Shape and Topology Constrained Image Segmentation with Stochastic Models.

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Thomas Zöller

aus

Wissen

Bonn, 2005

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn.
Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn
`http://hss.ulb.uni-bonn.de/diss_online` elektronisch publiziert.

1. Referent: Univ.-Prof. Dr. Joachim M. Buhmann
2. Referent: Univ.-Prof. Dr. Armin B. Cremers

Tag der Promotion: 05. 01. 2005

Acknowledgements

This thesis draws upon the advice, constructive criticism and support by so many people, that it would be impossible to mention everyone in this acknowledgment. Nevertheless I would like to express my gratitude to all of them.

Very special thanks go to Prof. Joachim M. Buhmann. First of all, I am deeply indebted to him for the thorough foundation in statistical pattern recognition and image processing / computer vision that I received from his lectures. Moreover, he created a unique research atmosphere in his very lively and active group. His scientific advice and the engaged discussions helped immensely to shape the ideas that finally led to this thesis. Additionally, he generously provided funding for the participation in international conferences and workshops.

Furthermore, I would like to thank Prof. Armin B. Cremers for the evaluation of this thesis. Apart from this, I greatly enjoyed the time at his well organized department. The competent support team and secretarial staff contributed a lot to this positive experience.

I really appreciated the lively discussions with Prof. Anil K. Jain during his visit as a Humbold fellow at our research group, from which I immensely benefitted. Thank you !

Over the years, my work has been funded by the DFG under grant # BU 914/3-1 (DFG-Schwerpunktprogramm Echtzeitoptimierung großer Systeme), by Infineon Technologies and by Infoterra GmbH. I am very grateful for that.

The Computer Vision group at the University of Berkeley, California under Prof. Jitendra Malik has put a considerable amount of time and effort in the compilation an publication of a common standard test-set for edge-detection and image segmentation algorithms. Without this test-bed, the quantitative performance assessment of the PDC segmentation method would not have been possible.

My fellow PhD students Bernd Fischer and Peter Orbanz took the burden of proof-reading large parts of this thesis. I am deeply indebted to them. Also, many thanks go to Dr. Lothar Hermes, with whom I closely collaborated especially during the development of the basic PDC approach.

Last, but definitely not least, I would like to thank my wife Simone for her patience, understanding and love. Without her support, this thesis would not have been written.

to my parents

Contents

List of Figures

List of Tables

Symbols and Abbreviations

Abbreviations

fig.	figure
eq.	equation
tab.	table
SA	Simulated Annealing
DA	Deterministic Annealing
MFA	Mean Field Annealing
ICM	Iterated Conditional Mode
RV	random variable

General Symbols

$\mathbb{N}$	natural numbers
$\mathbb{R}$	real numbers
$\mathbb{D}$	data domain
$\|\cdot\|$	L_2–Norm of $\cdot$
$\|\cdot\|$	L_1–Norm of $\cdot$ / set cardinality
$[a, b]$	closed interval of real numbers between a and b
(a, b)	open interval of real numbers between a and b
$\operatorname{argmax} f(\cdot)$	argument maximizing $f(\cdot)$
$\operatorname{argmin} f(\cdot)$	argument minimizing $f(\cdot)$
sgn	sign function
D_{KL}	Kullback–Leibler divergence
$\delta(\cdot, \cdot)$	Kronecker delta function
$\bigotimes$	Cartesian product

Probability Theory and Statistics

E	expectation value
$\hat{\mathrm{E}}$	empirical expectation value (mean)
D^2	variance
D	standard deviation
$\hat{\mathrm{D}}$	empirical standard deviation
P	probability
$\mathcal{L}$	likelihood function

Features and Image Representation

$\mathcal{I}$	input image
$\mathbf{s}$	image site
$\mathcal{S}$	set of all image sites
n	number of sites on the image grid
x	individual observation
r	number of observations per site
$\mathcal{X}$	set of observations
d	dimension of feature vectors
$\mathbf{n}$	local feature histogram
m	number of histogram bins
$\mathbf{f}$	spatial frequency
$\mathcal{G}$	Gabor wavelet

Clustering

c	cluster c
k	number of clusters
$\mathbf{m}$	cluster assignment mapping function
$\mathcal{M}$	combinatorial space of all mapping functions
Θ	continuous model parameters
T	topological coupling matrix
$\mathcal{N}_{\mathbf{s}}$	neighborhood system for site $\mathbf{s}$
λ^{t}	weighting factor for spatial coupling

Performance Evaluation

π	precision
ρ	recall
F	F-measure (harmonic mean of precision and recall)
REL	ground-truth boundary pixels
RET	inferred boundary pixels
$RETREL$	retrieved and relevant boundary pixels

Spatialy Constrained Clustering

Ω	sigma algebra of semantic categories
ω_f	foreground category
ω_b	background category
$\mathcal{A}$	set of aspect distributions
$\mathcal{A}_f$	set of foreground aspect distributions
$\mathcal{A}_b$	set of background aspect distributions
f	individual foreground aspect distribution
b	individual background aspect distribution
a	generic aspect model
h_f	local cost of foreground object
h_b	local cost of background
p^{ω_f}	foreground likelihood
p^{ω_b}	background likelihood
P_S	shape prior
P_{ω_f}	posterior probability of foreground object

Opimization

$\mathcal{F}$	generalized free energy
$\mathcal{H}$	cost function
H	entropy
T	computational temperature
$\mathcal{Z}$	partition function
I	coarse scale mapping function
q	expected value of an assignment
h	local costs / mean fields

Chapter 1

Introduction

Enabling the automated interpretation of visual scenes, which comprises the problem of machine vision, has been in the focus of artificial intelligence research for a long time. In his influential book *Vision*[Mar82], Marr identified three main stages in visual scene interpretation: 1) the primal sketch (edges, segments); 2) the $2\frac{1}{2}$d sketch (which makes explicit assumptions about surface orientations etc.); and 3) the full 3d sketch, containing information about the full three-dimensional structure of a scene. A human subject, as well as any higher animal, effortlessly performs this processing chain in a fraction of a second. Even if one does not subscribe to Marr's paradigm, one has to concede, that the problem of reliable interpretation of visual scenes is solved with remarkable performance by all higher organisms. Otherwise, the mere struggle for existence, that Darwin proposed as the driving force of biological evolution, would have eliminated those species, that fail to extract the relevant information concerning their natural environment.

Taking the immense computational power of modern computers into account, it seemed a trivial task to achieve and even to improve the performance of biological systems. This assumption led Minsky to frame the design of a general purpose computer vision system as a summer-project for one of his students in the late fifties of the previous century. Not only did Minsky's student fail to do his homework, the posed problem still remains largely unsolved even today after several decades of research by a growing community of computer-scientists, applied mathematicians, electrical engineers and researchers from related fields.

Clearly, substantial progress has been achieved, nevertheless, the majority of approaches capitalize on some narrow sub-domain or specific instances of the general computer vision problem. Even for arguably the most basic and fundamental processing step of *image segmentation*, on which image understanding and object recognition crucially rely upon for the provision of an intermediate representation of image content, most of the suggested algorithms lack the potential to extend them beyond a narrow scope of application. Examples of this phenomenon include thresholding techniques, which, despite having a long standing tradition in color image segmentation [SSWC88], completely fail on textured images. Moreover, split-and-merge techniques [HS85], still being widely applied in standard image segmentation algorithms, suffer from their inherent greedy strategy and

are not able to support non-local a priori information.

Apart from a principled interest in the nature of the vision process, the growing amount of pictorial material that we are faced with today demands for automated processing techniques. For example, large images data-bases, available today, together with the increasing ability to acquire images in digital form can only be coped with, when reliable techniques for content-based retrieval exist. Image processing techniques, in particular segmentation approaches, are needed in order to arrive at intermediate representations of image content, on which a meaningful retrieval strategy can operate.

Approaches to image segmentation which lack supervision information in the form of ground-truth labels are often formulated as data clustering problems [JD88]. Regardless of the particular nature of the image primitives in question, these methods share as a common trait that they search for a partition of pixels or pixel blocks with a high degree of homogeneity. The specific choice of a clustering algorithm, however, depends on the nature of the given image primitives which might be feature vectors, feature relations or feature histograms. In this thesis, I advocate to characterize an image site by the empirical color distributions extracted from its neighborhood, which is regarded as a robust and statistically reliable descriptor of local feature properties.

One way to design a clustering technique for this type of data is to apply a statistical test to the measured histograms. This processing step yields pairwise dissimilarity values, for which a multitude of grouping techniques can be found in the literature (e.g.[JD88, PHB00, SM00]). Alternatively, feature histograms can be grouped directly by histogram clustering [PTL93, PHB99]. This approach characterizes each cluster by a prototypical feature histogram, and it assigns local histograms measured at image sites to the nearest prototype. Closeness is measured by the *Kullback–Leibler–Divergence*. As a consequence, this method retains the efficiency of central clustering approaches like k-means clustering, but it avoids the restrictive assumption that features are vectors in a Euclidean space.

Discrete representations of feature distributions are often inadequate for the underlying type of measured data. Color, although discretized in standard image file formats is a continuous phenomenon. Also, the filter responses proposed in this contribution for the measurement of texture characteristics are real-valued entities. Moreover, histogram clustering in its original form is invariant to permutations of histogram bins. In computer vision where the histogramming process is prone to noise induced errors, this invariance neglects information about the order of bins and the distance of bin centers in feature space. As an example, consider two image sites, which differ only in one particular feature, in the sense that the corresponding values are assorted in neighboring, albeit different, bins by the histogramming process. For the associated cost function, the entries for the affected bins are maximally dissimilar. Consequently, the clustering approach does not have an incentive to attribute the corresponding sites to the same cluster. It could equally well merge two sites having a comparable difference of counts in

completely unrelated bins of the pertaining histograms, without incurring higher cost.

It is, therefore, suggested to replace the non-parametric density estimation via histograms by a continuous mixture model, which no longer suffers from the shortcomings. This constitutes the key modeling idea behind the Parametric Distributional Clustering (PDC) model, which is discussed in chapter 2. Despite being a well founded clustering technique, PDC also defines a generative model of image features. Moreover, the clustering cost function, which is derived in a strict maximum likelihood framework, can also be shown to achieve the minimum KL-divergence between the site-specific histograms and the associated continuous model distributions of the inferred clusters.

In this setting, the search for a good grouping solution is posed as a mixed combinatorial optimization problem. Due to the fact that the cost landscape may have a very jagged structure, powerful optimization techniques with regularization or smoothing behavior should be applied to avoid poor local minima. Consequently, deterministic annealing (DA) techniques are utilized, which have been shown to achieve superior solution quality [Hof97] compared to other optimization techniques. In order to further increase the computational efficiency, multi-scale techniques [Puz99] are brought to bear in order to speed up the corresponding computations. The validity of the achieved segmentation results is exemplified in a qualitative fashion by the provision of representative image partitions. Moreover, a thorough quantitative performance evaluation in terms of the quality measures precision, recall and F-value, well-known from the field of information retrieval, has been conducted.

In chapter 3, the basic PDC approach is augmented by topological constraints. PDC itself is a generic clustering procedure for histogram data. As such, it does not make any assumptions about the structure of the space, from which the grouped entities emerge. In image segmentation, however, one deals with image sites. In this domain, there clearly exists a topological structure in the sense that neighboring sites have a high probability of belonging to the same segment, unless artificially created images are considered, which are specifically designed to violate that property. Consequently the introduction of spatial topological constraints in the cost-function is a viable means of improving the quality of image partitions. This extension is the subject of section 3.1. As has already been stated, one of the main goals in the design of the PDC cost function was to overcome the permutation invariance of histogram bins. Apart from this, another invariance still persists, which affects the cluster labels. Exchanging the labeling of data-groups in the PDC approach does not affect the clustering solution. In order to address this second permutation invariance, topological constraints in the cluster-domain are introduced in section 3.2. This approach is inspired by Kohonen's self-organizing maps [Koh95], which have received considerable interest in the neural-computation community. Finally, section 3.3 takes the next logical step by demonstrating the joint application of both kinds of topological

constraints in the PDC framework.

All proposed PDC variants share as a common trait, that they are clustering procedures, and as such unsupervised learning approaches. For any learning algorithm, the problems of robustness towards small fluctuations in the data as well as the generalizability of inferred solutions to previously unseen instances of data-sets from the covered domain are highly relevant [Vap98]. For PDC, these issues are addressed in an empirical manner. Section 4.1 puts a focus on robustness. To this end, a variant of bootstrap sampling for image data is introduced. Apart from demonstrating the robustness of PDC-based image partitions, the bootstrap approach provides for a viable means of finding the most pronounced, and thus most important, boundaries between image regions. Section 4.2 examines the transferability of segmentation solutions between images of similar content. Here, a PDC model is inferred on the basis of a single image. Then, this model is applied without modification to another picture of comparable content, producing a segmentation of this second image.

Image segmentation solely based on low-level features clearly has its limitations. There is strong evidence, that a substantial performance increase can only be achieved by incorporating higher-level knowledge. For example, it is a well-known finding from the field of neuro-anatomy, that there is a considerable amount of neurons projecting from higher-level brain regions back to the visual cortex V1. In order to make a first step to bridge this apparent gap between low-level bottom-up approaches and higher-level knowledge about probable image configurations, the incorporation of shape knowledge in the segmentation process is addressed in chapter 5. Here, a probabilistic representation of the shape of objects from a certain semantic category is derived as a first step. Then, corresponding likelihood maps for the low-level features are established by developing the notion of aspect-sets. These two sources of information about the presence of objects from a given category of interest are fused in a Bayesian setting in order to arrive at image segmentations that faithfully capture the semantic content of the depicted scene.

Chapter 2

The Parametric Distributional Clustering Model

The objective of this chapter is to introduce the basic Parametric Distributional Clustering (PDC) model. The exposition is structured as follows: At first, the data acquisition process is examined in section 2.1. Then, the generative model for the observed feature values is introduced in section 2.2. With these prerequisites at hand, one can derive the complete data log-likelihood in section 2.3, which serves as the cost function for our clustering model. An interesting alternative interpretation of the PDC cost function is then introduced in section 2.4. The optimization of this objective function including the derivation of the E- and M-step equations is the subject of section 2.5. Moreover, closedness of the cost function under the operation of multi-scale operators is demonstrated in section 2.6, which is a mandatory pre-condition for the application of multi-scale optimization schemes. Finally, section 2.7 contains a discussion on implementation questions together with a comprehensive overview on segmentation results including a quantitative performance evaluation.

2.1 The Data Acquisition Process

Despite the fact that the PDC clustering model can be applied to any kind of histogram data, the focus will be put on the application domain of image processing.

While color features for all pixels are available in any standard image format in the form of RGB triplets, the situation concerning texture cues is not so simple. Despite the intuitive clarity of the notion of texture, no precise definition is universally agreed upon to date. Clearly, texture is not a phenomenon that is localized point-wise, as is the case with color measurements. It is inherently spatial in nature. Texture can be described as any form of regular or irregular pattern. Thus, the notion of texture includes all kinds of relatively small-scale non-uniformity present in images. For the experiments concerning segmentation

based on texture features, it is proposed to employ Gabor filters as derived in [MM96]. This family of filters consists in principle of complex sinusoids modulated by Gaussian functions. The real-part of a Gabor function can be found as an example in figure 2.1. Due to the construction of the filter function, Gabor

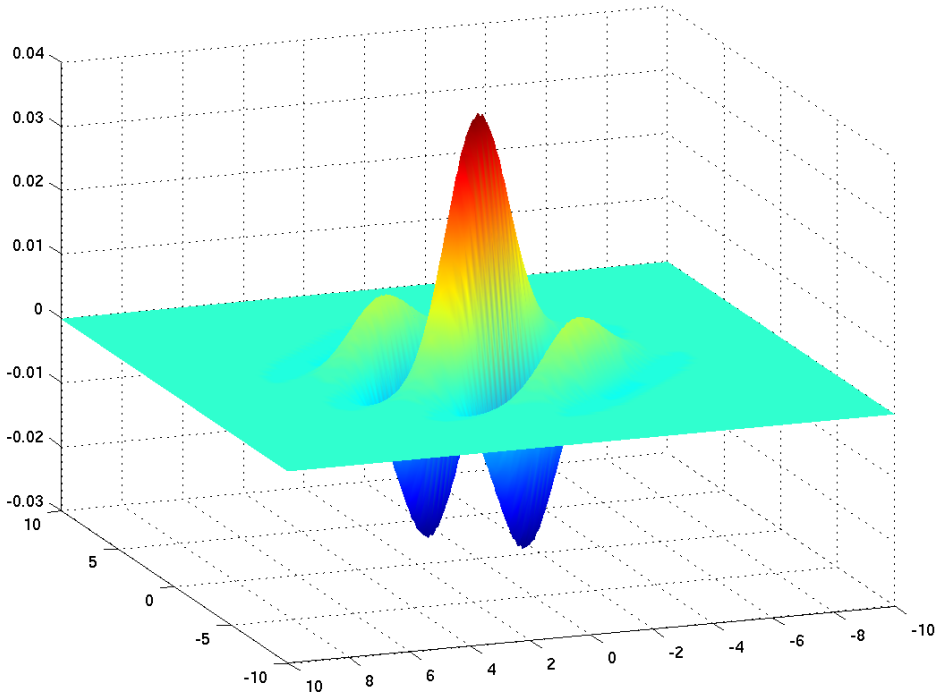


Figure 2.1: The real-part of an example Gabor-function.

filters provide responses that are spatially localized by virtue of the Gaussian modulation, while being tuned to a particular spatial frequency $\mathbf{f}$. The frequency tuning operation [tHR94] can be formalized in the context of scale-space theory [FtHRKV92], leading to the following functional form of the Gabor-filters at a particular location $\mathbf{x}$ in the image plane:

$$\mathcal{G}_{\sigma,\mathbf{f}}(\mathbf{x}) = \frac{1}{2\pi\sigma^2}\exp(-\mathbf{x}^t\mathbf{x}/2\sigma^2)\exp(2\pi i\mathbf{f}^t\mathbf{x}), \quad \mathbf{x}\in\mathbb{R}^2. \tag{2.1}$$

In the equation above, σ denotes the localization parameter, which is typically chosen proportional to the wavelength $\frac{1}{||\mathbf{f}||}$. It is impossible, to simultaneously localize a signal in the spatial and the frequency domain due to the fundamental uncertainty relation. It can, however, be shown, that Gabor filters are optimal in the sense that they achieve the lower bound for simultaneous localization [Dau85]. Moreover, they exhibit very good discrimination properties for a broad variety of

textures [BCG90, FS89, HPB98, JF91, MM96]. Psycho-physical and electrophysiological experiments suggest, that the human visual system also employs such a filtering method [FS89]. In accord with the majority of the pertaining literature, which puts a strong emphasis on response strengths, the phase information is discarded, retaining only the magnitudes of the filter-responses at any given image pixel as our texture features. Due to the scale and orientation selectivity of the Gabor functions, a bank of filters spanning two octaves in scale while using four orientations at each scale is employed. Thus, one arrives at an eight-dimensional texture feature vector for each image pixel.

The next question which now has to be answered is how to reliably represent local image content. Point measurements at the various pixel locations in the image suffer from noise, which is invariably present in all image data. In order to counter this problem, it is proposed to use histograms of feature measurements, which provide a more robust description of the prevailing local feature characteristics. These histograms are gathered at image locations, so-called sites $\mathbf{s}_\cdot$, which are arranged on a regular grid superimposed on the image. The set of all sites is denoted by $\mathcal{S} = \{\mathbf{s}_i : i = 1, \ldots, n\}$. Each site has a local neighborhood of pixels which are taken into account when computing the histogram for that particular site. The feature values at each pixel could in principle be arbitrary elements of the data domain $\mathbb{D} \subset \mathbb{R}^d$, where d denotes the dimensionality of the underlying feature-space. However all current image formats are based on some kind of quantization of this continuous domain in order to arrive at a discrete representation for the given image at hand. This corresponds to the introduction of coherent intervals of feature values $I \subset \mathbb{D}$, the so-called bins. When this quantization is done uniformly, the resolution solely depends on the number of bins m. An elementary observation x is then given by the measurement of a feature value inside the corresponding bin for the current image site. The set of all observations belonging to site $\mathbf{s}$ is denoted by $\mathcal{X}_\mathbf{s} = \{x_{\mathbf{s}1}, \ldots, x_{\mathbf{s}r}\}$, where observation sets from different sites are considered to be statistically independent. The limiting case of bin-size converging to zero is treated in [Her03].

Consequently, the parameters governing the data acquisition process are the number of sites on the grid, the size of the local neighborhood for each site and the number of bins covering the data domain. When these values are fixed, the histogram at site $\mathbf{s}$ is given by the number of occurrences for each elementary observation in the neighborhood of that site, denoted by $\mathbf{n}_\mathbf{s}$. The counts for bin j are then referenced as $\mathbf{n}_{\mathbf{s}j}$.

This process is visualized in fig. 2.2. Here, an example grey-level image is visualized together with a very coarse grid of sites, which are depicted as dots on the cross-points of the grid. For one particular site in the lower part of the image, the local neighborhood is shown in magnification in the lower left of the figure. The corresponding grey-level histogram for that image region is depicted on the lower right. These histograms form the input to the PDC clustering model, which is described in the following sections.

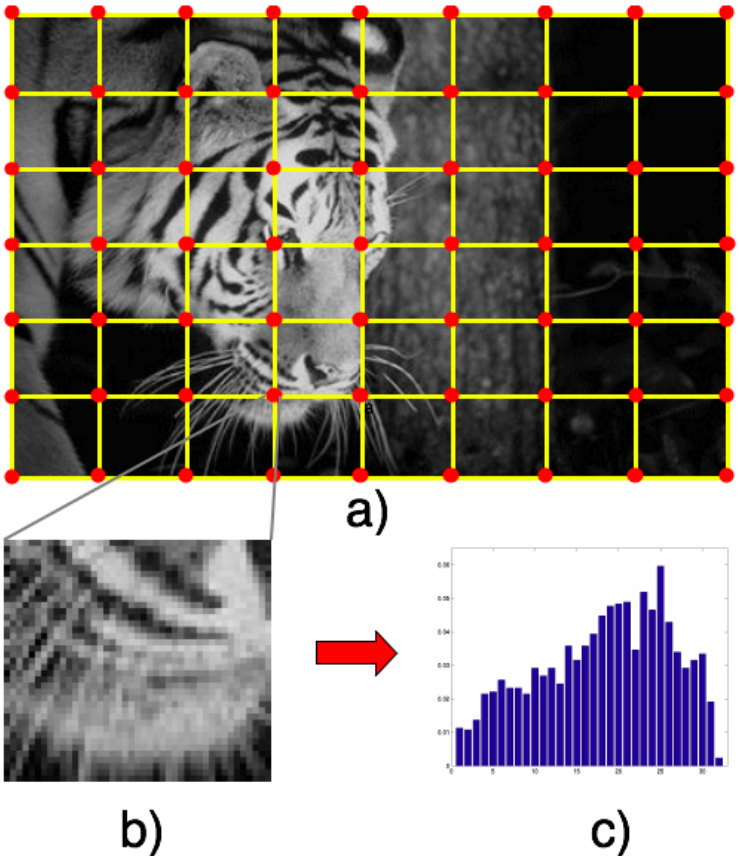


Figure 2.2: Illustration of the data acquisition process: a) image with overlaid grid, b) close-up of image data in the vicinity of a site, c) corresponding local histogram.

2.2 The Generative Model

One of the main advantages of generative models is given by the fact that they treat data-sets as the result of a stochastic sampling process. Therefore, such a model exhibits a high degree of descriptive power. Unlike many other approaches for image segmentation or clustering, the PDC model is based on a generative model for the observed data.

The basic assumption is that the given image, or more generally speaking, the dataset at hand, can be described by a fixed number of k different data sources, the so-called clusters, which in turn are characterized by Gaussian mixture distributions. Each of the n image sites is associated with one of these clusters by means of a mapping function $\mathbf{m} : \mathcal{S} \to \{1, \ldots, k\}$, which can also be regarded as an element in the combinatorial optimization space $\mathcal{M}_{n,k}$, where

$$\mathcal{M}_{n,k} = \bigotimes_{s \in \mathcal{S}} \{1, \ldots, k\} = \{1, \ldots, k\}^n \subset \mathbb{N}^n. \tag{2.2}$$

The explicit dependency on the indices n and k is dropped due to reasons of notational convenience, whenever it is clear from the context. The term $\mathbf{m}(\mathbf{s}) = c$ denotes that site $\mathbf{s}$ is mapped to cluster c.

Each site is equipped with a set of observations. The distribution of these feature values is considered to be independent from the site given its group membership assignment, i.e. the local feature measurements solely depend on the characteristics of the associated data-group in the clustering model and not on other properties of the site, e.g. its spatial position. Therefore, the measured observation histogram at a given grid position depends only on the cluster-specific distribution of features, which is assumed to be a Gaussian mixture model. Thus, the generative model for an individual observation $x_{\mathbf{s}} \in \mathcal{X}_{\mathbf{s}}$ given the group membership $\mathbf{m}(\mathbf{s})$ of its associated site $\mathbf{s}$ is defined as

$$p(x_{\mathbf{s}} |\, \mathbf{m}(\mathbf{s})) = \sum_{\alpha=1}^{l} p_{\alpha |\, \mathbf{m}(\mathbf{s})} g_\alpha(x). \tag{2.3}$$

Here, $g_\alpha(x) = g(x |\, \mu_\alpha, \mathbf{\Sigma}_\alpha)$ denotes a multivariate Gaussian distribution with mean μ_α and covariance matrix $\mathbf{\Sigma}_\alpha$. Although each of the cluster could in principal be equipped with its unique set of Gaussians, the Gaussians are restricted to form one common *alphabet* $\{g_\alpha : \alpha = 1, \ldots, l\}$ for all data groups in order to infer parsimonious models. The cluster specific distributions are then synthesized by a particular choice of mixture coefficients $p_{\alpha | c}$. As has already been stated in section 2.1, the observations are assumed to be discretized, which discards differences between observations falling into the same bin. Consequently, the probability of observing a feature value inside bin I_j is given by

$$\begin{aligned} p(j |\, \mathbf{m}(\mathbf{s})) &= \sum_{\alpha=1}^{l} p_{\alpha |\, \mathbf{m}(\mathbf{s})} G_\alpha(j), \quad \text{where} \\ G_\alpha(j) &= \int_{I_j} g(x |\, \mu_\alpha, \mathbf{\Sigma}_\alpha) dx. \end{aligned} \tag{2.4}$$

In terms of the PDC model, image formation can be regarded as a two stage process: In the first stage, the various sites, which in total comprise the image, are assigned to clusters according to a probability distribution $p(c)$. Once this assignment is completed, the mapping function $\mathbf{m}(\cdot)$ is fixed. The second stage consists of the sampling of feature values for the mapped sites according to the characteristic distributions of the clusters. Due to the fact that the feature distributions are statistically independent of the sites given their group assignment, this image formation process is Markovian. The corresponding graphical model is visualized in fig. 2.3.

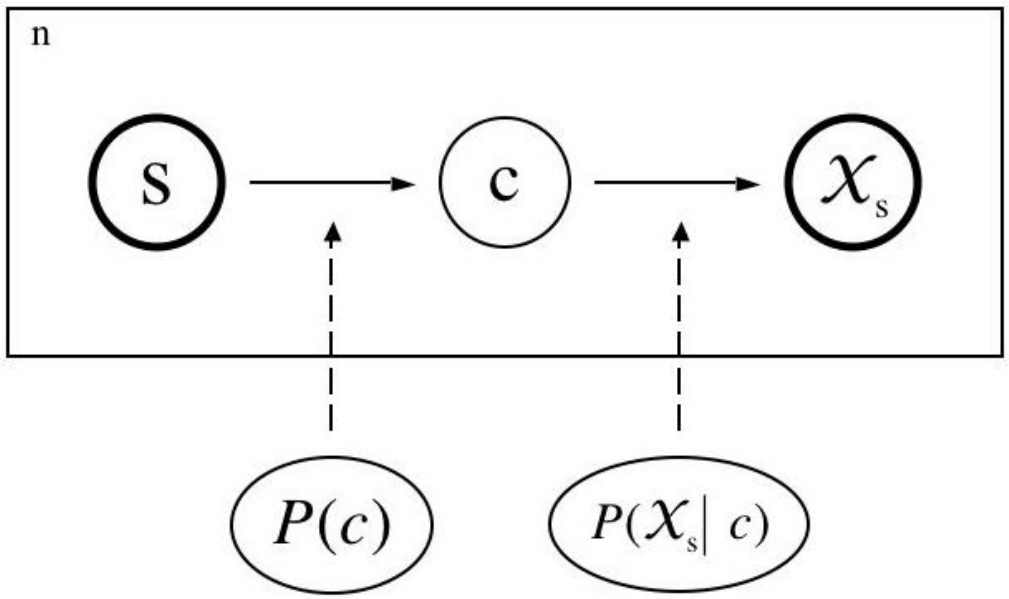


Figure 2.3: The image formation process as a graphical model. Random variables are denoted by circles, observed quantities are depicted bold. Statistical dependencies are expressed in the usual way via solid arrows, while frames indicate multiple instances. The probabilities governing the process are shown as ellipses.

2.3 The Likelihood Function

Having specified the generative model for the data acquisition process in the previous section, the complete data likelihood function is now being derived, i.e. the likelihood of the joint occurrence of all observed and unobserved variables given the parameters of the data generation process. As figure 2.3 indicates, the first step in the generative model is given by the assignment of image sites to cluster labels. This process is governed by the probabilities $p(c)$. Consequently, the probability of observing a certain group assignment at site $\mathbf{s}$ can be expressed as

$$p(\mathbf{m}(\mathbf{s})) = \sum_{c=1}^{k} \delta(c, \mathbf{m}(\mathbf{s}))p(c) = \prod_{c=1}^{k} p(c)^{\delta(c,\mathbf{m}(\mathbf{s}))}. \tag{2.5}$$

Here, the well-known Kronecker delta function of two variables is denoted by $\delta(\cdot,\cdot)$. The identity of the sum and the product in eq.(2.5) is valid due to the Boolean nature of the delta function which selects one out of k terms. As a consequence of the assumed conditional independence of sites given their group assignments, the probability of the complete configuration of site assignments $\mathbf{m}$

for a given image is given by

$$p(\mathbf{m}) = \prod_{\mathbf{s}=1}^{n} \prod_{c=1}^{k} p(c)^{\delta(c,\mathbf{m}(\mathbf{s}))}. \tag{2.6}$$

Having determined the association of sites to clusters, the second step of the generative model consists of the sampling of feature values according to the cluster-specific Gaussian mixture models. In the quantized representation, the set of observations at a given site $\mathbf{s}$ is summarized in the histogram $\mathbf{n}_\mathbf{s}$. Therefore, the probability of the observation set at the given site can be expressed as

$$p(\mathcal{X}_\mathbf{s} |\, \mathbf{m}(\mathbf{s})) = \prod_{j=1}^{m} \left(\sum_{\alpha=1}^{l} p_{\alpha |\, \mathbf{m}(\mathbf{s})} G_\alpha(j) \right)^{\mathbf{n}_{\mathbf{s}j}}. \tag{2.7}$$

Consequently, the probability of the complete set of observations at all sites given the cluster assignments is

$$p(\mathcal{X} |\, \mathbf{m}) = \prod_{\mathbf{s}=1}^{n} \prod_{c=1}^{k} \left[\prod_{j=1}^{m} \left(\sum_{\alpha=1}^{l} p_{\alpha |\, \mathbf{m}(\mathbf{s})} G_\alpha(j) \right)^{\mathbf{n}_{\mathbf{s}j}} \right]^{\delta(c,\mathbf{m}(\mathbf{s}))}. \tag{2.8}$$

Combining eq.(2.6) and eq.(2.8) one arrives at the complete data likelihood:

$$\begin{aligned} p(\mathcal{X}, \mathbf{m}) &= p(\mathbf{m}) p(\mathcal{X} |\, \mathbf{m}) \\ &= \prod_{\mathbf{s}=1}^{n} \prod_{c=1}^{k} \left[p(c) \prod_{j=1}^{m} \left(\sum_{\alpha=1}^{l} p_{\alpha |\, \mathbf{m}(\mathbf{s})} G_\alpha(j) \right)^{\mathbf{n}_{\mathbf{s}j}} \right]^{\delta(c,\mathbf{m}(\mathbf{s}))}. \end{aligned} \tag{2.9}$$

In order to simplify subsequent computations, the corresponding log-likelihood $\mathcal{L}(\mathcal{X}, \mathbf{m})$ will be considered, where the unwieldy products change into sums due the application of the logarithm. Apart from this, results from large deviation theory also suggest that the transformation of sums into products is beneficial. Large sums behave as Gaussian random variables with small variance whereas large products can have strange distributions, e.g. the most likely value is not equal to the average value. Please note that the use of the logarithm does not affect the location of the optima because it is a monotone transformation. As the goal is to pose the search for optimal parameter values as a minimization problem, the negative log-likelihood is regarded as the natural cost function $\mathcal{H}^{\mathrm{PDC}}$ of the PDC approach:

$$\begin{aligned} \mathcal{H}^{\mathrm{PDC}} &= -\mathcal{L}(\mathcal{X}, \mathbf{m}), \quad \text{where} \\ -\mathcal{L}(\mathcal{X}, \mathbf{m}) &= -\sum_{\mathbf{s}=1}^{n} \sum_{c=1}^{k} \delta(c, \mathbf{m}(\mathbf{s})) \left[\log p(c) + \sum_{j=1}^{m} \mathbf{n}_{\mathbf{s}j} \log \left(\sum_{\alpha=1}^{l} p_{\alpha |\, c} G_\alpha(j) \right) \right] \\ &= -\sum_{\mathbf{s}=1}^{n} \log p(\mathbf{m}(\mathbf{s})) - \sum_{\mathbf{s}=1}^{n} \sum_{j=1}^{m} \mathbf{n}_{\mathbf{s}j} \log \left(\sum_{\alpha=1}^{l} p_{\alpha |\, \mathbf{m}(\mathbf{s})} G_\alpha(j) \right). \end{aligned} \tag{2.10}$$

2.4 A different view on the PDC cost function

In the previous section, the cost function of the PDC approach has been derived from a maximum likelihood perspective on the image formation model. There exists, however, another interpretation of the cost function that provides interesting insights in the nature of the PDC image segmentation model. Please note, that the histogram of observations $\mathbf{n_s}$ at a given site $\mathbf{s}$, which has been used in the previous derivation of the cost function, can be turned into the empirical distribution $\hat{p}_\mathbf{s}$ of observations by mere re-normalization: $\hat{p}_\mathbf{s} = \frac{1}{r}\mathbf{n_s}$. Therefore, denoting the empirical distribution over bin j at site $\mathbf{s}$ by $\hat{p}_\mathbf{s}(j)$, the PDC cost function can be re-written as

$$\mathcal{H}^{\mathrm{PDC}} = -\sum_{\mathbf{s}=1}^{n} \log p(\mathbf{m}(\mathbf{s})) - r \sum_{\mathbf{s}=1}^{n} \sum_{j=1}^{m} \hat{p}_\mathbf{s}(j) \log \left(\sum_{\alpha=1}^{l} p_{\alpha|\,\mathbf{m}(\mathbf{s})} G_\alpha(j) \right). \tag{2.11}$$

In the next step, a constant (i.e. 0) is added to the cost function, arriving at

$$\begin{aligned} \mathcal{H}^{\mathrm{PDC}} = & \sum_{\mathbf{s}=1}^{n} \log p(\mathbf{m}(\mathbf{s})) \\ & - r \sum_{\mathbf{s}=1}^{n} \sum_{j=1}^{m} \hat{p}_\mathbf{s}(j) \log \left(\sum_{\alpha=1}^{l} p_{\alpha|\,\mathbf{m}(\mathbf{s})} G_\alpha(j) \right) \\ & + r \sum_{\mathbf{s}=1}^{n} \sum_{j=1}^{m} \hat{p}_\mathbf{s}(j) \log(\hat{p}_\mathbf{s}(j)) \\ & - r \sum_{\mathbf{s}=1}^{n} \sum_{j=1}^{m} \hat{p}_\mathbf{s}(j) \log(\hat{p}_\mathbf{s}(j)). \end{aligned} \tag{2.12}$$

With a little algebra, eq.(2.12) can be transformed into

$$\begin{aligned} \mathcal{H}^{\mathrm{PDC}} = & \sum_{\mathbf{s}=1}^{n} \log p(\mathbf{m}(\mathbf{s})) \\ & \underbrace{+ r \sum_{\mathbf{s}=1}^{n} \sum_{j=1}^{m} \hat{p}_\mathbf{s}(j) \left[\log(\hat{p}_\mathbf{s}(j)) - \log \left(\sum_{\alpha=1}^{l} p_{\alpha|\,\mathbf{m}(\mathbf{s})} G_\alpha(j) \right) \right]}_{(*)} \\ & \underbrace{- r \sum_{\mathbf{s}=1}^{n} \sum_{j=1}^{m} \hat{p}_\mathbf{s}(j) \log(\hat{p}_\mathbf{s}(j))}_{(**)}. \end{aligned} \tag{2.13}$$

Clearly, the term $(**)$ does not depend on the parameters of the cost function, but solely on the input data. Therefore, it can be safely dropped in the optimization process. Moreover, using the short-hand $\tilde{p}_\mathbf{s}(j) = \sum_{\alpha=1}^{l} p_{\alpha|\,\mathbf{m}(\mathbf{s})} G_\alpha(j)$ while

denoting the Kullback-Leibler divergence between two distribution by $D_{\mathrm{KL}}(\cdot||\cdot)$, one arrives at

$$\begin{aligned}(*) &= r \sum_{\mathbf{s}=1}^{n} \sum_{j=1}^{m} \hat{p}_{\mathbf{s}}(j) \log \left(\frac{\hat{p}_{\mathbf{s}}(j)}{\sum_{\alpha=1}^{l} p_{\alpha|\,\mathbf{m}(\mathbf{s})} G_{\alpha}(j)} \right) \\ &= r \sum_{\mathbf{s}=1}^{n} D_{\mathrm{KL}}(\hat{p}_{\mathbf{s}}(\cdot)||\tilde{p}_{\mathbf{s}}(\cdot)). \end{aligned} \tag{2.14}$$

Summing up the derivations given above, the PDC cost function can be expressed as the sum of Kullback-Leibler divergences between the empirical distributions of features in the local neighborhood of the sites and the prototypical Gaussian mixture distributions of the clusters, which are associated with the sites, and additive contributions from the negative logarithm of the cluster probabilities :

$$\mathcal{H}^{\mathrm{PDC}} = r \sum_{\mathbf{s}=1}^{n} D_{\mathrm{KL}}(\hat{p}_{\mathbf{s}}(\cdot)||\tilde{p}_{\mathbf{s}}(\cdot)) - \sum_{\mathbf{s}=1}^{n} \log p_{\mathbf{m}(\mathbf{s})}. \tag{2.15}$$

2.5 Model Identification

After having specified the cost function for the standard PDC approach in eq. (2.10), it remains to be shown, how its free parameters are adapted to the observed image data in order to arrive at the cluster-model and thus at the sought segmentation of the input image at hand. Determining the values of the free parameters is the key problem in model identification for a given data set, which is accomplished by maximum likelihood estimation. For the PDC approach, the set of free parameters is given by

1. the cluster mapping function $\mathbf{m}$,
2. the second order mixture coefficients (i.e. the cluster probabilities) $\{p_c : c = 1, \ldots, k\}$,
3. the mixture coefficients for the cluster-specific models $\{p_{\alpha|\,c} : c = 1, \ldots, k, \alpha = 1, \ldots, l\}$,
4. the means of the Gaussian alphabet $\{\mu_\alpha : \alpha = 1, \ldots, l\}$ and
5. the corresponding variances $\{\mathbf{\Sigma}_\alpha : \alpha = 1, \ldots, l\}$.

The means and variances of the Gaussian alphabet are initialized by a conventional mixture model estimation step which is performed prior to the PDC inference. Despite the fact that in principle all the parameters listed above could be re-estimated during the optimization of the PDC model, the variances $\{\mathbf{\Sigma}_\alpha : \alpha = 1, \ldots, l\}$ are excluded in order to alleviate the computational burden.

The remaining parameters fall into two distinct categories. On the one hand, one has the cluster assignment mapping, which is of a binary nature. Opposed to this, items 2 to 4 comprise the set of continuous model parameters which are summarized in the set Θ, i.e. $\Theta = \{p_c, p_{\alpha|c}, \mu_\alpha : 1 \leq c \leq k, 1 \leq \alpha \leq l\}$.

Due to the fact that one has to solve a statistical inference problem with hidden variables (i.e. the cluster assignments), the method of choice is the well known *Expectation–Maximization–Algorithm* (EM) [DLR77]. EM proceeds iteratively by computing posterior probabilities $P(\mathbf{m}|\,\Theta^{\text{old}})$ in the E-step and maximizing the *averaged complete data log–likelihood* $\mathrm{E}[\mathcal{L}(\Theta|\,\mathbf{m})]$ with respect to Θ in the M-step. Extending this interpretation, EM can be viewed as maximizing the following joint function of the parameters Θ and the hidden states $\mathbf{m}$ (see [CT84, Hat86, NH99]):

$$\mathcal{F}' = \mathrm{E}\left[\log p(\mathcal{X}, \mathbf{m}|\,\Theta) - \log p(\mathbf{m})\right]. \tag{2.16}$$

Apart from a difference in the sign, this equation is identical to the *generalized free energy* $\mathcal{F}$ at temperature $T = 1$ known from statistical physics. Inserting the corresponding cost function $\mathcal{H}^{\text{PDC}} = -\mathcal{L}$, the free energy for arbitrary temperatures T is given by the following expression:

$$\mathcal{F} = \mathrm{E}[\mathcal{H}^{\text{PDC}}] - T \cdot H. \tag{2.17}$$

Here, H denotes the entropy of the distribution over the states $\mathbf{m}$. This formal equivalence provides an interesting link to another well known optimization paradigm called *Deterministic Annealing* (DA) [RGF90]. The key idea of this approach is to combine the advantages of a temperature controlled stochastic optimization method with the efficiency of a purely deterministic computational scheme. A given combinatorial optimization problem over a discrete state space $\mathcal{M}$ is relaxed into a family of search problems in the space $\mathcal{P}(\mathcal{M})$ of probability distributions over that space. In this setting, the *generalized free energy* takes the role of the objective function. The temperature parameter T controls the influence of the entropic term, leading to a convex function in the limit of $T \to \infty$. At $T = 0$ the original problem is recovered. The optimization strategy starts at high temperature and it tracks local minima of the objective function while gradually lowering the computational temperature. An in-depth description of DA as well as its relation to Simulated Annealing (SA) techniques and Maximum Entropy (MA) methods can be found in [Hof97] and [Puz99].

