Eberhard Karls
Universität
Tübingen

Diplomarbeit

Biologically Motivated Image Preprocessing for Appearance–based Landmark Selection

Lorenz Gerstmayr

29. April 2005

Betreut von

Prof. Dr. Hanspeter A. Mallot
Lehrstuhl Kognitive Neurowissenschaft,
Fakultät für Biologie

Prof. Dr. Andreas Schilling
Graphisch Interaktive Systeme,
Fakultät für Kognitions– und Informationswissenschaften

Ich erkläre, dass ich die vorliegende Diplomarbeit selbstständig und nur mit den angegebenen Hilfsmitteln angefertigt habe und dass alle Stellen, die dem Wortlaut oder dem Sinn nach aus Arbeiten anderer Personen übernommen wurden, durch Quellenangaben als solche gekennzeichnet sind.

Tübingen, den 29. April 2005,

Zusammenfassung

Navigation ist eine besonders faszinierende Verhaltensweise, weil sie die Grundlage für viele weitere Verhaltensleistungen ist, die für Tiere überlebenswichtig sind. Da Tiere so sehr auf zuverlässige Navigation angewiesen sind, haben sich über die Zeit Navigationsstrategien und Sinnessysteme herausgebildet. Diese ermöglichen es die zur Navigation benötigen Informationen aus der Umwelt zu extrahieren, und sind optimal an die Umwelt der Tiere angepasst. Was bei Tieren so einfach aussieht, ist in der Robotik noch immer ein ungelöstes Problem, auch wenn in den letzten Jahren deutliche Fortschritte erzielt werden konnten.

Diese Arbeit ist Teil eines Projektes, das es sich zum Ziel gesetzt hat, die nötigen Steuerungs– und Navigations–Mechanismen für einen autonom fliegenden Zeppelin, der über bebauten Gebieten eingesetzt werden soll, zu entwickeln. Ein erster Teilschritt ist der Aufbau einer topologischen Karte, d.h. eines Netzwerkes aus bekannten Orten und deren Verbindungen. Dazu werden verlässliche und gut wiederzufindende Landmarken benötigt. Zur Vereinfachung der Problemstellung werden im Rahmen dieser Arbeit Luftbilder des Einsatzgebietes verwendet. In der vorliegenden Diplomarbeit wird ein Verfahren vorgestellt, wie in einem Luftbild Punkte, die als Landmarken dienen könnten, detektiert werden. Sie erweitert schon bestehende Algorithmen, die aus einer Menge an möglichen Landmarken die aussichtsreichsten Kandidaten extrahieren, zu einem vollständigen Landmarken–Selektions–System. Dieses bildet somit die Grundlage für den Aufbau der topologischen Karte und für alle weiteren Navigationsleistungen bildet.

Ein zentrales Problem dieser Arbeit besteht darin, die in den Luftbildern enthaltene Information zu reduzieren, ohne dabei zur Navigation relevante Informationen zu verwerfen. Da Linienzüge ein elementarer Bestandteil von Luftbildern sind, wurde eine Kantenrepräsentation als Ausgangspunkt für die Detektion möglicher Landmarken gewählt, die durch das Zusammentreffen unterschiedlich orientierter Konturen definiert sind.

Sowohl die Kantendetektion als auch die Detektion von Kreuzungspunkten erfolgt nach einem Modell für die visuelle Informationsverarbeitung beim Menschen. Besonderes Augenmerk wurde darauf gelegt, dass vor allem das Modell zur Kantendetektion an die statistischen Eigenschaften der zu verarbeitenden Bilder angepasst ist. Dazu wurden zunächst die statistischen Eigenschaften von Luftbildern mit denen von natürlichen Szenen (Landschaften, Pflanzen, Tiere) und von Menschen gemachten Objekten (Städte, Gebäude, Innenansichten) verglichen. Dadurch konnte gezeigt werden, dass Luftbilder durch lange Kantenzüge ohne Vorzugsorientierung charakterisiert sind. Natürliche Szenen zeigen hingegen kurze Kantenzüge jeglicher Orientierung und Szenen, die von Menschen gemachte Objekte zeigen, werden von kurzen Kantenzügen, die vor allem horizon-

tal und vertikal orientiert sind, dominiert. Ausgehend von diesen Eigenschaften konnten rezeptive Felder abgeleitet werden, die optimal an die Eigenschaften der jeweiligen Bildklasse angepasst sind. Verwendet man diese rezeptiven Felder, um Bilder einer anderen Bildklasse damit zu verarbeiten, werden zwar die grundlegenden Strukturen erkannt. Die Resultate, bei denen die Kantenrepräsentation die vorhandenen Kantenstrukturen am besten beschreibt, werden aber mit den angepassten rezeptiven Feldern erzielt.

Die durch die Detektion von Kreuzungspunkten selektierten Landmarken–Kandidaten werden durch die Anwendung der Algorithmen zur Landmarkenselektion weiter reduziert. Im Rahmen dieser Arbeit konnte gezeigt werden, dass die Algorithmen auch mit dem hier beschriebenen Verfahren zur Selektion von Landmarken–Kandidaten unterschiedliche Landmarken selektieren, die oft durch ein charakteristisches Muster aus Straßenzügen oder Gebäuden gekennzeichnet sind.

Insgesamt sind die Ergebnisse dieser Arbeit erfolgsversprechend, auch wenn einzelne Teilaspekte noch verbessert werden können. Als Ziele für weiterführende Teilschritte werden die Implementierung eines Modells zur Verstärkung zusammenhängender Kantenzüge, eine Analyse der Verlässlichkeit der selektierten Landmarken und Experimente mit einem Roboter oder einem detaillierten Simulator vorgeschlagen.

Abstract

Navigation is a fascinating behaviour because it builds the basis for many other, often very complex behaviours, which are essential for the survival of an animal. Since animals have to rely on navigation capabilities, navigation mechanisms as well as sensory systems have evolved to be optimally tuned to an animal's environment. They allow the animals to extract the relevant information from their environment. Even though navigation behaviour might look very simple, it is still an unsolved problem from the viewpoint of robotics.

This diploma thesis is part of a project which aims at developing the necessary control– and navigation–strategies for an autonomous floating vehicle that will be used over urban areas. A first step includes the selection of robust landmarks, which can then be used to build topological maps, i.e. a network of known places and their connections. For sake of simplicity, aerial images of the operational area are used for this work. Here, a method is proposed to detect keypoints in an aerial image that can serve as possible landmarks. Already existing work, which deals with the selection of distinctive landmarks from a set of landmark candidates, is extended to a complete landmark selection system.

One challenge of this work is to reduce the information contained in the aerial images without discarding information relevant for navigation taks. As contours can be identified as an essential building block of aerial images, for this work an edge representation was chosen. This also facilitates the detection of landmark candidates which then can be defined as points where several contours of different orientation intersect.

For the edge detection as well as for the junction detection, models for the visual information processing in humans are used. In particular, attention was payed to adapt the models to the statistical properties of the input images. Therefore, the statistical properties of urban aerial images were compared to those of natural scenes (landscapes, flowers, animals) and of manmade scenes (buildings, cities, indoor scenes). The comparison revealed that aerial images are characterized by long contours in various orientations. Natural and manmade scenes are characterized by short contours in various orientations and short contours in mainly horizontal and vertical directions, respectively. Based on these properties, class–specific receptive fields were derived which are optimally tuned to the statistical properties of the corresponding image class. If used to process images of a different image class, the main structures are detected. Nevertheless the edge representation describing the input image best is obtained by the class–specific receptive fields.

The final set of landmark candidates is obtained by reducing the number of junction points by assuring a certain inter–landmark distance. The most distinctive landmarks are then selected using existing landmark selection algorithms. The results of this work show that the algorithms, if used in conjunction with the proposed preprocessing stages,

select distinctive landmarks, which often contain a characteristic pattern formed by streets or buildings.

Although single processing stages can still be improved, the overall results of this work promise that the proposed processing stages can be used to select good landmarks for topological mapping and all navigation strategies beyond. As future work the implementation of an intermediate processing stage for contour grouping and enhancement, an empirical reliability analysis of the selected landmarks, and experiments with a real robot or at least a detailed simulator are proposed.

Acknowledgements

The Encyclopedia Galactica defines a robot as a mechanical aparatus designed to do the work of a man. The marketing division of Sirius Cybernetics defines a robot as "Your Plastic Pal Who's Fun To Be With".

Douglas Adams
The Hitchhiker's Guide to the Galaxy

Since this work is rather theoretical there is no robot to be grateful to. However, there are many people who made this work possible and a very pleasant and enriching one.

Let me start with Prof. Hanspeter Mallot, who was a valuable advisor, always had an open ear for questions, made the participtation in several meetings possible, and let me chose the topic of this diploma thesis freely. Thanks a lot also to Prof. Andreas Schilling, who agreed to supervise this thesis, and always showed a lot of interest.

I would also like to mention all the members of the department of Cognitive Neuroscience: thanks a lot for the pleasant and amicable atmosphere during work and beyond, for discussions, and for giving me the feeling of being part of the team. Thanks a lot to Olga Rodríguez–Sierra, who patiently stood my first trials in teaching somebody programming, and to Wolfgang Stürzl, whose phone calls from Australia always were a welcome break during the afternoons and a good and valuable chance for discussion.

I thank Prof. José Santos–Victor and Prof. Alexandre Bernardino from the ISR–Vislab at Lisbon, Portugal. During a great visit at their lab I could not only gather the experience of living and working in a foreign country but I also got in touch with the topic and could do the work on which this thesis builds. Their advice and ideas were always very helpful. Thanks a lot the "Vislabeers International Family" for – among many other things – providing me with good news about next generation robot vision researchers and invitations to PhD–parties around the world.

Further I would like to mention Thorsten Hansen, who is now with the Department of Psychology at the Giessen university and who developed in his PhD–thesis at the university of Ulm the models I used in this work. Thanks a lot for valuable advice and interest in my work.

I thank all my friends and fellow students for their friendship, discussions, or common activities which made my spare time a valuable source of recreation. I'm also deeply grateful towards my parents, who made my studies possible and always had confidence in me. Finally, I would like to thank my girl–friend Nina – very short but the more cordially – for everything and in particular for a great job in last–minute–proof–reading.

Contents

List of Figures

List of Tables

Abbreviations

BSS	Blind Source Separation
cdf	Cumulative Distribution Function
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transformation
DOI	Dominating Opponent Inhibition
DoG	Differential of Gaussians
EEG	Electroencephalogram
EPS	Ensemble Power Spectrum
ES	Evolution Strategies
EVD	Eigenvalue Decomposition
FFT	Fast Fourier Transformation
HMB	Hoffmeister–Bäck adaptation scheme (ES)
HWHH	Half Width at Half Height
IC	Independent Component
ICA	Independent Component Analysis
IPCA	Incremental Principal Component Analysis
KST	Kolmogorov–Smirnov–Test
LGN	Lateral Geniculate Nucleus
MEG	Magnetoencephalogram
MSA	Multiple Stepsize Adaptation (ES)
PCA	Principal Component Analysis
pdf	Probability density function
PoI	Point–of–Interest
ROI	Receiver Operator Characteristic
SSD	Sum of Squared Differences
V1	Primary Visual Cortex
V2	Secondary Visual Cortex

Mathematical Notation

General Identifiers, Special Functions

non–boldface letter	Scalar value or function
caligraphical letter	Sets or special probability density functions
x, y, z	Coordinates
t	Time
θ	Orientation
$\mathbf{I}(x, y)$	An image. Please watch out that image sizes are given in matrix notation
$\mathbf{I}(f_x, f_y)$	An image in the frequency domain
$a(.)$	A function
$A(f)$	The Fourier transformed of a
$S(f)$	The power spectrum of a
$h(.)$	Windowing function for Fourier Transformation
$\bar{c}$	Complex conjugate of a complex number c
$*$	Convolution operator
$\mathrm{abs}\,(.)$	Absolute value
$\exp(.)$	Exponential function
$\max(.)$	Maximum function
$\min(.)$	Maximum function
$[.]^+$	Half wave rectification, positive part
$[.]^-$	Half wave rectification, negative part
$\|.\|$	Modulos of a complex number

Linear Algebra

Uppercase bold letters	Matrices
Lowercase bold letters	Vectors
$\mathrm{diag}\,(.)$	A diagonal matrix
$\mathbf{1}$	The unit matrix
$\lambda_1, \ldots, \lambda_n$	Eigenvalues
$\mathbf{\Lambda}$	Diagonal matrix of eigenvalues $\lambda = \mathrm{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_n)$
$\mathbf{e}_1, \ldots, \mathbf{e}_n$	Eigenvectors

Statistics and Stochastic

$\mathbf{X}$	Random variable
$\mathbf{x}$	Realization of $\mathbf{X}$
$p(.)$	Probability density function (pdf)
$P(.)$	Probability of a certain event
$\mathcal{X}$	Uniform distribution on $[0, 1]$
$\mathcal{N}_0(\sigma)$	Gaussian distribution with zero mean and standard deviation σ
$\bar{\mathbf{x}}$	Mean of $\mathbf{x}$
$\mathrm{std}(.)$	Standard deviation
$E(.)$	Expected value
$\mathrm{cov}(.)$	Covariance matrix
$\mathrm{circvar}(.)$	Circular variance
$\mathrm{osgnf}(.)$	Orientation significance function

PCA and ICA

$\mathbf{x}$	A single observation
$\mathbf{X}$	$(m \times n)$ Matrix of n observations of m dimensions
$\tilde{\mathbf{X}}$	A zero–mean observation matrix
$\hat{\mathbf{X}}$	A whitened observation matrix
$\mathbf{s}$	Source signals
$\mathbf{T}$	Transformation matrix for PCA
$\mathbf{g}$	An observation in the eigenspace
$\mathbf{r}$	Residue vector
τ	Proportional variance
$\mathbf{A}$	ICA mixing matrix
$\mathbf{W}$	ICA separating matrix

Evolution Strategies

$\mathcal{G}(t)$	A population at time t
$\mathcal{I}_i(t)$	The individual i at time t
μ	The population size, i.e. the number of individuals
λ	Number of offspring individuals per parent individual
ρ	Number of recombinants
$\mathbf{p}$	The chromosome of an individual
$\mathbf{s}$	The strategy parameters of an individual
ω_{mut}	Strategy mutation operator
ξ_i	Scaling factor for mutating strategy parameter i
β	Strategy mutation parameter, regulates the stepsize adjustment

ω_{rec}	Recombination operator
ω_{sel}	Selection operator

Early Vision Models

$\mathbf{G}_\sigma$	Gaussian smoothing filter with standard deviation σ
$\mathbf{G}, \mathbf{G}^\theta$	Gabor filter kernel (with orientation θ)
K	Amplitude scaling factor for Gabor functions
s_1	Variance of the Gabor function along the tuning direction
s_2	Variance of the Gabor function perpendicular to the tuning direction
m_x, m_y	Center position of the Gabor function
θ	Tuning direction of the Gabor function
F	Spatial frequency parameter
φ	Phase shift
$\sigma_c, \sigma_s, \sigma_{\text{J}}$	Standard deviations for Gaussian smoothing filters
α_{LGN}	Activity decay rate for the LGN stage
β_{LGN}	Upper bound of activity in the LGN stage
γ_{LGN}	Lower bound of activity in the LGN stage
$\alpha_{\text{S}}, \beta_{\text{S}}, \gamma_{\text{S}}$	Shunting parameters for the simple cell stage
ξ	Strength of DOI
$\text{net}^+, \text{net}^-$	Excitatory and inhibitory activation
$\mathbf{I}_c, \mathbf{I}_s$	Center and surround streams
$\mathbf{X}_{\text{on}}, \mathbf{X}_{\text{off}}$	Contrast sensitive signals for on– and off–domain
$\mathbf{K}_{\text{on}}, \mathbf{K}_{\text{off}}$	LGN–response
$\mathbf{R}_{\text{on}}, \mathbf{R}_{\text{off}}$	On– and off–subfields of simple cells
$\mathbf{S}_{\text{ld}}, \mathbf{S}_{\text{dl}}$	Light–dark and dark–light sensitive simple cell
$\mathbf{C}_\theta$	Complex cell with preferred orientation θ
$\mathbf{C}_P$	Pooled complex cell response
$\mathbf{J}$	Junctioness map
α_{J}	Detection threshold as fraction of maximal response
δ_{J}	Neighborhood size for junction detection

Landmark Selection

n_C	Number of landmark candidates to detect
n_L	Number of landmarks to select
α_L	Distance to image boarder
β_L	Decay parameter for detection strength reduction
γ_L	Minimal distance between two landmarks
δ_L	Size of the sub–image representing a landmark
τ_L	Proportional variance to keep (profile–based algorithm)

1. Introduction

Navigation is a fascinating behavior because it builds the basis for many other, often very complex behaviors that are essential for the survival of animals or humans. Migration, territoriality, foraging, or mating are only view examples for such behaviors. Through the need of reliable navigation capabilities, an astonishing variety of navigation strategies has evolved in the animal kingdom and many of them are still poorly understood. The same holds true for the variety of sensory systems that are used to extract relevant information about the environment and the moving animal itself which can be used for navigation.

Probably the most well known examples for such navigation capabilities are migration in birds and homing in pigeons. During fall, many birds migrate from Northern or Central Europe to Southern Europe or Africa. Many travel for thousands of kilometers, often without any or almost any break, and many species cross the open see. According to Mouritsen (2001) the Arctic Tern, *Sterna paradisea*, breeds close to the North Pole and spends the winter close to Antarctica, therefore migrating for approximately 18000 km twice a year. To master this journeys birds can rely on sun compasses, star compasses, and magnetic compasses (Mouritsen, 2001; Wiltschko and Wiltschko, 2003). Although the magnetic compass in birds has been known for approximately 40 years, there is still an ongoing debate how the birds can sense the earth's magnetic field (Mouritsen et al., 2004; Mora et al., 2004). Also the relationships between the different compasses and how they are recalibrated is a field of ongoing research (Cochran et al., 2004). Following Wiltschko and Wiltschko (2003) there is also a change in the used navigation strategies with increasing experience. While young birds have to rely on innate navigation programs, experienced birds rather rely on things they have learned in the past including various landmarks.

The navigation capabilities of flying insects are even more amazing as they have smaller brains than birds. Many insights about flying insects have been gained on the Honeybee, *Apis mellifera*. A scout bee, which returned to the hive after finding an attractive food patch, communicates its nest mates the direction and distance to the food patch using the famous waggle dance. On their outward trip to the food patch the recruited bees can estimate the traveled distance from optic flow, i.e. by the extent to which the image of the environment they perceive changes (Si et al., 2003). In order to find back to their hive, especially young bees do learning flights, usually referred to as Turn Back and Look Behavior. On the outward flight the bees turn around, face the direction of the nest, and then proceed in arcing and circling flight patterns of increasing radius (Wei et al., 2002). Additionally, bees can rely on a polarization compass giving a reference direction (Labhart and Meyer, 2002) and on visual landmarks (Collett and Collett, 2002). For landmark navigation, the "snapshot" model was proposed according to which bees

store a retinotropic snapshot of the surrounding landmarks at the goal position. While approaching the goal the current retinal view is compared to the stored snapshot and the direction towards the goal is derived.

Compared to these amazing capabilities the navigation capabilities of robots still seem to be very poor, inaccurate and unreliable, although the computational power of a today's of–the–shelf computer is comparable to that of an insect brain (Webb, 2002) and today's navigation algorithms are much more flexible and robust than early work in this area. Therefore, animal behavior is often used as an inspiration for technical applications (see Webb (2000, 2001) for reviews). Thus, robotic researchers as well as neurobiologists, ethologists and psychologists work on related questions concerning how reliable navigation behaviors can be achieved, which sensor systems are used, how different sensor readings are integrated, and how the necessary knowledge about the environment can be represented.

The representation of the environment most favored in the literature is the so called cognitive map. It is a mental representation of features, objects, and locations also including their spatial relations (Golledge, 1999). The process of building such a cognitive map –or more generally of building a representation of an environment– is referred to as mapping.

This work deals with an important step in the mapping process: the selection of landmarks or rather the preselection of possible landmarks. Here, landmarks are unique image patches.

1.1. Outline of this Project

Before describing the aims of this diploma project in detail an overview over the project which it is a part of will be given and related aspects of mobile robot navigation will be outlined.

1.1.1. The RESCUE–Project

This diploma project can be seen as part of the RESCUE–project of the *Instituto de Sistemas e Robótica* located at the *Instituto Superior Técnico* in Lisbon, Portugal. The project dealt with cooperative navigation for rescue robots and aimed to develop robot systems that can assist humans in search and rescue missions in disaster areas like areas destructed by earthquakes, floods or terrorist attacks. For a summary see Lima et al. (2003), or Bernardino et al. (2003a,b, 2004), as well as the project homepage.[1]

One subproject includes the use of an autonomous blimp flying over the scenario. According to Lima et al. (2003), the advantages of aerial robots include that they can provide a wide view from a bird's position and are therefore able to gather information about areas which might not be reachable for human rescue troops. Based on this information these regions can be mapped, operators can better guide human or machine rescue troops, or the mission planning can be done fully autonomously.

[1] See `http://rescue.isr.ist.utl.pt`

This tasks can only be solved if the blimp is equipped with powerful and robust navigation and localization capabilities. Since dealing with a complete catastrophic scenario, which is highly dynamic, is still too difficult to handle due to moving rescue troops and the excavation of debris, navigation over urban areas was selected as first subgoal. However there are still many problems to master mainly related to changing weather conditions, wind causing a drifting of the blimp, and daylight changes. Therefore a practical setup was developed at the *Laboratoria de Visão* (VisLab):[2] an indoor blimp (approximately 0.8 m in diameter and 2 m in length) is flying over a huge poster of an aerial image. The setup is shown in figure 1.1. The blimp is equipped with a camera looking downwards to the "city". The camera image as well as the movement commands are exchanged via a radio link with a ground computer on which all the necessary computations are carried out.

Figure 1.1.: The experimental setup with the blimp

The camera is also used to control the blimp. Elementary steps for vision–based control include to explore the environment, to fuse the sensor readings to a consistent map of the environment, to recognize and extract landmarks from this map as well as their configurations, and to allow navigation between these selected reference points which includes recognizing them when they are approached from different directions and under different environmental conditions.

Since this work concerns more fundamental aspects of selecting good landmarks, it was done without robot experiments, neither simulated nor real ones. However, running robot experiments, e.g. by using the simulator developed in Metelo and Garcia (2003), to proof the developed algorithms is an important and necessary goal for future work.

[2]See `http://www.isr.ist.utl.pt/vislab`

1.1.2. Mapping and Localization

As sketched above, mapping and localization are fundamental building blocks for navigation strategies. According to standard robotics textbooks like Dudek and Jenkin (2000) or Siegwart and Nourbaksh (2004), the following standard approaches to mapping and localization can be distinguished. For a more detailed review on localization work see Gerstmayr et al. (2004b).

Geometric localization: Geometric methods all use a 2D or 3D model of the environment as map representation. The robot's position is then determined with respect to the map's coordinate system by matching the sensory input with the map. Geometric approaches are very exact, but also very error–prone and need a huge amount of storage. Some examples for geometrical localization include Dudek and Jugessur (2000), Kelly (2000), Artac et al. (2002a), or Shaw and Barnes (2002).

Topological localization: Topological navigation strategies use an adjacency graph as representation of the environment. The only information stored in the map are the different places and their topological relations. Usually metric information is discarded. Only the node closest to the robot's current position is determined. Therefore, the localization is less precise than for geometric navigation, less error–prone, and for many applications more flexible. Examples for topological localization include Franz et al. (1998), de Verdiere and Crowley (1998), Gaspar et al. (2000), Ulrich and Nourbakhsh (2000), Freitas et al. (2003) and Hübner (2005).

Hybrid localization: Hybrid methods try to combine the accuracy of geometric approaches with the flexibility of topological navigation. An example for a hybrid method is Bailey and Nebot (2001).

For the blimp, a topological approach as sketched in figure 1.2 was chosen. The topological map will be determined from a mosaic–based representation of the environment. Since the blimp is supposed to fly at such altitudes that the image depth can be neglected, mosaic–based maps can describe the environment appropriately. Such a map can be obtained from consecutive frames of a video sequence recorded when the blimp flies over the environment in an exploration phase. Therefore, the frame–to–frame correspondences have to be estimated from which the movements of the blimp can be estimated. For more details the reader is referred to Gracias and Santos-Victor (2001), Santos-Victor et al. (2001) and Gracias et al. (2003). The methods proposed for submarines can be applied to the blimp easily as submarines and blimps have similar kinematics. Since the work reported here is independent of robotic experiments a set of aerial images taken from an airplane is used instead of a mosaic–based map.

The main difficulty for extracting the topological map is to determine places that allow robust navigation. Such places are often called landmarks, the process of selecting appropriate places is referred to as landmark selection.

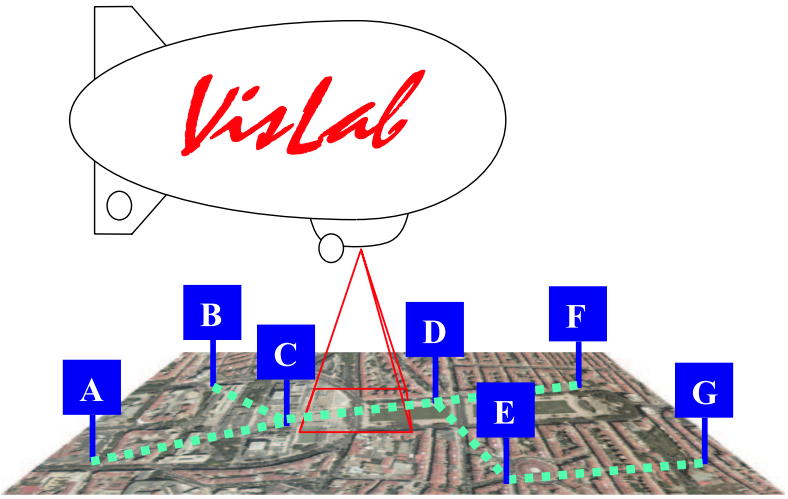


Figure 1.2.: Schematic drawing for topological navigation of an autonomous blimp

1.1.3. Landmark Selection

A landmark is a pattern in the sensor reading that can be used as a reference for navigation. This definition of a landmark was given by Thrun (1998) and is the most abstract one that can be found in the literature. More concrete definitions include the need for a concise geometric parameterization (Leonard and Durrant-Whyte, 1991), an underlying map (Deng et al., 1996) representing the exact positions of other landmarks (Mata et al., 2001), or the physical identifiability of a landmark (Mata et al., 2001). Although those definitions of the concept of a landmark are correct for the specific application, they often fail to cover all aspects of landmarks. Here, like in Knapek et al. (2000), Ohba and Ikeuchi (1997), or many other related works on vision–based navigation, a landmark is a subimage of a known image that is as dissimilar to all other considered subimages as possible. However, the properties of good landmarks are usually the same independent of the used definitions. The properties were summarized in Ohba and Ikeuchi (1997) and include detectability, uniqueness and reliability.

Detectability: The landmark has to be detectable within the sensory input.

Uniqueness: The detectability criterion does not guarantee the global uniqueness of a landmark. Thus, this property ensures that the structures to be used as landmark are discriminative.

Reliability: Discriminative landmarks can still be useless for navigation tasks if they are not stable over time, e.g. if they disappear or change their location. Additionally, a good landmark should enhance the robustness and reliability of the navigation behavior.

This definition is very technically oriented. Definitions from psychology also include these properties, but additionally take the relevance of a landmark into account Golledge

(1999); Janzen and van Turennout (2004). The relevance is an important aspect for the task an animal or a person has to solve. For example it could be shown that landmarks along a path are not as well remembered as landmarks at decision points or intersections. However, as the properties as proposed by Ohba and Ikeuchi (1997) are sufficient for the special task of mapping, this work will neglect the relevance of a landmark.

The main aspect, which all approaches to landmark selection have in common, is to reduce navigation uncertainty by extracting points that make navigation more robust. The advantages of selecting landmarks are better localization and a speed up of the computations because less landmarks are used. Some approaches like Thrun (1998) or Olson (2002) therefore use probabilistic approaches. Another possibility used for example by Sutherland and Thompson (1994) or by Burschka et al. (2003) is to use landmarks in areas where it is known a priori that the landmarks lead to small localization errors.

Most of the approaches use two steps: in the first step socalled landmark candidates are preselected using point–of–interest detectors such as corner detectors (Schmid and Mohr, 1997; Ohba and Ikeuchi, 1997; Little et al., 1998; Jugessur and Dudek, 2000; Knapek et al., 2000) or edge density (Bourque et al., 1998). In a second step, the pre–selected points are tested for reliability or uniqueness, and the candidates not satisfying that criteria are reject. Then, for vision–based localization, subimages around the selected points are used. In the second step related work like Ohba and Ikeuchi (1997), Schmid and Mohr (1997), or Knapek et al. (2000) compare image similarities and select landmarks that are as dissimilar as possible. In Ohba and Ikeuchi (1997) the number of landmarks is further reduced by discarding landmarks that are not stable to small changes of the viewpoint. Little et al. (1998) use stereo information to detect and discard corners that resulted from overlapping objects, keeping corners on planar surfaces. Jugessur and Dudek (2000) compute the standard deviation of the pixel values in the subwindows and only keep landmarks for which it is above a threshold. Johnson (2000) proposes a method for terrain matching. Landmarks that are located in terrains with high curvature or in planes are discarded because small changes in the sensor measurement result in great localization errors, or the landmarks are too similar, respectively. For a more detailed review of related work see Gerstmayr et al. (2004b).

1.1.4. Existing work

This work builds on the work presented in Gerstmayr et al. (2004a,b). There two algorithms for appearance–based landmark selection were proposed that try to select landmarks from a set of landmark candidates that are as dissimilar to each other as possible. Since the algorithms use Principal Component methods they are counted to the appearance–based navigation methods. The main advantages of using PCA (see section A.1) is that it tends to obtain compact and efficient representations of the global environment with good generalization capabilities. Although PCA got an established method in localization and object recognition Gerstmayr et al. (2004a) is the only known work that uses pure PCA features as criterion to select landmarks.

The existing work only addresses the second step of the landmark selection problem. The landmark candidates were obtained by dividing the whole image into a regular

grid. Further candidates were positioned at the adjoining points of the others, the overlap between adjacent landmark candidates was at most 25 %. The set of landmark candidates is visualized in figure 1.3.

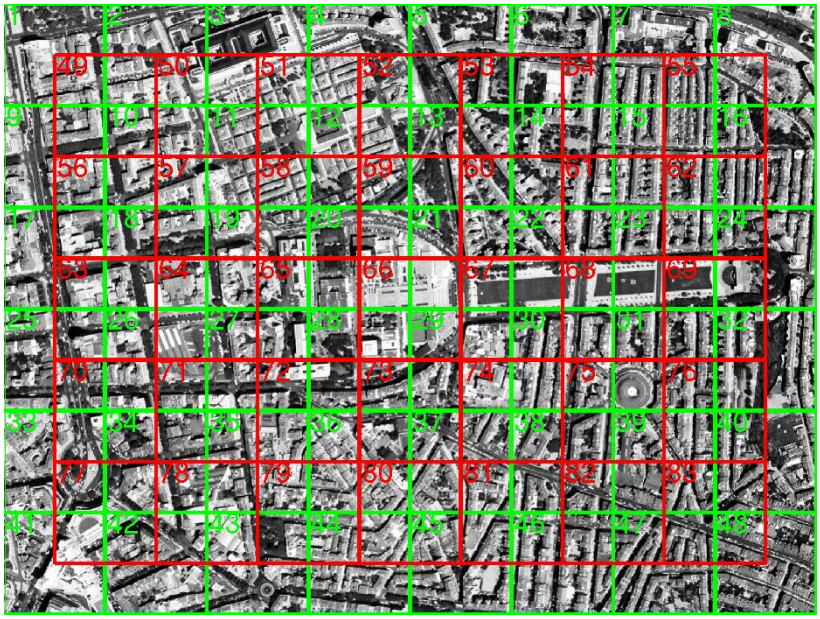


Figure 1.3.: The set of landmark candidates used in Gerstmayr et al. (2004a,b).

