seated at 50 cm viewing distance from the LCD monitor. In each trial, they had to indicate (by pressing keys on the keyboard) which of the two centers (left or right) was most salient, i.e. standing out most from the surround. They were encouraged to make a decision although they could also indicate that the two centers were equally salient.

4.4 Validation and Results

Here we present the computational results obtained for the test set for all methods (4.4.1). Then, the results are provided of the psychophysical experiment (section 4.4.2) and a comparison of the computational saliency models with the psychophysical results is given (section 4.4.3).

4.4.1 Color saliency on real-world images

Here we analyze the computational and human salience measures on the real-world large-scale image data set [121]. To evaluate saliency methods, we use the Hit and Miss index (a common comparison measure used in literature). Note that for each

image the size of the foreground (rectangle) is given. If the maximum of the saliency map falls inside the original rectangle, we have a hit, otherwise, a miss is registered.

We evaluate the human saliency ($\mathbf{M}_o^h$), computational global saliency ($\mathbf{M}_o^c$) and computational local saliency ($\mathbf{M}_l^c$). In addition to these transformations, we also show results obtained with the multi-scale computational local transformation ($\mathbf{M}_l^{\sigma c}$), the RGB edges without any transformation (RGBe), the Itti saliency method (Itti) and a random selection of the most salient location (Random). Table 4.3 summarizes the results obtained. From this table, it can be concluded that the results obtained with multi-scale contrast are better than others. A 5.8% increase in visual saliency accuracy is obtained. Further, using locally induced saliency provides better performance than computing the color (matrix) transformation based on color edges extracted from the whole image dataset (global). As expected, locally computing the transformation adapts better to the edge distribution for each image. Furthermore, the results show that locally induced saliency with multi-scale contrast provide the best performance.

Transformation	Hit	Miss
Global Human ($\mathbf{M}_o^h$)	87.1	12.9
Global Computational ($\mathbf{M}_o^c$)	87.9	12.1
Local Computational ($\mathbf{M}_l^c$)	89.6	10.4
Local multi-scale computational ($\mathbf{M}_l^{\sigma c}$)	95.2	4.8
RGB edges	81.4	18.6
Itti	88.2	11.8
Random	72.8	27.2

Table 4.3: Hit and Miss values obtained in the test set for all proposed saliency transformations as well as for RGB edges, Itti saliency measure [94] and a Random selection of the most salient location.

4.4.2 Psychophysics

In each trial, a subject indicated which center was most salient. Each center was in competition with the 12 others just once. In Figure 4.6, the relative saliency is shown obtained for all surrounds. Error bars indicate the standard error of the mean, on descending order, obtained by averaging over the 8 observers. The data did not indicate one or more of the observers to be an outlier.

Regarding S_L, Fig. 4.6 shows that center C_{ab} has the highest relative saliency. This is the expected result, because S_L has the largest variance in the L* dimension and C_{ab} has a color edge distribution boosted along both the *a and b* dimension, at the cost of reducing energy in the intensity edge (L*) distribution. So, center C_{ab} looks more strongly colored but with less luminance contrast, which is highly salient in the S_L surround. In contrast, the least salient center (C_L) has increased the energy in the intensity edges, at the cost of reducing energy in the color edges. However, since the surround S_L already has a distribution that dominates in the intensity edges, the

extra boosting in intensity edges does not result in visual saliency, as predicted for our saliency measure.

Surround S_a was created by rotating the axes of edge distributions such that the largest variance coincided with the a* axis of CIELAB space. This results in an increased edge distribution along the red-green axis of color space, *i.e.*, the colors along the red-green axis become more saturated, at the cost of a decreased edge intensity. Fig. 4.6 shows that for this background the most salient center is C_b and the least salient center is C_6. Note that there is no significant difference between the saliency of C_4 and C_6. Center C_b is most salient because it is boosted along the b* axis (the blue-yellow axis in color space) which is orthogonal to the boosted a* axis of the surround, at the cost of reduced energy in the b* and L* axes. Blue-yellow edges with decreased intensity edges are salient in a dominant red-green edge distribution. Center C_6 and C_4 are least salient in surround S_a because their γ-coefficient in the saliency transformation equals α_0, which is the largest ($\alpha_0 > \beta_0 > \gamma_0$). So, intensity edges are boosted most but do not show up as salient in the dominating red-green surround.

The results for surround S_b are described in a similar as that of S_a, but with the role of the red-green and yellow-blue axes interchanged. So, in short, C_a is most salient because it has boosted red-green edges (at the cost of blue-yellow and intensity), which stands out from the dominating blue-yellow surround.

The surround S_{eq} is characterized by equal amounts of energy in the edge distributions along the L*, a* and b* axes of CIELAB color space. Center C_a apparently is most salient, followed by C_{ab} and C_b, which are all chromatic transformations. The least salient centers are all intensity boostings. This is an important result: when the edge distributions in the three axes of color space are equal, the most salient change to that distribution is a chromatic one, i.e. an increase of edges along the a* or b* axis, or both, at the cost of a decrease of energy in intensity edges.

With respect to the natural surround S_L there remains one important question. Why was center C_1 not the most salient one? We recall that C_1 was expected to be most salient from a computational point of view. Figure 4.6 shows that C_1 and C_2 are not significantly different, and have a higher relative saliency than C_3 and C_4, and C_5 and C_6. So, C_1 has indeed the highest saliency with respect to the group of centers C_1 to C_6, but it is still outperformed by the chromatic transformations C_{ab}, C_a and C_b. The reason for this is that the latter transformations have maximized energy in one or two axes which exceeded the transformation of C_1.

4.4.3 Comparison of computational models with psychophysics

In this section, we compare the performance of the different saliency models on predicting the human response (the selection of the most salient center) in our psychophysical experiment. We apply a transformation to the matrices obtained in section 4.2 to convert them to L*a*b* space. For each subject ($s = 1..8$) and each computational model ($m = 1..5$) we computed the overall correspondence between the subject's selection and the model's selection of the most salient center. This

correspondence $Cor(s, m)$ is a value between 0 and 100 and is computed as follows:

$$Cor(s, m) = 100 \frac{\sum_{i=1}^{468} a_i}{468}, \tag{4.11}$$

where a_i denotes - per trial i - the agreement (either 0 or 1) between model and subject. Fig. 4.7 shows the correspondence for the 5 computational models. Trials in which subjects could not decide on the most salient center are left out of the computation.

It is clear from Fig. 4.7 that global computational saliency ($\mathbf{M}_o^c$) and global human saliency ($\mathbf{M}_o^h$) outperform the other models. Additional statistical testing (Statgraphics Centurion XV) indicate no significant difference between ($\mathbf{M}_o^c$) and ($\mathbf{M}_o^h$). At the 95% confidence level significant differences exist between ($\mathbf{M}_o^h$) and local computational saliency ($\mathbf{M}_l^c$) (p=1.1E-4), between ($\mathbf{M}_l^c$) and Itti (p=1.2E-3) and Itti and RGB (p=1.8E-15). We also computed the inter-observer agreement using Eq. 4.11 but with a_i replaced by w_i, where w_i represents the fraction (between 0 and 1) of subjects that gave the same response in each trial i. So, if 6 of the 8 subjects selected the same center, $w_i = 6/8$. This resulted in an observer agreement of 86.8%. In conclusion, our computational saliency methods (both local and global) are significantly better at predicting human saliency than Itti and Koch model, as showed in Fig. 4.7.

4.5 Conclusions

In this chapter, a model for saliency is proposed generated by multicontrast based on an early fusion of chromaticity and contrast. The information of these features is obtained by means of a local process based on Shannon's information theory.

Computational results obtained from a large-scale dataset confirms that an early fusion of these features results in an improvement on the prediction of saliency. Further, it can be derived that the proposed method provides very accurate performance to compute visual saliency with a Hit rate up to 95.2%.

From the psychophysical experiment, it can be derived that, for a uniformly distributed background, humans are more sensitive to chromatic changes than luminance variations. Further, it is shown that the proposed method performs significantly better at predicting human saliency than state-of-the-art models.

The method presented in this chapter is used to yield a non-supervised image segmentation. We compute the saliency of the segments resulting for a given parameter settings of RAD. In this way we can detect oversegmentation. Furthermore, we can combine the most salient segments of varying parameter settings. We detail this procedure in the next chapter.

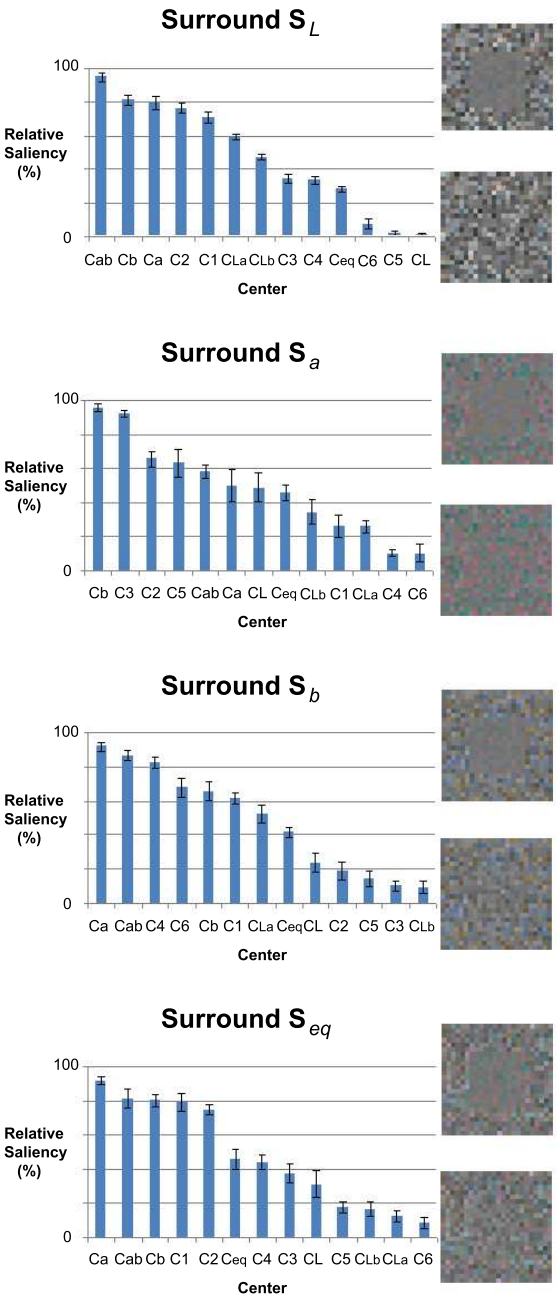


Figure 4.6: Relative saliency of the 13 centers for the surrounds S_L, S_a, S_b, S_{eq} averaged over observers. Error bars represent the standard error of the mean. The images on the right hand side show the most salient (top) and least salient (bottom) centers, in a small portion of the surround.

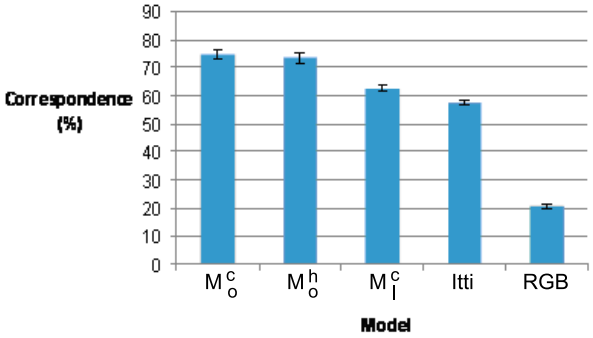


Figure 4.7: Correspondence as computed with eq. 4.11 between computational saliency and human saliency. The different computational models are sorted on descending correspondence. Error bars indicate standard error of the mean (8 subjects).

Chapter 5

Hybrid RAD Using Saliency and Prior Knowledge

The segmentation method (RAD) detailed in Chapter 3, models the shape of a single material reflectance in histogram-space. The method is based on a multilocal creaseness analysis of the histogram, which results in a set of ridges representing the material reflectances. The segmentation method derived from these ridges is robust to both shadow, shading and specularities, and texture in real-world images.
In this chapter we further complete the method by incorporating saliency-based prior-knowledge and spatial coherence by using the multi-scale color contrast saliency information method detailed in Chapter 4. Results obtained show that our method clearly outperforms state-of-the-art segmentation methods on a widely used segmentation benchmark, having as a main characteristic its excellent performance in the presence of shadows and highlights.

5.1 Introduction

The segmentation method detailed in Chapter 3 overcomes the shortcomings derived from the dichromatic reflection model [171] by introducing a ridge-based analysis of the histogram which better describes the shape of a single material reflectance. Our approach is more flexible than the dichromatic reflection model, thus solving cases as the one depicted in Fig. 5.1. The cast shadow on the floor provokes an abrupt change in the material reflectance representative of the floor. RAD is able to find a single ridge which includes this change. Such shape is not described by the dichromatic reflection model.

The main advantage of RAD is its performance in the presence of shadows and highlights. Fig 5.2 depicts two examples of it. In the first row we can see how RAD properly segments the face of the man. In the second row we show another example with a cast shadow. In this case, the shadows have a bluish color. Moreover, part of the shadow is in the grass. RAD is able to successfully segment this image.

RAD has a shortcoming though, namely, its risk of oversegmentation. Fig 5.3

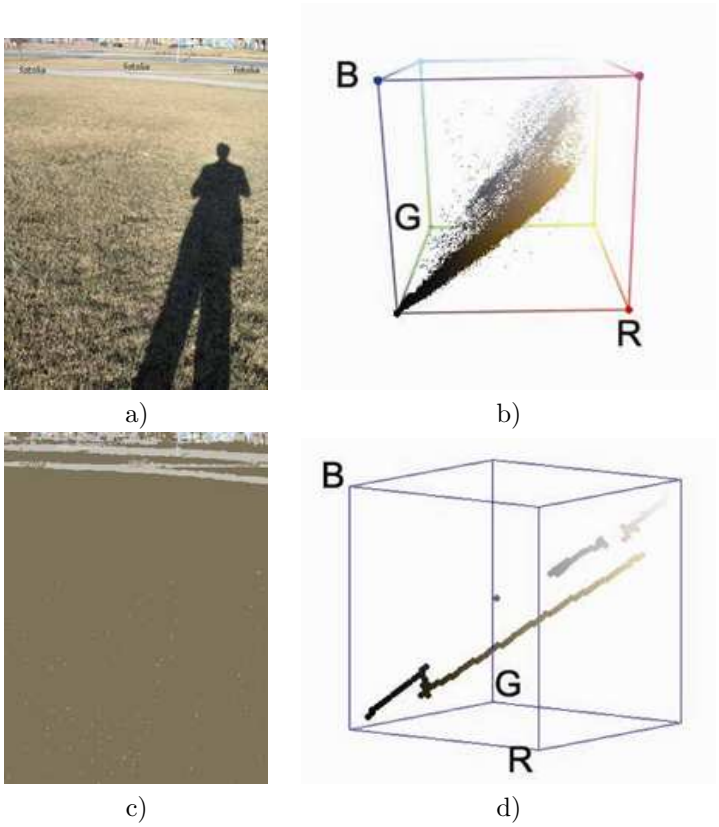


Figure 5.1: a) Original image. b) RGB histogram. c) Segmented image. d) The material reflectance representative of the floor can not be described for the dichromatic reflection model. RAD is able to find a single ridge (brownish) even with the change in direction due to the shadow.

shows two examples.

In this chapter we propose two extensions of RAD which are aimed to cope with this problem which has two main causes: the excessive flexibility of RAD and its lack of spatial coherence.

5.1.1 Shortcoming 1: Lack of physical preference

RAD has been thought to be a flexible method to avoid the shortcomings of the dichromatic reflection model. Nevertheless, there is an issue that is ignored for RAD: the directions of the ridges, even with its irregularities should approximately coincide with the directions of the dichromatic reflection model. We can see an example in Fig. 5.1. The ridge corresponding with the floor, mainly follows a direction from the black to the white areas of the RGB cube. Statistically this is to be expected. shading and shadows mainly causes changes in this same direction. The same surface is not sta-

Figure 5.2: Two examples of the good performance of RAD in the presence of shadows and highlights.

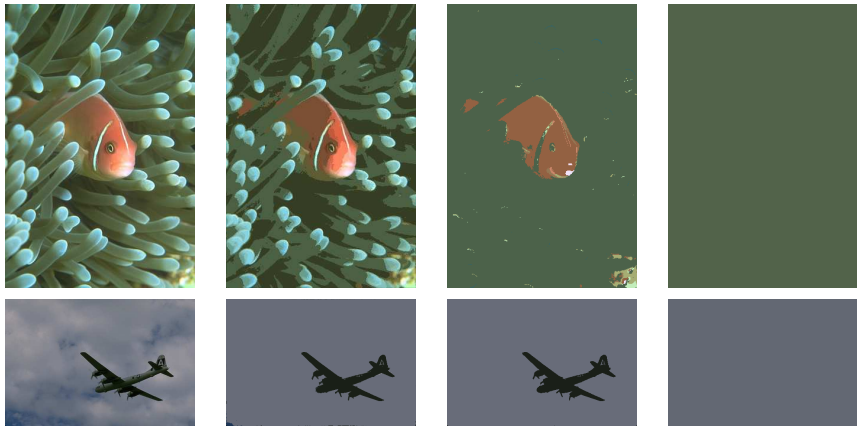


Figure 5.3: Examples of undersegmentation. For each image: original image. and segmentation with RAD with (σ_d,σ_i)={(0.8,0.05),(2.5,0.05),(2.5,1.5)}.

tistically expected to have changes in the chromatic direction, that is, perpendicular to the main diagonal of the RGB cube. So far, RAD do not includes this knowledge. Therefore, there is theoretically equally-probable having a ridge joining different colors than a ridge describing a surface including its shadows. In practice changes in chromaticity due to shadows and shading are more subtle than those in chromaticity

corresponding to different surfaces. Nonetheless, we have to give more likelihood to having ridges following directions similar to the main diagonal than ridges perpendicular to it. In Sec 5.2 we present a way to include such statistical knowledge. It turns RAD from a feature-based analysis method to a hybrid one by including also physical information. The method resulting is the physics-RAD (pRAD).