Setting $q_{\mathbf{s}c} = \mathrm{E}[\delta(c, \mathbf{m}(\mathbf{s}))] = p(\delta(c, \mathbf{m}(\mathbf{s})) = 1)$, the expected costs of a given configuration is given by:

$$\mathrm{E}[\mathcal{H}^{\text{PDC}}] = -\sum_{\mathbf{s}} \sum_{c} q_{\mathbf{s}c} \left[\log p_c + \sum_{j} \mathbf{n}_{\mathbf{s}j} \log \left(\sum_{\alpha} p_{\alpha|c} G_\alpha(j)\right)\right]. \tag{2.18}$$

2.5.1 E-Step–Equations

Maximizing eq. (2.18) with respect to $P(\mathcal{M})$, which basically recovers the E-Step of the EM–scheme, requires to evaluate the partial costs of assigning a site $\mathbf{s}$ to cluster c. The additive structure of the objective function allows us to determine these partial costs h as

$$h_{\mathbf{s}c} = -\log p_c - \sum_j \mathbf{n}_{\mathbf{s}j} \log \left(\sum_\alpha p_{\alpha|\,c} G_\alpha(j) \right) . \tag{2.19}$$

Utilizing the well known fact from statistical physics that the generalized free energy at a certain temperature is minimized by the corresponding Gibbs distribution, one arrives at the update equations for the various $q_{\mathbf{s}c}$:

$$q_{\mathbf{s}c} \propto \exp(-\frac{1}{T} h_{\mathbf{s}c}) = \exp\left(\frac{1}{T} \left(\log p_c + \sum_j \mathbf{n}_{\mathbf{s}j} \log \left(\sum_\alpha p_{\alpha|\,c} G_\alpha(j) \right) \right) \right) . \tag{2.20}$$

2.5.2 M-Step–Equations

In accordance with [NH99], the estimates for the class probabilities p_c must satisfy

$$\frac{\partial}{\partial p_c} \mathcal{F} - \lambda \cdot \left(\sum_{\tilde{c}=1}^{k} p_{\tilde{c}} - 1 \right) \;=\; 0 \;, \tag{2.21}$$

where λ is a Lagrange parameter enforcing a proper normalization of p_c. Expanding $\mathcal{F}$ and solving for p_c leads to the M-step formulae

$$p_c = \frac{1}{n} \sum_{\mathbf{s}=1}^{n} q_{\mathbf{s}c} \;\;,\; c = 1, \ldots, k. \tag{2.22}$$

While lacking a closed-form solution for the second set of parameters $p_{\alpha|\,c}$, their optimal values can be found by an iterated numerical optimization. Instead of directly solving

$$\frac{\partial}{\partial p_{\alpha|\,c}} \mathcal{F} - \lambda \cdot \left(\sum_{\gamma=1}^{l} p_{\gamma|\,c} - 1 \right) \;=\; 0 \;, \tag{2.23}$$

which would be the analog to eq. (2.21), we repeatedly select two Gaussian components α_1 and α_2. Keeping $p_{\gamma|\,c}$ fixed for $\gamma \notin \{\alpha_1, \alpha_2\}$, $p_{\alpha_2|\,c}$ is directly coupled to $p_{\alpha_1|\,c}$ via

$$p_{\alpha_2|\,c} \;=\; 1 - \sum_{\gamma \notin \{\alpha_1, \alpha_2\}} p_{\gamma|\,c} - p_{\alpha_1|\,c} \;, \tag{2.24}$$

so that only one free parameter remains. Inserting (2.24) into (2.23), one obtains

$$\frac{\partial}{\partial p_{\alpha_1|\,c}}\mathcal{F}\left(\alpha_1,\alpha_2\right) \;=\; -\sum_{j=1}^{m}\left(\sum_{\mathbf{s}=1}^{n} q_{\mathbf{s}c}\mathbf{n}_{\mathbf{s}j}\right)\frac{G_{\alpha_1}\left(j\right)-G_{\alpha_2}\left(j\right)}{\sum_{\gamma=1}^{l} p_{\gamma|\,c}G_{\gamma}\left(j\right)}\;, \qquad (2.25)$$

and

$$\frac{\partial^2}{\partial p_{\alpha_1|\,c}^2}\mathcal{F}\left(\alpha_1,\alpha_2\right) \;=\; \sum_{j=1}^{m}\left(\sum_{\mathbf{s}=1}^{n} q_{\mathbf{s}c}\mathbf{n}_{\mathbf{s}j}\right)\frac{\left(G_{\alpha_1}\left(j\right)-G_{\alpha_2}\left(j\right)\right)^2}{\left(\sum_{\gamma=1}^{l} p_{\gamma|\,c}G_{\gamma}\left(j\right)\right)^2}\geq 0\;. \qquad (2.26)$$

The joint optimization of α_1 and α_2, therefore, amounts to solving a one-dimensional convex optimization problem. The optimal value of α_1 is either located on the boundary of the interval $\left[0;1-\sum_{\gamma\notin\{\alpha_1,\alpha_2\}} p_{\gamma|\,c}\right]$, or is equal to the zero-crossing of (2.25). In the latter case, it can be determined by the Newton method or by an interval bisection algorithm, which were both found to achieve sufficient precision after few optimization steps. The computational demands of this algorithm are dominated by the evaluation of $\sum_{\mathbf{s}=1}^{n} q_{\mathbf{s}c}\mathbf{n}_{\mathbf{s}j}$, which is linear in the number of sites, n. The computation of the remaining parts of (2.25) scales with the number of clusters, k, and the number of bins, m, and can thus be done efficiently.

Some care should also be spent on the selection of α_1 and α_2. Although the free energy will monotonously decrease even if α_1 and α_2 are randomly drawn, the convergence can be enhanced by choosing, in each iteration, α_1 and α_2 such that $\left\|\frac{\partial}{\partial p_{\alpha_1|\,c}}\mathcal{F}\left(\alpha_1,\alpha_2\right)\right\|$ is maximum. To adjust the mixture distribution $p_{\alpha|\,c}$ for a fixed cluster c, it is usually sufficient to repeat the selection and subsequent optimization of pairs (α_1,α_2) for const. $\cdot\, l$ times, where const. is a small constant (e.g. const. $= 3$). Although the optimization process might not have found the exact position of the global cost minimum at this time (*incomplete M-step*), any further optimization is unlikely to substantially influence the M-step result, and can thus be skipped.

Finally it is possible to adapt the means μ_α. To improve the readability, the calculations are restricted to one-dimensional data (when operating in d dimensions, diagonal covariance matrices are assumed, so that the estimation of the d-dimensional vector μ_α reduces to d one-dimensional optimization problems). Denote by $x_j^{\ominus}$ and $x_j^{\oplus}$ the boundaries of the interval $I_j = \left[x_j^{\ominus};\, x_j^{\oplus}\right]$, so that $G_\alpha(j) = \int_{x_j^{\ominus}}^{x_j^{\oplus}} g_\alpha(x)dx$. μ_α can then be determined by gradient or Newton descent, the first derivative of $\mathcal{F}$ being given by

$$\frac{\partial}{\partial \mu_\alpha}\mathcal{F} = -\sum_{c=1}^{k}\sum_{j=1}^{m}\left(\sum_{\mathbf{s}=1}^{n} q_{\mathbf{s}c}\mathbf{n}_{\mathbf{s}j}\right) p_{\alpha|\,c}\frac{g_\alpha\left(x_j^{\ominus}\right)-g_\alpha\left(x_j^{\oplus}\right)}{\sum_{\gamma=1}^{l} p_{\gamma|\,c}G_\gamma\left(j\right)}\;. \qquad (2.27)$$

It has been observed in the segmentation experiments, that fixed means μ_α, initialized by a conventional mixture model procedure, produced satisfactory seg-

mentation results. Adapting the means, however, can improve the performance of PDC when used as a generative model.

2.6 Multi-Scale Techniques

As has been demonstrated in [Puz99], the performance of Deterministic Annealing, which is already good in comparison to stochastic search algorithms like Simulated Annealing, can be significantly improved by multi-scale techniques. The idea of multi-scale optimization is to lower the computational complexity by decreasing the number of considered entities in the object space. In most application domains for image segmentation it is a natural assumption, that neighboring image sites contain identical, or at least similar, feature histograms. This domain–inherent structure is exploited to create a pyramid of coarsened data and configuration spaces by tying neighboring assignment variables. However, this approach can only be brought to bear, if the relevant cost function is closed under the pertaining coarsening operations, which means that the cost functions for all coarsened levels in the hierarchy are of the same algebraic type as the original.

In keeping with the notations introduced in [Puz99], it is now shown, that the PDC cost function as given in eq.(2.10) is indeed closed under multi-scale coarsening for arbitrary coarsening structures. This amounts to prove, that the PDC cost functions for all layers of the coarsened assignment variables remain of the same algebraic type. To this end, the definition of the PDC cost function from eq.(2.10) is repeated first:

$$\mathcal{H}^{\mathrm{PDC}} = -\sum_{\mathbf{s}=1}^{n}\sum_{j=1}^{m} \mathbf{n}_{\mathbf{s}j} \log\left(\sum_{\alpha=1}^{l} p_{\alpha|\,\mathbf{m}(\mathbf{s})} G_{\alpha}(j)\right) - \sum_{\mathbf{s}=1}^{n} \log p(\mathbf{m}(\mathbf{s})). \tag{2.28}$$

Next, one has to define how the data belonging to sites on layer ℓ of the multi-scale hierarchy is transferred to the next coarser layer $\ell+1$. As one is concerned with non-normalized histograms, it is only natural to set the data on the coarser level as the sum of the observational data for all pertaining sites on the finer layer. Thus one arrives at the following recursive definition:

$$\mathbf{n}_{\mathbf{s}j}^{\ell+1} = \sum_{\mathbf{s}' \in I_{\ell}^{-1}(\mathbf{s})} \mathbf{n}_{\mathbf{s}'j}^{\ell}. \tag{2.29}$$

Having specified the way the feature information at finer layers in the multi-scale coarsening hierarchy is collected, one can now enter the recursive definition from eq.(2.29) in the equation describing the generation of coarse-grained cost functions for PDC clustering. Here, the set of sites at a given layer of the multi-scale structure is denoted by $\mathcal{S}_{(\cdot)}$, whereas the inverse coarsening map is denoted

by $I_{(\cdot)}^{-1}$:

$$\mathcal{H}_{\ell+1}^{\mathrm{PDC}} = -\sum_{\mathbf{s}\in\mathcal{S}_{\ell+1}} \sum_{\mathbf{s}'\in I_\ell^{-1}(\mathbf{s})} \sum_{j=1}^{m} \mathbf{n}_{\mathbf{s}'j}^{\ell} \log\left(\sum_{\alpha=1}^{l} p_{\alpha|\,\mathbf{m}(\mathbf{s})} G_\alpha(j)\right) - \sum_{\mathbf{s}\in\mathcal{S}_{\ell+1}} \log p(\mathbf{m}(\mathbf{s}))$$

$$= -\sum_{\mathbf{s}\in\mathcal{S}_{\ell+1}} \sum_{j=1}^{m} \mathbf{n}_{\mathbf{s}j}^{\ell+1} \log\left(\sum_{\alpha=1}^{l} p_{\alpha|\,\mathbf{m}(\mathbf{s})} G_\alpha(j)\right) - \sum_{\mathbf{s}\in\mathcal{S}_{\ell+1}} \log p(\mathbf{m}(\mathbf{s})). \quad (2.30)$$

As the equation given above clearly demonstrates, the algebraic form of the original PDC cost function $\mathcal{H}^{\mathrm{PDC}}$ is preserved on all levels of the multi-scale hierarchy. Consequently, multi-scale DA optimization as described in [Puz99] can be applied to speed up the optimization of the PDC cost function and thus the inference of segmentation solutions. Analogous results can be obtained for all PDC variants discussed in subsequent chapters.

2.7 Experimental Results

2.7.1 Implementation Details

For the evaluation of the basic PDC approach, two test scenarios have been examined. One is the segmentation of images according to color information alone, the other is given by the segmentation taking color and texture features into account. In the first case, the elementary features are given by the three dimensional color vectors at each pixel. In the latter case, this information is augmented by the magnitudes of a Gabor filter-bank as described in section 2.1. Therefore, one arrives at an eleven-dimensional feature vector at each image pixel. For every image in the test-set, a mixture model with 32 clusters has then been trained on these vectorial features to estimate the Gaussian alphabet. For reasons of computational efficiency, the alphabets have not been altered in the subsequent PDC model estimation.

As has been stated in section 2.1, the input data for the PDC approach to image segmentation are histograms of feature values taken at image sites which are located on a regular grid. For the experiments described here, a fine grid with an inter-site spacing of two pixels has been used. The local neighborhoods from which the histograms have been drawn had a size of 11×11 pixels, while discretizing the feature values for each dimension into 48 bins.

2.7.2 Test-Set and Evaluation Methodology

Judging the quality of a given segmentation can be considered a difficult problem, because a certain amount of subjectivity concerning the assessment can not be

entirely avoided. Even when this limitation is neglected, ground truth is unavailable in most cases. Furthermore the segmentation is often only one item in a larger context of processing steps. In those cases it is only natural, as Borra and Sakar point out [BS97], to judge the segmentation quality with respect to the overall task.

In contrast to this view, Martin et al. examined human image segmentation [MFTM01a] experimentally. Their results indicate a remarkable consistency in the segmentation of given images among different human observers. This finding motivates their ongoing effort to construct a database of human segmented images from the Corel collection [Cor] for evaluation purposes, which is publicly available together with a set of software tools which aid the analysis process. This set of images has been chosen as our testbed, making a direct comparison between our segmentation model and human performance possible.

It is therefore advisable, to briefly describe this evaluation framework before introducing the corresponding results. A more detailed description can be found in [MFM04]. The data-base consists of two sets of images. The first one contains 200 images which can be used to tune the parameters of the analyzed segmentation method. Consequently, these pictures are labeled as the training set. As ample experiential knowledge of the behavior of the PDC method on real-world images has already been gathered, this set has been entirely neglected in our performance assessment. The second collection, including 100 images, comprises the test set. For each item in the latter set, 5-10 segmentations by human subjects are available. The instructions for these individuals were brief: *You will be presented a photographic image. Divide the image into some number of segments, where the segments represent "things" or "parts of things" in the scene. The number of segments is up to you, as it depends on the image. Something between 2 and 30 is likely to be appropriate. It is important, that all of the segments have approximately the same importance.* These instructions clearly encourage the subjects to mark image regions based on semantic knowledge. Therefore, achieving a comparable performance by automated methods operating on low-level features alone can be considered a hard task.

The quality assessment of the machine-inferred boundaries between sub-parts of the depicted scenes is based on the performance measures *precision*, π, and *recall*, ρ, which have evolved into a standard evaluation technique in the information retrieval community [Rij79]. Their use in the quantification of edge detector performance dates back to the work of Abdou and Pratt [AP79]. Denoting the set of retrieved edge pixels which are also true boundary elements according to the ground-truth by $RETREL$ and the set of true boundary pixels by REL, while labelling the set of retrieved boundary elements by RET, one arrives at the following definition:

$$\rho = \frac{|RETREL|}{|RET|}, \qquad \pi = \frac{|RETREL|}{|REL|}. \tag{2.31}$$

Phrased in probabilistic terms, recall is the probability that a valid boundary pixel gets detected by the algorithm in question, where precision denotes the probability that a detected edge element lies at a true segment border. Clearly, there is a trade-off between the quantities recall and precision. If one can define a relative weight α between these entities, the so-called *F-measure* (c.f. [MFM04] or [Rij79]) provides a summary of precision and recall in those cases, where a single performance measure is required or sufficient. This entity is defined as follows:

$$F = \frac{\pi\rho}{\alpha\rho + (1-\alpha)\pi}. \tag{2.32}$$

As one can see from eq.(2.32), the *F-measure* is the harmonic mean of precision and recall for a particular weighting factor α, which is set to 0.5 for the evaluations presented in the following sections.

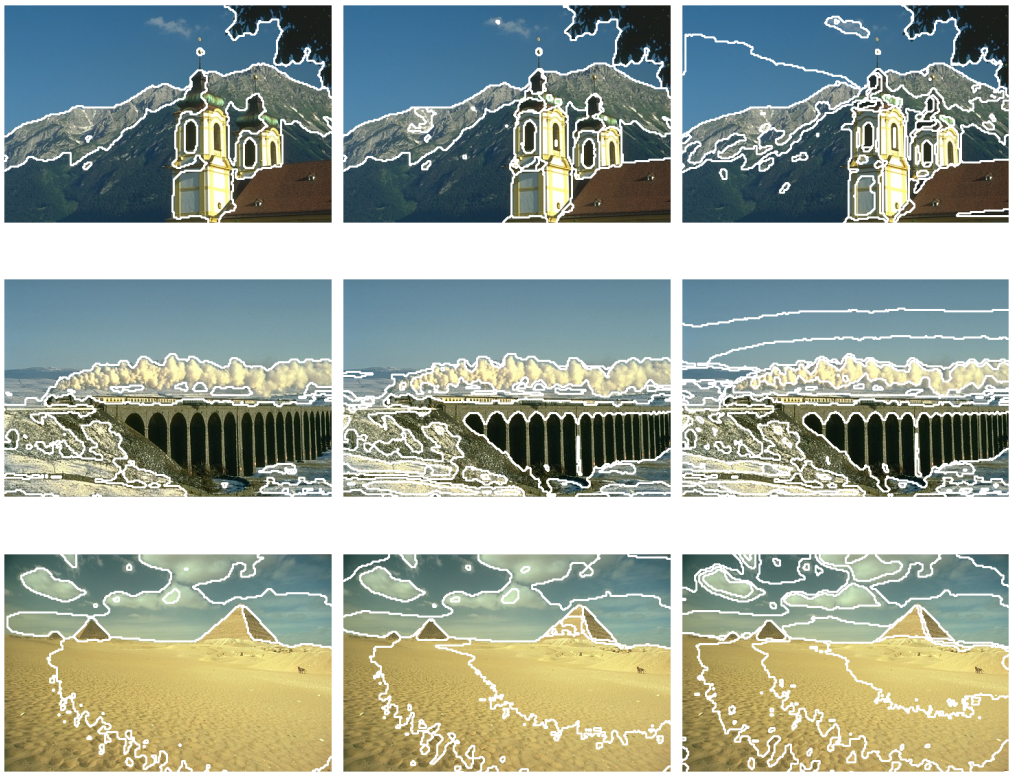
In order to make the notion of precision and recall and their summary statistic, the *F-measure* operational, the correspondence between the machine generated boundary pixels and those of the human segmentations has to be established. To this end, the correspondence problem is posed as a minimum cost bipartite assignment problem. In this scenario, the matching costs are given by the relative distance in the image plane between the border pixels. Furthermore, this enables one to declare all edge pixels, whose distance is beyond a pre-defined threshold to be non-hits.

As the data-base of human segmentations provides more than one segmentation solution for each of the images, a method for combining the results of different comparisons has to be found. Certainly, such an approach should have the desirable property to match human intuition of quality assessment by visual inspection as closely as possible. In particular, the expected properties of a boundary detector – detection, localization and uniqueness – should be encouraged. Therefore, as a first step, the human boundary maps are related to the machine output individually. The hit-rate is then averaged over the edge maps of the different human subjects. Consequently, perfect recall can only be achieved, when the machine generated segment borders match all of the boundaries found by the human observers.

Moreover, the whole evaluation approach is geared towards probabilistic edge maps. In this setting, the edge strengths are thresholded in order to arrive at a collection of binary boundary maps. The performance measurement as outlined above is then carried out at all levels of edge probability. In such a way, one arrives at precision and recall curves for each image as a function of the threshold. As PDC produces hard segment borders, this step cannot be carried out in our setting. Consequently, one can only provide single performance measures for each image and in summary for each of the tested algorithms, instead of complete precision and recall graphs.

2.7.3 Color Segmentation

In this section, the application of the PDC approach to image segmentation based solely on color features will be examined. With the settings for the data acquisition and pre-processing as described in section 2.7.1, the only remaining parameter, which has to be determined by user interaction, is the number of clusters. Please note, that this quantity refers to the partition of the data-space and therefore does not directly correspond to the number of segments in the PDC solutions. Clearly, the latter number is related to the former in the sense that higher cluster numbers always lead to higher numbers of image segments. In order to provide a reasonable overview concerning the performance of PDC color segmentation, solutions with three, five and eight clusters have been computed for each of the images in the Berkeley test-set.

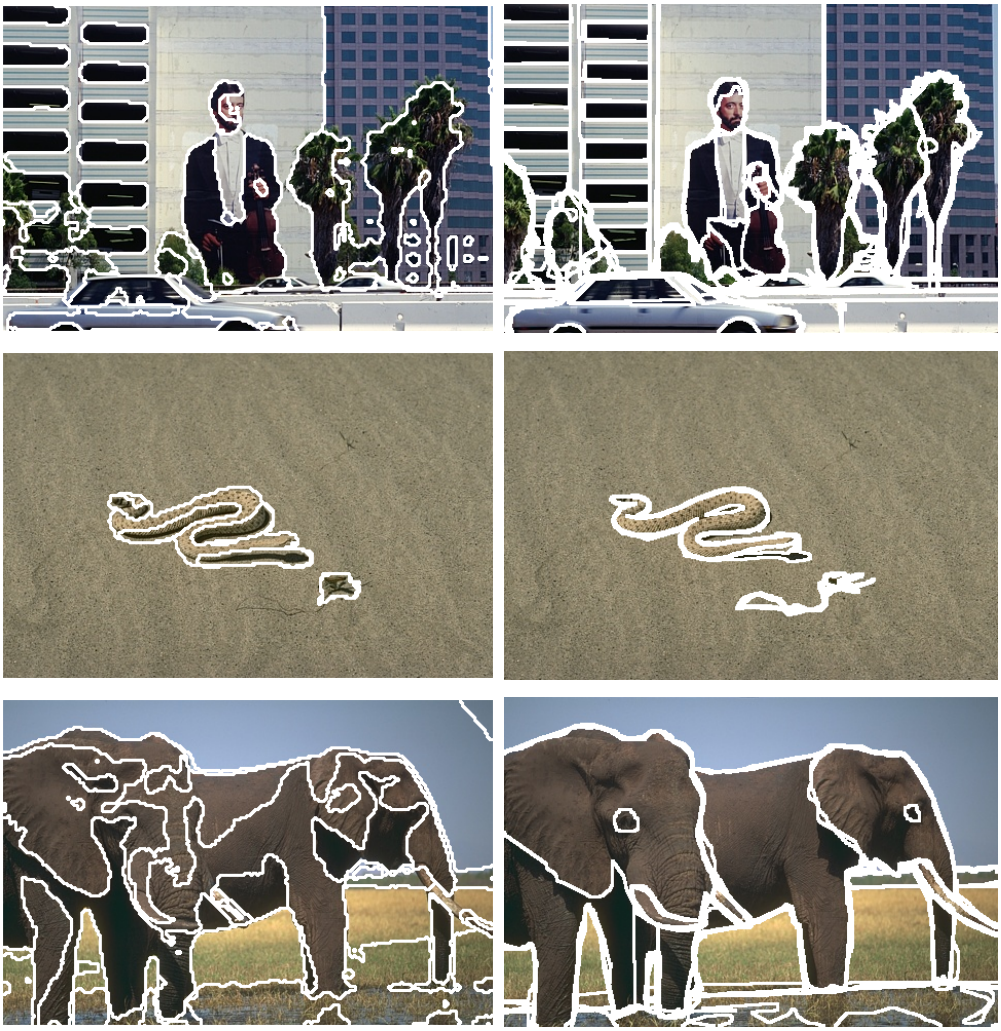


(a) (b) (c)

Figure 2.4: Effects of the number of data groups on the image segmentations: a) three, b) five and c) eight clusters.

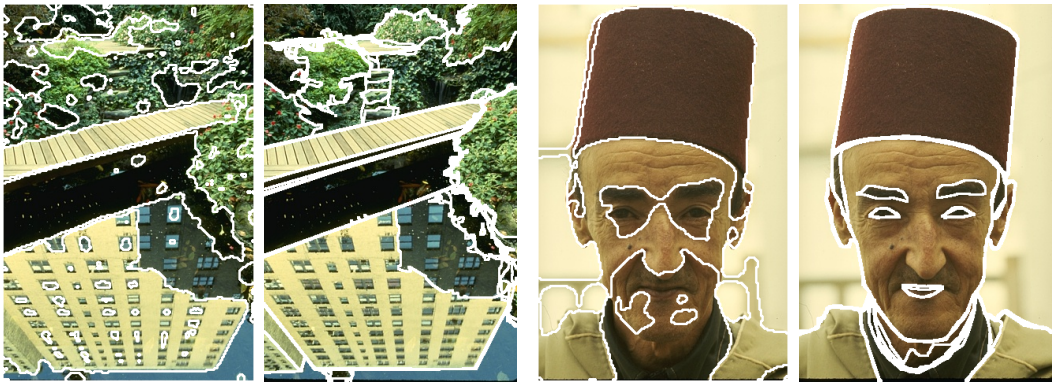
The effects of increasing the number of clusters for a given image at hand is demonstrated on four examples depicted in figure 2.4, where the borders between segments are indicated by thick white lines. Here one can observe, that

pronounced differences in terms of feature content result in the formation of separate clusters and thus in distinct segments even when only a small number of data groups is inferred. When the number of clusters is increased, more and more details of the structure, which is inherently present in the image data, emerge as unique segments in the PDC solutions. When a certain number of data groups is exceeded, one can notice the phenomenon of over-segmentation, i.e. the segmentation solutions show far too much detail than a human observer would deem appropriate given the semantic image content. Clearly, the point on the grouping granularity axis at which this actually happens is highly depending on the given image at hand.



(a) (b)

Figure 2.5: a) PDC color segmentations, b) overlaid segmentations of the human subjects.



(a) (b) (a) (b)

Figure 2.6: a) PDC color segmentations with three data groups, b) overlaid segmentations of the human subjects.

In order to provide a qualitative overview on the segmentation performance of the PDC approach for color features, a representative set of example segmentation results is depicted in the figures 2.5 and 2.6 under item a). For all of the pictures shown here, the segmentations are based on clustering solutions with three data groups, except for the image of the elephants in 2.5, for which the PDC segmentation with five clusters is depicted. Item b) in the aforementioned figures exhibit the results of the human subjects for the particular images. As there is always more than one human segmentation for each image in the Berkeley test-set, the results of the different subjects are overlaid. Due to small variations in boundary location, some of the edges are depicted by heavy white lines.

The first two segmentation results depicted in figure 2.5 show a remarkable consistency with the human segmentations for those images in question. For the city-scene example, some of the semantically motivated segment boundaries (e.g. treating the two neighboring trees on the right as independent regions, or differentiating between structural parts of the building on the left) could not be recovered by PDC, simply because of the fact that they are not warranted by the low-level color information. On the other hand, some small segments that have been inferred by PDC do not show up in the human image partitions because they are not semantically meaningful (e.g. around the lower parts of the trees on the right). The picture of the snake is even more striking, as the human image partitions are much more inconsistent with respect to the small object in front of the animal than the PDC solution. The last example in this collection, however, has been included here in order to demonstrate that there are cases, in which the partitions motivated by the image semantics and those which stem from the feature information noticeably disagree. Although the PDC segments spanning the group of elephants as a whole are well separated from the image background, the interior structure extracted by PDC and the one found by the human subjects are in pronounced discord.

Figure 2.6 shows another two comparison results between human and PDC segmentations. In the first image, which depicts a bridge over a pond, the PDC segmentation solution corresponds well to the human segmentations. However, small differences of the sort that has already been described in the commentary on the first image in figure 2.5 can also be noted here. In terms of segmentation quality, the results for the second picture, showing a portrait, are comparable to those of the last one in figure 2.5 in the sense that feature values and semantic content give rise to different image partitions.

# clusters	$\hat{\mathrm{E}}[\rho]$	$\hat{\mathrm{D}}[\rho]$	$\hat{\mathrm{E}}[\pi]$	$\hat{\mathrm{D}}[\pi]$	$\hat{\mathrm{E}}$[F-value]	$\hat{\mathrm{D}}$[F-value]
3	55.0	11.9	50.0	15.4	50.4	10.5
5	71.4	10.0	43.1	13.8	51.9	11.4
8	82.3	7.1	37.7	12.1	50.2	11.5

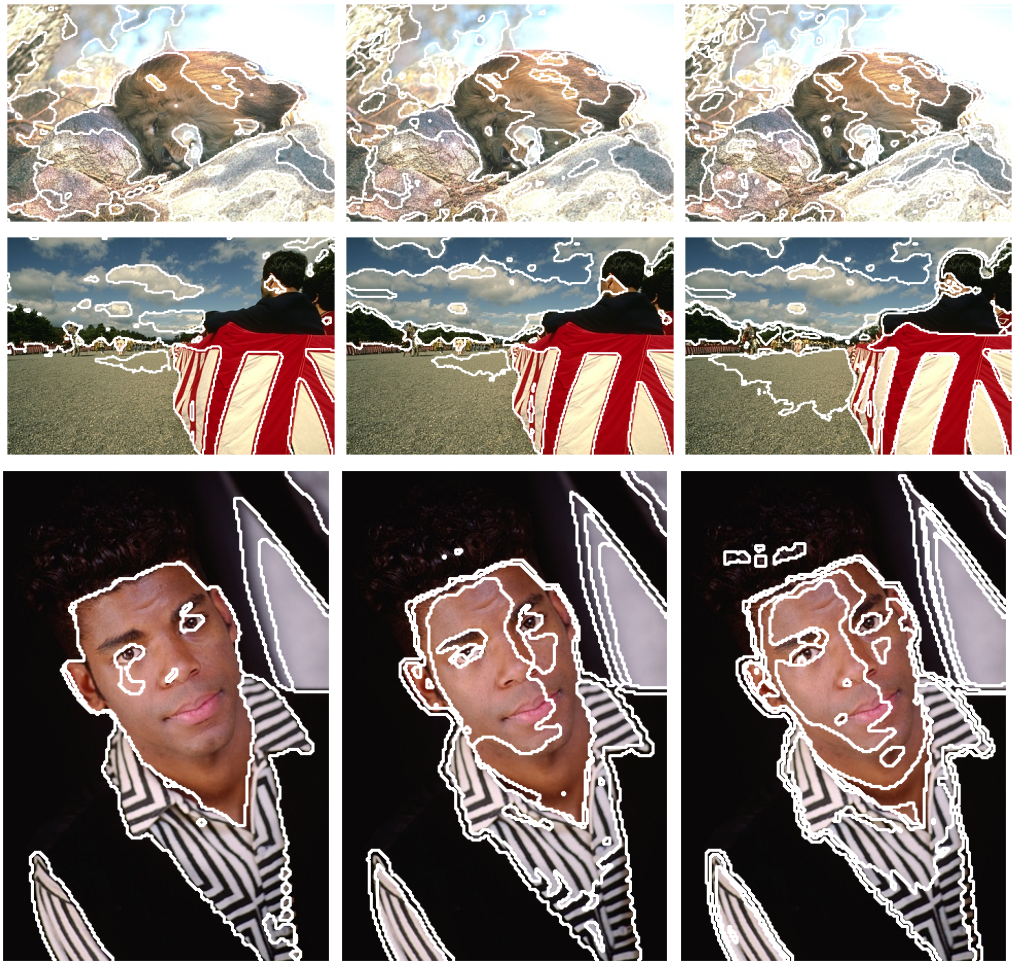
Table 2.1: Recall, Precision and F-value summary for PDC color segmentation experiments with three, five and eight clusters. Values have been multiplied by 10^2 for better readability.

For the quantitative analysis, all results of the color segmentation experiments with three, five and eight clusters respectively have been subjected to the evaluation in terms of precision, recall and F-values. The achieved performance is summarized in table 2.1, where the means (denoted by $\hat{\mathrm{E}}[\cdot]$) and standard deviations (denoted by $\hat{\mathrm{D}}[\cdot]$) for the segmentation quality indicators are given, which have been computed over the entire set of 100 test images. A more detailed graphical account of the corresponding distributions of the performance measures is given in the appendix. As expected, the recall measure increases with the growing number of inferred data groups. This is due to the fact that the number of segment boundaries grows accordingly. The precision, however, consecutively decreases, strongly indicating that the additional edges which show up at higher cluster numbers fail to provide additional information, which is relevant in terms of the semantically motivated image segmentations of the human subjects. According to the F-measure, the jointly optimal results in terms of precision and recall are generated for image partitions with five clusters. At a first glance, the overall performance of the PDC approach seems relatively low. Keeping in mind, however, that the median F-value for the human subjects lies at 0.8 (c.f. [MFM04]), one has to assert, that PDC achieves more than half of the segmentation quality than the human subjects. Taking the discussion in the paragraph on the qualitative evaluation of the PDC results into account, this can be considered rather remarkable. While PDC solely operates on low-level features, a human observer has many higher level knowledge sources to draw upon when producing an image segmentation. Moreover, the subjects who contributed to the image data-base were explicitly instructed to find partitions which are semantically meaningful.

2.7.4 Combined Color and Texture Segmentation

After having examined the performance of the PDC model for image segmentation based on color features alone, the results for image partitions which are based on the combination of color and texture features as outlined in section 2.7.1 are now being described.

The presentation of the corresponding findings is oriented at the scheme of the previous section. Consequently, the behavior of PDC solutions for growing numbers of inferred clusters is examined first. As in the case of the color image partitions, segmentation solutions for three, five and eight clusters have been determined for the 100 images in the Berkeley test-suite. An exemplary comparison of image partitions can be found in figure 2.7.



(a) (b) (c)

Figure 2.7: Effects of the number of data groups on the image segmentations: a) three, b) five and c) eight clusters.

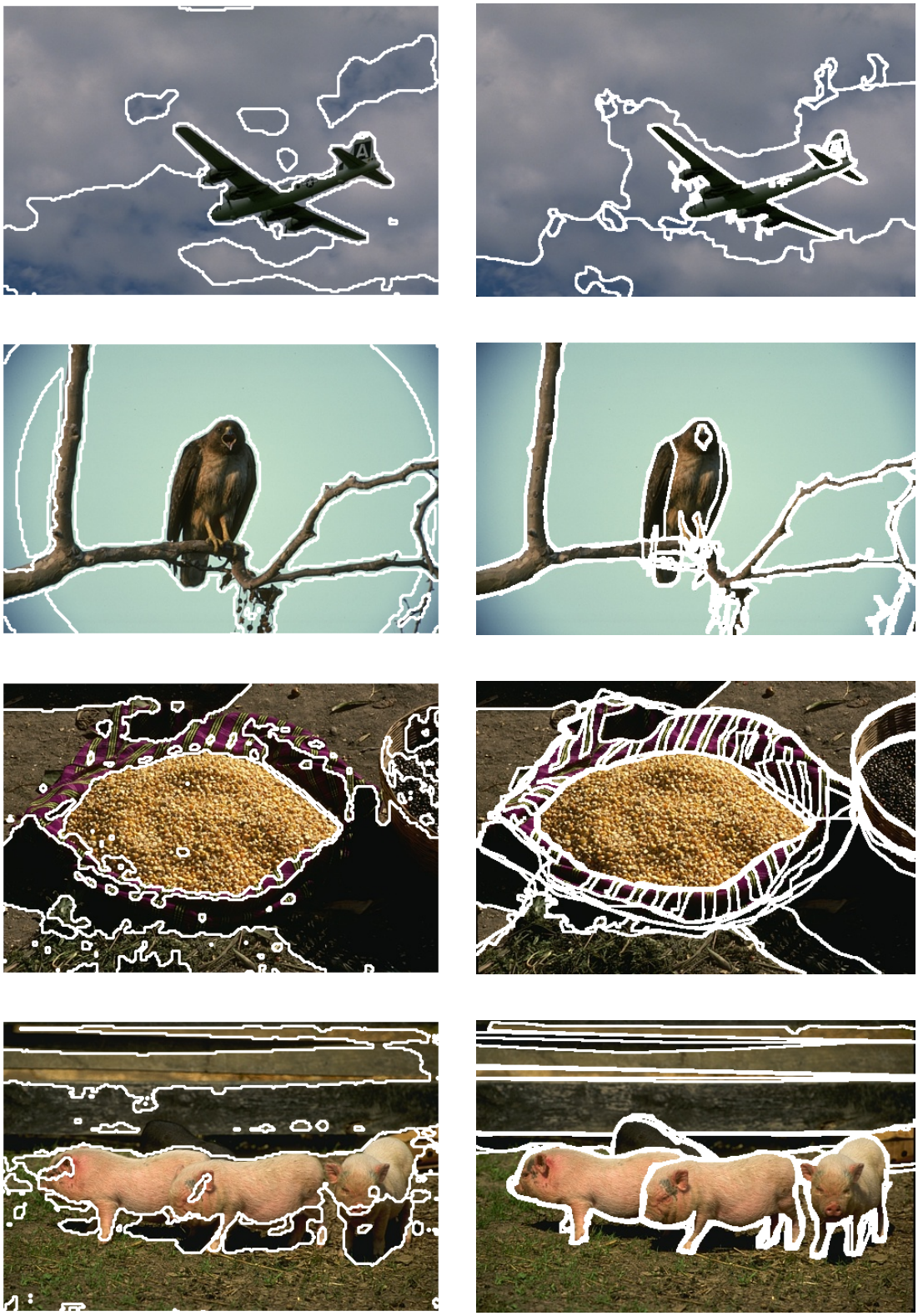
In principle, the same effects as in the case of color only partitions can again be observed here. Markable feature differences lead to deviating group assignments and therefore to the formation of different segments even when the number of inferred clusters is small. With the rising number of clusters, the PDC model gains the necessary additional degrees of freedom for the faithful representation of minor feature differences in terms of independent data groups and thus as individual segments in the image partition. Moreover, in comparison with the corresponding findings for color only partitions, one can observe that the tendency of generating fragmented segments at higher cluster numbers is somewhat alleviated. This result is explained by the large support of the Gabor filters, which effectively operates as a spatial smoothing of features. The downside of this property is given by the fact that image sites at the boundary between regions of differing feature content possess ambiguous statistical descriptions, as the pertaining histograms integrate information from both sides of the boundary. This effect is mainly responsible for the formation of double, or even triple, borders between image segments, which are also visible in the depicted results.

A qualitative comparison for the combined PDC color and texture segmentations is shown in figure 2.8. Here, image partitions with three clusters are opposed to their counterparts from the set of overlaid human segmentations. As in the previously given examples for color PDC, the overall agreement between human and machine generated boundaries is remarkable. For the picture in the first row of figure 2.8 one has to assess, that the PDC solution is even better than the summarized human image segmentations, due to their pronounced inconsistency concerning the partition of the background. However, there are again semantically motivated segment borders in the segmentations of the test-subjects, which cannot be reproduced by PDC because they do not correspond to differing feature content. The pictures in the second, third and fourth row provide good examples of this phenomenon.

# clusters	$\hat{\mathrm{E}}[\rho]$	$\hat{\mathrm{D}}[\rho]$	$\hat{\mathrm{E}}[\pi]$	$\hat{\mathrm{D}}[\pi]$	$\hat{\mathrm{E}}$[F-value]	$\hat{\mathrm{D}}$[F-value]
3	52.8	11.8	50.0	16.1	49.3	10.7
5	69.9	9.0	43.2	13.8	51.7	11.0
8	80.7	7.7	38.1	12.4	50.3	11.7

Table 2.2: Recall, Precision and F-value summary for combined color & texture PDC segmentation experiments with three, five and eight clusters. Values have been multiplied by 10^2 for better readability.

As the judgment of the qualitative results is always subjective, a quantitative analysis of the PDC segmentation performance has been conducted for the case of combined color and texture features as well. The findings are summarized in table 2.2. Boxplots of the pertaining distributions are given in the appendix.



(a) (b)

Figure 2.8: a) Combined color & texture PDC segmentations, b) overlaid segmentations of human subjects.

Although the incorporation of the additional information given by the texture features has been expected to improve the segmentation performance, the quantitative analysis shows, that this anticipation has not been warranted. The main reason for this failure is given by the fact that the test-suite of pictures does not contain a sufficient amount of examples, for which texture provides a relevant segmentation cue. The vast majority of samples can be sufficiently well described in terms of the color features alone. In order to demonstrate, that the incorporation of texture features can indeed lead to an improved segmentation performance, an additional collection of results for other images taken from the Corel image gallery is shown in figure 2.9. Here, the image content is chosen such that the texture features contain a large amount of relevant information in terms of the data grouping structure and thus for the image partition problem. On the left hand side, results for PDC segmentations based solely on color features are depicted, whereas the right hand side exhibits those image partitions which are based on the joint color and texture feature set. In all cases, segmentations with three clusters are shown.

In the first example in figure 2.9, one can observe, that in the color-only segmentation, the image part covering the tiger is bleeding out into the background near the animals tail. This effect is caused by the shadow that covers the tiger as well as the image region in the lower left. Opposed to this, the segmentation based on color and texture cues does not fail to identify the tiger as an independent image part.

The next picture shows two zebras, where color PDC fails to segment out the animals from the background. Including both color and texture features in the data-set, the results become markably different. The zebra on the right hand side is found with high accuracy as an unique image part. The segment for the animal on the left hand side still covers some minor part of the background. The segmentation is, however, much better in comparison with the image partition ignoring the texture cues.

In principle, the same observation holds true for the next example, in which two groups of zebras are shown. In the case of PDC segmentation based only on color features, the segments covering the groups of animals also contain substantial parts of the background. Taking the texture features into account, the zebras are identified precisely as individual segments. Moreover, the image partition as a whole is far less cluttered.

Finally, the last row exhibits the test case of a single zebra in front of a savannah background. Again, a faithful representation of the animal as an image segment of its own can not be achieved in the case of color-only PDC. When the texture features are included in the data-set, one finds that the zebra is inferred as an individual image part, for which the corresponding segment boundaries closely match the zebra's contour. Again, the full-featured segmentation solution for this image is much smoother and less contaminated by clutter than in the case of PDC image partition based solely on the color information.

(a) (b)

Figure 2.9: a) PDC color segmentations, b) Combined PDC color & texture segmentations.

2.8 Summary

In this chapter, a novel model for unsupervised image segmentation has been proposed. It is based on robust measurements of local image characteristics given by feature histograms. As one of the main contributions, it contains a continuous model for the group-specific distributions. In contrast to many other approaches, our method thus explicitly models the noise-induced errors in the histogramming of image content by the introduction of continuous model distributions for the inferred data groups. Being based on the theoretically sound maximum likelihood framework, our approach makes all modeling assumptions explicit in the cost function of the corresponding generative model. Moreover, there exists an informative interpretation of the PDC cost function in terms of the minimization of KL-divergences between the empirical distributions of features in the local vicinity of sites and the prototypical Gaussian mixture distributions of the data groups. Finally, the experimental results demonstrate the good performance of our model, often yielding close to human segmentation quality on the testbed. Moreover, a comprehensive quantitative performance evaluation of PDC has been conducted.

2.9 Bibliographic Remarks

For the problem of image segmentation, a large variety of approaches has been proposed. Correspondingly, a vast amount of literature on this topic exists today. Among the most well-known methods are region splitting and merging techniques [GPGSRC01, Pav72, VNBF00] including the *watershed transform* [Ols97, Wei98]. Graph partitioning, especially the *normalized cut* is also very popular [MSBL99, MBSL99, MBLS01, SM00, SBB01]. Moreover, variational methods, in particular in the level-set-framework, provide the basis of many segmentation algorithms [MS95, PD99, PD02, Par02]. Apart from this, image partition techniques applying concepts from the field of Markov chain Monte-Carlo (MCMC), e.g. [TZ02], and information theory [Lec89, NNF01, NNB$^+$01] are proposed. Novel image data domains like synthetic aperture radar outputs further stimulate the development of segmentation approaches [Fjø99, DMF01]. Using clustering techniques as a means of image partitioning also has a long standing history, recent approaches include [BCGM98, PF00, FZB01].

Data clustering as an unsupervised learning technique is extensively treated in [JD88] and other textbooks like [DHS01] and [Web99]. Concerning the optimization of objective functions in the DA framework, [Hof97, HPB98] provide excellent references. Multiscale-annealing as a means to speed-up DA computations is covered in [Puz99, PB99].

Numerous approaches to the quantitative measurement of texture properties have been proposed in the literature. The variety of methods reaches from the

parameters of explicit Markov random field models [CJ83], over fractal dimension [CS95], and wavelet coefficients [PTH96], to the means and variances of filter-bank outputs [DHW94, JF91].

The pioneering work of [PTL93] introduced the concept of histogram-based clustering. This idea has then been developed further in [PHB99] and [HP98]. Being build on these foundations, PDC using color features has been published in [HZB02], while the incorporation of the texture modality is described in [ZHB02]. Apart from the interpretation of the PDC cost function as the KL-divergence between site-specific empirical feature distribution and the cluster-specific prototypical distribution, another interesting connection to the *Information Bottleneck* principle [TPB99, TS01] is established in [Her03].

The validation of image segmentation results is a critical problem for the future development of this research field. Most often, only a small number of experimental results are given in the presentation of a new algorithm. Judging the quality of these results is frequently depending entirely on subjective criteria. Although some efforts are made towards a more objective performance assessment (c.f. [YMB77, CK97, CP02, SG95, SJK02]), most of the proposed approaches are tailored towards a certain application domain or employ disputable quality measures. Related survey articles are [BS97, Zha96]. On the other hand, the computer vision group at UC Berkeley offers a large test-bed of example images together with quality measures that clearly proved their worth in the validation of information retrieval approaches [MFTM01b, MFTM01a, MFM04].

Chapter 3

Incorporating Topological Constraints

In the preceding chapter, the basic Parametric Distributional Clustering (PDC) model has been introduced. Although the presentation of the approach has been geared towards the main application domain of image segmentation, PDC itself is a generic clustering model for any kind of histogram data. In this chapter, three modifications of PDC will be described, which constrain the data-grouping method in a topological way. First, the introduction of topological constraints in the spatial domain is examined, effectively coupling assignments for neighboring sites on the image grid. This approach will prove exceptionally valuable in relation to the performance of PDC as an image segmentation tool, due to the fact that one thereby captures the inherent smoothness of natural images. Inspired by approaches like Kohonen's self-organizing maps, topological constraints in cluster-space will then be examined. This leads to image segmentations which possess the property that similarity between prototypical feature distributions for the various clusters are made explicit in the ordering of data-groups in the sense that similar clusters are neighbors in the index space of data-groups. Finally, both kinds of topological couplings are combined in a single framework in the last section of this chapter.

3.1 Spatial Topology

The basic PDC model treats objects as independent entities, e.g. the grid sites for image segmentation are independent random variables without any mutual interactions. This is warranted, whenever no other relation between the relevant objects exists, which is not already accounted for by the similarity of the associated feature histograms. In the case of image segmentation, this assumption is not valid. Letting aside artificially created images, pictures of natural scenes are usually characterized by large homogeneous parts or regions. Therefore, neigh-

boring sites on the image grid tend to have a significantly larger probability of belonging to the same segment than sites which are separated by a considerable distance. Consequently, a better segmentation performance can be expected by taking this spatial topological relation into account.