The first algorithm, called profile–based algorithm, selects landmarks by comparing the pairwise image dissimilarity between the landmark candidates. A good landmark is a landmark candidate that is as dissimilar as possible to all others. The pairwise comparison of candidates is done in an eigenspace computed by PCA. The second algorithm, called IPCA–based algorithm, uses incremental PCA (IPCA, see section A.1.2) to iteratively increment an already existing eigenspace by adding landmark candidates that can not be expressed accurately in the existing eigenspace. Such landmark candidates are dissimilar to all the other already selected landmarks. The results of Gerstmayr et al. (2004a,b) reveal that none of the algorithms is superior to the other. Both select distinctive landmarks that often contain a unique pattern formed by streets or buildings. The profile–based algorithm does not select landmarks in repetitive areas, whereas the IPCA–based algorithm covers also these regions with landmarks.

1.1.5. Goals of this work

The goal of this work is to overcome the drawbacks of the already existing work. Its main drawback is that it only addresses the second step of landmark selection. As features like corners, edge density, or symmetry have a rather uniform distribution, standard point–of–interest operators can not be applied for aerial images because their discriminative power is limited too much. Additionally simple solutions like downscaling the images did not lead to robust and reliable interest points.

Therefore, the main goal of this work is to find a meaningful representation – or a method for image preprocessing – for the aerial images that robustly extracts the key information contained in aerial images. This representation can facilitate the implementation of a robust point–of–interest (PoI) operator and can further make the selected landmarks more reliable, thus enhancing the navigation and localization capabilities of the blimp.

In order to approach such a meaningful representation and to extract the key information contained in aerial images, it is worthwhile to have a look at related literature from geographical psychology. In Haken and Portugali (2003) it was shown that remembered city elements that are e.g. important for navigation tasks are exactly those that convey the highest quantity of information (Shannon and Weaver, 1949). Thus, their results also motivate the definition of landmark used for this work. The authors apply Shannon information to geometric building blocks of cities[3] which are motivated by the five building blocks proposed in Lynch (1960):

Paths: Paths are channels along which traffic flows such as streets, railroads, or walkways.

Nodes: Nodes are strategic points in the city formed by intersecting paths.

Edges: Edges are boundaries between different structures of the edge.

Districts: Districts are larger–scale structures combining areas with a common character such as construction style or purpose.

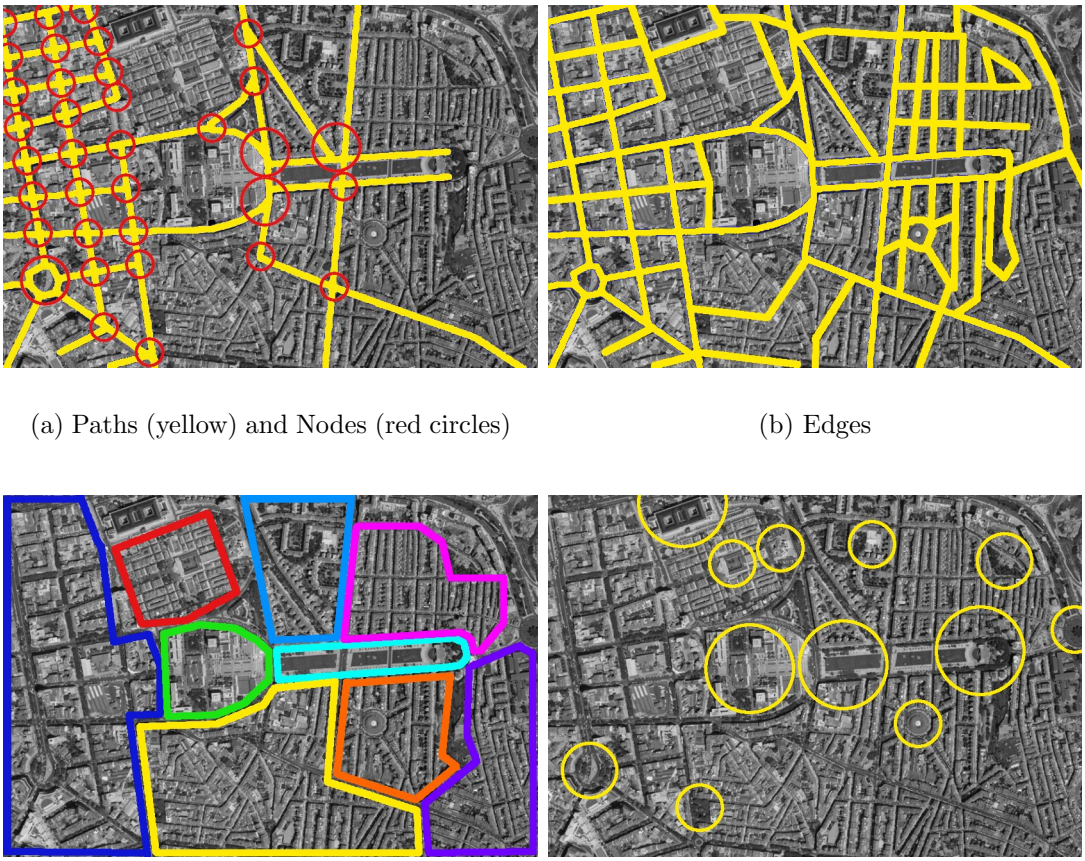
Landmarks: Landmarks are outstanding physical objects like characteristic buildings or mountains that are used as references for navigation.

As shown in figure 1.4 these building blocks can also be established in aerial images. However, these building blocks are already at a rather high level and can be further abstracted to more general urban elements proposed including points, lines, and surfaces (Golledge, 1999). Thus nodes and landmarks can be combined to points whereas paths, edges and the boarders of districts can be fused to lines. Since for aerial images all these low–level structures are closely related to contours in the image, contours were chosen as key information contained.

An alternative motivation for that decision are the Gestalt laws which are early attempts to describe and explain grouping and segregation of visual perception (Spillmann and Ehrenstein, 2003). Among the different laws especially the law of good continuity, which states that collinear parts of a stimulus tend to be grouped together, could explain why the pattern formed by the streets and other long contours in the aerial images pop out for human observers.

As biologically motivated approaches to contour detection and enhancement have achieved very good results, a model for contour detection in the human visual system

[3]For the cited work the authors neglect semantic urban elements that have a personal, cultural or symbolic values for people and that were shown to influence human behavior in cities



(a) Paths (yellow) and Nodes (red circles)

(b) Edges

(c) Districts

(d) Landmarks

Figure 1.4.: Lynch's elements shaping a city

is used. The employed model is based on Hansen (2002) because the models proposed there can be implemented and combined in a very modular way. Motivated by the assumption proposed by Barlow (2001) or Olshausen (2003), telling that any sensory system is optimally tuned to the statistical properties of the stimulus occurring most frequently, the statistical properties of aerial images will be investigated and the findings will be used to optimally tune the contour detection model for aerial images.

This approach has to be understood as prior module for the landmark selection task. Adapting the preprocessing system to the statistics of the environment is assumed to yield an optimal representation of the environment. This representation can facilitate landmark selection, which selects from the optimal representation the most outstanding features.

1.1.6. Related Research Areas

For sake of completeness two additional aspects of related work shall be outlined briefly.

Such work includes projects on unmanned aerial vehicles (UAV). For this thesis, the project described in Hygounenc et al. (2004) is most related. It deals with Simultaneous Localization and Mapping (SLAM) in outdoor environments and also relies on a blimp. Projects using low–level navigation strategies often rely on optic flow for navigation and control of the UAV. For recent reviews see Ruffier and Franceschini (2005) or Muratet et al. (2005).

The other related field to mention includes road detection systems as they are used in geographical information systems. For recent reviews see Mena (2003) and Zhang (2004). In general these models fuse knowledge derived from several sources of information including aerial images of different resolution, satellite images, 3D surface maps, and already existing maps. These approaches also include detailed statistical models about cars, streets, street markings, railroads, vegetation, or buildings, which are then used to proof the knowledge obtained so far. One particular road extraction model is sketched in figure 1.5. It is based on the detection of edges and step by step generates more hypotheses which are then fused to a consistent estimation. This work does not implement such a road extraction network. The first reason is that such models are computationally not tractable due to the runtime requirements of robot navigation. The second reason is that the objective of this work is not the exact detection of roads. It aims at finding simple means of information processing that allow robust navigation.

Before pointing out the structure of this thesis an introduction to the processing of visual information in the brain will be given.

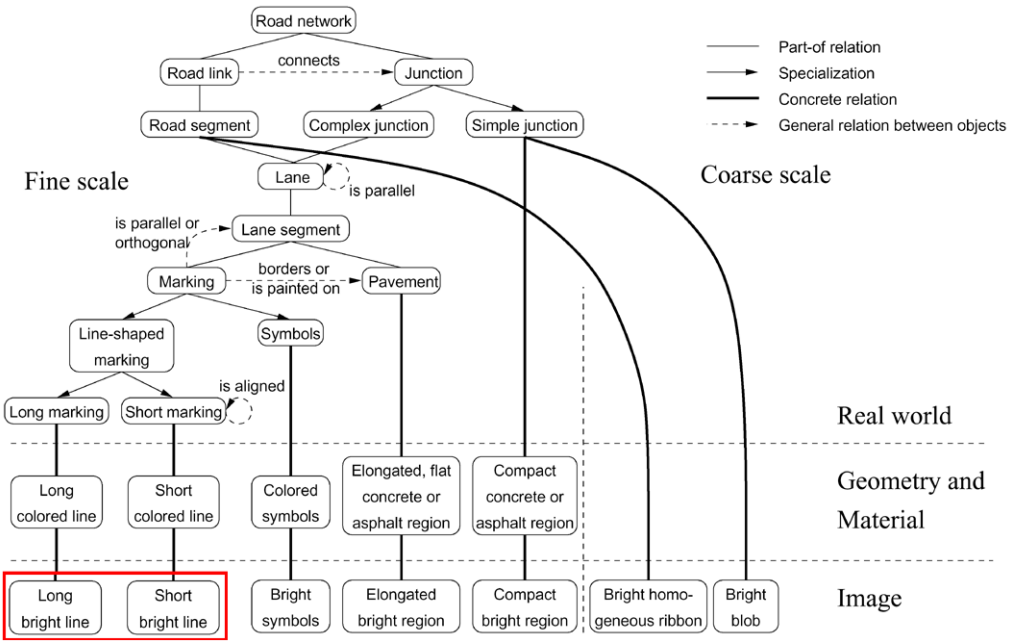


Figure 1.5.: Sketch of a road extraction network for urban areas. The model also starts with edge detection (marked in red) and then fuses several hypothesis to a consistent road map. From Hinz and Baumgartner (2003).

1.2. Processing of Visual Information by the Brain

Humans get most of their sensory perception by vision. During the time fascinating organs and incredible processing facilities have evolved that are still far to complex to be fully mimicked by computers – although according to Olshausen (2003) their computational power with respect to speed and memory capacity has grown for a factor of 1000 over the last 20 years. For the next paragraphs the most important aspects of the human visual system and the underlying aspects of neural information coding that are necessary to understand this thesis shall be summarized.

1.2.1. The Human Visual System

This summary is based on Kandel et al. (2000), reviews about the cutting–edge research in visual neurosciences can be found in Chalupa and Weber (2003a,b).

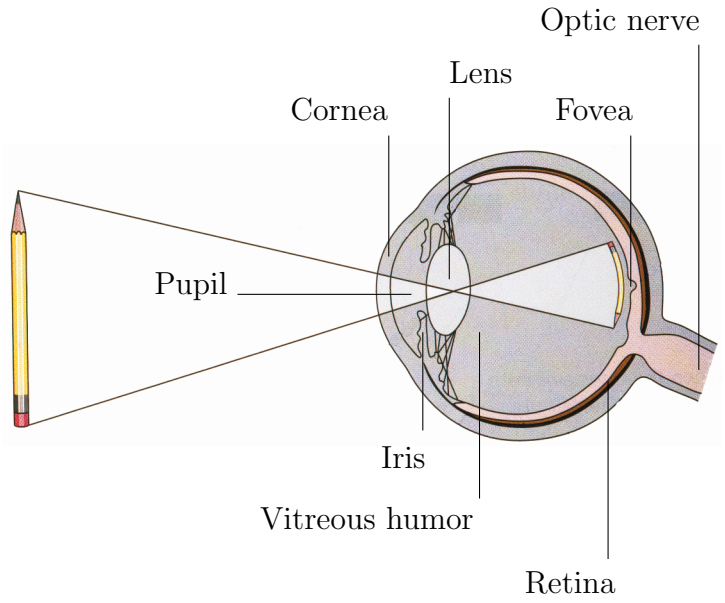
1.2.1.1. Retina

Light reflected or emitted from objects, which are a part of the world, passes the cornea, the lens, and the vitreous humor of the eye before reaching the photoreceptors located in the retina. The retina is the first processing stage and contains two types of photoreceptors transforming physical light intensities to electric potentials: rods and cones. Rods are designed to detect dim light because they amplify signals stronger than cones do, they reinforce each other, and their signals are pooled by bipolar cells. Rods optimally respond to light with a wavelength of 496 nm.

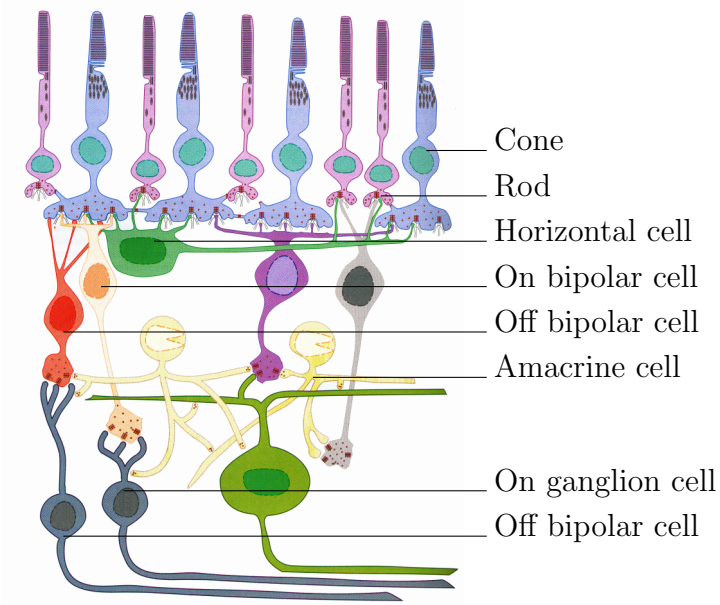
The cones mediate color vision. Although there are roughly 20 times more rods than cones, the cones have a much better spatial resolution than the rod system because they are concentrated in the fovea, the part with the best resolution in which no rods are, and because less cones are pooled by bipolar cells. The color sensitivity is due to different visual pigments that are sensitive for different parts of the light spectrum. There are three subtypes of cones that are optimally tuned for short (419 nm), middle (531 nm) and longer (559 nm) wavelengths. Therefore, they are often called S–, M–, and L–cones or blue–, green– and red–cones respectively.

Like for the photoreceptors there also exist two types of bipolar cells: rod and cone bipolar cells depending on the type of the photoreceptor projecting to them. Since the cone system is not as complex as the rod system the following description will focus on the cone system. Cone bipolar cells can be further divided into on–center and off–center cells. On–center cells depolarize if light is falling onto the center of their receptive fields and hyperpolarize by light falling onto the surrounding cones. The opposite holds for off–center cells. Each cone synapses with both on–center and off–center bipolar cells. The potential of surrounding cones is mediated by horizontal cells.

The antagonistic center–surround organization and the circular receptive fields of bipolar cells are the first processing steps towards contrast perception and can also be found in ganglion cells. An on–center ganglion cell optimally responds to a bright spot surrounded by a dark annulus while an off–center ganglion cell is best tuned to the opposite



(a) Schematic drawing of the eye. Adapted from Kandel et al. (2000).



(b) Summary diagram of cell types in the retina. Unlabeled cell types are not relevant for the understanding of this thesis. Adapted from Zigmond et al. (1999).

Figure 1.6.: The human eye

kind of stimulus. Ganglion cells transduce the potential changes of photoreceptors and bipolar cells into action potentials. Their axons form the optic nerve and project in parallel pathways for on–center and off–center ganglion cells to the LGN. Amacrine cells mediate between different ganglion cells.

The ganglion cells can also be divided into P– and M–cells. P–cells constitute about 80% of all ganglion cells. They can be further divided into a red–green and a blue–yellow type. The red–green cells receive input from L– and M–cones, the blue–yellow type receives input from all cone types. Within both classes different connection patterns result in the characteristic on–center and off-center response properties. The M–ganglion cells are achromatic and have larger receptive fields than the P–cells.

1.2.1.2. Lateral Geniculate Nucleus

The LGN is part of the thalamus and the major target of retinal ganglion cells. It has a complex layered structure and shows a precise topographic organization, such that adjacent cells have adjacent retinal receptive fields. The properties of LGN–cells are quite comparable to retinal ganglion cells with respect to their receptive field shape and spectral tuning.

The LGN receives most of its input from cortical feedback neurons. Therefore it is much more than a simple relay station between the retina and the cortex. It has been proposed that the LGN is involved in attentional mechanisms and selection of salient information.

On–center and off–center receptive fields have been successfully modeled by DoG–filters as proposed by Marr (1982). The antagonistic structure of an on–cell is represented by subtracting a surround Gaussian with large standard deviation from a center Gaussian with smaller standard deviation, or for an off–cell by subtracting a Gaussian with small standard deviation from a Gaussian with larger standard deviation. The result is a mexican–hut–like function visualized in figure 1.7. The positive part models the on–subfield of the receptive fields and its negative part models the off–subfield.

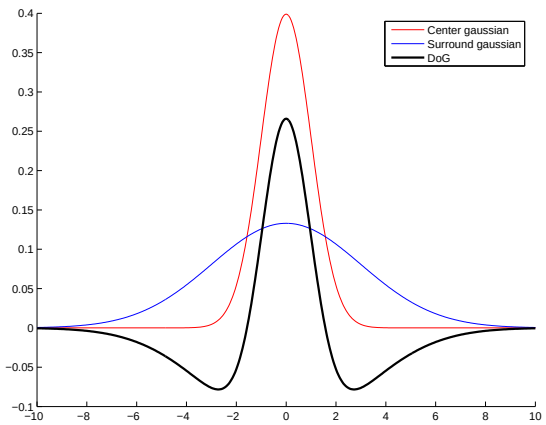


Figure 1.7.: Difference of Gaussians (DoG). The red, blue and black lines represent the center Gaussian, the surround Gaussian, and the DoG function, respectively.

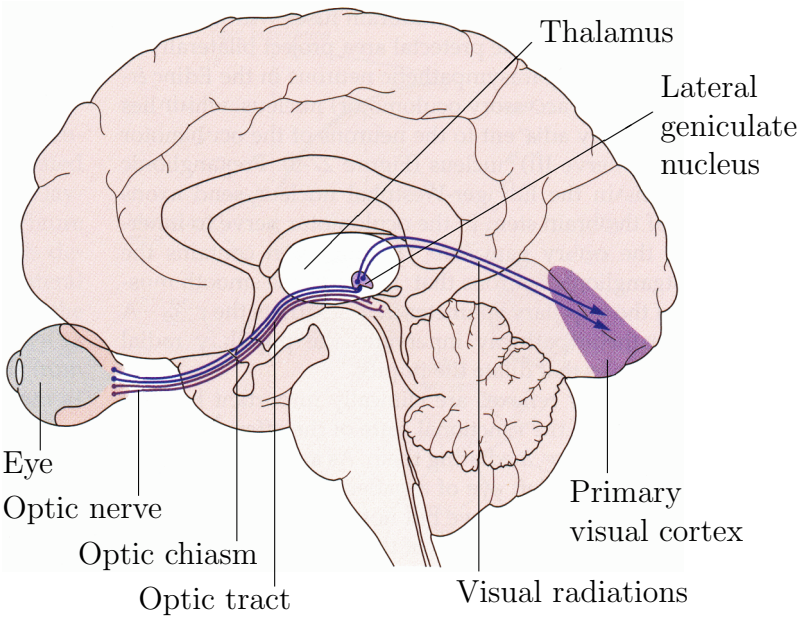


Figure 1.8.: The visual pathway in humans. Adapted from Kandel et al. (2000).

1.2.1.3. Primary Visual Cortex

The LGN projects along the optic radiations to V1 which has a layered structure based on anatomic criteria. Each layer has specific afferent and efferent connections determing the layer's functional properties. Additionally, the topographic representation of the LGN cells is kept. The mapping is highly nonlinear because the representation of the fovea is much larger than the representation of the periphery. The mapping can be modeled by a log–polar transformation.

In V1, most receptive fields are elongated with an axis of preferred orientation. Therefore, they are optimally tuned to detect bars and lines. The cell types can be categorized into two major groups: simple cells and complex cells. Simple cells are characterized by alternating, elongated subfields with a preferred orientation. Therefore, the most effective stimulus is a light/dark patch that coincides with the simple cell's subfields. Orientation selectivity and sensitivity to dark/light and light/dark transitions arises because each location of the visual field is analyzed by many different simple cells with different properties. These properties are determined by pooling a large number of properly aligned LGN–cells with certain properties: the simple–cell's on–subfield receives input from adjacent on–center cells, the off–subfield receives input from off–center cells.

Simple cells are often modeled by Gabor functions which are, as depicted in figure 1.9, the product of a Gaussian envelope and a sinusoidal function resembling a plane wave pattern:

$$G(x,y) = \frac{K}{2\pi s_1 s_2} \exp\left(-\frac{1}{2}\left(\left(\frac{x'}{s_1}\right)^2 + \left(\frac{y'}{s_2}\right)^2\right)\right) \cos\left(2\pi F x' + \varphi\right) \tag{1.1}$$

with

$$x' = (x - m_x)\cos\theta + (y - m_y)\sin\theta \tag{1.2}$$
$$y' = (y - m_y)\cos\theta - (x - m_x)\sin\theta \tag{1.3}$$

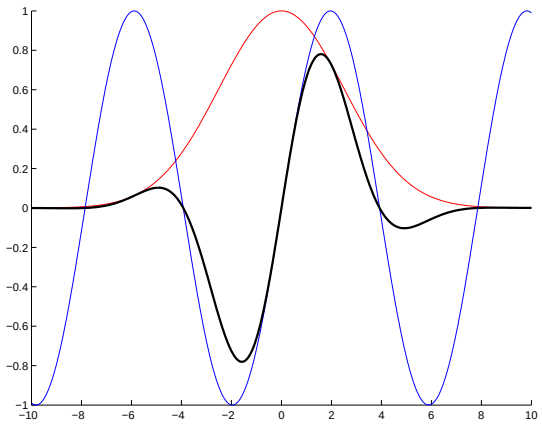


Figure 1.9.: Gabor Function as product of a sinusoidal and a Gaussian. The red, blue and black lines denote the Gaussian envelope, the sinusoidal and the Gabor function, respectively.

The standard deviations s_1 and s_2 are oriented along the direction θ and $\theta + \frac{\pi}{2}$, respectively. Usually a sine function generates an anti–symmetric Gabor functions and cosine functions are used for the symmetric case. Since in (1.1) a phase shift φ was introduced a cosine function can be used for both cases. Anti–symmetric functions are generated for $\varphi = \pm\frac{\pi}{2}$, symmetric functions are generated for $\varphi = 0$ or $\varphi = \pi$. The so–called n_1–value, which is defined as $n_1 = Fs_1$, is a rough estimate for the number of lobes of a Gabor function. The larger it gets the more alternating on– and off–subfields the corresponding receptive field has. In the literature a related measure, the bandwidth b, is often used which is defined as $b = 2\pi n_1$. The n_2–value, defined as $n_2 = Fn_2$, is a measure for the elongation of a Gabor function. The larger it gets the more elongated is the receptive field. The other parameters are explained in table 1.1.

Like for DoG–functions on–subregions are modeled by positive areas, off–subregions by negative areas. Gabor functions have been proven to be an optimal trade–off between being localized in space and frequency (Daugman, 1985). According to Hansen (2002), this reflects the duality of simple cells being also located as well as frequency selective.

Complex cells also have a preferred axis of orientation and elongated subfields, but their receptive fields cannot be divided into on– and off–subfields. They pool simple cells of different contrast polarity with the same preferred orientation. Therefore, their main property is contrast invariance which according to DeAngelis and Anzai (2003) cannot be described by a linear system.

Most of the simple cells and complex cells respond best to achromatic stimuli and only few cells respond most to color variations. This fact is due to the relatively small amount of information carried by color variances.

Table 1.1.: Parameters for Gabor functions

Parameter	Meaning	Unit
m_x	Center position in x–direction	Pixel
m_y	Center position in y–direction	Pixel
θ	Orientation, $0 \leq \theta < \pi$	rad
s_1	Standard deviation along main orientation θ	Pixel
s_2	Standard deviation perpendicular to main orientation θ. For practical use it holds $s_1 < s_2$.	Pixel
F	Spatial frequency parameter. The spatial frequency is the reciprocal of the wavelength $F = \omega^{-1}$	Cycles per pixel
K	Amplitude scaling factor. Usually $K = 1$	–
φ	Phase shift.	rad

V1 not only has a layered structure but also a columnar organization. Each of those columns spans all cortical layers and contains for example neurons with a preferred axis of orientation. Adjacent orientation columns show a shift of the their preferred orientation. Therefore, several orientation columns form a hypercolumn which can perform an analysis of a certain region in the visual field for all orientations.

Additionally, there are horizontal long–range connections in the V1 that can influence response properties of a cell by mediating the cell activity depending on the activation of a cell from outside the classical receptive field. It has been shown that horizontal long range connections contribute to contrast and contour processing.

1.2.2. Relevant Aspects of Neural Coding

As it has been shown in the previous section the human visual system can be understood as an hierarchical information processing system. This system progressively extracts information by analyzing local features and combines these features to higher level representations. In this section the focus will be on how to optimally encode the visual input and how this encoding can be modeled mathematically.

1.2.2.1. The Role of Redundancy Reduction

Shortly after C. Shannon came up with his famous work on information theory and channel capacity (Shannon and Weaver, 1949), people like F. Attneave and H. Barlow started to apply these findings to visual neuroscience (Attneave, 1954; Barlow, 1961). They argued that stimuli perceived from the natural environment have to be redundant and that the sensory processing should be optimally adapted to deal with the occurring redundancies. These redundancies might be a valuable source of knowledge about the environment. If ignored, they might lead to errors in estimating probabilities of hypotheses about the environment (Barlow, 2001). During that time, the key idea was that all redundancy should be reduced because it is unnecessary information.

A couple of decades later, Attneaves's and Barlow's ideas were more testable because of the progresses in computer hardware as well as in neuroscience. In Barlow (2001) is mentioned that Shannon's work is still valid and correct, but the brain does not deal with information in terms of classical communication engineering. So the concepts have to be more flexible for neuroscience. It is further argued that discovering the statistical structure of sensory input is still an important aspect. Nevertheless, the best way to deal with the information is not necessarily redundancy reduction but rather to code the information depending on the use that is made out of it.

1.2.2.2. Perception as Statistical Inference

Before discussing which coding strategy leads to meaningful representations in the visual system, it is worth to formalize how the brain can infer knowledge about the environment from the perceived data.

As mentioned in Olshausen (2003), this task is extremely difficult as there is no unique solution for the mapping from environmental properties to sensor activations because the 3D structure of the world is perceived as 2D image. Therefore, the eye can only do probabilistic inference about the world, since some object configurations in the world leading to the same sensory input are more likely than others.

Following Barlow (2001) and Olshausen (2003) this process of statistical inference can be modeled in terms of Bayesian inference (Duda et al., 2001): A certain state E of the environment results in a certain state of receptor activity A. Then, the conditional distribution $P(A|E)$ describes the probability of the activation pattern A given the state E. Further on, the receptor has some knowledge about which properties of the environment are more likely than others. This knowledge is modeled in the prior distributions about the environment $P(E)$ and the activation patterns $P(A)$. To do inference one can now follow Bayes' rule

$$P(E|A) = \frac{P(A|E)P(E)}{P(A)} \tag{1.4}$$

and compute the state E that is most likely to lead to the given receptor activation A.

The above formalization of the problem makes clear that robust and reliable inference is only possible if the underlying model represents the statistical properties of the environment and receptor activity properly. Thereupon, as mentioned in Torralba and Oliva (2003), these regularities are a relevant source of information concerning top–down and contextual priming in the visual system. The importance of analyzing the statistical properties of a certain environment is even stressed if one considers that the number of possible images is extremely large. But in contrast, the number of images that arise from a certain environment are almost infinitely small, far from random, and showing a large degree of characteristic structure (Ruderman, 1994; van der Schaaf and van Hateren, 1996; Olshausen and Field, 1996; Srivasta et al., 2003).

Closely related to characterizing properties of the environment is the question to which kind of stimulus the processing system is optimally tuned. In Hyvärinen and Hoyer (2001) is mentioned that no statistical signal processing system can be optimally tuned to process any input. The reasons given in Simoncelli and Olshausen (2001) or

Hyvärinen et al. (2003a) include that the visual system is important for survival and reproduction. Thus it should be optimally tuned to the kind of stimulus occurring most frequently. The tuning can be a result from evolution or adaptation and learning in the individual development. The experiments reported e.g. in Blakemore and Cooper (1970) give hints that adaptation during the development is more essential.

Therefore visual neuroscience can gain new insights from modeling the properties of an environment, applying different coding principles, and than testing which one leads to the most meaningful representation of the data and best fits the responses measured in real neurons.

1.2.2.3. Principles of Neural Coding

From the viewpoint of computational neuroscience coding schemes can be roughly organized into two groups, namely compact coding and sparse–dispersed coding. The definition given here is based on Willmore et al. (2000).

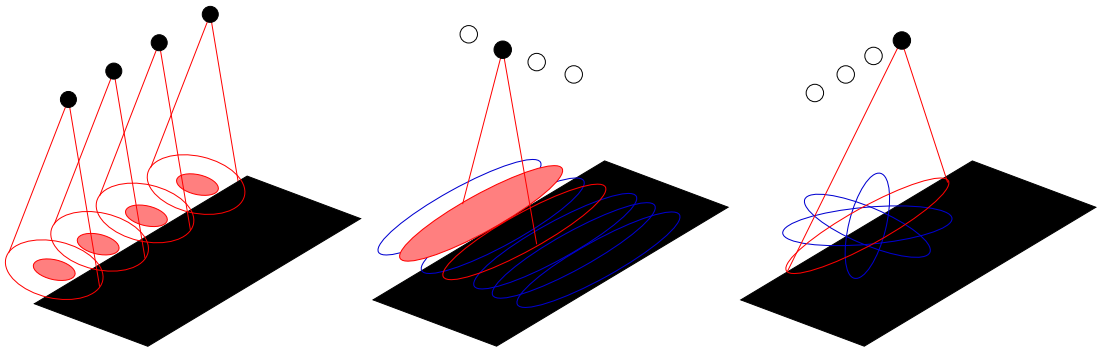
In compact coding the number of neurons needed to encode information accurately is minimized. This perfectly fits into the signal processing theories of efficient coding. Mathematically compact coding is closely related to PCA (see section A.1) because it seeks a transformation of the input such that the significant parts of the information are coded in only little dimensions or (in terms of neurons) in only few units. Compact codes are also distributed, which means that each unit is involved in representing many different entities.