5.1.2 Shortcoming 2: lack of spatial coherence

RAD analyzes the RGB space representation of the image. the spatial coherence of an image, that is, the spatial relation of the pixels in the image space, is not included. Spatial coherence can be used to determine whether a segmentation should be considered oversegmenation or not. A common way to do it is using image saliency. In Chapter 4 we have described a method for image saliency focused on chromatic transitions. Since RAD segments are based on colour, the analysis of the chromatic transitions as representative of the image colors and its spatial positions should be a coherent method to validate RAD segmentation. In Sec 5.3 we describe how to use our saliency measure to perform such validation. Moreover, we also detailed in Sec 5.3 how to use saliency to perform a multi-scale segmentation. The method resulting is the spatial-RAD (sRAD).

5.2 Adding Physical Preference (pRAD)

The dichromatic model predicts pixels of a single colored object to form a line passing through the origin as long as no specular reflection is present. In case of specular reflection, it models pixels of both body and specular reflection to form a plane. However, applying these geometrical models to the pixel values often leads to unsatisfying results because of the many deviations causing the body reflectance pixels neither to lie on a line, nor the combined body and specular pixels to lie in a plane. In the previous section, we therefore proposed a method to extract ridges from histogram space, based on the observation that ridges capture the essential structure predicted by the dichromatic model while being more robust to slight deviations from the ideal case. These ridges are allowed to have any orientation. However, the dichromatic results suggest the orientation of body reflection and of specular reflection, to be more likely than others. In this section, we will incorporate this additional information into the RAD method, and propose the *physic-based RAD* called *pRAD*.

The general structure which a single colored object forms in histogram space, is a ridge in the radial direction caused by shadow and shading variations with in the higher intensity regions of the RGB cube some branches in the illuminant direction caused by specularities. Changes in the chromatic direction, perpendicular to these two directions are seldom. Due to blurring effects, caused by for example out of focus, relative motion between camera and object, and aberrations in the optical system, ridges in the chromatic direction are formed between different surface reflectances. These ridges which might result in undesired segmentation results. To suppress ridges in the less probable orientations and favor ridges in probable ones, we propose to exploit the image statistics of ridge orientations. This statistic is captured by computing the normalized tensor representation $\hat{S}$ of the color histograms

generated by a set of images in a train data set with,

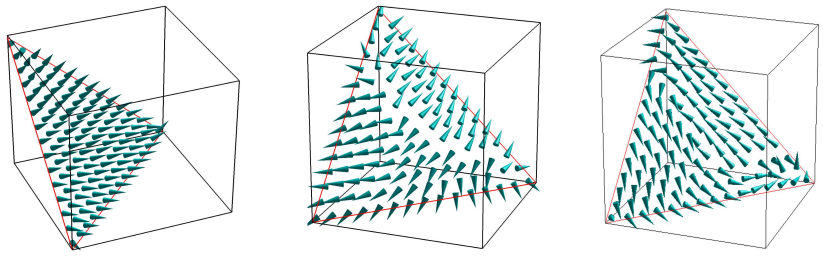
$$\begin{aligned} \overline{\hat{S}(\mathbf{x},\sigma)} &= \sum_{i \in T} \hat{S}_i(\mathbf{x},\sigma) \\ \hat{S}_i(\mathbf{x},\sigma) &= \frac{S(\mathbf{x},\sigma)}{\|S(\mathbf{x},\sigma)\|} \end{aligned} \tag{5.1}$$

where T is the set of indexes of the train data. We normalize the tensors with

$$\|S(\mathbf{x},\sigma)\| = N(x,\sigma_i) * \left(\nabla\Omega^{\mathbf{t}}(\mathbf{x},\sigma_d) \cdot \nabla\Omega(\mathbf{x},\sigma_d)\right) \tag{5.2}$$

since we are only interested in the orientation of the ridges, not their strength (note that the transpose operates on the first gradient here, whereas in Eq. 3.2 it operates on the second). The outcome $\overline{\hat{S}}$ is a tensor field, which for each RGB value in the histogram indicates the relative likelihood of the orientations of ridges passing through this point.

The tensor field $\overline{\hat{S}}$, which does not require human segmentation, is learned on the complete COREL dataset of over 40.000 images [31]. In Fig.5.4 we have depicted the three eigenvectors of the tensor field for a slice of the RGB-cube, namely the chromatic plane ($R+G+B=1$). The dominant orientation (Fig.5.4a) coincides with the intensity direction. The orientation of the second and third eigenvector is less clear. However, in general the second eigenvector points outwards from the center of the chromatic plane (Fig.5.4b). This is called the saturation direction, since changes in this direction cause colors to become more or less saturated. The least variations is found in the angular direction in the chromatic plane (Fig.5.4c), coinciding with hue changes. This is what we expect since most physical changes such as shadows, shading, and specularities (when white) do not cause any hue changes of the MR.



a) b) c)

Figure 5.4: From left to right: first, second and third dominant orientations of the tensor field computed using 40.000 images of the COREL dataset.

This prior knowledge can be incorporated in the ridge extraction framework proposed in the previous section by using instead of Eq. 3.2 the following equation,

$$S^{\lambda}(\mathbf{x},\sigma) = (1-\lambda)\, S(\mathbf{x},\sigma) + \lambda \|S(\mathbf{x},\sigma)\| \overline{\hat{S}(\mathbf{x},\sigma)} \tag{5.3}$$

where $\lambda = [0,1]$ regulates the influence of the prior knowledge represented by $\overline{\hat{S}}$. For example, $\lambda = 0.25$, indicates that 75% of the strength of tensors in S^{λ} is based on

the image to be segmented, and 25% of strength is defined by prior knowledge. The regulation parameter λ is learned from a training data set. In our experiments on the Berkeley training set. We found $\lambda = 0.33$ to yield the optimal results (optimization is based on the GCE score). Further results obtained with pRAD are detailed in section 5.4.

In conclusion, we proposed a method to favor ridges in orientation commonly seen in real-world images, and suppress ridges in the less probable orientations. It is important to note that the extra computational cost of pRAD is negligible with respect to RAD, since $\overline{\hat{S}}$ is precomputed.

5.3 Multi-scale segmentation adding image spatial coherence (sRAD)

With the addition of the physical information (pRAD) we add robustness to the method, which has a better behavior in those cases where the geometry of the objects and the light causes shadows, shading and highlights for a single colored object. In this section, we propose two further improvements. Firstly, the optimal parameter setting was found to vary for each image. To automatically obtain a good segmentation, we propose to combine the segmentations at various parameter settings. Secondly, RAD tends to oversegmentation in textures formed by multiple chromaticities. To overcome this problem, we propose to use mutliscale contrast (see section 4.2.1). We call this method *sRAD* (spatial RAD) or in combination with pRAD, it is called *spRAD*.

The idea to combine different sub-segmentations to build a combined segmentation has been investigated before [50] [142] [160]. The objective is to take the strengths of each segmentation while avoiding its weakness. Roughly, it implies to determine a measure of the goodness of a segment. JSEG algorithm [46], for instance, propose the *J-measure* that is based on the variance of the pixels belonging to a class-map (color quantization). This measure of goodness is computed at different scales forming the set of images that have to be combined. This is achieved by a region growing algorithm. The resulting image (based on goodness, not on chromaticity) tends to be oversegmented. Hence, a merge algorithm based on Euclidean distance of the histogram of each neighboring region is applied. Actually, this way to merge regions is commonly applied, e.g. also for Mean Shift segmentation, another algorithm which tends to oversegment. A graph-based approach to merge oversegmented images is presented in [160]. Other measures to describe the correctness of a segment are the homogram proposed in [35], the spatial-color compactness degree [130], a calculus based in the Bhattacharyya distance [50] or a probabilistic approach as explained in [142]. The method introduced in this chapter to combine sub-segmentations belongs to those methods that use contrast as a criteria of the goodness of a segmentation (e.g. [70][87]).

5.3.1 Combining sub-segmentations

With RAD and pRAD, we can segment at different feature-space scales by changing σ_d (feature-space smoothness). The optimal σ_d value varies depending on the image.

Therefore, whereas a single value can be found as the optimal when considering a whole dataset, results can be improved by selecting a different value depending on the image. Moreover, good segments can be found at different scales. Hence, we propose a method to consider segments at different feature-space scales for any single image. For the selection of segments we use a multi-scale contrast representation of the image which suppresses texture edges. Its computation is further explained in section 4.2.1.

The procedure follows two steps. First, we perform a set of segmentations at different feature-space scales (named sub-segmentations) of the same image. An example of these sub-segmentations is showed in Fig. 5.5, second column. The number refers to the value of σ_d used. Afterwards, we select the best segments of this sub-segmentations using the multiscale contrast image (Fig. 5.5) to build the final segmentation. The selection is based on a ranking which is computed by summing the contrast underlying the edges of the segments normalized for the perimeter of the segment. This operation, a combination of each sub-segmentation with the multiscale image is represented with $\otimes$. In Fig. 5.5, we show the ranking for every segment selected by spRAD with a gray-value codification: the lighter the color, the higher the rank position. Once the most contrasted segment (first in the ranking) has been added to the combined segmentation, the contrast already contained in this segment is removed from the multiscale image and the ranking is done again. The numbers appearing in the images of the Fig. 5.5 (third row and spRAD segmentation), illustrate at what scales were selected the segments that form the combined segmentation (spRAD).

sRAD is evaluated on the Berkeley dataset using the GCE error measure. Comparison based on GCE requires methods to have a similar number of segments [135]. In the Berkeley dataset, human segmentations have an average of eight segments. Additionally, the results of all methods presented in section 5.4 have a similar number of segments (between 7 and 10). Therefore, we will select the nine first ranked segments to build the final segmentation. It is interesting to note that the combined segmentation was found to outperform all of the sub-segmentations from which it was formed, showing the validity of the approach.

5.3.2 Multiscale Color Contrast

In the previous section, we introduced a multiscale color contrast image as a selection criteria to combine various segmentations. The relevance of the segments is computed by summing the contrast underlying the edges of the segment (normalized for the perimeter of the segment). To obtain a good segmentation, we need the contrast-image to suppress shadow and specular edges, as well as spurious texture edges.

Textures tend to be present at certain scales, but exhibit weak contrast at other scales. For this reason, we propose to use a multi-scale contrast image. This multiscale chromatic contrast is computed as a linear combination of the Gaussian pyramid image, commonly used in saliency (*e.g.* [94]), according to:

$$s\left(\mathbf{x}\right) = \sum_{\sigma \in \Sigma} \sum_{\mathbf{x}' \in N(\mathbf{x})} \left\| \mathbf{M}^{\sigma} \left(\mathbf{f}^{\sigma}\left(\mathbf{x}\right) - \mathbf{f}^{\sigma}\left(\mathbf{x}'\right) \right) \right\|^{2} \tag{5.4}$$

where $\mathbf{f}^\sigma$ denotes the Gaussian smoothed image at scale σ chosen from $\Sigma = \{1, 2, 4, 6, 8, 10, 12, 14\}$. $N(\mathbf{x})$ is a 9x9 neighborhood window. The approach is similar to [121].

To prevent the re-introduction of shadow and specular edges, we apply a color boosting matrix $\mathbf{M}$ in Eq. 5.4 [198]. This approach was originally proposed to amplify salient chromatic edges in the image, and thereby indirectly suppressing shadows and specularities. The boosting matrix is computed with

$$\mathbf{M}^\sigma = \left(\mathrm{diag}\left(\bar{\mathbf{o}}_{\mathbf{x}}^\sigma\right)\right)^{-1} \mathbf{U}, \tag{5.5}$$

where

$$\begin{aligned} \mathbf{o}_{\mathbf{x}}^\sigma(\mathbf{x}) &= \mathbf{U}\mathbf{f}_{\mathbf{x}}^\sigma(\mathbf{x}) \\ \bar{\mathbf{o}}_{\mathbf{x}}^\sigma &= \sqrt{\sum_{\mathbf{x}\in \mathrm{X}} \left(\mathbf{o}_{\mathbf{x}}^\sigma(\mathbf{x})\right)^2} \end{aligned} \tag{5.6}$$

where the summation is over all pixels in the data set X (in our experiments the COREL data set), and $\mathbf{U}$ the transformation from RGB to opponent color space is given by

$$\mathbf{U} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}. \tag{5.7}$$

The boosting matrix $\mathbf{M}$ normalizes the derivatives in each of the opponent color channels with the average derivative energy in that opponent channel in the data set. As was shown in [198] most derivative energy is along the intensity axis (the third opponent axis $O3$) and only little variations in the chromatic directions (the first and second opponent axes $O1$ and $O2$). Therefore, multiplication with the boosting matrix emphasizes salient chromatic edges and suppresses achromatic edges.

The multiscale approach helps to minimize oversegmentation in textured parts of the image. Fig.5.6 shows two examples of the improved behavior when adding spatial coherence to the method. We can see how the flower-texture is assigned to a single segment by sRAD, whereas RAD and pRAD assign multiple labels to this texture. The same occurs in the second row with the plants.

5.4 Results and performance evaluation

In this section, the performance of RAD is compared with pRAD, sRAD and spRAD. Finally, our method is compared on the Berkeley data set against a set of state-of-the-art segmentation methods.

5.4.1 Results obtained with pRAD, sRAD and spRAD

As explained before, the addition of physical-based prior knowledge requires to select a proper value for λ. We obtained that the best value for λ is 0.33. This prior knowledge is added following equations 5.2 and 5.3. The addition of the prior knowledge aims to favor ridges following the statistically expected directions of a surface reflectance (mainly achromatic changes), at the same time that we suppress those ridges formed by different surface reflectances. These effects can be observed in Fig. 5.6, first row.

Table 5.1: Global Constancy Error for our different proposals.

	human	spRAD	pRAD	sRAD	RAD
GCE index	0.080	0.1678	0.1780	0.1860	0.2048

Table 5.2: Global Constancy Error for several state-of the-art methods: seed [142], fow [61], MS, and nCuts [173]. Values taken from [142] and [221].

	human	spRAD	seed	fow	MS	nCuts
GCE index	0.080	0.1678	0.209	0.214	0.2598	0.336

Whereas RAD joins the purple flowers with the grass, pRAD correctly finds a ridge for purple flowers and another for the grass. More qualitative examples are showed in Fig. 5.7. First row: pRAD is able to find a segment for the gray little stones in the top-left part of the images. Second and fourth rows: in both cases, pRAD finds a single surface reflectance for all the rocks, whereas RAD clearly oversegment these rocks.

As a second adaptation to RAD we propose sRAD which uses the information contained in the image to yield a segmentation enhancing the multicontrast of the image. It can be performed using subsegmentations generated either by RAD (then the method is called sRAD), or by pRAD (then the method is called spRAD). This segmentation is less affected by textures, since they have a weak effect in a multiscale analysis. Fig. 5.6 shows two clear examples of it. We can see how, RAD segments incorrectly the red and yellow flowers and oversegment the grass on the second row. pRAD find better segments in both cases, but still with an oversegmentation. spRAD, instead, find a single segment for the red flowers, the purple ones, and the floor of the second row, that is, produces a non-oversegmented images. In the examples showed in Fig. 5.7 the same effect can be observed. For instance, in the second row, we can see how spRAD is able to find a single segment for the trees and segment for the rocks. In the third row it is showed an example of the improvement achieved by combining multiple segmentations. RAD and pRAD, can not find a segment for every mountain due to a clear blurring effect, whereas spRAD produces a better segmentation. We point out that results obtained with sRAD are worse than the ones obtained with spRAD, as can be observed in these examples. This is the expected result, since the subsegmentations used by spRAD are better than the ones used by sRAD.

Quantitative results using the GCE score are presented in section 5.4.2.

5.4.2 Comparison to State of the Art

In this section we show more quantitative results obtained with our segmentation method. Table 5.1 shows results obtained with RAD, sRAD, pRAD and spRAD. It can be seen how each improvement outperforms the other proposals, being the combination of sRAD and pRAD, namely, spRAD, the one obtaining the best performance. Table 5.2 shows GCE values for several state-of-the-art methods. These values are taken from [142] and [200]. For both RAD and MS we present the results obtained with the best parameter settings. For RAD, best results were obtained with

(σ_d,σ_i)={(2.5,0.05)}. The mean number of SR found using RAD has been 5, but it is not directly translated in 5 segments on segmented images. Often, some segments of few pixels appear due to chromaticity of surfaces. CGE evaluation favors oversegmentation [135]. Hence, to make feasible a comparison with other methods using GCE, we have performed the segmentation without considering segments of an area lower than 2% of the image area. In this case, the mean number of segments for the 200 test images is 6.98 (7 segments). The number of segments for the other methods varies from 5 to 12, including pRAD. Finally, for RAD and spRAD we show results obtained by generating a combined segmentation having 9 segments. Furthermore, we stand out that results obtained with spRAD, outperform all results obtained with its sub-segmentations. These sub-segmentations, have GCE values going from 0.1780 to 0.2205.

As can be seen our final approach, spRAD, obtains the best results. Furthermore, it should be noted that the method is substantially faster than the *seed* and the $nCuts$ [173] method. In addition, the results obtained with the MS need an additional step. Namely, a final combination step, which requires a new threshold value, is used to fuse adjacent segments in the segmented image if their chromatic difference is lower than the threshold (without pre- an postprocessing MS obtains a score of 0.2972).