3.1.1 The Cost Function for Spatially Coupled PDC

In order to make the notion of local spatial homogeneity mathematically precise, the following topological cost function is suggested, which explicitly couples assignments in a local neighborhood $\mathcal{N}_{\mathbf{s}}$ around each site $\mathbf{s}$ on the grid:

$$\begin{aligned}\mathcal{H}^{\mathrm{s}} &= \sum_{\mathbf{s}=1}^{n} \sum_{\tilde{\mathbf{s}} \in N(\mathbf{s})} \lambda^{\mathrm{t}} (1 - \delta(\mathbf{m}(\tilde{\mathbf{s}}), \mathbf{m}(\mathbf{s}))) \\ &= \sum_{\mathbf{s}=1}^{n} \sum_{c=1}^{k} \delta(\mathbf{m}(\mathbf{s}), c) \lambda^{\mathrm{t}} \sum_{\tilde{\mathbf{s}} \in \mathcal{N}_{\mathbf{s}}} (1 - \delta(\mathbf{m}(\tilde{\mathbf{s}}), c)). \end{aligned} \tag{3.1}$$

In the equation above, the weighting factor λ^{t} is introduced to control the extend, to which the homogeneity of local assignments is enforced. On the other hand, the basic unconstrained PDC cost function is of the following form:

$$\begin{aligned}\mathcal{H}^{\mathrm{PDC}} &= \sum_{\mathbf{s}=1}^{n} \sum_{c=1}^{k} \delta(c, \mathbf{m}(\mathbf{s})) h_{\mathbf{s}c}, \quad \text{where} \\ h_{\mathbf{s}c} &= -\sum_{j} \mathbf{n}_{\mathbf{s}j} \log \left(\sum_{\alpha} p_{\alpha|\, c} G_{\alpha}(j) \right) - \log p_c. \end{aligned} \tag{3.2}$$

The cost function $\mathcal{H}^{\mathrm{sPDC}}$ of a PDC model in which highly fragmented image partitions are discouraged can thus be described by the combination of the two terms given in eq.(3.1) and eq.(3.2):

$$\begin{aligned}\mathcal{H}^{\mathrm{sPDC}} &= \mathcal{H}^{\mathrm{PDC}} + \mathcal{H}^{\mathrm{s}} \\ &= \sum_{\mathbf{s}=1}^{n} \sum_{c=1}^{k} \delta(c, \mathbf{m}(\mathbf{s})) h_{\mathbf{s}c} + \sum_{\mathbf{s}=1}^{n} \sum_{c=1}^{k} \delta(\mathbf{m}(\mathbf{s}), c) \lambda^{\mathrm{t}} \sum_{\tilde{\mathbf{s}} \in \mathcal{N}_{\mathbf{s}}} (1 - \delta(\mathbf{m}(\tilde{\mathbf{s}}), c)) \\ &= \sum_{\mathbf{s}=1}^{n} \sum_{c=1}^{k} \delta(c, \mathbf{m}(\mathbf{s})) \underbrace{\left(h_{\mathbf{s}c} + \lambda^{\mathrm{t}} \sum_{\tilde{\mathbf{s}} \in \mathcal{N}_{\mathbf{s}}} (1 - \delta(\mathbf{m}(\tilde{\mathbf{s}}), c)) \right)}_{:= h^{\mathrm{s}}_{\mathbf{s}c}}. \end{aligned} \tag{3.3}$$

The abbreviation sPDC has been chosen for this clustering model because it describes a spatially *s*mooth variant of PDC. As the derivation in the equation above shows, one is able to formulate the cost function for PDC with spatial

topological coupling in direct analogy to the algebraic form of eq.(3.2). For $\mathcal{H}^{\mathrm{sPDC}}$, however, the local assignment cost for a site $\mathbf{s}$ do not only depend on the data and the continuous model parameters, but also on the assignment of the other sites in the local neighborhood.

3.1.2 Model Identification for sPDC

The identification of the relevant model parameters follows the same theoretical framework, which has already been used in the optimization of the PDC cost function (c.f. 2.5). Consequently, the derivation starts with the generalized free energy of the sPDC objective function:

$$\mathcal{F}^{\mathrm{sPDC}} = \mathrm{E}[\mathcal{H}^{\mathrm{sPDC}}] - T \cdot H. \tag{3.4}$$

Setting $q_{\mathbf{s}c} = \mathrm{E}[\delta(c, \mathbf{m}(\mathbf{s}))] = p(\delta(c, \mathbf{m}(\mathbf{s})) = 1)$, one arrives at

$$\mathrm{E}[\mathcal{H}^{\mathrm{sPDC}}] = \sum_{\mathbf{s}=1}^{n} \sum_{c=1}^{k} q_{\mathbf{s}c} \left(h_{\mathbf{s}c} + \lambda^{\mathrm{t}} \sum_{\tilde{\mathbf{s}} \in \mathcal{N}_{\mathbf{s}}} (1 - q_{\tilde{\mathbf{s}}c}) \right). \tag{3.5}$$

sPDC E-Step Equations: The computation of the cluster assignment probabilities $q^{\mathrm{s}}_{\mathbf{s}c}$ is based on the local costs $h^{\mathrm{s}}_{\mathbf{s}c}$ as defined in eq.(3.3). In the case of sPDC, these local costs are true meanfields due to the topological coupling of neighboring sites. Therefore, the assignment probabilities are computed according to a factorial model for the corresponding probability distribution which approximates the true non-factorial distribution (c.f. [Hof97] for details of the mean field approximation and mean field annealing (MFA)). Nevertheless, the assignment probabilities are given by the corresponding Gibbs distributions in the following way:

$$q_{\mathbf{s}c} \propto \exp(-\frac{1}{T} h^{\mathrm{s}}_{\mathbf{s}c}). \tag{3.6}$$

sPDC M-Step Equations: In complete analogy to the derivation of the M-step formulae for PDC, the cluster probabilities are computed by

$$p_c = \frac{1}{n} \sum_{\mathbf{s}=1}^{n} q_{\mathbf{s}c} \ , \ c = 1, \ldots, k. \tag{3.7}$$

For the cluster-specific mixture weights $p_{\alpha|\,c}$, the numeric optimization scheme outlined in 2.5.2 has to be applied. This amounts to repeatedly selecting pairs of coefficients $p_{\alpha_1|\,c}$ and $p_{\alpha_2|\,c}$ for which the optimization procedure is carried out. The optimal value of $p_{\alpha_1|\,\nu}$ either satisfies

$$-\sum_{c=1}^{k} \sum_{j=1}^{m} \left(\sum_{\mathbf{s}=1}^{n} q_{\mathbf{s}c} \mathbf{n}_{\mathbf{s}j} \right) \frac{G_{\alpha_1}(j) - G_{\alpha_2}(j)}{\sum_{\gamma=1}^{l} p_{\gamma|\,c} G_{\gamma}(j)} = 0, \tag{3.8}$$

and can be determined by interval-bisection, or – if no unique optimum can be found in the interval $\left[0; 1 - \sum_{\gamma \notin \{\alpha_1, \alpha_2\}} p_{\gamma|c}\right]$ – it is equal to one of the interval boundaries. Once $p_{\alpha_1|c}$ has been determined, $p_{\alpha_2|c}$ can be computed from

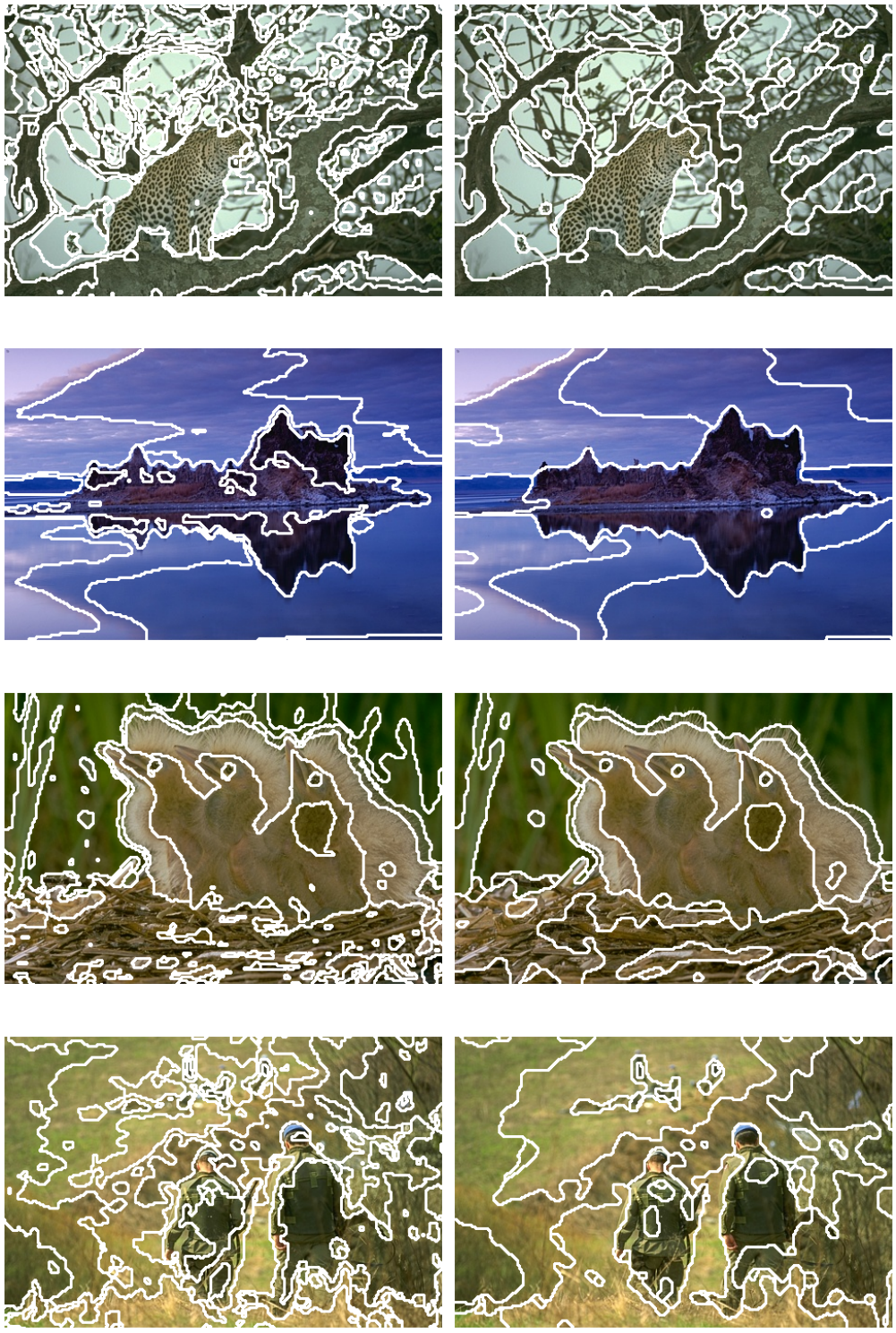
$$p_{\alpha_2|c} = 1 - \sum_{\gamma \notin \{\alpha_1, \alpha_2\}} p_{\gamma|c} - p_{\alpha_1|c} \,. \tag{3.9}$$

The optimization of the Gaussian means μ_α can also be accomplished in accordance with the approach outlined at the end of section 2.5.2. However, the initialization of the Gaussian alphabet with a standard mixture model estimation procedure has proven to yield satisfactory results for all practical applications. Thus, the adaptation of these parameters has not been utilized in the experiments, which are described in the ensuing section.

3.1.3 Experimental Evaluation

For the experiments which were conducted to test the sPDC approach to image segmentation, the basic settings for the data-acquisition and pre-processing steps are as described in section 2.7.1. Although the incorporation of texture features has not significantly improved the segmentation performance on the Berkeley test-bed, they have been included nevertheless due to the fact that the utilization of these segmentation cues has been shown useful for those cases, where texture is a prominent characteristic of the depicted scenes. Consequently, it can be considered inappropriate for a general purpose segmentation approach to neglect this information source. Moreover, the scheme of producing image partitions based on data-groupings with three, five, and eight clusters is abided by in order to get an impression of the image partition results at a reasonable variety of grouping granularities. The influence of the local assignment coupling on the segmentation solution is dependent on two parameters: the size of the local neighborhood $\mathcal{N}_{\mathbf{s}}$ around a site $\mathbf{s}$, and the weighting factor λ^{t}. For all of the experiments described in this section, the neighborhood size has been fixed to 5×5 sites, which leads to a reasonable amount of local smoothing. The setting of the weighting factor will be discussed in the next paragraph in the context of the comparison between spatially constrained and unconstrained segmentation results.

At first, a direct comparison of segmentation results for unconstrained PDC and sPDC is examined in order to demonstrate the qualitative performance difference between these two approaches. A collection of results is depicted in figure 3.1. Here, the first column shows image partitions by basic PDC, whereas the second column depicts the segmentations of the spatially coupled sPDC method. In the first row at the top of fig. 3.1, an example of a very irregular image is shown. The maze of branches together with the leopard, whose fur provides for a very effective natural camouflage, poses a very hard segmentation problem.



(a) (b)

Figure 3.1: Image partitions with five clusters based on color and texture features. a) PDC segmentations, b) sPDC segmentations.

Consequently, the image partition produced by the standard PDC approach is highly fragmented. Opposed to this, the incorporation of the spatial constraints on the group assignment variables leads to a much smoother result. Nevertheless, the major image structures like the leopard and the bigger branches are still represented very accurately. The next row depicts an image, which is much simpler in the sense that it it is inherently smooth. Yet, the application of the basic PDC method still leads to a certain number of small segments, while producing double edges at the boundaries between different image regions. These adverse effects are almost totally cured in the sPDC result. Image regions are separated by single borders and only one small region survives the smoothing induced by the spatial constraints. The third row contains another example of an image with a highly irregular feature content. Especially the straw in the birds nest provides for a large variability in terms of the measured image features. It does not come as a surprise that PDC, which solely optimizes the structural properties of the clustering solution in feature space, produces a fragmented image partition. Utilizing spatial constraints, however, sPDC is able to infer a much less cluttered segmentation solution. Finally, the last image in this collection again shows a rather difficult segmentation problem. As expected, the basic PDC approach leads to a highly irregular segmentation, while sPDC provides a substantially less cluttered image partition result. However, even the sPDC segmentation shows a certain amount of fragmentation. Clearly the influence of the spatial coupling is dependent on the magnitude of the weighting factor λ^{t} in the sPDC cost function. This factor has been set on an experiential basis, trying to strike for a balance between spatial smoothing and structure preservation of the clustering solution in feature space. In this respect, setting $\lambda^{\mathrm{t}} = 0.15 \cdot \bar{h}_{\mathbf{s}}$, where $\bar{h}_{\mathbf{s}}$ denotes the mean of the local costs $h_{\mathbf{s}c}$ at site $\mathbf{s}$, has proven to be a reasonable choice. As this setting has been fixed for all depicted segmentation results, a perfect solution for all images cannot reasonably be expected. Nevertheless, the yielded improvement is clearly visible from all of the presented image partition results.

In the next collection of segmentation results for sPDC shown in figure 3.2, a qualitative overview of segmentations with five clusters opposed to image partitions of the subjects in the Berkeley study is given. In the first image in this figure, a segmentation result is depicted for which the human subjects have produced largely differing segmentations concerning the objects in the foreground. sPDC on the other hand reliably separated the rock and the tree from the sky background. It fails, however, to identify the bushes in the foreground as independent segments but incorporates them in larger structures covering wider areas of the depicted scene. Moreover, the sky gets split into different image parts due to the apparent gradient in the blue hues. These differences are typical examples of segmentation based on low-level features alone as opposed to partitions including semantic knowledge. The overall agreement between the sPDC result and the human made segmentations is still very good.

(a) (b)

Figure 3.2: Qualitative performance overview of sPDC vs. the results of the human subjects. a) sPDC image partitions, b) overlaid segmentations of the human subjects.

The next picture in the second row shows a result, in which sPDC accurately infers the semantically meaningful image parts like the group of trees in the background and the buffalo in the foreground. Given the granularity of the clustering solution, the segregation of the background into the various parts is in good agreement with the local feature content. On the other hand, the corresponding background segmentations of the human subjects seem somewhat subjective and alternative solutions might also be acceptable. This observation is especially obvious when one focuses on the mountains in the far background of the depicted image. In the third row, we see a case in which sPDC also achieved a very good representation of the semantic content of the depicted scene, as the camel-rider and his animal are accurately separated from the background. The semantically motivated fine-structure of this image part and the separation of the pyramid from the background cannot be reproduced by sPDC. First of all, these partitions are not completely warranted from the feature information. Second, the data on which the machine segmentations are based is gathered from a spatial support of some extend. This fact is pronounced in sPDC by the spatial coupling of assignments. Therefore, there exists a certain limit of resolution or scale, at which independent image parts can still reliably be inferred. This effect also prevents sPDC to identify some of the bigger stones in the foreground as independent image regions, like the human subjects did. Keeping these limitations in mind, the correspondence between the sPDC segmentation and the human image partition is remarkable. Finally, the last image in this figure depicts an image, for which the agreement between the human and machine generated segmentations is even better. Here, even some of the internal fine-structuring of the depicted bird into various segments is largely identical. Moreover, despite some minor differences, the segregation of the background is in very good correspondence. Summing up the findings, one has to conclude that the smoother results of the sPDC approach correspond much better to the human image partitions than the more cluttered segmentations of the basic PDC method.

Apart from the impression that a direct comparison between human and machine results as given in figure 3.2 provides, the quantitative performance of sPDC on the Berkeley data-base has also been determined. The measured values for the segmentation quality indicators precision, recall, and F-value are shown in table 3.1. A more detailed graphical account of the corresponding distributions in form of a box-plot can be found in the appendix. The first important observation is that the recall values for all clusters numbers are noticeably lower than in the case of unconstrained PDC for color and texture features. This finding is explained by the fact that the topological coupling in the spatial domain significantly reduces the number of inferred segment boundaries and thus the clutter in the image partition results. This effect is amply visible in fig. 3.1 and has also already been noted in the corresponding discussion of the depicted results. Clearly, the probability of detecting a valid boundary pixel decreases to some extend when the number of inferred segment borders is remarkably lower. On the other hand, the

# clusters	$\hat{\mathrm{E}}[\rho]$	$\hat{\mathrm{D}}[\rho]$	$\hat{\mathrm{E}}[\pi]$	$\hat{\mathrm{D}}[\pi]$	$\hat{\mathrm{E}}$[F-value]	$\hat{\mathrm{D}}$[F-value]
3	44.2	11.3	56.3	16.5	47.8	10.2
5	55.4	11.2	50.9	15.0	51.5	10.8
8	62.8	10.0	46.6	13.1	52.1	10.0

Table 3.1: Recall, Precision and F-value summary for sPDC segmentation experiments with three, five and eight clusters. Values have been multiplied by 10^2 for better readability.

precision values, and thus the probability that a detected edge pixel lies at a true boundary between different image parts is noticeably increased. Consequently, the incorporation of spatial topological constraints seems to provide a strong advantage whenever the maximization of precision is required from the context in which the segmentation is applied. Moreover, judging the segmentation quality on a subjective basis, it seems that high precision values better correspond to the intuitive notion of a good segmentation result than the recall measure, which produces high values even for very fragmented image partitions. Keeping the variance of the precision values in mind, one has to observe that the precision distributions for the unconstrained case and the spatially coupled sPDC are not completely separated. Therefore, in order to determine the significance of the precision increase, a statistical test on the differences between the precisions for the constrained and unconstrained PDC variant has to be applied. In particular, the null-hypotheses that the empirical distribution of $\Delta\pi = \pi(sPDC) - \pi(PDC)$ for the Berkeley test-bed has mean zero is evaluated.

To this end, the statistics toolbox of the *Matlab* mathematical programming environment offers a test function, which is derived from the standard t-test known from the theory of statistical hypothesis testing. In this particular function, an empirical distribution of unknown variance is tested for the null-hypothesis of having a certain mean, which has to be specified as an input parameter. Moreover, that function provides for an additional parameter, which can be used to explicitly test the null-hypothesis against the case that the empirical distribution has a mean greater than the one suggested by the null-hypothesis. As this test function is ideally suited for the test that one has to perform on the difference values, it has been utilized to determine the significance of the increase in terms of the precision measure for spatially coupled PDC. The test has been performed for all data-grouping granularities, i.e. the results for sPDC with three, five, and eight clusters were tested against their unconstrained counterparts. In all cases, the null hypothesis of zero-mean for the precision differences could be rejected at a significance level of 0.01. The empirical distributions for the precision differences is depicted in figure 3.3. As one can see from this diagram, almost all of the probability mass is located in the interval above zero, clearly demonstrating that

the spatial coupling of assignment variables indeed offers a pronounced advantage over the standard PDC approach.

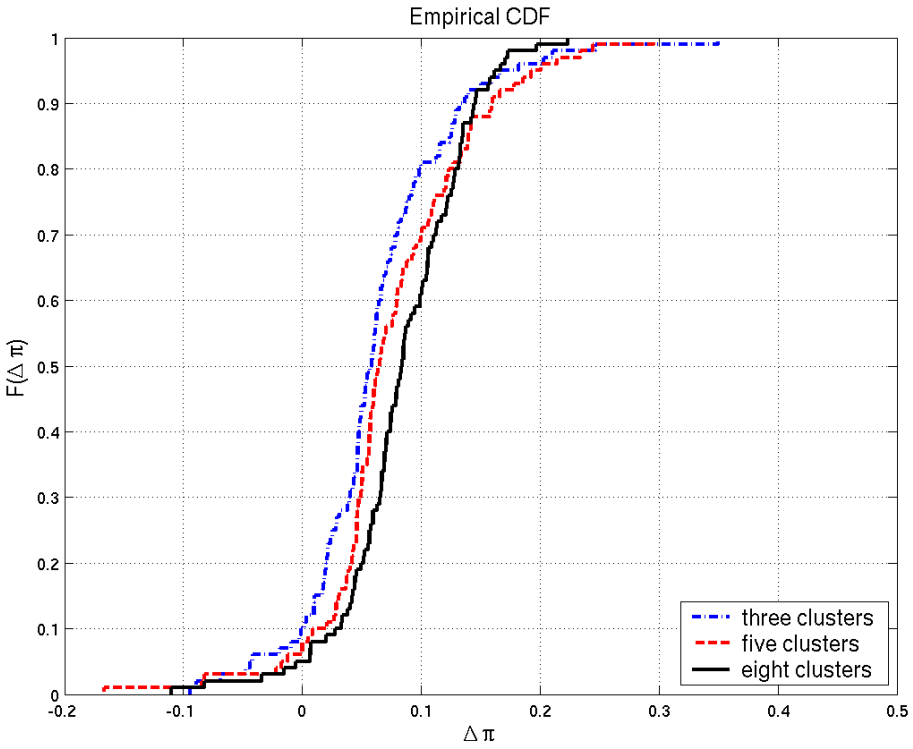


Figure 3.3: Empirical cumulative density functions of the precision differences between spatially constrained and standard PDC with three, five, and eight clusters.

3.2 Topology in Cluster-Space

The main contribution of this section is the consideration of the locality of segments in the space of mixture models. This is taken into account by the topological coupling of cluster assignments analogous to self-organizing maps in Neural Computation. In cases where the prototypical mixture distributions of clusters become similar, the noise, which is invariably present in the observational data, may lead to ambiguous cluster assignments which in turn are characterized by noise dependent transition probabilities between those data-groups. Therefore, the permutation symmetry of cluster indices is broken and a topology in the domain of data-groups is induced. Consequently, one can expect a better model adaptation by taking the resulting additional distortions into account when estimating the most likely group assignments. Thus, a connection to the field of

Self Organizing Maps [Koh95] is established. An analogous approach for the ACM histogram clustering model has first been suggested in [Buh98]. In figure 3.4 the underlying idea is visualized by depicting three data groups (each one being a mixture model with three components) in feature space. Neighboring clusters are connected by the topological coupling (visualized as arcs), thereby forming a chain in cluster space. Enforcing the data grouping solution to obey to such a simple neighborhood structure, similarities between clusters can be read of directly from the indices of the groups. Moreover, each cluster-specific mixture model is optimized in such a way, that it also to some extend accounts for measurements from neighboring data-groups.

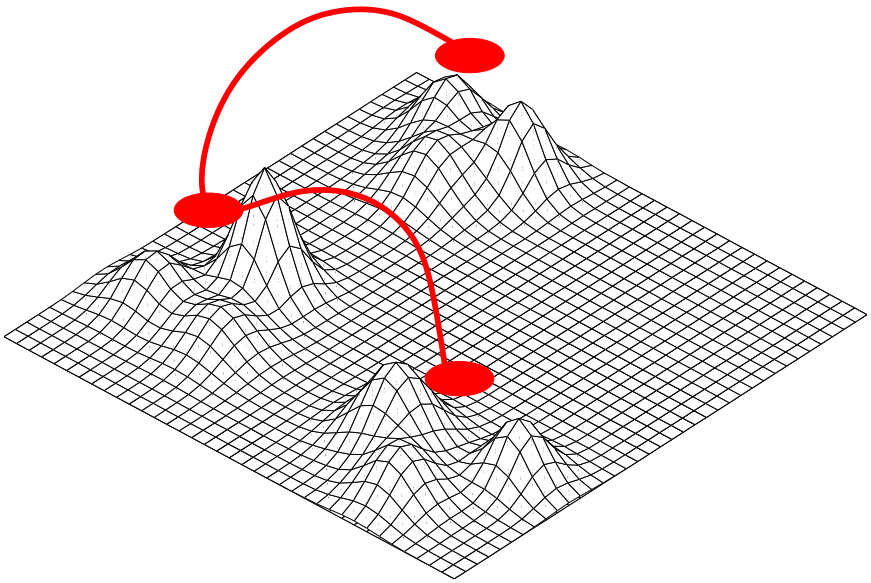


Figure 3.4: Topological coupling between neighboring clusters.

3.2.1 The Segmentation Model

For the derivation of the objective function for PDC with topological coupling in cluster-space, one starts from the cost function for the basic PDC model, which is repeated here in order to increase the clarity of presentation:

$$\mathcal{H}^{\mathrm{PDC}} = \sum_{\mathbf{s}=1}^{n} \sum_{c=1}^{k} \delta(c, \mathbf{m}(\mathbf{s})) h_{\mathbf{s}c}, \quad \text{where}$$

$$h_{\mathbf{s}c} = -\sum_{j} \mathbf{n}_{\mathbf{s}j} \log \left(\sum_{\alpha} p_{\alpha|c} G_{\alpha}(j) \right) - \log p_c. \tag{3.10}$$

As has been stated in the introduction of this section, the cost function as given above does not take the locality of data-groups in cluster-space into account.

Consequently, the probability of confusing any two indices $\tilde{c}$ and c has to be specified, which is accomplished by the introduction of the matrix $T_{c\tilde{c}}$. The coupling itself is formulated on the level of the local grouping cost, i.e. the following transition has to be made:

$$\begin{aligned} h_{\mathbf{s}c} &\rightarrow h^{\mathrm{T}}_{\mathbf{s}c}, \quad \text{where} \\ h^{\mathrm{T}}_{\mathbf{s}c} &= \sum_{\tilde{c}}^{k} T_{c\tilde{c}} h_{\mathbf{s}\tilde{c}}. \end{aligned} \tag{3.11}$$

Thus, the objective function for PDC with topological coupling in cluster-space (TPDC) is given by

$$\mathcal{H}^{\mathrm{TPDC}} = \sum_{\mathbf{s}=1}^{n} \sum_{c=1}^{k} \delta(c, \mathbf{m}(\mathbf{s})) h^{\mathrm{T}}_{\mathbf{s}c}. \tag{3.12}$$

As one can immediately see from the equation above, the TPDC cost function is still linear in the assignments of sites to clusters. That is, the assignment of any particular site is independent of the assignments of all other sites given the continuous model parameters. On the other hand, one can re-formulate the objective function in such a way that the topological coupling is made explicit in terms of the assignment variables, while the problem itself remains linear. To this end, one has

$$\begin{aligned} \mathcal{H}^{\mathrm{TPDC}} &= \sum_{\mathbf{s}=1}^{n} \sum_{c=1}^{k} \delta(c, \mathbf{m}(\mathbf{s})) \sum_{\tilde{c}}^{k} T_{c\tilde{c}} h_{\mathbf{s}\tilde{c}} \\ &= \sum_{\mathbf{s}=1}^{n} \sum_{c=1}^{k} \sum_{\tilde{c}}^{k} \delta(c, \mathbf{m}(\mathbf{s})) T_{c\tilde{c}} h_{\mathbf{s}\tilde{c}} \\ &= \sum_{\mathbf{s}=1}^{n} \sum_{\tilde{c}}^{k} \sum_{c=1}^{k} T_{c\tilde{c}} \delta(c, \mathbf{m}(\mathbf{s})) h_{\mathbf{s}\tilde{c}}. \end{aligned} \tag{3.13}$$

Here, the transition from the second to the last row is just a re-arrangement of the summation order. Setting

$$\delta^{\mathrm{T}}(\tilde{c}, \mathbf{m}(\mathbf{s})) = \sum_{c=1}^{k} T_{c\tilde{c}} \delta(c, \mathbf{m}(\mathbf{s})), \tag{3.14}$$

while exchanging the cluster summation indices c and $\tilde{c}$, one arrives at

$$\mathcal{H}^{\mathrm{TPDC}} = \sum_{\mathbf{s}=1}^{n} \sum_{c=1}^{k} \delta^{\mathrm{T}}(c, \mathbf{m}(\mathbf{s})) h_{\mathbf{s}c}. \tag{3.15}$$

3.2.2 TPDC Model Identification:

The identification of the correct settings for all relevant model parameters is treated like in the cases of PDC and sPDC. Thus, the alternating optimization scheme of EM is embedded in the DA framework. Consequently, one has to start from the generalized free energy:

$$\mathcal{F}^{\mathrm{TPDC}} = \mathrm{E}[\mathcal{H}^{\mathrm{TPDC}}] - T \cdot H. \tag{3.16}$$

In the equation above, H again denotes the entropy of the distribution over the states in $\mathcal{M}$, while the computational temperature is given by T. For TPDC, the expected cost are given by

$$\begin{aligned} \mathrm{E}[\mathcal{H}^{\mathrm{TPDC}}] &= \sum_{\mathbf{s}=1}^{n} \sum_{c=1}^{k} q_{\mathbf{s}c} h^{\mathrm{T}}_{\mathbf{s}c} \\ &= -\sum_{\mathbf{s}=1}^{n} \sum_{c=1}^{k} q_{\mathbf{s}c} \sum_{\tilde{c}}^{k} T_{c\tilde{c}} h_{\mathbf{s}\tilde{c}}. \end{aligned} \tag{3.17}$$

TPDC E-Step–Equations: Maximizing eq.(3.16) with respect to $P(\mathcal{M})$ corresponds to the E-Step of the EM–scheme. To this end, the partial costs of assigning a site $\mathbf{s}$ to cluster c, i.e. $h^{\mathrm{T}}_{\mathbf{s}c}$ in eq.(3.17), have to be evaluated. Setting $q_{\mathbf{s}c} = \mathrm{E}[\delta(c, \mathbf{m}(\mathbf{s}))] = p(\delta(c, \mathbf{m}(\mathbf{s})) = 1)$ and utilizing the well known fact from statistical physics that the generalized free energy at a certain temperature is minimized by the corresponding Gibbs distribution, i.e. $q_{\mathbf{s}c} \propto \exp(-\frac{1}{T} h^{\mathrm{T}}_{\mathbf{s}c})$, one arrives at the following set of update equations:

$$q_{\mathbf{s}c} \propto \exp\left(-\frac{1}{T} \sum_{\tilde{c}}^{k} T_{c\tilde{c}} h_{\mathbf{s}\tilde{c}} \right). \tag{3.18}$$

TPDC M-Step–Equations: In the M-step, one has to solve the stationary equations for the model parameters p_c and $p_{\alpha|c}$, given the estimated assignment probabilities $q_{\mathbf{s}c}$. For p_c, a closed-form solution is obtained by taking the derivative of eq.(3.16) while introducing Lagrange parameters to ensure a proper normalization:

$$p_c = \frac{1}{n} \sum_{\mathbf{s}=1}^{n} q_{\mathbf{s}c}\ , \quad c = 1, \ldots, k. \tag{3.19}$$

The second set of parameters $p_{\alpha|c}$ is updated by the iterative selection of pairs of Gaussian components α_1, α_2, as has already been described in the previous sections on the M-Step equations for the other PDC variants. Fixing $p_{\gamma|c}$ for all $\gamma \notin \{\alpha_1, \alpha_2\}$, the optimal value of $p_{\alpha_1|c}$ is determined by solving the following

equation via interval bisection:

$$-\sum_{c=1}^{k}\sum_{j=1}^{m}\left(\sum_{\mathbf{s}=1}^{n} q_{\mathbf{s}c}\mathbf{n}_{\mathbf{s}j}\right)\frac{G_{\alpha_1}(j)-G_{\alpha_2}(j)}{\sum_{\gamma=1}^{l} p_{\gamma|c}G_{\gamma}(j)}=0. \tag{3.20}$$

If no unique optimum can be found in this way, i.e. it is outside the interval $\left[0;1-\sum_{\gamma\notin\{\alpha_1,\alpha_2\}},p_{\gamma|c}\right]$, the mixture coefficient is equal to one of the interval boundaries. Once $p_{\alpha_1|c}$ has been determined, $p_{\alpha_2|c}$ can be computed from

$$p_{\alpha_2|c}=1-\sum_{\gamma\notin\{\alpha_1,\alpha_2\}} p_{\gamma|c}-p_{\alpha_1|c}\,. \tag{3.21}$$

3.2.3 Experimental Results

So far, the principle *ansatz* for PDC with topological constraints in cluster-space has been outlined. For concrete experiments, however, the topology has to be given by specifying a coupling matrix $T_{c\tilde{c}}$. Clearly, there are numerous possible topologies which could be explored. For the sake of conciseness in description, two choices have been made for the experiments which are presented in this section. One is a chain-like relationship $T_{c\tilde{c}}^{\text{chain}}$ between the clusters, the other is a circular structure $T_{c\tilde{c}}^{\text{circle}}$, which couples the data-group with the highest index back to the one with the lowest. These coupling structures are made mathematically precise by the following definitions for the cluster confusion matrices:

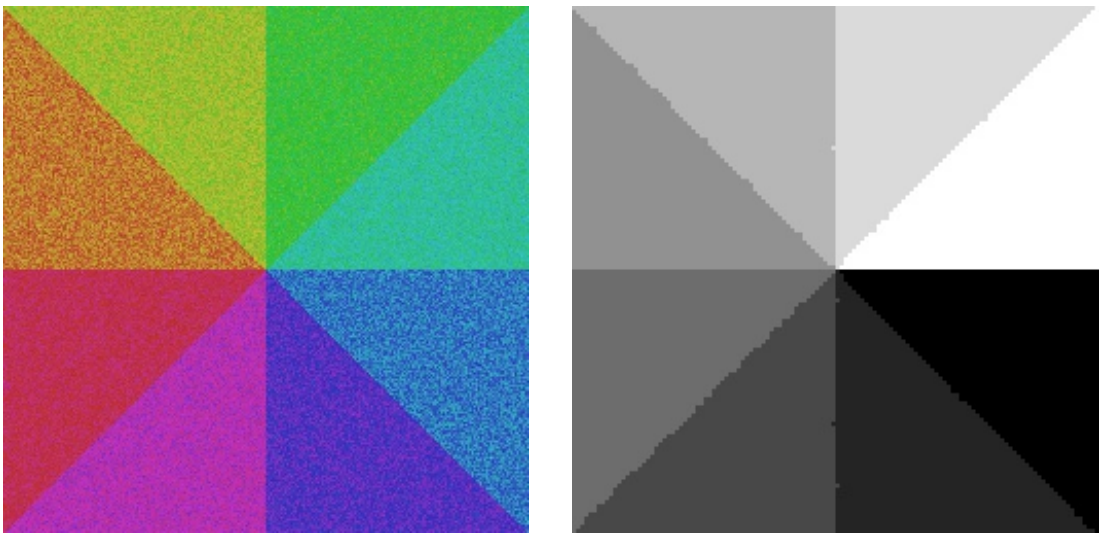
$$T_{\tilde{c}c}^{\text{chain}}=\begin{pmatrix}1-\epsilon & \epsilon & 0 & \dots & 0\\ \epsilon & 1-2\epsilon & \epsilon & \ddots & \vdots\\ 0 & \ddots & \ddots & \ddots & 0\\ 0 & \ddots & \ddots & \epsilon & 1-\epsilon\end{pmatrix}, \tag{3.22}$$

and

$$T_{\tilde{c}c}^{\text{circle}}=\begin{pmatrix}1-2\epsilon & \epsilon & 0 & \dots & \epsilon\\ \epsilon & 1-2\epsilon & \epsilon & \ddots & \vdots\\ 0 & \ddots & \ddots & \ddots & 0\\ \epsilon & \ddots & \ddots & \epsilon & 1-2\epsilon\end{pmatrix}. \tag{3.23}$$

In both of the equations given above, the topological structure is determined by the layout of the coupling matrices. The strength of the coupling, however, can be influenced by the parameter ϵ, which provides for an additional degree of freedom in the specification of the concrete TPDC cost function.

Chain Topology: In order to evaluate TPDC for the chain structure in cluster-space, i.e. TPDC utilizes the matrix T^{chain} for the specification of the topological structure, a segmentation result for an artificial data set is examined first. Due to the fact that this first test should be kept simple, it is chosen to work with color features only. The image data itself has been specified in HSV color space, because it is linear in the sense that perceptual color differences correspond to Euclidean distances of the color vectors, which does not hold true for the RGB color space. This linearity alleviated the design of a generative data model with a certain inherent neighboring structure in terms of the usual Euclidean distances between the modes of the Gaussian alphabet, which is at same time meaningful in terms of perceptual differences. The data set consists of eight groups, such that neighboring segments in the image have colors generated by neighboring mixture distributions in color space. To be precise, the H channel is divided into 24 equally spaced angles while keeping the S and V channels fixed to 0.75. Thus one arrives at a set of 24 color vectors, which defines the alphabet of Gaussians for the image. The image itself is divided into eight regions. For each of these regions, a prototypical mixture model is given by taking three consecutive Gaussians from the underlying alphabet with equal weighting. Any two consecutive mixtures thus have two Gaussians lying very closely together in feature space. With this construction, the topological structure for the data groups is given by a chain. From this model, an image is generated by sampling from the region specific mixture distributions. The result of this procedure is depicted on the left hand side of fig. 3.5. Here, the input image is opposed to the segmentation result of



(a) (b)

Figure 3.5: TPDC applied to artificial test-data: a) image data, b) segmentation.

TPDC, where the parameter for the strength of the topological coupling has been set to $\epsilon = 0.08$. For the visualization of the image partition, group memberships are encoded by grey values in such a way that two consecutive indices have consecutive shades of grey. As can be seen from the grouping result, the chain topology of the clusters has been correctly identified.

Encouraged by this positive result on the artificially created data, TPDC has been applied to a set of 100 images from the Corel Image Gallery taking both color and texture features into account. One of the corresponding segmentations is depicted in fig. 3.6. Others can be found in [ZB02]. For this particular clustering solution, the number of data-groups has been fixed to four while limiting the set of Gaussian modes to 16 and setting $\epsilon = 0.1$. The resulting segments are grey level coded like in the case of the artificial data set, thus visualizing the topological structure in cluster space. The first segment covers the cat and the part of the tree with low texture content. Following in the chain topology of clusters is the segment in light grey, covering the rest of the tree (with more pronounced texture). Then comes the segment of the smaller branches of the tree above and below the cat, depicted in dark grey. Due to the finite support of the local histograms, the corresponding region statistics contain both contributions from the branches themselves and from the blue sky in the background. Topologically, this segment should lie between the tree and the sky segment, which is exactly what our algorithm yields. Finally, there is the sky segment in black, which lies topologically correct at the other end of the inferred chain structure. This interpretation is backed up by the underlying prototypical mixture models of the various groups. In order to visualize the distributions, the fact has been utilized that keeping the alphabet of Gaussians fixed, the mixtures are characterized completely by their coefficients $p_{\alpha|c}$, which are depicted in the diagrams. The visually apparent topological relations between the inferred data groups can be easily verified by looking at salient correspondences in the distributions, which are depicted by arrows.

Circular Topology: For the segmentation experiments using TPDC to infer a circular topology in cluster-space as defined by the coupling matrix given in eq.(3.23), the settings for the data acquisition and pre-processing stages were kept fixed as described in section 2.7.1, while fixing the coupling strength to $\epsilon = 0.08$. Moreover, the Berkeley image data-base has been used as the test-bed for this particular variant of TPDC. In order to provide a qualitative overview concerning the influence of the topological coupling on the arrangement of clusters in index space, fig. 3.7 shows two representative image partition solutions with five clusters. In the left column of this figure, outline visualizations of the segmentations are shown. The right column shows the underlying TPDC clustering solutions in form of color-coded images. The color code itself is explained by the schematic in the first row of fig. 3.7, which also shows the topological relations between

the clusters. For the depicted images, texture is negligible. The first example is a portrait image. The dominating colors are brownish hues of varying lightness, while the hat and throat areas show a darker tone, which in case of the hat has a reddish touch. The relations between the color distributions of the inferred segments are perfectly represented in the topological structure in index space. The segment marked in black in the color-coded visualization of the segmentation result mostly covers areas in between the ecru parts of the image (blue cluster) and the dark regions (yellow cluster). Therefore, a mixture statistic of these image parts is predominant for the black cluster. For the data-groups marked in blue, green, red, and yellow, the lightness of the brown hues, which are covered by the respective regions in the picture, is continually decreasing, making the corresponding clusters perfect neighbors in the sense of the topological coupling. The second picture in fig. 3.7 shows a mushroom on forest soil. Here, the first cluster covers the image areas of a light brown hue tending to ecru, whereas the blue segment, which immediately follows in the topological ordering mainly covers the ecru parts. The next segment in the topological structure is the one which is color coded in green. The characteristic property of the corresponding image region is its dark brown color. The similarity to the preceding segment is given by the fact that the brown hue is related to ecru, while the link to the following segment is visible in terms of the lower lightness. The next cluster (shown in red) primarily covers the darkest regions of the picture, while also containing the overshadowed parts of the moss in the image background. Finally, the cluster, being depicted in yellow, contains the regions which cover the lighter part of the moss. This increased lightness then justifies the closing of the circular topological structure to the black segment.

For the quantitative performance evaluation of the circularly coupled TPDC, segmentation experiments for the Berkeley test-bed have been run for three, five and eight clusters in the TPDC data-grouping. For all cluster numbers, the coupling strength has been fixed to $\epsilon = 0.08$. The achieved recall, precision and F-measure values are summarized in table 3.2. Corresponding box-plots are given in the appendix. As this table clearly shows, the segmentation performance indicators are not better than in the case of topologically unconstrained PDC. One thus has to conclude that the introduction of couplings between clusters is not a viable means of progressing towards image segmentation methods which come closer to human performance standards for this task. On the other hand, the examples presented in this section amply demonstrate that the similarity between the inferred clusters in terms of their feature distributions is faithfully reflected in the index space of cluster numbers. This additional information which supplements the mere segmentation solution might prove very valuable for potential application contexts.

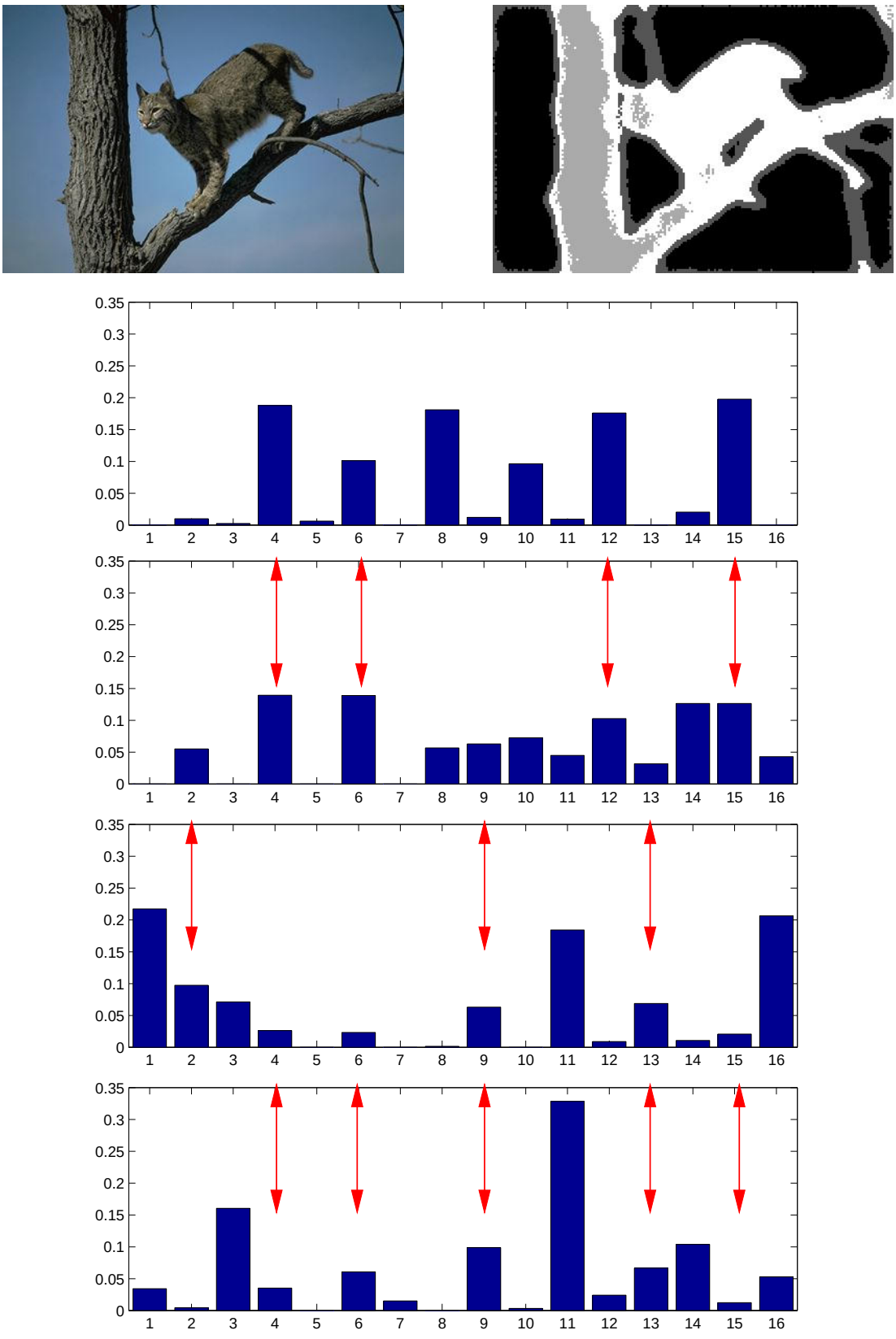


Figure 3.6: TPDC with chain topology applied to real-world data: Input image, segmentation, and model parameters.