A sparse–dispersed code represents certain aspects of the information by a different, relatively small subset of all units in the population. It therefore fuses the advantages of a sparse code, requiring that at any time only few coding units are active, and of a dispersed code, which requires that all units contribute equally to the overall coding. In the related literature the term sparse code is frequently used in the sense of a sparse–dispersed coding. Mathematically sparse–dispersed coding can be modeled by ICA (see section A.2), a method maximizing statistical independence of the resulting codes.

Also population coding and "grandmother–neurons" can be explained in terms of sparse–dispersed coding: For population coding, neurons are optimally tuned for a certain stimulus. They show weaker responses for similar stimuli and do not respond at all for stimuli being very different to the optimal one. Population encoding occurs if few units respond (differently) to a certain kind of stimulus. Grandmother–neurons are an extreme form of sparse–dispersed coding, in which each unit is only active for one and only one aspect of the input information.

1.2.2.4. Coding Principles in the Visual Pathway

Following Olshausen (2003) it can be stated that sparse–dispersed coding is the preferred coding strategy in the visual system. By pooling cells of lower stages the information can be represented using less active neurons in higher stages. This idea is depicted in figure 1.10. The main difference between the retina and the LGN on the on hand and the cortex on the other hand is how the areas deal with the occurring redundancy.



(a) Redundancy reduction in the retina by pooling over several photoreceptors.

(b) Simple cells reduce redundancy by pooling several properly aligned center–surround cells. Thus the edge is represented with fewer active units. Additionally the resulting code is sparse as only the optimally tuned simple cell is active.

(c) Orientation selectivity of simple cells is an example for increasing the redundancy as there are several cells each coding a specific orientation. Also the code is sparse as only the optimally tuned unit is active.

Figure 1.10.: Coding principles in the visual pathway. Filled circles denote active units, red and blue lines the shape of active and inactive receptive fields. Adapted from Olshausen (2003)

In the retina redundancy has to be reduced because 100 million photoreceptors project onto 1 million ganglion cells and the retina is further constrained by the limited number of axon fibers that can leave the eye via the optic nerve.

In contrast to the retina, redundancy is actually increased in the cortex. As V1 has much more output than input neurons coming from the LGN and as assuming an approximately constant bandwidth of the neurons, information cannot be created in the cortex. This overcomplete representation allows each output to carry a very specific interpretation of the input pattern occurring at a certain location of the input image with a particular scale and orientation. However, each cortical simple cell also sparsifies the signal by pooling several properly aligned cells in LGN. This sparse overcomplete representation makes further analysis like contour grouping much easier because less neurons have to be taken into account to model their relationships. Another advantage mentioned in Barlow (2001) is that redundancy is more useful for error avoidance.

Sparse–dispersed coding is also the coding principle underlying complex cells. It can be argued that their characteristic properties of being phase invariant can be achieved by pooling certain simple cells, again leading to a more meaningful representation. The latter makes it also easier for further processing steps to derive information, e.g. about contours.

1.2.2.5. Modeling Sparse–Dispersed Coding

In order to find out transformations of the input image to a meaningful representation in several works like Bell and Sejnowski (1997), van Hateren and van der Schaaf (1998), Olshausen and Field (1997), Hoyer and Hyvärinen (2000), Hoyer and Hyvärinen (2002) or Hyvärinen et al. (2003a) a linear superposition model has been assumed. As sketched in figure 1.11, the model describes an image $\mathbf{I}$ as a linear combination of n basis images ψ_i with amplitudes a_i:

$$\mathbf{I} = \sum_{i=1}^{n} a_i \psi + \nu. \tag{1.5}$$

Here, ν denotes all that parts of an image that cannot be expressed by the linear combination like nonlinear effects or image noise. For purely linear models ν is neglected. For sure the linear model cannot cover all aspects of images. For example in in Zetsche and Röhrbein (2001) it is argued that linear mechanisms cannot describe images and cortical mechanisms properly. However, the linear superposition model seems to cover the most important aspects of image representation reasonable well.

Figure 1.11.: Linear superposition model. An image is represented as a linear combination of several basis images.

In works like Bell and Sejnowski (1997), Hoyer and Hyvärinen (2000), Srivasta et al. (2003) it has been pointed out that a transformation only decorrelating the input data (for example by applying a PCA–like transformation) is not sufficient to describe the visual information processing. Tightly coupled with this result is that images cannot be described by Gaussian distributions and higher–order dependencies play an important role in image statistics. Therefore, stronger criteria have to be used in order to estimate the base images.

One such criterion often used is the sparseness of the resulting code. That means that the basis functions ψ should allow to describe each image $\mathbf{I}$ by only few basis functions. As consequence only few of the coefficients a_i are nonzero, while most of them are zero. The resulting basis functions share the same properties like V1 simple cells, i.e. they are spatially localized, oriented, and bandpass. It also has been shown, e.g. by van Hateren and van der Schaaf (1998), that they can be described by Gabor functions reasonably well.

Rather similar results can be achieved by ICA (see section A.2) which searches a basis such that the transformed information is a linear combination of sources that are as independent as possible. One frequently used approach to maximize the independence of the resulting codes is to maximize the forth–order moment, the kurtosis. Since the distribution of a sparse–dispersed code is highly kurtotic Olshausen and Field (1997) could proof the equivalence between ICA and sparse–coding approaches.

In the last couple of years ICA has been used to model several aspects of neural coding in areas related to visual information processing: Early work like Bell and Sejnowski (1997) and van Hateren and van der Schaaf (1998) showed that the independent components of natural images are a sort of edge filters that can be described by Gabor filters and therefore resemble the receptive fields of simple cells. Olshausen and Field (1997) modeled the increase of redundancy in V1 by estimating a sparse dispersed code with an overcomplete basis set. For a review and a comparison of these and other similar studies see Willmore et al. (2000). ICA–models have also been used to explain the spatial and color tuning properties of V1 neurons (Caywood et al., 2004) and of binocular receptive fields tuned for different disparities (Hoyer and Hyvärinen, 2000).

These simple ICA–models have also been extended to non–negative sparse coding which is biologically more plausible since no negative activations can appear in the nervous system (Hoyer, 2002). Work like Hyvärinen and Hoyer (2000), Hyvärinen et al. (2001a) or Hyvärinen and Hoyer (2001) could explain the emergence of invariance properties comparable to those of complex cells and of topographic orderings like in the visual cortex. In Hoyer and Hyvärinen (2002) a sparse coding network is used to learn contour coding on top of a given complex–cell response. Reviews of these extensions can be found in Hyvärinen et al. (2003a,b) or Inki (2004).

Based on the striking resemblance of all these findings to receptive fields in V1 and visual information processing one can conclude that sparse–dispersed coding plays an important role. However, it should be stressed that all the theoretical findings reported here are just mathematical models describing certain aspects of visual information processing. It does not have to mean that the "algorithm" implemented by cortical neurons is similar to the algorithms of these models (Hoyer and Hyvärinen, 2000).

1.3. Outline of this Thesis

As outlined in section 1.1.5 the main contribution of this work is a point of interest operator that is supposed to work on an edge representation of the input image. For the edge and junction–point detection models inspired by the human visual system will be used. Every biological perception system is optimally tuned to the environment. So, a further key aspect of the work described here is to take the statistical properties of the input images as well as some basic assumptions about neural information processing into account to determine the model's parameters. In order to achieve these objectives, the results of different sub–parts had to be combined. These are sketched in figure 1.12 and described in the following.

In section 2.1 the statistical properties of different image classes, namely natural scenes, manmade scenes, and aerial images of urban areas, are compared by computing Power Spectra. The results reveal that the statistical properties of these classes are indeed different. This motivates a more detailed analysis how the statistical properties of the environment influence the shape of the receptive fields.

The analysis is presented in section 2.2. By computing the ICA for a huge number of image patches the shape of receptive fields can be estimated. A statistical comparison

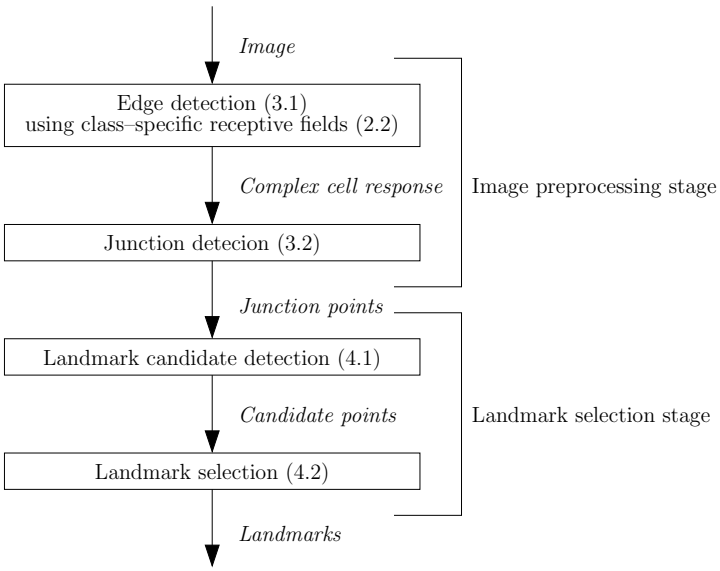


Figure 1.12.: Sketch of the different processing stages. The numbers refer to section numbers, the information passed from stage to stage is denoted in italics.

of the estimated receptive fields was done after fitting Gabor functions. It reveals significant differences between the image classes, especially with respect to the elongation and the distribution of preferred orientations. Based on these findings, class–specific receptive fields for the image classes are derived: Receptive fields tuned to natural scenes are sensitive for short edges in various orientations, those tuned to manmade scenes are sensitive to short contours which are mainly horizontally and vertically oriented, and the receptive fields tuned to aerial images are sensitive to long contours in various orientations.

The class–specific receptive fields are used as edge–detection filters in the simple cell model proposed in section 3.1. The model is a feedforward model including mutual inhibition as well as various other nonlinearities to sharpen the orientation tuning and to suppress noise. The model parameters are determined by using an optimization that tries to increase the distinctiveness between responses obtained for image elements lying on an edge and those not lying on an edge. The results show that the edge representation obtained by the simple cell model describes the image structures best if processed with the receptive fields adapted to the image class.

Based on the edge representation, the next step (presented in section 3.2) includes the detection of junctions. Therefore, an implicit model is used that detects junction points by detecting hypercolumns showing activity in various orientation channels. So far, the junction detection model is the bottle–neck of this work. It is very difficult to set the detection threshold in order to achieve good and meaningful detection results. Hopefully these drawbacks can be overcome in future work by implementing a grouping mechanism that enhances the responses to collinear edges and can further reduce noise in the edge representation.

In chapter 4 the biologically motivated models for early vision mechanisms are combined to solve the landmark selection problem. In section 4.1, a detection algorithm for landmark candidates is proposed which is based on the junction detection model. Due to the dense distribution of junction points in the aerial images and the problems with setting the detection threshold, additional distance constraints had to be added. However, the detection algorithm selects landmark candidates with high changes in the image texture.

The detected candidates are used as input for the existing landmark selection algorithms and an extension of the profile–based algorithm to select landmarks based on ICA features. Section 4.2 proofs the results of previous works showing that the landmark selection algorithms select distinctive landmarks that are often characterized by different patterns formed by streets or buildings.

The thesis closes with a final discussion (chapter 5). Future working directions beside the implementation of the contour grouping mechanism include experiments with a real robot or a detailed simulation as well as an empirical analysis of the landmark's reliability.

2. Estimation of Class–Specific Receptive Fields

Sensory systems have evolved to optimally process the kind of stimuli occurring most frequently and to transform the input into a meaningful representation facilitating further processing (Barlow, 2001; Simoncelli and Olshausen, 2001). For this reason the receptive fields of the neurons "implementing" that transformation are supposed to be optimally adapted to the statistical properties of the input, i.e. the perceived parts of the environment. In this chapter it is studied how the environment influences the shape of the receptive fields for different environment types. The "environments" are images from natural and manmade scenes and aerial images of urban Portuguese areas. The image classes are characterized by short contours which are in various orientations, short contours mostly horizontally and vertically oriented, and long contours in various orientations, respectively. In section 2.1 it is shown by analyzing averaged Power Spectra of the image classes that these classes have different statistical properties. Based on these findings in section 2.2 class–specific receptive fields, i.e. receptive fields which are optimally tuned to the statistical properties of the image class, will be estimated. The main parts of this section have been presented as a poster at the Eighth Tübingen Perception Conference (Gerstmayr and Mallot, 2005).

2.1. Analysis of Ensemble Power Spectra

In order to show that used image classes have different statistical properties, power spectra were computed and the standard analyses suggested in the related literature were done. The power spectrum of a function is equivalent to the Fourier transformed of the autocorrelation function therefore giving insight into the correlation, i.e. the second order statistics, between different points in the image.

2.1.1. Experiments

The used images of the natural and the manmade classes were taken from the Corel Stock Photo Library showing landscapes, flowers, or animals and houses, cities, industrial facilities, or indoor–scenes, respectively. The aerial images are showing urban areas including coast and river lines, parks and city limits and were provided by the *Laboratoria de Visão* of the *Instituto de Sistemas e Robótica* at Lisbon, Portugal. A representative collection of the images is shown in figure 2.10. The natural and the manmade class contained $\rho_{n,m} = 2520$ images each, the aerial class contained $\rho_a = 2459$ images. All

images were grayscale images sized 256×256 pixels. Each of the manmade and natural images was taken from the center of an image sized 768×512 pixels. The aerial training images have been taken at regular grid positions from urban aerial images of different sizes, but the same resolution with a maximum overlap between two images of 50 %.

For each image the power spectrum $S(f_x, f_y)$ defined as the squared modulus of the 2D–DFT of the image $I(x, y)$ was computed according to

$$S(f_x, f_y) = |I(f_x, f_y)|^2 \tag{2.1}$$

$$= \left| \sum_{k_2=0}^{N-1} \sum_{k_1=0}^{N-1} I(k_1, k_2) h(k_1, k_2) \exp\left(\frac{2\pi i k_1 f_x}{N}\right) \exp\left(\frac{2\pi i k_2 f_y}{N}\right) \right|^2 \tag{2.2}$$

with $N = 256$ and $h(x, y)$ the Blackman–Harris Window function as defined in equation (A.52).

As a first experiment Ensemble Power Spectra (EPS), defined as the average spectrum of all the ρ spectra of an image class,

$$EPS = \frac{1}{\rho} \sum_{i=1}^{\rho} S_i \tag{2.3}$$

were computed. The contour plots of the results are shown in figure 2.1.

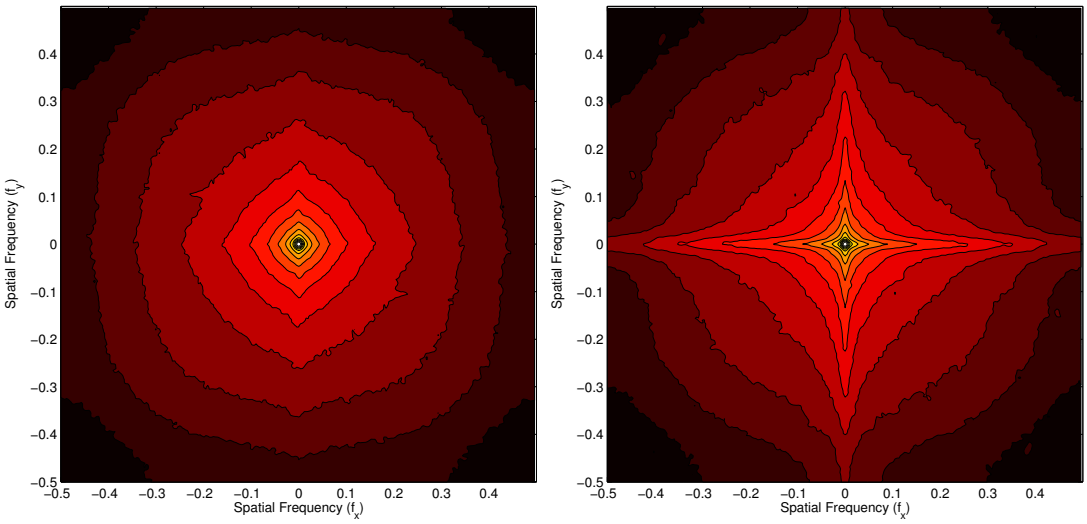
As a second analysis the power spectrum has been transformed to polar coordinates (f, θ) with 16 angular and 24 radial sectors. All the elements within a sector have been averaged, elements that were further away from the origin than the maximal radius of 128 pixels have been discarded.

In Ruderman (1994), van der Schaaf and van Hateren (1996), Oliva and Torralba (2001), Balboa and Gryzwacz (2003), and many others it has been shown that the average energy decay can be modeled as function of the orientation θ and the frequency f

$$S(f, \theta) = \frac{b(\theta)}{f^{-a(\theta)}} \tag{2.4}$$

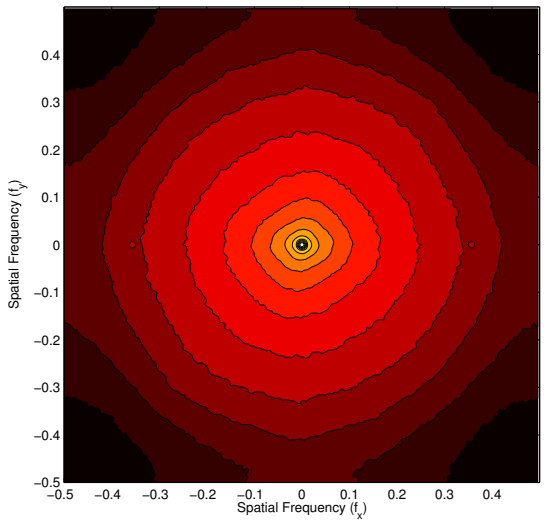
with $\alpha \approx 2 \; \forall \; \theta$ and $b(\theta)$ being a scaling function. This property arises from the scale invariance of natural images (Simoncelli and Olshausen, 2001; Mumford and Gidas, 2001; Srivasta et al., 2003). Using logarithms the right hand side can be written as a linear equation with slope $\alpha(\theta)$ and $\beta(\theta)$ being the intercept of the y–axis.

Since within the scope of this paragraph it shall only be shown that aerial images are different from other classes, there was no such detailed evaluation like in Ruderman (1994) or van der Schaaf and van Hateren (1996). In order to analyze the frequency decay, α and β were computed for the average energy of all orientations therefore assuming the EPS to be isotropic. The frequency bins for the six highest frequencies have not been taken into account for the evaluation since their energies cannot be estimated reliably. The results are summarized in table 2.1 and figure 2.2. To evaluate the frequency decay in dependence of the orientation, the mean energy has been computed for all possible orientations by averaging over the frequencies. The results are shown in figure 2.3.



(a) "Natural" images

(b) "Manmade" images

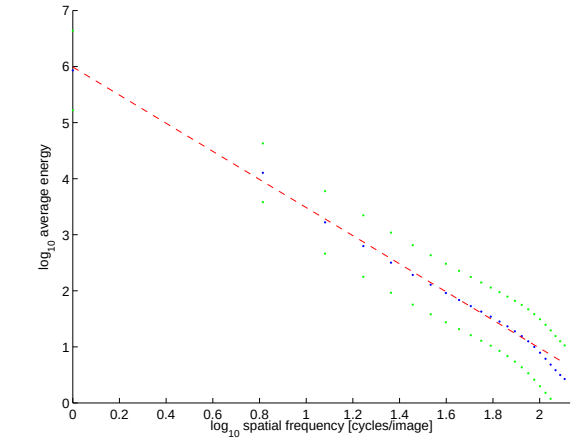


(c) Aerial images

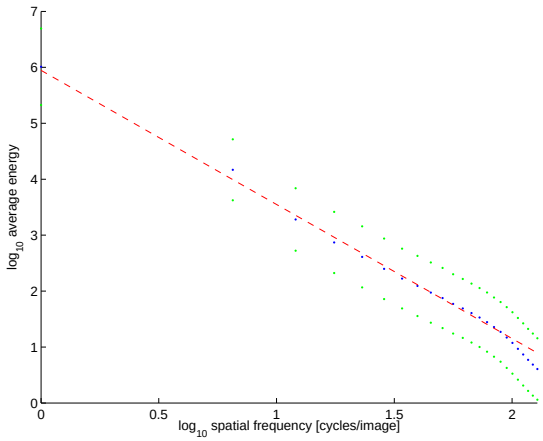
Figure 2.1.: Contour plots of ensemble power spectra (EPS) for different image classes. Black codes zero energy, the brighter the color is, the higher gets the energy.

Table 2.1.: Results of the linear fit to describe power spectra of different image classes.

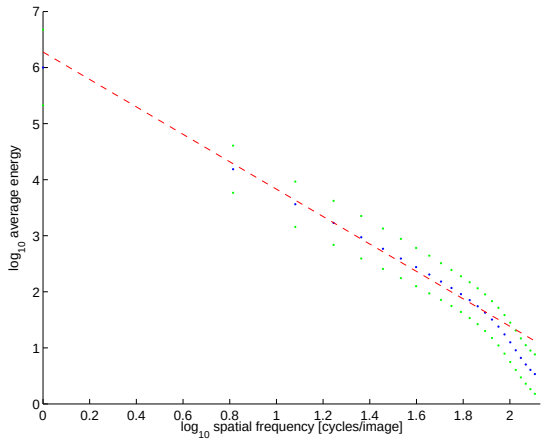
Category	α	β
Natural	-2.5075	5.9921
Manmade	-2.4037	5.9621
Aerial	-2.4462	6.2768



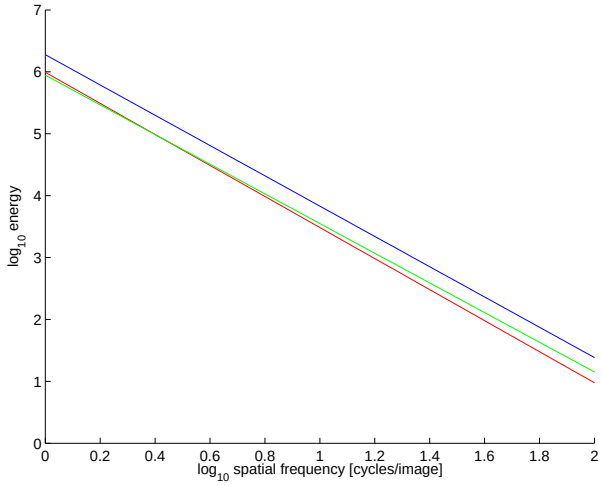
(a) “Natural” images



(b) “Manmade” images



(c) Aerial images



(d) Comparison of fitted lines

Figure 2.2.: Energy decay against spatial frequency. In plots (a) to (c) the blue dots represent the mean energy for a certain spatial frequency, the red dots are the mean plus/minus the standard deviation. The red dashed line is the fitted line. Plot (d) compares the fitted lines for the different image classes considered within this work, blue, red and green lines denote the line for the natural, the manmade, and the aerial image class, respectively.

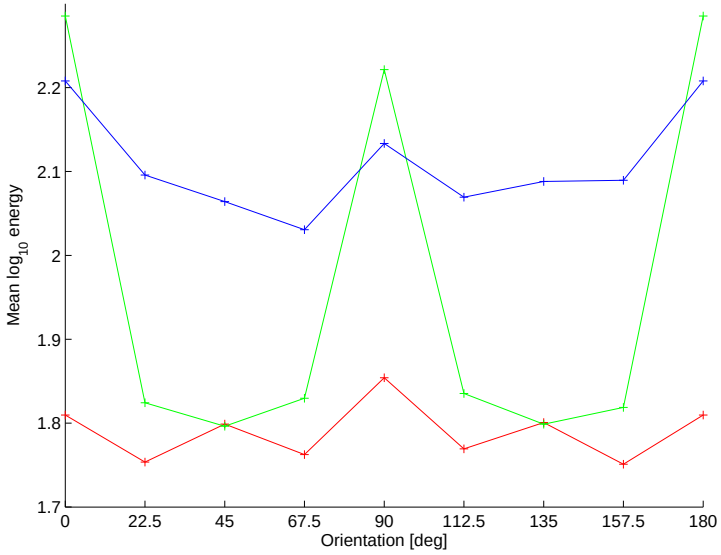


Figure 2.3.: Mean energy against orientation. The blue, red, and green line denotes the natural, the manmade, and the aerial image class, respectively.

2.1.2. Results and Discussion

The contour plots in figure 2.1 show that the EPS for natural and manmade scenes have the typical shapes also described in related literature like Oliva and Torralba (2001) or Torralba and Oliva (2003): The EPS of manmade structures has large tails for horizontal and vertical orientations. The spectrum for natural scenes is almost isotropic with a slight tendency towards horizontal end vertical oriented structures. For the aerial image class the spectrum looks almost isotropic showing that there is no preferred orientation in the used image dataset.

These findings are strengthened by analyzing the mean energy in relation to the orientation (figure 2.3). Again, manmade scenes show large peaks for horizontal (0°) and vertical (90°) orientation, while for oblique orientations the mean energy is almost identically low. The natural scene's spectrum reveals peaks for horizontally, vertically, as well as for diagonally (45°) oriented structures. All in all, the difference between the peaks and the minima is not as big as for manmade images. The same holds for aerial images, except that there are no peaks for diagonal orientations. Instead the oblique orientations are approximately all at the same level.

Determing the parameters α and β of equation (2.4) leads to the values given in table 2.1. The results show that the EPS for natural and aerial images decay stronger than the EPS of manmade images, which is again due to the vertical and horizontal tails. Since β is largest for aerial images it is a hint that aerial images contain the most variance. This should be due to the special structure of aerial images being somehow more repetitive.

The computed values are in accordance with values reported in Ruderman (1994), van der Schaaf and van Hateren (1996) or Mumford and Gidas (2001), although the mean values of α reported there are all smaller than the values presented here. According

to Mumford and Gidas (2001) this might be due to that fact that these studies took extreme care of calibrating their vision systems and of correcting all the factors that can influence the shape of the spectrum. For the computations presented here, simply the JPEG–images as provided by the image databases were used. Since the JPEG–algorithm (Pennebaker and Mitchell, 1993) does a DCT and cuts off high frequencies, it probably influences the shape of the spectra. Anyway, since all the images in all classes have been stored as JPEG–images, these effects should not influence the comparability between the different image classes. It should only have an impact on the comparison to studies with images chosen more carefully.

2.1.3. Conclusions

From the results shown above one can conclude that the three different image classes do have different statistical properties. The images of the manmade class are characterized by horizontal and vertical orientations that are due to horizontal and vertical structures dominating manmade environments. The images of the natural class contain structures in different orientations with only little preference for vertical and horizontal orientations resulting from the horizon or objects following gravity like trees. The EPS for aerial images is almost isotropic because the images do not show structures that are aligned to any reference direction.

The findings are in accordance with the results presented in Oliva and Torralba (2001), Torralba and Oliva (2003) and Balboa and Gryzwacz (2003) showing strong differences depending on the environment thus making environment–specific adaptations of receptive fields for different environment types very likely. A description of these adaptation will be given in the following section.

2.2. Estimation of Receptive Fields

In the following section receptive fields of simple cells will be estimated for the three different image classes using ICA and a statistical analysis of the results will be done. Thereon, representative class–specific receptive fields will be derived.

2.2.1. Experiments

2.2.1.1. ICA of Grayscale Images

From the image database selected for the experiments described in section 2.1 a total of 63000 image patches per image class sized 20×20 pixels were selected from random positions. For preprocessing the data was transformed to a 150–dimensional eigenspace and was whitened.

The dimensionality of the eigenspace was been chosen comparably to related works like Hoyer and Hyvärinen (2000) or Hyvärinen and Hoyer (2000), and in order to cover approximately 98 % of the total information. The size of the image patches has been

chosen larger than in related works since the computational power has increased and the aerial images are characterized by longer contours.

The aerial image patches have additionally been scaled down before applying PCA by a factor of 0.5 leading to a resolution of $16\,\mathrm{m}^2$ per pixel. Combined with the work presented in Gerstmayr et al. (2004a,b) this leads to an image size of the blimp's on board camera of 50×50 pixels. In previous work image sizes of about 10×10 pixels have been proofen to be optimally for landmark selection tasks. Such a high resolution has been chosen since smaller image sizes would be too small for the filters needed for the contour detection and enhancement mechanisms. After downscaling the resulting image patches have been smoothed with a Gaussian smoothing filter with standard deviation $\sigma = \frac{2}{3}$ to make the proportional variance for 150 dimensions comparable to those of the natural and manmade class. The eigenimages for the three different image classes are shown in figure 2.12 at the end of this chapter.

ICA was computed using the MATLAB implementation of the FastICA–algorithm (see Hyvärinen (1999a) and section A.2.3) with the hyperbolic tangent as nonlinearity and for fine–tuning, a random initialization, and stabilization being enabled. The parameters not mentioned here remained unchanged. The resulting independent basis vectors are visualized in figure 2.13.

Since the resulting basis vectors resemble the receptive fields of cortical simple cells, Gabor functions have been fitted using a Least Squares Approximation. Therefore, the error function

$$\eta = \sum_{x=-9}^{10} \sum_{y=-9}^{10} \left(a_i(x,y) - G(x,y)\right)^2 \tag{2.5}$$

was minimized where $a_i(x,y)$ is the i–th basis vector, i.e. the i–th column of the mixing matrix A reshaped to an image sized 20×20 and $G(x,y)$ is the Gabor function as defined in equation (1.1) whose parameters are changed in order to optimally fit the basis vector. The optimization was done with regard to the following constraints

$$-100 \leq m_x \leq 100 \tag{2.6a}$$
$$-100 \leq m_y \leq 100 \tag{2.6b}$$
$$0 < \theta \leq \pi \tag{2.6c}$$
$$0 \leq s_1 \leq s_2 \leq 100 \tag{2.6d}$$
$$0 \leq F \leq 100 \tag{2.6e}$$
$$0 \leq K \leq 100 \tag{2.6f}$$
$$-\pi < \varphi \leq \pi. \tag{2.6g}$$

which have been chosen rather loosely to allow a good approximation even to Gabor functions with high spatial frequency and centered outside the image patch.

To compute the nonlinear regression an Evolution Strategy (ES, see A.6) has been used with $\sigma = 4$ independent subpopulations of $\mu = 10$ individuals producing $\lambda = 12$ offspring and $\rho = 3$ recombined individuals. The mutation was done using the HMB–scheme with intermediate recombination and a Comma–evolution strategy. The resulting parameters have been analyzed statistically, the results are shown in figures 2.4 to 2.8 and table 2.2.

2.2.1.2. ICA of Color Images

For an human observer the transitions from red roofs to dark gray streets form a very striking contrast in the aerial images. Thus, one would expect that color information plays a more important role in aerial images than it does in natural images. Therefore, ICA was computed following Hoyer and Hyvärinen (2000) for colored versions of the images used in section 2.1. Some of the images are shown in figure 2.11. Image regions sized 256×256 pixels were selected at the image center of the database images. From these regions a total of 63000 image patches sized 20×20 pixels were picked at random positions. The image patches have been described as a 1200–dimensional vector (400 pixels times 3 color channels). For aerial images the same additional preprocessing as described above for grayscale images was done. Again, a 150–dimensional eigenspace has been computed which covered about 98 % of the proportional variance contained in the data. The resulting eigenimages are shown in figure 2.14. After the transformation to the eigenspace the data was sphered and ICA was computed as described above. The independent basis vectors are visualized in figure 2.15. For the colored images no further computations and analyses were performed subsequent to the ICA.