Finally, when comparing the different versions of RAD, we can see how, each of them improve in a coherent way the results obtained with the basic version of RAD. Thus, we can see how pRAD clearly outperforms results obtained with RAD, at the same computational cost. It makes pRAD, the best version when looking for a fast method of segmentation. Further, spRAD outperforms all the other methods. Nonetheless, its computational cost is much higher, since it computes five subsegmentations, a multicontrast image and a ranking of all the segments obtained.

5.5 Conclusions

This chapter introduces a new segmentation method, called pRAD, that extracts the ridges formed by a surface reflectance. This method is robust against discontinuities appearing in image histograms due to compression and acquisition conditions. Furthermore, those strong discontinuities, related with the physical illumination effects are correctly treated due to the topological treatment of the histogram and the addition of prior knowledge. As a consequence, the presented method yields better results than Mean Shift on a widely used image dataset and error measure. Additionally, even with neither preprocessing nor postprocessing steps, pRAD has a better performance than the state-of-the-art methods. Furthermore, we have proposed an improvement of pRAD, called spRAD, consisting in the addition of the spatial coherence to be less affected by texture edges and avoiding oversegmentation. spRAD outperforms results obtained with pRAD but at higher computational cost. Results obtained with pRAD point out that the chromatic information is an important cue on human segmentation. Additionally, the elapsed time for pRAD is not affected by its parameters. Due to that it becomes a faster method than Mean Shift and the other state-of-the-art methods.

spRAD is based on the saliency method detailed in Chapter 4. In the next chap-

ter we further evaluate the potential of using the saliency of image derivatives in segmentation evaluation.

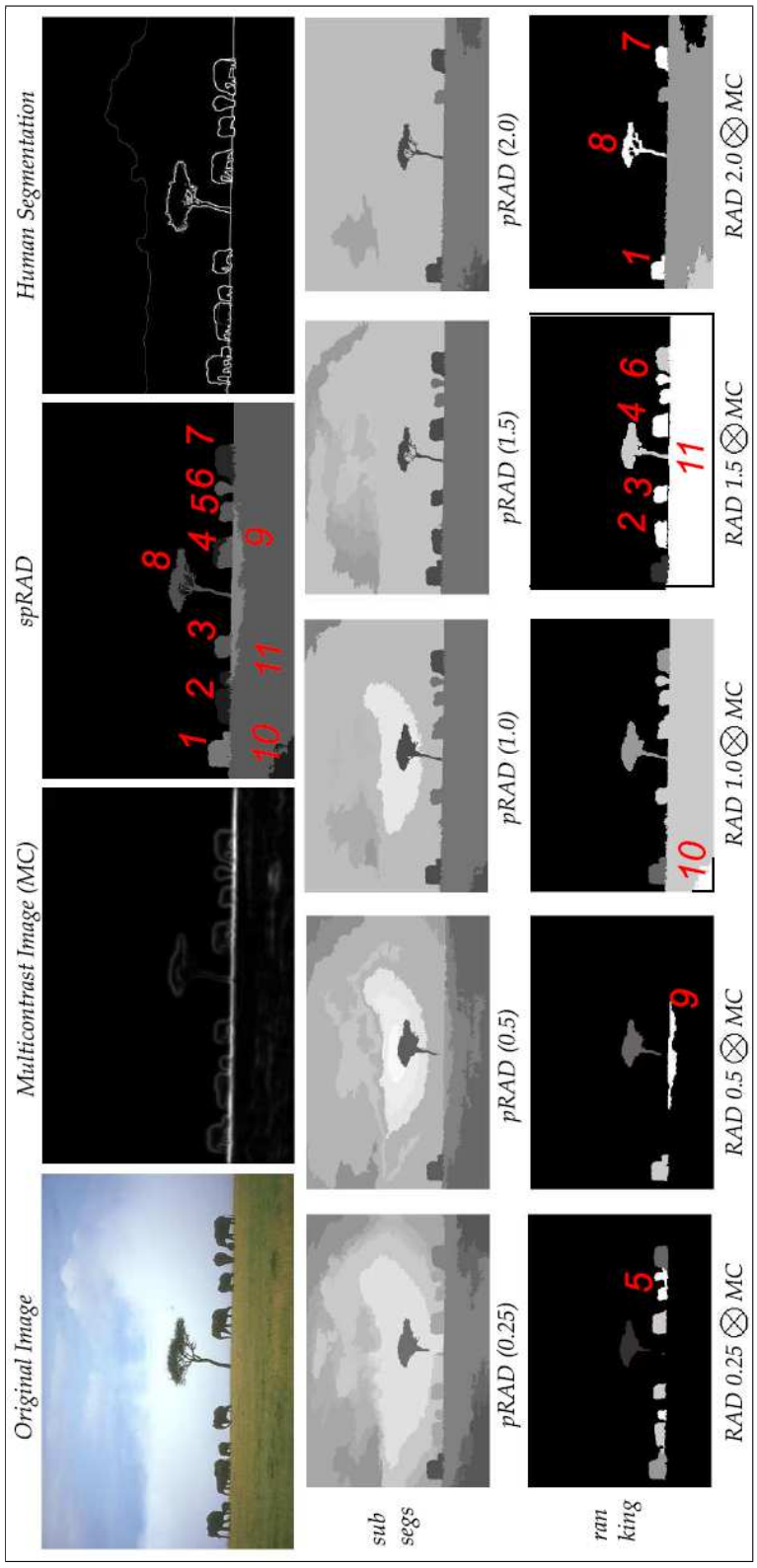


Figure 5.5: spRAD segmentation: from an original image, we generate a multiscale image (MC) and a number of sub-segmentations with different parameters of pRAD (second row). The goodness of a segment is computed with by summing the contrast underlying the edges of the segments normalized for the perimeter of the segment ($RAD \otimes MC$). The best segments will form the combined segmentation (spRAD).

Figure 5.6: Examples of the best performance of spRAD in textured images. spRAD assigns a single segment far all the flowers. A similar effect occurs with the plants of the images showed in the second row.

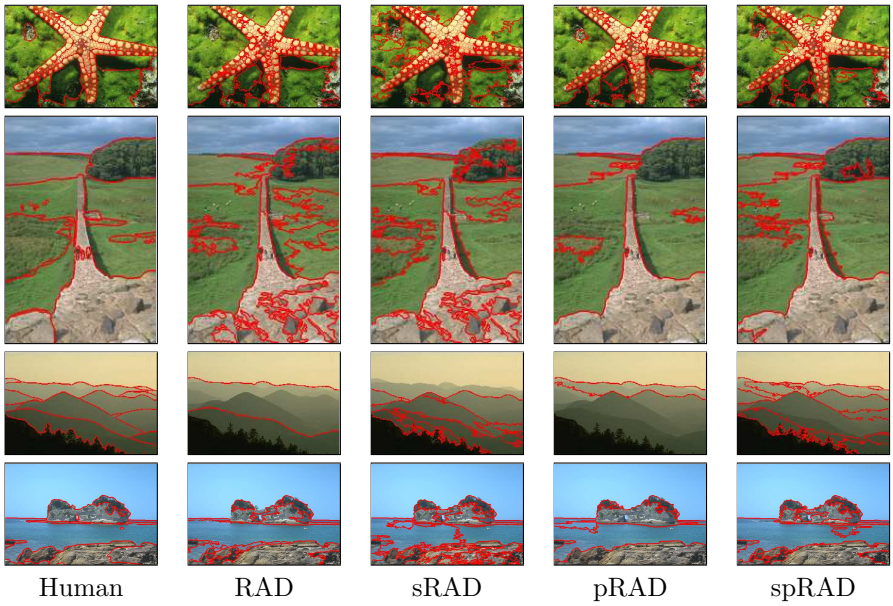


Figure 5.7: Examples of segmentation. First column: From first to last column, respectively: Human segmentation, RAD, sRAD, pRAD and spRAD. It can be observed that when adding spatial coherence, the segmentations have a closer similarity with human segmentation.

Chapter 6

Unsupervised Evaluation of Color Image Segmentation

A wide variety of segmentation approaches have been introduced along last years. Commonly these methods can be adapted to image content by changing a set of parameters which determine segmentation coarseness. Currently, one of the main challenges in segmentation is to perform such adaptation in a non-supervised manner. Due to that, applying a segmentation method without a previous, time-consuming supervision usually leads to inconsistent results. Saliency approaches have been shown to yield a good performance in unsupervised segmentation evaluation. Moreover, saliency based approaches facilitate a good guess about if objects in the scene are properly segmented, therefore making potentially easier ulterior stages such as inclusion of top-down information for object detection.
The saliency method detailed in Chapter4 has been shown to improve results obtained by RAD as described in Chapter 5. In this chapter we use our saliency method for unsupervised segmentation evaluation. Our approach is compared with a ground truth and a state-of-the-art saliency-based evaluation method by using diverse segmentation approaches and parameter settings. Results obtained show how our approach is successfully applied for non-supervised segmentation evaluation, helping in one of the main challenges on segmentation so far.

6.1 Introduction

Image segmentation aims to partition an image in a set of non-overlapped regions, called segments [34, 126]. Segmentation coarseness is determined by a set of parameters to better adapt results to image content. The segmentation coarseness required for a low-resolution image with a single object and few chromatic variations is opposed as that of a high-resolution image with multiple objects. A good segmentation method should be able to perform a correct segmentation in both scenarios by changing its settings. One of the main challenges in image segmentation is how to find out which parameters shall adapt the segmentation to each scenario in a non-supervised

manner. By doing so, undesired effects such as over and undersegmentation and other inconsistencies in general purpose segmentation, are minimized.

Some approaches has been proposed to perform this unsupervised evaluation. For instance, the JSEG method introduced in [46], which is a two-step segmentation schema. First, a clustering of the color space is performed. Afterwards, a criterion of *good* segmentation is applied using the spatial coherence of the image, *i.e.*, the information of the spatial relation existing between the pixels in the image space. Other measures to describe the goodness of a segment are the homogram proposed in [35], a calculus based in the Bhattacharyya distance [50] or a probabilistic approach as explained in [142].

Along with specific proposals, a family of methods based on image saliency have been shown to yield a good performance in unsupervised segmentation evaluation [70, 87, 130]. These biologically-inspired approaches have as an interesting characteristic that they facilitate a good guess about if objects in the scene are properly segmented, therefore making easier ulterior stages such as inclusion of top-down information for object detection.

The method detailed in [130] proposes that a good segmentation region should be formed by strongly connected pixels with homogeneous colors. This approach follows a similar idea as the one introduced in [87], which uses the color distinctiveness as a measure of goodness. The authors define a measure of color saliency of a segment, which considers its color distinctiveness, that is, its difference with the surrounding segments. Saliency is also used in other approaches as in [70], where a segmentation is considered *good* if it includes the most salient object of the image. The authors in the same article present a ground truth of the most salient objects in a set of images. Nonetheless, these methods have not been tested in common segmentation datasets such as the Berkeley segmentation one [135]. In this work we perform such evaluation.

The method proposed in this chapter is an extension of the color-boosting algorithm introduced in [198]. The method uses the saliency of the color image derivatives in the opponent chromatic space in a a multi-scale, center-surround schema [199] as commonly used in saliency algorithms [94] [121]. We evaluate the correspondence between the most salient edges and the edges of the segmented image.

Our method is evaluated using the ground-truth facilitated in the Berkeley dataset [135] using the Boundary Displacement Error measure [89][62]. We also compare our proposal with the Heidemann's saliency-based segmentation evaluation method introduced in [87]. To this aim, we use a set of segmentations obtained with the Mean Shift algorithm [39], the Efficient Graph-Based segmentation method [58] and the Ridge-Based Analysis of a Distribution method [200]. Results obtained show how our approach successfully evaluates the goodness of the segmentation methods used and clearly outperforms the state-of-the-art method presented in [87].

This chapter is organized as follows: in section 6.2 we explain the method used for segmentation evaluation. In section 6.3 we outline Heidemann's method. Afterwards, in section 6.4 we present the methodology used to evaluate our approach. Subsequently, sections 6.5 and 6.6 presents results obtained and the conclusions extracted respectively.

6.2 The saliency of the image derivatives

In this chapter, for segmentation evaluation, we propose to use a saliency method based on chromatic transitions [199]. This method computes the saliency of the image derivatives. As a results, we form the Boosting-based Images (BI). We transform these derivatives to a new space where the most salient transitions are enhanced by considering its information content [198]. The modelling of saliency based on the information content of the image has been assessed in several approaches [100] [63][132]. This theory holds that saliency is inversely related to the number of occurrences of a feature, in our case, chromatic transitions. Thus, those colors which barely appear in the image are the most informative and, therefore, the most salient. Chromatic information is closely related with contrast [22] [66]. Highly contrasted objects/surfaces are expected to be segmented [70] [87]. The most salient edges will be used to rank a set of segmentations.

The saliency method used in this work has two main improvements compared with the one originally proposed in [198]. First, it is not computed globally (for the whole dataset), but locally (for a single image). It makes the method more adaptable to any image singularities. Furthermore, we propose another biologically-inspired mechanism to improve results obtained with Color Boosting. To generate an image in concordance with the multi-scale way to process images of the HVS, we build a pyramid of Gaussians [78] [94] [121]. Then we compute the color boosting transformation at each level and, finally we perform a center-surround calculus to build an image which is less affected by textured parts of the image keeping those most informative and contrasted objects of the scene.

6.2.1 Color Boosting

Color boosting is used to find the most informative chromatic transitions of an image. This method is based on the self information of the chromatic transition (first order derivatives of the image). It is showed in [198] that Color Boosting improves the color distinctiveness in a framework of interest points detection.

The color saliency method introduced by Van de Weijer *et al.* in [198] is inspired by the notion that a feature's saliency reflects its information content. Consider an image $\mathbf{f} = (R, G, B)^{\mathbf{t}}$. The information content, I, of an image derivative $\mathbf{f_x}$, according to information theory, is given by the logarithm of its probability p:

$$I = -log(p(\mathbf{f_x})). \tag{6.1}$$

Hence, color image derivatives which are equally frequent have equal information content. We choose to map the derivatives to a new space where isosalient derivatives have equal norms:

$$p(\mathbf{f}_x) = p(\mathbf{f}'_x) \leftrightarrow |g(\mathbf{f}_x)| = |g(\mathbf{f}'_x)|. \tag{6.2}$$

The saliency function g transfers color image derivatives to a space where their norm is proportional to their information content.

It can be seen in [198] that the derivatives form an ellipsoid-like distribution. The longest axis corresponds with the luminance direction. This indicates that equal displacements are more informative along the color directions (perpendicular to the

luminance) than in the luminance direction. In the original work these statistics are computed in the opponent. Then, a single color boosting transformation is obtained from the statistics computed on a whole dataset, which might be used for any image. As in [199] we use a more general transformation to compute g in that it is not computed in a dataset but for any single image. The improvements obtained are shown in [199] in the framework of saliency.

Let the distribution of the ellipsoid to be described by the covariance matrix $\mathbf{M}$ between color channels:

$$\mathbf{M} = \overline{\mathbf{f_x}\left(\mathbf{f_x}\right)^{\mathbf{t}}} = \begin{pmatrix} \overline{R_{\mathbf{x}}R_{\mathbf{x}}} & \overline{R_{\mathbf{x}}G_{\mathbf{x}}} & \overline{R_{\mathbf{x}}B_{\mathbf{x}}} \\ \overline{R_{\mathbf{x}}G_{\mathbf{x}}} & \overline{G_{\mathbf{x}}G_{\mathbf{x}}} & \overline{G_{\mathbf{x}}B_{\mathbf{x}}} \\ \overline{R_{\mathbf{x}}B_{\mathbf{x}}} & \overline{G_{\mathbf{x}}B_{\mathbf{x}}} & \overline{B_{\mathbf{x}}B_{\mathbf{x}}} \end{pmatrix} \tag{6.3}$$

where the matrix elements are computed by

$$\overline{R_{\mathbf{x}}R_{\mathbf{x}}} = \sum_{\mathbf{x}\in X} R_{\mathbf{x}}\left(\mathbf{x}\right) R_{\mathbf{x}}\left(\mathbf{x}\right) \tag{6.4}$$

where X is the set of pixels coordinates $\mathbf{x}$ in an image. Matrix $\mathbf{M}$ describes the derivatives energy in any direction $\hat{v}$. This energy is computed by $E(\hat{v}) = \hat{v}\mathbf{M}\hat{v}^{\mathbf{t}}$. Matrix $\mathbf{M}$ can be decomposed into eigenvector matrix $\mathbf{U}$ and eigenvalue matrix $\mathbf{\Lambda}$ according to $\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{\Lambda}\mathbf{U}^{\mathbf{t}}$. This provides us with the saliency function g:

$$\mathbf{g}\left(\mathbf{f_x}\right) = \mathbf{\Lambda}^{-1}\mathbf{U}^{\mathbf{t}}\mathbf{f_x}. \tag{6.5}$$

Substitution of Eq. 6.5 into Eq. 6.3 yields

$$\mathbf{g}\left(\mathbf{f_x}\right)\left(\mathbf{g}\left(\mathbf{f_x}\right)\right)^{\mathbf{t}} = \mathbf{\Lambda}^{-1}\mathbf{U}^{\mathbf{t}}\mathbf{U}\mathbf{\Lambda}\mathbf{\Lambda}\mathbf{U}^{\mathbf{t}}\mathbf{U}\mathbf{\Lambda}^{-1} = I, \tag{6.6}$$

meaning that the covariance matrix of the transformed image is equal to the identity matrix. This implies that the derivative energy in the transformed space is equal in all directions. In this case, the matrix $\mathbf{U}^{\mathbf{t}}$ corresponds with the transformation matrix to the Opponent color space.