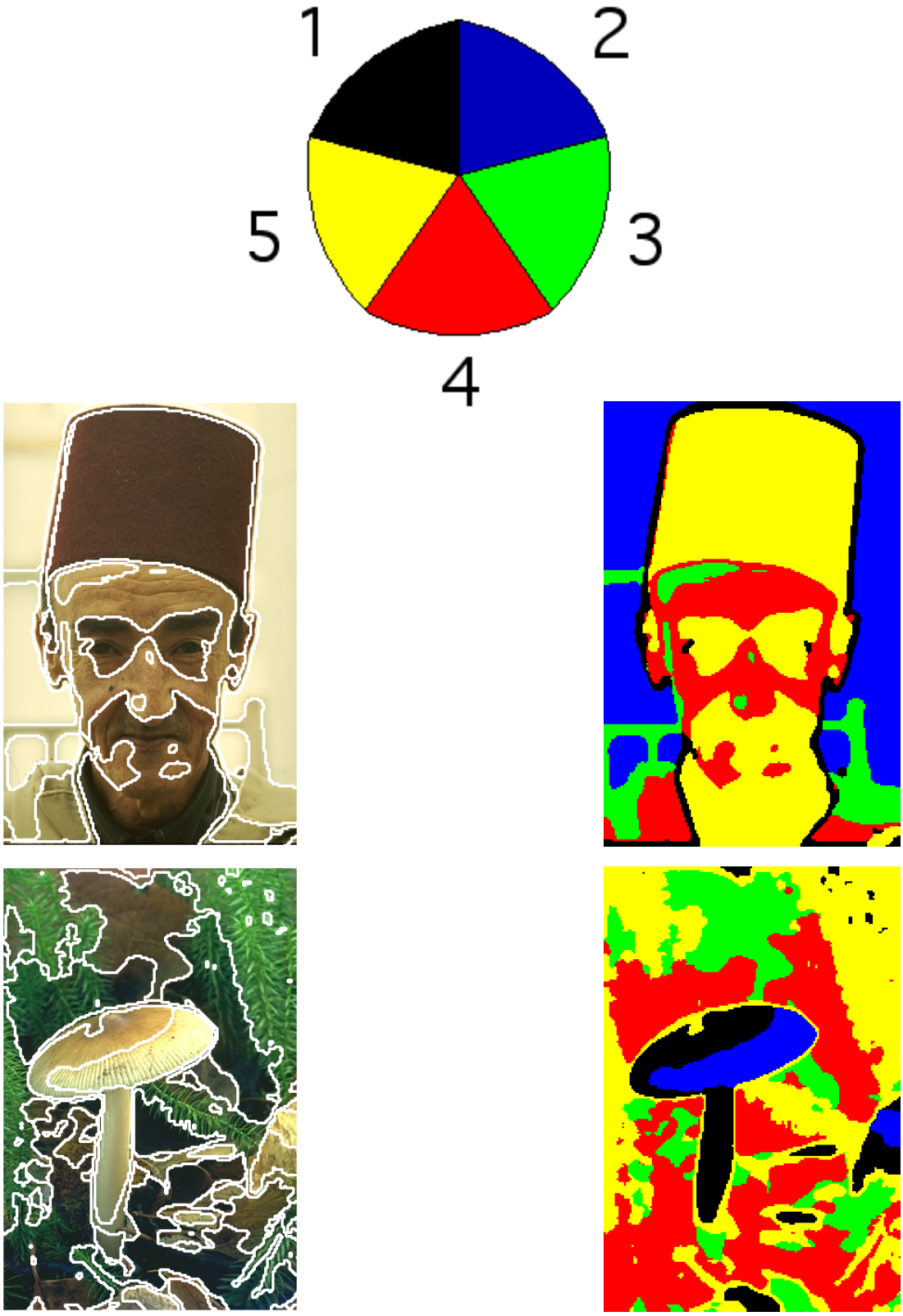


Figure 3.7: TPDC segmentation results with five clusters. a) outline visualization, b) color coded cluster assignments.

# clusters	$\hat{\mathrm{E}}[\rho]$	$\hat{\mathrm{D}}[\rho]$	$\hat{\mathrm{E}}[\pi]$	$\hat{\mathrm{D}}[\pi]$	$\hat{\mathrm{E}}$[F-value]	$\hat{\mathrm{D}}$[F-value]
3	53.0	11.4	49.1	15.4	49.1	10.6
5	67.3	11.3	43.6	13.9	51.3	11.5
8	78.0	9.2	38.6	12.0	50.2	11.0

Table 3.2: Recall, Precision and F-value summary for TPDC segmentation experiments with three, five and eight clusters. Values have been multiplied by 10^2 for better readability.

3.3 Combining Spatial and Group Topology

After having introduced topological constraints in the spatial domain as well as in the cluster realm, it suggests itself to merge these two separate approaches into a variant of PDC that equally enforces both types of topological couplings. Thereby, one can combine the advantage of an explicitly given topology in the space of feature distributions with the increased precision of the spatially smoothed segmentations.

3.3.1 The sTPDC Model

In principle, the construction of the cost function for Parametric Distributional Clustering with topological coupling of data-groups as well as coupling in the spatial domain is analogous to the derivation of sPDC in section 3.1.1. Here, however, the TPDC objective function provides the starting point:

$$\begin{aligned} \mathcal{H}^{\mathrm{TPDC}} &= \sum_{\mathbf{s}=1}^{n} \sum_{c}^{k} \delta(\mathbf{m}(\mathbf{s}), c) h^{\mathrm{T}}, \quad \text{where} \\ h^{\mathrm{T}} &= \sum_{\tilde{c}}^{k} T_{c\tilde{c}} h_{\mathbf{s}\tilde{c}}. \end{aligned} \tag{3.24}$$

In the equation above, the notations, which have already been introduced before are used again. I.e. h^{T} denotes the local assignment cost for TPDC, while $h_{\mathbf{s}\tilde{c}}$ stands for their counterparts in the basic PDC model. Finally, $T_{c\tilde{c}}$ denotes the matrix of coupling strengths between the clusters. The spatial coupling of assignments can now be encouraged as before by setting

$$\mathcal{H}^{\mathrm{s}} = \sum_{\mathbf{s}=1}^{n} \sum_{c=1}^{k} \delta(\mathbf{m}(\mathbf{s}), c) \lambda^{\mathrm{t}} \sum_{\tilde{\mathbf{s}} \in \mathcal{N}_{\mathbf{s}}} (1 - \delta(\mathbf{m}(\tilde{\mathbf{s}}), c)). \tag{3.25}$$

The two types of topological constraints are combined in the following way:

$$\begin{aligned}\mathcal{H}^{\mathrm{sTPDC}} &= \mathcal{H}^{\mathrm{TPDC}} + \mathcal{H}^{\mathrm{s}} \\ &= \sum_{\mathbf{s}=1}^{n}\sum_{c}^{k}\delta(\mathbf{m}(\mathbf{s}),c)h^{\mathrm{T}} + \sum_{\mathbf{s}=1}^{n}\sum_{c=1}^{k}\delta(\mathbf{m}(\mathbf{s}),c)\lambda^{\mathrm{t}}\sum_{\tilde{\mathbf{s}}\in\mathcal{N}_{\mathbf{s}}}(1-\delta(\mathbf{m}(\tilde{\mathbf{s}}),c)) \\ &= \sum_{\mathbf{s}=1}^{n}\sum_{c=1}^{k}\delta(c,\mathbf{m}(\mathbf{s}))\underbrace{\left(h^{\mathrm{T}}_{\mathbf{s}c} + \lambda^{\mathrm{t}}\sum_{\tilde{\mathbf{s}}\in\mathcal{N}_{\mathbf{s}}}(1-\delta(\mathbf{m}(\tilde{\mathbf{s}}),c))\right)}_{:=h^{\mathrm{tT}}_{\mathbf{s}c}}. \end{aligned} \tag{3.26}$$

3.3.2 sTPDC Model Identification

The optimal values for the free parameters in the sTPDC model are found according to the scheme which has already been used to derive the update equations of PDC, sPDC, and TPDC. Again, one starts from the generalized free energy functional of the problem at hand. In the case of sTPDC, one has

$$\mathcal{F}^{\mathrm{sTPDC}} = \mathrm{E}[\mathcal{H}^{\mathrm{sTPDC}}] - T\cdot H. \tag{3.27}$$

Denoting the assignment probabilities by $q_{\mathbf{s}c} = \mathrm{E}[\delta(c,\mathbf{m}(\mathbf{s}))] = p(\delta(c,\mathbf{m}(\mathbf{s}))=1)$, the expected cost are given by

$$\begin{aligned}\mathrm{E}[\mathcal{H}^{sTPDC}] &= \sum_{\mathbf{s}=1}^{n}\sum_{c=1}^{k} q_{\mathbf{s}c}h^{\mathrm{tT}}_{\mathbf{s}c} \\ &= \sum_{\mathbf{s}=1}^{n}\sum_{c=1}^{k} q_{\mathbf{s}c}\left(h^{\mathrm{T}}_{\mathbf{s}c} + \lambda^{\mathrm{t}}\sum_{\mathbf{s}\prime\in\mathcal{N}_{\mathbf{s}}}(1-q_{\tilde{\mathbf{s}}c})\right). \end{aligned} \tag{3.28}$$

sTPDC E-Step–Equations: The assignment probabilities $q_{\mathbf{s}c}$ are found by evaluating the Gibbs distribution corresponding to the local costs of the sTPDC model. Consequently, one arrives at

$$q_{\mathbf{s}c} \propto \exp\left(-\frac{1}{T}\left(h^{\mathrm{T}}_{\mathbf{s}c} + \lambda^{\mathrm{t}}\sum_{\tilde{\mathbf{s}}\in\mathcal{N}_{\mathbf{s}}}(1-q_{\tilde{\mathbf{s}}c})\right)\right). \tag{3.29}$$

sTPDC M-Step–Equations: For the second order mixture coefficients p_c, update equations in closed form can be derived by differentiating eq.(3.27). Again, Lagrange parameters have to be introduced in order to enforce a proper normalization. This procedure leads to

$$p_c = \frac{1}{n}\sum_{\mathbf{s}=1}^{n} q_{\mathbf{s}c}\ ,\ c=1,\ldots,k. \tag{3.30}$$

The coefficients of the Gaussian mixture models of the various clusters, $p_{\alpha|c}$, can again only be computed via a numeric optimization step. To this end, pairs of coefficients $p_{\alpha_1|c}$ and $p_{\alpha_2|c}$ are repeatedly selected. The optimal value of $p_{\alpha_1|\nu}$ is characterized by

$$-\sum_{c=1}^{k}\sum_{j=1}^{m}\left(\sum_{\mathbf{s}=1}^{n} q_{\mathbf{s}c}\mathbf{n}_{\mathbf{s}j}\right)\frac{G_{\alpha_1}(j)-G_{\alpha_2}(j)}{\sum_{\gamma=1}^{l} p_{\gamma|c}G_{\gamma}(j)}=0, \tag{3.31}$$

where the root is again determined by interval-bisection. If no unique optimum can be found inside the interval $\left[0;1-\sum_{\gamma\notin\{\alpha_1,\alpha_2\}} p_{\gamma|c}\right]$, $p_{\alpha_1|\nu}$ is equal to one of the interval boundaries. Once the optimal value of $p_{\alpha_1|c}$ has been computed, $p_{\alpha_2|c}$ must satisfy

$$p_{\alpha_2|c}=1-\sum_{\gamma\notin\{\alpha_1,\alpha_2\}} p_{\gamma|c}-p_{\alpha_1|c}\,. \tag{3.32}$$

3.3.3 Experimental Evaluation

In order to ensure the comparability of experimental results, the settings for the data acquisition and pre-processing stages have been kept fixed as described in section 2.7.1. Moreover, the practice of computing image partitions based on data groupings with three, five and eight clusters has been continued. For the topology in cluster space, the circular structure defined by $T^{\text{circle}}_{\tilde{c}c}$ in eq.(3.23) has been used with $\epsilon = 0.08$. Due to the fact that the influence of the spatial coupling as well as the cluster ordering according to the enforced topological constraints has already been outlined in the previous sections, the presentation of qualitative segmentation results for sTPDC is limited to the comparison with those of the human subjects of the Berkeley study. To this end, the figures 3.8 and 3.9 offer two collections of representative results for sTPDC segmentations and their counterparts from the set of human generated results.

(a) (b) (a) (b)

Figure 3.8: a) sTPDC segmentations with three (left) and eight (right) data groups, b) overlaid segmentations of the human subjects.

In the first picture of fig. 3.8 one has a striking example of the phenomenon that segmentations motivated by the low-level feature content and those arrived at by semantic considerations do not necessarily agree. Although the sTPDC result based on a grouping with five clusters captures the differences in terms of feature content very well while avoiding fragmented over-segmentation of image details due to the spatial coupling of assignments, it does not infer the statue, let alone its sub-parts, as independent regions. For the second picture, feature- and semantic-based image partition are in much better accordance. Here, the depicted sTPDC segmentation is based on a cluster-model with eight data groups. Nevertheless, the semantically motivated sub-structure in the human segmentation (e.g. separating architecturally different parts of the building in the background and the segregation of the group of flowers in individual blossoms and stems) can not be reproduced by sTPDC, simply because it is not supported by the low-level feature information.

(a) (b)

Figure 3.9: a) sTPDC segmentations, b) overlaid segmentations of the human subjects.

In fig. 3.9, the first two sTPDC results are based on data-groupings with five clusters, while the last image has been partitioned according to a sTPDC model with eight data-groups. Again, the overall agreement with the human segmentations is very good. As in the previous examples, however, minor differences of sTPDC as compared to the human segmentations are also visible in these examples.

In order to arrive at quantitative measurements of sTPDC's image segmentation performance, the precision, recall, and F-measure determination for data partitions with three, five, and eight clusters has been conducted for all of the 100 images in the test-bed. The pertaining results are summarized in table 3.3, whereas a box-plot of the performance indicator distributions is shown in the appendix. Here, one can note again the decrease in recall compared to the spatially unconstrained case of TPDC. On the other hand, the precision values are noticeably better. As in the case of sPDC vs. standard PDC, the distributions for the precision values at the various cluster granularities are not completely separated due to the comparatively large variances. One, therefore, has to resort to the technique that has already been used in the determination of the significance of the precision increase in the case of sPDC as opposed to basic PDC. Consequently, the differences $\Delta\pi = \pi(sTPDC) - \pi(TPDC)$ have been computed for all inferred cluster numbers. Then, the t-test variant from the *Matlab* statistics toolbox is utilized again, which has already been described in section 3.1.3, in order to test against the null-hypothesis of zero-mean for the empirical distributions of difference values. For all numbers of clusters, the null-hypothesis could be rejected at a significance level of 0.01, proving that the incorporation of spatial topological constraints has again increased the performance in terms of the precision measure. The empirical distributions of the precision differences are shown in fig. 3.10, visually demonstrating that the vast majority of the probability mass is located above zero.

# clusters	$\hat{\mathrm{E}}[\rho]$	$\hat{\mathrm{D}}[\rho]$	$\hat{\mathrm{E}}[\pi]$	$\hat{\mathrm{D}}[\pi]$	$\hat{\mathrm{E}}$[F-value]	$\hat{\mathrm{D}}$[F-value]
3	41.0	11.7	45.0	12.7	41.6	9.7
5	49.3	11.4	44.4	12.7	45.5	10.1
8	54.6	13.1	43.4	12.3	47.4	10.5

Table 3.3: Recall, Precision and F-value summary for sTPDC segmentation experiments with three, five and eight clusters. Values have been multiplied by 10^2 for better readability.

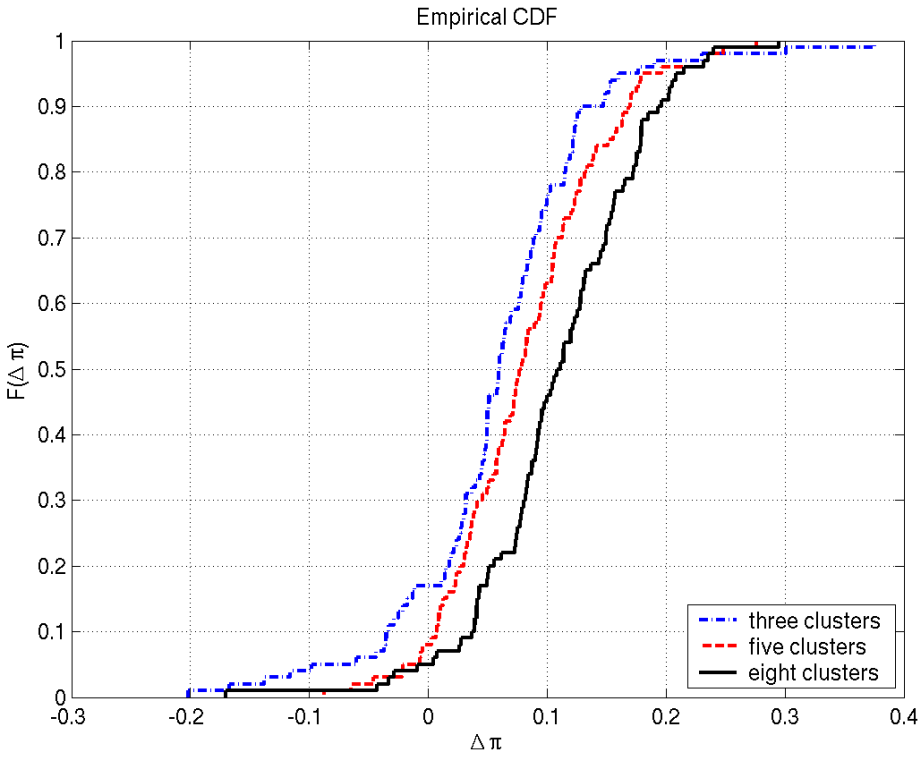


Figure 3.10: Empirical cumulative density functions of the precision differences between sTPDC and TPDC with three, five, and eight clusters.

3.4 Summary

In this chapter, three different variants of PDC with topological constraints have been introduced. The first one (sPDC) addressed the characteristic property of natural images to consist of large continuous regions by discouraging highly fragmented image partitions. For this method, the significant increase in performance has been demonstrated visually by the depiction of image partition results and statistically by evaluating the gain in terms of the precision quality measure.

The second derivative (TPDC) of the basic PDC approach explicitly accounts for the similarity between the Gaussian mixture models of different clusters by enforcing a topological coupling of data-groups. For the experiments described herein, two reasonable choices for the topology have been explored, namely a chain-like and a circular structure, while the coupling strengths have been fixed. An interesting direction for further investigations is given by the possibility to learn the topology of the cluster coupling together with the respective coupling strengths. As the strength of the coupling should be related to the difference of the group specific mixture models, choosing the former entities proportional to the Kullback-Leibler divergences between the feature distributions of the clusters

provides for a promising possibility. Concerning the quantitative performance evaluation, it has to be admitted that the achieved results stayed virtually the same as compared to standard PDC. Nevertheless, the correct identification of the neighborhood structure in cluster space can be a valuable additional information for some application contexts. Here, one might think of an image-database solution, where the user selects a segment and wants to quickly retrieve similar image parts. For standard PDC, KL-divergences between all inferred group models would have to be computed, for TPDC, a simple look-up of neighboring indices is sufficient.

Finally, both kinds of topological constraints were unified into a single framework. The corresponding experimental evaluation demonstrated by qualitative and quantitative means that the resulting sTPDC approach combines the good segmentation performance of spatially coupled sPDC with the explicit representation of cluster similarities.

3.5 Bibliographic Remarks:

The application of topological constraints in the spatial domain lies at the core of Markov random field approaches to image segmentation [GG84, DMZB99, KBZ96, SS02, SBK$^+$02]. Text books on this subject are [Li01] and [Win02]. The idea of self-organizing maps was first introduced by Kohonen [Koh81]. A comprehensive overview including the subsequent developments is provided by [Koh95]. Convergence properties of self-organizing maps are discussed in [YA98], while the application to clustering problems is the subject of [MJ96, BK93]. The TPDC approach as presented in this chapter, was first published in [ZB02].

Chapter 4

Robustness and Generalization

In the preceding two chapters, various methods for image segmentation by variants of parametric distributional clustering have been introduced. All these approaches share as a common trait that the segmentation solution is achieved by a clustering of image sites according to the similarities of the corresponding local feature histograms. Therefore, a clustering procedure and thus an unsupervised learning problem comprises the core of all the discussed segmentation methods. The quality of the proposed approaches has been demonstrated by examples as well as by the statistical evaluation of performance measures, comparing the machine generated image partitions to those chosen by human subjects. For many learning procedure, there exist, however, other quality criteria, which are not directly connected to the results on single instances of the processed datasets. In particular, the stability of the inferred clustering solutions against small variations in the processed data and the transferability of data groupings to new problem instances, i.e. the generalization ability, are widely considered in the machine learning community. In principle, these questions are relevant for all PDC variants introduced in this thesis. For conciseness of description, however, the focus of attention is put on the sPDC approach, as it has proven to be superior in terms of the quality measure evaluation. The problem of stability of sPDC with respect to small changes in the given image data will be treated in the first section of this chapter. Then, section 4.2 will deal with the transfer of segmentation solutions between images of similar content.

4.1 Bootstrap Resampling

4.1.1 The Resampling Strategy

The problem of over-fitting is of major importance for all machine learning tasks, regardless whether they are supervised, i.e. ground-truth label information is available, or unsupervised, i.e. one has to rely solely on the measured features. The learning procedure is supposed to infer the structural characteristics of the

dataset in question while avoiding to represent the statistical fluctuations, and thus the noise, which is invariably present in all measurements. The main assumption is, that the data at hand has been generated by a source which is characterized by a certain distribution law. Consequently, the measured feature information is considered to be the result of a sampling process from this distribution. Ideally, the learning algorithm would have access to that source, from which individual samples could be drawn in an arbitrary and thus possibly infinite amount. According to the law of large numbers, fluctuations are eventually averaged out, facilitating the estimation of the true structural properties. For most applications, however, this ideal situation is not given. On the contrary, the amount of data is usually limited and often enough rather scarce.

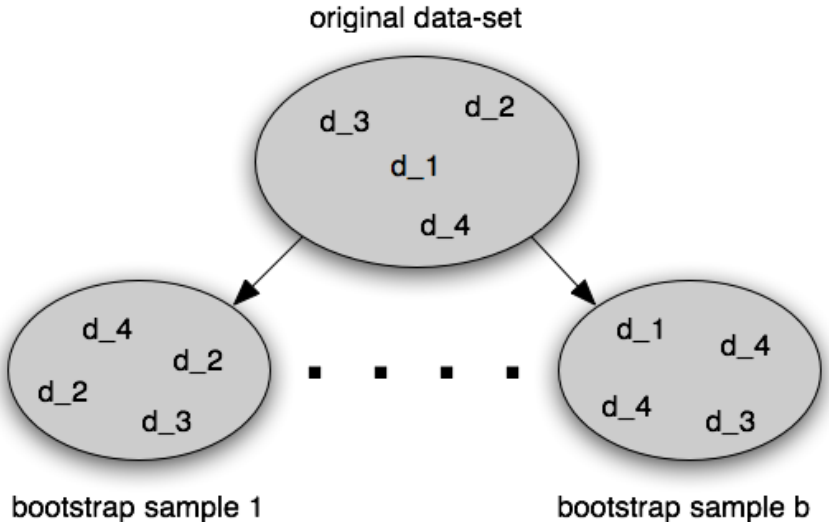


Figure 4.1: Visualization of the bootstrap sampling process.

In order to alleviate the problem, resampling techniques can be used to generate multiple instances of the given dataset at hand. One of the most prominent techniques in this respect is the bootstrap[1] method. Suppose one is given a data set $\mathbb{D}$ with n elements. Any new instance of the set, the bootstrap sample, is then created by drawing n items from $\mathbb{D}$ with replacement. Due to the fact that $\mathbb{D}$ itself contains only n items, there will almost always be duplications, while some of the items will be missing in the sampled dataset. The selection process is independently repeated B times in order to produce B bootstrap datasets. This procedure is visualized in fig. 4.1. These re-sampled datasets can then be utilized, for example, to improve the estimation of some statistic θ. In this case, the bootstrap estimate $\hat{\theta}^*$ is simply the mean of the B individual estimates of the

[1]The name "bootstrap" comes from subsequent versions of Rudolph Erich Raspe's novel "The Adventures of Baron Münchhausen". In one of these tales, the hero was able to pull himself up onto his horse by lifting his bootstraps.

statistic $\hat{\theta}^{*(b)}$ based on the bootstrap samples, i.e.:

$$\hat{\theta}^{*} = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^{*(b)}. \tag{4.1}$$

Our goal is to utilize the bootstrap framework in order to assess the stability of segmentation solutions generated by the sPDC approach with respect to variations in the input image data. To this end, the bootstrap sampling technique has to be adapted to the case of image data. Here, the direct application of resampling by drawing with replacement cannot be applied. This is due to the fact that some of the pixels will be drawn more than once, while others are not chosen at all. Therefore, the basic bootstrap sampling scheme will lead to images in which a certain fraction of pixels that are not set (i.e. blank, or black pixels). Based on such image data, a reasonable application of sPDC, or any other segmentation technique, is impossible. In order to remedy this adverse situation, it is proposed to utilize a two stage sampling process. In the first step, the basic bootstrap resampling technique is applied. Then, in the ensuing second stage, the holes in the sampled image data are filled by a local re-shuffling of pixels. To be precise, for any empty pixel, a replacement is randomly drawn from the 3×3 vicinity of that pixel in the original image data. A visual description of this procedure can be found in fig. 4.2.

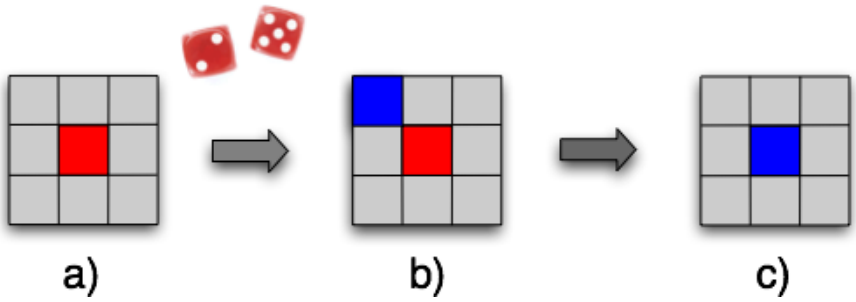


Figure 4.2: Selection of replacements for pixels which have been left empty in the first bootstrap resampling stage: a) empty pixel (red), b) a replacement has been determined via sampling from the vicinity in the original image (blue), c) the gap in the bootstrap-sample is closed by the chosen replacement.

4.1.2 Evaluation Results

In order to gain insight in the stability of sPDC solutions, a collection of seven example images from the Berkeley test data-base has been selected. For each of these pictures, 100 samples have been generated by means of the bootstrap-based strategy outlined in the previous section. Then, sPDC with five clusters has been applied to all of these images. The segmentation solutions for each

image were aggregated into a probabilistic edge map for that particular image in question. This has been achieved by counting the number of times that any pixel has been located at a boundary between different image regions and dividing this sum by the total number of resampling instances. Consequently, those edges which turned out to be most stable across the samples have received the highest probability mass. As one can see from eq.(4.1), the averaging procedure is in complete analogy to the generic derivation of bootstrap estimates of statistics.

The achieved results are shown in figures 4.4 to 4.10. These depictions are structured as follows: On the left hand side of the top row, the overlaid segmentations of the human subjects from the Berkeley evaluation study are shown. Next to it, a visualization of the edge probabilities is given, in which higher probability mass is indicated by darker tones of grey. Below these images, a graph depicting the achieved recall, precision and F-measure values is shown. To arrive at this graph, the threshold concerning the normalized frequency, from which on an edge is counted in the evaluation process, is varied between the extremal values zero and one. Thus, the curve is parametrized by edge probability threshold. The best F-measure value along the graph together with the corresponding recall and precision values at which it has been achieved and the pertaining threshold are shown in the legend. The general behavior of precision and recall as functions of the detection threshold t and the mutual relationship between these functions has been extensively studied in the information science literature (e.g. [BG94, GK89]). Generally, $\rho(t)$ monotonously decreases, while $\pi(t)$ monotonously increases as t approaches 1 (c.f. [Zhu04] for a nice derivation of this connection). Consequently, one would expect a monotonous decrease of the combined precision recall graphs in the result figures. As the graphs clearly show, this presumption is violated in some of the presented examples. Unfortunately, the cause of this irregularities could not be determined. The evaluation of precision and recall relies on code that is provided as part of the utilized test suite. Apart from the fact that crucial parts of the code are undocumented and thus hard to read and understand, altering the code amounts to a deviation from the public test platform. Therefore, this route was not pursued as it would have endangered the comparability of the achieved results.

The first example picture (given in fig. 4.4) shows a pile of corn on a blanked. To its right, one can find a bowl of other fruits, whereas the lower part of the image is dominated by some small twigs on the soil. As the aggregated edge image clearly shows, the most stable boundaries are those between the most dissimilar image parts. In particular, the border of the corn pile, the outline of the large bowl on the right, the shadowed regions, the area of the small branches at the bottom of the image, and the soil segments on the upper right and the left of the picture produce very stable segment boundaries across all segmentations of the re-sampled images. On the other hand, the fine structure of the blanket, on which the corn is piled up is mostly averaged out, indicating that the sPDC segmentations produced differing segmentations on the different bootstrap in-

stances of the picture. Table 4.1 compares the best F-measure value achieved by resampling and the pertaining recall and precision values with the corresponding results of a sPDC segmentation with five clusters on the original input picture. This table documents, that the resampling approach has been able to rise all of the performance measures slightly but noticeably.

	ρ	π	F-measure
single sPDC segmentation	0.56	0.61	0.59
aggregated bootstrap edges	0.58	0.68	0.63
gain (%)	4	11	7

Table 4.1: Resampling evaluation summary for image 58060.

The next image (fig. 4.5) shows three piglets on muddy, dark brown soil. For this picture, the stable edges for the segmentation of the far background are in excellent accordance with those arrived at by the human subjects. Opposed to this, the piglets are not reliably inferred as independent segments in the sPDC solutions for the re-sampled image instances. On the contrary, sPDC seems to prefer a segmentation, in which the upper light pink part of the piglets is separated from the lower parts of the animals, which show a darker hue of pink due to dirt and shadows. Moreover, the hard shadows below the animals are reliably inferred as separate image parts. This partition cannot be found in the human segmentations, as shadow seems not to be a semantically relevant category for the individuals who produced the reference image partitions. A comparison of the performance measures for the aggregated bootstrap edges and the single sPDC segmentation of the discussed image can be found in tab. 4.2. While the recall quality stayed the same, the precision increased 17 %. As a consequence, the F-measure improved by 7 %.

	ρ	π	F-measure
single sPDC segmentation	0.58	0.53	0.56
aggregated bootstrap edges	0.58	0.62	0.60
gain (%)	0	17	7

Table 4.2: Resampling evaluation summary for image 66053.

The third example, shown in fig. 4.6, poses a very hard segmentation problem for sPDC. Here, a leopard is depicted sitting in a tree amidst a maze of branches and twigs. The color impression of this image is dominated by grey hues ranging from lighter to darker shades. Moreover, the fur of the animal and the branches

generate rather similar responses of the texture filters. Consequently, the segmentations of the bootstrap instances of the picture show a large extend of variation. Nevertheless, the outlines of the major branches turn out to be comparatively stable across the various segmentations. The aggregated edges along the boundary of the animal receive only medium probability values, demonstrating, that a considerable amount of insecurity exists concerning the inference of the leopard segment. This situation is even aggravated for the smaller twigs of the tree and the parts of the sky in the background that show through the branches. These adverse conditions are reflected in the achieved performance measures for the aggregation result which are compared to the corresponding values for the single sPDC segmentation of the original image in tab. 4.3. Opposed to the other results for the re-sampled images, the recall is increased, while one has to acknowledge a decrease in the precision performance measure. Nevertheless, the summary measure of the F-value is still increased compared to the single sPDC segmentation.

	ρ	π	F-measure
single sPDC segmentation	0.53	0.55	0.54
aggregated bootstrap edges	0.65	0.51	0.57
gain (%)	22	-7	6

Table 4.3: Resampling evaluation summary for image 134035.

In the fourth image (fig. 4.7), that has been chosen to demonstrate the effects of bootstrap-based resampling strategies in image segmentation, one can see a group of zebras in front of a savannah background. Here, the difference between the foreground objects and the background is much more pronounced than in the previous example. However, the segmentations of the human subjects in the Berkeley study are rather peculiar in my view. Some of the bigger stripes in the fur pattern of the zebras are treated as individual image parts, while the different regions of the background (i.e. grassy soil and small trees) are merged into one large segment. As the aggregation result clearly demonstrates, the outlines of the animals are reliably inferred. This is especially valid for the upper segment borders along the back of the zebras. The separation of the top of the image, which is due to its dark color, is also very stable across all re-sampled image instances. On the other hand, some variations in the segmentation of the medium range background, containing some bushes and small trees is apparent from the aggregated edges, while other segregations in this image region turn out to be comparatively stable. Moreover, there are some boundaries in the interior of the zebra image part, which also receive a considerable amount of probability mass. These regions, covering major parts of the animal's bodies, are separated from the legs, because the horizontal stripe pattern of the limbs changes

into a vertical pattern with lower frequency on the torso. Therefore, the set of texture filters which respond most strongly to these patterns is different, thereby justifying a partition into separate segments. Again, the achieved performance measures on the aggregated edges are compared to their counterparts from a single sPDC segmentation in tab. 4.4. Here, one observes an improvement of all quality measures. Most remarkably, however, is the increase in precision of 52%. Together with the slightly improved recall, this leads to a raise of the summary F-measure of 28%.

	ρ	π	F-measure
single sPDC segmentation	0.54	0.40	0.46
aggregated bootstrap edges	0.57	0.61	0.59
gain (%)	6	52	28

Table 4.4: Resampling evaluation summary for image 253027.

The fifth example, shown in figure 4.8, consists of a picture which shows two elephants standing on dry grass-land. The upper half of the image is covered by blue sky. Between the soil and the sky, one can see a forest in the far background. The separation of that forest from the sky as well as from the grass is reliable inferred in the majority of sPDC segmentations of the re-sampled image instances. Moreover, the animals are well separated from the other parts of the image, as one can see from the high edge probabilities at the outline of the group of elephants. In the interior of that region, stable borders can be found between shadowed, and thus darker areas, and those parts which exhibit a slightly lighter shade of brown. Obviously, there has been some amount of variation concerning the exact localization of the upper part of that segment's outline, as one can see from the fuzzy edge probabilities. Furthermore, on can observe that no stable segmentation of the grass region directly below the animals has been found, which is indicated by the lack of pronounced edge probabilities in that area. The results of the quantitative performance comparison between the probabilistic edge map and the single sPDC segmentation with five clusters of the original input image from the test-set is given in tab. 4.5. Here, one has to note a decrease in terms of the recall of 10%. This is contrasted to a gain of the precision performance measure of 33%. Overall, this amounts to an increase of the F-measure value of 10%.

On the next image that has been evaluated (fig. 4.9), a camel rider is depicted in front of a pyramid. The upper part of the background is given by a mildly clouded sky, whereas the lower regions are dominated by the depiction of desert sand. The large dissimilarity between the light brown hues of the sand and the pyramid, and the bluish grey of the sky leads to a very stable separation of these image parts. Moreover, the rider and his animal are segregated from the

	ρ	π	F-measure
single sPDC segmentation	0.73	0.55	0.63
aggregated bootstrap edges	0.66	0.73	0.69
gain (%)	-10	33	10

Table 4.5: Resampling evaluation summary for image 296059.

other areas with high reliability. Only at the legs, which are very similar in color impression to the surrounding darker parts of the sand, one can observe some decrease in terms of edge probability. The inference of the shadow as an independent segment also shows a comparatively large amount of stability. Opposed to this, the separation of the sky region into clouded and unclouded parts, as can be seen in the human segmentations, could not be reliably reproduced by sPDC. Apart from this, the separation of the pyramid from the desert sand, which is visible in the image partitions of the subjects in the Berkeley segmentation study, also could not be achieved by sPDC, as it is purely semantically motivated. The same holds true for the identification of the small rocks on the sand as image parts in their own right. Comparing the quantitative performance measures, one has to observe a decrease of recall as large as 4%. This is contrasted, however, by the highly remarkable gain of 63% in terms of the precision quality measure. The joint criterion of the F-value consequently increases by 24%.

	ρ	π	F-measure
single sPDC segmentation	0.57	0.52	0.54
aggregated bootstrap edges	0.55	0.85	0.67
gain (%)	-4	63	24

Table 4.6: Resampling evaluation summary for image 299086.

Finally, the last example depicted in fig. 4.10, shows two polo-players hunting after a ball, which can be seen in the lower left corner of the image. In the background, there is a small house on the left, surrounded by dark green forest. At the very top of the picture, a small strip of sky is also visible. As one can see from the pertaining edge probability map, both players together with their horses are reliably segmented. Furthermore, the shadows below the animals are identified as image parts in the majority of sPDC segmentations. The white ball also comprises a very stable segment. Apart from this, the separation between the dark green forest region and the grass area has a large probability too. The upper boundary of the forest region is separated from the sky with a somewhat lesser, but still comparatively high probability. As far as the house in the back-

ground is concerned, the white areas around the windows are reliably inferred as independent segments. Opposed to this, the distinction between the roof and the forest is much less stable. This finding is explained by the large similarity in terms of feature content between the dark brown roof and the dark green forest. The evaluation of the quantitative performance measures, depicted in tab. 4.7, shows a 10% decrease of recall. The precision, on the other hand, is raised by as much as 29% compared to the single sPDC segmentation of the original image from the Berkeley data-base. Combined in the summary F-measure, a gain of 8% is achieved.

	ρ	π	F-measure
single sPDC segmentation	0.70	0.52	0.60
aggregated bootstrap edges	0.63	0.67	0.65
gain (%)	-10	29	8

Table 4.7: Resampling evaluation summary for image 361010.

Summing up the findings of the experiments described above, the conclusion is that sPDC is a very stable segmentation method in the sense that boundaries between image regions of pronouncedly varying feature content are reliably inferred. This is reflected in the high edge probabilities between such image parts, which are amply visible from the depictions of the edge probability maps. Often these most stable segment borders correspond well to those, which are semantically meaningful. This is the explanation for the remarkable increase in precision, which could be found for all but one example (i.e. image 134035). The latter image must be interpreted as an outlier. Most likely, the complicated image structure is more susceptible to the distortions produced by the two-stage sampling scheme. Overall, the precision values increased about 28% on average. The situation for the recall quality measure is somewhat less clear. On the one hand, one has three examples for which one has to note a decrease, whereas on the other hand, three examples exhibit a small, but nevertheless noticeable, increase. Finally, for one example, there has been no difference at all between the result for the aggregated edges from the resampling experiments and the corresponding findings for single sPDC segmentation. In contrast to the precision evaluation, a decrease of recall could have been expected for the aggregated edges. Due to the averaging process, unstable segment boundaries are diminished. Consequently, the number of strong edge pixels can be expected to be lesser than in the case of single sPDC segmentation, where all borders between image parts receive probability one. This connection in turn leads to a smaller probability, that a retrieved edge pixel is also relevant, which is exactly what the recall value measures. Nevertheless, one finds an averaged overall increase of recall of about 1%. For the summary statistic of the F-measure, the gain in terms of precision and recall

amounts to a mean increase of nearly 13%. The averaged recall-precision-graph for all the evaluated examples is given in fig. 4.3.

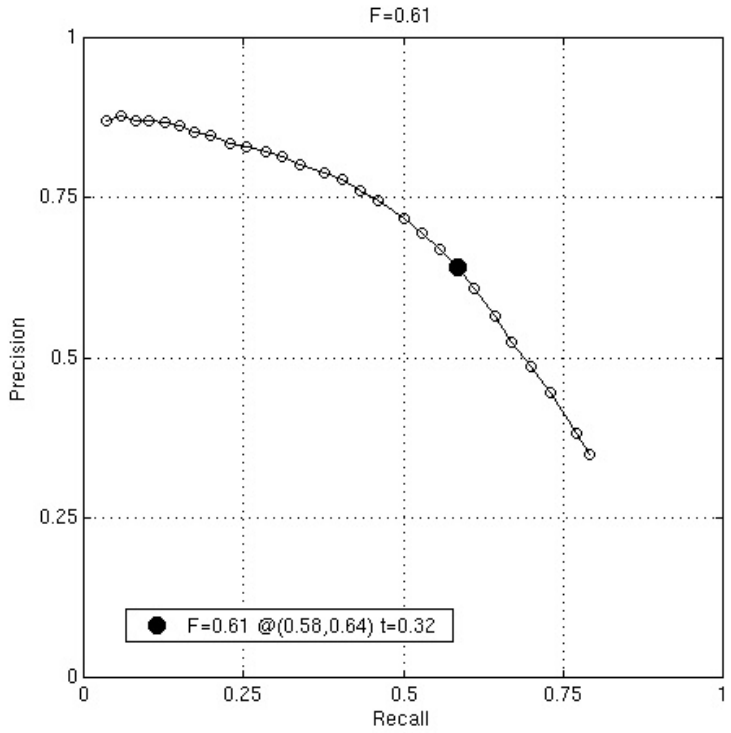
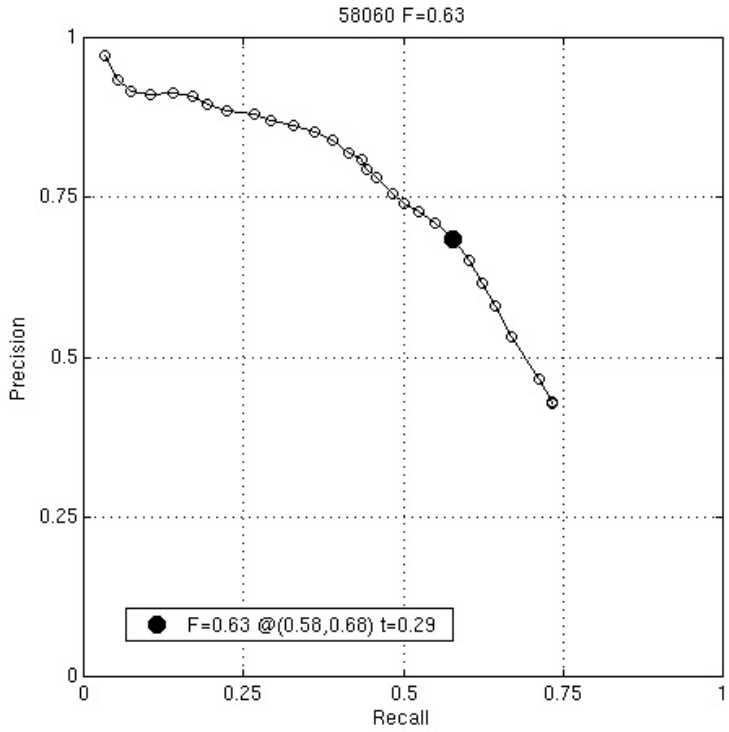


Figure 4.3: Joint recall-precision-curve for the discussed resampling examples.