2.2.2. Results

2.2.2.1. ICA of Grayscale Images

The eigenvectors shown in figure 2.12 for the three different image classes are not visually distinctable and the eigenvectors generally resemble 2D–Fourier bases. The first few eigenvectors consist of low frequency patterns. The higher eigenvectors, covering less variance, more and more resemble to wave patterns of higher spatial frequencies.

The ICA basis vectors visualized in figure 2.13 resemble the receptive fields of V1 simple cells as it was expected. When having a close look, it seems that more natural and manmade basis vectors resemble Gabor functions which are more elongated and have higher spatial frequency.

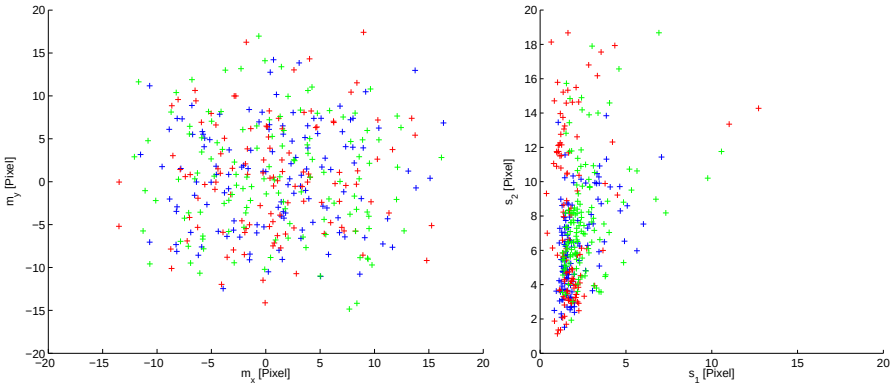
To better qualify these differences, Gabor functions have been fitted to the basis vectors. The mean fitting error was 0.2324 (0.1567), 0.2776 (0.1521), and 0.2099 (0.1619) with standard deviations given in parentheses for the natural, the manmade, and the aerial image class, respectively. From the analysis components have been excluded that could not have been fitted well, that were located close to the boarder of the basis vectors, and that had a great variance in either direction. Mathematically formulated only patches that hold the following constraints were included in the analysis:

$$\text{abs}\,(m_x) \leq 17.5 \tag{2.7a}$$
$$\text{abs}\,(m_y) \leq 17.5 \tag{2.7b}$$
$$s_x \leq 20.0 \tag{2.7c}$$
$$s_y \leq 20.0 \tag{2.7d}$$
$$\eta \leq 0.50. \tag{2.7e}$$

By excluding patches that did not hold the constraints a total of 25, 35, and 23 basis vectors have been discarded. The average fitting error of the remaining vectors were 0.2060 (0.1141), 0.2776 (0.1253), and 0.1742 (0.1214) for the natural, the manmade, and the aerial images, respectively, with standard deviations shown in parentheses.

For the remaining patches relative frequencies for the various parameters have been computed. The underlying distributions have been compared by computing significance levels using the Kolmogorov–Smirnov–Test (KST, see Press et al. (2003) and section A.4). The results are summarized in table 2.2.

Figure 2.4(a) shows the center positions (m_x, m_y) of the Gabor functions within the basis image. The center positions are not significantly different for the three different image classes and most center positions are in the area $[-10, 10] \times [-10, 10]$. Within this area there is no particular distribution recognizable.



(a) Center positions

(b) Variance in y–direction vs. variance in x–direction

Figure 2.4.: Scatter plots for center positions and standard deviations of fitted Gabor functions. Blue, red, and green crosses denote the natural, manmade and aerial image class, respectively.

A scatter plot of the variance s_2 perpendicular to the tuning direction θ versus the variance s_1 along tuning direction θ is given in plot 2.4(b). As one would expect from the constraints in equation (2.6) $s_1 < s_2$ holds for all analyzed Gabor functions. There is no striking difference between the different image classes, most of the values are in the area $[1, 3] \times [2, 9]$. Figure 2.5(a) shows the relative frequencies of the standard deviation s_1 along the tuning direction. The distributions for natural and manmade structures both have a peak at $s_1 \approx 1.2$ and are both relatively narrow. For aerial images the distributions is clearly more narrow with a maximum for $s_1 \approx 1.75$. Additionally, it differs highly significant from those of natural and manmade images. The standard deviation s_2 perpendicular to the tuning direction all differ significantly from each other. The distribution of the natural image class raises fast in $[0, 4]$ and then decreases slowly. For manmade scenes a maximum is reached for $s_2 = 3$ and for $s_2 \approx 12$, indicating that

Table 2.2.: Summary of significance levels for analyzing the fitted Gabor functions

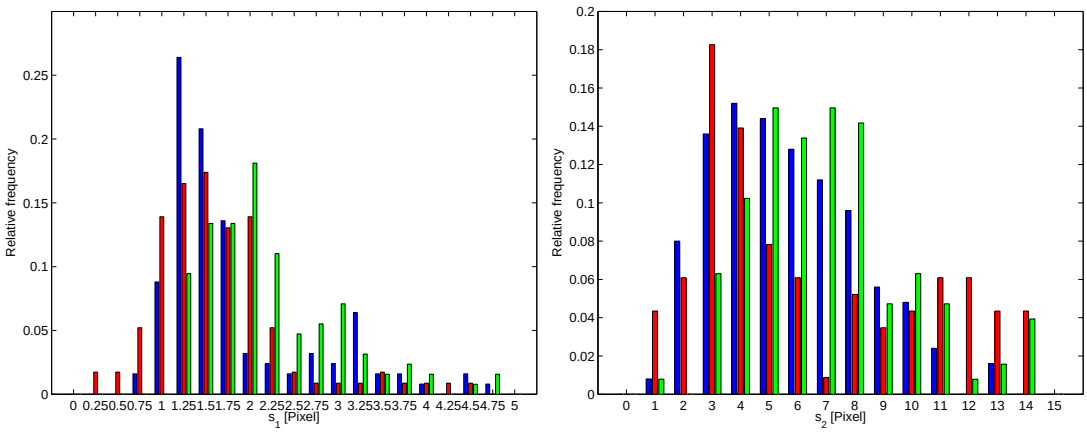
Parameter	Compared image classes	P	Significance
m_x	natural vs. manmade	0.8143	—
	natural vs. aerial	0.7363	—
	manmade vs. aerial	0.8030	—
m_y	natural vs. manmade	0.4795	—
	natural vs. aerial	0.8973	—
	manmade vs. aerial	0.7580	—
s_1	natural vs. manmade	0.1454	—
	natural vs. aerial	< 0.0001	$> 99.99\,\%$
	manmade vs. aerial	< 0.0001	$> 99.99\,\%$
s_2	natural vs. manmade	0.0001	$99.99\,\%$
	natural vs. aerial	0.0023	$99.77\,\%$
	manmade vs. aerial	0.0001	$99.99\,\%$
θ	natural vs. manmade	0.0622	$93.78\,\%$
	natural vs. aerial	0.6967	—
	manmade vs. aerial	0.0064	$99.36\,\%$
φ	natural vs. manmade	0.9649	—
	natural vs. aerial	0.9535	—
	manmade vs. aerial	0.5667	—
F	natural vs. manmade	0.0443	$95.57\,\%$
	natural vs. aerial	0.0010	$99.90\,\%$
	manmade vs. aerial	0.0022	$99.78\,\%$
n_1	natural vs. manmade	0.0269	$97.31\,\%$
	natural vs. aerial	0.0012	$99.88\,\%$
	manmade vs. aerial	0.0003	$99.97\,\%$
	natural vs. monkey	< 0.0001	$> 99.99\,\%$
	manmade vs. monkey	< 0.0001	$> 99.99\,\%$
	aerial vs. monkey	0.0334	$96.66\,\%$
n_2	natural vs. manmade	0.0499	$95.01\,\%$
	natural vs. aerial	0.0052	$99.48\,\%$
	manmade vs. aerial	< 0.0001	$> 99.99\,\%$
	natural vs. monkey	0.0024	$99.76\,\%$
	manmade vs. monkey	0.1905	—
	aerial vs. monkey	< 0.0001	$> 99.99\,\%$

there are receptive fields tuned for shorter and for very long contours. The distribution for aerial images increases moderately and has a wide, almost plateau–like peak for $[5, 8]$.

Figure 2.6(a) visualizes the distributions for the various orientations found in the fitted Gabor functions. It strikes out that the distribution for the natural and aerial image class are almost equally distributed but the distribution of the manmade class has strong preferences for horizontal ($\theta = 0$) and vertical ($\theta = \frac{1}{2}\pi$) orientations and minor preferences for diagonal directions. These results are strengthened by the computed KST: the distribution for manmade images differs significantly from those of the natural and aerial image class.

The distributions of the phase φ shown in figure 2.6(b) do not differ significantly. For all image classes most of the Gabor functions are antisymmetric with phase $\varphi = \pm\frac{1}{2}\pi$. The proportions of negative and positive antisymmetric Gabor functions are only approximately the same.

The spatial frequency parameter F visualized in figure 2.7(a) differs significantly. However, all distributions are bimodal showing maxima for $F \approx 0.15$ cycles per pixel and for smaller spatial frequencies. With a value of $F = 0.15$ cycles per pixel the antisymmetric Gabor functions have one on– and one off–subfield.

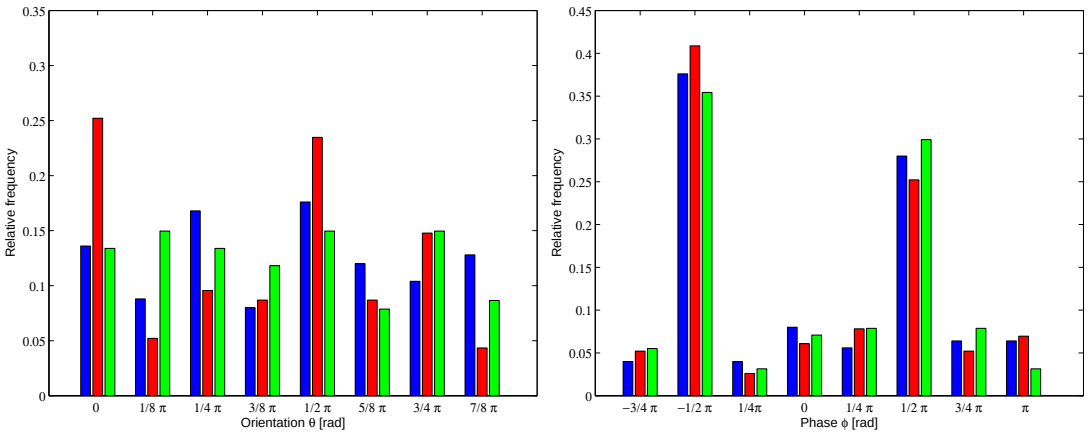


(a) Relative frequencies of s_1

(b) Relative frequencies of s_2

Figure 2.5.: Relative frequencies for standard deviations s_1 and s_2. Blue, red, and green bars denote the natural, manmade and aerial image class, respectively

Figure 2.7(b) is a scatter plot of n_2 versus n_1 for the three different image classes and data of electrophysiological studies on Macaque monkeys as presented in Ringach (2002). For an overview over the current state of electrophysiology in V1 see also Ringach (2004). In comparison to the receptive fields estimated by ICA the receptive fields of monkeys seem to be less elongated because the maximum for n_2–values is smaller for the monkey data than for the estimations. Since also the n_1–values for the monkey data are larger than the computed ones, the receptive fields of monkeys are also wider or have more lobes.



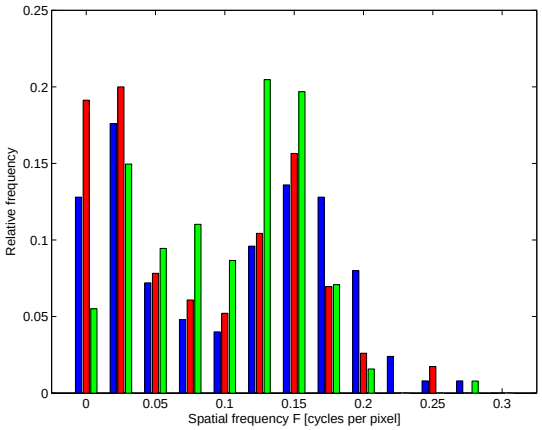
(a) Relative frequencies of orientation θ

(b) Relative frequencies of phase parameter φ

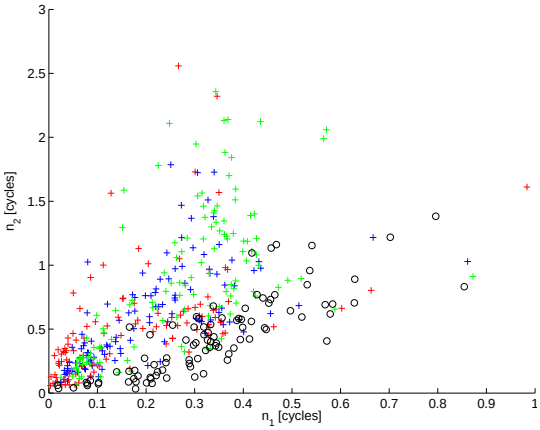
Figure 2.6.: Relative frequencies for different orientations θ and phase parameters φ. Blue, red, and green bars denote the natural, manmade and aerial image class, respectively

The relative frequencies of n_1 and n_2 values are visualized in figure 2.8(a) and 2.8(b). Both plots include again data provided by Ringach (2002). The distribution of n_1 values for the monkey data increases, reaching a maximum at 0.3 cycles, and then decreases again. Although showing peaks around 0.3, too, the distributions for the examined image classes differ from the monkey data as they show peaks for very small values of n_1. These are probably due to some very broadly tuned Gabor functions. Also the decrease for values greater than the maximum is steeper. All distributions show pairwise significances. The same holds for the n_2 values except for the comparison between manmade scenes and monkey data. For manmade images there is a slight increase in the range $[0, 0.5]$, while for natural and aerial images the frequencies are almost constant in the range $[0, 0.5]$ and $[0, 1]$, respectively.

Using these results, class–specific receptive fields have been derived from the distributions shown above. They are meant to be average or representative receptive fields of the corresponding image class. The parameters were manually adapted to the several parameter distributions by finding appropriate tradeoffs. The class–specific Gabor functions are visualized in figure 2.9 and the corresponding parameters are summarized in table 2.3. As for the manmade image class a clear preference of horizontal and vertical structures is identifiable, the number of orientations $|\theta|$ to be used for the simple cell model described in section 3.1 was chosen to be 4. This choice was based on the assumption that due to the dominance for horizontal and vertical structures less oriented filters are needed to filter the images. The orientations are equally distributed in $|\theta| + 1$ steps over $[0, \pi]$. The exact number of orientations needed for an isotropic edge detector depends on the width of the simple cell's tuning curve (see 3.1.3). The width strongly

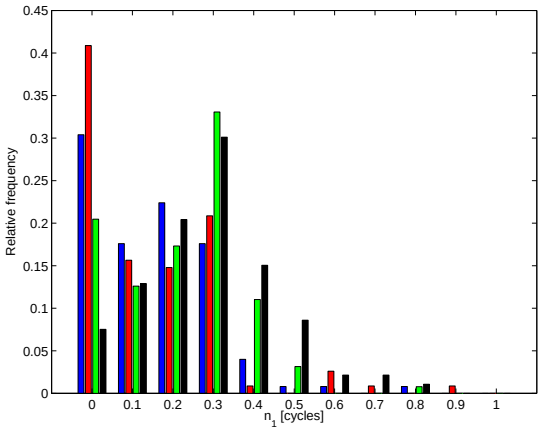


(a) Relative frequencies of spatial frequency F. Blue, red, and green bars denote the natural, manmade and aerial image class, respectively.

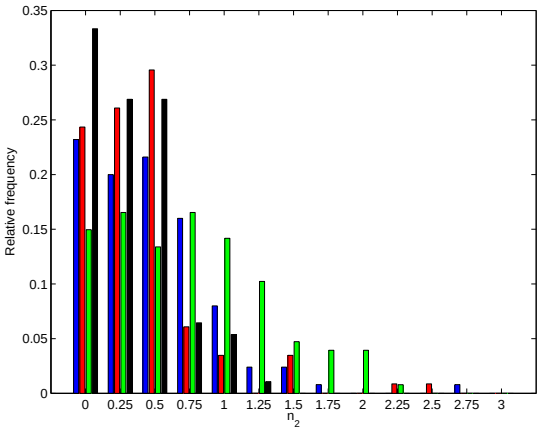


(b) Scatter plot for n_2 versus n_1. Blue, red, and green crosses denote the natural, manmade and aerial image class, respectively. Black circles denote data of Macaque monkey as provided by Ringach (2002).

Figure 2.7.: Relative frequencies of spatial frequency F and scatter plot for n–values



(a) Relative frequencies of $n_1 = Fs_1$



(b) Relative frequencies of $n_2 = Fs_2$

Figure 2.8.: Relative frequencies for n_1 and n_2. Blue, red, and green bars denote the natural, manmade, and aerial image class, respectively. Black bars denote data of Macaque monkey as provided by Ringach (2002)

depends on the used simple cell model. Thus, the choice here is somehow arbitrarily.

Table 2.3.: Overview over parameters for the class–specific receptive fields

Parameter	Natural	Manmade	Aerial
(m_x, m_y)	$(0, 0)$	$(0, 0)$	$(0, 0)$
s_1	1.25	1.50	1.75
s_2	5.00	4.50	8.00
φ	$\frac{1}{2}\pi$	$\frac{1}{2}\pi$	$\frac{1}{2}\pi$
$\lvert\theta\rvert$	8	4	8
F	0.15	0.15	0.15
K	1.00	1.00	1.00
n_1	0.188	0.225	0.263
n_2	0.750	0.675	1.200

(a) Natural class (b) Manmade class (c) Aerial class

Figure 2.9.: Class–specific receptive fields for several image classes

2.2.2.2. ICA on Color Images

Most of the eigenvectors for the PCA of color images (see figure 2.14) are achromatic. The resulting patterns that are similar to those of grayscale images and again resemble 2D–Fourier basis. The smaller fraction of the eigenvectors contain color information. Unlike for the gray–scale eigenvectors, the first eigenvector of all image classes does not represent a circle with a diameter of approximately the patch width, but is rather a completely black image patch.

For the natural image class the second eigenvector is a completely lightblue patch, the fifth eigenvector is completely purple. The other chromatic eigenvectors mainly code two different types of color opponencies: The first type codes lightblue–orange transitions, and the second one represents green–purple opponencies. The chromatic regions are not uniform but also show some sort of wave pattern. The eigenvectors for the manmade image class are comparable to those of the natural class. Again, the second eigenvector

is a completely lightblue patch, the tenth eigenvector represents again a purple patch. The other chromatic eigenvectors code lightblue–orange and green–purple transitions.

The chromatic eigenvectors for the aerial images are different from those of the other classes. The first 18 eigenvectors are achromatic and there are no completely uniform patches except the first eigenvector. Except for the fortieth eigenvector, which shows a green–yellow transition, all chromatic eigenvectors represent red–cyan opponencies.

Comparing the obtained ICA basis vectors reveals that the vectors for aerial images are beside one exception all achromatic. For the other image classes there are more achromatic than chromatic patterns. The achromatic patches resemble various Gabor functions. The chromatic patches can be divided into few uniform patches, a majority of horizontal or vertical color transitions, and few more complex transitions. The transitions include red–green, purple–green, blue–yellow, blue–orange, and blue–yellow–orange for natural scenes and red–green, red–blue, and blue–orange for manmade structures.

2.2.3. Discussion

2.2.3.1. ICA on Grayscale Images

Discussion of the Results

Both, the computed eigenvectors (figure 2.12) as well as the independent basis vectors (figure 2.13) are in good accordance with the related literature like Olshausen and Field (1996), Hoyer and Hyvärinen (2000), and many others.

Gabor function were fitted to the basis vectors reasonable well allowing a detailed analysis of the differences between the image classes. The results revealed that the fitted Gabor functions for natural and manmade structures are not as elongated as those for aerial images. Further the functions for manmade images were oriented mainly horizontally and vertically, while for natural and aerial images the orientations were equally distributed. Thus, the receptive fields reflect the statistical properties of the images.

The comparison of the independent basis vectors or the independent component filters for different environments has received only little attention so far. The work which is closest to this thesis is Zhang and Mei (2003). There, the experiments reported in Blakemore and Cooper (1970) were simulated and ICA for natural and aerial images was computed. In the paper neither the type of the aerial images is specified closer nor the distributions of the fitted Gabor functions are shown. However, the authors claim that the ICA filters for the aerial images do have more horizontal and vertical orientations than those of natural images. It is likely that these findings are due to the characteristics of the used training images, which might contain areas with dominating perpendicular structures possibly aligned along a common reference direction. Further, the authors do not mention any comparison on the elongation of the subfields.

Another related study is Zieghaus and Lang (2003) in which the authors computed ICA using several neural network based algorithms for a set of natural and urban images and fitted Gabor– and Haar–wavelets, respectively. However, the paper focuses on the comparison between different ICA algorithms and not on the comparison of different

image classes. Therefore, and because important information and data is missing in the paper a more substantiated comparison to the work presented here is not possible.

For the sake of completeness two other works exploiting the difference of ICA filters or basis vectors for different input data shall be mentioned: the first is H. Le Borgne (2001) in which a classification method based on the difference of ICA filters is proposed. The second work is Lee et al. (2000) or Lee and Lewicki (2002). As by product of their actual objective, an ICA mixture model, the authors report different ICA filters for different image classes: printed text yields ICA filters representing bars of different length and width capturing high–frequency patterns, whereas natural images yield Gabor–like ICA filters.

So far none of the related work has shown the actual distribution of the parameters for the fitted Gabor functions, or has derived something comparable to the class–specific receptive fields. Further on results from image statistics have not been used in a biologically motivated computer vision approach.

Discussion of the Methods

Deriving the parameters for the class–specific receptive fields as described in section 2.2.2 is the simplest approach possible. However, each of the parameters was assumed to be independent from all others which is not the case for "real" receptive fields as argued in Ringach (2002): every simple cell's receptive field is describable by a Gabor function, but not every Gabor function describes a receptive field. A more exact estimation, e.g. by means of Bayesian statistics, would have been difficult to realize since much more information about the distributions and their dependencies would have been needed. Another important question is whether the estimates could be more reliable if the basis vectors would not contribute equally but would be weighted according to their importance for the overall coding. It sounds reasonable that due to the whitening step unimportant basis vectors can bias the estimation of the distributions. However, such a weighting would be difficult to realize since ICA does not allow to order the basis vectors with respect to their importance. On the other hand, in a sparse–dispersed coding each neuron should contribute equally to the overall coding. Thus, it is justified that each basis vector also contributes equally to the estimation of the distributions.

Another important aspect to discuss is the question whether the receptive fields should be compared to the basis vectors or to the independent component filters. For this work receptive fields have been compared to independent component basis vectors, although many related studies like Bell and Sejnowski (1997), van Hateren and van der Schaaf (1998), or Zhang and Mei (2003) argued that the receptive fields should be compared to the independent component filters and not to the basis vectors.

The reason why receptive fields should be compared to ICA filters is that it is the separating matrix $\mathbf{W}$ transforming an observation or sensory input $\mathbf{x}$ to an independent signal $\mathbf{s}$ (see equation (A.22)). However, this work follows the argumentation of Hoyer and Hyvärinen (2000) and Hyvärinen and Hoyer (2001) comparing receptive fields to the independent component basis vectors arguing that each of the basis vectors $\mathbf{a}_i$ forms some sort of "optimal stimulus" giving a non–zero response if and only if the input stimulus equals $\mathbf{a}_i$. Additionally, it is argued that the argumentation is stressed if inhibition

mechanisms like gain control (Ringach, 2004) are taken into account that suppress neural responses if a large number of neurons are simultaneously activated. Additionally, that the visualization for the filters is not as straightforward. Caywood et al. (2004) mention that for grayscale images there are no striking differences between ICA filters and basis vectors. In Hyvärinen and Hoyer (2001) it has been shown that the basis vectors $\mathbf{a}_i$ are a low–pass filtered version of the filters $\mathbf{w}_i$ having the same orientation, location and frequency tuning properties.

ICA as Model for V1

Another important point is whether a linear transformation is sufficient to explain the properties of cortical cells or not. It has been pointed out in van Hateren and van der Schaaf (1998) or Zetsche and Röhrbein (2001) that independent components are not completely independent, but only as independent as possible by a linear transformation. ICA got an established method for investigating information theoretical questions and it can explain neural coding in V1 reasonably well. Though it neglects nonlinearities, temporal aspects, and various adaptation mechanisms. After Hoyer and Hyvärinen (2000) this might be due to the fact that nonlinearities, such as rectification and shunting interactions, involved in V1 can be thought of as operating on top of the linear representation. However, in Ringach (2004) it is argued that these nonlinearities should be taken into account more carefully because they do influence the neural information processing.

Closely related to that argument is the question whether ICA is an appropriate method to estimate receptive fields. Although it is by the time used for almost 10 years to model information theoretic aspects in V1, it is still a sophisticated method. van Hateren and van der Schaaf (1998) describe a detailed comparison between receptive fields estimated by ICA and receptive fields as measured by electrophysiological studies like Jones and Palmer (1987a,b,c) or DeAngelis et al. (1993a,b). The authors of van Hateren and van der Schaaf (1998) conclude that ICA gives qualitative predications about receptive fields which can be compared reasonably well with neurophysiological data. The only notable exception mentioned is the spatial frequency of the estimated receptive fields. Compared to electrophysiological data ICA filters show less variability. Following the authors this is due to the fact that ICA works at a scale which is close to the sampling grid of the images. Based on this property, the authors also report that ICA leads to receptive fields aligned with the sampling grid of the images. In their study they proofed by rotating the scenes that the orientation preference does not result from horizontal and vertical structures in the training images. For the natural and the aerial image class no such preference was found in the described experiments of this work. Therefore, it is likely that the preferences for horizontal and vertical orientations found in manmade scenes are due to the statistical properties of the environment, and not to the sampling grid.

In Ringach (2002) the data estimated by ICA in Olshausen (2001) is compared to receptive fields measured by reverse correlation methods of Macaque monkeys. One result of the paper is that the phase φ for the monkey data clusters nicely into two clusters corresponding to even– and odd–symmetric receptive fields. However, ICA seems to prefer antisymmetric receptive fields as most of the receptive fields described in

Olshausen (2001) or this work are antisymmetric. Since symmetric receptive fields could be separated into at least two antisymmetric receptive fields one could argue that ICA reduces the redundancy contained in these fields. Like for the data computed in this work, the data of Olshausen (2001) does not match the n–values reported for monkeys in Ringach (2002). While there is mentioned that the estimated data contains more subfields than the monkey's receptive fields, i.e. that n_1 is larger for the estimated receptive fields, the receptive fields for all image classes estimated within this work are much more narrow, i.e. n_1 is smaller, than those of monkeys. This discrepancy might be due to the different ICA algorithms used in Olshausen (2001) and this work. Additionally, the receptive fields of all image classes computed within this work are more elongated than the ones found in monkeys. The comparison between Olshausen (2001) and the measured receptive fields does not reveal any difference. To summarize this part of the discussion one can state that ICA does not reproduce all properties of simple cell's receptive fields exactly, but it estimates most properties reasonable well and it is the most appropriate linear method available.

The last point to discuss is why a standard ICA–approach has been used for this work and not a biologically more plausible extension. Such extensions include Hyvärinen et al. (2001a, 2003b) or a generalized ICA approach like the one suggested in Utsugi (2002). Since none of these models can cover all aspects of the visual cortex and since these extensions have not been proven to give more plausible receptive fields, a standard ICA approach has been used. It is an interesting question for future work to compare the shapes of the receptive fields computed by the FastICA algorithm and several extensions.

2.2.3.2. ICA of Color Images

In the last years ICA on color images has received quite some attention, but various studies came up with different results. Computing ICA on color images is according to Caywood et al. (2004) quite sensitive to the used images, to the preprocessing, and to the ICA–algorithm. In Wachtler et al. (2001) and Doi et al. (2003) LMS input images, i.e. input images coded with spectral sensitivities of the human L–, M–, and S–cones, have been used. Following Caywood et al. (2004) their results are quite suitable to explain certain aspects of neural color information processing in V1 like the emergence of chromatic and achromatic processing units, color opponencies, as well as their distributions. In contrary, studies like Tailor et al. (2000) or Hoyer and Hyvärinen (2000) used RGB–coded images. They argue that the difference between LMS– and RGB–coding is not essential and that their studies also could explain visual information processing. However, Caywood et al. (2004) pointed out that the results of these studies are not comparable to those obtained by the studies mentioned above. Therefore, the results presented here are only compared to the latter studies which used RGB–images. Since the overall goal of this work was not to explain color information processing in the brain, but rather to compare different image classes with different color information, it is not important to use LMS–coding.

As there are no other related studies that computed ICA on aerial color images or any other non–natural image class the results for natural scenes will be discussed first.

The results presented here are not fully in accordance with those of related work. With regard to the eigenvectors of natural images, studies like Buchsbaum and Gottschalk (1983), Ruderman et al. (1998), or Hoyer and Hyvärinen (2000) report red–green and blue–yellow opponencies instead of green–purple and lightblue–orange opponencies found in the experiments described in this thesis. The relation of chromatic and achromatic eigenvectors seems to be approximately the same as in the reviewed literature. An exact comparison to related studies is difficult since the proportion of chromatic eigenvectors obtained is dependent on the number of used eigenvectors. This is due to the fact that most of the variance in the images is covered by achromatic eigenvectors and chromatic eigenvectors covering little variance are discarded by reducing the dimensionality.

The results for the ICA differ significantly from those shown in Tailor et al. (2000). There, a much larger proportion of chromacy is mentioned. In contrary to the work presented here, Tailor et al. (2000) used all possible dimensions for the ICA step such that the complete information contained in the images was taken into account. The basis vectors shown in Hoyer and Hyvärinen (2000) are quite comparable to the results shown above, except for the occurring opponencies.

In Caywood et al. (2004) a disagreement is mentioned whether the independent component filters (Tailor et al., 2000; Doi et al., 2003) or the independent basis vectors (Hoyer and Hyvärinen, 2000; Wachtler et al., 2001) should be used. Since the differences between basis functions and filters does not have an impact on the comparison between the image classes and since the rest of the work focuses on grayscale images, here the independent basis vectors are shown, although Caywood et al. (2004) argue that for color images there are striking differences between filters and basis vectors.

In comparison to manmade and natural images on the one hand side and aerial images on the other, the most striking difference is that the first 18 eigenvectors of aerial image patches and almost all independent basis vectors are achromatic. Further the emergence of the red–cyan opponency is worth discussing. PCA and ICA are both unsupervised learning methods finding the directions of the greatest variance or the directions corresponding to maximally independent signals. One could argue that transitions from red roofs to other elements in the aerial image do play an important role and therefore influence the statistical properties of the images. PCA and ICA discover these properties leading to results differing from other images. However, this argumentation is somehow restricted as the results for natural scenes are not fully compliant with results of other works. In further work it is necessary to explain why the occurring color opponencies were found in natural and manmade images. Since several trials (with the same input images) always let to the same results and since the computation is rather straightforward, an error in the used programs seems to be unlikely but cannot be fully ruled out. A good starting point for further investigations would be to recompute the analysis using training images that have also been used for related studies or to analyze the color distributions of the used training sets. This could reveal if the findings are due to the particular color statistics of the training set.