We use the modulus of the transformed image to build a gray-scale image which will be used to evaluate a segmentation, we call this image BI (boosting-based image). Figs. 6.1c-f show four examples where derivatives have been computed at four different scales. When comparing BI with the ground-truth in Fig. 6.1b, we can appreciate a high correspondence between them. Edges drawn by humans are clearly visible in BIs. The main problem is that there is also too much non-significative information mainly related with the textures formed by the rocks and the clouds.

To better suppress spurious transition we propose to use a biologically inspired mechanism as the center-surround image using a multi-scale schema.

6.2.2 Multi-scale, center-surround boosting

We compute the boosting image at several scales by building a pyramid of Gaussians [78]. Afterwards we generate a center-surround image which considers pixel differences

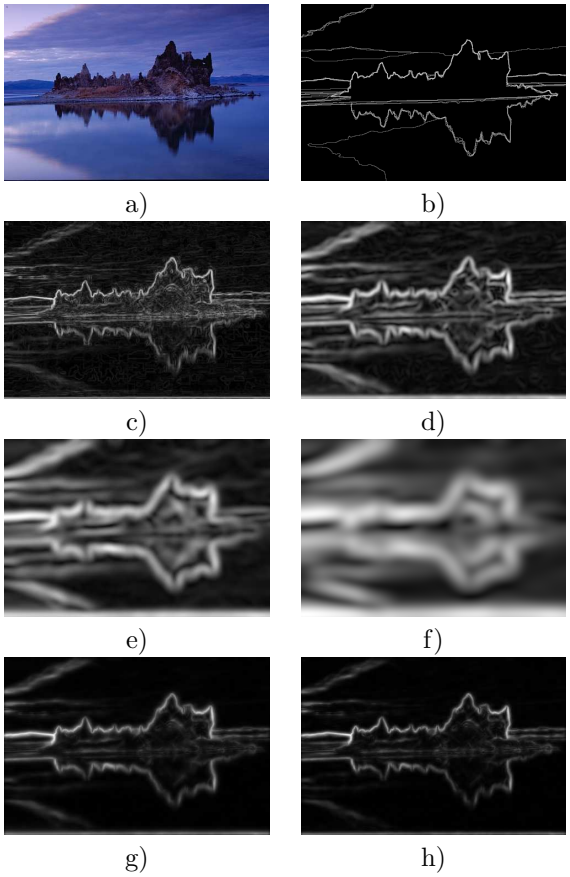


Figure 6.1: a)Original image from [135]. b)Ground truth [135]. c-f)Single-scale boosting images computed at *sigma* = 1, *sigma* = 4, *sigma* = 8, *sigma* = 16 respectively. g) Boosting with $\sigma = \{1, 2, 4, 6, 8, 10, 12, 14\}$ as proposed in [121]. h) Boosting following human perceptual octaves, $\sigma = \{1, 2, 4, 8, 16\}$.

at multiple scales. With this approach, those transitions which are representative at multiple scales are strongly detected.

Maps computed at multiple scales can be combined into a single saliency map as follows:

$$s(\mathbf{x}) = \sum_{\sigma \in \Sigma} \sum_{\mathbf{x}' \in N(\mathbf{x})} \left\| \mathbf{M}^{\sigma} \left(\mathbf{f}^{\sigma}(\mathbf{x}) - \mathbf{f}^{\sigma}(\mathbf{x}') \right) \right\| \tag{6.7}$$

where $\mathbf{f}^{\sigma}$ denotes the Gaussian smoothed image at scale σ, and $\sigma = \{1, 2, 4, 8, 16\}$. $N(\mathbf{x})$ is a 9x9 neighborhood window. $\mathbf{M}^{\sigma}$ is the transformation matrix computed from Gaussian derivatives of scale σ computed as explained in section 6.2.1. Note that leaving out $\mathbf{M}$ from Eq. 6.7 results in the multi-scale contrast approach proposed by Liu et al. [121]. Two examples of a multi-scale color saliency map is given in Fig. 6.1g,h. The boosting image depicted in Fig. 6.1g corresponds with the set of scales

proposed in [121], namely, $\sigma = \{1, 2, 4, 6, 8, 10, 12, 14\}$ whereas Fig.6.1h shows the BI image computed at scales $\sigma = \{1, 2, 4, 8, 16\}$ which corresponds with the human visual system, based on octaves [15]. The latter one, since has a biological explanation is the one that we use in our approach.

6.2.3 Applying boosting for evaluation

BI images can not be directly used for our aim. Even with the multiscale approach there is information in the whole image. nonetheless, most salient edges are much more clear than with the single scale approach. It allows the application of a threshold. In our case, we use values falling into the ten topmost percentiles of the BI image. Afterwards, we compute the skeleton of the images. The distance between the borders of the image and the skeleton of the BI images will be used to evaluate the segmentation.

Fig 6.4 shows three examples of the final BI. In Sec.6.4 we give further details about the evaluation.

6.3 Heidemann's color saliency

The approach of Heidemann introduced in [87] proposes a goodness function for color segmentation, which allows to predict whether the segmented regions will be stable against noise, variation of lighting, and change of viewpoint. Color saliency is defined from the average border contrast of the segmented image. Experiments for three different algorithms show that the performance is independent of the particular functional principle of segmentation. Thus, the measure can be applied for the automatic and context-free optimization of segmentation parameters.

The measure proposed is based on the color distinctiveness of the regions of the segmented image. Thus, as larger the (Euclidean) distance between neighboring regions, the better is the segmentation. Given an image $\mathcal{I}$ having three chromatic channels for each pixel (x, y), we compute a segmentation from which $\mathcal{I}$ is divided in N_R non-overlapped regions. The *region color* is defined as the mean color of this region in the original image.

The *region saliency* $S_R(R_i)$ is defined as the average color difference of R_i to the neighboring regions. Concretely, let the boundary of R_i be given as a set $B(R_i)$ consisting of $N_B(R_i)$ different pixels. Then $S_R(R_i)$ is calculated along the boundary as

$$S_R(R_i) = \frac{1}{N_B(R_i)} \sum_{(x,y)\in B(R_i)} \frac{1}{N_{diff}(x,y)} \times \sum_{R_j(x',y')|(x',y')\in Neigh4(x,y)} \| \overline{C}(R_i) - \overline{C}(R_j) \| . \quad (6.8)$$

Here, $\|\|$. denotes the color distance measure for the particular segmentation algorithm used. For color spaces such as RGB, $L*u*v*$ or $L*a*b*$ the Euclidean distance is used.

The first sum in Eq. 6.8 is over all boundary pixels (x, y). The second sum goes over the pixels (x', y') within a 4-neighborhood of (x, y) being denoted by $Neigh4(x, y)$. To each neighboring pixel (x', y') the corresponding region $R_j(x', y')$ has to be found, so that the Euclidean distance between the region colors $\overline{C}(R_i)$ and $\overline{C}(R_j)$ can be calculated. $N_{diff}(x, y)$ denotes the number of pixels of $Neigh4(x, y)$ that belong to a different region, not to R_i. This factor is introduced to avoid dilution of the average distance in case there is, e.g. only one neighboring pixel which belongs to a different region. $N_{diff}(x, y)$ is at least 1 since (x, y) is part of the boundary, the maximum value is $N_{diff}(x, y) = 4$ in the case that (x, y) is a region consisting of an isolated pixel.

The Saliency measure of an image $\mathcal{I}$ denoted by $S(\mathcal{I})$ is given by the average over all its regions

$$S(\mathcal{I}) = \frac{1}{N_R} \sum_{R_i \in I} S_R(R_i) \tag{6.9}$$

Summarizing, $S(\mathcal{I})$ is a measure of the color distinctiveness of the regions of a segmented image.

In the next section we explain a another way to include color distinctiveness to and contrast to decide the goodness of a segmentation.

6.4 BI evaluation

The performance of BI has been evaluated using the ground truth facilitated in [135]. In addition, our approach is compared with a state-of-the-art method introduced by Heidemann in [87].

6.4.1 Ground truth and error measure

The ground-truth used is formed by 300 images labelled by 6 users [135].

Along with the ground-truth, an error measure called Global Constancy Error, was also proposed. The limitation with the Global Constancy Error is that it can just compare two segmentations if they have a similar number of segments. Such an error measure is not valid in our evaluation, since we expect to evaluate in a fully-non supervised way a set of segmentations having a number of segments which goes from few hundreds to just 8 segments. Moreover, BI images draw the most salient transitions on the images but not forming closed objects. Fig. 6.4 shows some example of BI images after thresholding. This information can be successfully used with the Boundary Displacement Error (BDE). This method, introduced in [89], evaluates the precision of the extracted region boundaries [62]. Let B be the estimated boundary and G the ground-truth boundary. The method uses two distance distribution signatures from the estimated to the ground truth borders, denoted by D_G^B and vice versa, denoted by D_B^G. For two sets of boundary points B_1 and B_2, $D_{B_1}^{B_2}$ is a discrete function whose distribution characterizes the discrepancy, measured in distance, from B_1 to B_2. The authors define the error measure as the minimum absolute Euclidian distance. $D_{B_1}^{B_2}$ which can be established from the distance histogram from individual

$x \in B_1$ to B_2. It can be estimated through a distance transformation with respect to B_2.

6.4.2 Segmentation methods used

To perform the evaluation we have selected three segmentation methods, namely, the Efficient Graph-based method [58] (EG), the Ridges-based Analysis of a Distribution (RAD) [200] and the Mean Shift (MS) [39]. These methods perform the segmentation in three different ways, namely, image-based, feature-based (histogram), and using both image and histogram space respectively.

The efficient graph-based method performs the segmentation in the image space and has a public available code. We have selected 6 different sets of parameters. From the parameters recommended for the authors to yield a good performance, we have empirically selected a set of parameters to go to a slight oversegmentation to an undersegmentation. These parameters are (k,σ)= {(250,0.5) , (250,2.5) , (250,5) , (500,5) , (1000,0.5) , (1000,5)}.

RAD performs the segmentation in the histogram space and has been demonstrated to yield state-of-the-art results. We have also performed 6 sets of segmentations with RAD following the same criteria, that is, to go from a slight oversegmentation to an slight undersegmntation. In this case we have (σ_d,σ_i)={ (0.5,0.05) , (0.5,0.8) , (0.8,1.5) , (1,0.5) , (1.5,0.05), (1.5,1.5) }.

Finally, MS performs the segmentation using a combination of the histogram space and the image one. MS also has a public available version, called EDISON [37]. In this case the parameters have been $(h_s, h_r) = \{(7,3),(7,19),(13,7),(17,23),(20,25),(30,35)\}$ as suggested in [200]

Examples of results obtained with these methods are depicted in Fig.6.2.

6.5 Results obtained

In this section BI is compared with a the public available ground-truth [135] and with the sate-of-the-art Heidemann method introduced in [87], ranking 3 segmentation methods with 6 sets of parameters each. The error measure used is the boundary displacement error (BDE) [89][62].

Fig.6.3 shows some examples of BI images for a qualitative evaluation of our proposal. From these examples it stands out the high correlation between BI and the ground-truth. Second and fourth rows illustrate particularly interesting examples. The former shows how BI draws the borders corresponding with the two people and the structure on the back despite the plants of the floor, which are considered not informative. In the fourth row, the giraffes have a color fairly similar to the floor. Even in this case, our approach generates an image with the giraffes and the horizon's line. Furthermore the clouds are also detected as low informative. Finally, the last row is an example of a case where BI do have a lower similarity with the ground-truth. In this case, the information content of the animal is similar to some parts of the background. Nonetheless, the borders of the animal are correctly drawn.

Finally, some examples of the BI images after thresholding are shown in fig.6.4.

Table 6.1: First Column: segmentation method and parameters used. Second and third columns: single-scale BI. Fourth: multi-scale BI. Last column: ground-truth. Multi-scale BI ranks the segmentation as the ground truth does so.

	$BI_{\sigma=1}$	$BI_{\sigma=4}$	$BI_{\sigma=\{1,2,4,8,16\}}$	Grnd Truth
MS (13,7)	2.85 (#2)	4.78 (#2)	7.36 **(#1)**	13.54 **(#1)**
MS (7,19)	3.28 (#3)	5.55 (#4)	7,52 **(#2)**	13,65 **(#2)**
MS (7,3)	4.25 (#4)	6.25 (#3)	8,40 **(#3)**	14,21 **(#3)**
MS (17,23)	2.43 (#1)	3.19 (#1)	8,46 **(#4)**	14,39 **(#4)**
MS (20,25)	5.06 (#5)	6.91 (#5)	8,85 **(#5)**	14,94 **(#5)**
MS (30,35)	5.65 (#6)	7.40 (#6)	11,4 **(#6)**	17,46 **(#6)**
RAD (1.5,1.5)	3,13 (#2)	5,15 (#2)	7,49 **(#1)**	13.18 **(#1)**
RAD (0.8,1.5)	2,59 (#1)	3,37 (#1)	7,969 **(#2)**	13,50 **(#2)**
RAD (1.5,0.05)	3.18 (#3)	3.87 (#3)	9,43 **(#3)**	14,43 **(#3)**
RAD (0.5,0.8)	5,30 (#5)	6,72 (#5)	10.41 **(#4)**	15.43 **(#4)**
RAD (0.5,0.05)	4,87 (#4)	6,11 (#4)	10,43 **(#5)**	15,65 **(#5)**
RAD (1,0.5)	5.73 (#6)	10.40(#6)	10,67 **(#6)**	16,24 **(#6)**
EG (250,2.5)	4.45 (#3)	7.36 (#2)	6,85 **(#1)**	12,91 **(#1)**
EG (1000,0.5)	3.79 (#2)	8.17 (#3)	6,99 **(#2)**	13.17 **(#2)**
EG (250,0.5)	6.52 (#4)	10.71 (#4)	7,94 **(#3)**	13.59 **(#3)**
EG (250,5)	2.53 (#1)	3.99 (#1)	8,61**(#4)**	14,86 **(#4)**
EG (500,5)	14.61 (#5	19.43 (#5)	11,07 **(#5)**	18.95 **(#5)**
EG (1000,5)	23.73 (#6)	29.25 (#6)	17,40 **(#6)**	26.88 **(#6)**

Table 6.2: First Column: segmentation method and parameters used. Second column: Heidemann. Third column: multi-scale BI. Last column: ground-truth. Multi-scale BI ranks the segmentation as the ground truth does so, clearly outperforming results yield by Heidemann.

	Heidemann	$BI_{\sigma=\{1,2,4,8,16\}}$	Ground Truth
MS (13,7)	2,10 (#4)	7.3693 **(#1)**	13.5484 **(#1)**
MS (7,19)	1,07 (#5)	7,5255 **(#2)**	13,6511 **(#2)**
MS (7,3)	0,26 (#6)	8,4033 **(#3)**	14,2119 **(#3)**
MS (17,23)	2,96 (#3)	8,4622 **(#4)**	14,3994 **(#4)**
MS (20,25)	3,83 (#2)	8,8568 **(#5)**	14,9489 **(#5)**
MS (30,35)	6,90 (#1)	11,416 **(#6)**	17,461 **(#6)**
RAD (1.5,1.5)	2,91 (#4)	7,4923 **(#1)**	13.1845 **(#1)**
RAD (0.8,1.5)	1,93 (#6)	7,969 **(#2)**	13,501 **(#2)**
RAD (1.5,0.05)	2,90 (#5)	9,4345 **(#3)**	14,4385 **(#3)**
RAD (0.5,0.8)	23,9 (#1)	10.4127 **(#4)**	15.4366 **(#4)**
RAD (0.5,0.05)	16,4 (#3)	10,4354 **(#5)**	15,6537 **(#5)**
RAD (1,0.5)	16,5 (#2)	10,6759 **(#6)**	16,2466 **(#6)**
EG (250,2.5)	14,6 (#4)	6,8534 **(#1)**	12,9145 **(#1)**
EG (1000,0.5)	18,0 (#2)	6,9964 **(#2)**	13.1711 **(#2)**
EG (250,0.5)	3,15 (#6)	7,9487 **(#3)**	13.5904 **(#3)**
EG (250,5)	18,1 (#1)	8,6135**(#4)**	14,8632 **(#4)**
EG (500,5)	16,2 (#3)	11,0774 **(#5)**	18.9582 **(#5)**
EG (1000,5)	10,5 (#5)	17,4058 **(#6)**	26.8829 **(#6)**

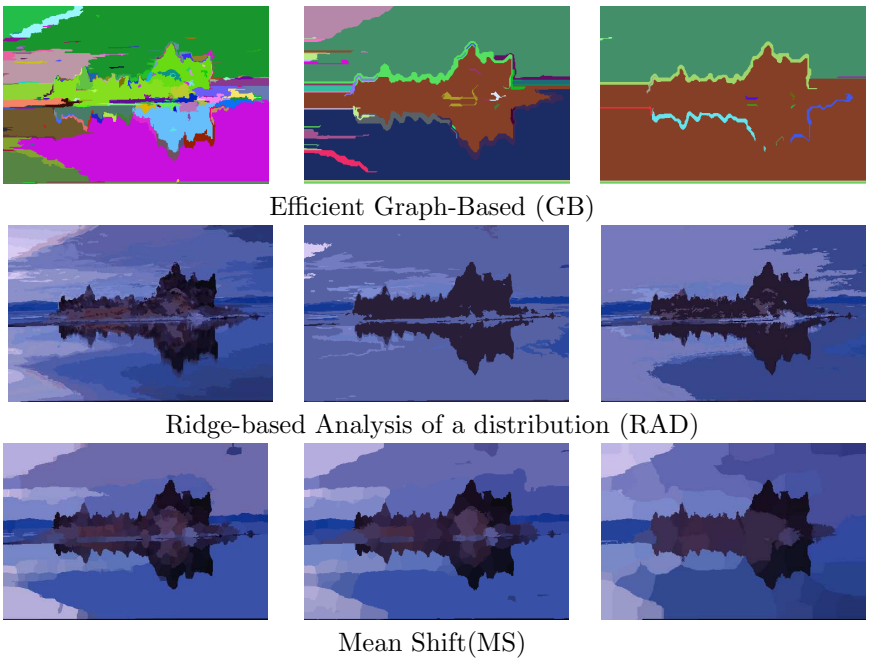


Figure 6.2: Examples of segmentation results. The original image and the ground-truth are showed in Fig.6.1a,b respectively.