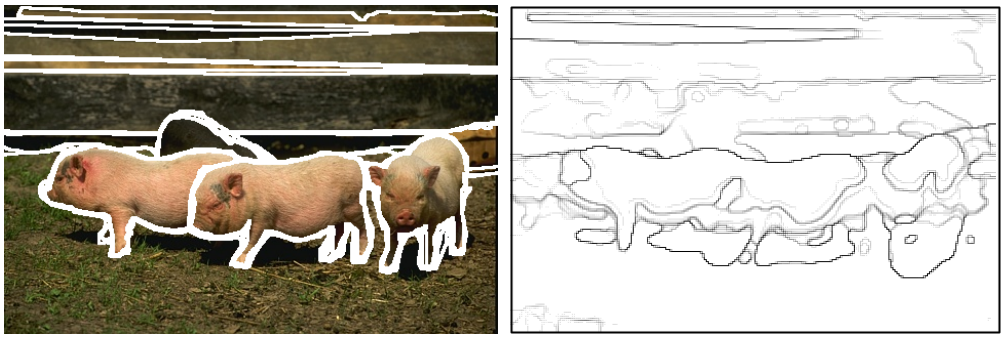
Finally, a word of caution seems to be in place. When interpreting the results for the quantitative performance evaluation, one has to distinguish two effects which influence the recall, precision and F-value performance measures. One is given by the fact that by means of the averaging of boundaries over the set of bootstrap-samples, stable edges are emphasized, while those that resulted from optimization artifacts or specific noise characteristics of the input image are diminished. This gain is actually achieved by using the resampling strategy. On the other hand, however, the aggregated edges are not any longer true image partitions due to the fact that no closed coherent segments are inferred. This relaxation of the segmentation constraint also accounts for some part of the performance increase. Unfortunately, there is no simple way of isolating these two kinds of influences on the quality measures. One could think of a bootstrap-enhanced version of PDC, which constructs segmentations by computing the averaged group assignments over all sample images. However, the implementation of such a procedure is rather involved, as it requires to find corresponding assignments across different permutations of cluster-indices (c.f. [FB02]). Thus, it could not be further pursued in this thesis due to time constraints.

(a) (b)

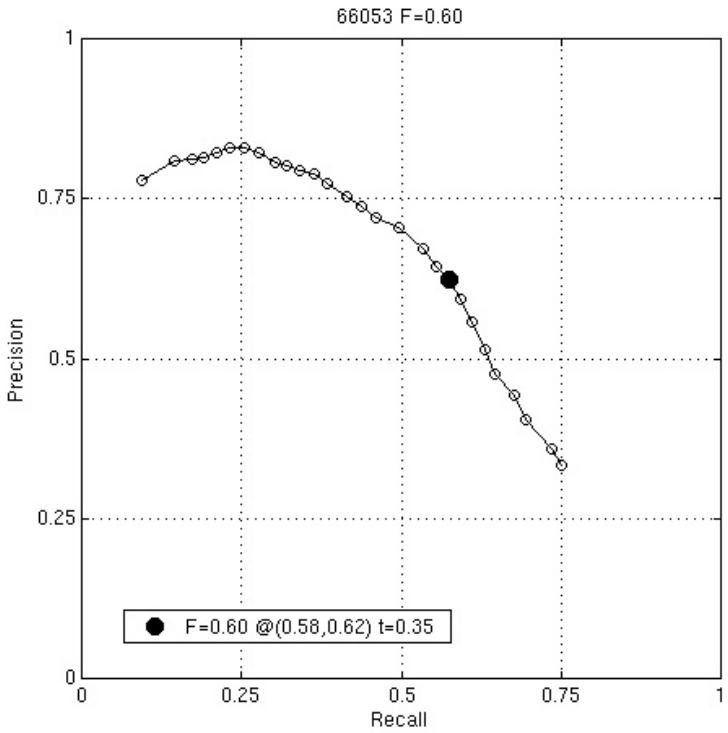


(c)

Figure 4.4: Human segmentation compared to aggregated bootstrap edges for image 58060. a) overlaid human segmentations, b) aggregated bootstrap edges, c) precision-recall-curve.

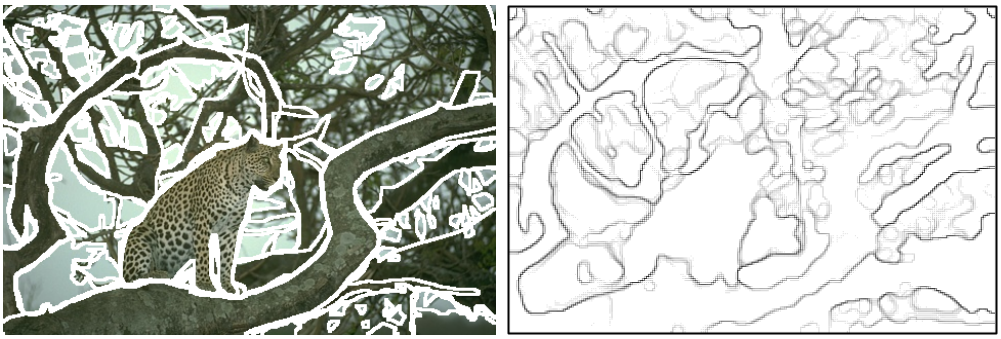


(a) (b)

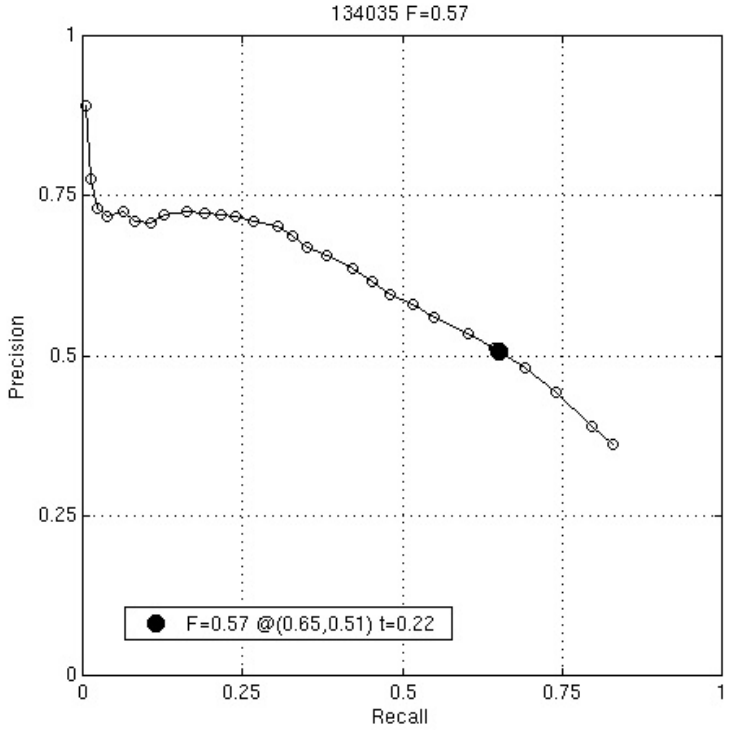


(c)

Figure 4.5: Human segmentation compared to aggregated bootstrap edges for image 66053. a) overlaid human segmentations, b) aggregated bootstrap edges, c) precision-recall-curve.

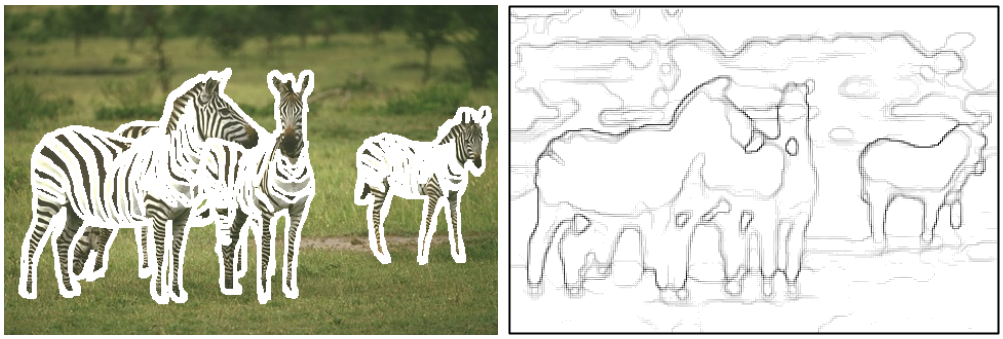


(a) (b)

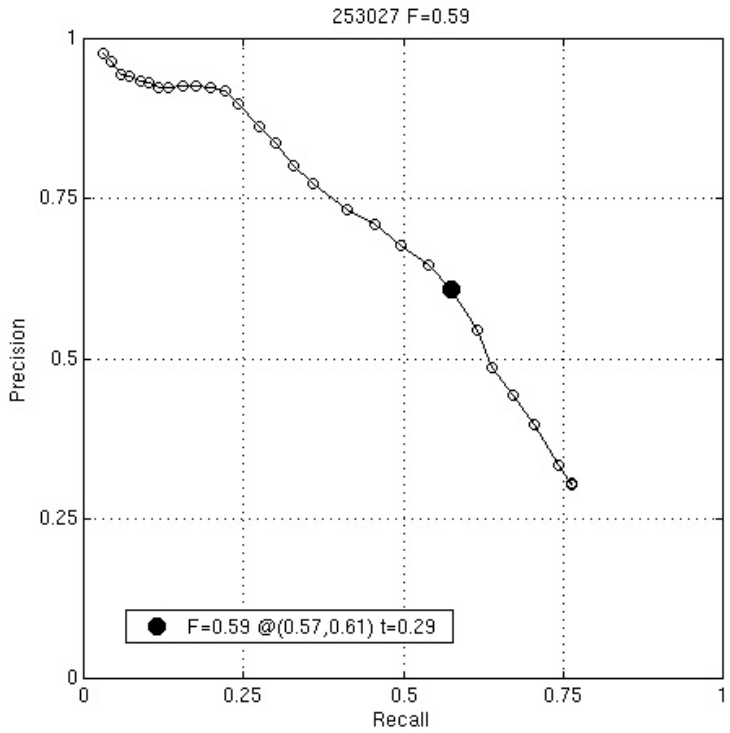


(c)

Figure 4.6: Human segmentation compared to aggregated bootstrap edges for image 134035. a) overlaid human segmentations, b) aggregated bootstrap edges, c) precision-recall-curve.

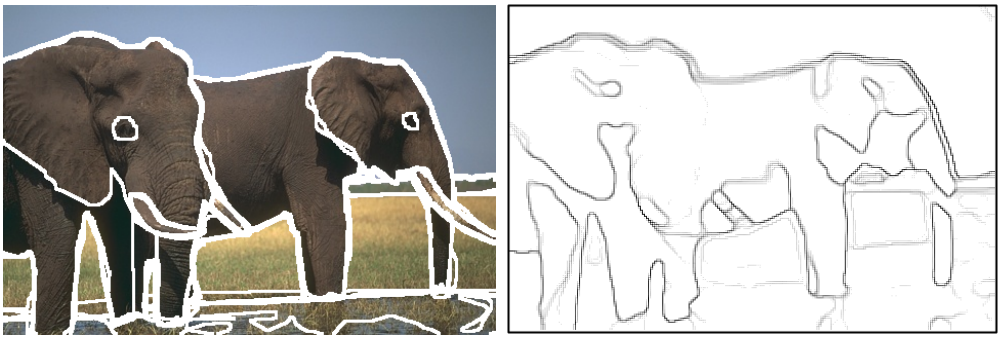


(a) (b)

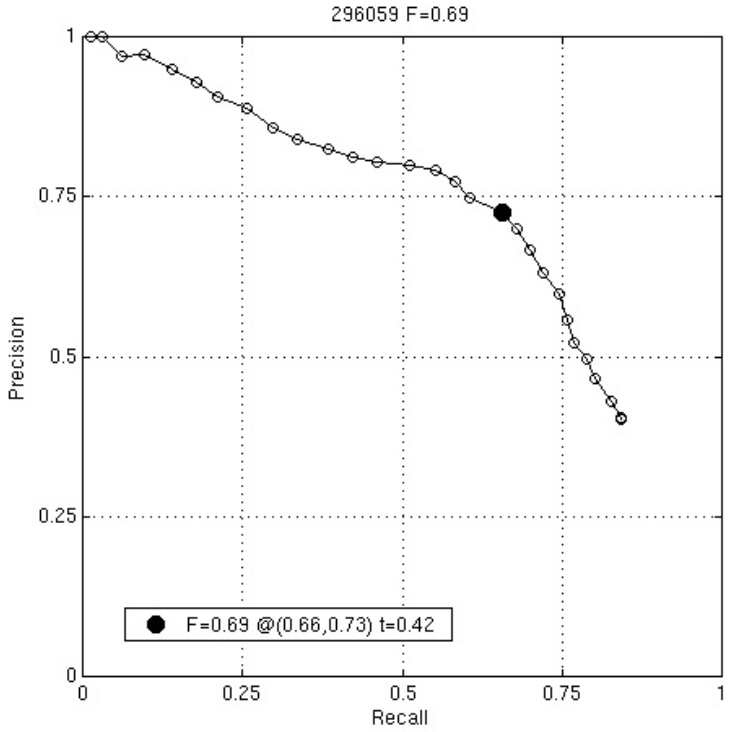


(c)

Figure 4.7: Human segmentation compared to aggregated bootstrap edges for image 253027. a) overlaid human segmentations, b) aggregated bootstrap edges, c) precision-recall-curve.

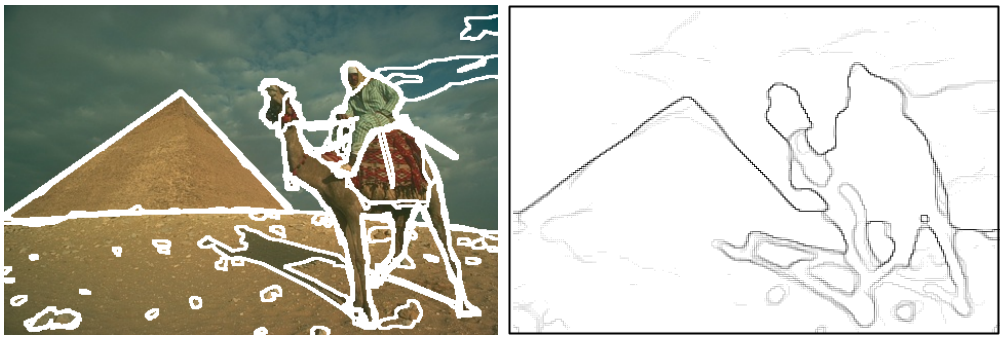


(a) (b)

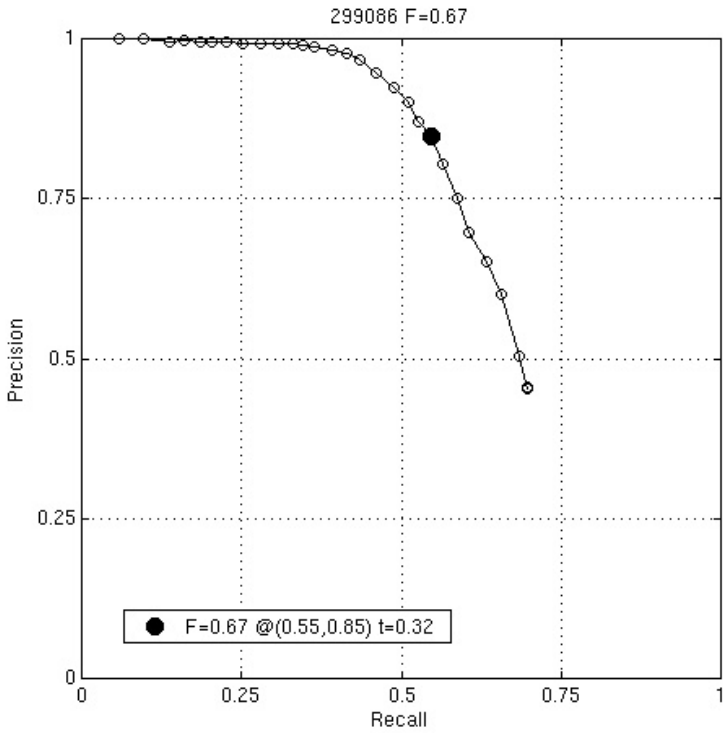


(c)

Figure 4.8: Human segmentation compared to aggregated bootstrap edges for image 296059. a) overlaid human segmentations, b) aggregated bootstrap edges, c) precision-recall-curve.

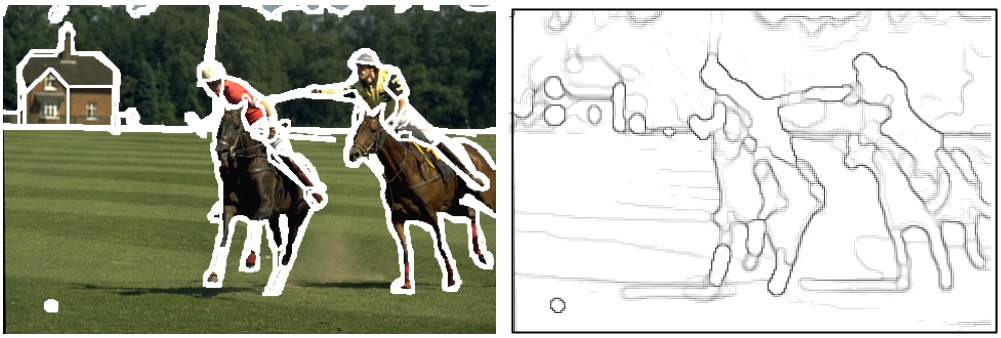


(a) (b)

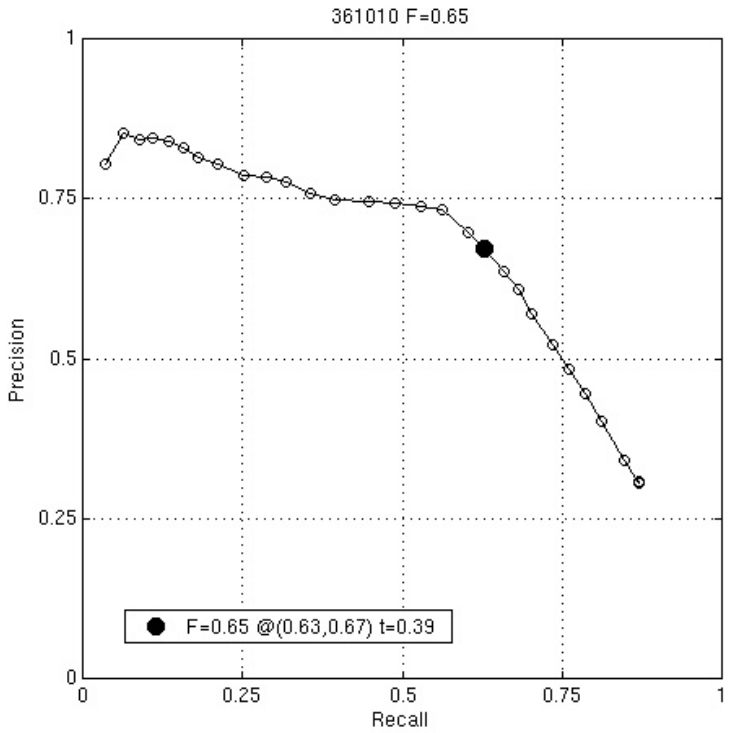


(c)

Figure 4.9: Human segmentation compared to aggregated bootstrap edges for image 299086. a) overlaid human segmentations, b) aggregated bootstrap edges, c) precision-recall-curve.



(a) (b)



(c)

Figure 4.10: Human segmentation compared to aggregated bootstrap edges for image 361010. a) overlaid human segmentations, b) aggregated bootstrap edges, c) precision-recall-curve.

4.2 Generalizing Segmentation Solutions

4.2.1 The Generalization Problem

Apart from, but closely related to, the issue of stability of inferred grouping or classification solutions against random fluctuations of the input data, another major concern for any learning algorithm is the question of generalizability. Commonly, a grouping or classification model is determined on the basis of a given test-set of representative objects from the domain of interest. Once the model is fixed, it is applied to new, previously unseen, examples. Consequently, the performance on the latter set is usually of much greater importance, than the quality achieved on the training set. As has already been pointed out in the discussion of the resampling approach, any dataset does not only contain structural information about the nature of the source, but also random fluctuations. Adapting the learning algorithm perfectly to the training data thus inevitably results in modeling the aforementioned noise. Consequently, the performance on unseen examples deteriorates, which constitutes the problem of over-fitting. In the case of image segmentation, many algorithms are proposed, which arrive at an image partition by other means than learning procedures. Moreover, in the majority of cases, only the performance in terms of the segmentation quality on the given image data is deemed to be of interest. Thus, the application of a trained model to new data is not widely discussed in the pertaining literature. PDC and its variants, however, are clustering approaches, representing the data-groups via continuous models of feature distributions, and as such they are learning algorithms. Therefore, the question to which extend the model fitting stage is able to identify the true structural properties of these distributions, while ignoring the noise, is most relevant. In this section, this problem is addressed in an empirical manner.

4.2.2 Experimental Setup and Results

In order to evaluate the generalization ability of sPDC solutions, experiments concerning the transfer of models between similar images have been conducted. As the Berkeley test-set is geared towards offering a great variety of different images, it has not been considered suitable for this kind of experimental set-up. Instead, a collection of five pairs of images from the Corel image gallery has been compiled, which comprises four examples from the sub-directory Zebras and three image pairs from the sub-directory Cheetahs, Leopards and Jaguars. For the zebra category, sPDC solutions with four clusters have been produced, while the results for the other set are based on segmentations with five clusters. The settings for the data-acquisition and pre-processing steps are the same as for the sPDC experiments discussed in section 3.1.3. For each image pair, the experimental procedure is as follows: In the first step, sPDC segmentation solutions have

been computed independently for both pictures. The relevant model parameters for the solution of the first image, i.e. the Gaussian alphabet and the specific mixture weights for each data-group, have been recorded. The second step then consisted of applying the sPDC model from the first image to the second. To this end, the data-acquisition process of the second image has been run, followed by a complete E-Step computation yielding the optimal assignments of all sites in the second image given the continuous model parameters of the sPDC result from the first picture. In this way, a transfer of segmentation solutions from the first to the second image in each pair has been achieved under the assumption, that everything which can be learned and transferred from one image to the other is contained in the prototypical group distributions. A pseudo-code version of the transfer algorithm is shown below. Note that the random permutations are applied in order to prevent artifacts from a fixed optimization order.

Algorithm 1 Transfer of sPDC segmentation solutions.

Require: sPDC model for first image has been computed
Ensure: segmentation of image two according to the sPDC model of image one
  load continuous sPDC model parameters of first image
  run data extraction procedure for second image
  repeat
    generate a random permutation Π of $\{1, \ldots, n\}$
    for $i = 1$ to n **do**
      set $\mathbf{s} = \Pi(i)$
      for $c = 1$ to k **do**
        calculate meanfields $h^{\mathrm{s}}_{\mathbf{s}c}$ according to eq.(3.3)
      end for
      for $c = 1$ to k **do**
        update assignments according to eq.(3.6)
      end for
    end for
  until converged

In the ensuing examples, the first row shows the partition solution of the first image in the top row. Then, the segmentation of the second image is depicted in row two. Finally, the result of the solution transfer is shown in the last row. For all segmentations, outline visualizations are given in the first column, while the second column exhibits a color coded depiction.

The first image in the first example pair, given in fig. 4.11, shows a herd of zebras at a small river. The image region covered by the group of animals is essentially covered by the green cluster, visible in the color coded depiction of the segmentation result. The black cluster contains regions of high lightness, in which also zebras are present. The light brown parts of the image at the lower border are attributed to the red cluster, while the darker regions are assigned

to the blue data-group. In the second picture, one sees another group of zebras at approximately the same scale standing on a grass plain. The picture also contains regions of a light brown hue as well as darker regions especially around the legs of the animals and in the lower corners. These properties make it an ideal candidate for the application of the sPDC model from the first picture. Although the segmentation is of slightly lesser quality than the result of the specifically optimized model for the second picture, the characteristics of the inferred clusters of the first image are perfectly reflected in the results of the model transfer shown in the last row of the figure. The region containing the torsos of the zebras is attributed to the green and black clusters, which also covered the corresponding areas in the first image. Moreover, the red data-group covers the image parts of light brown and green hues, whereas the darker areas are assigned to the blue cluster, which is again in very good agreement with the partition of the first picture.

In the next pair of example images shown in fig. 4.12, single zebras are depicted. In the first image, the region occupied by the animal is essentially represented by the blue cluster in the pertaining sPDC clustering solution. The region of lighter brown and green hues is attributed to the data-group visualized in black. Those parts of the image that are dominated by brown hues of medium lightness are assigned to the red cluster, whereas the darkest image regions correspond to the data-group shown in green. The zebra which is depicted in the second image possesses a strongly differing fur pattern than the one in the first picture. The spatial frequency of stripes is much higher. Moreover, there is a noticeably lesser amount of horizontal stripes. Additionally, the second picture is more closely focused on the animal than in the first case. In summary, the measured texture features, and thus their distributions show a certain amount of variation. Despite these restrictions, the transfer of the segmentation model from the first image to the data of the second still yields a very satisfactory result. The zebra is, except for some small regions, attributed to the blue cluster, which also represented the image data of the animal in the first segmentation. The correspondence of the other clusters is very good too. The data-group depicted in green also covers the darkest parts of the second picture. The black cluster, being responsible for the lightest regions in image one, also represents the lightest parts of image two. Furthermore, the data-group for the segment characterized by intermediate brown tones in the first image is used to encode those regions of the second picture, for which the pertaining brown hues lie between the dark colors of the shadow regions and the light brown or almost ecru hues of the black cluster.

For the third example set, which is depicted in fig. 4.13, the model transfer starts out from the same image as in the preceding experiment. Here, however, the pertaining sPDC clustering solution is transferred to another image. The latter picture shows two zebras, a dam and its foal, standing on a paltry plain. Obviously, the dissimilarity between the two images in this example set is rather high. Not only are the two zebras in the second picture farther away from the

observer, the foal also exhibits a fur pattern with higher spatial frequency compared to a full-grown zebra. Furthermore, the surrounding background sceneries in the two images are rather distinct. Nevertheless, the transfer of clusters models turns out to work very well. The vast majority of the area covered by the two zebras is assigned to the data-group color-coded by blue. This is the same cluster to which the zebra region in the first image belongs. Apart from this, the other cluster associations also match those of image one. The regions of light grayish brown in the second picture are assigned to the black cluster, the one which encoded the light brown regions in image one. The red cluster, to which the regions of medium green and brown hues in the first image are attributed to is also responsible for the areas of medium green and brown tints in image two. Finally, the darkest regions in both images are attributed to the green cluster.

The fourth example, depicted in fig. 4.14, shows a distant group of zebras. The corresponding region in the image is mainly attributed to the red cluster, while some minor parts are also assigned to the black cluster. The other image parts are dominated by shades of brown, grey, and green hues of varying lightness. The lightest parts are covered by the blue data group, while the darkest regions are assigned to the black cluster. The areas of intermediate lightness belong to data-group color coded by green. The second image of this example pair also depicts a small group of zebras, which are slightly closer to the observer. The background of this image mainly exhibits grayish hues of brown and green. The outcome of the model transfer is again shown in the third row of the image collection. Here, one has once again to note, that the transfer of segmentation solution provides for a very good result. The zebras are assigned to the red clusters, just as in image one. Moreover, the correspondence between the other group-assignments is as expected. The darkest image regions are attributed to the black clusters, whereas the lightest ones are associated with the data-group color-coded in blue. Again, the green cluster is responsible for the segment of intermediate lightness.

In the fifth example, given in fig. 4.15, pictures from the *Cheetahs, Leopards, and Jaguars* sub-directory from the Corel image gallery are treated. Consequently, the first image shows a cheetah walking over a grassy plain. The area occupied by the animal itself is mostly attributed to the black cluster. Some parts, which are covered by shadows, however, are assigned to the red cluster, which otherwise covers the darkest green parts of the image. Those areas which exhibit light green an brown hues belong to the data-group color-coded in green. The other two clusters separate two regions of varying green hues. In the second image, one can observe a far less amount of green hues, instead, large parts of the image exhibit the light brown and ecru tones of dry soil. These colors correspond much better to the color of the cheetahs fur than the green grass in the first picture. Nevertheless, the animal is found as an independent segment with great accuracy. The solution transfer is not trivial in this case, as the cheetah in the second image is somewhat closer to the observer than the one in the first, leading to different responses of the texture filter-bank. Moreover, as has just been

stated, the first image, in contrast to the second, does not contain light brown regions except for the cheetah itself. Nevertheless, the transfer of the sPDC model distributions from the first image to the second yields very satisfactory results. Most of the area occupied by the animal is associated to the black cluster, just as in the segmentation of the first image. Furthermore, the darkest regions of brown and green hues are assigned to the red data-group, in perfect correspondence to the clustering solution of the first picture. The visual inspection of the other image parts shows, that the assignments of remaining regions match those of the first image in this example pair too.

In the example pair given in fig. 4.16, the same image as in the previous experiment is utilized as the starting-point for the model transfer. This time, however, its pertaining sPDC model is applied to a different image. Here, the similarity between the depicted scenes is certainly larger than in the previous experiment, consequently, an even better transfer result can be expected. Indeed, as the segmentation visualizations show, the image region assigned to the black cluster in the transfered solution fits the area covered by the cheetah as tightly as the region of the corresponding data-group in the sPDC model that has been trained on image two. Moreover, the correspondence between the background clusters in terms of the visual impression of the associated regions is also remarkable.

Finally, the seventh image pair, given in fig. 4.17, shows jaguars in their natural environment. In the first picture, one sees a single animal in front of a small cave-like structure. As the background solely consists of the same type of rock, the colors are just varying hues of brown, reaching from ecru in the areas of incident sunlight to almost black in the shadowed regions. As the segmentation result shows, sPDC is very well able to identify the jaguar as a segment of its own. The other four data-groups mainly differentiate between the different brown to black tones of the background. The second image in this pair is remarkably different. Here, the animal is depicted amidst a variegated background. The range of colors includes the bluish hues of the water, green tones, and varying shades of brown. Although this scenery is very cluttered, the jaguar is identified as an independent segment in the pertaining sPDC segmentation solution. Despite the tremendous differences between these images, the transfer of the segmentation model from image one to the second picture works very well. The animal is attributed to the data-group color-coded in black, just as one would expect from the segmentation result of the first image. Clearly, the identification precision of the jaguar's segment outline does not quite reach the one from the adapted model. But this could hardly be expected given the differences between the pictures. Moreover, despite the fact that the other four clusters in the sPDC solution for the first image had only to differentiate between varying brown hues, the segmentation of the background in image two is very reasonable.

Summing up the results for the experiments discussed above, one has to conclude, that the sPDC approach to image segmentation is very well able to capture the structural statistical properties of image regions in the inferred Gaussian mix-

ture models of the pertaining clusters. The transfer of solutions yields satisfactory results even for image pairs which exhibit a considerable amount of variation in terms of the particular feature distributions.

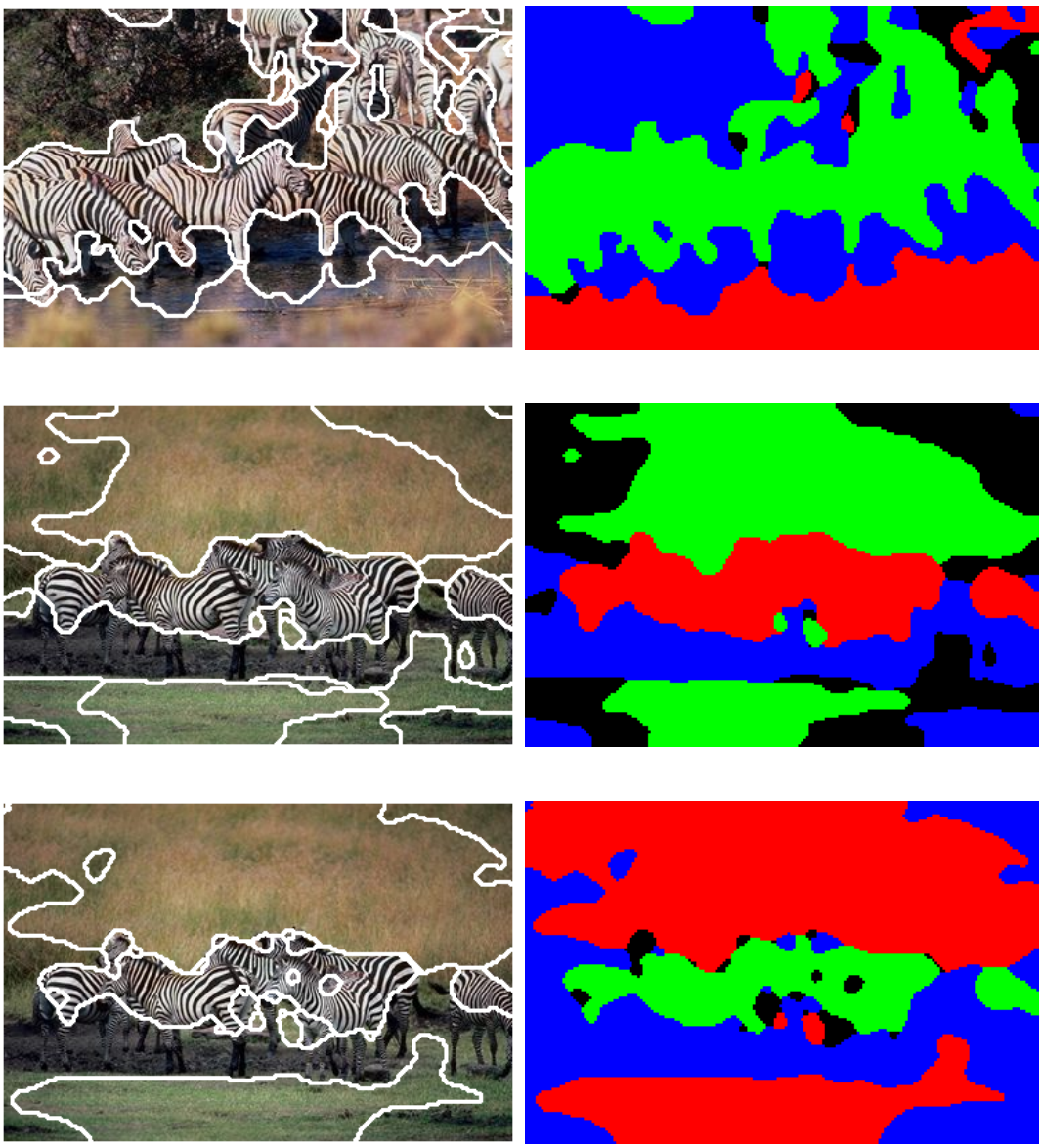


Figure 4.11: Generalizing sPDC solutions, example image pair one.

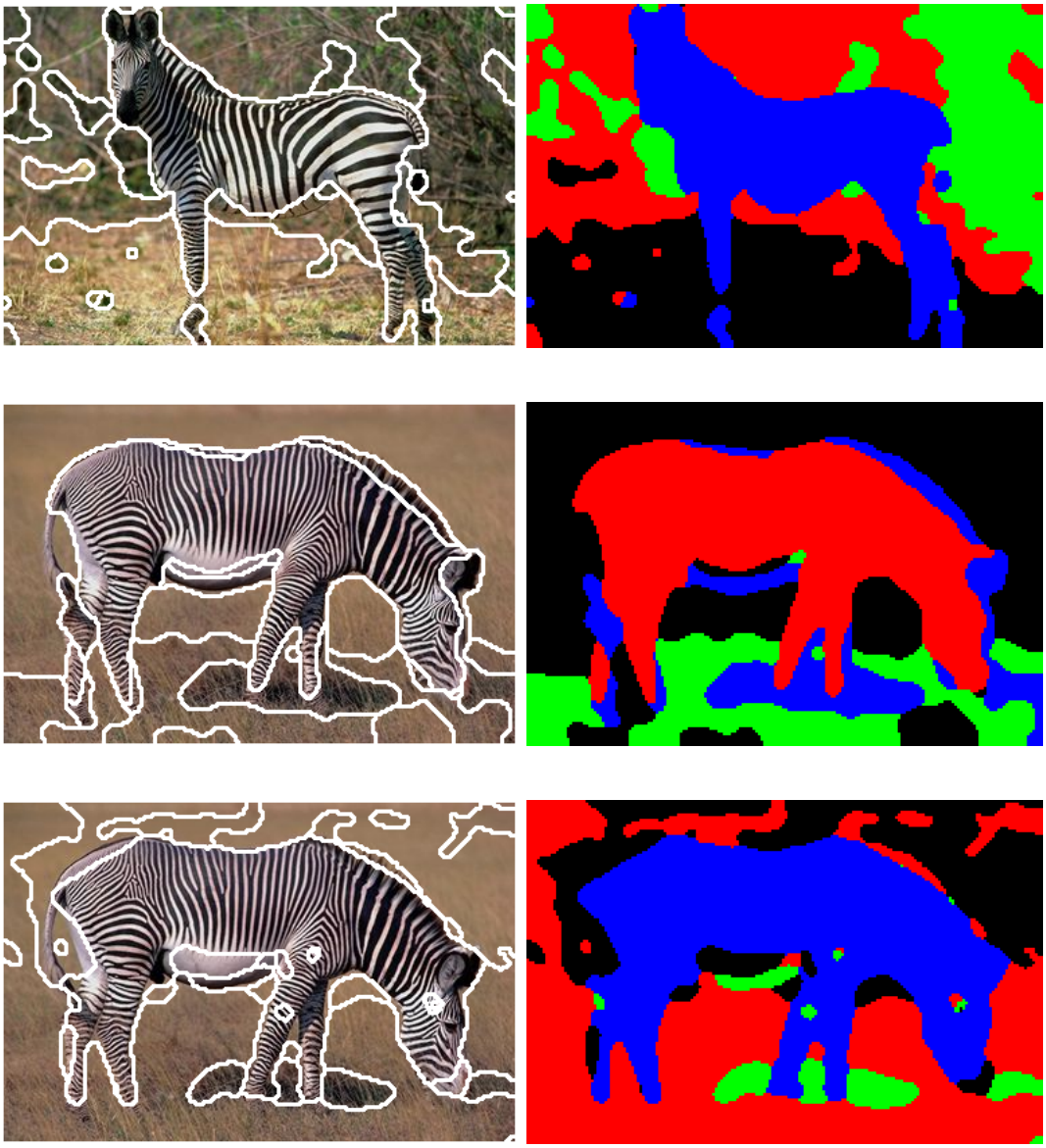


Figure 4.12: Generalizing sPDC solutions, example image pair two.

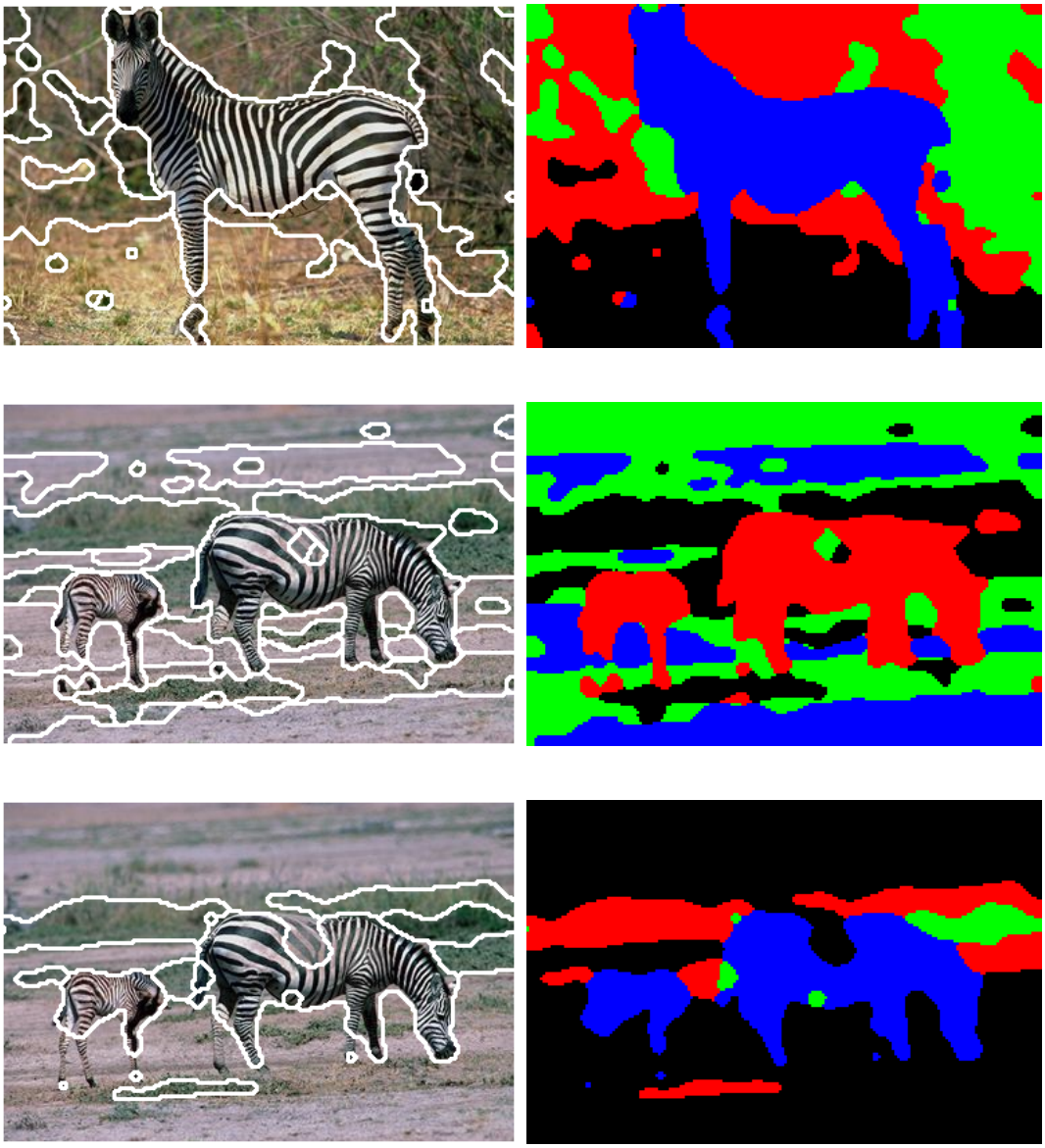


Figure 4.13: Generalizing sPDC solutions, example image pair three.

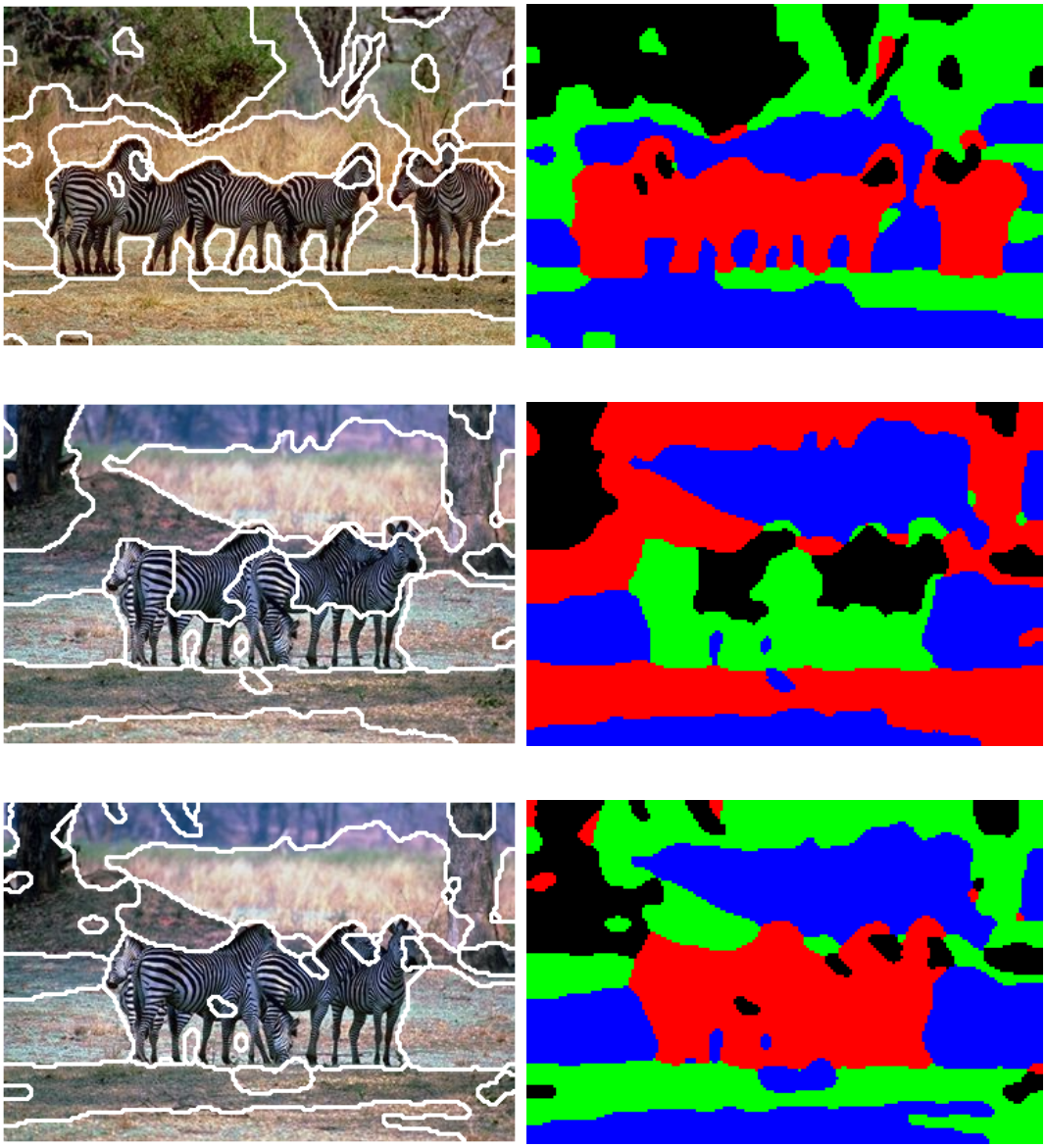


Figure 4.14: Generalizing sPDC solutions, example image pair four.