2.2.4. Conclusions

2.2.4.1. ICA on Grayscale Images

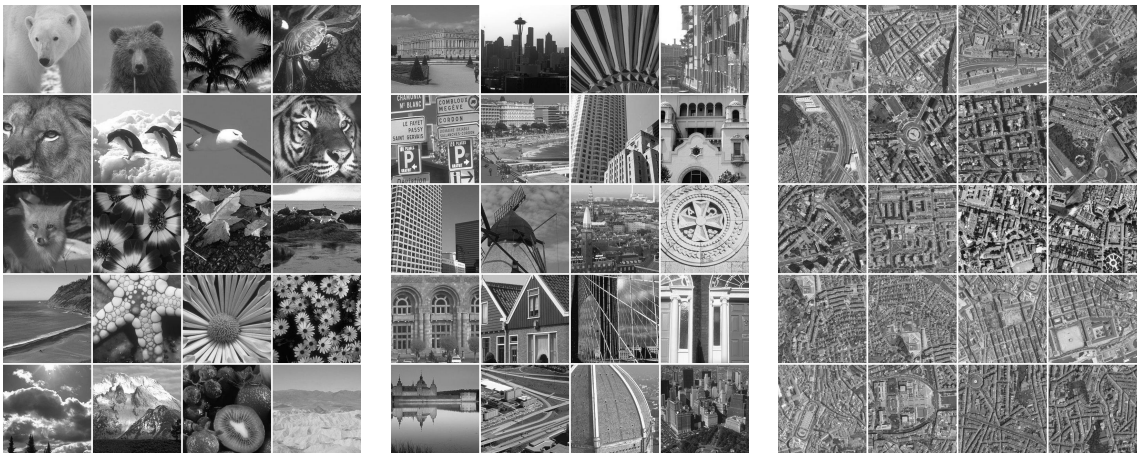
The main conclusion one can draw from the presented results is that statistical properties of the environment are reflected in the shape of receptive fields. This substantiates the theory that sensory systems are optimally tuned to the kind of stimuli occurring most frequently. By comparing the parameter distributions of the fitted Gabor functions it showed out that the images contained in the natural, manmade, and aerial image class are characterized by short contours in various directions, short contours in mainly horizontal and vertical directions, and long contours in various directions, respectively. From this distributions class–specific receptive fields were derived that are optimally tuned to the statistical properties of the environment. They will be used in the chapter 3 of this thesis for contour and junction detection.

2.2.4.2. ICA on Color Images

From the results for the analysis of color images one can conclude that color plays a less important role in aerial images because almost all independent basis vectors and the first 18 eigenvectors are achromatic. The results for the natural image class are not fully accordant with the results of other studies, especially not with Hoyer and Hyvärinen (2000) which is the closest of all reviewed works. Thus, it is difficult to draw further conclusions. The emergence of a red–cyan opponency might be a hint that transitions from red structures in aerial images (i.e. roofs) to other structures (e.g. streets or gardens) play an important role and influence the statistical properties of aerial color images. However, to fully proof or at least to strengthen this argumentation, the results for natural images should be more comparable to those of other studies. Since the results obtained for grayscale images were more promising, the open questions concerning color image statistics have been left for future work.

2.3. Chapter Summary

The main conclusion of this chapter is that aerial images differ in their statistical properties from natural and manmade scenes. This was shown by analyzing EPS in section 2.1 and by computing ICA for a large set of training images of three different image classes, namely natural, manmade and aerial images, in section 2.2. From the results of this step one can also conclude that the statistical properties of the environment are reflected in the shape of the optimally tuned receptive fields. These findings strengthen the hypothesis that any perceptual system is tuned to its environment. They also motivate to use the derived class–specific receptive fields in conjunction with a model of contour detection in the human visual system.

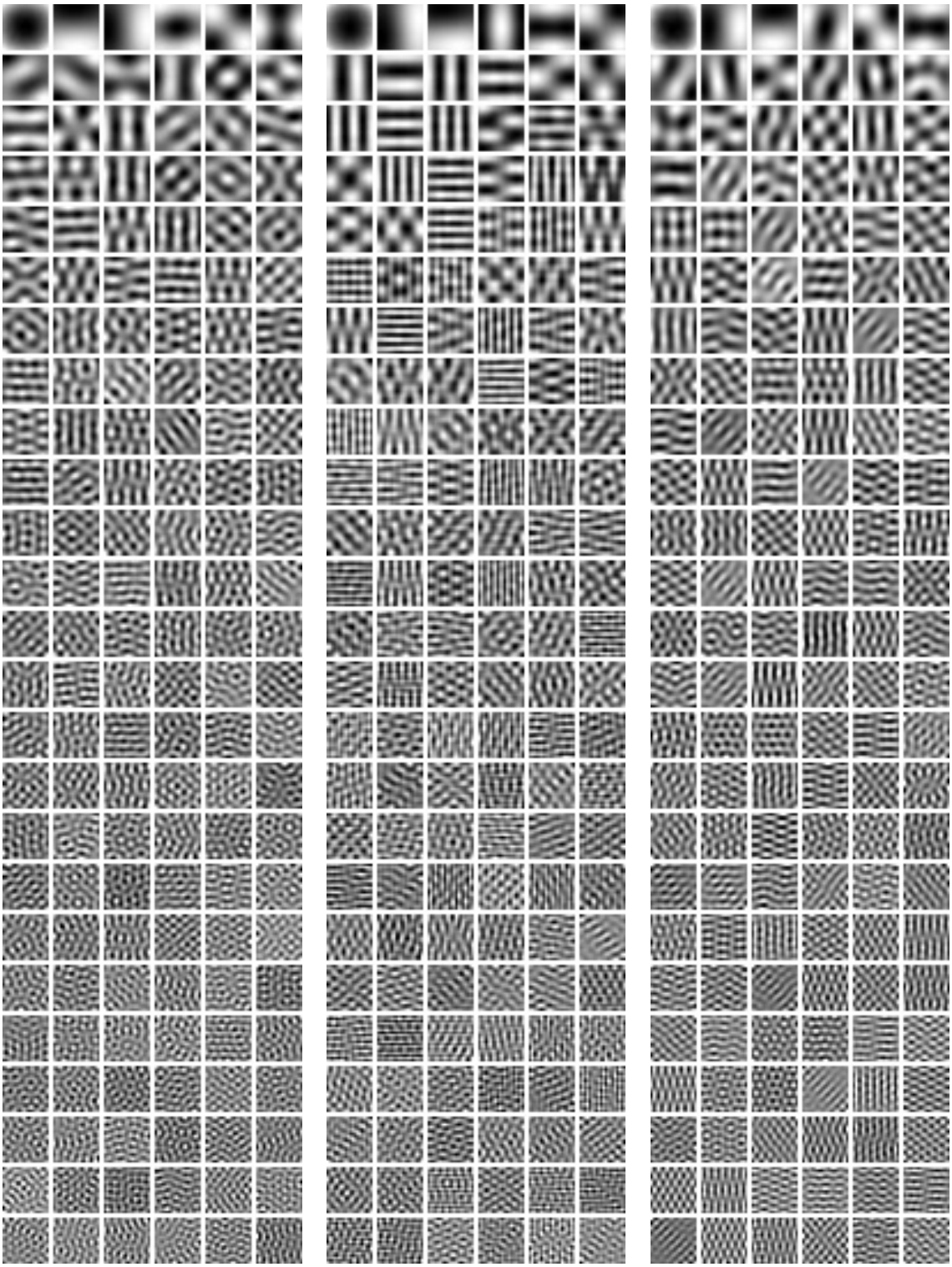


(a) “Natural” images (b) “Manmade” images (c) Aerial images

Figure 2.10.: Representative grayscale images of the three image classes

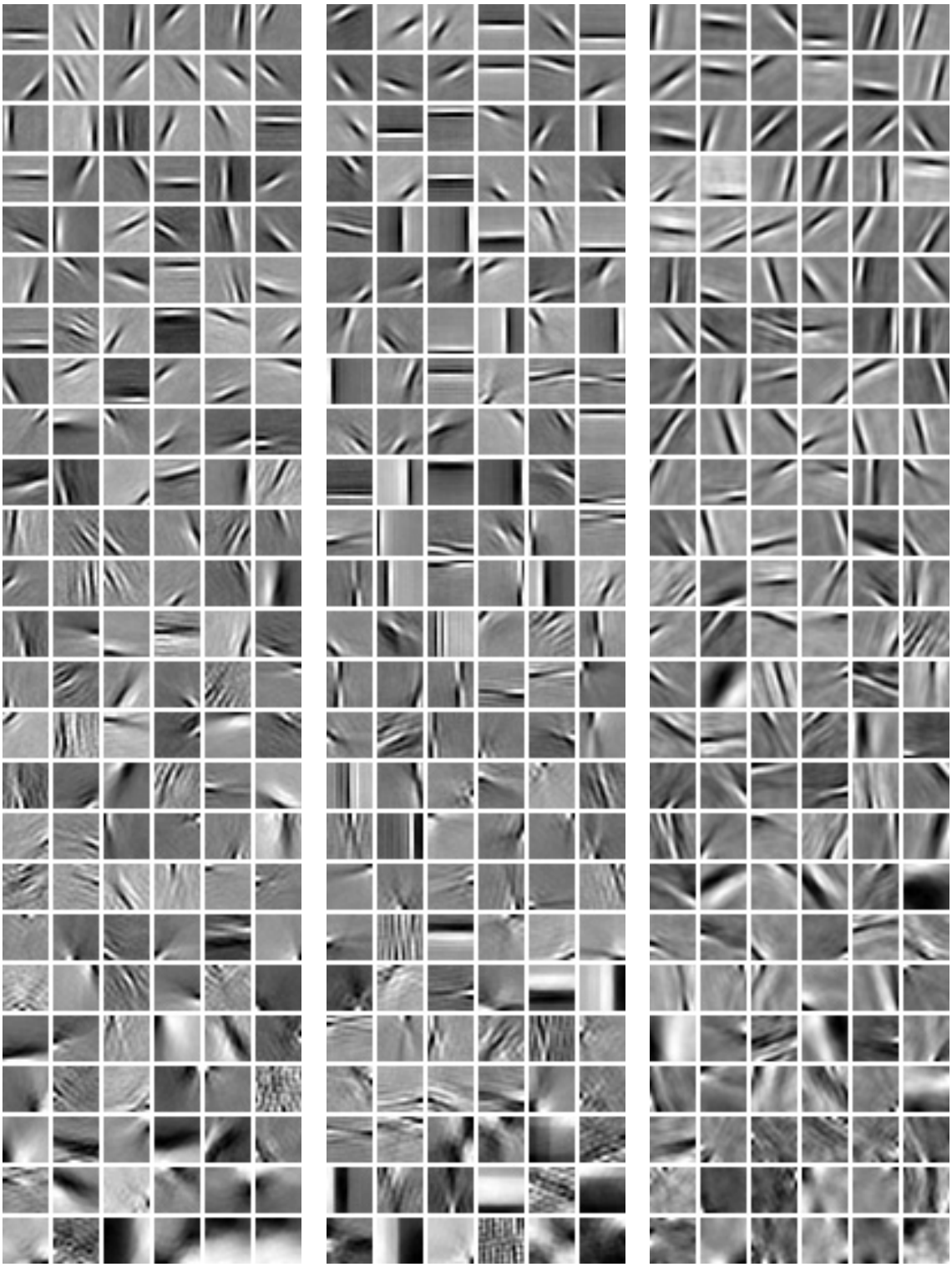
(a) “Natural” images (b) “Manmade” images (c) Aerial images

Figure 2.11.: Representative color images of the three image classes



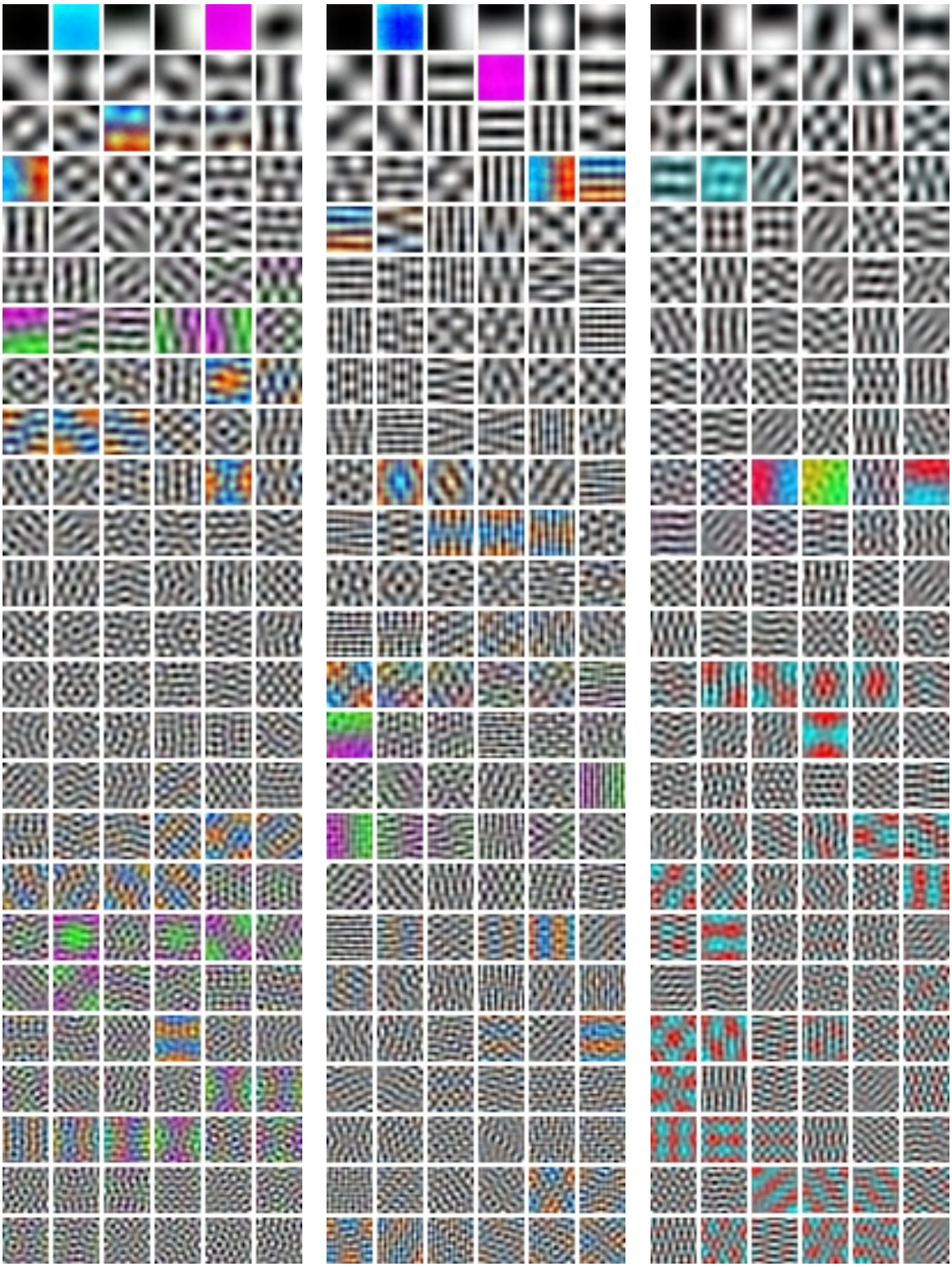
(a) “Natural” images (b) “Manmade” images (c) Aerial images

Figure 2.12.: The first 150 eigenvectors for different classes of grayscale images



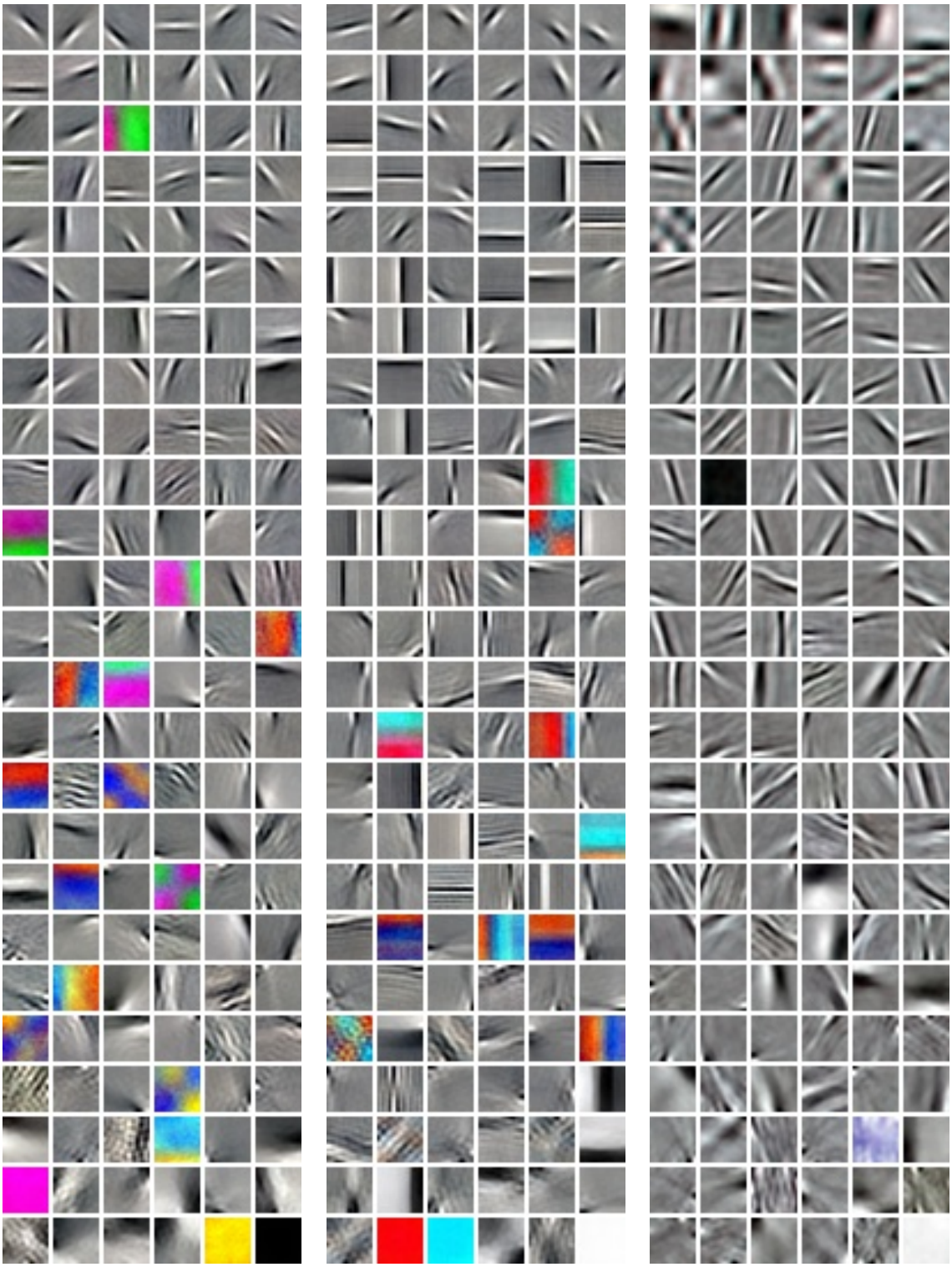
(a) “Natural” images (b) “Manmade” images (c) Aerial images

Figure 2.13.: ICA basis vectors for different classes of grayscale images



(a) “Natural” images (b) “Manmade” images (c) Aerial images

Figure 2.14.: The first 150 eigenvectors for different classes of color images



(a) “Natural” images (b) “Manmade” images (c) Aerial images

Figure 2.15.: ICA basis vectors for different classes of color images

3. Biologically Motivated Image Preprocessing

3.1. Contrast Processing

This section explains the use of class–specific receptive fields derived in section 2.2 in conjunction with the simple cell model proposed in Hansen and Neumann (2004b) to implement a biologically motivated edge detector. Before describing the model and the obtained results, an overview over related work will be given.

3.1.1. Related Work on Simple Cell Models

The following review is neither very detailed nor complete, as most of the proposed models are computationally not tractable for computer vision applications because they use compartmental–modeling or spiking–neurons (Dayan and Abbott, 2001) to explain results from electrophysiology or often model only one hypercolumn. More detailed reviews of such models can be found in Carandini et al. (1999), Hansen (2002), and Ursino and Cara (2004). According to Hansen (2002) three different types of models can be distinguished:

Pure Feedforward Models

These models are just implementations of the model described by Hubel and Wiesel (1962). However, due to the insight gained in neuroscience pure feedforward models are no longer of interest for modeling V1 because they are not sufficient to exactly describe the visual system.

Feedforward Models with Inhibition

This kind of model additionally takes inhibition into account. Inhibition is usually non–linear and often implemented as divisive normalization, i.e. the activation of a neuron is inhibited by the total activation of neurons tuned to similar and contrary spatial frequencies or orientations. Neurons contributing to the inhibition thus form a suppressive field (Carandini, 2004) or a non–classical receptive field Grigourescu et al. (2003). As nonlinearities play an important role in V1 (DeAngelis and Anzai, 2003) this kind of models have received a lot of interest in recent years by works like Carandini et al. (1997) or Tolhurst and Heeger (1997a,b). The models proposed in Neumann et al. (1999) and Hansen and Neumann (2004b), on which the model described in this section builds, are also feedforward models with inhibition. Another example from the computer vision domain includes Grigourescu et al. (2003). There, an extended energy operator for

pooling simple cell responses is proposed that includes an inhibition of neighboring edge elements which are are not in the preferred direction or which are outside the receptive field of the Gabor filters.

Recurrent Networks

These models also include feedback from the same cortical areas (short–range interactions) or from other cortical areas (long–range interactions) that can iteratively enhance edges. Therefore, they take into account how this mechanisms sharpens orientation tuning and influences contrast invariance. Examples from electrophysiology include Somers et al. (1995) and Carandini and Ringach (1997). Examples from computer vision include Neumann and Sepp (1999), Kolesnik et al. (2002), and Hansen and Neumann (2004a), which model the long–range integration between V1 and V2. Recurrent simple cell networks are also used by models for texture or color segmentation. There, they are used as preprocessing in order to detect different regions separated by edges. Such models include Li (1998, 1999) and Kokkinos et al. (2004).

3.1.2. Proposed Simple Cell Model

The used model is based on Hansen and Neumann (2004b) or Hansen (2002) except for the contrast detection filters. Here, the class–specific receptive fields estimated in section 2.2 are used. The model follows the general structure of the visual pathway as depicted in section 1.2.1.

3.1.2.1. Center Surround Stage

The LGN responses are modeled as rectified and non–linearly filtered signals of the input image. The input image $\mathbf{I}$ with luminance values in the range $[0, 1]$ is first processed by a center–surround mechanism similar to LGN cells. The center stream $\mathbf{I}_c$ and the surround stream $\mathbf{I}_s$ are obtained by filtering $\mathbf{I}$ with isotropic Gaussians of standard deviation σ_c and σ_s, respectively:

$$\mathbf{I}_c = \mathbf{I} * \mathbf{G}_{\sigma_c} \tag{3.1}$$

$$\mathbf{I}_s = \mathbf{I} * \mathbf{G}_{\sigma_s}. \tag{3.2}$$

For the implementation the size of the mask for the Gaussians was determined by $6\sigma + 1$. The choice of the model parameters is described in detail in section 3.1.2.4. These responses form the input for a shunting interaction, which is a sort of divisive inhibition yielding a bounded activity following the Weber–Fechner law. The change of activation $u_{i,j}$ for a pixel $\mathbf{I}(i, j)$ over time t can be written as

$$\frac{\partial u_{i,j}}{\partial t} = -\alpha_{\mathrm{LGN}} u_{i,j} + (\beta_{\mathrm{LGN}} - u_{i,j})\mathrm{net}^{+}_{i,j} - (\gamma_{\mathrm{LGN}} + u_{i,j})\mathrm{net}^{-}_{i,j} \tag{3.3}$$

where $\text{net}^+_{i,j}$ and $\text{net}^-_{i,j}$ are the excitation and inhibition of the neuron, and α_{LGN} denotes the activity decay rate, which is according to Kokkinos et al. (2004) related to the leakage conductance of the neuron. The lower and upper bound of the activity is influenced by γ_{LGN} and β_{LGN}. For a more detailed discussion on this type of shunting interaction see Neumann (1996) and Kokkinos et al. (2004).

Equation 3.3 is assumed to quickly reach steady state and is solved at equilibrium, i.e. for $\frac{\partial u_{i,j}}{\partial t} = 0$. For that case the above equation can be rewritten as a function of $\text{net}^+_{i,j}$ and $\text{net}^-_{i,j}$:

$$u_{i,j}(\text{net}^+_{i,j}, \text{net}^-_{i,j}) = \frac{\beta_{\text{LGN}}\text{net}^+_{i,j} - \gamma_{\text{LGN}}\text{net}^-_{i,j}}{\alpha_{\text{LGN}} + \text{net}^+_{i,j} + \text{net}^-_{i,j}} \tag{3.4}$$

The shunting interaction is modeled both for the on–center and the off–center domain. For the first, excitatory input is provided by $\mathbf{I}_c$ and inhibitory input is provided by $\mathbf{I}_s$. For the off–center domain the reverse holds true:

$$\mathbf{X}_{\text{on}} = U(\mathbf{I}_c, \mathbf{I}_s) \tag{3.5}$$
$$\mathbf{X}_{\text{off}} = U(\mathbf{I}_s, \mathbf{I}_c). \tag{3.6}$$

The LGN responses $\mathbf{K}_{\text{on}}$ and $\mathbf{K}_{\text{off}}$ result from mutually inhibiting the opposite domains and rectification

$$\mathbf{K}_{\text{on}} = [\mathbf{X}_{\text{on}} - \mathbf{X}_{\text{off}}]^+ \tag{3.7}$$
$$\mathbf{K}_{\text{off}} = [\mathbf{X}_{\text{off}} - \mathbf{X}_{\text{on}}]^+ \tag{3.8}$$

The contrast signals form the input for the next processing stage, the simple cell stage. Models using opponent inhibition are often referred to as "push–pull" models. Opponent inhibition is also an elementary building block for the simple cell stages.

3.1.2.2. Simple Cell Stage

The simple cells are modeled by Gabor functions and not as proposed in Hansen and Neumann (2004b) by elongated weighting functions consisting of five isotropic Gaussians aligned along the axis of preferred orientation. The advantage of this receptive fields is a better adaptation to edge detection, whereas Gabor filters are a tradeoff between edge detectors and spatial frequency detectors. One can argue that Gabor functions more related to the shape of receptive fields found in V1. Therefore, and because the results of section 2.2 shall be used to tune the simple cell model to the statistical properties of the input images, Gabor function have been used for this work.

In order to detect light/dark transitions a Gabor kernel $\mathbf{G}^{\text{ld}}_\theta$ with preferred orientation θ as given in equation (1.1) is used. The Gabor kernel detecting dark/light transitions is given by the negative of the light/dark kernel: $\mathbf{G}^{\text{dl}}_\theta = -\mathbf{G}^{\text{ld}}_\theta$. Since the equations for the light/dark and the dark/light case are analogous the indices are left out for the simple cell stage.

The receptive fields have been modeled based on Gabor functions as described in Troyer et al. (2001). Positive values of the Gabor function give the connection strength for on–subfields $\mathbf{G}^{\theta}_{\text{on}}$, negative regions are used to model the off–subfields $\mathbf{G}^{\theta}_{\text{off}}$:

$$\mathbf{G}^{\theta}_{\text{on}} = \left[\mathbf{G}^{\theta}\right]^{+} \tag{3.9}$$

$$\mathbf{G}^{\theta}_{\text{off}} = \text{abs}\left(\left[\mathbf{G}^{\theta}\right]^{-}\right). \tag{3.10}$$

The input activation of the on– and off–subfields, $\mathbf{R}_{\text{on}}$ and $\mathbf{R}_{\text{off}}$, is modeled as weighted difference of the contrast signals $\mathbf{K}_{\text{on}}$ and $\mathbf{K}_{\text{off}}$ convolved with the subfield masks $\mathbf{G}^{\theta}_{\text{on}}$ and $\mathbf{G}^{\theta}_{\text{off}}$:

$$\mathbf{R}^{\theta}_{\text{on}} = \left[(\mathbf{K}_{\text{on}} - \xi\mathbf{K}_{\text{off}}) * \mathbf{G}^{\theta}_{\text{on}}\right]^{+} \tag{3.11}$$

$$\mathbf{R}^{\theta}_{\text{off}} = \left[(\mathbf{K}_{\text{off}} - \xi\mathbf{K}_{\text{on}}) * \mathbf{G}^{\theta}_{\text{off}}\right]^{+}. \tag{3.12}$$

From that processing stage on all responses have to be computed for several discrete orientations θ. For convenience the index for the orientations is left out as all the following stages can be computed independently from each other.

For the activation $\mathbf{R}_{\text{on}}$ and $\mathbf{R}_{\text{off}}$ the contrast signals $\mathbf{K}_{\text{on}}$ and $\mathbf{K}_{\text{off}}$ compete using the DOI–scheme as proposed in Hansen and Neumann (2004b), which uses a stronger weighting of the inhibitory input from the opponent pathway, i.e. $\xi > 1$. The strong inhibition can suppress weak excitations resulting in a sharper tuning curve and an enhanced robustness against noise. The DOI scheme is motivated by recent findings in electrophysiology like Hirsch et al. (1998, 2003) revealing that strong inhibitory contributions can overwhelm excitatory input.

In Hansen (2002) as well as in Hansen and Neumann (2004b) several simple cell models are proposed depending on whether the excitatory and inhibitory input are weighted equally ($\xi = 1$) and whether the input activations $\mathbf{R}_{\text{on}}$ and $\mathbf{R}_{\text{off}}$ are combined in a linear or a non–linear way. This discussion will focus on the non–linear simple cell model with DOI, the other models were also implemented and will briefly be discussed in appendix B.

The nonlinear simple cell model is sketched in figure 3.1. The model first proposed in Neumann et al. (1999) and later enhanced in Hansen (2002) consists of the three intermediate stages $\mathbf{S}^{(1)}_{\text{on}}$, $\mathbf{S}^{(2)}_{\text{on}}$ and $\tilde{\mathbf{S}}$. The main building block of the circuit are the excitatory connections $\mathbf{R}_{\text{on/off}} \rightarrow \mathbf{S}^{(2)}_{\text{on/off}} \rightarrow \tilde{\mathbf{S}}$ providing the excitatory signals to the simple cell $\tilde{\mathbf{S}}$ from its subfields $\mathbf{R}_{\text{on}}$ and $\mathbf{R}_{\text{off}}$.

The further connections in the model introduce nonlinearities and contribute to make the model more selective for light/dark or dark/light contrasts, respectively. The channels $\mathbf{R}_{\text{on/off}} \rightarrow \mathbf{S}^{(1)}_{\text{on/off}} \multimap \mathbf{S}^{(2)}_{\text{on/off}}$ implement a self–normalization by inhibiting $\mathbf{S}^{(2)}_{\text{on/off}}$, thus giving an upper bound for the activity.