Scores obtained using the boundary displacement error (BDE) [89][62] as an error measure among the 300 images of the Berkeley dataset [135] are shown in Tables 6.1 and 6.2. Table 6.1 summarizes the performance of the single scale and the multi-scale approaches of BI. In the first row it is shown the 3 segmentations used, namely, mean shift (MS) [39], efficient graph-based segmentation (EG) [58], and ridge-based analysis of a distribution segmentation method (RAD) [200]. For each segmentation method we perform six different segmentations which parameters setting are showed within brackets. Single-scale BI (second and third columns) and multi-scale BI (fourth column) are compared with a ground truth [135] (last column). For each combination of segmentation and ground-truth, the score obtained with BDE is displayed. Using this score we can rank the segmentation methods, as displayed within brackets ((#1),...,(#6)). It is shown how the ground truth ranks the segmentations in the same way multi-scale BI does so, for all segmentation methods and sets of parameters. Furthermore, Table 6.1 also illustrates the necessity of using a center-surround multi-scale framework. When generating BI at a single scale, the method does not succeed in its ranking in all the cases. Table 6.2, shows a comparison between multi-scale BI (third column) and Heidemann (second column). Note that results obtained by Heidemann approach does no correspond with those obtained with the ground truth. This table shows how BI clearly outperforms Heidemann in segmentation evaluation.

Results obtained in a general ranking are depicted in Fig. 6.5. The segmentation methods in the abscissa axis are sorted with the BDE value obtained by the ground

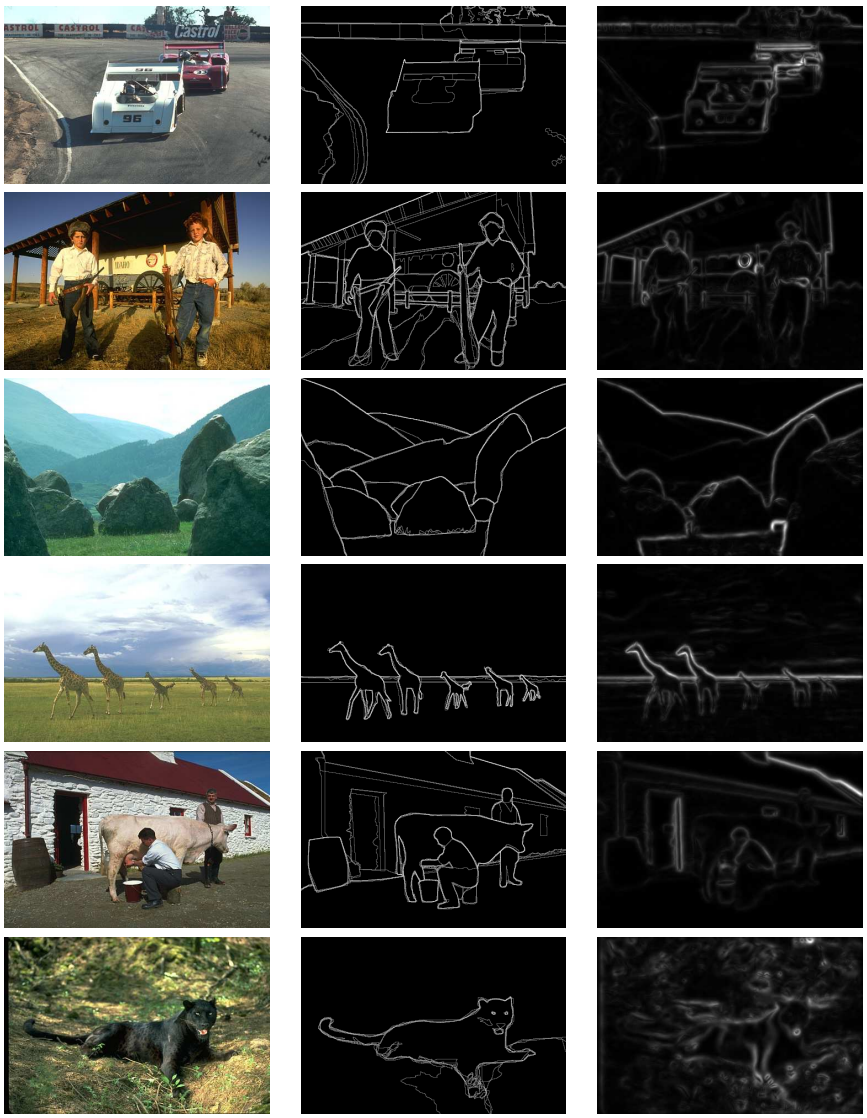


Figure 6.3: The similarity of a human made ground-truth and our approach BI can be seen in these examples. Second and fourth rows show a particulary challenging image where our approach yields good results. Last row show an example where BI is not that close to human segmentation. Nonetheless, the animal's borders are correctly drawn.

truth, shown in the ordinate axis. It is also displays the BDE scores obtained by our approach (BI). It can be seen that those methods incorrectly ranked by BI as the third and the fourth methods, have a virtually equal performance. Furthermore, in

Figure 6.4: Examples of BI after applying a threshold and finding the skeleton.

can be seen that by multiplying the BDE error obtained by BI, it is obtained almost the BDE scores derived from the ground truth.

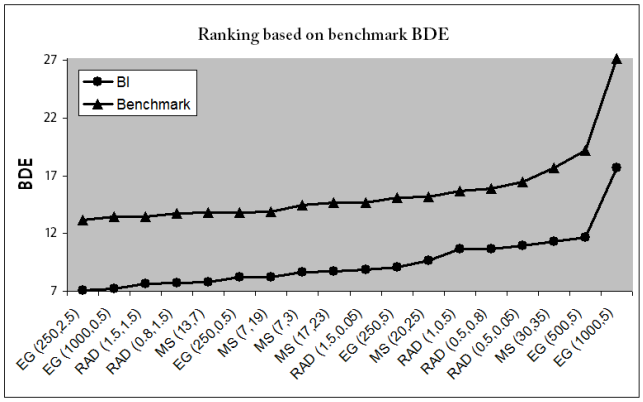


Figure 6.5: General ranking based on the BDE score obtained with the ground truth, along with corresponding values of BDE score obtained by our approach (BI). It shows that the confusions in the ranking as obtained by BI corresponds to method with virtually equal performance.

6.6 Discussion and further work

In this chapter a new approach for unsupervised segmentation evaluation has been introduced. It uses the saliency of the chromatic transitions in an image. The most salient transitions on the image, as computed by multi-scale boosting [199] shows a remarkably resemblance with a ground truth [135] used for evaluating the performance of our approach. Results obtained shows how using our approach we rank a set of segmentations with diverse parameter settings in a very similar way as using the ground truth. Differences are obtained by a minimal error in the BDE error measure. Furthermore our approach clearly outperforms results obtained with a state-of-the-art method.

A method to avoid the threshold step in our approach might yield better results. It could be performed by finding the ridges of maximum energy in the multi-scale BI images.

Chapter 7

Conclusions

Image segmentation is still a challenging task in Computer Vision, despite the burgeoning of multiple approaches in the recent years. A remarkable example of this is the role of segmentation in object recognition and classification. Segmentation techniques, used as a preprocessing step, are often replaced by superpixels approaches or superpixels versions of segmentation algorithms. Several important questions in segmentation remain open. Moreover, there are some questions that cannot be formulated without taking into account a degree of uncertainty and faultiness. Probably, the main question about segmentation is:

'**what is a *correct* segmentation?**'.

Nevertheless, it might be argued that such question is pointless for there is no correct segmentation, but only a segmentation which is suitable for a given problem. This point of view remains unclear, although some authors might firmly disagree. This issue leads to the next essential question:

'can we talk about general purpose segmentation without knowing if there is a correct (general) segmentation?'.

In order to shed light onto this second question we need to think over the first question. Probably there is not a categorical, irrefutable answer to it, but we can be sure that there is a *correct* segmentation for a given specific problem, *e.g.*, face segmentation. In this case, we would be looking for specific colors (assuming or not a correct color constancy algorithm involved), known shapes, and so on. Not in vain, segmentation in computer vision is an active research area. **A segmentation method can be suitable when looking for uniform textures, another method can be better when looking for faces, or cars, or people, and so on**. Undoubtedly, a proper segmentation can most effectively assist in recognizing any object or action on a scene. Hence, the aforementioned questions should be reformulated into the following question: 'can we find a suitable segmentation method for our framework?'. A segmentation method is essentially a technique **which aims to**

find regions in the image sharing specific properties such as color, shape or texture. Segmentation does not aim to find specific objects in a scene or to identify actions in a video sequence. This dissertation follows the same chain of thoughts. Concretely, we have proposed a multidisciplinary approach to deal with chromaticity and illumination in a scene. We have adopted a technique of medical imaging, the MLSEC-ST operator, which we have therefore applied to the image histogram.In our final proposal, called spRAD, we have included a saliency measure to cope with chromaticity. spRAD state-of-the-art performance has been also verified. Nevertheless, although spRAD proved to outperform other segmentation methods, we cannot firmly assert that it can replace other segmentation methods. For this reason, the final part of this dissertation is to be read in the light of an effort to facilitate an unsupervised method of segmentation evaluation. In this sense the main strengths of spRAD are:

- Its good behavior in the presence of shadows and highlights. spRAD describes in a robust manner light changes in a scene.

- Results obtained in textures. Changes in light in the textures do not lead to oversegmentation. Smooth-confusing textures in the scene with similar colors (such as foliage) can be successfully segmented.

Regarding the weaknesses of spRAD, we can affirm that:

- It may result in undersegmentation when the scene presents poor chromatic variety.

The final proposal, spRAD, gives a suitable framework to overcome this problem. We might include a set of subsegmentations based on the intensity channel and a two dimensional color spaces as the rgb (see Appendix A).

As already mentioned in the introduction of this dissertation, the necessity of including top-down information in segmentation is an interesting discussion topic. The question is:

'**does segmentation can be correct without top-down information?**'.

This is indeed a common discussion in other computer vision fields. In the case of segmentation, the answer comes from the aforementioned reasoning . Segmentation is correct as far as it fits to a specific framework. spRAD, for instance, is a good segmentation method for textures and scenarios with a variety of colors. Top-down information is required when we want to segment semantic objects in a scene. However, even in that case, a correct initial segmentation, commonly derived from bottom-up methods, is to be applied. Thus, it is necessary, for different reasons, to have a robust bottom-up segmentation schema. Top-down information is indeed a part of a global schema for complex computer vision tasks. This global schema needs a good knowledge of bottom-up information in the scenario to yield a comprehensive solution to a complex task such as action recognition.

7.1 Contributions of this dissertation

In this thesis we have presented a new hybrid segmentation method. The main idea lies in the use of the MLSEC-ST operator, initially thought for finding ridges in medical imaging, as a method for analyzing a color histogram. Ridges found for our approach describe a single material reflectance. This approach overcomes the main shortcomings of the dichromatic reflection model which was an inspiration for our work.

The first approach presented, called RAD, is further completed by the addition of physics-based statistics to suppress spurious ridges, forming a second approach called pRAD. Finally we have included image coherence into our segmentation model by using the saliency of the chromatic transitions. This last approach, called spRAD is a non-supervised segmentation method which overcomes the other two proposals and state-of-the-art segmentation methods.

The saliency method used as a basis of sRAD is a new proposal also presented in this dissertation. It has been validated in a computational way as well as by means of a psychophysical experiment. The good performance of our saliency method yielded by sRAD, has motivated the proposal of a method of non-supervised segmentation evaluation. This also outperforms state-of-the-art methods of segmentation evaluation based on saliency.

The main contributions of this work are:

- Overcoming of the main drawbacks of the dichromatic reflection model. RAD presents a more flexible behavior in practice than the dichromatic reflection model which leads to a more robust extraction of a single material reflectance.
- We have presented a new hybrid segmentation model which takes the strengths of the three main categories of segmentation methods. Feature space information is analyzed by the MLSEC-ST operator. Physics-based cues are introduced by a statistical analysis of the directions of the material reflectances. Image coherence is added by means of a saliency-based approach. The resulting segmentation method, spRAD, overcomes state-of-the-art segmentation methods. We point out that the combined segmentation obtained with spRAD outperforms all results obtained with its sub-segmentations.
- spRAD is a fully non-supervised segmentation method, which is one of the main challenges in segmentation so far.
- A new saliency method has been presented and validated both computationally and psychophysically. Our experiments suggest that red-green transitions are more salient than blue-yellow and intensity ones.
- The saliency approach has been used as a non-supervised segmentation evaluation method which overcomes a state-of-the-art segmentation method and allow to rank a set of segmentations as when using a human-made ground-truth.
- Another contribution is the application of the MLSEC-ST operator, formerly introduced as a method to find ridges in gray-scale medical imagery, as a histogram analysis method.

7.2 Further work

Further lines of research derived from this dissertation has been already pointed out. Those, along with some additional interesting sources of study are summarized below.

Image segmentation

spRAD can be further improved by introducing subsegmentations based on intensity or two-dimensional chromatic spaces such as rgb.

The combination of subsegmentation in spRAD might be performed using more complex techniques which might improve results. For instance, instead of applying to the saliency images, an adaptive threshold aimed in order to find closed regions which would maximize energy, a simple threshold has been replaced by an adaptive threshold, since the latter is expected to have a better correspondence with objects in a scene.

Saliency

Some interesting conclusions, arising from the psychophysical experiments, deserve further attention. The apparent higher responses of subjects to red-green transitions suggest that we are indeed more sensible to such wavelengths which form our visual system.

RAD as a general manifold analysis method

In this work RAD has been presented as a segmentation method, although it can be considered as a generic manifold analysis method. Some efforts to find out its potential as a general technique have been already performed. RAD has been applied to forming a color space adaptive to image content in [201]. Furthermore, RAD is being used for a color constancy approach with very promising results.

Appendix A

Appendix A: Colour spaces brief discussion.

Along this thesis we have presented results and methods working in RGB, Luv and opponent chromatic spaces. In this Appendix we explain the most common chromatic spaces used in computer vision. Some more recent spaces has been proposed which are claimed to overcome main drawbacks associated with spaces such as RGB or Luv. Nonetheless, a detailed discussion about all chromatis spaces is out of the scope of this appendix.

Though lots of colour spaces has been propose in order to find a correct representation of colour [57, 133, 2], in this section we will just describe the most common and used of them. We divide the most common colour models in *device dependent* and *device independent* colour spaces.

A.1 Device Dependent colour spaces.

Here are included all those spaces where the position of a colour in them, is directly calculated in function of the trichromatic values received from the sensor. This kind of spaces also belongs to the category of non-perceptual uniform colour spaces, i.e., its relative distances between colours do not reflect the perceptual differences. It means that when we move a distance d in this spaces, depending the direction taken, the perceptual differences are not the same. It is a shortcoming in the feature space based segmentation algorithms, because some perceptual low differences, can be treated as high perceptual differences, and the final segmentation can be perceptually inconsistent.

A.1.1 RGB space

Probably the most used colour space is the RGB space due to acquisition and display devices used to work with this three chromatic representation. Red, green and blue components are the sum of the respective sensitivity functions and the incident light and are based in the following equations:

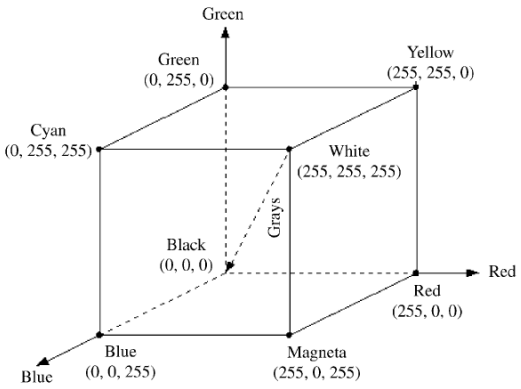


Figure A.1: RGB graphical representation. Main diagonal from black (0,0,0) to white (255,255,255) represents the gray values from low to high intensities.

$$\begin{aligned} R &= \int_{\lambda} S(\lambda) f_R(\lambda) d\lambda \\ G &= \int_{\lambda} S(\lambda) f_G(\lambda) d\lambda \\ B &= \int_{\lambda} S(\lambda) f_B(\lambda) d\lambda \end{aligned} \tag{A.1}$$

Where $S(\lambda)$ is the light spectrum, λ is the wavelength and f_R, f_G and f_B are the sensitivity functions for the R,G and B sensors respectively. Graphically, this representation is a three-dimensional cube with R,G and B as coordinates, commonly with values from 0 to 255. In this cube, $[0, 0, 0]$ are black while white is in the opposite vertex, i.e., $[255, 255, 255]$. The diagonal from black to white represents the gray-values from low to high intensities. Figure A.1 shows an example. Hence, by taking perpendicular planes to gray-diagonal, we obtain planes with constant lightness.