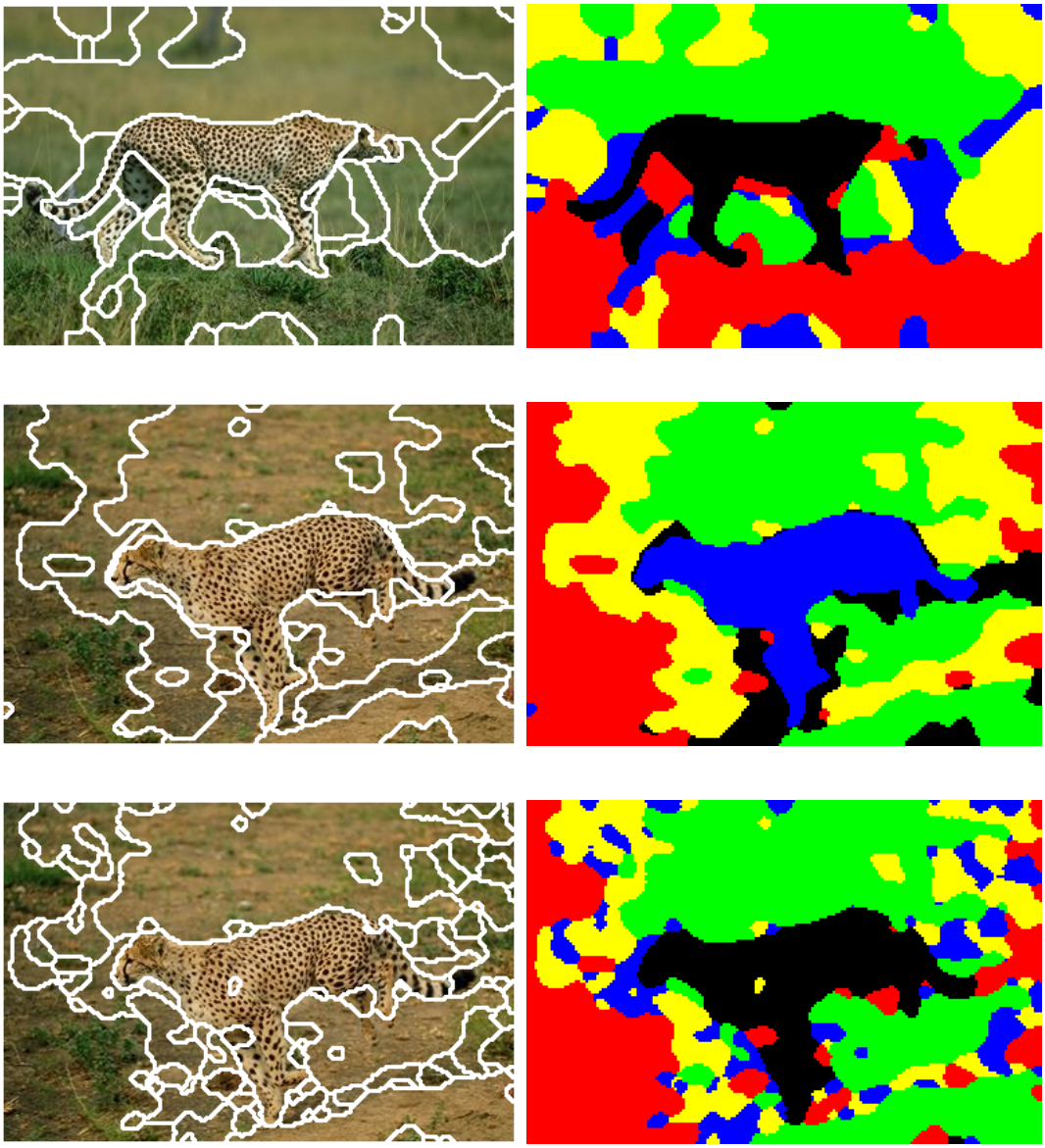


Figure 4.15: Generalizing sPDC solutions, example image pair five.

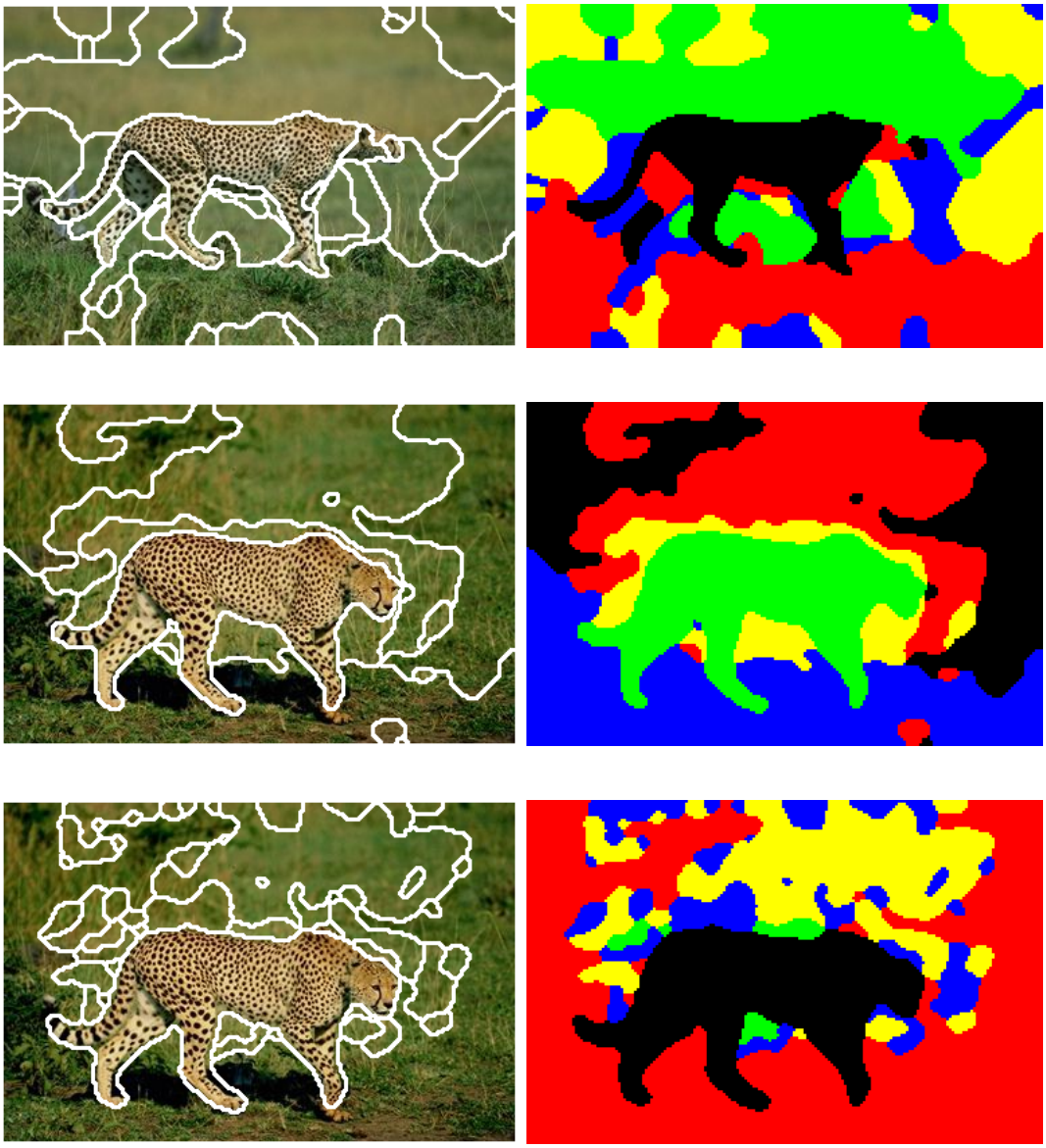


Figure 4.16: Generalizing sPDC solutions, example image pair six.

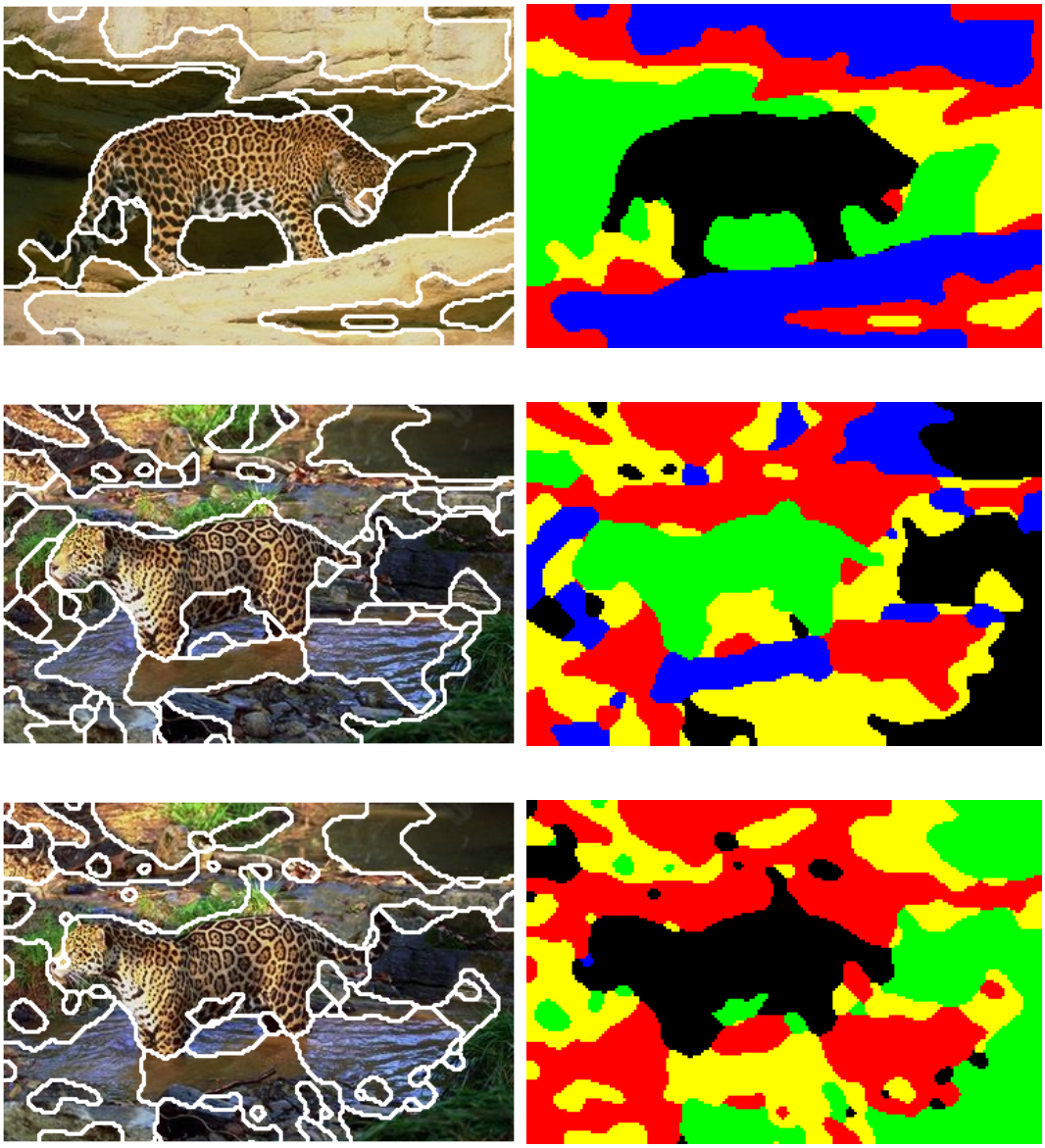


Figure 4.17: Generalizing sPDC solutions, example image pair seven.

4.3 Summary

In this chapter, two major subjects of learning theory, i.e. the stability of inferred solutions against small variations in the data, as well as the applicability of learned models to new problem instances, have been empirically evaluated for the sPDC approach to image segmentation. In both cases, sPDC showed a remarkable performance. The stability experiments via a variant of bootstrap resampling for image data clearly demonstrated, that sPDC is capable of producing robust segmentations of the given image data at hand. All major boundaries between different image regions have been reliably inferred. Moreover, the averaging process leading to aggregated probabilistic edge-maps of the processed images yielded a considerable performance increase in terms of the utilized quality measures precision and F-value, while leaving the recall virtually unchanged. This finding is explained by the suppression of those segment borders, that did not prove to be stable across segmentations of different image instances. Consequently, the proposed resampling approach provides for an interesting means of boosting segmentation performance. The only drawback of such a procedure is given by the fact that the probabilistic edge maps do not constitute valid image partitions, since border pieces and edge sets and not necessarily closed, coherent segments are inferred.

The experiments concerning the generalization ability of sPDC demonstrate, that the inferred model distributions for each cluster faithfully capture the statistical properties of the associated regions. Otherwise, the remarkable results concerning the transfer of segmentation solutions between images of different albeit similar content could not have been achieved. In the given examples for the transfer of segmentation models, not only the cluster of the foreground object, but also those covering the less specific background regions have been reasonably matched to the characteristics of the images, to which they have been applied.

4.4 Bibliographic Remarks

The foundations of statistical learning theory including an in-depth discussion of the problems concerning robustness and generalization ability of machine learning approaches are laid by Vapnik in [Vap95, Vap98]. The Bootstrap method has been invented by Efron [Efr82, ET84]. Practical applications are discussed in [DH97, Goo99]. Bootstrap error estimation is treated in [JDC98]. The related approaches of *bagging* and *boosting* are introduced in [Bre96] and [Fre95] respectively. The application of these methods to the classification of remote sensing data is discussed in [CHD01]. Finally, [FB02, RBLB02, RLBB02] develop resampling methods for cluster validation and model order selection.

Chapter 5

Shape Constrained Segmentation

5.1 Introduction

The segmentation of images, e.g. as discussed in the previous chapters, is of prime importance in low level vision. Besides being a lively field of research in its on right, it can also be used to arrive at intermediate representations of image content, which provides an abstracted and simplified model of the depicted scene compared to the raw image data itself. Having such a representation at hand, higher level processing, e.g. object recognition and content based image retrieval, is greatly facilitated.

Full advantage of such an approach is apparent, if the segmentations are in good agreement with the semantic content of the depicted scene. Unfortunately, this precondition is frequently violated. Although considerable progress has been achieved and powerful algorithms are available today, human segmentation quality is still unmatched, especially when real world imagery defines the application domain. One of the reasons for this discrepancy is given by the fact that most of the algorithms in that field rely solely on low-level features, whereas human segmentation has many higher level knowledge sources to utilize. The central motivation for this chapter arises from the goal to bridge this apparent gap. To this end, a novel way of incorporating prior knowledge on the shape of certain objects of interest in the segmentation process is introduced to facilitate a semantically correct segmentation.

The structure of this chapter is as follows: First, the probabilistic representation of shape, which subsequently functions as a prior for the shape constrained segmentation approach (SCS), is described in section 5.2. Then, section 5.3 will expand the idea of model transfer as outlined in the previous chapter to the notion of *aspect sets*. These are collections of prototypical feature distributions for some given semantic category that reliably capture the essential statistical properties thereof. From these aspect sets, likelihoods for the presence of an object from a certain semantic category are derived. With these prerequisites at hand, the shape constrained segmentation approach is formulated in a Bayesian setting in section 5.4, combining the category likelihoods with the *a priori* shape information to a posterior occurrence probability of an object from the given semantic category. Representative experimental results are discussed in section 5.5.

5.2 Representing Shape Knowledge

In order to integrate shape knowledge into the segmentation process, the problem of adequate representation has to be addressed. An answer to this question is highly dependent on the kind of objects one is interested in and the application domain in general. Rigid objects which are always presented in the same pose certainly demand a less variegated description than articulated objects that are depicted in a rich variety of possibly different poses.

Although the method of shape constrained segmentation presented here is by itself very generic, its application context is restricted to the identification of certain species of wildcat (cheetahs and jaguars) in images of their natural environment (c.f. section 5.5 for details). This setting clearly demands for a highly adaptive shape representation, as both the multitude of possible poses and the non-rigidity of the objects in question exclude static approaches. It is also apparent that no single shape representation can at the same time be totally pose invariant while being still discriminative enough to capture the specific details which are necessary for a faithful recognition of the given object of interest. For the sake of simplicity, the emphasis is put on one specific class of poses, namely wildcats in a sideward standing position. Even under this seemingly limiting constraint, the images show a remarkably large diversity of shape instances due to the non-rigidity of the objects in question. Consequently, it is proposed to represent the shape information in a probabilistic way. This approach has the further advantage that it fits well into the statistical framework of the PDC method, as will be shown in section 5.4.

In order to arrive at such a representation, a number of representative images of the objects are first segmented in a sketchy way. This step could also be automated by using a segmentation algorithm. After that, these segments are centered with respect to the particular image at hand in order to avoid influences of the relative position on the final shape representation. In a next step, distances of every pixel in a given image to the now centered region depicting the object of interest are computed for all the images in the set of representatives by applying a chamfer transform [SHB98]. These distances themselves could be interpreted as shape probabilities after an appropriate normalization. This, however, would yield unsatisfactory results due to the fact that the linear gradient puts too much emphasis on remote regions.

Therefore, a zero-mean Gaussian model is applied to the distances, which ensures an exponential decay of the probabilities for sites which are not part of the shape areas. A detailed description of the parameters is given in section 5.5. In order to join the shape information contained in the images of the representative set into a single shape description, these individual probabilistic shape models are fused by an averaging process. Finally, the combined probabilistic representation is subjected to a blurring operation to reduce inhomogeneities which might still persist after the averaging. The processing pipeline for the construction of the shape representation is outlined in figure 5.1.

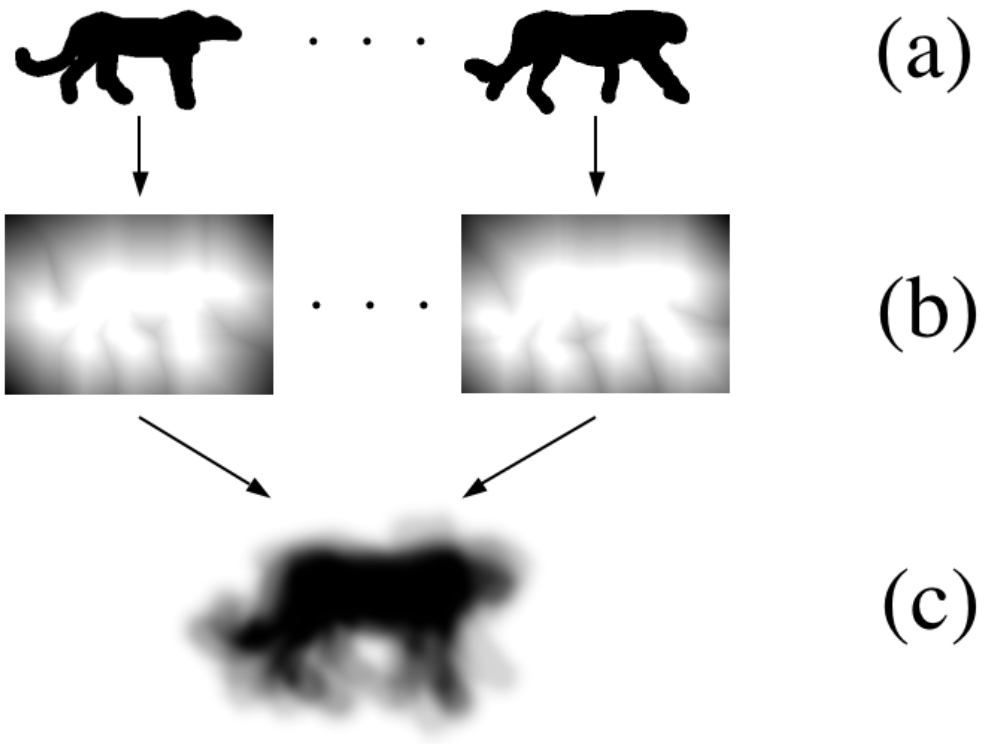


Figure 5.1: Prior shape model construction: a) sketchy hand segmentations, b) chamfer transforms, c) final shape prior.

The approach for shape prior generation described above certainly requires some amount of human interaction. Nevertheless, this is not considered a critical problem. One can think about a bootstrap procedure, where initial segments without shape information are used to form the first shape templates; then these boundaries are averaged after a rough alignment and the automatically synthesized shape priors are then used to identify new shape constrained segments. Such an approach would not require any human interaction at all.

5.3 Aspect Sets

For real-world applications it is evident that given images not only contain instances of the sought-after objects, but also large amounts of background pixels. Although this background usually is composed of clutter with little discernible shape properties, it can embody a broad variety of different distributions of elementary image features. The PDC segmentation method and its variants characterize image parts by mixtures of Gaussians which define prototypical distributions for the measured features. Hence it concisely summarizes the statistical

properties of image regions. Once PDC segmentations for a given database of images at hand have been inferred, representative feature distributions for the semantic category of interest as well as the background can be selected by user interaction.

Encouraged by the success of the model transfer as documented in the previous chapter, this idea is now taken one step further. Despite the satisfactory results for the direct transfer of segmentation solutions from one image to another, this approach evidently has its limitations. Clearly, the concrete form of such model distributions for the foreground object is dependent on a number of influences like viewing angle, lighting conditions, and the distance between observer and the depicted object, which may vary widely over different images in a dataset. Thus it is appropriate to represent a semantic category by more than one distribution, where these statistical models are interpreted as different aspects of a category. This situation is even more pronounced for the background, as it is a much less well-defined concept. Therefore, it is proposed to summarize manually chosen prototypical feature distributions for background and foreground in a database $\mathcal{A}$, which is comprised of the set of foreground distributions $\mathcal{A}_f$ of size n_f and the set of background distributions $\mathcal{A}_b$ of size n_b:

$$\begin{aligned} &\mathcal{A} = \mathcal{A}_f \cup \mathcal{A}_b, \text{ where} \\ &\mathcal{A}_f = \{a_1^f, \ldots, a_{n_f}^f\}, \quad \mathcal{A}_b = \{a_1^b, \ldots, a_{n_b}^b\}. \end{aligned} \tag{5.1}$$

In the derivations given below, a generic aspect in the set $\mathcal{A}$ is denoted by a. The division of the aspect set corresponds to the dichotomy of the set of semantic categories $\Omega = \{\omega_f, \omega_b\}$ into a semantic category representing the foreground object ω_f and another one for the category background ω_b. When the approach is presented with a new input image, it is first processed by the feature extraction procedure of the PDC framework, i.e. histograms for all image sites are computed.

The goal is now to find the most probable assignment of sites to the semantic categories given the measured feature histograms. This inference problem is based on independence assumptions that are summarized in the following graphical model for the data generation process, which is valid for all image sites: $\Omega \rightarrow \mathcal{A} \rightarrow \mathcal{X}$. A visualization of this graphical model is depicted in fig. 5.2. In the first step, a semantic category for the given site is determined by a probability distribution $p(\Omega)$. Then, an aspect is chosen with probability $p(\mathcal{A}|\,\Omega)$. Finally, a concrete realization of the observation data is drawn according to the distribution $p(\mathcal{X}|\,\mathcal{A})$. Thus one arrives at the following representation for the joint probability:

$$p(\Omega, \mathcal{A}, \mathcal{X}) = p(\Omega) \cdot p(\mathcal{A}|\,\Omega) \cdot p(\mathcal{X}|\,\mathcal{A}). \tag{5.2}$$

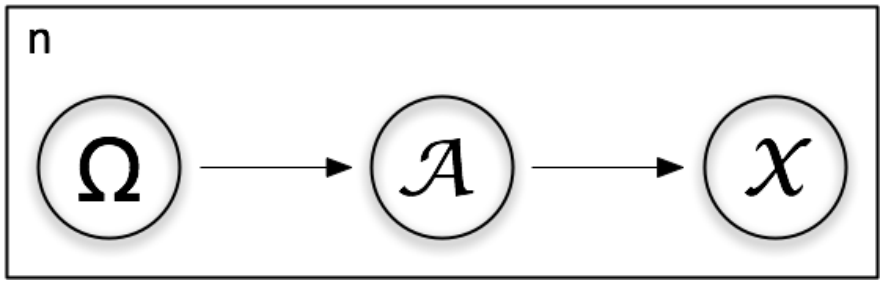


Figure 5.2: Graphical model of the SCS approach.

From this equation, twofold application of Bayes rule leads to

$$\begin{aligned} p(\Omega, \mathcal{A}, \mathcal{X}) &= p(\Omega) \cdot p(\mathcal{A}|\, \Omega) \cdot \frac{p(\mathcal{A}|\, \mathcal{X}) p(\mathcal{X})}{p(\mathcal{A})} \\ &= p(\Omega) \cdot \frac{p(\Omega|\, \mathcal{A}) p(\mathcal{A})}{p(\Omega)} \cdot \frac{p(\mathcal{A}|\, \mathcal{X}) p(\mathcal{X})}{p(\mathcal{A})} \\ &= p(\Omega|\, \mathcal{A}) \cdot p(\mathcal{A}|\, \mathcal{X}) \cdot p(\mathcal{X}). \end{aligned} \tag{5.3}$$

The computations detailed in the equation above can be interpreted as the reversal of the arrow direction in the graphical model given in fig. 5.2. Thereby, the perspective on the model changes from a generative to an inference oriented one. Moreover, one has

$$\begin{aligned} p(\Omega|\, \mathcal{X} = \mathcal{X}_\mathbf{s}) &= \frac{p(\Omega, \mathcal{X} = \mathcal{X}_\mathbf{s})}{p(\mathcal{X} = \mathcal{X}_\mathbf{s})}, \quad \text{where} \\ p(\Omega, \mathcal{X} = \mathcal{X}_\mathbf{s}) &= \sum_{a \in \mathcal{A}} p(\Omega, \mathcal{X} = \mathcal{X}_\mathbf{s}, \mathcal{A} = a) \\ = \sum_{a \in \mathcal{A}} p(\Omega|\, \mathcal{A} = a) \cdot p(\mathcal{A} = a|\, \mathcal{X} = \mathcal{X}_\mathbf{s}) \cdot p(\mathcal{X} = \mathcal{X}_\mathbf{s}). \end{aligned} \tag{5.4}$$

Therefore, the probability distribution of the semantic models for a site given a feature measurement is described by the following expression:

$$p(\Omega|\, \mathcal{X} = \mathcal{X}_\mathbf{s}) = \sum_{a \in \mathcal{A}} p(\Omega|\, \mathcal{A} = a) \cdot p(\mathcal{A} = a|\, \mathcal{X} = \mathcal{X}_\mathbf{s}). \tag{5.5}$$

Due to the fact that the set of aspect distributions is divided in the disjunct sets of foreground and background aspects, one obtains

$$p(\Omega = \omega_f|\, \mathcal{A} = a) = \begin{cases} 1 & \text{if } a \in \mathcal{A}_f, \\ 0 & \text{otherwise} \end{cases},$$

$$p(\Omega = \omega_b|\, \mathcal{A} = a) = \begin{cases} 1 & \text{if } a \in \mathcal{A}_b, \\ 0 & \text{otherwise} \end{cases}.$$

In order to evaluate eq.(5.5), the probability of an aspect model given the set of observations has to be specified. To this end, the application of Bayes rule leads to

$$p(\mathcal{A} = a|\, \mathcal{X} = \mathcal{X}_{\mathbf{s}}) = \frac{p(\mathcal{X} = \mathcal{X}_{\mathbf{s}}|\, \mathcal{A} = a) \cdot p(\mathcal{A} = a)}{\sum_{\tilde{a} \in \mathcal{A}} p(\mathcal{X} = \mathcal{X}_{\mathbf{s}}|\, \mathcal{A} = \tilde{a}) \cdot p(\mathcal{A} = \tilde{a})}. \tag{5.6}$$

As one can see from eq.(2.20), the probability of a concrete realization of the set of observations at a certain site given its assignment to one of the aspects is given by the exponential of the local PDC costs for that aspect. Therefore, $p(\mathcal{X} = \mathcal{X}_{\mathbf{s}}|\, \mathcal{A} = a) = \exp(-h_{\mathbf{s}a})$. Furthermore, the a priori probability of an aspect model is considered to be uniform. I.e., $p(\mathcal{A} = a) = 1/n_a$, where n_a denotes the total number of aspects in the set $\mathcal{A}$. Thus one arrives at

$$p(\mathcal{A} = a|\, \mathcal{X} = \mathcal{X}_{\mathbf{s}}) = \frac{\exp(-h_{\mathbf{s}a})}{\sum_{\tilde{a} \in \mathcal{A}} \exp(-h_{\mathbf{s}\tilde{a}})}. \tag{5.7}$$

Inserting equations (5.3) and (5.7) in eq. (5.5), one sees that the probabilities of the two semantic categories given the feature observations are described by the following expressions:

$$\begin{aligned} p^{\omega_f}(\mathbf{s}) = p(\Omega = \omega_f|\, \mathcal{X} = \mathcal{X}_{\mathbf{s}}) &= \sum_{a \in \mathcal{A}_f} \frac{\exp(-h_{\mathbf{s}a})}{\sum_{\tilde{a} \in \mathcal{A}} \exp(-h_{\mathbf{s}\tilde{a}})} \\ &\text{and analogously} \\ p^{\omega_b}(\mathbf{s}) = p(\Omega = \omega_b|\, \mathcal{X} = \mathcal{X}_{\mathbf{s}}) &= \sum_{a \in \mathcal{A}_b} \frac{\exp(-h_{\mathbf{s}a})}{\sum_{\tilde{a} \in \mathcal{A}} \exp(-h_{\mathbf{s}\tilde{a}})}. \end{aligned} \tag{5.8}$$

Please note that the short-hand notations $p^{\omega_f}(\mathbf{s})$ and $p^{\omega_b}(\mathbf{s})$ have been introduced for the probabilities of the semantic categories for a given image site $\mathbf{s}$ in the equation above to facilitate later reference.

5.4 Combining Shape and Segmentation

Up to this point, a statistical representation of shape together with likelihood maps for the semantic categories *foreground object* and *background* have been established. Now, these two sources of information on object identity have to be fused to infer a segmentation of a given input image into areas which correspond to these categories. As both types of information are provided in a probabilistic form, it is most natural to combine them in the framework of Bayesian statistics. To this end, the shape information concerning the foreground object is interpreted as the prior and will henceforth be denoted by P_S. Formally, the posterior probability P_{ω_f} of the foreground semantic category at site $\mathbf{s}$ is computed by

$$P_{\omega_f}(s) = \frac{p^{\omega_f}(\mathbf{s}) \cdot P_S(\mathbf{s})}{p^{\omega_f}(\mathbf{s}) \cdot P_S(\mathbf{s}) + p^{\omega_b}(\mathbf{s}) \cdot (1 - P_S(\mathbf{s}))}. \tag{5.9}$$

However, to achieve satisfactory recognition performance, the shape information can be successfully applied in the way given in eq. (5.9) only if the prior probabilities are in accord with the likelihood p^{ω_f}. In order to strictly adhere to the Bayesian paradigm, the prior would have to be formulated as a distribution over entities like position, orientation as well as scale. Due to the fact that the considered objects can occur in both orientations at almost any location in the image at any reasonable scale, imposing something different from a flat prior would overly constrain the applicability of the proposed method on unseen examples. Furthermore, the computational evaluation would be even more demanding than it already is, endangering the practical application of the approach in cases where short response times are crucial.

Thus one seeks for the best alignment of the shape probabilities with those regions in the image which display high likelihood values for the foreground object. This step requires to shift the cloud of prior probability values to a new location such that the foreground object is optimally covered. Denoting the shifted prior probabilities by $\overrightarrow{P_S}^{\mathbf{c}}$ and its center by $\mathbf{c}$, the optimal position $\mathbf{c}^*$ can be found by solving the maximization problem

$$\mathbf{c}^* = \max_{\mathbf{c}} \sum_{\mathbf{s}} p^{\omega_f}(\mathbf{s}) \cdot \overrightarrow{P_S}^{\mathbf{c}}. \tag{5.10}$$

Moreover, since the shape prior is not mirror symmetric, both possible orientations are checked during the alignment. The decision concerning the most probable configuration is met on the basis of highest posterior probability. Small changes in scale of the object of interest do not significantly affect the segmentation results due to the probabilistic prior representation. However, if these size variations exceed a certain limit, they must be accounted for by an appropriate scaling of the prior. Clearly, enlarging the integrated prior probability mass by a factor λ always produces a higher posterior for the given alignment. Therefore, a direct comparison of the posterior probabilities over different scales can not guide a meaningful decision for the scale parameter. Under the condition that the scaled prior exclusively covers areas of high likelihood ($p^{\omega_f} \approx 1$), the sum of the posterior probability mass also scales according to λ. In order to take the scaling-behavior of the posterior alignment score into account, the selection of an appropriate prior starts with the smallest reasonable size. Alignment scores are then computed for larger versions, where the range of possible sizes is coarsely sampled. A larger prior is accepted, if it improves the score by at least $0.8 \cdot \lambda$, i.e. it almost exclusively covers areas of high likelihood.

One has to be aware of the fact that from a pure Bayesian viewpoint the priors as proposed above have been adapted to the data and, therefore, do no longer qualify as true priors. However, the term shape prior still be considered appropriate, since the a priori selected form is not changed by the adaptation step but only its relative position to the image. The processing pipeline of the shape constrained segmentation procedure is outlined in figure 5.3.

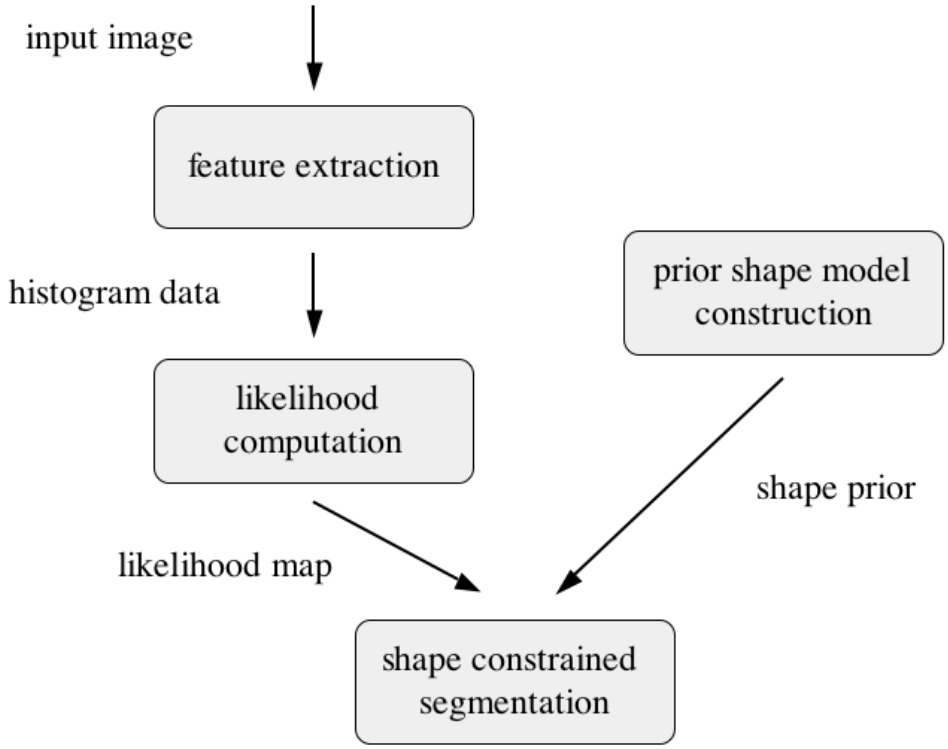


Figure 5.3: SCS processing pipeline.

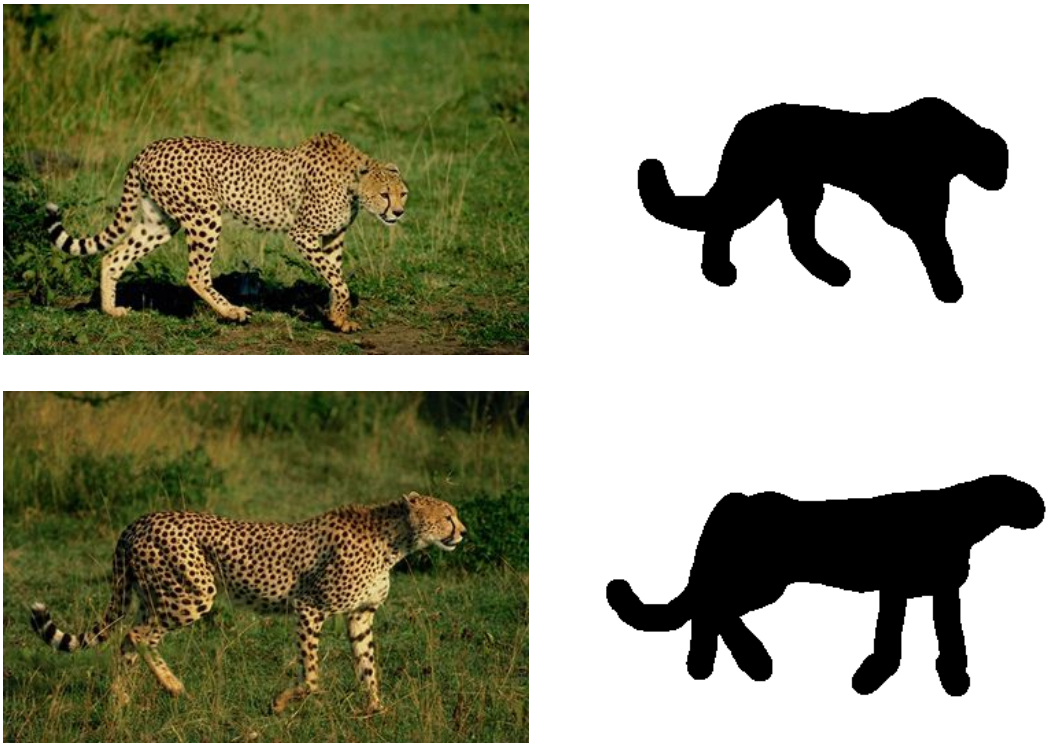
5.5 Implementation and Experimental Results

5.5.1 Dataset and Features

The images, on which the shape constrained segmentation approach has been tested, come from the Corel image database. In particular, the set *Cheetahs, Leopards and Jaguars* has been used. It contains one hundred images of wildcat from the aforementioned species, where 34 of them depict cheetahs. Out of this group, 14 show a single animal in a sideward view, so they where selected as the actual testbed for our method. The input data for the PDC-based approach to image segmentation are histograms of feature values taken at image sites lying on a regular grid. For the experiments which are described here, a fine grid with an inter-site spacing of two pixels has been used. The local neighborhoods from which the histograms have been drawn have a size of 11×11 pixels. The features which were subjected to the histogramming procedure are the values from the three color channels of the original input images together with the magnitudes of the Gabor filter-bank that has already been used for the segmentation experiments in the previous chapters. Consequently, this procedure yields an eleven-dimensional feature vector at each pixel.

5.5.2 Shape Prior Construction

As a first step in the construction of a shape model for a standing cheetah in sideward view, rather sketchy hand-segmentations for 6 images have been made using a standard painting application. Some examples are shown in the top row of figure 5.1 as well as in fig. 5.4. The latter figure in particular demonstrates the relation between the images from the test-bed and the manually constructed segmentations of the objects of interest. These hand-segmentations where then



(a) (b)

Figure 5.4: Sketchy hand-segmentations in shape prior construction: a) input image, b) hand-segmentation.

centered, in order to arrive at a common standard configuration. After that, the now centered images have been processed by a distance transform (chamfering). In the next step, a Gaussian probability function with $\sigma = 10$ has been applied to the distances, transforming them into probabilities while leading to a steep decay of values in the outer regions of the images. Having averaged the shape probabilities, an additional Gaussian blurring with a stencil size of 10×10 pixels has been applied. The final probabilistic shape model resulting from this procedure is depicted in the last row of figure 5.1, where large probabilities are shown in dark colors. Please note, however, that the procedure described above

is not the only possible way of arriving at a probabilistic shape description for the shape constrained segmentation approach. One could as well start with a single hand-drawn sketch of the object of interest which captures its essential shape properties, applying the distance transform and the Gaussian model to this single image. In such a way, the shape constrained segmentation approach can be utilized in a content based image retrieval system with user interaction.

5.5.3 Aspect Model Generation

As has been outlined before, apart from the probabilistic shape information, one important building block of the SCS approach is given by the computation of likelihood maps for the semantic categories. Consequently, the system has to be provided with a database of probability distributions encoding the statistical properties of features for these categories. In order to produce such a database, the test set of images has been segmented by a PDC model of ten clusters. Each segmented image thus led to a set of ten distributions. Eight out of the total of 14 images were selected as sources of aspect models. From these remaining images, a set of 12 distributions for the foreground, and another 12 distributions for the background have been selected by user interaction. Here, the number of images as well the number of aspects can be freely chosen. The goal behind the selection process was to acquire enough models in order to faithfully capture the variety inherent in the instances of the semantic categories while still demanding a good generalization performance on unseen examples.

As has been demonstrated in section 5.3, the computation of aspect likelihoods relies on the local PDC costs of the given input data at a site. Therefore, it is very instructive to examine the relationship between these two entities. The transition from the local PDC costs to the likelihoods of a certain semantic category serves two main purposes. First and foremost, one arrives at a probabilistic measure of association between the regarded site and the prototypical distributions for the relevant semantic categories. This association is indispensable for the later combination with shape-information in the Bayesian setting described in section 5.4. Furthermore, the aspect likelihoods enable a comparison of aspect association strengths across different sites. This property does not hold for the local PDC costs themselves. Experiments showed that for each site, the local PDC costs are in good accordance with the expectations in the sense that a site which belongs to the semantic category of interest almost always has lesser PDC costs with respect to the aspects from the foreground-set, than to the background. The converse holds true for the case of sites from background regions of the test-images. On the other hand, when one compares local PDC costs across different sites, the surprising finding was that sites from the background at times had lesser PDC costs to the foreground aspect set than other sites that clearly belonged to foreground regions. This means that, although the relative ordering of aspect costs at any site is in good correspondence to the true semantic category, the

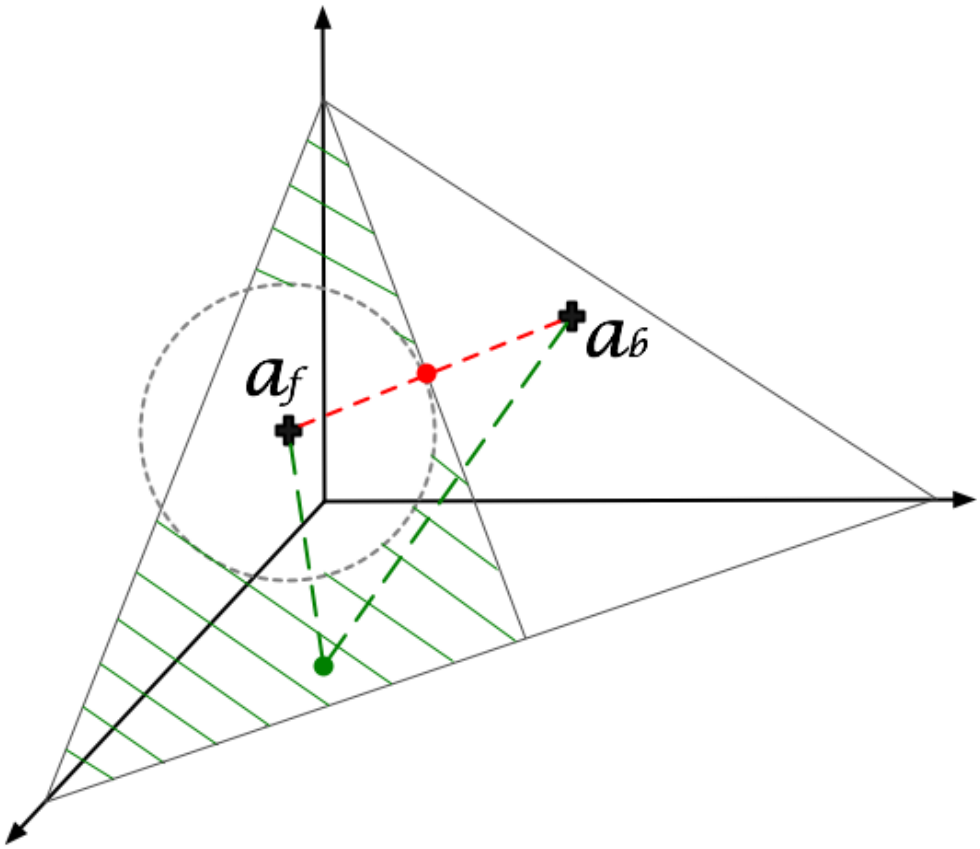


Figure 5.5: Symbolic depiction of the geometry in PDC cost-space.

costs at different sites are not comparable with each other, due to the lack of a common scale.