The model also includes cross–channel inhibition $\mathbf{R}_{\text{on}} \multimap \mathbf{S}^{(1)}_{\text{off}}$ and $\mathbf{R}_{\text{off}} \multimap \mathbf{S}^{(1)}_{\text{on}}$. The cross–channel connections implement a soft AND–gate or a disinhibition because they inhibit the inhibitory contributions of $\mathbf{S}^{(1)}_{\text{on/off}}$. Therefore, the simple cell response is

amplified if both subfields are simultaneously activated. The stages $\mathbf{S}^{(1)}_{\text{on/off}}$ and $\mathbf{S}^{(2)}_{\text{on/off}}$ can be modeled by a shunting interaction of the type

$$\frac{\partial u_{i,j}}{\partial t} = -\alpha_{\text{S}} u_{i,j} + \text{net}^{+}_{i,j} - \beta_{\text{S}} u_{i,j} \text{net}^{-}_{i,j}, \tag{3.13}$$

which is again assumed to reach steady state quickly. Thus, the channels can be written as

$$\mathbf{S}^{(1)}_{\text{on/off}} = \frac{\mathbf{R}_{\text{on/off}}}{\alpha_{\text{S}} + \beta_{\text{S}} \mathbf{R}_{\text{off/on}}} \tag{3.14}$$

$$\mathbf{S}^{(2)}_{\text{on/off}} = \frac{\mathbf{R}_{\text{on/off}}}{\gamma_{\text{S}} + \delta_{\text{S}} \mathbf{R}_{\text{off/on}}}. \tag{3.15}$$

The simple cell activity $\tilde{\mathbf{S}}$ results from linearly pooling the activations of the on– and the off–channel:

$$\tilde{\mathbf{S}} = \mathbf{S}^{(2)}_{\text{on}} + \mathbf{S}^{(2)}_{\text{off}}. \tag{3.16}$$

By combining the above equations and with $\delta_{\text{S}} = \beta_{\text{S}} \gamma_{\text{S}}$ the simple cell response now reads

$$\tilde{\mathbf{S}} = \frac{\alpha_{\text{S}}(\mathbf{R}_{\text{on}} + \mathbf{R}_{\text{off}}) + 2\beta_{\text{S}}(\mathbf{R}_{\text{on}} \mathbf{R}_{\text{off}})}{\alpha_{\text{S}} \gamma_{\text{S}} + \beta_{\text{S}} \gamma_{\text{S}} (\mathbf{R}_{\text{on}} + \mathbf{R}_{\text{off}})}. \tag{3.17}$$

As final processing step the simple cell activations are mutually inhibited according to

$$\mathbf{S}_{\text{ld}} = \left[\tilde{\mathbf{S}}_{\text{ld}} - \tilde{\mathbf{S}}_{\text{dl}} \right]^{+} \tag{3.18}$$

$$\mathbf{S}_{\text{dl}} = \left[\tilde{\mathbf{S}}_{\text{dl}} - \tilde{\mathbf{S}}_{\text{ld}} \right]^{+}. \tag{3.19}$$

3.1.2.3. Complex Cell Stage

As depicted in section 1.2.1, complex cells are independent of the contrast polarity. They can be modeled by pooling simple cells sensitive to light/dark and dark/light transitions:

$$\mathbf{C}_{\theta} = \mathbf{S}_{\text{ld},\theta} + \mathbf{S}_{\text{dl},\theta}. \tag{3.20}$$

For visualization purposes the pooled complex cell response is used for which all the contributions of the different orientation channels are summed:

$$\mathbf{C}_{\text{p}} = \sum_{\theta} \mathbf{C}_{\theta}. \tag{3.21}$$

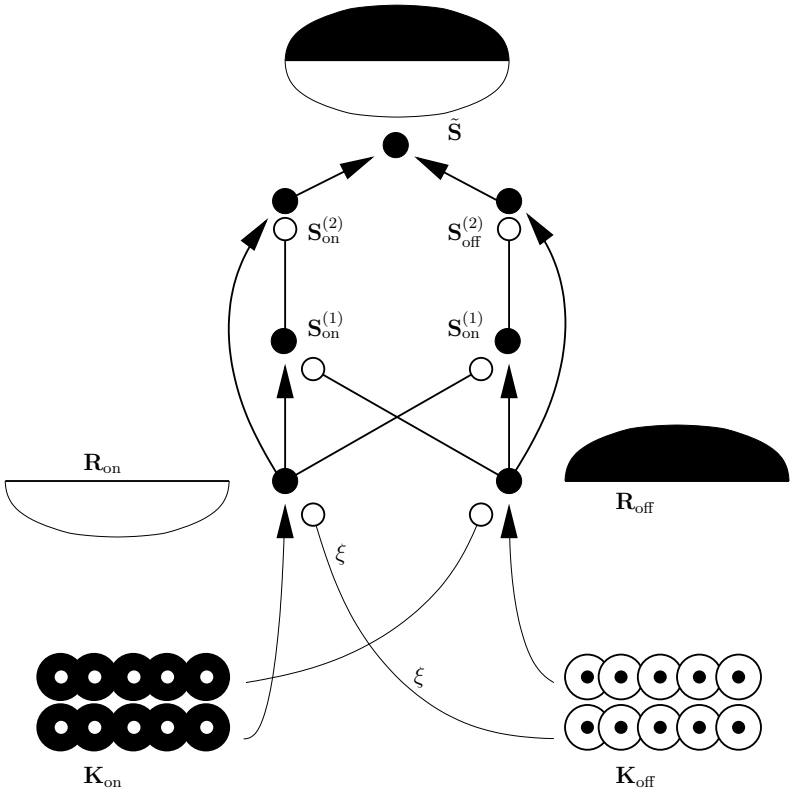


Figure 3.1.: Sketch of the nonlinear simple cell circuit. Filled circles denote neurons, arrows denote excitatory connections, unfilled circles at the end of a connection denote inhibition. Adapted from Hansen and Neumann (2004b)

3.1.2.4. Determing the Model Parameters

The single stages of the model were implemented in MATLAB. The most difficult part was the tuning of the model parameters since the values proposed in Hansen (2002) could not be used due to the different edge detection filters. One possibility to tune the parameters would have been to manually balance the excitatory and inhibitory contributions occurring in several stages. However, an optimization method was chosen as a sensory system is supposed to optimally process the input and to transform it to a meaningful representation.

Up to that step the simple cell model is nothing else than an edge detector. In the literature about edge detectors the following criteria for an optimal edge detector are formulated: good detection, good localization and good response (Canny, 1986). A standard approach to measure the quality of edge detectors is based on Geman and Jedynak (1996) and evaluates the Chernoff bound (Cover and Thomas, 1991). Here, only the brief idea underlying the optimization will be sketched, for a more detailed introduction see section A.5. The complex cell response can be divided into two disjunctive sets of pixels. The pixels lying on an edge are in further referred to as on–edge pixels, and the pixels not lying on an edge are in further referred to as off–edge pixels. Both, the response

for on–edge as well as for off–edge pixels form characteristic probability distributions $p(.|\text{on–edge})$ and $p(.|\text{off–edge})$. The edge detector works well if the two distributions can be distinguished easily which is the case if a classificator has a small classification error. In that case inference about the world as outlined in section 1.2.2.2 is facilitated by the visual system. For this purpose the log–likelihood classificator can be used whose classification error decreases exponentially by

$$\epsilon = \exp\left(-NC(p(.|\text{on–edge}), p(.|\text{off–edge}))\right) \tag{3.22}$$

where N is the length of a sequence of samples drawn either from the on–edge or the off–edge pixels.

$$C\left(p,q\right) = -\min_{0\leq\lambda\leq 1} \log \sum_{j=1}^{m} p^{\lambda}(y_j) q^{1-\lambda}(y_j) \tag{3.23}$$

is the Chernoff bound between the distributions p and q which can take the discrete values $y_i; i = 1 \ldots m$. Thus, the two distributions can be separated more easily if the Chernoff bound is large.

Therefore, the model parameters for the simple cell circuit were tuned in the optimization step such that the response to a stimulus maximizes the Chernoff bound. As stimulus a test stimulus similar to the one depicted in figure 3.2 was used. The width of each bar was 120 pixels, the height were 2600 pixel. The contrast for the optimization was 0.085, a Gaussian noise of standard deviation $\sigma = 0.05$ was added. The contrast was chosen such that the models can detect it well but not perfect at the beginning of the optimization steps.

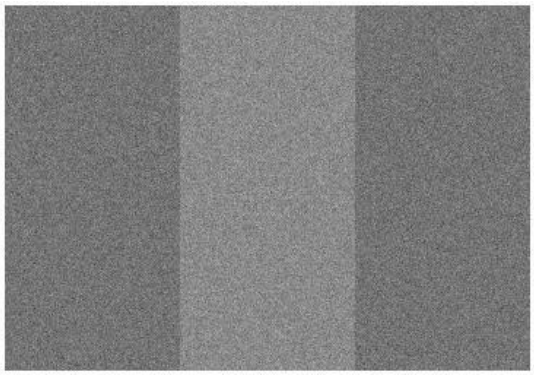


Figure 3.2.: Sketch of the stimulus used for optimization

The optimization was done using an ES with $\mu = 10$, $\sigma = 1$, $\lambda = 10$, $\rho = 3$, a HMB strategy mutation with $\beta = 3$, intermediate recombination, and a Comma–evolution strategy. For every optimization step the complex cell response $\mathbf{C}_\theta$ for the test stimulus was computed, the on–edge and off-edge distributions were estimated, and the Chernoff bound was computed and used as fitness function. For the on–edge distributions only the responses for $\theta = 0$ around the step edges of the pattern were used, all other responses

were counted as off–edge pixels. This method should ensure that only complex cell responses of preferred orientation influence the on–edge distribution, whereas responses of other orientations influence the off–edge distributions. Thus, responses of cells not optimally tuned for the direction of the edge decreased the distinctiveness of the two distributions and sparse–disperse codes of the complex cell response were preferred.

The optimization was done using the following constraints

$$4.0 \leq \alpha_{\mathrm{LGN}} \leq 15.0 \tag{3.24a}$$
$$0.1 \leq \beta_{\mathrm{LGN}} \leq 15.0 \tag{3.24b}$$
$$0.1 \leq \gamma_{\mathrm{LGN}} \leq 15.0 \tag{3.24c}$$
$$0.01 \leq \alpha_{\mathrm{S}} \leq 0.1 \tag{3.24d}$$
$$1000.0 \leq \beta_{\mathrm{S}} \leq 20000.0 \tag{3.24e}$$
$$0.01 \leq \gamma_{\mathrm{S}} \leq 15.0. \tag{3.24f}$$

The other model parameters were fixed to $\sigma_c = 1$, $\sigma_s = 3$ and $\xi = 3$ as they always approached the limits in early trials of the optimization process. The results of the optimization for the nonlinear simple cell model with DOI are summarized in table 3.1.

Table 3.1.: Model parameters for the nonlinear model with DOI

Parameter	Natural filters	Manmade filters	Aerial filters
σ_c	1.000	1.000	1.000
σ_s	3.000	3.000	3.000
α_{LGN}	14.517	12.604	14.323
β_{LGN}	0.573	0.011	0.048
γ_{LGN}	0.048	0.013	0.107
ξ	3.000	3.000	3.000
α_{S}	14.935	13.358	11.050
β_{S}	1039.600	9999.800	9995.700
γ_{S}	0.424	9.546	0.121

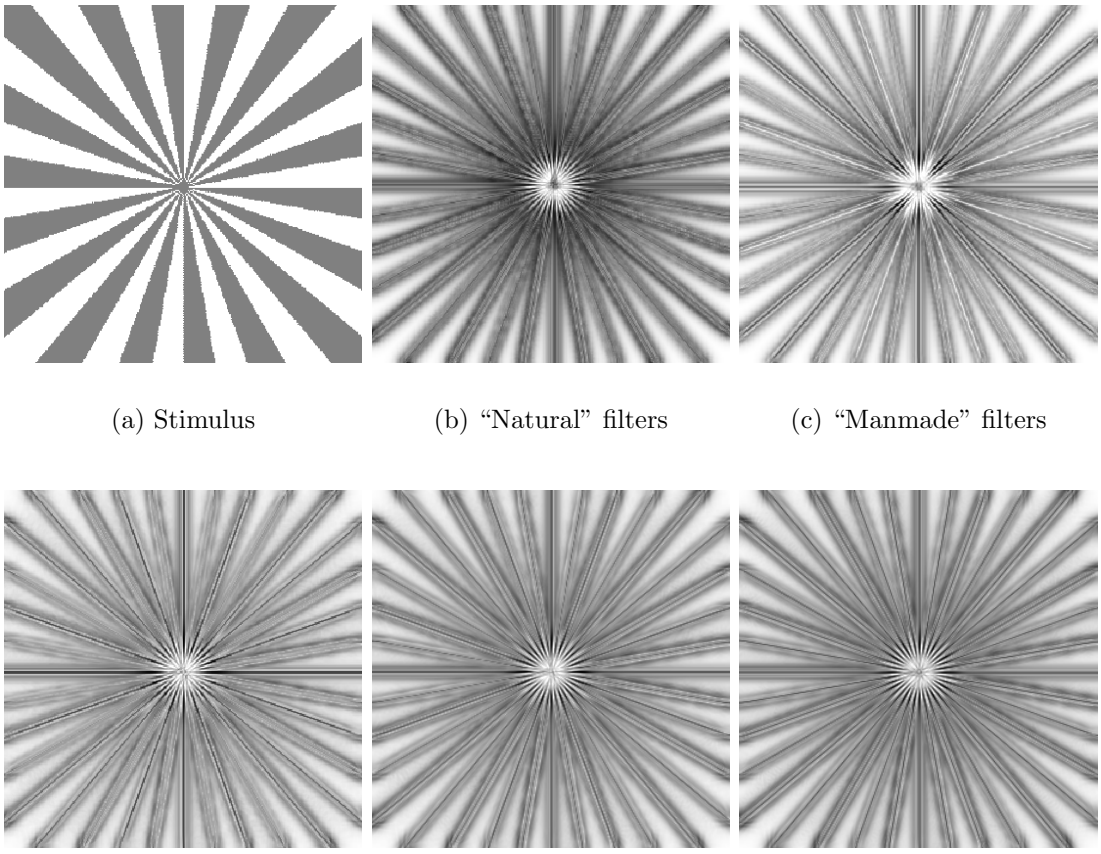
3.1.3. Results and Discussion

This section includes a couple of simulations that are supposed to point out important properties of the model. It closes with results of processing real images for the three image classes described in chapter 2. All responses shown in the following section visualize the pooled complex cell response $\mathbf{C}_P$ and are normalized to the range $[0, 1]$. White regions show areas of no response, black ones denotes maximum response.

3.1.3.1. Artificial Test Images

Siemens Star

For a first test the simple cell model was applied to the Siemens star. The Siemens star is a test pattern known from photography where it is used to test the limits of resolution for optical systems or films. Here, it can show how well the simple cell model can detect edges in various directions, if the response strength is comparable for all directions, and, especially for the center region of the star, how well it can resolve fine patterns. The results of the simulation are shown in figure 3.3.



(a) Stimulus (b) "Natural" filters (c) "Manmade" filters

(d) "Aerial" filters, 8 orientations (e) "Aerial" filters, 10 orientations (f) "Aerial" filters, 12 orientations

Figure 3.3.: Edge detection results for the Siemens star. Each complex cell response was normalized to $[0, 1]$.

The results for processing the Siemens star with receptive fields optimized for natural images show that all orientations are detected with an approximately equal response. Even fine structures at the center of the star can be detected reasonably well. For the filters optimized for manmade scenes and aerial images it appears that the network does

not respond equally to contours in various directions. The filters optimized for manmade scenes are tuned for $\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$. As these orientations are detected better than oblique directions, the result is due to the limited number of used orientations $|\theta|$. Since in manmade scenes there is a strong preference for horizontal and vertical directions (see section 2.1) the number of orientations was not increased to see how the model can deal with realistic images of manmade scenes. Thus, the following results for the manmade image class have to be understood as tests and not as perfect results.

For the simple–cell model optimized for aerial images, the number of orientations was corrected, as one can conclude from the statistical analysis that aerial images do not have an orientation preference. The results for processing the Siemens star with 10 and 12 different orientations show that for both simulations the network responds equally to all orientations. Therefore, $|\theta| = 10$ orientations are sufficient to get good detection results. All further results shown here and in subsequent sections were processed with 10 different orientations.

A nice result for this simulation is that in none of the simulations a difference between the response to light/dark and dark/light transitions occurs and that all can detect fine structures reasonable well. The latter is because of the fact that the used Gabor filters are comparable narrow, i.e. s_1 is rather small. However, for all models the edges are surrounded by a region of more or less strong response. This is due to the fact that in the neighborhood various cells tuned for a similar orientation respond to the edge.

Response to Small Contrasts

In order to analyze the responses to small contrasts a pattern of alternating light/dark and dark/light transitions of increasing contrast was used, comparable to the one proposed in Hansen (2002). The contrast was increased from 0.05 on in steps of 0.05 to 0.145. White noise of standard deviation $\sigma = 0.05$ was added. Each patch of the pattern was sized (500×120) pixels. The results are visualized in figure 3.4 at the end of this section.

For all three models the detection threshold is approximately the same: for a contrast of 0.06 the edges appear as closed contours. For the model with receptive fields optimized to natural and manmade images first edge elements get visible for a contrast level of 0.04, whereas for the model with receptive fields tuned to aerial images only a contrast level of 0.03 is needed. This might be due to the better signal to noise ratio because for the third model the off–edge regions appear brighter in the complex cell response. The contrast values of this work are comparable to the results obtained in Hansen and Neumann (2004b): there, fragmented edges appear for a contrast of 0.045, for closed contours a contrast of 0.07 is necessary.

3.1.3.2. Model Properties

In this paragraph two important properties of the model will be analyzed: the tuning curve and the influence of the DOI.

Tuning Curves

To measure the tuning curve a single step edge which was corrupted by white noise of $\sigma = 0.05$ was analyzed by the simple cell network. The receptive fields were tuned to $|\theta| = 21$ different orientations distributed equally in the range $\left[-\frac{\pi}{5}, \frac{\pi}{5}\right]$, thus sampling the range in steps of 3.6°. Along the edge the mean response of the complex cells for each orientation was computed in order to determine the tuning curve. The tuning curves for the simple cell model with receptive fields optimized to the different image classes and for different contrast levels are shown in figure 3.5.

For all tested models the orientation tuning of the simple cell model is contrast invariant, i.e. the width of the tuning curve is not dependent on the contrast. According to Hansen (2002) this is a property which can only be obtained by simple cell models with inhibition. The shape and the width of the tuning curves varies for the different models tested. For the filters optimized to aerial images the tuning curve orientation selectivity is more restricted than for the other two models. This model only shows activities for orientations deviating ±18.0° from the preferred orientation. The manmade and the natural model show activities for orientations deviating ±25° and ±29°, respectively. Usually the width of the tuning curves is determined by the Half Width at Half Height (HWHH), which measures half the width at the curve's inflection points. The HWHHs of the models are 10.1°, 13.8°, and 6.8° for the receptive fields tuned for natural, manmade, and aerial images, respectively. These values are smaller than the HWHH of 14.7° reported in Hansen (2002). With respect to the shape it has to be mentioned that the tuning curve for the natural and aerial filters are tapering, whereas the one for the manmade filters have a very round peak.

The different width of the tuning curves is probably also responsible for the results obtained for the Siemens star (see figure 3.3). As the receptive fields for natural images are tuned very broadly, $|\theta| = 8$ different orientations are sufficient to detect all orientations equally well, whereas for the aerial filters, which are tuned much narrower, more orientations are needed. Due to the broad tuning, the region surrounding the actual edge is wider and the responses are stronger than for the other image classes.

Influence of DOI to Noise

As mentioned above the DOI has an important influence on the reduction of noise. To measure this influence the complex cell response for an image of constant intensity corrupted with a certain amount of noise was computed. The mean response in dependence of the DOI–parameter ξ, which was varied over a wide range for the measurements, was analyzed. The plots are shown in figure 3.6.

On a first glance all curves have the same shape as they all decay and converge against zero. For all curves the results show that weighting the inhibitory contributions stronger than the excitatory ones reduces the noise contained in the complex cell responses. The main difference between the filter families is how fast the curves decay, i.e. for which ξ they reach a response level where the complex cell response is independent of the noise. For the natural and manmade filters this is only the case for $\xi \geq 8$, whereas for the aerial filters the response is noise–independent for $\xi \geq 4$. These findings might explain the worse signal to noise ratios reported above for analyzing the responses to small contrasts.

However, as pointed out in Hansen (2002), ξ should not be chosen arbitrarily large. As there is a tradeoff between edge detection and noise reduction, ξ was fixed for $\xi = 3.0$ for all models. Although a direct comparison is difficult it seems that the model proposed in Hansen and Neumann (2004b) is more robust against noise. This should mainly be due to the different kind of filters used there. For a further analysis of the DOI–scheme the reader is referred to Hansen (2002).

3.1.3.3. Edge Detection for Real Images

A central aspect of this work is to tune the receptive fields of the simple cells to the statistical properties of the stimuli which it is supposed to process. Thus, in the following paragraph a comparison of the complex cell responses resulting from processing an image with class–specific filters and the filters of the other classes will be given. It has to be mentioned that all evaluations presented here are based on visual inspection of the results, which is a very subjective method. Additionally, the evaluation was done on the computer screen that allowed a better comparison of the results than it is the case for the printed version of the thesis.

Natural Scenes

On the first glance all three receptive field classes lead to good detection results. The shape of the animals is represented best when processed with the filters tuned for natural images. By applying the other filters the smooth and round contours like the penguin's head get more jaggered and do not look very naturally. Especially the eyes of the penguins get almost rhombic if processed by the receptive fields tuned to manmade scenes. This is again due to the problems of the model to detect oblique oriented edges. The very low contrast edge arising from the left penguin occluding the right one is detected very weakly with all models. However, as in figure 3.7(d) the noise resulting from the bird's feathers is very low, the edge is most outstanding in this image.

Similar results are obtained for figure 3.8. By processing the image with filters adapted to aerial images the vertical structures of the tree are detected very well, all other structures are detected well, although the results for the other filters appear more realistic and do not drop as many details, especially in the horizontal structures of the tree or the animal's leg.

Manmade Scenes

Figure 3.9 shows the results for an image with almost only horizontal and vertical edges. There is no obvious difference between the results obtained by receptive fields tuned to natural and manmade scenes, respectively. This is due to the structure of the image and the similarly parameterized Gabor functions used to model the receptive fields. If the image is processed with the filters tuned to aerial images the main shapes of the buildings are detected well, but some details like the vertical structures of the long flat house directly at the shore are not detected, whereas the other two receptive field types can detect it very weakly.

The image showing the row of houses is also characterized by vertical and horizontal structures but oblique oriented edges play a more important role, especially for details.

The results are visualized in figure 3.10. When processed with receptive fields adapted to natural images, the main structures as well as many details are detected and the round structures are represented smoothly. For processing with filters tuned to manmade scenes, some differences can be revealed, especially for the second building from the right. There, the oblique structures of the roof and the gable are not detected. Also, the round structures are not represented well. When filtered with the receptive fields optimized for aerial images there are some details missing again, e.g. the decorations at the second gable from the right. However, the main forms are detected well, round contours look smooth.

Aerial Images

The results for aerial test images are shown in figures 3.11 to 3.14. The differences between the several test images are not very obvious. For all test images the model with receptive fields tuned to aerial images extracted mainly long contours, which are assumed to be one of the key informations of aerial images. However, small round structures, such as the roundabout close to the lower right corner in figure 3.11, are not detected very well. For such fine structures, the Gabor filters tuned to aerial images are too elongated. On the other hand, when using the receptive fields optimized for natural images the resulting representation contains still a lot of noise and jitter, although the main structures are detected. The results for processing the images with filters tuned to manmade scenes were not applied to aerial images as the results are comparable to those obtained for natural filters except for problems with detecting oblique contours.

3.1.4. Conclusions

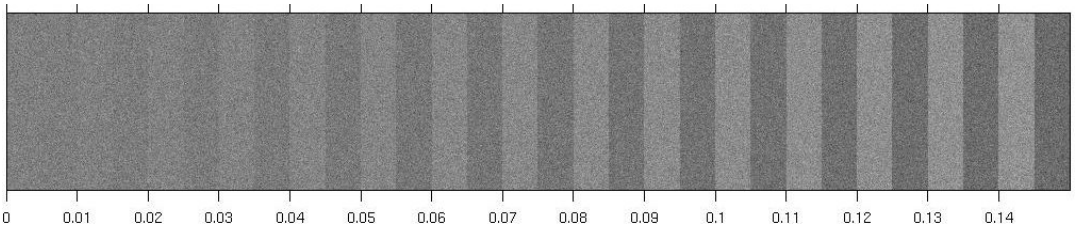
The simple cell model detects edges and contours robustly. It seems that the implemented model is not as robust against noise and the orientation tuning is not as sharp as for the original model proposed in Hansen (2002) or Hansen and Neumann (2004b). This is due to the different orientation selective filters used here. The filters used in Hansen (2002) are comparable to the filters proposed in Canny (1986). They are optimized for edge detection purposes and are not a tradeoff between edge and frequency detectors as it is the case for Gabor filters. In future work it might be worthwhile to implement such an optimization step as it was proposed in Lähdesmäki et al. (2001) in order to tune an edge detector to cave inscriptions.

The analysis of real photographies revealed that taking the statistical properties of the environment and the receptive fields into account helps to enhance the quality of the edge detection results. The rough edge structures of the images is also obtained when the test images are processed with filters optimized for a different image class. However, the best results, i.e. the edge representation looking most naturally or the representation containing the least amount of noise and jitter, is always obtained by filtering with optimized filters. One can also conclude that when applying receptive fields from the "aerial" class to images of the "natural" and "manmade" class aerial receptive fields to natural and manmade scenes fine detail is often not detected. Applying filters tuned for natural scenes to aerial images results in a complex cell response with too many

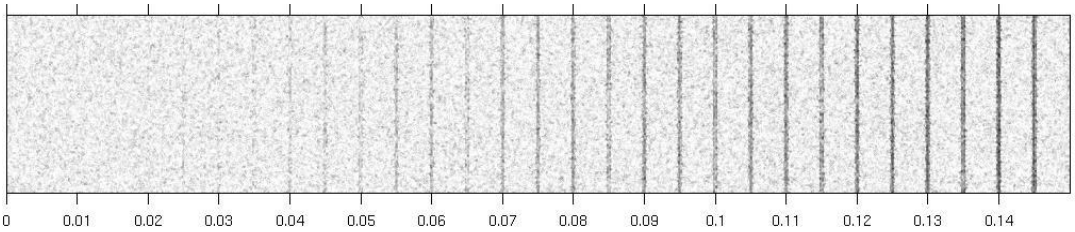
details and a lot of jitter making the further processing for the landmark selection and navigation task hard or even not possible. Although the problems are due to the different scales of the used filters, it is still a nice result that ICA adapts to this different scales. This relationship between scale space theory (Lindeberg, 1999) and the tuning of Gabor filters estimated by ICA might also be a future working direction.

The results for the manmade scenes that were filtered with receptive fields of only $|\theta| = 4$ different orientations revealed that, although the simple cell network does not detect oblique contours as well as horizontal, vertical, and diagonal contours, this property influences mainly detailed structures. Therefore, it does not decrease the quality of the overall edge detection too much. However, to model an isotropic edge detector $|\theta|$ has to be enlarged.

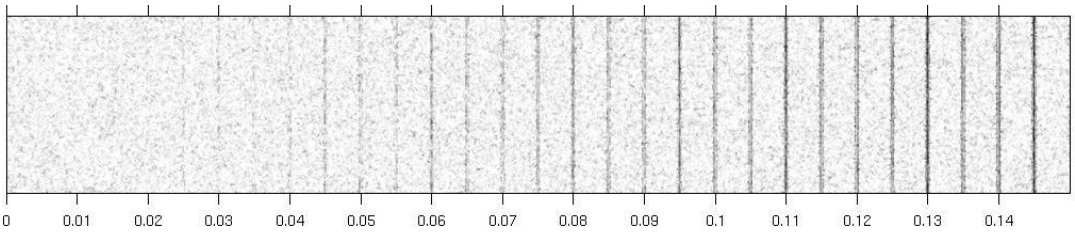
Although the edge detection works it could be worthwhile to implement an iterative contour enhancement and grouping mechanism as proposed in Neumann and Sepp (1999) or Hansen and Neumann (2004a). These models for the recurrent V1–V2 interactions enhance collinear edge elements and inhibit edge elements of other orientations. Thus, the grouping mechanism can close gaps in the edge representation, sharpen the edge response, and suppress noise. All properties could be useful to enhance results of the edge detection network. An interesting part of this work could include the investigation how the statistical properties of the contours vary for different image classes. Thereupon, one could tune the contour grouping to the specific image class by adapting the range of enhancement or the range of inhibition. Works that could be used as starting point for these investigations include Geisler et al. (2001), Sigman et al. (2001), or Hoyer and Hyvärinen (2002) that analyzed the statistical properties of contours in natural images.



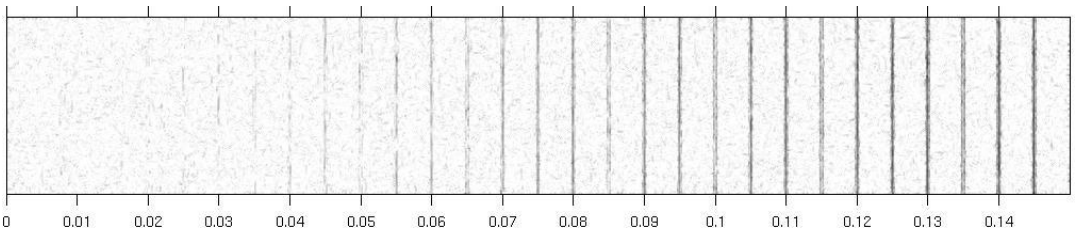
(a) Stimulus



(b) “Natural” filters

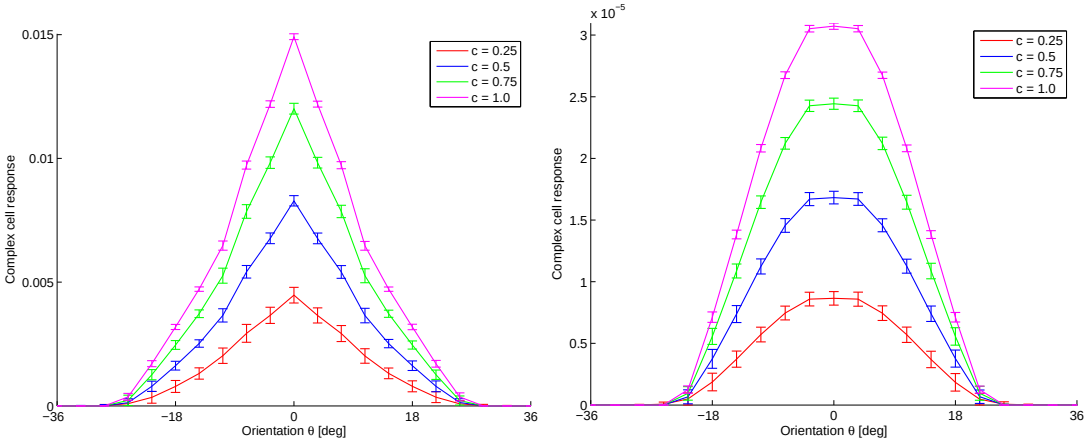


(c) “Manmade” filters



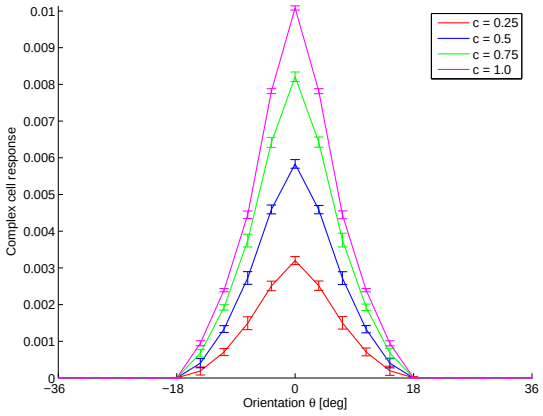
(d) “Aerial” filters, 10 orientations

Figure 3.4.: Response properties to small contrasts. As the responses for the different models are normalized to $[0, 1]$ only the signal to noise ratio between the plots can be compared.