The disadvantages of RGB space are, first, that this is a device dependent space, due to its values depend of the sensitivity functions; second, its high correlation between its components and third, the fact that this is not a perceptual uniform representation.

Standard RGB.

In spite of drawbacks related to RGB space, nowadays, the standard RGB (sRGB) representation has become a standard. The difference between RGB and sRGB is the *gamma correction* introduced to improve visualization in common displays and environments by raising R, G and B channels to a γth power; commonly we find $\gamma = 1.2$. All disadvantages of RGB are inherit in sRGB, so, no improvements far the visualization ones are introduced. Furthermore this representation distort colours.

The theoretical advantage is that digital devices adjust their outputs to this standard space. However, there exist strong differences in practice: the same scene can looks quite different depending the acquisition device.

Normalized RGB and chromatic coordinates.

Another representation directly related with RGB, is the normalized RGB (Nrgb) which tries to be a representation independent of intensity by following the next equations:

$$\begin{aligned} r &= \frac{R}{R+G+B} \\ g &= \frac{G}{R+G+B} \\ b &= \frac{B}{R+G+B} \end{aligned} \tag{A.2}$$

Because Red, Green and Blue values have a high correlation, RGB cube have a non-linear and desired behavior under illumination changes. In Nrgb space this effect is not completely removed. On existing literature, is common to assume that this does not happen and talk about Nrgb as a way to achieve a real intensity independent representation. Moreover, Nrgb introduces noise in low intensities due to the non-linear transformation from RGB.

A.1.2 Opponent colour space.

This chromatic representation borns from the observation that some some colour mixtures never appear. An observer can define a colour perception as 'reddish-brown', but never as 'reddish green' or yellowish-blue'. In fact, it seems to be that postreceptor retina cells responds to opponent colour stimulus in spite of a c simply combination of three basic colours. Moreover, some studies appoint that blue-yellow channel is a good choice to fins shadows in a scene, and it would be useful in segmentation tasks.

Not a unique model for opponent representation exists, though all of them share the same idea, i.e., describe an image by means of three channels: Yellow-Blue (YB) channel, Red-Green(RG) channel, and Intensity channel. A simple model is defined as follows:

$$\begin{aligned} RG &= R - G \\ YB &= \frac{2B - R - G}{2} \\ I &= \frac{R+G}{2} \end{aligned} \tag{A.3}$$

It is also common to find in existing literature $YB = 2B - R - G$ and $I = R + G$.

Another common representation for the the opponent colour space is its logarithmic version, and also its chromatic representation (without I channel [10]:

$$\begin{aligned} RG &= log(\frac{RG}{B^2}) \\ YB &= log(B) - \frac{log(R) + log(G}{2} \\ I &= log(G) \end{aligned} \tag{A.4}$$

The opponent space is very interesting when focusin in cromathic differences.

A.1.3 YIQ

This model is used in the NTSC television format in USA, Japan and Central America. The Y component represents the luminance information, and is the only component used by black-and-white television receivers. I and Q represent the chrominance information.

This model is defined with the following conversion from RGB:

$$\begin{pmatrix} Y \\ I \\ Q \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.273 & -0.322 \\ 0.212 & -0.522 & 0.315 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{A.5}$$

This model has been used mainly in works form the USA community though it is not considered an usual space.

A.1.4 Ohta $I_1I_2I_3$ and Karhunen Loeve

This space, directly related with segmentation tasks, was presented in [148]. The authors propose an alternative to Karhunen Loeve transformation (K-L transformation). K-L Transformation extract three colour features, (X_1, X_2 and X_3) by means of an eigensystem analysis of the histogram. More specifically, let $\sum$ be the covariance matrix matrix of the RGB distribution, λ_1, λ_2 and λ_3 with $\lambda_1 \geq \lambda_2 \geq \lambda_3$ be the eigenvalues of $\sum$. Let $W_i = (\omega_{Ri}, \omega_{Gi}, \omega_{Bi})^t$ for $i = 1, 2, 3$ be the eigenvectors of $\sum$ corresponding to λ_1, λ_2 and λ_3. Then X_1, X_2 and X_3 are defined as:

$$X_i = \omega_{Ri}R + \omega_{Gi}G + \omega_{Bi}B \tag{A.6}$$

Analyzing 109 color features in eight color pictures, Ohta et. al. find three effective colour features which can be used instead of X_1, X_2, X_3, i.e., I_1, I_2 and I_3. This three colour features are defined in terms of RGB values:

$$\begin{aligned} I_1 &= \frac{R+G+B}{3} \\ I_1 &= \frac{R-B}{2} \\ I_3 &= \frac{2G-R-B}{4} \end{aligned} \tag{A.7}$$

Even though this space was proposed as the best option for segmentation tasks, there do not exist conclusive works in this sense.

A.1.5 HSI , HLS and HSV.

These colour spaces, tries to represent colour information by a more intuitive way. In fact, these spaces are based on the human perception.

The dominant wavelength of colour is represented by hue (H) component, the purity of color is represented by saturation component (S) and, finally, the darkness or lightness is represented by means of intensity component (I). Let MAX be the maximum of R,G and B, and MIN be the minimum of R,G an B, then to convert from RGB to HSI space:

$$I = \frac{R+G+B}{3}$$

A simplification of this, widely used to do this conversion is presented in equation A.9.

$$\begin{aligned} H &= \arctan\left(\frac{\sqrt{3(G-B)}}{(R-G)+(R-B)}\right) \\ S &= 1-\frac{min(R,G,B)}{I} \\ I &= \frac{R+G+B}{3} \end{aligned} \tag{A.8}$$

If HSI, and HSL are the same colour space, there exist a difference between this two spaces and HSV. Instead of *intensity* (I) or *lightness* L, HSV uses *value* (V) which is defined as follows:

$$V = max(R,G,B) \tag{A.9}$$

A graphical representation of this two colour spaces is showed in figure A.2.

This space is interesting because it assigns at every axis an intuitive feature of colour Its main shortcoming is that one of her axis depends of the angle and it produces a instability at low saturations and, in general, its geometry becomes difficult to apply it in image segmentation.

A.1.6 CMY and CMYK

In contraposition with the spaces presented before, this space is not represented by means of a combination of RGB values. Furthermore, this space have the characteristic to be a subtractive colour space in opposition with RGB space which is an additive space. In an additive space, the colours are achieved by adding colours to the black one. In a subtractive space, we subtract colour to the black one. Hence:

$$(R,G,B) = (1,1,1) - (C,M,Y) \tag{A.10}$$

The difference between CMY and CMYK is that the second spaces adds the black component, K, because is really hard to achieved just with C (cyan), M (magenta) and Y (yellow).

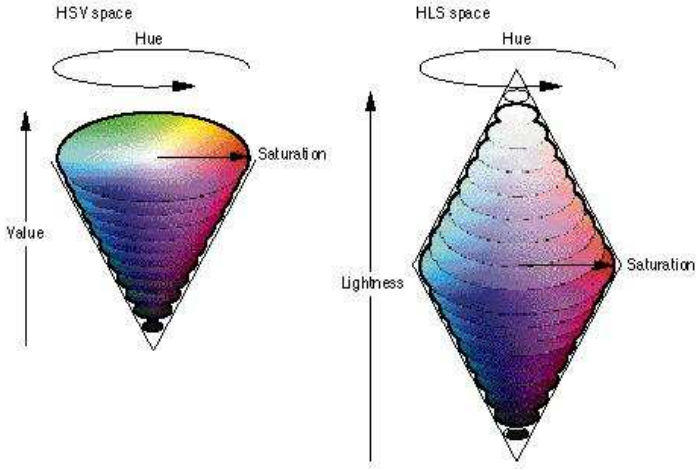


Figure A.2: Graphical representation of HSV and HSI-HLS colour spaces.

This space would be probably the less used in segmentation tasks because is focused in printing tasks.

A.2 Device Independent colour spaces.

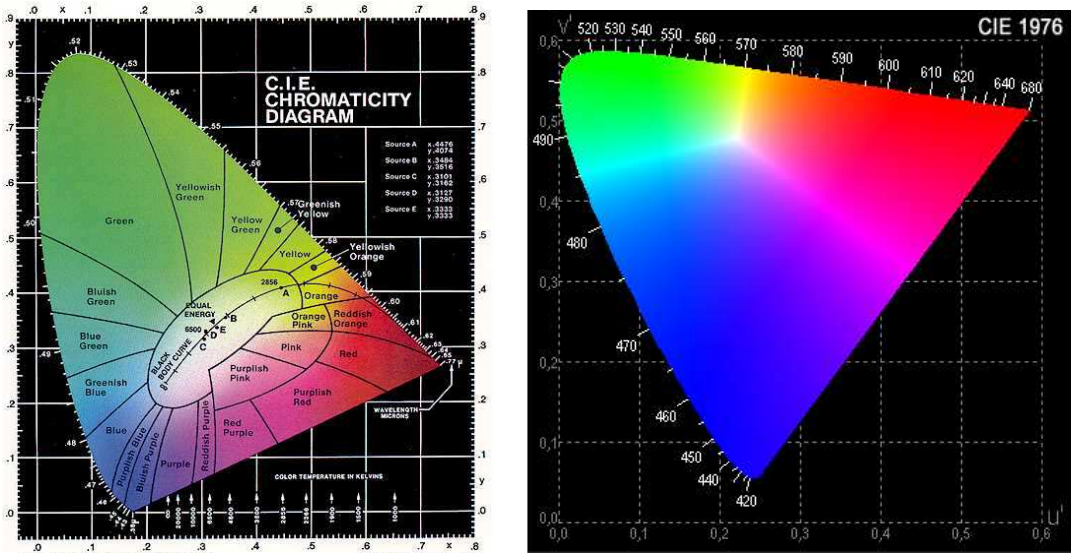
This kind of spaces such as CIE colour spaces, does not depend of the sensor, but the values of an standard observer. These spaces are deeply treated in [57]. Some other device independent have been proposed last years such as ATD or LLAB. In spite of that, in this section we just present the main and widely used device independent colour spaces.

A.2.1 CIE 1931 (XYZ)

The first not-device dependent space presented was the CIE 1931. The idea was to create an standard chromatic space to avoid all drawbacks of the device dependent spaces. To do it, Guild and Wright made some experiments with 10 people to extract these standard primitives. As a result, they define the CIE 1931 (or CIE XYZ) chromatic space as follows:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.49 & 0.31 & 0.2 \\ 0.17697 & 0.81240 & 0.01063 \\ 0.0 & 0.01 & 0.99 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{A.11}$$

A graphical representation of CIE XYZ space is showed in figure A.3.



a) b)

Figure A.3: (a)Graphical representation of CIE 1931 (XYZ) colour space. (b) CIE 1976 ($L^*u^* * v^*$) colour space.

A.2.2 CIE 1976 ($L^*u^*v^*$) and CIE 1976 ($L^*a^*b^*$)

The main problem with that space is that it is not a perceptual uniform colour space. Due that , the proposal of some new spaces appear, such as CIE 1964 ($U^*V^*W^*$) and the widely used spaces CIE 1976 ($L^*u^*v^*$) and CIE 1976 ($L^*a^*b^*$).

CIE 1976 ($L^*u^*v^*$) was proposed to be used as the new device independent and perceptual uniform space as a modification of XYZ coordinates as follows:

$$\begin{aligned} u &= \frac{4X}{X+15Y+3Z} \\ v &= \frac{6Y}{X+15Y+3Z} \end{aligned} \tag{A.12}$$

This first definition had problems of non-.preceptual uniformity in yellowish, orange and reddish areas. Hence, the next correction was proposed:

$$\begin{aligned} u' &= u \\ v' &= \frac{3v}{2} \end{aligned} \tag{A.13}$$

Finally:

$$\begin{aligned} L^* &= \begin{cases} 116\sqrt[3]{Y/Y_n}, & \frac{Y}{Y_n} > 0.008856 \\ 903\sqrt[3]{Y/Y_n}, & \frac{Y}{Y_n} \leq 0.008856 \end{cases} \\ u^* &= 13L*(u' - u'_n) \\ v^* &= 13L*(v' - v'_n) \end{aligned} \tag{A.14}$$

where u'_n and v'_n corresponds to the coordinates of a reference white.

CIE 1976 ($L^*a^*b^*$), basically proposed for industrial ends, is a also a modification of CIE XYZ. This space is defined as follows:

$$\begin{aligned} L^* &= \begin{cases} 116\sqrt[3]{Y/Y_n}, & \frac{Y}{Y_n} > 0.008856 \\ 903\sqrt[3]{Y/Y_n}, & \frac{Y}{Y_n} \leq 0.008856 \end{cases} \\ u^* &= 500(\sqrt[3]{X/X_n} - \sqrt[3]{Y/Y_n}) \\ v^* &= 200(\sqrt[3]{Y/Y_n} - \sqrt[3]{Z/Z_n}) \end{aligned} \tag{A.15}$$

List of Publications

This dissertation has led to the following communications:

Journal Papers

- Eduard Vazquez, Ramon Baldrich, Joost van de Weijer and Maria Vanrell. Describing Reflectances for Colour Segmentation Robust to Shadows, Highlights and Textures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5), 2011.
- Eduard Vazquez, Theo Gevers, M. Lucassen, Joost van de Weijer and Ramon Baldrich. Saliency of Color Image Derivatives: A Comparison between Computational Models and Human Perception. *Journal of the Optical Society of America A*, 27(3):613-621, 2010.

Conference Contributions

- Eduard Vazquez and Ramon Baldrich. Unsupervised evaluation of color image segmentation based on the saliency of the color derivatives. Submitted.
- Naila Murray and Eduard Vazquez. Lacuna Restoration: How to choose a neutral colour?. *Proceedings of CREATE 2010*. Gjovik, Norway.
- Eduard Vazquez and Ramon Baldrich. Non-supervised goodness measure for image segmentation. *Proceedings of CREATE 2010*. Gjovik, Norway.
- Eduard Vazquez, Joost van de Weijer and Ramon Baldrich. Image Segmentation in the Presence of Shadows and Highligts. *10th European Conference on Computer Vision ECCV08*. LNCS 5305:1–14, 2008.
- Partha Pratim Roy, Eduard Vazquez, Josep Llados, Ramon Baldrich and Umapada Pal. A System to Segment Text and Symbols from Color Maps. *Graphics Recognition. Recent Advances and New Opportunities*. LNCS 5046:245–256, 2008.
- Eduard Vazquez and Ramon Baldrich. Colour Image Segmentation in Presence of Shadows. *4th European Conference on Colour in Graphics, Imaging and Vision, CGIV08*. pp.383–387. 2008

- Partha Pratim Roy, Eduard Vazquez, Josep Llados, Ramon Baldrich and Umapada Pal. A System to Retrieve Text/Symbols from Color Maps using Connected Component and Skeleton Analysis. *Seventh IAPR International Workshop on Graphics Recognition GREC 2007*, pp.79–82, 2007.

- Eduard Vazquez, Ramon Baldrich, Javier Vazquez, Maria Vanrell. Topological histogram reduction towards colour segmentation. *3rd Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 2007)*. LNCS 4477:55–62, 2007.

- Javier Vazquez, Maria Vanrell, Anna Salvatella, Eduard Vazquez. *Artificial Intelligence Research and Developmen*. pp.205–212, 2007.

- Eduard Vazquez, Francesc Tous, Ramon Baldrich, Maria Vanrell. n-Dimensional Distribution Reduction Preserving its Structure. *Artificial Intelligence Research and Development*. 146:167–175, 2006

Internal Workshops & Technical Reports

- Eduard Vazquez. Distribution Characterization using Topological Features. Application to Colour Image Processing. *CVC Technical Report* Num. 107, 2009.

- Eduard Vazquez, Theo Gevers, M. Lucassen, Joost van de Weijer and Ramon Baldrich. Psychophysical Comparison of Human Perception and Rarity in the Framework of Saliency. *Proceedings of the Third CVC Internal Workshop on the Progress of Research & Development: Current Challenges in Computer Vision, CVCRD08*, pages 51–55, 2008.

- Eduard Vazquez and Ramon Baldrich. Topological Colour Image Segmentation. *Proceedings of the Second CVC Workshop, Computer Vision: Advances in Research & Development, CVCRD07*, pp.1–6, 2007.

- Eduard Vazquez, Francesc Tous, Ramon Baldrich, Maria Vanrell. Dimensional Distribution Reduction Preserving its Structure. *Proceedings of the First CVC Internal Workshop on the Progress of Research & Development, CVCRD06*, pages 48–53, 2006.

Bibliography

[1] W. Abd-Almageed and L. Davis. Density Estimation Using Mixtures of Mixtures of Gaussians. *9th European Conference on Computer Vision*, 2006.

[2] Alaa E. Abdel-Hakim and Aly A. Farag. Color segmentation using an eigen color representation. *Information Fusion, 2005 8th International Conference on*, 2(2-3):1576–1583, 2005.

[3] J. Abonyi, F. Szeifert, and R. Babuska. Modified gath-geva fuzzy clustering for identification of takagi-sugeno fuzzy models, 2001.

[4] R. Adams and L. Bischof. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6):641–647, 1994.

[5] Shilpa Agarwal, Shweta Madasu, Madasu Hanmandlu, and Shantaram Vasikarla. A comparison of some clustering techniques via color segmentation. In *ITCC '05: Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'05) - Volume II*, pages 147–153, Washington, DC, USA, 2005. IEEE Computer Society.

[6] R.K. Ahuja. *Network flows*. PhD thesis, Massachusetts Institute of Technology, Cambridge, 1993.

[7] R. Bajcsy, S.W. Lee, and A. Leonardis. Color image segmentation with detection of highlights and local illumination induced by inter-reflections. In *Color*, page 204. Jones and Bartlett Publishers, Inc., 1992.