After numerous experimental investigations, the source of this problem could be tracked down to the employed mode alphabet. For reasons of simplicity, a common alphabet has been used for all PDC clustering solutions, from which aspect distributions have been derived. In order to construct this common alphabet, the set of chosen example images has been coarsely subsampled. From the resulting joint set of feature vectors, the mode alphabet has been generated by a mixture model estimation step with the prescribed number of modes. Clearly, the images of the test set contained far more feature vectors from the background, than from the foreground category. Consequently, the background has been far better represented by the mode alphabet. In order to remedy this situation, another mode alphabet has been constructed, containing sixteen modes that were solely computed based on background feature vectors and another sixteen modes that were derived explicitly from data of foreground regions. Using this adapted alphabet, the problem described above did no longer occur. These findings are exemplified by three examples given in fig. 5.6. In the top row, the example

images are depicted. Then, the second row shows the aspect likelihoods derived according to the approach described in section 5.3. For all subsequent visualizations in this figure, higher values of the depicted entities are indicated by darker shades of grey.

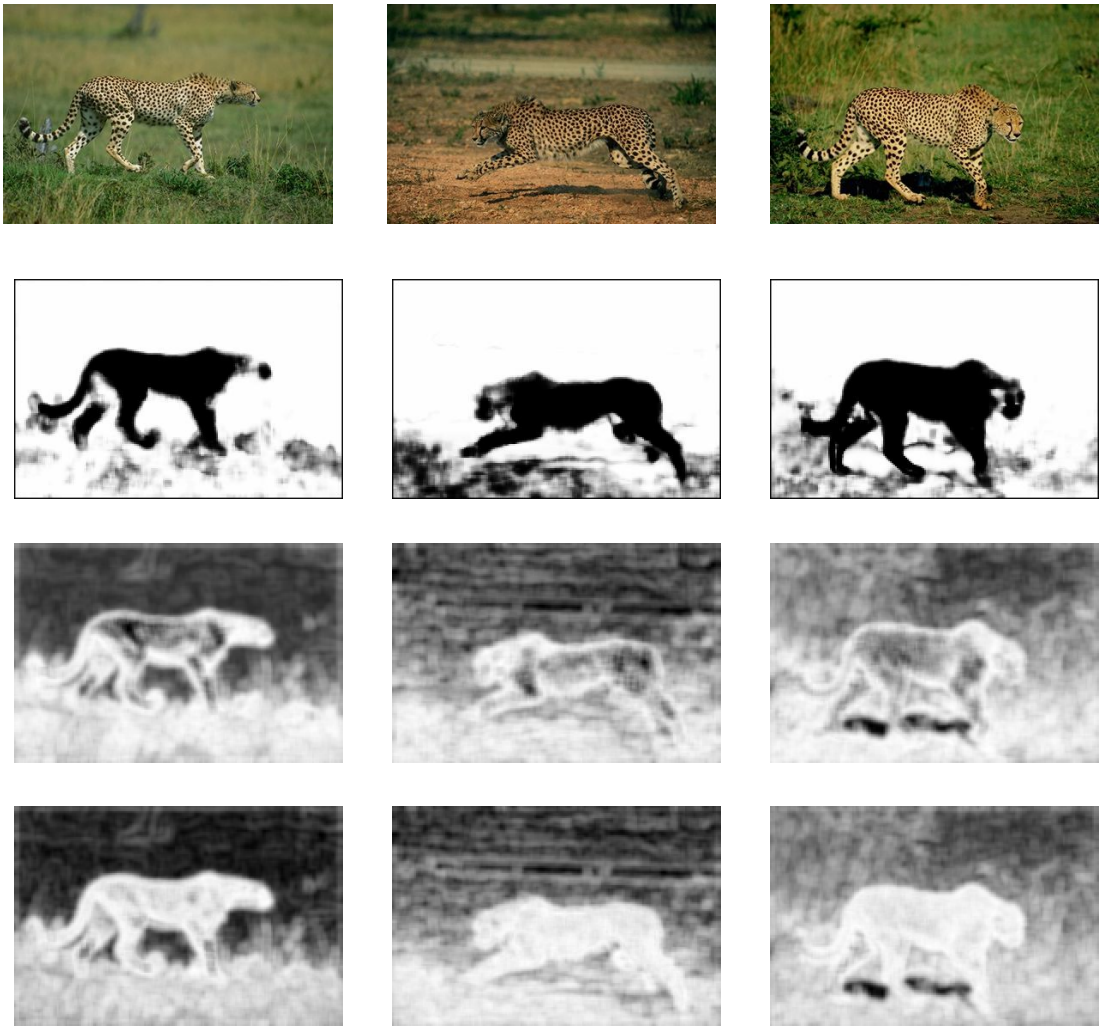


Figure 5.6: Aspect likelihoods in comparison to PDC costs. First row: test images; second row: aspect likelihoods; third row: PDC costs for non-adapted mode alphabet; fourth row: PDC costs for adapted mode alphabet. For all visualizations, higher values are indicated by darker shades of grey.

Although a non-adapted mode alphabet has been used, the likelihoods are not only in good accordance with the true semantic categories, moreover, they are also comparable across different sites due to the fact that they are true (i.e. normed) probabilities. The third row shows visualizations of the PDC costs to the nearest aspect from the cheetah category. Here, one can clearly see that the cheetah

region is largely inhabited by sites that have higher minimum PDC cost to the cheetah aspects, than large portions of the background sites. Finally, the last row depicts the results when using the adapted mode alphabet as described above. Now, the magnitude of cost terms across different sites is in good accordance to the true semantics of the input images.

The observed geometry in PDC cost-space is depicted symbolically in fig. 5.5. Here, an aspect from the foreground as well as one from the background are shown as black crosses. A site from the background is shown as a red dot, while a corresponding one from the foreground is visualized as a green dot. The distances (i.e. the assignment costs) to the aspects are indicated by dashed lines. Although the relative cost for the foreground site to the foreground aspect is lower than its distance to the background aspect, it is still larger than than the corresponding distance of the background site. The same holds true for all data-points inside the hatched region.

Summing up the discussion above, the transition from PDC costs to aspect likelihoods not only enables the formulation of the shape constrained segmentation approach in a Bayesian setting, it also leads to association strengths of sites to aspects that refer to a common scale, even if no specially adapted mode-alphabet is used.

5.5.4 Prior Alignment

Clearly, the proper alignment of the shape prior is of crucial importance for the successful application of the shape constrained segmentation approach. In this respect, the ability to successfully adapt to changes in the scale of the object plays an important role. Therefore, two examples of the scaled prior alignment are provided in figure 5.7. Due to the probabilistic nature of the shape prior, it possesses an inherent robustness towards small changes in the scale of the objects of interest. Consequently, the option of choosing a scaled prior has been adopted only for one example in the data set. This particular image shows a close-up of a cheetah, thereby warranting the selection of an enlarged shape prior. That image together with the prior alignment and the final shape constrained segmentation is shown in the first column of figure 5.7. Considering the fact that this image has not been used for the construction of the shape prior, nor has it contributed to the aspect data-base, the segmentation quality is very good. In order to provide another example for the scale selection method, one image of the original data-set has been manually enlarged to 130 % of its original size. After that, this example has been cropped to the size of the other pictures in the test-set. The alignment process has then been run for that image, leading to the results depicted in the second column of fig. 5.7. In this example, an image of a relatively distant, and thus small, cheetah has been used. As the alignment procedure is able to cover almost the complete area depicting the cheetah with the original prior, the shape representation has not been enlarged by the alignment procedure.

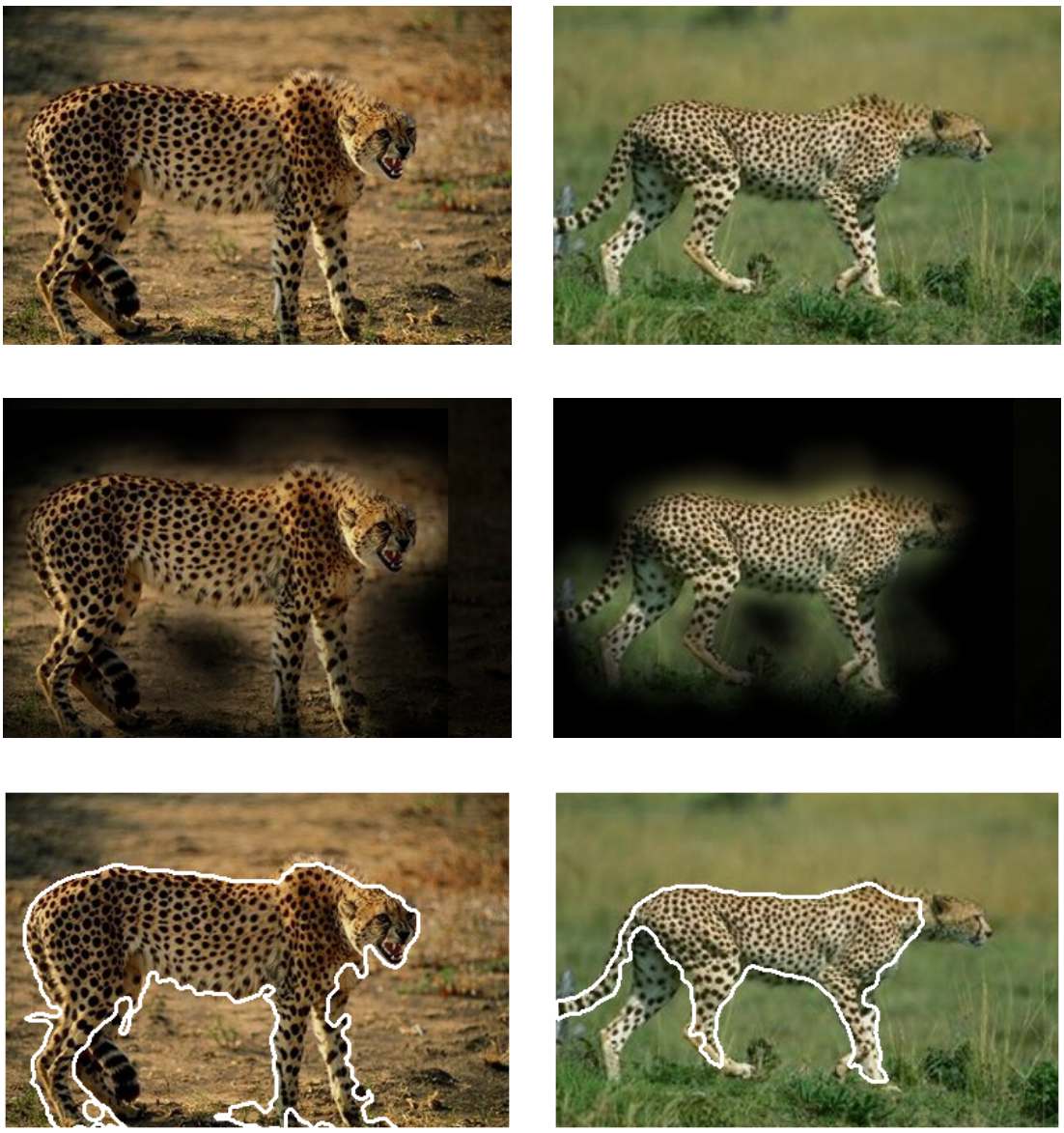


Figure 5.7: Results of the scaled prior alignment procedure.

5.5.5 Shape Constrained Image Segmentation

To turn the probability assessment for the semantic categories into a segmentation for the given input image at hand, each image site is assigned a label according to the maximum of the posterior probability values for foreground and background respectively. After computing this labeling, one sweep of a post-processing step has been applied to the segmentations in which each site is relabeled according to the majority of assignments in a local neighborhood of 7×7 sites. Thereby, one

arrives at a mildly smoothed version of the original segmentation. Even after this step, spurious small regions (consisting of less than 500 sites) might still persist. Thus, another post-processing step is applied in which all regions with areas below the aforementioned threshold are eliminated. The result of this procedure for a collection of images is displayed in figures 5.8 and 5.9. In both figures, the input images are shown in the first row, followed by the visualization of the prior alignment in the second row. Finally, the shape constrained segmentation results are depicted in the third row.

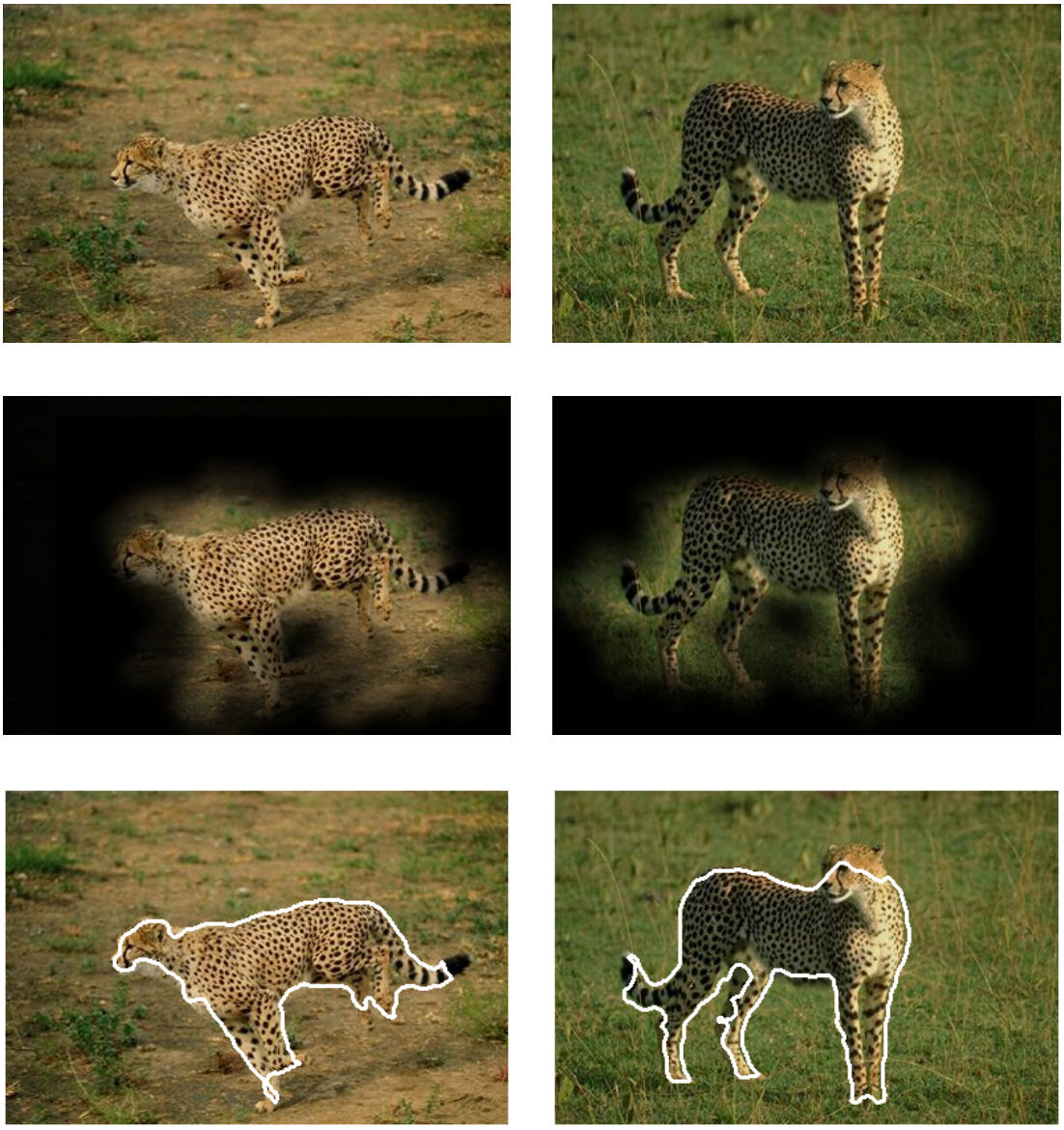


Figure 5.8: Shape constrained segmentation results, (example set one).

The first example in figure 5.8 displays a cheetah from which aspect distributions have been included in the aspect data-base. Please note, however, that it is rather similar in color to its brownish background. Moreover, it exhibits an leaping posture which is not well accounted for by the shape prior. Despite these restrictions, the final SCS aggregation result precisely segments the animal from the remaining image. The second picture in this set shows a cheetah which is better silhouetted against the background. It has been used as a model for the shape prior and aspect distributions have been drawn from a PDC segmentation for this image. However, the posture of the head differs from the majority of examples used for generating the prior, thus this information has been virtually lost in the averaging process. This shows in up in the final segmentation insofar as a small part of the head is missed. Apart from this restriction, the cat is very precisely segmented from the background.

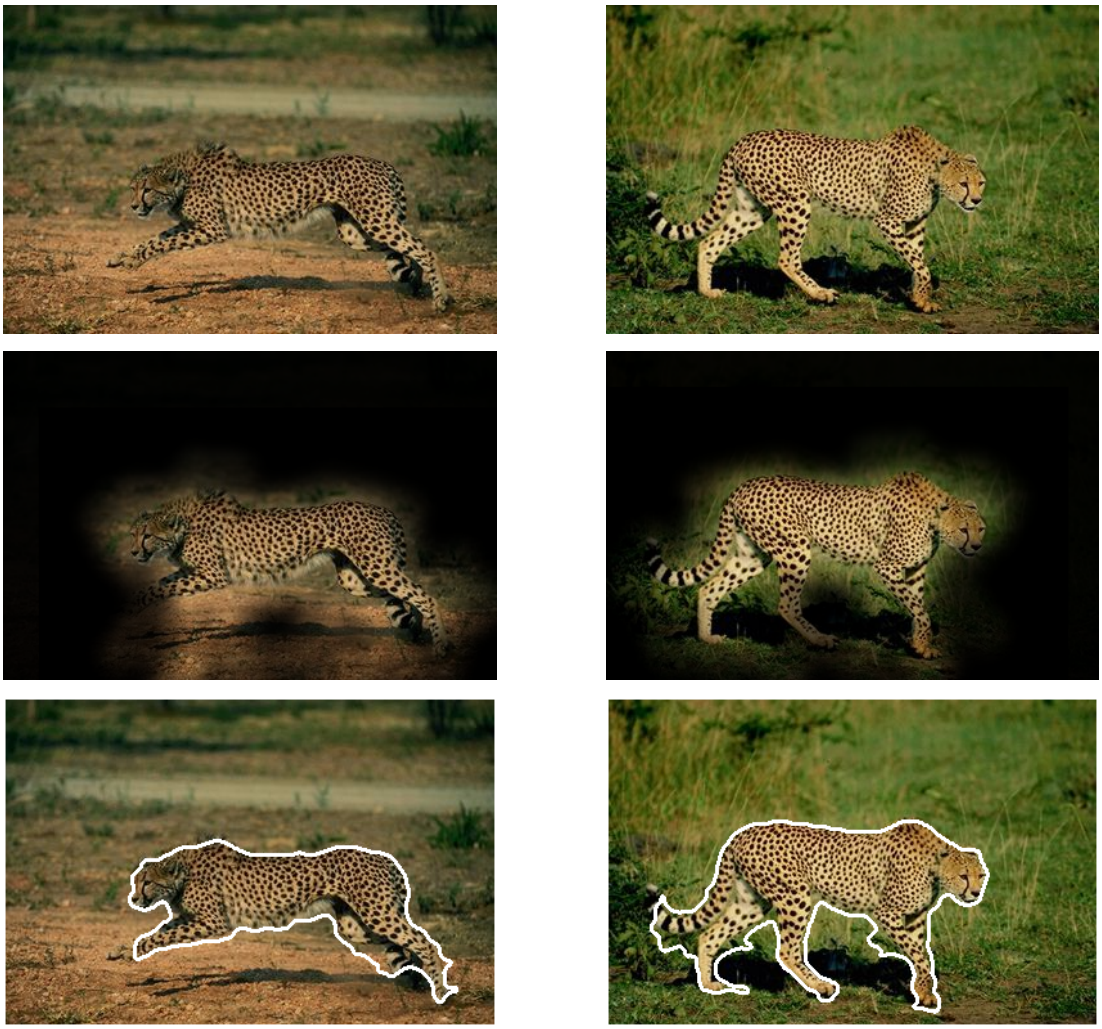


Figure 5.9: Shape constrained segmentation results (example set two).

The first column of the second result collection in figure 5.9 shows another example of an image for which no aspect distribution is in the set $\mathcal{A}_f$. Moreover, its hand-segmentation has not been utilized in the formation of the shape prior. Still, the shape prior alignment process is able to identify the correct orientation, i.e. the selected prior model is facing leftward, and the area covered by the animal is represented well. Due to the leaping posture of the wildcat, a minor part of its forelegs is missed by the prior. This, however, does not adversely affect the final segmentation based on the posterior probabilities. Only the very small area of the front feet is not accounted for. This result is contrasted to the outcome of the shape constrained segmentation process for the image in the second column. Here, aspect models have been taken from the image. Although it also contributed to the shape prior construction, the stepping position of the right hind leg did not prevail in the averaged prior. Consequently, the lower part of this leg receives comparatively little prior probability mass. Nevertheless, as the last picture in this column shows, the segmentation achieves a good representation of the cheetah, although a minor amount of the background is also attributed to the wildcat. In these parts, small branches of plants account for a rather strong response of the texture filters which is noticeably higher than in other regions of the background. Moreover, the applied Gabor filters have a certain spatial extend proportional to their scale, thereby gathering texture information not only for one site, but also for other neighboring locations on the site grid. Therefore, a precise localization of segment borders becomes rather difficult, as the scale of the responding filter gets larger.

To demonstrate the gain in segmentation performance due to the incorporation of shape constraints, figure 5.10 shows a comparison between the pure PDC segmentation without shape information and the results achieved by the SCS method. In the first row, the original image data is depicted. The second row shows the segmentations produced by the PDC approach using color and texture features while partitioning the data into ten clusters. Finally, the last row exhibits the results of the application of the SCS method.

In all of the results the images neither contributed to the set of aspect models, nor have they been used in the generation of the shape prior. As one can see from these examples, the PDC segmentation is able to divide the images into homogeneous regions based on the low-level features. It fails, however, in identifying the cheetahs as contiguous segments. These findings are in contrast to the results of the shape constrained segmentation approach. The images in the first column of figure 5.10 show results for a particularly interesting case. Here, the leaping pose of the animal is remarkably different from the pose of the prior. Nevertheless, the majority of the area belonging to the wildcat is preserved by the prior. However, some minor amount of background is also contained in the segment of the cheetah. This observation is explained by the fact that the prior, not being perfectly adapted to the posture of the animal, puts a considerable amount of probability mass in this area. Moreover, the low-level features *color*

and *texture* are not very discriminative in this case. The color impression of the dry mud and that of the animal is very similar. Due to the small irregularities on the background plain, the texture filters also respond in a similar way. Bearing these restrictions in mind, the cheetah is identified with a surprising accuracy.

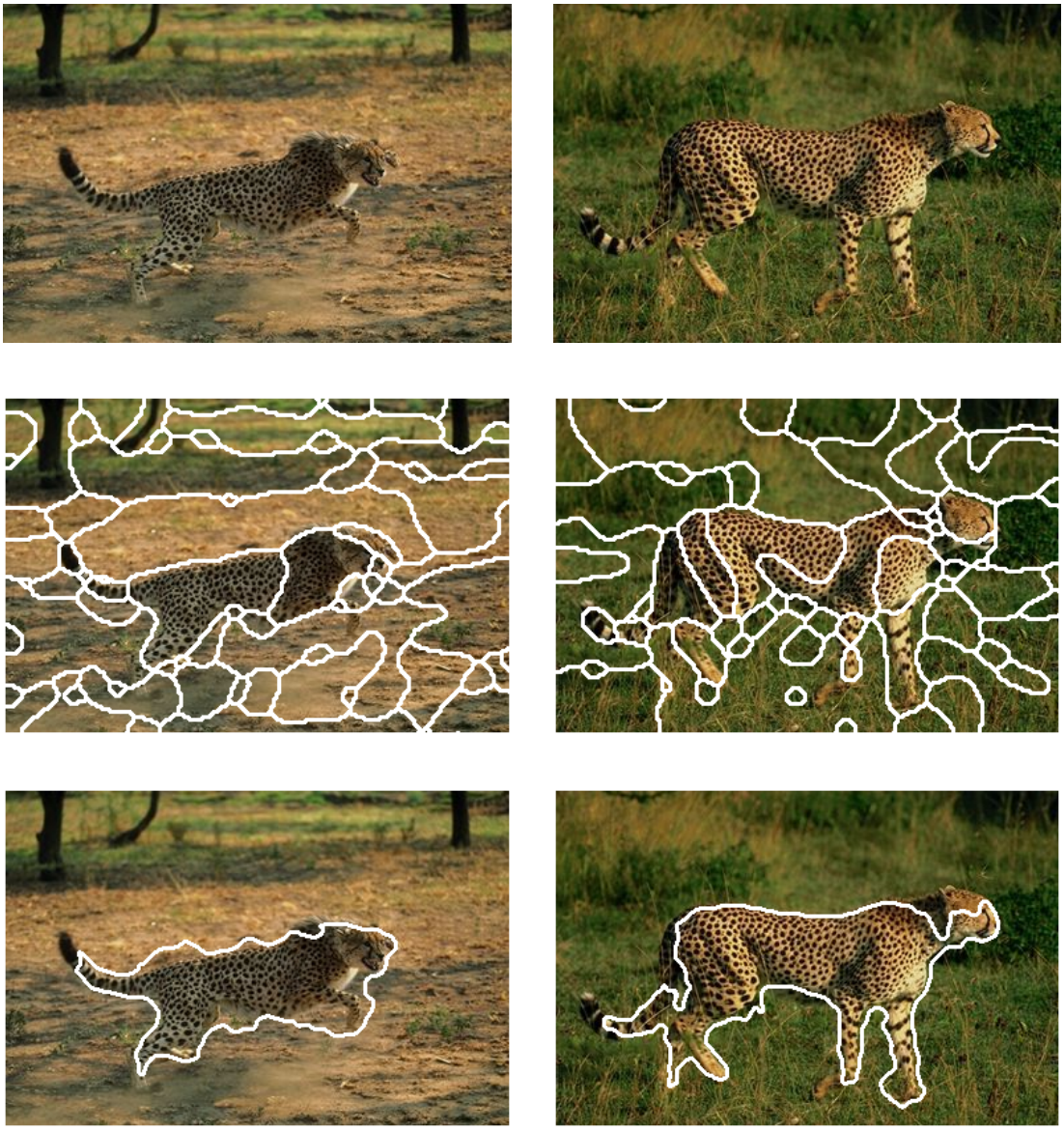


Figure 5.10: Comparison between segmentations with and without shape constraints (example set one).

The example in the second column depicts segmentations for a cheetah in front of a dark green background, while some parts of the animal are covered by shadows. These differences in terms of the low-level features lead to the very frag-

mented results achieved by the PDC approach. These adversary conditions also apply to the SCS method, because no aspect models of this cheetah contributed to the aspect data-base. Moreover, the pose of the animal does not closely match the form of the shape prior due to the fact that the second foreleg comprises an individual part of the cheetahs silhouette. The shape constrained segmentation approach is nevertheless able to group most part of the animal as the foreground segment. Only small parts at the head and tail containing no texture information are missed.

5.5.6 Generalization To Other Semantic Categories

Up to this point, the focus of the experimental evaluation has been put on finding animals from the zoologic category of cheetahs. In order to demonstrate, that the proposed SCS approach is of a generic nature, a second test-bed has been created. Here, the category of interest is given by jaguars. Although this setting is similar to the first one in the sense that the relevant objects are also wildcats, the fur pattern and the typical surrounding background are noticeably different. Therefore, it can be used to demonstrate the application of SCS to novel domains.

As a first step, a second aspect set has been constructed in the same way, as for the cheetah examples. This time, however, sPDC with eight clusters has been utilized in order to arrive at the segmentations from which the aspect distributions have been selected. The data-base of jaguar images consists of sixteen pictures. From the pertaining segmentations, ten cluster-specific distributions for fore- and background respectively, were chosen to form the aspect-set. Emphasizing the generic nature of the approach, no adapted shape-prior for jaguars has been derived. Instead, the same shape-prior as for the cheetah examples has been used. Ignoring the special shape properties of jaguars, a noticeably lesser segmentation quality can be expected.

Nevertheless, figure 5.11 demonstrates that this reduction in quality does not take place. On the contrary, bearing the restriction in the form of the non-adapted prior in mind, SCS is still able to identify the objects of interest with remarkable accuracy. Aspect distribution have been taken from all images, except for those in the last row.

In the first example on the left of the topmost row, the contour of animal has been precisely found by SCS, only a very small part of the tail and the snout are missed due to their high similarity to the surrounding background. The result for the neighboring image on the right of the first row is even more striking. Although the identification of the jaguar's outline around the hind legs is not so precise as in the previous example, the segmentations is still very good, keeping the strongly differing pose of the depicted animal to the shape-prior in mind.

The SCS segmentations in the middle row exhibit a quality that is comparable the the one of the first image in this result collection. Only very small parts of the animals are missed in the segmentation. Again, these latter regions are

either very similar to the background (e.g. at the heads and feet of the jaguars), or constitute narrow, elongated structures (e.g. the tail of the animal in the first picture of the middle row), which makes them hard to identify because of the spatial extend of the applied texture filters.

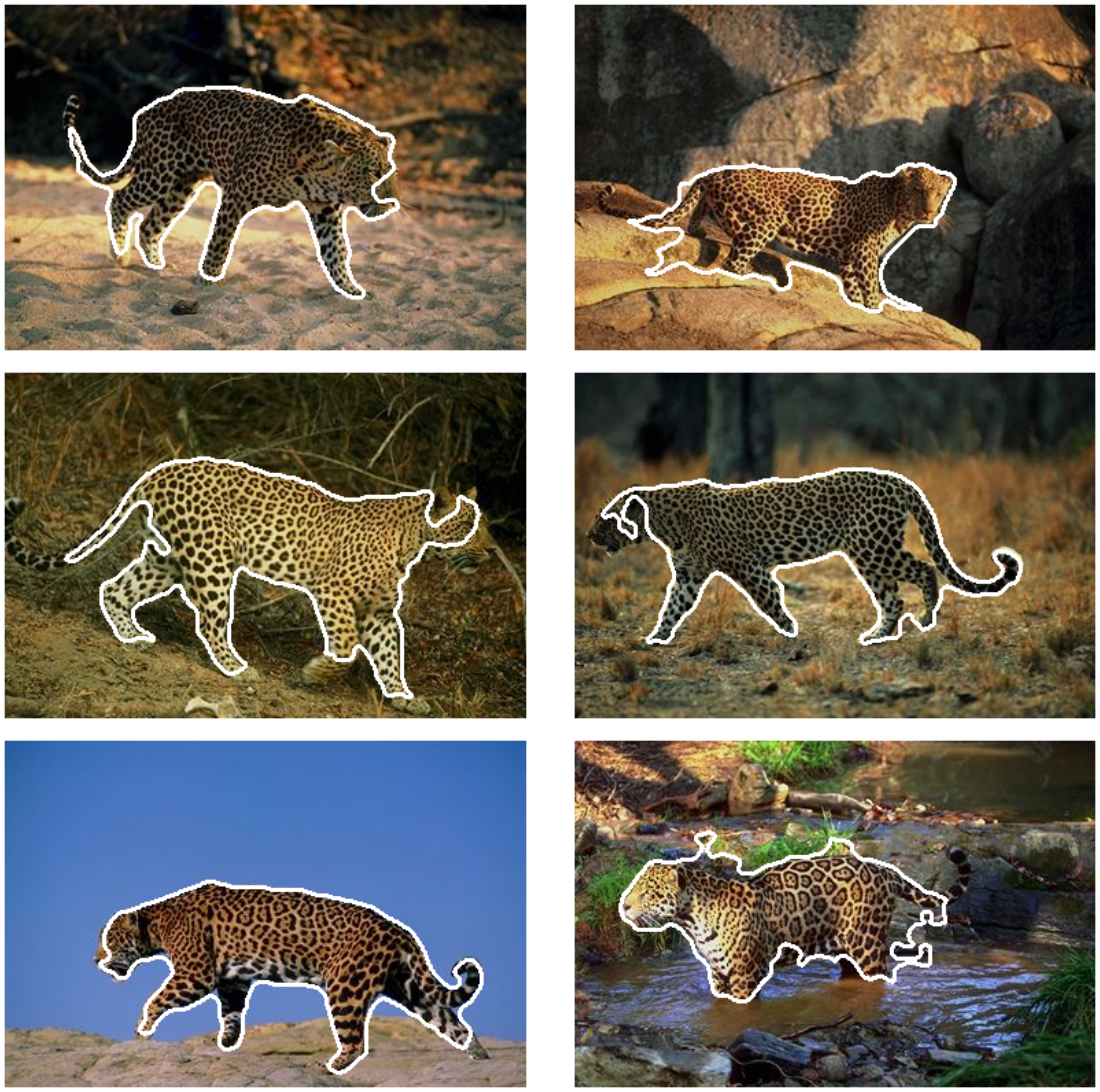


Figure 5.11: Shape constrained segmentation results (example set three)

The results for the examples in the bottom row of fig. 5.11 are noteworthy, because no aspect distributions have been taken from the pertaining segmentations. The first image of this row shows an animal that is depicted in front of a blue sky background. Thus, identification can be considered an easier task than for the other examples. As a matter of fact, the SCS segmentation for this particular image shows a very good quality. This finding again demonstrates the good representation quality of the relevant features of a given semantic category by the aspect distribution set.

The last image in this result collection exhibits a segmentation that, at a first glance, does not keep up with the good quality of the other examples. On the other hand, it is the hardest test-case in the compilation of SCS results. First of all, the depicted jaguar is standing amidst a very irregular and cluttered background. Moreover, no aspect distribution has been selected from the corresponding PDC segmentation, consequently, the computation of the likelihood for the semantic category jaguar has to rely on the similarity to the other feature distributions in the aspect set. Apart from this, the pose of the jaguar is noticeably different from the shape information encoded in the prior. Taking all of these restrictions into account, the segmentation reached by SCS is still comparatively good.

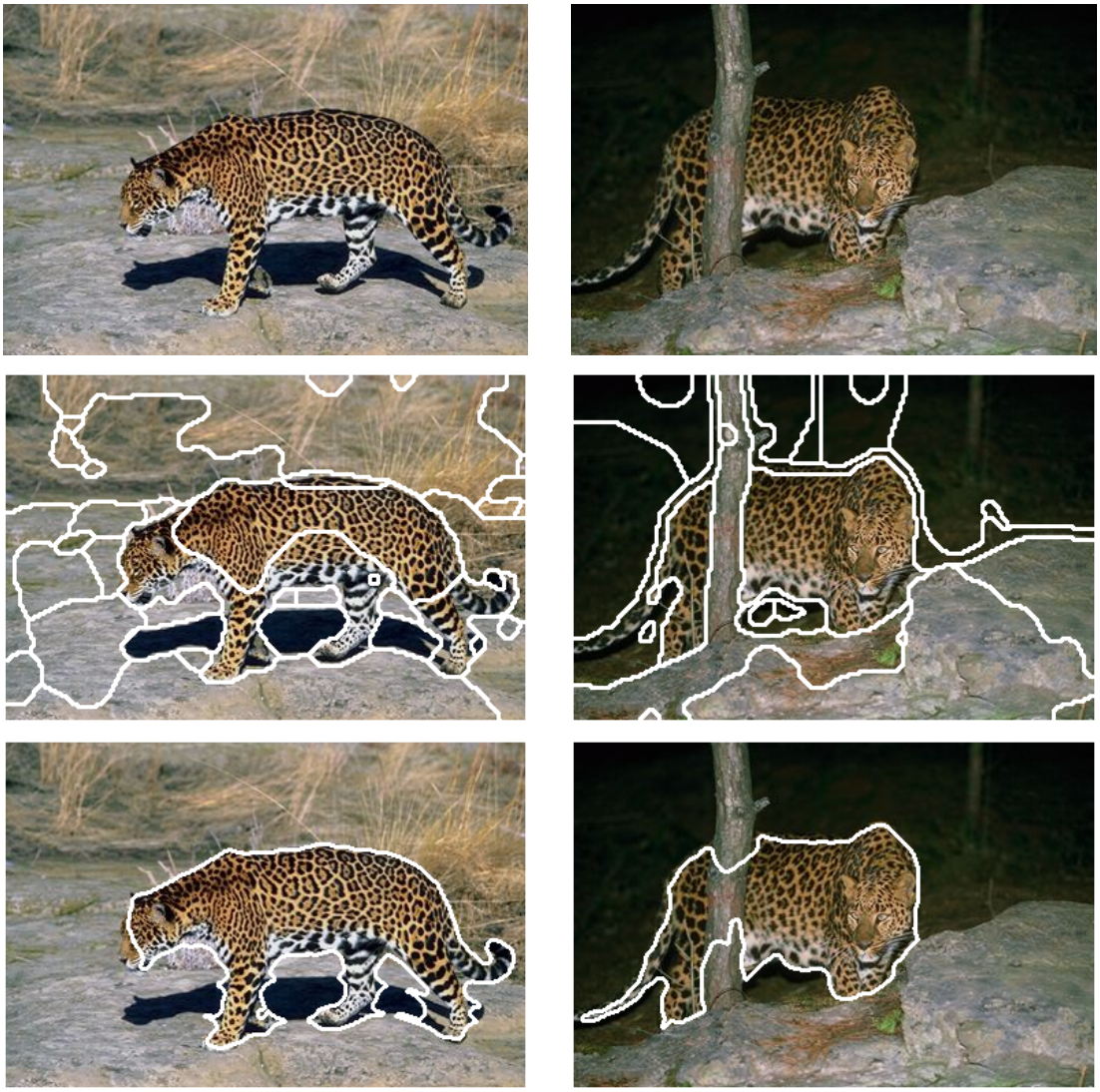


Figure 5.12: Comparison between segmentations with and without shape constraints (example set two).

In the next collection of SCS results for the jaguar data-set, the shape constrained segmentation is opposed to sPDC image partitions with eight clusters. In the first row, the input image is shown. Row two then depicts the sPDC partitions in segment outline visualization, while the last row shows the results of SCS. From the segmentation of the image in the left column, no cluster model has been included in the aspect set. The picture on the right hand side, however, has contributed to the aspect set. For the first picture, the sPDC segmentation closely matches the feature content. However, the animal is not inferred as one contiguous segment, but is assigned to multiple clusters. Opposed to this, SCS is capable of inferring the jaguar as one segment. Only a small fraction at the snout is missed. This error is most likely caused by the shape-prior, which favors a more rounded outline of the head area. The second example is particularly interesting, because the depicted wildcat is standing behind a tree. Clearly, the sPDC segmentation assigns the area covered by the tree to a different cluster than the jaguar. Consequently, the animal is represented by two different segments. In contrast to this, the SCS image partition is able to identify the jaguar as a coherent segment by virtue of the shape-prior information. Thereby, the jaguar's outline is reliably inferred despite the occlusion.

5.6 Summary

The central motivation for the work described in this chapter stems from the apparent shortcomings of low-level segmentation techniques to reliably capture the semantic properties of images. Consequently, bringing higher-level knowledge sources to bear seems a promising way of significantly enhancing segmentation performance. This conjecture is particularly evident in those cases, where the image partition is only an intermediate step in a chain of processes finally leading to a full scene interpretation. Therefore, I proposed the incorporation of higher level knowledge in the form of shape information to augment the PDC segmentation approach. The resulting SCS method is formulated in terms of the sound theoretical framework of Bayesian statistics, representing both shape and elementary image features in a probabilistic way. Experiments demonstrating the effectiveness of the method have been conducted on a hard real-world data-set (i.e. cheetahs in images of their natural environment). Taking the large diversity of the images both in terms of shape and of color / texture features into account, satisfactory results have been achieved. Although only one pose has been used in the experiments, the generalization to new poses is straightforward. All that needs to be done is the provision of a shape prior, which can be arrived at either by averaging over a number of sketchy segmentations (as described in section 5.2) or by taking a user-drawn shape as the starting point for the computation of prior probabilities. The successful application of the SCS method to objects from a different, although similar, semantic category (i.e. jaguars) has been exemplified,

thereby showing the extensibility to novel target objects. There exists, however, a drawback that needs to be addressed: The approach is based on the assumption that the objects of interest exhibit a characteristic distribution of elementary image features like color and texture. Clearly, there exist object categories, for which this assumption is violated, e.g. cars. The question of how to successfully incorporate shape knowledge in those cases provides for an interesting direction of further research.

5.7 Bibliographic Remarks

Shape representation, analysis and matching constitutes a very large and active field of research. A good introductory book is [dFCC01]. Current developments are described in [RS98, LL99, LL00, MA02, BMP02]. A relatively recent survey of the relevant literature can be found in [Lon98]. Shape constrained image segmentation has recently matured to a very vivid research area too. Rousson and Paragios [RP02] proposed a relevant approach in the level set framework. In their model, the region boundary is represented by the zero level set of an embedding function. Interesting contributions to this domain can also be found in the work of Cremers et al. [CKS02] and [Cre03]. Here, the segmentation problem is formulated as a variational approach, where the latter publication is concerned with using motion information as a segmentation criterion. The application domain of video segmentation is also addressed by Jehan-Besson et al. in [JBBAF03]. The pertinent work of Zhu and Yuille [ZY96], which addresses the problem of shape constrained segmentation from a Bayesian– / MDL– point of view, should also be noted here. In another recent publication, Galun et al. [GSBB03] use small scale shape properties of image patches in order to guide a texture-based segmentation approach.

Chapter 6

Conclusion

The central theme of this thesis has been to develop robust algorithms for the task of image segmentation. To achieve this goal, all stages of the segmentation process have to be addressed. The first question which has to be answered is how to reliably measure image content. In this respect, the use of local histograms as proposed in this thesis provides for an substantial advantage over point-wise measurements of image features, due to the fact that the latter entities are much more strongly affected by noise, which is invariably present in all observational data.

Once the form of the input data has been specified, corresponding segmentation algorithms can be developed. All approaches that have been proposed in this thesis are based on the sound modeling of the image formation process, i.e. parametric probabilistic models are designed, from which images instances can be generated by sampling. Therefore, all modeling assumptions are made explict and cost functions for the segmentation problem can be derived in the maximum likelihood framework. In my view, this decisive feature sets the image partition techniques described in this contribution apart from the vast majority of competing techniques that are based on some kind of heuristics. Moreover, all elementary segmentation techniques from the PDC family are completely unsupervised, i.e. no label information has to be supplied in the training step of the pertaining models. This feature contributes to the broad applicability of these methods, as the labeling of sufficient amounts of data is not only inconvenient, but usually also implies prohibitively high costs. Furthermore, the methodological approach that is advocated here leads to a clean separation of the modeling step and the subsequent optimization. The latter problem has been addressed by the application of highly robust Deterministic Annealing (DA) techniques where the computational demands were further alleviated through the application of multi-scale techniques.

The basic PDC method was explicitly designed to overcome the insensitivity of previous approaches to histogram clustering concerning the permutation of bins. Based on the standard PDC technique, improved variants have been derived, which incorporate topological assumptions in the corresponding cost functions. Natural images possess the property of being composed by large, homogeneous

regions. This valuable prior information clearly should not be neglected in the design of segmentation techniques. Consequently, the sPDC approach emphasizes the spatial topological relations between the different images sites. Thereby, highly fragmented and irregular image partitions are substantially suppressed, leading to segmentations that show a significant increase in precision, i.e. the probability, that an inferred boundary between different image regions is valid. On the other hand, the TPDC approach enforces topological relations in the cluster-domain. This technique has been inspired by self-organizing maps, discovered and popularized by Kohonen, which stimulated a flurry of research papers in the neural network community. By means of TPDC, the similarity of cluster models can be directly read off the cluster labels, thereby making the neighborhood relationship in the cluster domain readily accessible from the indices of the data-groups. Furthermore, the two types of topological constraints are simultaneously obeyed by segmentation solutions of the sTPDC technique.

Quantitative measurements and their evaluation provide the basis of the scientific method. Unfortunately, assessing the performance of image segmentation techniques is treated in a qualitative and subjective way in the majority of the pertaining literature. In contrast to this attitude, the thorough quantitative evaluation of all segmentation methods proposed in this thesis is strongly emphasized. It has been demonstrated, that the PDC family of image partition techniques is capable of producing good results throughout the test-runs. The sPDC approach, which achieves the highest quality among all PDC variants, is able to arrive at an F-measure of 0.5 at 0.54 recall and 0.51 precision when averaged over the different cluster numbers in the experimental setting. The human subjects in the Berkeley study reach an F-measure of 0.8, when results of varying subjects are compared with each other. The only competitor, for which published test-statistics were found, is given by the combination of three edge detectors (brightness, color, and texture gradient), which have been trained in a supervised fashion on the segment boundaries produced by the human subjects [MFM04]. Moreover, the best way of combining these aforementioned cues in a single detector was also found by means of supervised learning. This joint boundary detector achieves an F-measure of 0.67 with a recall of 0.71 and precision as high as 0.64. Summing up, sPDC is on average capable of reaching 63% human performance and 75% of the performance of the best known competitor. The latter quality comparison might not seem satisfactory at a first glance. On the other hand, it has to be kept in mind, that the combined edge detector has two distinct advantages when it comes to the evaluation of quality measures, which are at the same time two disadvantages concerning its general applicability. First of all, it is not a true segmentation technique as it does not produce coherent image segments. In many application domains, the latter feature is of prime importance. Second, it relies on supervision information in the training phase. No matter how well-chosen the training set might be, there are always novel application scenarios, which have not been accounted for in the detector's training. Consequently, the performance

is prone to deteriorate in those settings. Furthermore, additional training data (i.e. human segmentations) are not readily available. Methods from the PDC family, on the other hand, do not need any additional information that is not already contained in the measured histograms. Moreover, apart from the very general assumptions that constitute the PDC approach, no further suppositions concerning the image data are made.