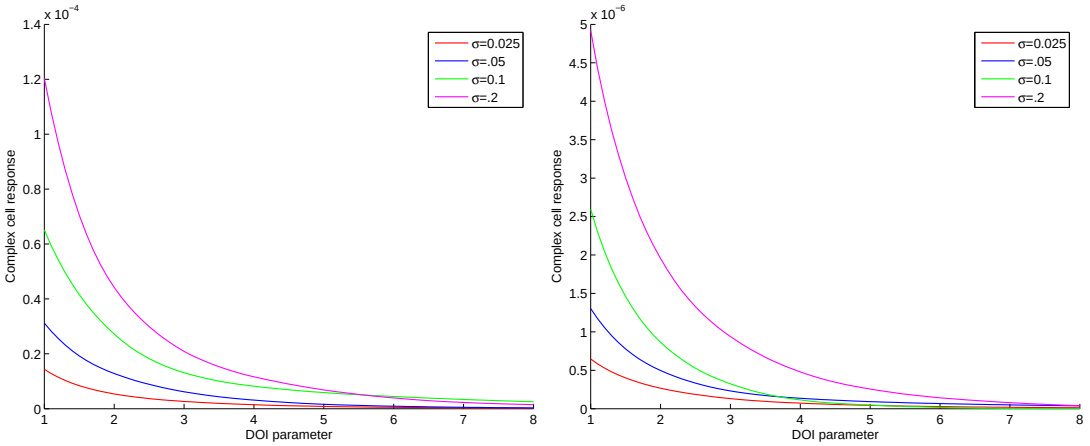
(a) "Natural" filters

(b) "Manmade" filters



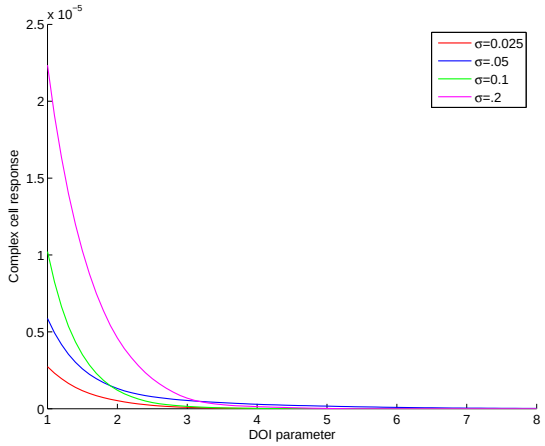
(c) "Aerial" filters, 10 orientations

Figure 3.5.: Tuning curves. The tuning curves were measured for several contrast levels and 21 orientations distributed equally in $[-36\,°, 36\,°]$. Plotted are the responses of complex cells with preferred orientation θ to a vertical edge (i.e. $\theta = 0$ of contrast c which was corrupted with noise of $\sigma = 0.05$. Error bars denote the standard deviation of the response



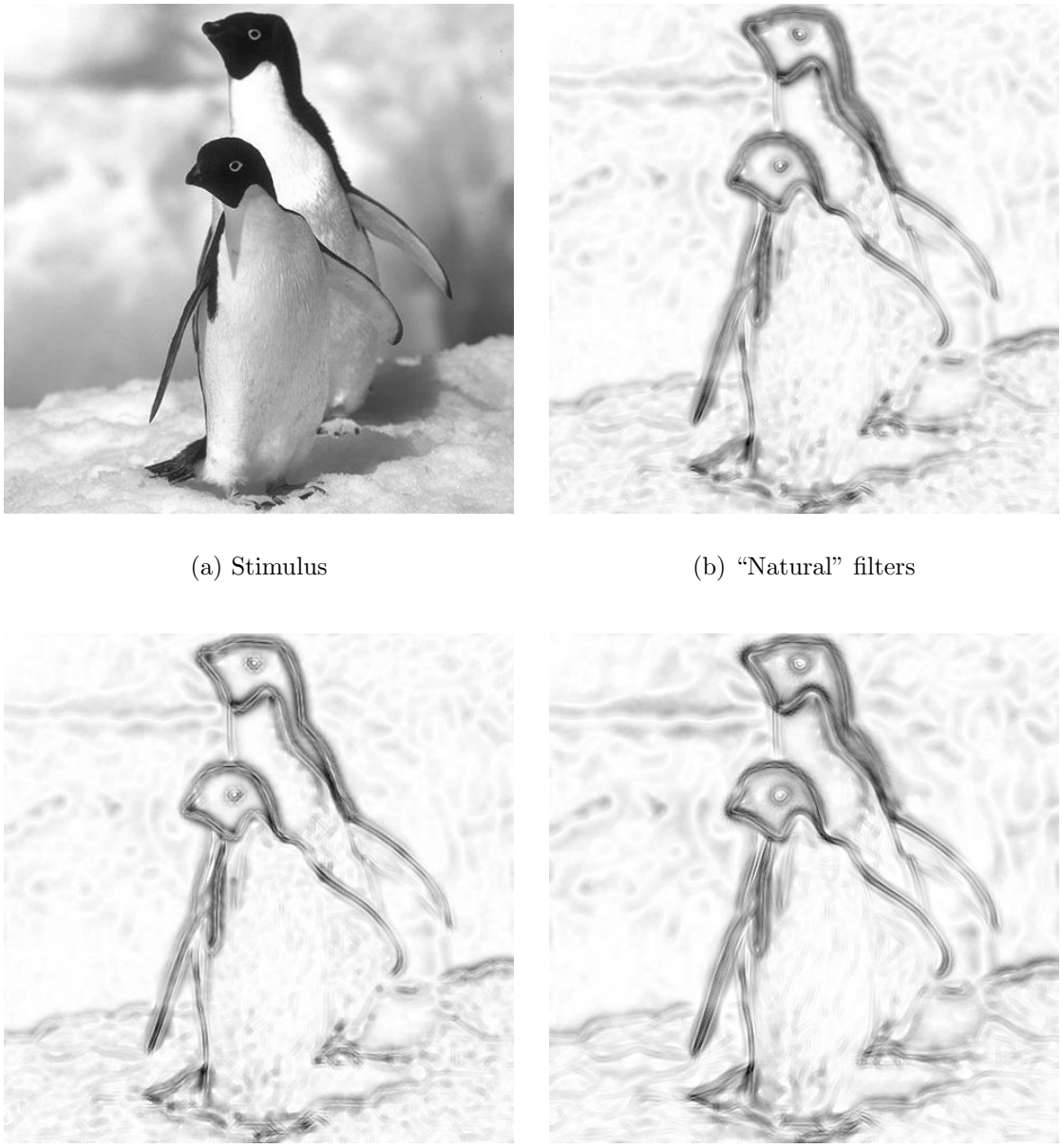
(a) "Natural" filters

(b) "Manmade" filters



(c) "Aerial" filters

Figure 3.6.: Influence of DOI to noise. A unique patch corrupted with different levels of noise was used as test pattern. The mean complex cell response to the patch is plotted over the change of ξ.



(a) Stimulus

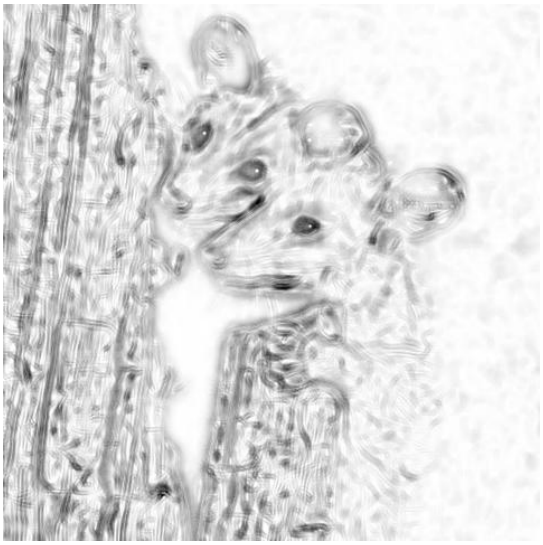
(b) “Natural” filters

(c) “Manmade” filters

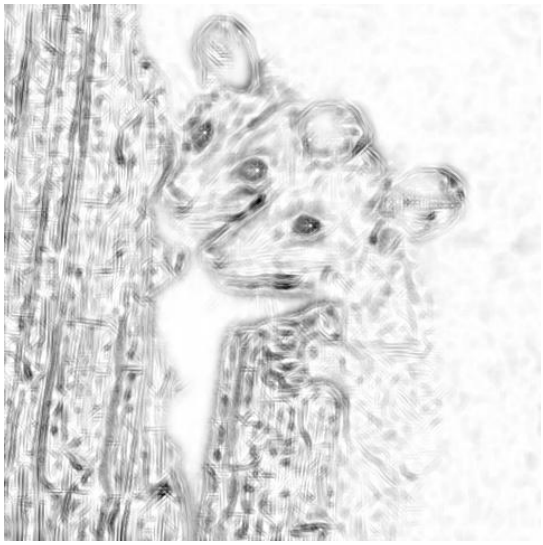
(d) “Aerial” filters

Figure 3.7.: Edge detection results for a natural scene, example 1

(a) Stimulus



(b) “Natural” filters



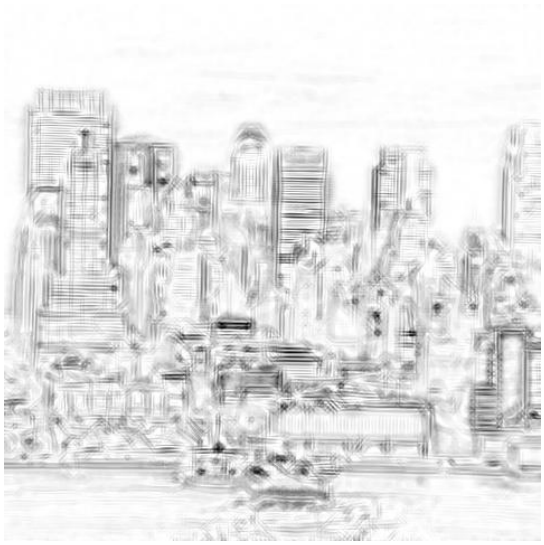
(c) “Manmade” filters

(d) “Aerial” filters

Figure 3.8.: Edge detection results for a natural scene, example 2

(a) Stimulus

(b) “Natural” filters



(c) “Manmade” filters

(d) “Aerial” filters

Figure 3.9.: Edge detection results for a manmade scene, example 1

(a) Stimulus

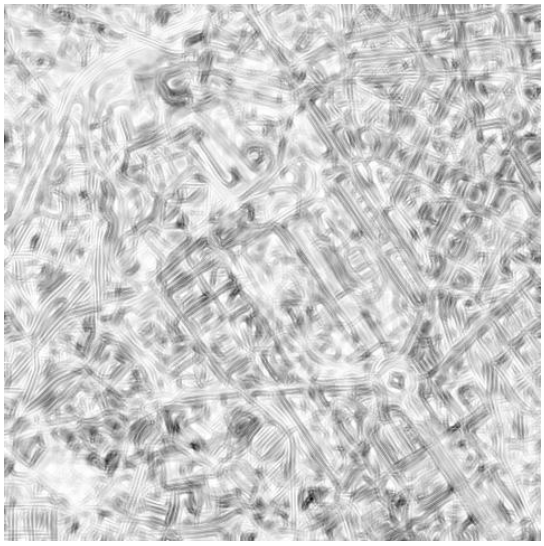
(b) “Natural” filters

(c) “Manmade” filters

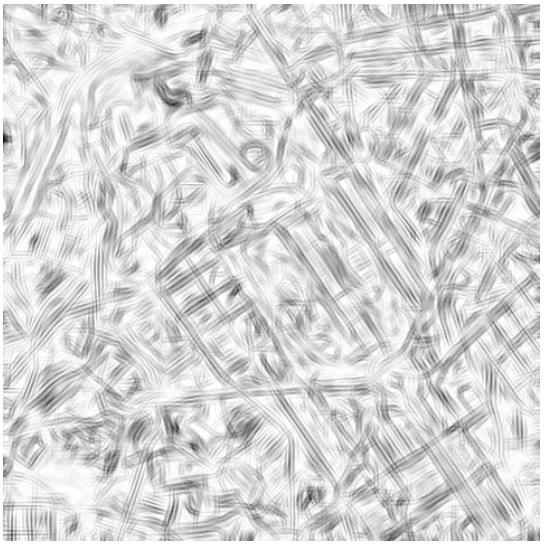
(d) “Aerial” filters

Figure 3.10.: Edge detection results for a manmade scene, example 2

(a) Stimulus



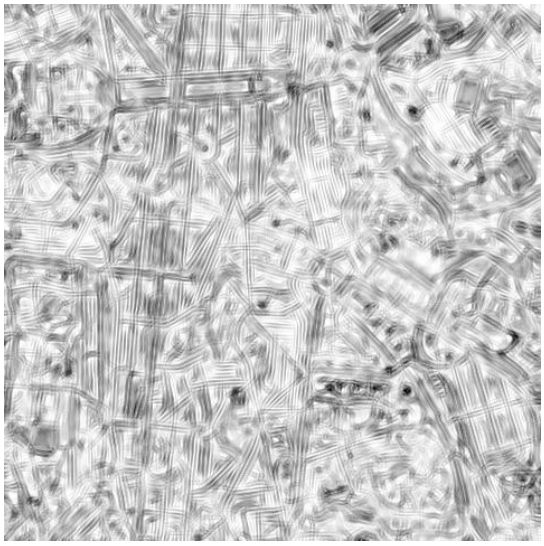
(b) “Natural” filters



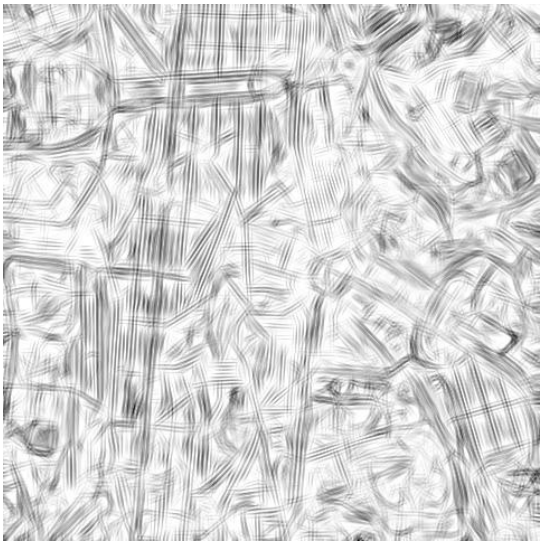
(c) “Aerial” filters

Figure 3.11.: Edge detection results for an urban aerial image, example 1

(a) Stimulus



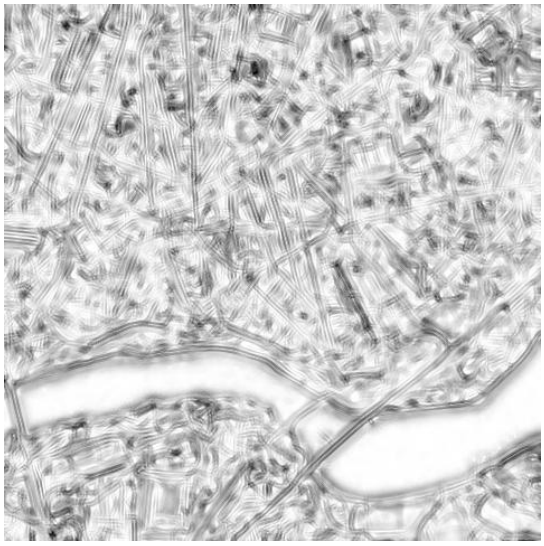
(b) “Natural” filters



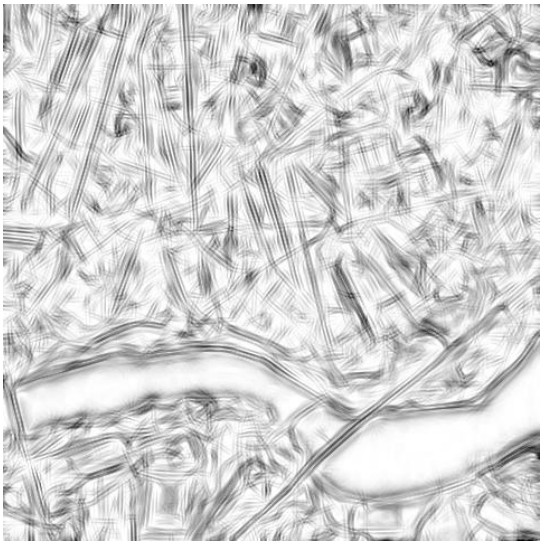
(c) “Aerial” filters

Figure 3.12.: Edge detection results for an urban aerial image, example 2

(a) Stimulus



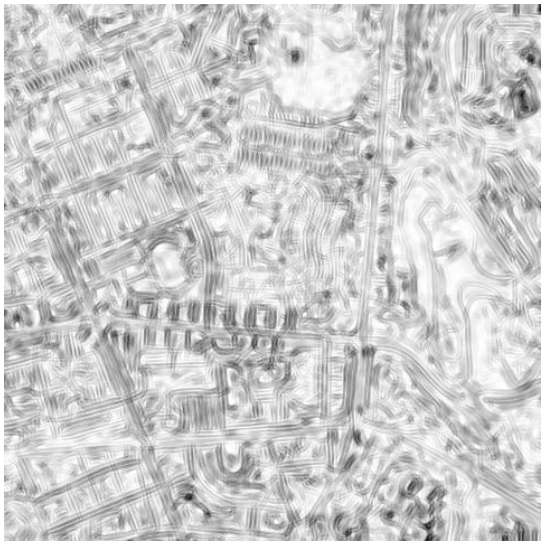
(b) “Natural” filters



(c) “Aerial” filters

Figure 3.13.: Edge detection results for an urban aerial image, example 3

(a) Stimulus



(b) “Natural” filters

(c) “Aerial” filters

Figure 3.14.: Edge detection results for an urban aerial image, example 4

3.2. Junction Detection

In this section the complex cell response is used to detect junctions and corners. Formally, the term junction refers to image configurations with one or more ending edge segment, whereas a corner is a junction with one of the angles between line segments larger than 180° (Würtz and Lourens, 2000). Here, both terms are used interchangeably. The detected junctions will be used to select landmark candidates as input for the landmark selection algorithms proposed in Gerstmayr et al. (2004a,b). Since the corner detector is used as point–of–interest (PoI) operator, the review will include related work on PoI operators in general and not only on edge detectors.

3.2.1. Related Work

The general advantages of PoI–selection are that computation time is saved and that robustness is improved because only the relevant parts of the image have to be processed. Additionally, only few but very distinctive and promising cues are taken into account (Triggs, 2004). To only pay attention to few outstanding points instead of the whole image is motivated by psychophysiological findings (Loy and Zelinsky, 2003).

Following Schmid et al. (2000), the following three approaches to PoI–detection can be distinguished:

Intensity–based Methods: These methods have in common that they are based on detecting contrast variations of different orientations. PoI–operators analyzing the structure tensor like Triggs (2004) are often based on Harris and Stephens (1988) or Shi and Tomasi (1994). An alternative intensity–based method is the symmetry transform. Based on gradient information works like Reisfeld et al. (1995) and Loy and Zelinsky (2003) detect points of high radial symmetry. Other intensity–based approaches evaluate Gaussian curvature, edge density (Bourque et al., 1998), or model the visual pathway in humans. Using the complex cell responses, corners and junctions can be detected using so–called end–stopped cells (Heitger et al., 1992; Würtz and Lourens, 2000) or implicitly like in Hansen and Neumann (2002, 2004a) or in this work.

Contour–based Methods: These methods are build on top of an edge detection scheme and detect characteristic patterns in the change of curvature.

Parametric Methods: Parametric models try to fit junction models to the image. These methods are computationally rather expensive as they include many image comparisons and an optimization step.

PoI–detection is a field of ongoing and very active research in Computer Vision. For reviews see Schmid et al. (2000), Sim et al. (2003), and Loy and Zelinsky (2003). A subtopic, which currently receives special interest, are PoI–detectors which are invariant to rotation, scale, illumination, and position (Triggs, 2004; Lowe, 2004).

3.2.2. The Model

Corners or junctions can be characterized by high activity in several orientation domains at a model hypercolumn of the complex cell response as computed by equation (3.20). The model proposed in Hansen and Neumann (2002, 2004a), which was inspired by Zucker et al. (1989). It does not rely on the distributed activity in several domains of a hypercolumn and does not need an explicit representation of corners or junctions. It is faster than a parametric model approach because there is no optimization step. Additionally, it can deal adequately with image structures of more than two intersecting edges, which is not the case for approaches based on the structure tensor (Medioni et al., 2000).

For a given complex cell response $\mathbf{C}_\theta$ a measure for the "junctioness" is given by

$$\mathbf{J} = \operatorname{circvar}(\mathbf{C}_\theta)^2 \sum_\theta \mathbf{C}_\theta \tag{3.25}$$

The function "circvar" is a measure for the circular variance at a certain location of the complex cell response and takes values in the range $[0, 1]$. A low value indicates activity in a single orientation domain. The ratio between low and high circular variances is further enhanced by the squaring operation. The circular variance is complementary to the orientation significance function:

$$\operatorname{circvar}(\mathbf{C}_\theta) = 1 - \operatorname{osgnf}(\mathbf{C}_\theta) \tag{3.26}$$

$$= 1 - \frac{|\sum_\theta \mathbf{C}_\theta \exp(2i\theta)|}{\sum_\theta \mathbf{C}_\theta} \tag{3.27}$$

The circular variance function is also used in electrophysiological work like Ringach et al. (2002) and Gur et al. (2004) to analyze orientation tuning in V1. The model is not restricted to operate on the complex cell response but can work on any phase–invariant and orientation selective cell response like the result of long–range integrations such as proposed by Neumann and Sepp (1999) and Hansen and Neumann (2004a).

In order to localize the corner points, the junction map $\mathbf{J}$ is blurred with a Gaussian smoothing filter of standard deviation σ_J. Local maxima are computed within a neighborhood of size δ_J. Points in $\mathbf{J}$ are marked as corner points if they reach the local maximum and their strength exceeds a fraction α_J of the overall maximum response.

3.2.3. Results and Discussion

3.2.3.1. Localization Accuracy for Different Junction Types

In order to find out how the model detects various junction types and how accurate the localization of the detected junctions is, it was applied to several artificial images showing various junction types proposed in the literature. Each junction was located in the center of the images sized 128×128 pixels and the distance from the center to the closest detected junction was computed using the Euclidean distance. The stimuli and

complex cell responses are shown in figure 3.15, the localization accuracy is summarized in table 3.2. Only for the X–junction the localization of the detected junctions are correct. The computed distance of 0.707 pixels is due to discretization errors of the junction detection. For all other junction types the localization error is comparable large. This might be due to the Gaussian smoothing filter which was parameterized with a rather large standard deviation of $\sigma_{\mathrm{J}} = 9$ in order to average over several maxima. Otherwise, double responses would have obtained. It seems that the model works best if filters optimized for aerial images are used. The localization accuracy of the original model presented in Hansen (2002) is slightly better for the junction detection model applied to the complex cell response. The detection is significantly more accurate for the model applied to the response after the iterative grouping mechanism which models the long range interaction between V1 and V2 (Hansen and Neumann, 2004a). As the localization accuracy is also a problem for the corner detection quality of real images, implementing the contour–grouping mechanisms seems to be useful to improve the edge detection quality.

Table 3.2.: Localization accuracy for different junction types. Shown are the Euclidean distances of the closest junction to the center position of the image which is the correct localization of the junction. The first three columns show the results for different class–specific receptive fields, the forth and fifth column show the results of Hansen (2002). The forth column shows the results for applying the junction detection model to the pure complex cell response, the fifth column shows the accuracy for the junction detection after contour grouping.

Type	Natural	Manmade	Aerial	Hansen	Hansen LR
L	1.581	7.106	2.915	3.54	0.70
X	0.707	0.707	0.707	0.00	0.00
T	7.778	7.106	4.743	2.92	0.05
Y	5.701	5.523	4.528	2.96	1.58
W	2.121	1.581	2.121	2.55	0.70
Ψ	7.906	7.649	4.528	6.50	3.50
mean	4.299	4.945	3.257	3.08	1.09
std	3.228	3.042	1.634	2.08	1.32

3.2.3.2. Junction Detection for Real Images

In the following paragraphs the junction detection results for the testimages shown in section 3.1.3 were used to evaluate the model. Again, the complex cell response computed with receptive fields optimized for natural and aerial images will be compared by applying the junction detection with the same parameters. Results for manmade scenes are not shown, as the results would not reveal any new insights or problems. As in section 3.1.3 it shall be stressed that the evaluation is based on visual inspection.

Natural Scenes

The results for natural scenes are depicted in figure 3.16 and 3.17. Again, it was necessary to choose $\sigma_J = 7$ and $\delta_J = 7$ rather large, in order to prevent double responses. Since the features are not to close together in the natural scenes, this choice only had an influence of the localization accuracy. Like for the artificial test images it is not too accurate, as can be seen e.g. in figure 3.16 at the transition between snow and background or the penguin's heads.

Additionally, it was very difficult to define an appropriate threshold. The chosen threshold of $\alpha_J = 0.2$ was therefore chosen more or less arbitrarily. For decreasing α_J the number of detected features increases extremely with also increasing the number of false negative and false positives, i.e. missed junctions and detected junctions at places where no junction is expected. If α_J is increased, only very little junctions are detected, which are additionally often located at unexpected and for an observer not very meaningful positions. This means that the junctioness values $\mathbf{J}$ vary over a large range and that there is no guarantee that relevant or meaningful structures are detected with high junctioness values. Examples of false positives in figure 3.16 include the background structures and the snow, which an human observer would not rate as corner or junction. False negatives include the tips of the wings, the front penguin's head occluding the back penguin's body, and the front penguin's tail end.

When comparing the results for processing the image with filters tuned for natural and aerial images there is no big difference visible for figure 3.16. Both detectors seem to have the same problems mentioned above. However, for figure 3.17 the junction detection of the natural filter's complex cell response gives significantly more junctions and corners than for the aerial filter's response. Both detect the main features in the image which include the animal's nose, eyes, ears, and paw. The main difficulties arise in the tree and the animal's coat. There, it is difficult to judge from just looking at the results which model performs better. It seems that – as expected from the results of the edge detection – the results for the complex cell response obtained with receptive fields adapted to natural scenes include small detailed as well as large–scale features. The results for the complex cell response obtained by processing with filters tuned to aerial images do only include small–scale features. These results are not surprising, as the filters tuned for aerial images work at a larger scale than those for aerial images.

Aerial Images

The results for aerial images are shown in figures 3.18 to 3.21. For this image class it is even more difficult to find appropriate parameters. On the one hand the features are distributed more densely than in natural and manmade scenes. On the other hand larger corners and junctions resulting from larger structures as roundabouts or intersections, which look very important for human observers, only have very small junctioness values. Therefore, $\alpha_J = 0.0$, $\sigma_J = 5$, $\delta_J = 5$ were chosen.

For the images processed with filters tuned to natural images this parameter choice is for sure not optimal. However, there are only few responses very close together, mainly in areas where the edge detector resolved small details and not only the main image structures dominated by streets, river lines, or parks. Most of the detected features

do indeed correspond to meaningful structures in the image, although the features are rather low–level features and not high–level features like crossroads formed by larger roads.

The junction detection model seems to give slightly less junctions for the complex cell response obtained from filtering the images with filters tuned to aerial images. Also the distance between two detected junctions seems to be larger which is due to the larger scale of the filters tuned to aerial images. This filters produce a "cleaner" representation of the aerial images that does not contain as much jitter and low–scale details.

The large amount of features in the images makes it difficult to analyze the performance of the junction detector in more detail. There are some sporadic false negative and false positives among the detections. It is difficult to interprete their influence on the detector's performance. Another problem which influences the quality is the worse localization accuracy, which is very obvious in figure 3.20. There, the corners and junctions of the river line are all detected in the river.

3.2.4. Conclusions and Future Work

The biggest problem of the model is the inaccurate localization of the detected junctions and corners. This is due to the rather large standard deviations σ_J needed to smooth the junctioness values $\mathbf{J}$ to avoid double responses. In Hansen (2002) it is shown that the contour–grouping mechanism as intermediate processing stage between edge and junction detection can increase localization accuracy significantly. Thus, it is an important step for future work to implement this mechanism as intermediate stage between the edge and corner detection.

As the contour grouping mechanism also increases contour saliency, closes gaps in contours, and enhances the contour strength up to a saturation level (Hansen, 2002) one can hope that this mechanism also helps to overcome the other drawback discussed above: the huge variety of junctioness values which makes it extremely difficult to set a detection threshold δ_J. Additionally, it has been shown in Hansen and Neumann (2004a) that the grouping mechanism can reduce the number of false detections significantly.

Although the results can still be improved, it is likely that the junction detection model can be used as PoI–detector to preselect a set of landmark candidates.

A point not analyzed in detail yet is the influence of the class–specific receptive fields to the results of the junction detection. There is a tendency that the model detects less junctions and corners if applied to the complex cell response obtained by filters tuned to aerial images than if applied to the complex cell response resulting from filtering with filters adapted to natural scenes. However, the responses were only analyzed by visual inspection which is for sure not sufficient for a well–founded comparison. A good approach for a statistical evaluation is approach of Hansen (2002) using ROC–curves. A ROC–curve visualizes the proportion of true positive responses against the proportion of false positive responses for a varying decision threshold (Duda et al., 2001). The proportions are obtained by comparing the results of the junction detection to a ground truth.

3.3. Chapter Summary

In this chapter the models for the biologically–motivated image preprocessing that will be used in chapter 4 to preselect landmark candidates were described. For the preprocessing the class–specific receptive fields derived in chapter 2 have been used.

In section 3.1 a simple cell model for edge detection has been proposed which uses the class–specific receptive fields. The results revealed that the receptive fields have an influence on the edge detection quality. Although the main structures are also detected if the image is processed with filters not tuned to the characteristics of the input image, the best results, i.e. the results looking most smooth or containing the least amount of jitter, are always obtained by processing with class–specific receptive fields. Another nice result was that the edge detector with receptive fields tuned to manmade structures can detect the main structures, although it is anisotropic and does not detect oblique structures. However, one can argue that these do not play an important role in aerial images.

The corner and junction detection model proposed in section 3.2 can detect corners and junctions in the image, although it still needs some improvements. Therefore, it is worth to implement a contour grouping mechanism as intermediate processing stage between the edge and junction detection. The main drawbacks of the junction detection model, which can hopefully be overcome in future work, are that its localization properties are rather inaccurate and that the tuning of the model parameters is difficult.