[8] M. Bar, KS Kassam, AS Ghuman, J. Boshyan, AM Schmid, AM Dale, MS Hamalainen, K. Marinkovic, DL Schacter, BR Rosen, et al. Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, 103(2):449, 2006.

[9] K. Barnard, Q. Fan, R. Swaminathan, A. Hoogs, R. Collins, P. Rondot, and J. Kaufhold. Evaluation of localized semantics: data, methodology, and experiments. *International Journal of Computer Vision*, 77(1):199–217, 2008.

[10] Jeff Berens and Graham D. Finlayson. Log-opponent chromaticity coding of color space. *icpr*, 01:1206, 2000.

[11] JC Bezdek and R. Ehrlich. FCM: The fuzzy c-means clustering algorithm. *Comp. Geosci.*, 10(2):191–203, 1984.

[12] Arijit Bishnu, Partha Bhowmick, Sabyasachi Dey, Bhargab B. Bhattacharya, Malay K. Kundu, C. A. Murthy, and Tinku Acharya. Combinatorial classification of pixels for ridge extraction in a gray-scale fingerprint image. In *ICVGIP*, 2002.

[13] A. Blake, M. Isard, et al. *Active contours*. Springer London, 2000.

[14] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr. Interactive image segmentation using an adaptive GMMRF model. *Lecture Notes in Computer Science*, pages 428–441, 2004.

[15] C. Blakemore and FW Campbell. On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *The Journal of Physiology*, 203(1):237, 1969.

[16] D.M. Blei, A.Y. Ng, and M.I. Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022, 2003.

[17] E. Borenstein, E. Sharon, and S. Ullman. Combining top-down and bottom-up segmentation. In *Proceedings IEEE workshop on Perceptual Organization in Computer Vision, CVPR 2004*. Citeseer, 2004.

[18] C.A. Bouman and M. Shapiro. A multiscale random field model for Bayesian image segmentation. *IEEE Transactions on Image Processing*, 3(2):162–177, 1994.

[19] S.T. Bow. *Pattern recognition and image preprocessing*. CRC, 2002.

[20] Y. Boykov and G. Funka-Lea. Graph cuts and efficient nd image segmentation. *International Journal of Computer Vision*, 70(2):109–131, 2006.

[21] X. Bresson, S. Esedoglu, P. Vandergheynst, J.P. Thiran, and S. Osher. Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and Vision*, 28(2):151–167, 2007.

[22] Neil Bruce and John Tsotsos. Saliency based on information maximization. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems 18*, pages 155–162. MIT Press, Cambridge, MA, 2006.

[23] W. Cai, S. Chen, and D. Zhang. Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. *Pattern Recognition*, 40(3):825–838, 2007.

[24] L. Cao and L. Fei-Fei. Spatially coherent latent topic model for concurrent object segmentation and classification. In *Proc. ICCV*, 2007.

[25] J.S. Cardoso and L. Corte-Real. Toward a generic evaluation of image segmentation. *IEEE Transactions on Image Processing*, 14(11):1773–1782, 2005.

[26] T.H. Carr and D. Dagenbach. Semantic priming and repetition priming from masked words: Evidence for a center-surround attentional mechanism in perceptual recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(2):341–350, 1990.

[27] M. Carrasco, S. Ling, and S. Read. Attention alters appearance. *Nature Neuroscience*, 7(3):308–313, 2004.

[28] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International journal of computer vision*, 22(1):61–79, 1997.

[29] J.E. Cates, R.T. Whitaker, and G.M. Jones. Case study: an evaluation of user-assisted hierarchical watershed segmentation. *Medical Image Analysis*, 9(6):566–578, 2005.

[30] S. Chabrier, B. Emile, H. Laurent, C. Rosenberger, and P. Marche. Unsupervised evaluation of image segmentation application to multi-spectral images. *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, 1, 2004.

[31] O. Chapelle, P. Haffner, and V. Vapnik. Support vector machines for histogram-based image classification. *IEEE Trans. on Neural Networks*, 10(5):1055–1064, 1999.

[32] J. Chen, T.N. Pappas, A. Mojsilovic, and B.E. Rogowitz. Adaptive perceptual color-texture image segmentation. *IEEE Transactions on Image Processing*, 14(10):1524–1536, 2005.

[33] PC Chen and T. Pavlidis. Image segmentation as an estimation problem. *Computer Graphics and Image Processing*, 12(2):153–172, 1980.

[34] H.D. Cheng, X.H. Jiang, Y. Sun, and J. Wang. Color image segmentation:advances and prospects. *Pattern Recognition*, 34(6):2259–2281, 2001.

[35] H.D. Cheng, X.H. Jiang, and J. Wang. Color image segmentation based on homogram thresholding and region merging. *Pattern recognition.*, 35(2):373–393, 2002.

[36] S.B. Choi, S.W. Ban, and M. Lee. Biologically motivated visual attention system using bottom-up saliency map and top-down inhibition. *Neural Information Processing-Letters and Review*, 2(1), 2004.

[37] C. Christoudias, B. Georgescu, and P. Meer. Synergism in low level vision. *International Conference on Pattern Recognition*, 4:150–155, 2002.

[38] K.S. Chuang, H.L. Tzeng, S. Chen, J. Wu, and T.J. Chen. Fuzzy c-means clustering with spatial information for image segmentation. *Computerized Medical Imaging and Graphics*, 30(1):9–15, 2006.

[39] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603–619, 2002.

[40] Daniel Cremers, Mikael Rousson, and Rachid Deriche. A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. *Int. J. Comput. Vision*, 72(2):195–215, 2007.

[41] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision, ECCV*, volume 1, page 22. Citeseer, 2004.

[42] F. Cutzu and J.K. Tsotsos. The selective tuning model of attention: psychophysical evidence for a suppressive annulus around an attended item. *Vision Research*, 43(2):205–219, 2003.

[43] D. Dagenbach and T.H. Carr. Inhibitory processes in perceptual recognition: Evidence for a center-surround attentional mechanism. *Inhibitory processes in attention, memory, and language*, pages 327–357, 1994.

[44] X. Y. Dai and J. Maeda. Unsupervised segmentation of natural images. *Optical Review*, 9(5):197–201, 2002.

[45] M. de Brecht and J. Saiki. A neural network implementation of a saliency map model. *Neural Networks*, 19(10):1467–1474, 2006.

[46] Y. Deng and B.S. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(8):800–810, 2001.

[47] KS Deshmukh and GN Shinde. An adaptive color image segmentation. *ELCVIA*, 5(4):12, 2005.

[48] R. Desimone and J. Duncan. Neural mechanisms of selective visual attention. *Annual review of neuroscience*, 18(1):193–222, 1995.

[49] J.C. Devaux, P. Gouton, and F. Truchetet. Aerial colour image segmentation by karhunen-loeve transform. *IEEE International Conference on Pattern Recognition*, pages 309–312.

[50] M. Donoser and H. Bischof. ROI-SEG: Unsupervised Color Segmentation by Combining Differently Focused Sub Results. *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8, 2007.

[51] J. Driver and G.C. Baylis. Attention and visual object segmentation. *The attentive brain*, pages 299–325, 1998.

[52] M. Egmont-Petersen, D. De Ridder, and H. Handels. Image processing with neural networks: a review. *Pattern Recognition*, 35(10):2279–2301, 2002.

[53] L. Elazary and L. Itti. Interesting objects are visually salient. *Journal of Vision*, 8(3):3, 2008.

[54] KK Evans and A. Treisman. Perception of objects in natural scenes: is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance*, 31:1476–1492, 2005.

[55] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2009 (VOC2009) Results. http://www.pascal-network.org/challenges/VOC/voc2009/workshop/index.html.

[56] M. Everingham, A. Zisserman, C. Williams, L. Van Gool, M. Allan, C. Bishop, O. Chapelle, N. Dalal, T. Deselaers, G. Dorko, et al. The 2005 pascal visual object classes challenge. 2006.

[57] M.D. Fairchild. *Color appearance models*. Addison-Wesley Reading, Mass, 1998.

[58] P. Felzenszwalb and D. Huttenlocher. Efficient graph-based image segmentation. *Intl. Journal of Computer Vision*, 59(2), 2004.

[59] M.J. Fenske, E. Aminoff, N. Gronau, and M. Bar. Top-down facilitation of visual object recognition: object-based and context-based contributions. *Visual Perception: Fundamentals of awareness: multi-sensory integration and high-order perception*, page 3, 2006.

[60] J. Folkesson, O.F. Olsen, P. Pettersen, E. Dam, and C. Christiansen. Combining binary classifiers for automatic cartilage segmentation in knee mri. *Lecture Notes in Computer Science*, 3765:230, 2005.

[61] C. Fowlkes, D. Martin, and J. Malik. Learning affinity functions for image segmentation: combining patch-based and gradient-based approaches. *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, 2, 2003.

[62] Jordi Freixenet, Xavier Muoz, D. Raba, Joan Mart, and Xavier Cuf. Yet another survey on image segmentation: Region and boundary information integration. In *ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part III*, pages 408–422, 2002.

[63] G. Fritz, C. Seifert, L. Paletta, and H. Bischof. Attentive Object Detection Using an Information Theoretic Saliency Measure. *Attention And Performance in Computational Vision: Second International Workshop, WAPCV 2004: Prague, Czech Republic, May 15, 2004: Revised Selected Papers*, pages 29–41, 2005.

[64] Keinosuke Fukunaga and Larry D. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *Information Theory, IEEE Transactions on*, 121(1):32–40, 1975.

[65] D. Gao, V. Mahadevan, and N. Vasconcelos. On the plausibility of the discriminant center-surround hypothesis for visual saliency. *Journal of Vision*, 8(7), 2008.

[66] D. Gao and N. Vasconcelos. Discriminant saliency for visual recognition from cluttered scenes. *Advances in neural information processing systems*, 17:481–488, 2005.

[67] D. Gao and J. Zhou. Adaptive background estimation for real-time traffic monitoring. In *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*, pages 330–333. IEEE, 2002.

[68] I. Gath and A. B. Gev. Unsupervised optimal fuzzy clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(7):773–780, 1989.

[69] J. M. Gauch and S. M. Pizer. Multiresolution analysis of ridges and valleys in grey-scale images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(6):635–646, 1993.

[70] F. Ge, S. Wang, and T. Liu. Image-Segmentation Evaluation From the Perspective of Salient Object Extraction. *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Volume 1*, pages 1146–1153, 2006.

[71] F. Ge, S. Wang, and T. Liu. New benchmark for image segmentation evaluation. *Journal of Electronic Imaging*, 16:033011, 2007.

[72] J. M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders. The Amsterdam library of object images. *Int. J. Comput. Vision*, 61(1):103–112, 2005.

[73] S. Ghosal and R. Mehrotra. Range surface characterization and segmentation using neural networks. *Pattern Recognition*, 28(5):711–727, 1995.

[74] B.S. Gibson and Y. Jiang. Surprise! An unexpected color singleton does not capture attention in visual search. *Psychological Science*, 9(3):176, 1998.

[75] C.D. Gilbert and M. Sigman. Brain states: top-down influences in sensory processing. *Neuron*, 54(5):677–696, 2007.

[76] L. Gorelick and R. Basri. Shape Based Detection and Top-Down Delineation Using Image Segments. *International Journal of Computer Vision*, 83(3):211–232, 2009.

[77] L. Grady. Random walks for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1768, 2006.

[78] H. Greenspan, S. Belongie, R. Goodman, P. Perona, S. Rakshit, and CH Anderson. Overcomplete steerable pyramid filters and rotation invariance. *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, pages 222–228, 1994.

[79] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. 2007.

[80] M. Guironnet, N. Guyader, D. Pellerin, P. Ladret, L.I. et des Signaux, and F. Grenoble. Spatio-temporal attention model for video content analysis. In *IEEE International Conference on Image Processing, 2005. ICIP 2005*, volume 3, 2005.

[81] G. D. Guo, S. Yu, and S. D. Ma. Unsupervised segmentation of color images. In *IEEE International conference on Image processing (ICIP'98)*, volume 3.

[82] A. Hanbury and B. Marcotegui. Waterfall segmentation of complex scenes. *Lecture Notes in Computer Science*, 3851:888, 2006.

[83] Jonathan Harel, Christof Koch, and Pietro Perona. Graph-based visual saliency. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 545–552. 2007.

[84] K. Haris, SN Efstratiadis, N. Maglaveras, and AK Katsaggelos. Hybrid image segmentation using watersheds and fast region merging. *Image Processing, IEEE Transactions on*, 7(12):1684–1699, 1998.

[85] X. He, R.S. Zemel, and M.A. Carreira-Perpinan. Multiscale conditional random fields for image labeling. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004.

[86] X. He, R.S. Zemel, and D. Ray. Learning and incorporating top-down cues in image segmentation. *Lecture Notes in Computer Science*, 3951:338, 2006.

[87] G. Heidemann. Color segmentation robust to brightness variations by using B-spline curve modeling. 26(2):211–227, 2008.

[88] T. Hofmann. Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, 42(1):177–196, 2001.

[89] Q. Huang and B. Dom. Quantitative methods of evaluating image segmentation. *IEEE International Conference on Image Processing*, 3:53–56, 1995.

[90] H. Hugli, T. Jost, and N. Ouerhani. Model performance for visual attention in real 3D color scenes. *Artificial Intelligence and Knowledge Engineering Applications: A Bioinspired Approach*, pages 469–478, 2005.

[91] L. Itti. Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, 12(6):1093–1123, 2005.

[92] L. Itti and P. Baldi. Bayesian surprise attracts human attention. *Vision research*, 49(10):1295–1306, 2009.

[93] L. Itti and C. Koch. Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203, 2001.

[94] L. Itti, C. Koch, and E. Niebur. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, pages 1254–1259, 1998.

[95] Ramesh Jain, Rangachar Kasturi, and Brian G. Schunck. *Machine vision*. McGraw-Hill, Inc., New York, NY, USA, 1995.

[96] X. Jiang, C. Marti, C. Irniger, and H. Bunke. Distance measures for image segmentation evaluation. *EURASIP Journal on Applied Signal Processing*, 15, 2006.

[97] M. Jones and T. Poggio. Model-based matching by linear combinations of prototypes. *Massachusetts Institute of Technology, Cambridge, MA*, 1996.

[98] T. Jost, N. Ouerhani, R. Wartburg, R. Müri, and H. Hügli. Assessing the contribution of color in visual attention. *Computer Vision and Image Understanding*, 100(1-2):107–123, 2005.

[99] T. Kadir and M. Brady. Saliency, Scale and Image Description. *International Journal of Computer Vision*, 45(2):83–105, 2001.

[100] T. Kadir, A. Zisserman, and M. Brady. An affine invariant salient region detector. *European Conference on Computer Vision*, 1:228–241, 2004.

[101] D. Karatzas and A. Antonacopoulos. Text extraction from web images based on a split-and-merge segmentation method using colour perception. In *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume*, volume 2, pages 634–637. Citeseer.

[102] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988.

[103] Z. Kato and T.C. Pong. A Markov random field image segmentation model for color textured images. *Image and Vision Computing*, 24(10):1103–1114, 2006.

[104] C. Kim, B.J. You, M.H. Jeong, and H. Kim. Color segmentation robust to brightness variations by using B-spline curve modeling. *Pattern Recognition*, 41(1):22–37, 2008.

[105] D.W. Kim, K.H. Lee, and D. Lee. A novel initialization scheme for the fuzzy c-means algorithm for color clustering. *Pattern Recognition Letters*, 25(2):227–237, 2004.

[106] G.J. Klinker and S.A. Shafer. A physical approach to color image understanding. *Int. Journal of Computer Vision*, 4:7–38, 1990.

[107] E.I. Knudsen. Fundamental components of attention. *Annual Review of Neuroscience*, 30:57–78, 2007.

[108] C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Hum Neurobiol*, 4(4):219–27, 1985.

[109] AR Koene and L. Zhaoping. Feature-specific interactions in salience from combined feature contrasts: Evidence for a bottom-up saliency map in V1. *Journal of Vision*, 7(7):6, 2007.

[110] J. Krummenacher, H.J. Muller, and D. Heller. Visual search for dimensionally redundant pop-out targets: Evidence for parallel-coactive processing of dimensions. *Perception & Psychophysics*, 63(5), 2001.

[111] M.P. Kumar, P.H.S. Torr, and A. Zisserman. OBJCUT: Efficient Segmentation using Top-Down and Bottom-Up Cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009.

[112] K. Kveraga, A.S. Ghuman, and M. Bar. Top-down predictions in the cognitive brain. *Brain and Cognition*, 65(2):145–168, 2007.

[113] Sarang Lakare. 3d segmentation techniques for medical volumes., 2000.

[114] C. Laurent, N. Laurent, M. Maurizot, and T. Dorval. In depth analysis and evaluation of saliency-based color image indexing methods using wavelet salient features. *Multimedia Tools and Applications*, 31(1):73–94, 2006.

[115] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau. A coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5):802–817, 2006.

[116] K.W. Lee, H. Buxton, and J. Feng. Cue-guided search: a computational model of selective attention. *IEEE transactions on neural networks*, 16(4):910–924, 2005.

[117] B. Leibe, A. Leonardis, and B. Schiele. Robust object detection with interleaved categorization and segmentation. *International Journal of Computer Vision*, 77(1):259–289, 2008.

[118] A. Levin and Y. Weiss. Learning to combine bottom-up and top-down segmentation. *International Journal of Computer Vision*, 81(1):105–118, 2009.

[119] N. Li and Y. F. Li. Feature encoding for unsupervised segmentation of color images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 33(3):438–447, 2003.