The goal of all machine learning techniques is to faithfully represent the structural properties of the given input data, while avoiding to fit the random fluctuations, which are inevitably present in all observational data. As all PDC variants are clustering approaches, and as such unsupervised learning techniques, the questions of robustness against small fluctuations in the input data and the generalizability of solutions to new data-sets from the same domain are of prime importance. In this thesis, these issues have been addressed in an empirical manner, giving comprehensive example sets for both problems. It has been shown, that the PDC framework is indeed capable of producing highly robust image segmentations. Moreover, re-sampling techniques like the described bootstrap-variant for images have been demonstrated to be viable approaches for the further improvement of PDC based image segmentation. Generalization of PDC models inferred on one picture, to another, for which only the site assignments have been re-computed, were also examined. The comparison of the corresponding results clearly demonstrates, that PDC-based approaches infer the structural properties of the given data-sets while avoiding to fit the random fluctuations.

Arguably, the strict bottom-up approach to image segmentation, which solely relies on low-level features, will never achieve human performance levels on this task. On the one hand, there are findings from the field of neuro-anatomy, concerning the large amount of connections from higher brain regions back to the early stages of visual signal processing (area V1). Moreover, the large body of psycho-visual experiments (e.g. from the *Gestalt* school) also provides strong evidence for the application of higher-level knowledge in human visual scene interpretation. Consequently, neglecting these information sources about probable image configurations must be considered a significant drawback. Clearly, knowledge about the shape of the objects of interest in a segmentation or recognition task belongs to this category of useful higher-level knowledge. In the context of PDC-based segmentation, a probabilistic representation of shape has been constructed. Furthermore, likelihood maps for given objects of interest were derived from the PDC cost function. Interpreting the shape information as a prior for the segmentation task, it has been combined with the likelihoods in a Bayesian setting. The resulting posterior probability for the occurrence of an object of a specified semantic category has been demonstrated to achieve excellent segmentation quality on very hard testbeds of images from the Corel gallery.

Appendix A

Box-Plots for recall, precision, and F-measure distributions

The aim of this appendix is to give an insight into the distribution of the quality measures *recall*, *precision*, and *F-value* in the quantitative performance assessment for the various PDC variants discussed in this thesis. The empirical means and variances of these quality indicators, computed over all of the images in the Berkeley test-set, have been given in the respective sections in the form of tables. Going beyond the first two central moments, the so-called box-plot gives a very informative graphical account of a sample's distribution. Therefore, this visualization method has been applied to the recorded quality measures in order to produce the results shown in the subsequent pages. In a box-plot, the mean is denoted by the central horizontal line inside the box. The outer limits of the box indicate the lower and upper quartiles. Going out from both ends of the box, whiskers are marking the extend of the sample. Any data-point outside this range is considered an outlier and is marked by a cross. If no data-point happens to be outside the interval indicated by the whiskers, a dot is placed at the lower whisker.

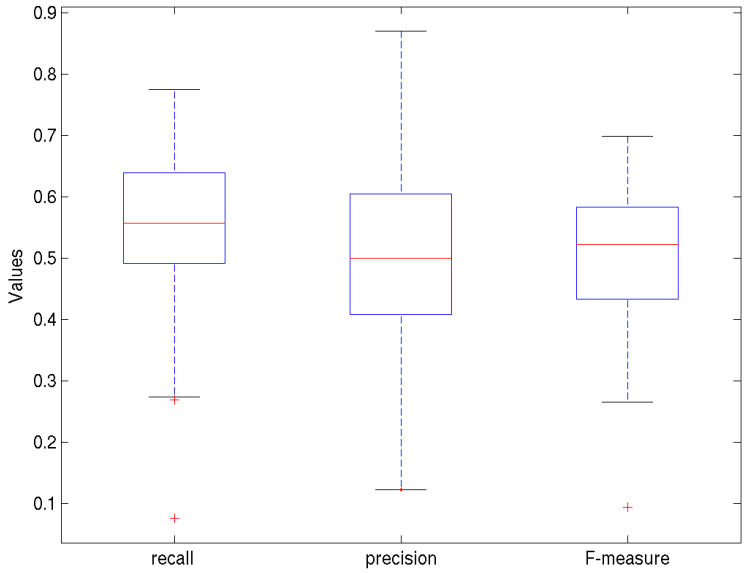


Figure A.1: Recall, precision and F-measure distributions for color-only PDC with three clusters.

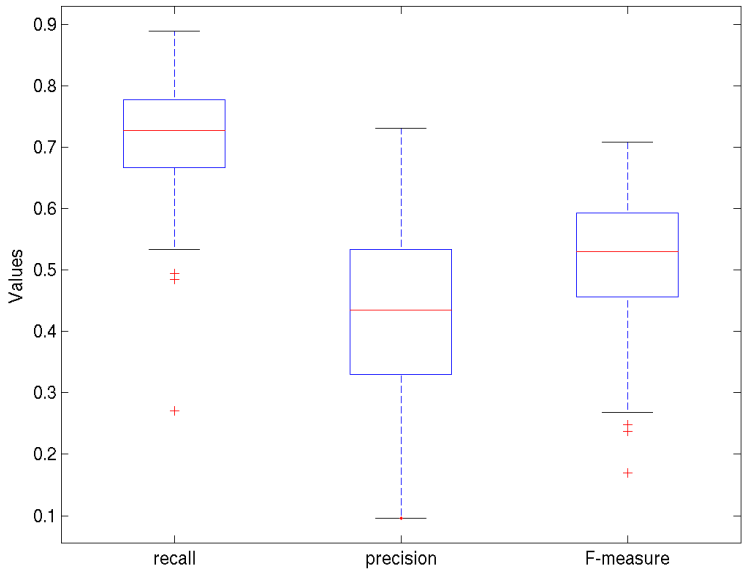


Figure A.2: Recall, precision and F-measure distributions for color-only PDC with five clusters.

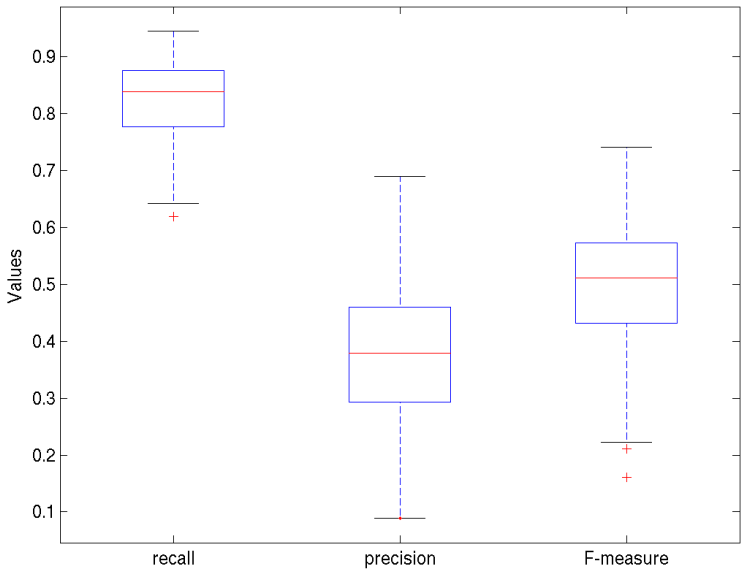


Figure A.3: Recall, precision and F-measure distributions for color-only PDC with eight clusters.

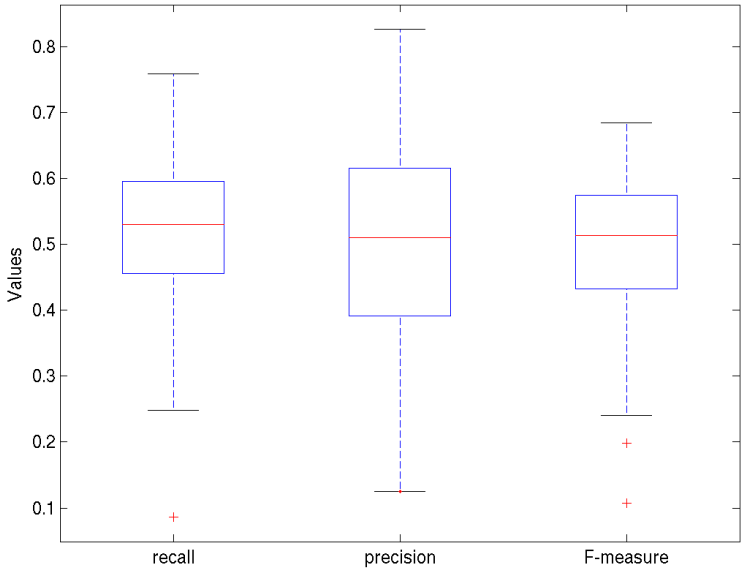


Figure A.4: Recall, precision and F-measure distributions for PDC with three clusters using color and texture features.

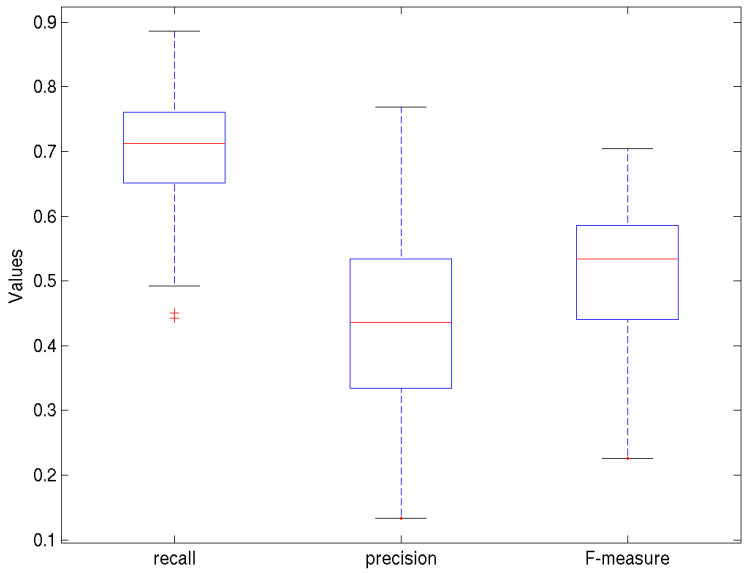


Figure A.5: Recall, precision and F-measure distributions for PDC with five clusters using color and texture features.

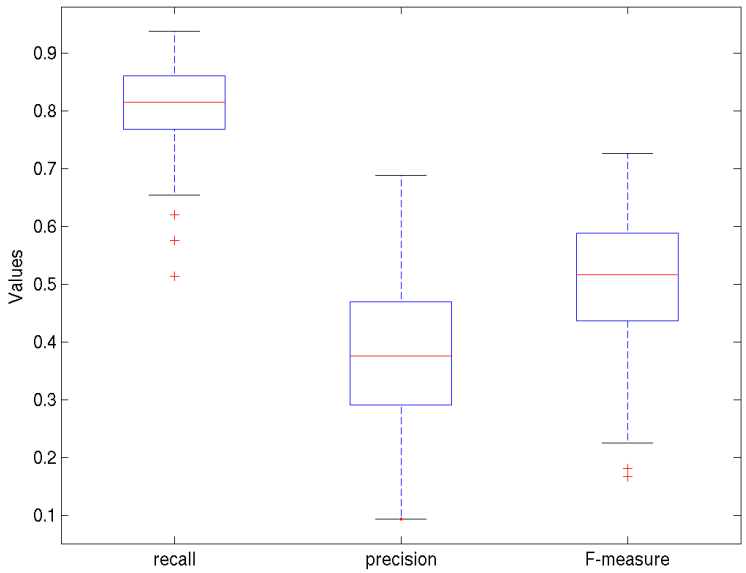


Figure A.6: Recall, precision and F-measure distributions for PDC with eight clusters using color and texture features.

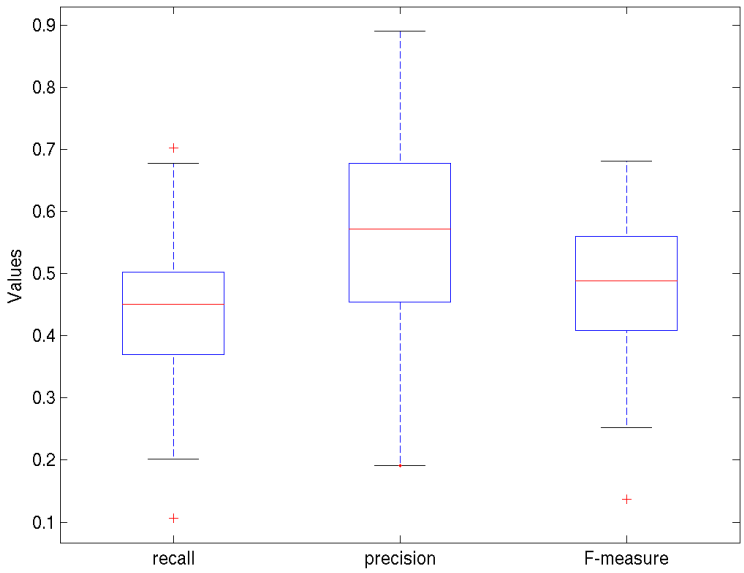


Figure A.7: Recall, precision and F-measure distributions for sPDC with three clusters using color and texture features.

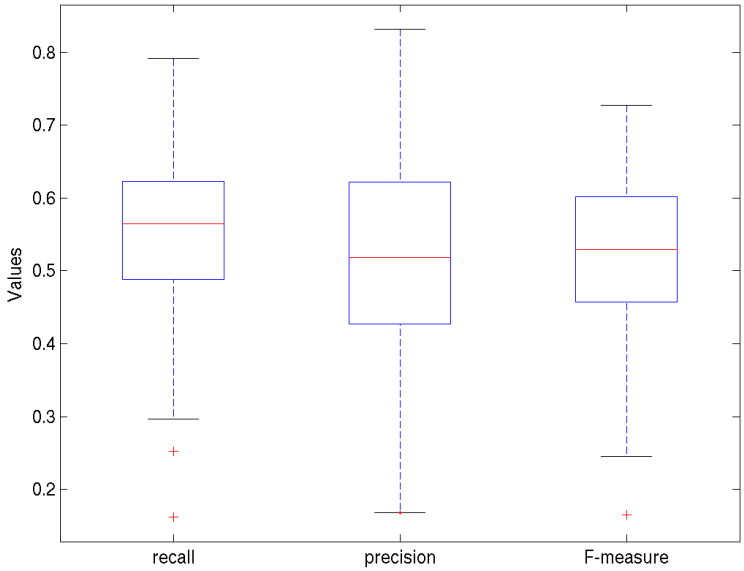


Figure A.8: Recall, precision and F-measure distributions for sPDC with five clusters using color and texture features.

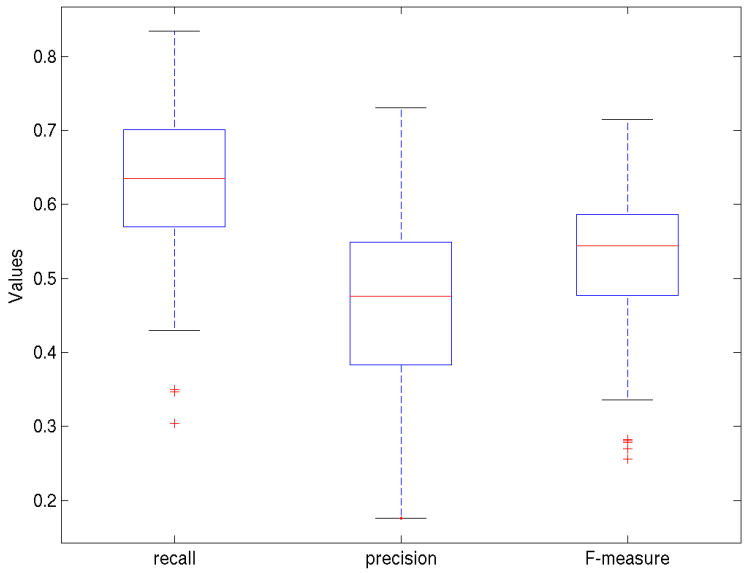


Figure A.9: Recall, precision and F-measure distributions for sPDC with eight clusters using color and texture features.

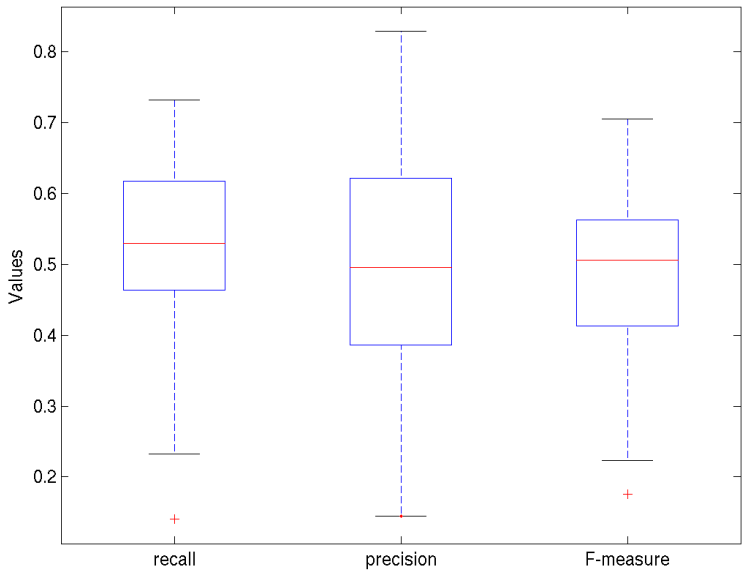


Figure A.10: Recall, precision and F-measure distributions for TPDC with three clusters using color and texture features.

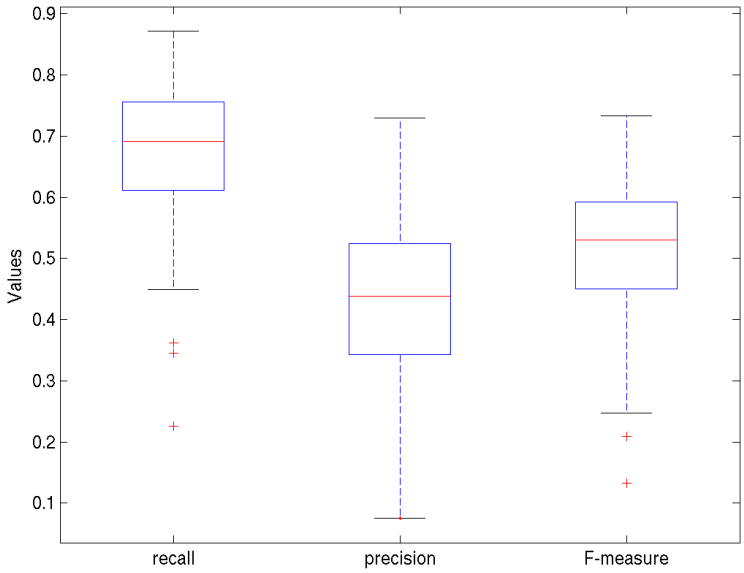


Figure A.11: Recall, precision and F-measure distributions for TPDC with five clusters using color and texture features.

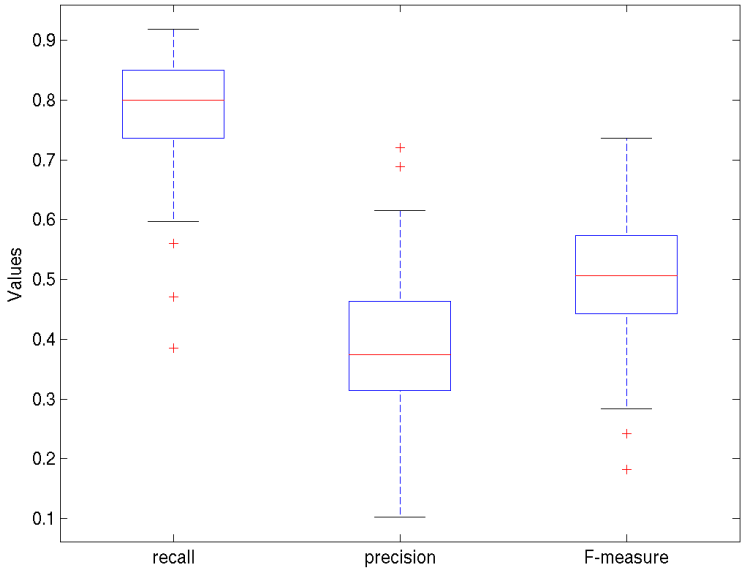


Figure A.12: Recall, precision and F-measure distributions for TPDC with eight clusters using color and texture features.

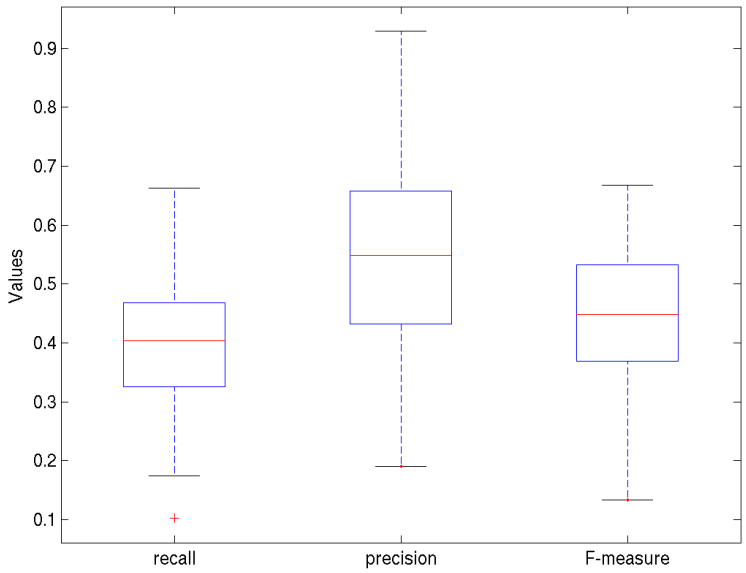


Figure A.13: Recall, precision and F-measure distributions for sTPDC with three clusters using color and texture features.

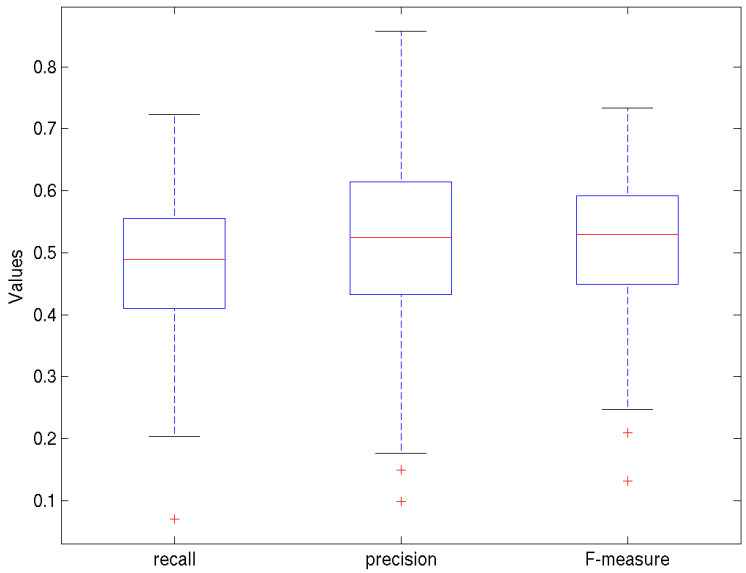


Figure A.14: Recall, precision and F-measure distributions for sTPDC with five clusters using color and texture features.

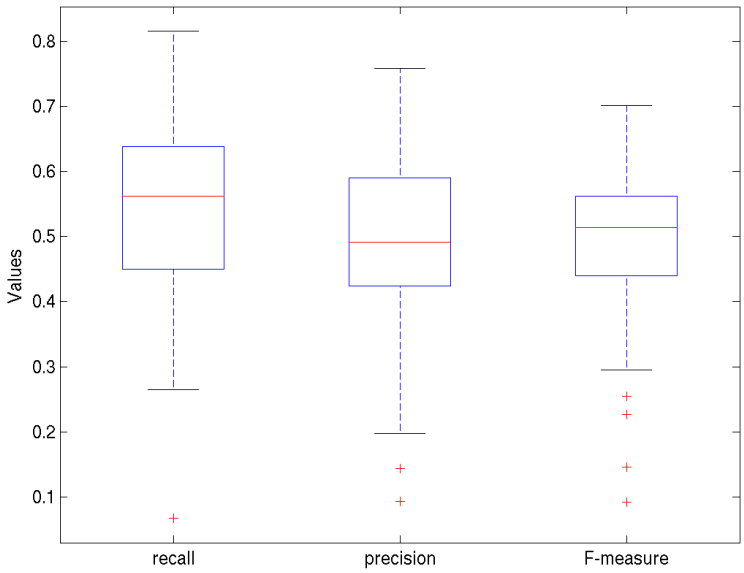


Figure A.15: Recall, precision and F-measure distributions for sTPDC with eight clusters using color and texture features.

Bibliography

[AP79] I. Abdou and W. Pratt. Quantitative design and evaluation of enhancement/thresholding edge detectors. *Proceedings of the IEEE*, 67(5):753–763, May 1979.

[BCG90] A. Bovik, M. Clark, and W. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):55–73, 1990.

[BCGM98] Serge Belongie, Chad Carson, Hayit Greenspan, and Jitendra Malik. Color- and texture-based image segmentation using EM and its application to content-based image retrieval. In *Proc. of the International Conference on Computer Vision*, Bombay, 1998.

[BG94] M. Buckland and F. Gey. The relationship between recall and precision. *Journal of the American Society for Information Science*, 45(1):12–19, 1994.

[BK93] J. M. Buhmann and H. Kühnel. Vector quantization with complexity costs. *IEEE Transactions on Information Theory*, 39(4):1133–1145, July 1993.

[BMP02] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape context. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 24(4):509–522, 2002.

[Bre96] L. Breiman. Bagging predictors. *Machine Learning*, 26(2):123–140, 1996.

[BS97] S. Borra and S. Sakar. A framework for performance characterization of intermediate–level grouping modules. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(11):1306–1312, 1997.

[Buh98] Joachim M. Buhmann. Exploratory data analysis. Tutorial at the Conference on Advances in Neural Information Processing Systems (NIPS) '98, 1998.

[CHD01] Jonathan Cheung-Wai Chan, Chengquan Huang, and Ruth DeFries. Enhanced algorithm performance for land cover classification from remotely sensed data using bagging and boosting. *IEEE Transactions on Geoscience and Remote Sensing*, 39(3):693–695, March 2001.

[CJ83] G. Cross and A. Jain. Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5:25–39, 1983.

[CK97] Vikram Chalana and Yongmin Kim. A methodology for evaluation of boundary detection algorithms on medical images. *IEEE Transactions on Medical Imaging*, 16(5):642–652, 1997.

[CKS02] D. Cremers, T. Kohlberger, and C. Schnörr. Nonlinear shape statistics in Mumford–Shah based segmentation. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Computer Vision - ECCV 2002*, volume 2 of *Lecture Notes In Computer Science (LNCS) 2352*, pages 78–92. Springer, 2002.

[Cor] Corel photo gallery, http://www.corel.com.

[CP02] Paulo Correia and Fernando Pereira. Standalone objective segmentation quality evaluation. *Journal on Applied Signal Processing*, 2002(4):389–400, 2002.

[Cre03] D. Cremers. A variational framework for image segmentation combining motion estimation and shape regularization. In *CVPR*, 2003.

[CS95] B. Chaudhuri and N. Sakar. Texture segmentation using fractal dimension. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(1):72–77, 1995.

[CT84] I. Csizàr and G. Tusnady. Information geometry and alternating minimization procedures. In E. J. Dudewicz et al, editor, *Recent Results in Estimation Theory and Related Topics*, Statisitcs and Decisions, Supplement Issue No. 1. 1984.

[Dau85] J. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America, A*, 2(7):1160–1169, 1985.

[dFCC01] L. da Fontoura Costa and R. M. Cesar. *Shape Analysis and Classification – Theory and Practice*. CRC Press, 2001.

[DH97] A. C. Davsion and D. V. Hinkley, editors. *Bootstrap Methods and their Application*. Cambridge University Press, Cambridge, UK, 1997.

[DHS01] Richard O. Duda, Peter E. Hart, and David G. Storck. *Pattern Classification*. John Wiley & Sons, 2nd edition, 2001.

[DHW94] D. Dunn, W. Higgins, and J. Wakeley. Texture segmentation using 2-d Gabor elementary functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):130–149, 1994.

[DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39:1–38, 1977.

[DMF01] Yunhan Dong, Anthony K. Milne, and Bruce C. Forster. Segmentation and classification of vegetated areas using polarimetric SAR image data. *IEEE Trans. on Geoscience and Remote Sensing*, 39(2):321–329, February 2001.

[DMZB99] Xavier Descombes, Robin D. Morris, Josiane Zerubia, and Marc Berthod. Estimation of Markov random field prior parameters using Markov chain Monte-Carlo maximum likelihood. *IEEE Transactions on Image Processing*, 8(7):954–963, 1999.

[Efr82] B. Efron. *The Jackknife, the Bootstrap and other Resampling Plans*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1982.

[ET84] B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall / CRC, 1984.

[FB02] Bernd Fischer and Joachim M. Buhmann. Data resampling for path based clustering. In Luc van Gool, editor, *Pattern Recognition - Symposium of the DAGM 2002*, volume 2449 of *LNCS*, pages 206–214. Springer, 2002.

[Fjø99] Roger Fjørtoft. Segmentation of SAR images by edge detection: Strategy, optimization and implementation. In *Proc. of the Norwegian Signal Processing Symposium (NORSIG'99)*, pages 70–75, 1999.

[Fre95] Y. Freund. Boosting a weak learning algorithm by majority. *Information and Computation*, 121(2):256–285, 1995.

[FS89] I. Fogel and D. Sagi. Gabor filters as texture discriminators. *Biological Cybernetics*, 61:103–113, 1989.

[FtHRKV92] L. M . J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever. Families of tuned scale-space kernels. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision*, Lecture Notes In Computer Science 588, pages 19–23. Springer, 1992.

[FZB01] Bernd Fischer, Thomas Zöller, and Joachim M. Buhmann. Path based pairwise data clustering with application to texture segmentation. In Mario Figueiredo, Josiane Zerubia, and Anil K. Jain, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 2134 of *LNCS*, pages 235–250. Springer, 2001.

[GG84] Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

[GK89] M. Gordon and M. Kochen. Recall-precision trade-off: A derivation. *Journal of the American Society for Information Science*, 40(3):145–151, 1989.

[Goo99] P. I. Good, editor. *Resampling Methods, A Practical Guide to Data Analysis*. Birkhäuser, Boston, MA, 1999.

[GPGSRC01] Daniel Gatica-Perez, Chuang Gu, Ming-Ting Sun, and Salvador Ruiz-Correa. Extensive partition operators, gray-level connected operators, and region merging / classification segmentation algorithms: Theoretical links. *IEEE Transactions on Image Processing*, 10(9):1332–1345, 2001.

[GSBB03] Meirav Galun, Eitan Sharon, Ronen Basri, and Achi Brandt. Texture segmentation by multiscale aggregation of filter responses and shape elements. In *ICCV*, 2003.

[Hat86] R. J. Hathaway. Another interpretation of the EM algorithm for mixture distributions. *Statistics and Probability Letters*, 4:53–56, 1986.

[Her03] L. Hermes. *Entropy-Based Segmentation and its Application to Remote Sensing*, volume 727 of *Fortschritt-Berichte VDI Reihe 10*. VDI Verlag, 2003.

[Hof97] Thomas Hofmann. *Data Clustering and Beyond: A Deterministic Annealing Framework for Exploratory Data Analysis*. Shaker Verlag, Aachen, 1997.

[HP98] Thomas Hofmann and Jan Puzicha. Mixture models for co-occurrence and histogram data. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, volume 1, pages 192–194, 1998.

[HPB98] T. Hofmann, J. Puzicha, and J. M. Buhmann. Unsupervised texture segmentation in a deterministic annealing framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):803–818, August 1998.

[HS85] R. Haralick and L. Shapiro. Survey of segmentation techniques. *Computer Vision, Graphics and Image Processing*, 29:100–132, 1985.

[HZB02] L. Hermes, T. Zöller, and J. M. Buhmann. Parametric distributional clustering for color image segmentation. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Computer Vision - ECCV 2002*, volume 3 of *Lecture Notes In Computer Science (LNCS) 2352*, pages 577–591. Springer, 2002.

[JBBAF03] S. Jehan-Besson, M. Barlaud, G. Aubert, and O. Faugeras. Shape gradients for histogram segmentation using active contours. In *ICCV*, 2003.

[JD88] A. Jain and R. Dubes. *Algorithms for Clustering Data*. Prentice Hall, Englewood Cliffs, NJ 07632, 1988.

[JDC98] A. K. Jain, R. C. Dubes, and C. C. Chen. Bootstrap techniques for error estimation. *IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI)*, 9(5):628–633, 1998.

[JF91] A. Jain and F. Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern Recognition*, 24(12):1167–1186, 1991.

[KBZ96] Zoltan Kato, Marc Berthod, and Josiane Zerubia. A hierarchical Markov random field model and multitemperature annealing for parallel image classification. *Graphical Models and Image Processing*, 58(1):18–37, 1996.

[Koh81] T. Kohonen. Self-organizing formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69, 1981.

[Koh95] T. Kohonen. *Self Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer, Berlin, Heidelberg, 1995.

[Lec89] Yvan G. Leclerc. Constructing simple stable descriptions for image partitioning. *International Journal of Computer Vision*, 3:73–102, 1989.

[Li01] S. Z. Li. *Markov Random Field Modeling in Image Analysis*. Computer Science Workbench. Springer, 2001.

[LL99] Longin Jan Latecki and Rolf Lakämper. Application of planar shape comparison to object retrieval in image databases. *Pattern Recognition (to appear)*, 1999.

[LL00] Longin Jan Latecki and Rolf Lakämper. Shape similarity measure based on correspondence of visual parts. *IEEE Transactions on PAMI*, 22(10):1185–1190, October 2000.

[Lon98] S. Loncaric. A survey of shape analysis techniques. *Pattern Recognition*, 31(8):31–41, 1998.

[MA02] F. Mokhtarian and S. Abbesi. Shape similarity retrieval under affine transforms. *Pattern Recognition*, 35:31–41, 2002.

[Mar82] David Marr. *Vision*. W. H. Freeman and Company, 1982.

[MBLS01] Jitendra Malik, Serge Belongie, Thomas Leung, and Jianbo Shi. Contour and texture analysis for image segmentation. *International Journal of Computer Vision*, 43(1):7–27, 2001.

[MBSL99] Jitendra Malik, Serge Belongie, Jianbo Shi, and Thomas Leung. Textons, contours and regions: Cue integration in image segmentation. In *Proc. of the IEEE International Conference on Computer Vision*, pages 918–925, Corfu, 1999.

[MFM04] David Martin, Charless Fowlkes, and Jitendra Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *to appear in: IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004.

[MFTM01a] 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. Technical Report UCB/CSD-1-1133, Computer Science Division (EECS), University of California, Berkeley, January 2001.

[MFTM01b] 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. In *ICCV*, Vancouver, July 2001.

[MJ96] J. Mao and A. K. Jain. A self-organizing network for hyperellipsoidal clustering. *IEEE Transactions on Neural Networks*, 7(1), 1996.

[MM96] B. Manjunath and W. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):837–842, 1996.

[MS95] Jean-Michel Morel and Sergio Solimini. *Variational Methods for Image Segmentation*. Birkhäuser, 1995.

[MSBL99] Jitendra Malik, Jianbo Shi, Serge Belongie, and Thomas Leung. Grouping in the normalized cut framework. In D.A. Forsyth, J. L. Mundy, V. Di Gesu, and R. Cipolla, editors, *Shape, Contour and Grouping in Computer Vision*, volume 1681 of *Lecture Notes in Computer Science (LNCS)*, pages 155–164. Springer, 1999.

[NH99] R. M. Neal and G. E. Hinton. A view of the EM algorithm that justifies incremental, sparse, and other variants. In M. I. Jordan, editor, *Learning in Graphical Models*. MIT Press, 1999.

[NNB+01] Unoma Ndili, Robert Nowak, Richard Baraniuk, Hyeokho Choi, and Mário Figueiredo. Coding theoretic approach to SAR image segmentation. In *Proc. of the SPIE 15th Annual International Symposium on Aerospace/Defense Sensing, Simulation, and Controls. Algorithms for Synthetic Aperture Radar Imagery VIII*, volume 4382 of *SPIE*, 2001.

[NNF01] Unoma Ndili, Robert D. Nowak, and Mário A. T Figueiredo. Coding theoretic approach to image segmentation. In *Proceedings of the ICIP'01*, 2001.

[Ols97] Ole F. Olsen. Multi-scale watershed segmentation. In J. Sporring, M. Nielsen, L. Florack, and P. Johansen, editors, *Gaussian Scale-Space Theory*, volume 8 of *Computational Imaging and Vision*. Kluwer, 1997.

[Par02] Nikos Paragios. A variational approach for the segmentation of the left ventricle in cardiac image analysis. *International Journal of Computer Vision*, 50(3):345–362, 2002.

[Pav72] Theo Pavlidis. Segmentation of picture and maps through functional approximation. *Computer Graphics and Image Processing*, 1:360–372, 1972.

[PB99] J. Puzicha and J. M. Buhmann. Multiscale annealing for unsupervised image segmentation. *Computer Vision and Image Understanding*, 76(3):213–230, 1999.

[PD99] Nikos Paragios and Rachid Deriche. Coupled geodesic active regions for image segmentation. Technical Report 3783, INRIA, Sophia Antipolis, 1999.

[PD02] Nikos Paragios and Rachid Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *International Journal of Computer Vision*, 46(3):223–247, 2002.

[PF00] Eric J. Pauwels and Greet Frederix. Image segmentation by nonparametric clustering based on the Kolmogorov-Smirnov distance. In David Vernon, editor, *Computer Vision – ECCV 2000*, volume 1843 of *LNCS*, pages 85–99. Springer, 2000.

[PHB99] J. Puzicha, T. Hofmann, and J. M. Buhmann. Histogram clustering for unsupervised segmentation and image retrieval. *Pattern Recognition Letters*, 20:899–909, 1999.

[PHB00] J. Puzicha, T. Hofmann, and J. M. Buhmann. A theory of proximity based clustering: Structure detection by optimization. *Pattern Recognition*, 2000.

[PTH96] O. Pichler, A. Teuner, and B. Hosticka. A comparison of texture feature extraction using adaptive Gabor filtering, pyramnidal and tree-structured wavelet transforms. *Pattern Recognition*, 29(5):733–742, 1996.

[PTL93] F. Pereira, N. Tishby, and L. Lee. Distributional clustering of english words. In *30th International Meeting of the Association of Computational Linguistics*, pages 183–190, Columbus, Ohio, 1993.

[Puz99] Jan Puzicha. *Multiscale Annealing for Grouping, Segmentation and Image Quantization*, volume 601 of *Fortschritt-Berichte VDI Reihe 10*. VDI Verlag, 1999.

[RBLB02] Volker Roth, Mikio L. Braun, Tilman Lange, and Joachim M. Buhmann. Stability-based model order selection in clustering with applications to gene expression data. In José R. Dorronsoro, editor, *Artificial Neural Networks – ICANN 2002*, volume 2415 of *LNCS*, pages 607–612. Springer, 2002.

[RGF90] K. Rose, E. Gurewitz, and G. Fox. A deterministic annealing approach to clustering. *Pattern Recognition Letters*, 11:589–594, 1990.

[Rij79] C. van Rijsbergen. *Information Retrieval*. Dept. of Comp. Sci. Univ. of Glasgow, 2 edition, 1979.

[RLBB02] Volker Roth, Tilman Lange, Mikio L. Braun, and Joachim M. Buhmann. A resampling approach to cluster validation. In W. Härdle and B. Rönz, editors, *Computational Statistics – COMPSTAT 2002*. Physica Verlag, Heidelberg, 2002.

[RP02] M. Rousson and N. Paragios. Shape priors for level set representations. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Computer Vision - ECCV 2002*, volume 2 of *Lecture Notes In Computer Science (LNCS) 2352*, pages 78–92. Springer, 2002.

[RS98] Calyampudi R. Rao and Shailaja Suryawanshi. Statistical analysis of shape through triangulation of landmarks: A study of sexual dimorphism in hominids. *Proceedings of the National Academy of Sciences (PNAS)*, 95(8):4121–4125, 1998.

[SBB01] Eitan Sharon, Achi Brandt, and Ronen Basri. Segmentation and boundary detection using multiscale intensity measurements. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 469–476, 2001.

[SBK+02] Anjan Sarkar, Manoj K. Biswas, B. Kartikeyan, Kumar Vikash, K. L. Majumder, and D. K. Pal. A MRF model-based segmentation approach to classification for multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 40(5):1102–1113, 2002.

[SG95] Theo E. Schouten and Maurice S. Klein Gebbinck. Quality measures for image segmentation using generated images. In Jacky Desachy, editor, *Proceedings Image and Signal Processing for Remote Sensing II*, volume 2579 of *SPIE*, pages 411–422, 1995.

[SHB98] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Understanding, and Machine Vision*. PWS, Boston, 2nd edition, 1998.

[SJK02] Cián W. Shaffrey, Ian H. Jermyn, and Nick G. Kingsbury. Psychovisual evaluation of image segmentation algorithms. In *Proc. Advanced Concepts for Intelligent Vision Systems*, 2002.

[SM00] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[SS02] Anne H. Schistad Solberg. Flexible nonlinear contextual classification. In Maria Petrou, editor, *Proc. of the 2nd Pattern Recognition for Remote Sensing Workshop*, pages 94–100. BMVA Press, 2002.

[SSWC88] P. K. Sahoo, S. Soltani, A. K . C. Wong, and Y. C. Chen. A survey of thresholding techniques. *Computer Vision, Graphics and Image Processing*, 41:233–260, 1988.

[tHR94] B. M. ter Haar Romeny, editor. *Geometry-Driven Diffusion in Computer Vision*. Kluwer Academic Publishers, 1994.

[TPB99] N. Tishby, F. Pereira, and W. Bialek. The information bottleneck method. In *Proc. of the 37th annual Allerton Conference on Communication, Control, and Computing*, 1999.

[TS01] N. Tishby and N. Slonim. Data clustering by Markovian relaxation and the information bottleneck method. In *Advances in Neural Information Processing Sytems*, volume 13. NIPS, 2001. to appear.

[TZ02] Zhouwen Tu and Song-Chun Zhu. Image segmentation by data-driven Markov chain Monte Carlo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):657–673, 2002.

[Vap95] Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1995.

[Vap98] Vladimir N. Vapnik. *Statistical Learning Theory*. John Wiley & Sons, 1998.

[VNBF00] Vidya Venkatachalam, Robert D. Nowak, Richard G. Baraniuk, and Mário A. T. Figueiredo. Unsupervised SAR image segmentation using recursive partitioning. In *Proc. of the SPIE 14th Annual International Symposium on Aerospace / Defense Sensing, Simulations, and Controls. Algorithms for Synthetic Aperture Radar Imagery VII*, volume 4053 of *SPIE*, 2000.

[Web99] Andrew Webb. *Statistical Pattern Recognition*. Arnold, 1999.

[Wei98] Joachim Weickert. Fast segmentation methods based on partial differential equations and the watershed transformation. In P. Levi, R.-J. Ahlers, F. May, and M. Schanz, editors, *Mustererkennung 1998*, Informatik aktuell, pages 93–100. Springer, 1998.

[Win02] G. Winkler. *Image Analysis, Random Fields and Markov Chain Monte-Carlo Methods*. Springer, 2nd edition, 2002.

[YA98] H. Yin and N. M. Allison. On the distribution and convergence of feature space in self-organizing maps. *Neural Computation*, 7(6):1178–1187, 1998.

[YMB77] William A. Yasnoff, Jack K. Mui, and James W. Bacus. Error measures for scene segmentation. *Pattern Recognition*, 9:217–231, 1977.

[ZB02] T. Zöller and J. M. Buhmann. Self-organized clustering of mixture models for combined color and texture segmentation. In *Proceedings of Texture 2002, The second international workshop on texture analysis and segmentation*, pages 163–167, 2002.

[Zha96] Yu J. Zhang. A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29(8):1335–1346, 1996.

[ZHB02] Thomas Zöller, Lothar Hermes, and Joachim M. Buhmann. Combined color and texture segmentation by parametric distributional clustering. In *Proc. of the International Conference on Pattern Recognition (ICPR)*, volume 2, pages 627–630, 2002.

[Zhu04] Mu Zhu. Recall, precision, and average precision. Technical report, Dept. of Statistics and Actuarial Science, University of Waterloo, CA, 2004.

[ZY96] S.-C. Zhu and A. L. Yuille. Region competition: Unifying snakes, region growing and Bayes / MDL for multi-band image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9), 1996.