[120] Z. Li. A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, 6(1):9–16, 2002.

[121] T. Liu, J. Sun, N.N. Zheng, X. Tang, and H.Y. Shum. Learning to Detect A Salient Object. *Proceedings of IEEE Computer Society Conference on Computer and Vision Pattern Recognition (CVPR)*, pages 1–8, 2007.

[122] Antonio M. López, David Lloret, Joan Serrat, and Juan J. Villanueva. Multilocal creaseness based on the level-set extrinsic curvature. *Computer Vision and Image Understanding: CVIU*, 77(2):111–144, February 2000.

[123] Antonio M. López, Felipe Lumbreras, Joan Serrat, and Juan J. Villanueva. Evaluation of methods for ridge and valley detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(4):327–335, 1999.

[124] D.G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the International Conference on Computer Vision.*, volume 2, pages 1150–1157, 1999.

[125] M.P. Lucassen, P. Bijl, and J. Roelofsen. The perception of static colored noise: Detection and masking described by CIE94. *Color Research and Application*, 33(3):178, 2008.

[126] L. Lucchese and S.K. Mitra. Color image segmentation: A state-of-the-art survey. In *INSA-A: Proceedings of the Indian National Science Academy*, pages 207–221, 2001.

[127] Luca Lucchese and Sanjit K. Mitra. Unsupervised low-frequency driven segmentation of color images. In *ICIP (3)*, pages 240–244, 1999.

[128] Y.F. Ma and H.J. Zhang. Contrast-based image attention analysis by using fuzzy growing. In *Proceedings of the eleventh ACM international conference on Multimedia*, page 381. ACM, 2003.

[129] Yi Ma, H. Derksen, Wei Hong, and J. Wright. Segmentation of Multivariate Mixed Data via Lossy Data Coding and Compression. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(9):1546 – 1562, 2007.

[130] L. Macaire, N. Vandenbroucke, and J.G. Postaire. Color image segmentation by analysis of subset connectedness and color homogeneity properties. *Computer Vision and Image Understanding*, 102(1):105–116, 2006.

[131] J. Malik, S. Belongie, T. Leung, and J. Shi. Contour and texture analysis for image segmentation. *International Journal of Computer Vision*, 43(1):7–27, 2001.

[132] M. Mancas, D. Unay, B. Gosselin, and B. Macq. Computational attention for defect localisation. In *Proceedings of ICVS Workshop on Computational Attention & Applications (WCAA'07)*. Citeseer.

[133] Jurij Tasic Marko Tkalcic. Colour spaces - perceptual, historical and applicational background. In Marko Tkalcic Baldomir Zajc, editor, *Eurocon 2003 Proceedings*. IEEE Region 8, September 2003.

[134] A. Martin, H. Laanaya, and A. Arnold-Bos. Evaluation for uncertain image classification and segmentation. *Pattern Recognition*, 39(11):1987–1995, 2006.

[135] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. *Proc. Eighth Int'l Conf. Computer Vision,*, 2:416–423, 2001.

[136] B.A. Maxwell and S.A. Shafer. A framework for segmentation using physical models of image formation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 361–361, 1994.

[137] B.A. Maxwell and S.A. Shafer. Physics-based segmentation of complex objects using multiple hypotheses of image formation. *Computer Vision and Image Understanding*, 65(2):269–295, 1997.

[138] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: a survey. *Medical image analysis*, 1(2):91–108, 1996.

[139] M. Meila. Comparing clusterings. *Proceedings of the Conference on Computational Learning Theory (COLT)*, 2003.

[140] M. Meilă. Comparing clusteringsan information based distance. *Journal of Multivariate Analysis*, 98(5):873–895, 2007.

[141] O. Michailovich, Y. Rathi, and A. Tannenbaum. Image segmentation using active contours driven by the bhattacharyya gradient flow. *IEEE Transactions on Image Processing*, 16(11):2787, 2007.

[142] B. Micusık and A. Hanbury. Automatic image segmentation by positioning a seed. *Proc. European Conference on Computer Vision (ECCV)*, 2006.

[143] E.N. Mortensen and W.A. Barrett. Interactive segmentation with intelligent scissors. *Graphical Models and Image Processing*, 60(5):349–384, 1998.

[144] V. Navalpakkam and L. Itti. Modeling the influence of task on attention. *Vision Research*, 45(2):205–231, 2005.

[145] A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm, 2001.

[146] D.P. Nikolaev and P.P. Nikolayev. Linear color segmentation and its implementation. *Computer Vision and Image Understanding*, 94(1-3):115–139, 2004.

[147] H.C. Nothdurft. Salience from feature contrast: additivity across dimensions. *Vision Research*, 40(10-12):1183–1201, 2000.

[148] Y. Ohta, Takeo Kanade, and T. Sakai. Color information for region segmentation. *Computer Graphics and Image Processing*, 13(3):222 – 241, July 1980.

[149] Slvia D. Olabarriaga and Arnold W. M. Smeulders. Setting the mind for intelligent interactive segmentation: Overview, requirements, and framework. In *IPMI '97: Proceedings of the 15th International Conference on Information Processing in Medical Imaging*, pages 417–422, London, UK, 1997. Springer-Verlag.

[150] Ido Omer and Michael Werman. Color lines: Image specific color representation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 04)*, volume II, pages 946–953. IEEE, June 2004.

[151] C.K. Ong and T. Matsuyama. Robust color segmentation using the dichromatic reflection model. In *International Conference on Pattern Recognition*, volume 14, pages 780–784, 1998.

[152] A. Ortiz and G. Oliver. On the use of the overlapping area matrix for image segmentation evaluation: A survey and new performance measures. *Pattern Recognition Letters*, 27(16):1916–1926, 2006.

[153] N. Ouerhani, R. von Wartburg, H. Hugli, and R. Muri. Empirical validation of the saliency-based model of visual attention. *Electronic Letters on Computer Vision and Image Analysis*, 3(1):13–24, 2004.

[154] N. P. Pal and S. K. Pal. A review on image segmentation techniques. *Pattern Recognition*, 26(9):1277–1294, 1993.

[155] N.R. Pal, K. Pal, J.M. Keller, and J.C. Bezdek. A possibilistic fuzzy c-means clustering algorithm. *IEEE Transactions on Fuzzy Systems*, 13(4):517–530, 2005.

[156] C. Pantofaru and M. Hebert. A comparison of image segmentation algorithms. Technical Report CMU-RI-TR-05-40, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, September 2005.

[157] T. N. Pappas. An adaptive clustering algorithm for image segmentation. *IEEE Transactions on Signal Processing*, pages 901–914, 1992.

[158] DJ Parkhurst and E. Niebur. Scene content selected by active vision. *Spatial Vision*, 16(2):125–154, 2003.

[159] T. Pavlidis and Y.T. Liow. Integrating region growing and edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(3):225–233, 1990.

[160] J.C. Pichel, D.E. Singh, and F.F. Rivera. Image segmentation based on merging of sub-optimal segmentations. *Pattern recognition letters*, 27(10):1105–1116, 2006.

[161] W.M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850, 1971.

[162] S. Ray and R.H. Turi. Determination of number of clusters in k-means clustering and application in colour image segmentation. In *Proceedings of the 4th international conference on advances in pattern recognition and digital techniques*, pages 137–143, 1999.

[163] BC Russell, WT Freeman, AA Efros, J. Sivic, and A. Zisserman. Using multiple segmentations to discover objects and their extent in image collections. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 284–291, 2006.

[164] B.C. Russell, A. Torralba, K.P. Murphy, and W.T. Freeman. LabelMe: A Database and Web-Based Tool for Image Annotation. *International Journal of Computer Vision*, 77(1):157–173, 2008.

[165] M. Sāenz, G.T. Buraas, and G.M. Boynton. Global feature-based attention for motion and color. *Vision Research*, 43(6):629–637, 2003.

[166] P. Schmid. Segmentation of digitized dermatoscopic images by two-dimensional color clustering. *IEEE Trans. on Medical Imaging*, 18(2):164–171, 1999.

[167] B.J. Scholl, Z.W. Pylyshyn, and J. Feldman. What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*, 80(1-2):159–177, 2001.

[168] N. Sebe, Q. Tian, E. Loupias, M. Lew, and T. Huang. Evaluation of salient point techniques. *Image and Video Retrieval*, pages 269–288, 2003.

[169] M. Sezgin and B. Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146–168, 2004.

[170] L. Shafarenko, M. Petrou, and J. Kittler. Automatic watershed segmentation of randomly textured color images. *Image Processing, IEEE Transactions*, 6(11):1530–1544, 1997.

[171] S.A. Shafer. Using color to seperate reflection components. *COLOR research and application*, 10(4):210–218, Winter 1985.

[172] C.W. Shaffrey, I.H. Jermyn, and N.G. Kingsbury. Psychovisual evaluation of image segmentation algorithms. *Proceedings of advanced concepts for intelligent vision systems*, 2002.

[173] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[174] F. Shic and B. Scassellati. A behavioral analysis of computational models of visual attention. *International Journal of Computer Vision*, 73(2):159–177, 2007.

[175] F.Y. Shih and S. Cheng. Automatic seeded region growing for color image segmentation. *Image and Vision Computing*, 23(10):877–886, 2005.

[176] A. Shokoufandeh, I. Marsic, and S.J. Dickinson. View-based object recognition using saliency maps. *Image and Vision Computing*, 17(5-6):445–460, 1999.

[177] J. Shotton, M. Johnson, and R. Cipolla. Semantic texton forests for image categorization and segmentation. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[178] A.K. Sinop and L. Grady. A seeded image segmentation framework unifying graph cuts and random walker which yields a new algorithm. In *IEEE 11th International Conference on Computer Vision*, pages 1–8. Citeseer, 2007.

[179] J. Sivic, B.C. Russell, A. Efros, A. Zisserman, and W.T. Freeman. Discovering object categories in image collections. In *Proc. ICCV*, volume 2, 2005.

[180] Wladyslaw Skarbek and Andreas Koschan. Colour image segmentation — a survey. Technical report, Institute for Technical Informatics, Technical University of Berlin, October 1994.

[181] R.J. Snowden. Visual attention to color: Parvocellular guidance of attentional resources? *Psychological Science*, 13(2):180–184, 2002.

[182] L. Spirkovska. A summary on image segmentation techniques. *NASA Technical Memorandum 104022*, 1993.

[183] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken. Ridge-based vessel segmentation in color images of the retina. *Medical Imaging, IEEE Transactions on*, 23(4):501–509, 2004.

[184] M. Stokes, M. Anderson, S. Chandrasekar, and R. Motta. A standard default color space for the internet-srgb. *Microsoft and Hewlett-Packard Joint Report, Version*, 1, 1996.

[185] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE transactions on systems, man, and cybernetics*, 15(1):116–132, 1985.

[186] K. Takahashi and K. Abe. Color image segmentation using isodata clustering algorithm. *Trans. of the Institute of Electronics, Information and Communication Engineers D-II*, J82D-II(4):751–762, 1999.

[187] B.W. Tatler, R.J. Baddeley, and I.D. Gilchrist. Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, 45(5):643–659, 2005.

[188] J. Theeuwes. Abrupt luminance change pops out; abrupt color change does not. *Perception & psychophysics*, 57(5):637–644, 1995.

[189] E.B. Titchener. *Lectures on the elementary psychology of feeling and attention*. 2005.

[190] O.J. Tobias and R. Seara. Image segmentation by histogram thresholding using fuzzy sets. *IEEE Transactions on Image Processing*, 11(12):1457–1465, 2002.

[191] S. Todorovic and N. Ahuja. Extracting Subimages of an Unknown Category from a Set of Images. In *IEEE Conf. Computer Vision and Pattern Recognition*, volume 1, pages 927–934, 2006.

[192] A. Torralba, A. Oliva, M.S. Castelhano, and J.M. Henderson. Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, 113(4):766–786, 2006.

[193] J.K. Tsotsos, S.M. Culhane, W.Y. Kei Wai, Y. Lai, N. Davis, and F. Nuflo. Modeling visual attention via selective tuning. *Artificial intelligence*, 78(1-2):507–545, 1995.

[194] M. Turatto and G. Galfano. Attentional capture by color without any relevant attentional set. *Perception & Psychophysics*, 63(2):286, 2001.

[195] S. Ullman. Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences*, 11(2):58–64, 2007.

[196] R. Unnikrishnan and M. Hebert. Measures of Similarity. *Seventh IEEE Workshop on Applications of Computer Vision*, pages 394–400, 2005.

[197] R. Unnikrishnan, C. Pantofaru, and M. Hebert. A measure for objective evaluation of image segmentation algorithms. In *Proceedings of the 2005 IEEE Conference on Computer Vision and Pattern Recognition (CVPR05), Workshop on Empirical Evaluation Methods in Computer Vision*, volume 3, pages 34–41, 2005.

[198] J. van de Weijer, T. Gevers, and A.D. Bagdanov. Boosting Color Saliency in Image Feature Detection. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 28(1):150–156, 2006.

[199] E. Vazquez, T. Gevers, M. Lucassen, J. van de Weijer, and R. Baldrich. Saliency of color image derivatives: a comparison between computational models and human perception. *Journal of the Optical Society of America A*, 27(3):613–621, 2010.

[200] E. Vazquez, J. van de Weijer, and R. Baldrich. Image Segmentation in the Presence of Shadows and Highlights. volume 5305, pages 1–14. Springer, 2008.

[201] J. Vazquez, A. Salvatella, E. Vazquez, and M. Vanrell. A colour space based on the image content. In *Artificial Intelligence Research and Development*, volume 163, pages 205–212. October 2007.

[202] S.P. Vecera and M.J. Farah. Is visual image segmentation a bottom-up or an interactive process?. *Perception & Psychophysics*, 59(8):1280–1296, 1997.

[203] A. Verikas, K. Malmqvist, and L. Bergman. Colour image segmentation by modular neural network. *Pattern Recognition Letters*, 18(2):173–185, 1997.

[204] D. Verma and M. Meila. A comparison of spectral clustering algorithmstechnical report uw-cse-03-05-01, university of washington.

[205] S. Vicente, V. Kolmogorov, and C. Rother. Graph cut based image segmentation with connectivity priors. In *CVPR, June*, volume 8, 2008.

[206] L. Vincent and P. Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):583–598, 1991.

[207] D. Walther and C. Koch. Modeling attention to salient proto-objects. *Neural Networks*, 19(9):1395–1407, 2006.

[208] L. Wang and T. Pavlidis. Direct gray-scale extraction of features for character recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(10):1053–1067, 1993.

[209] S. Wang, F. Chung, and F. Xiong. A novel image thresholding method based on Parzen window estimate. *Pattern Recognition*, 41(1):117–129, 2008.

[210] S. Wang and JM Siskind. Image segmentation with ratio cut. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(6):675–690, 2003.

[211] F.A. Wichmann, L.T. Sharpe, and K.R. Gegenfurtner. The Contributions of Color to Recognition Memory for Natural Scenes. *Learning, Memory*, 28(3):509–520, 2002.

[212] J. Winn and N. Jojic. Locus: Learning object classes with unsupervised segmentation. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 756–763, 2005.

[213] J.M. Wolfe. Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2):202–238, 1994.

[214] J.M. Wolfe. Guided Search 4.0: Current Progress with a model of visual search. *Integrated models of cognitive systems*, 25:9479–9487, 2007.

[215] J.M. Wolfe, K.R. Cave, and S.L. Franzel. Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human perception and performance*, 15(3):419–433, 1989.

[216] JM Wolfe and TS Horowitz. What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscince*, 5(6):495–501, 2004.

[217] Z. Wu and R. Leahy. An optimal graph theoretic approach to data clustering: theory andits application to image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(11):635–646, 1993.

[218] G. Wyszecki and W.S. Stiles. Color science: concepts and methods, quantitative data and formulae. 2000.

[219] C. Xu and J.L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on image processing*, 7(3):359–369, 1998.

[220] N. Xu, N. Ahuja, and R. Bansal. Object segmentation using graph cuts based active contours. *Computer Vision and Image Understanding*, 107(3):210–224, 2007.

[221] Y. Yang, J. Wright, S. Sastry, and Yi. Ma. Unsupervised segmentation of natural images via lossy data compression., 2007.

[222] B. Yao, X. Yang, S.C. Zhu, E.Z. City, H.B. Province, and PR China. Introduction to a large-scale general purpose ground truth database: methodology, annotation tool and benchmarks. In *Energy Minimization Methods in Computer Vision and Pattern Recognition: 6th International Conference, EMMCVPR 2007, Ezhou, China, August 27-29, 2007, Proceedings*, page 169. Springer-Verlag New York Inc, 2007.

[223] X. Yuan, D. Goldman, A. Moghaddamzadeh, and N. Bourbakis. Segmentation of Colour Images with Highlights and Shadows sing Fuzzy-like Reasoning. *Pattern Analysis & Applications*, 4(4):272–282, 2001.

[224] H. Zhang, J.E. Fritts, and S.A. Goldman. Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding*, 110(2):260–280, 2008.

[225] YJ Zhang. A survey on evaluation methods for image segmentation. *Pattern recognition*, 29(8):1335–1346, 1996.

[226] Y.J. Zhang. A review of recent evaluation methods for image segmentation. In *Sixth International Symposium on Signal Processing and its Applications*, volume 1, pages 148–151, 2001.

[227] L. Zhaoping. The primary visual cortex creates a bottom-up saliency map. *Neurobiology of attention*, pages 570–575, 2005.

Final Acknowledgment

This work has been partially supported by the CICYT projects TIN2006-15694-C02-02 and DPI2007-614452 (Ministerio Ciencia y Tecnología) and by the Spanish research programme Consolider Ingenio 2010: MIPRCV (CSD2007-00018).