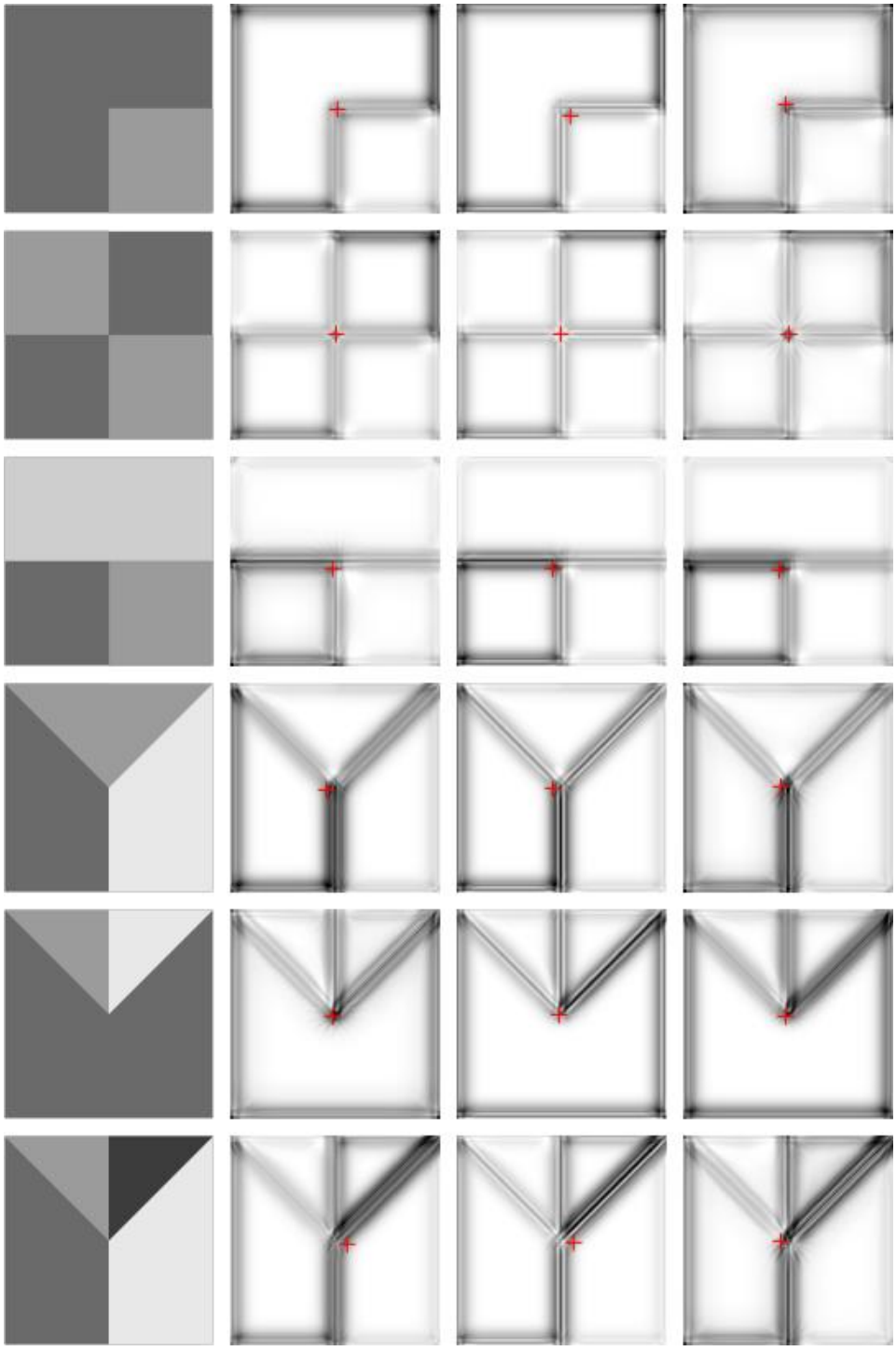
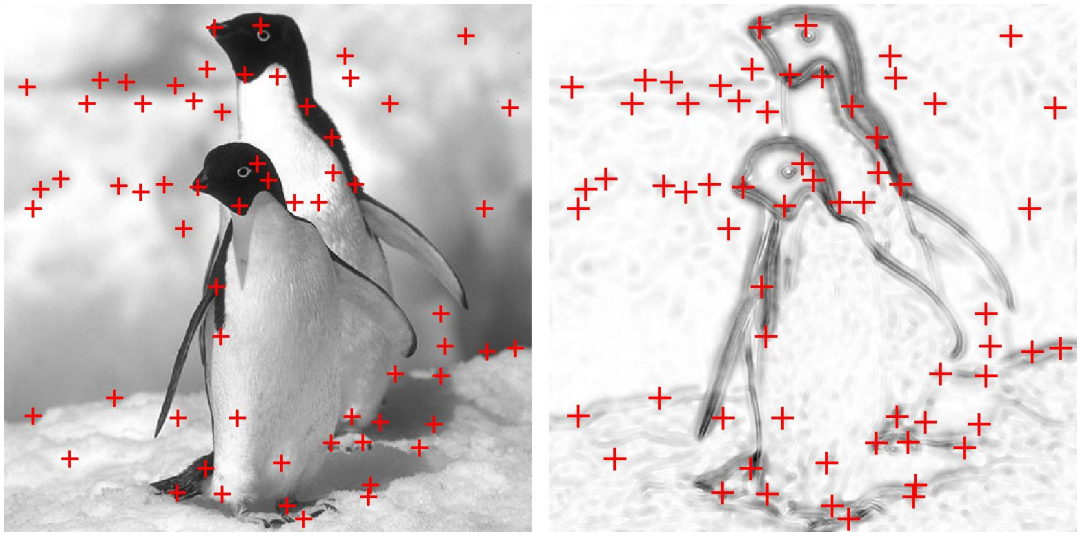
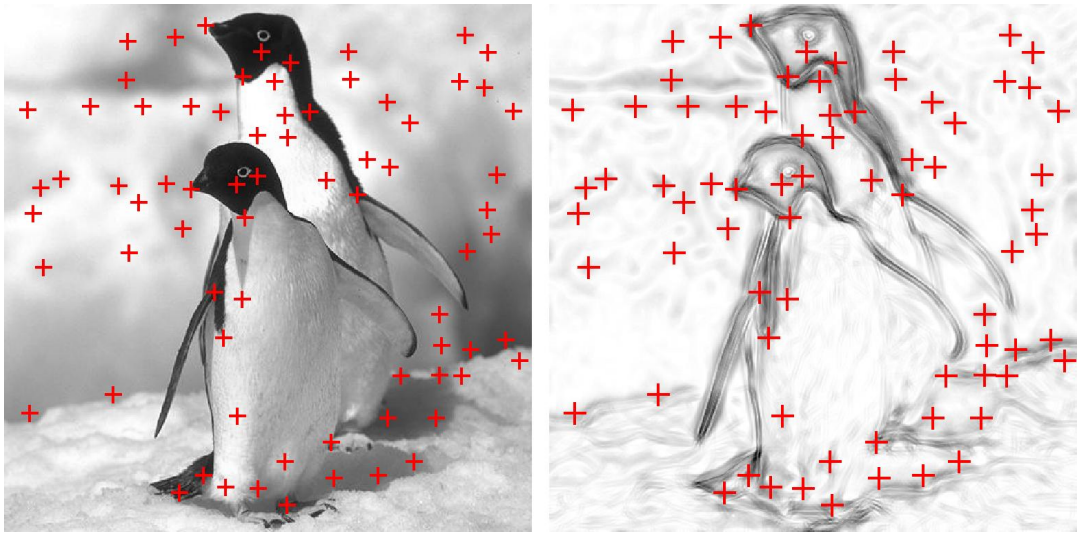


Figure 3.15.: Localization accuracy for different junction types. The rows show the different junction types, namely L–, X–, T–, Y–, W–, and Ψ–junctions. The columns show the stimulus and the results obtained with receptive fields for natural, manmade, and aerial images, respectively. The parameters were $\sigma_{\mathrm{J}} = 9$, $\delta_{\mathrm{J}} = 15$, and $\alpha_{\mathrm{J}} = 0.2$.

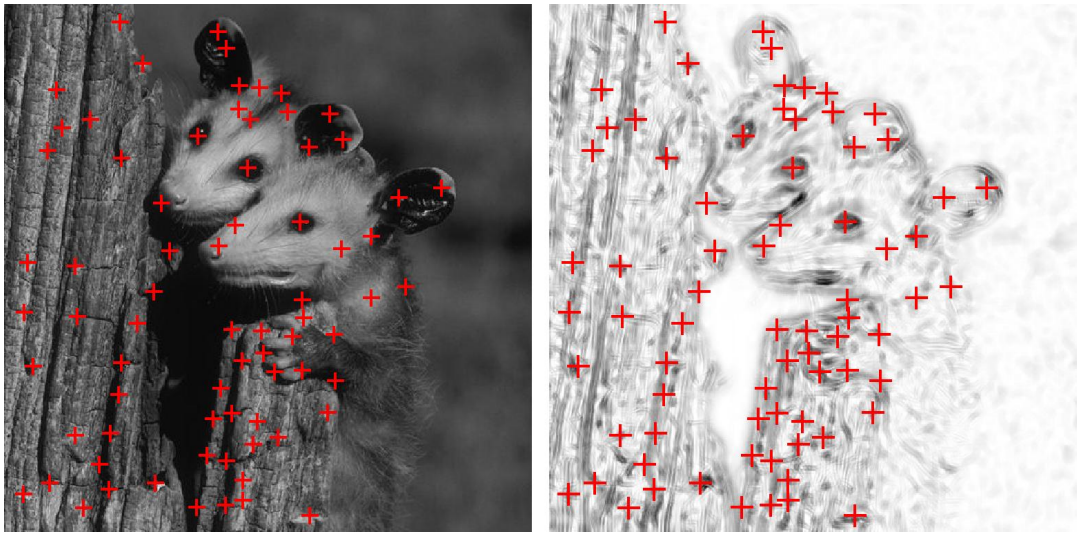


(a) “Natural” filters

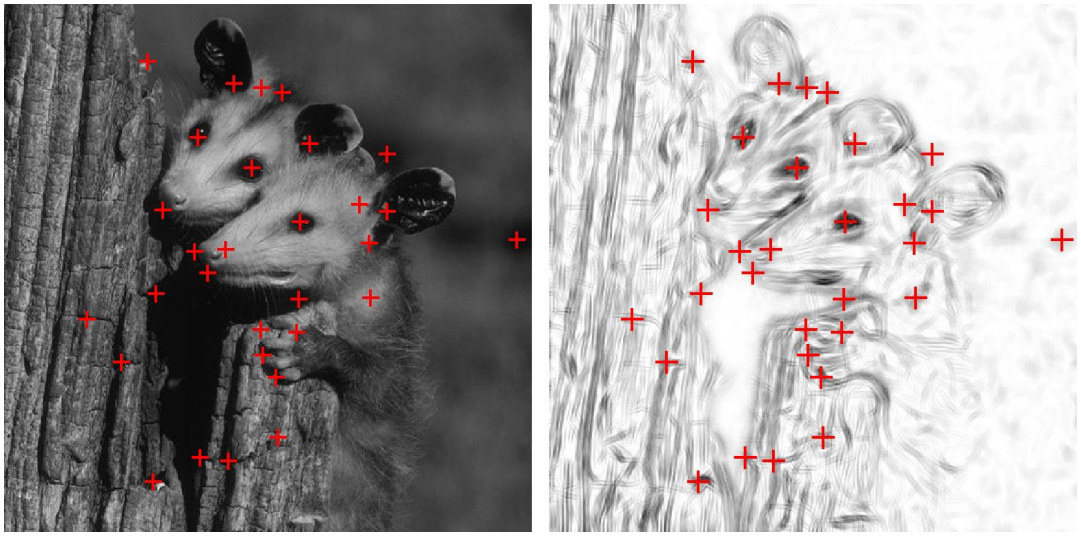


(b) “Aerial” filters

Figure 3.16.: Corner detection for a natural scene, example 1

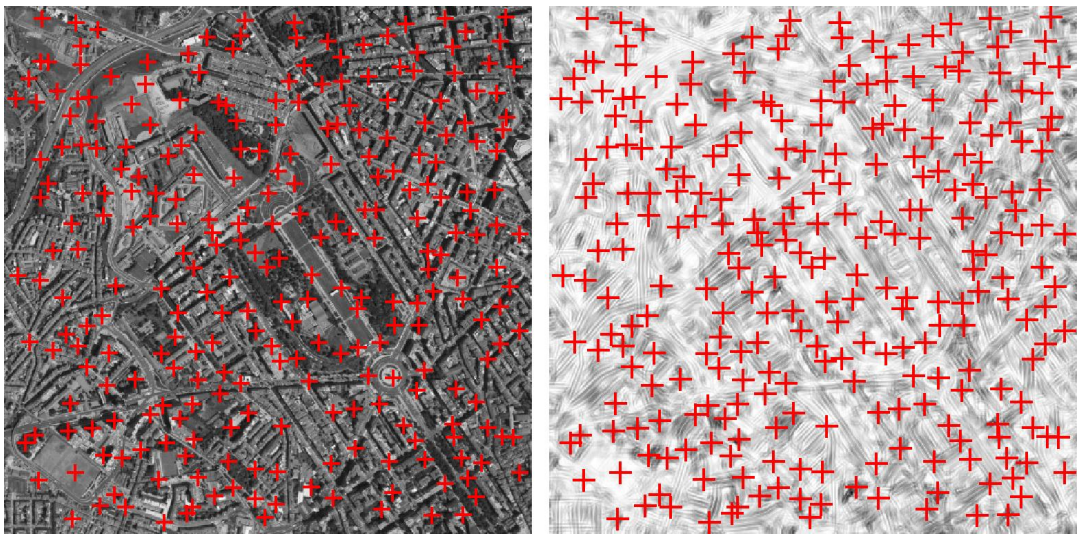


(a) “Natural” filters

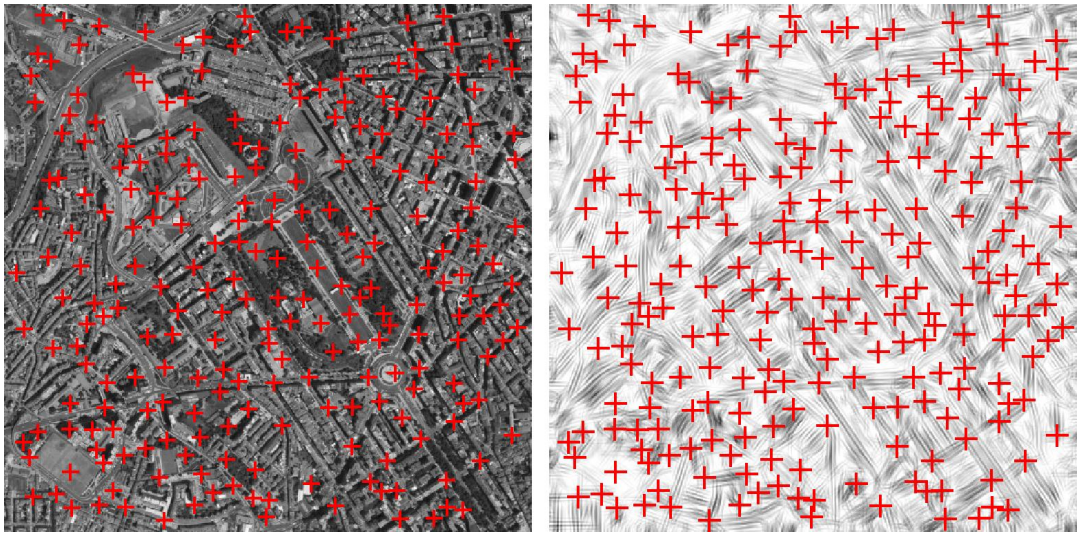


(b) “Aerial” filters

Figure 3.17.: Corner detection for a natural scene, example 2

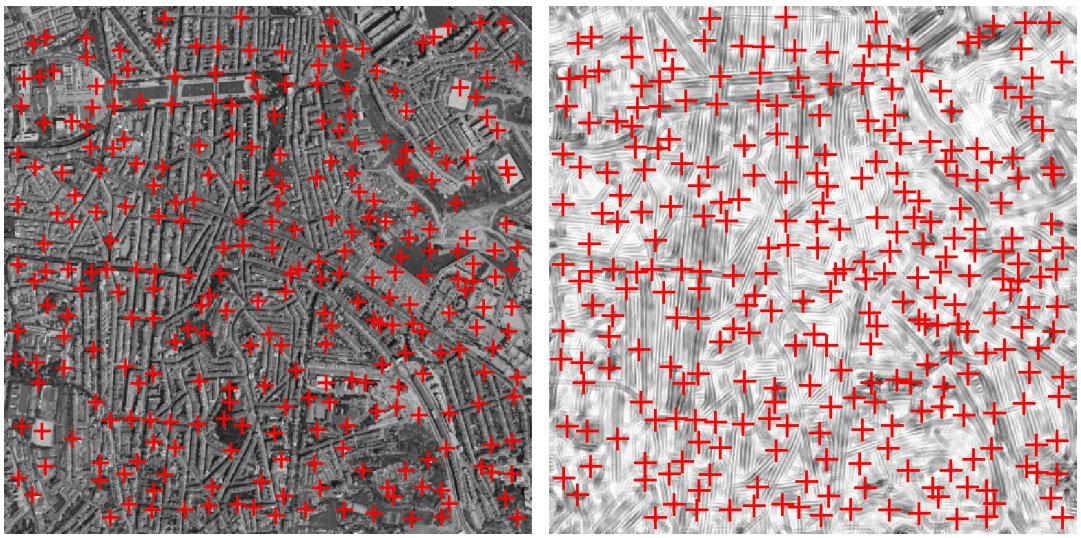


(a) “Natural” filters

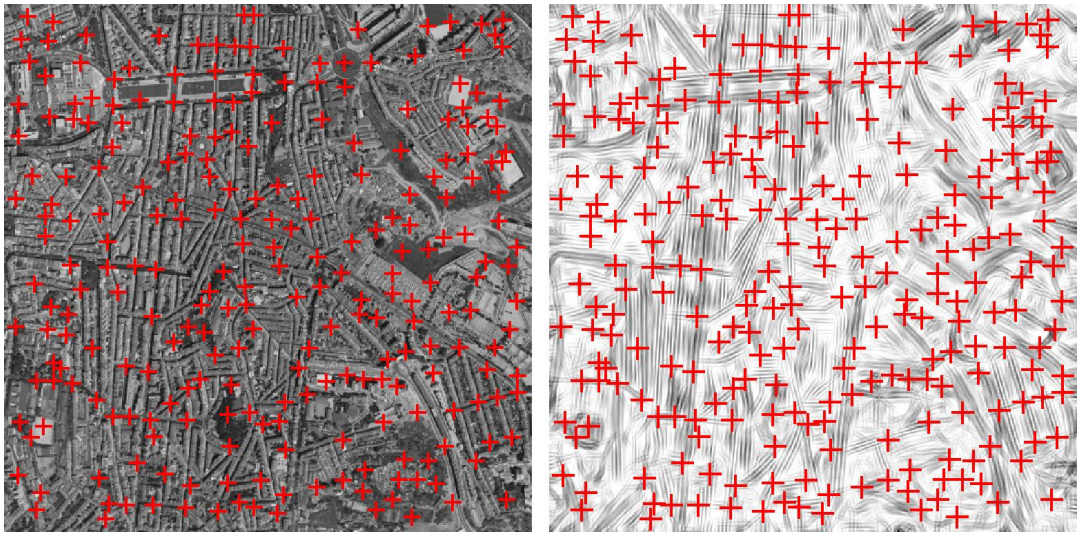


(b) “Aerial” filters

Figure 3.18.: Corner detection for an aerial image, example 1

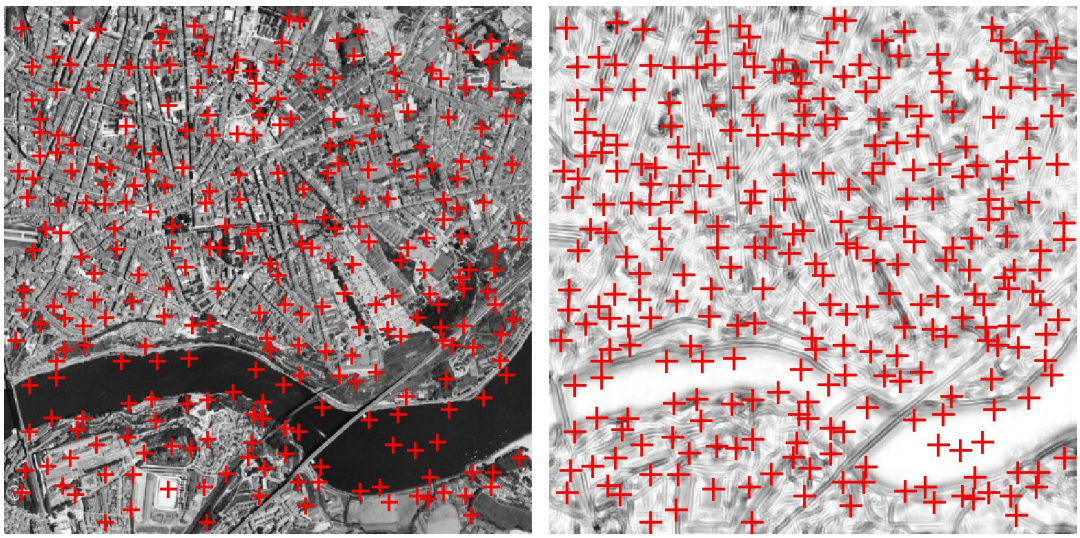


(a) “Natural” filters

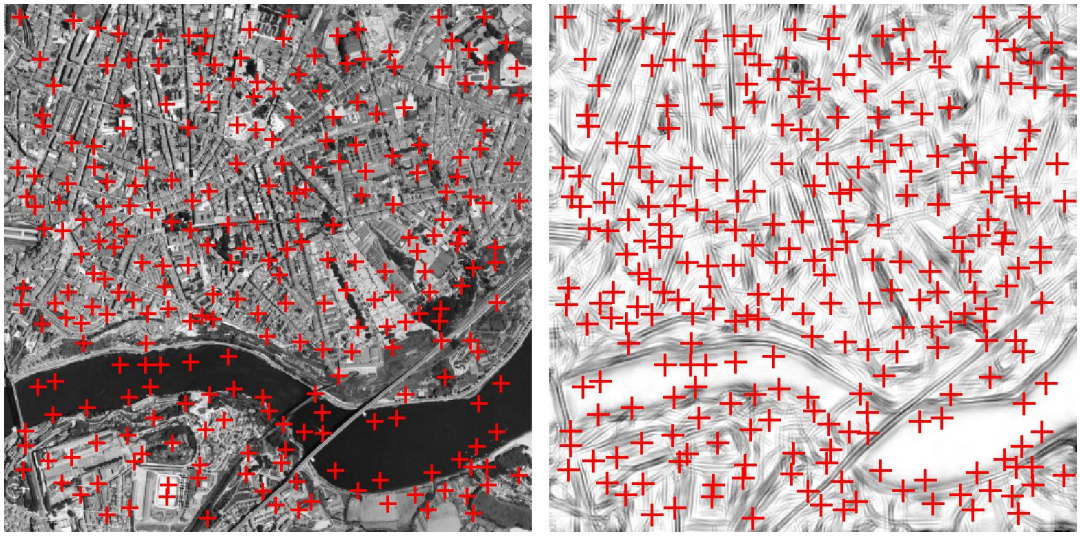


(b) “Aerial” filters

Figure 3.19.: Corner detection for an aerial image, example 2

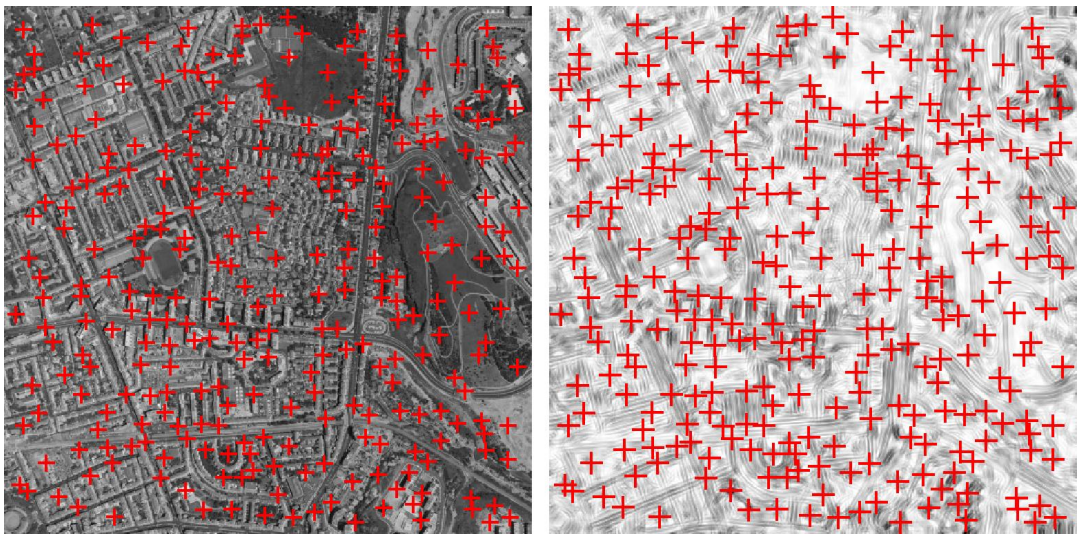


(a) “Natural” filters

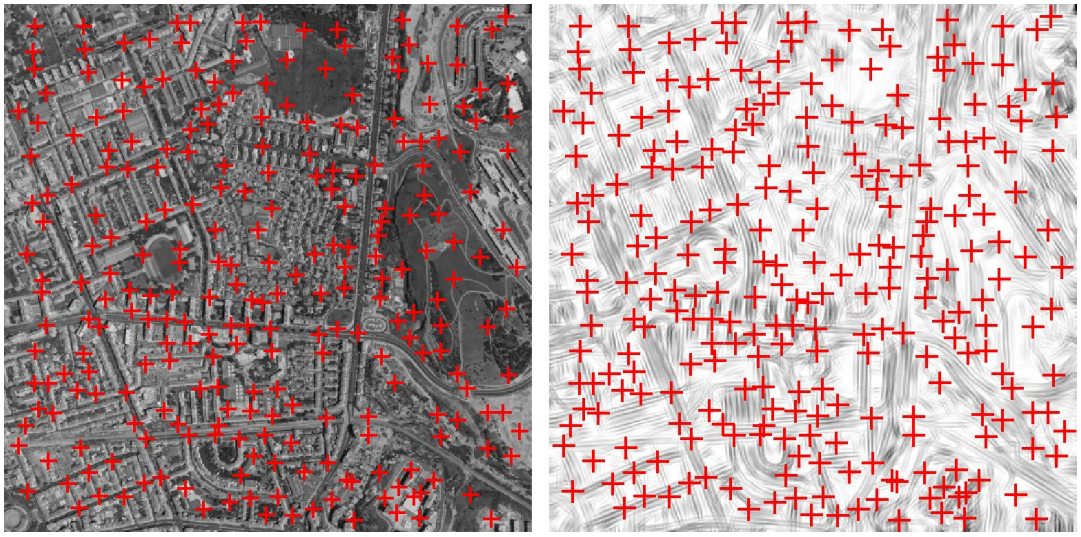


(b) “Aerial” filters

Figure 3.20.: Corner detection for an aerial image, example 3



(a) “Natural” filters



(b) “Aerial” filters

Figure 3.21.: Corner detection for an aerial image, example 4

4. Landmark Selection

In this section the results of the previous chapter will be used to overcome the drawbacks of the existing work on landmark selection sketched in section 1.1.4. First, a Point–of–Interest (PoI) detector based on the edge and junction detection models will be described. In section 4.1 an algorithm to reduce the number of junction points to an appropriate number of landmark candidate points is proposed. It assures a minimum distance between the landmarks. In section 4.2 sub–images around the detected landmark candidate points are used as landmark candidates. These are used as input for the landmark selection algorithms proposed in Gerstmayr et al. (2004a,b), which select the most distinctive landmarks.

All algorithms will only be applied to aerial images since the main goal of this thesis was to implement a preprocessing system for the landmark selection task of the flying blimp. However, all of them can be applied to other tasks like object recognition or image retrieval and are not at all restricted to aerial images.

4.1. Point–of–Interest Detection

4.1.1. Proposed Algorithm

The proposed algorithm uses the early vision models described in chapter 3 to detect edges and junctions in the image $\mathbf{I}$. The entries of the junctioness matrix $\mathbf{J}$ are transformed to image coordinates $(\mathbf{x}, \mathbf{y})$ representing junctions. Junctions which are closer to the boarder than α_L pixels are discarded. As discussed in section 3.2, it is very difficult to set an appropriate threshold α_{J}, which achieves good detection results. Additionally the features are distributed very densely over the aerial image. Thus, it is difficult to reduce the number of detected junctions by increasing α_{J}, but passing two landmarks that are too close together to the landmark selection algorithms will decrease their distinctive power. Therefore, a further reduction step is included. The landmark candidate points are selected from the set of junction points iteratively. In each step the candidate (x_m, y_m) with maximum detection strength s_m is selected. For all other junction points (x_i, y_i) a reduction factor ν for the detection strength s_i at the current iteration t is computed:

$$\nu_i = [1 - \exp\left(-\beta_{\mathrm{L}}(d_i - \gamma_{\mathrm{L}})\right)]^{+} \tag{4.1}$$

There, d_i is the Euclidean distance between the selected candidate (x_m, y_m) and the junction point (x_i, y_i). The detection strength used in the next iteration step is given by $s(t+1) = \nu_i s_i(t)$. This means that within a given range γ_{L} around the selected candidate all other detection strengths are set to zero preventing two candidate points to be too

close together. For a transition region the detection strength is only reduced to a certain degree decreasing exponentially with the distance d and depending on the choice of β_{L}. This helps to select junction points with a low detection strength which are far away from already selected candidates. The iteration stops if the number of candidate points to select n_C is reached or if the detection strength of all candidates is reduced to zero. A pseudocode notation of the algorithm is depicted in figure 4.1.

Figure 4.1.: Landmark candidate detection algorithm

Input: An image $\mathbf{I}$,
all the necessary model parameters
Output: A list of n_C coordinates $(\mathbf{x}, \mathbf{y})$
begin
$\mathbf{C}_\theta = \mathrm{ComputeComplexCellResponse}(\mathbf{I})$;
$\mathbf{J} = \mathrm{ComputeJunctions}(\mathbf{C}_\theta, \sigma_{\mathrm{J}}, \alpha_{\mathrm{J}}, \delta_{\mathrm{J}})$;
$(\mathbf{x}', \mathbf{y}', \mathbf{s}') = \mathrm{Junctionmap2Coordinates}(\mathbf{J})$;
$(\mathbf{x}'', \mathbf{y}'', \mathbf{s}) = \mathrm{CheckMargins}(\mathbf{x}', \mathbf{y}', \mathbf{s}', \alpha_{\mathrm{L}})$;
$n_{\mathrm{J}} = |\mathbf{x}''|$;
for $i = 1$ ***to*** n_c **do**
$m = \arg\max_{n_{\mathrm{J}}}(\mathbf{s})$;
if $s_m = 0.0$ **then**
$n_c = i$;
break;
end
$x_i = x''_m$;
$y_i = y''_m$;
for $j = 1$ ***to*** n_c **do**
$d = \sqrt{(x''_j - x_i)^2 + (y''_j - y_i)^2}$;
$s_j = s_j \cdot [1 - \exp(-\beta_{\mathrm{L}}(d - \gamma_{\mathrm{L}}))]^+$;
end
end
end

4.1.2. Results and Discussion

The results for some test images are shown in figures 4.2 to 4.3. All test images are sized (700×1075) pixels with a resolution of $16\,\mathrm{m}^2$ per pixel. Thus, the images cover an area of 4300×2800 meters. For all shown test images the model parameters were chosen as follows: $\sigma_{\mathrm{J}} = 5$, $\delta_{\mathrm{J}} = 5$, $\alpha_J = 0.015$, $\alpha_L = 50$, $n_C = 150$, $\beta_L = 0.25$, and $\gamma_L = 30$. This choice results in approximately 700 junction points denoted in the plots by red points, which are then reduced to $n_C = 150$ landmark candidate points denoted by yellow points. As input histogram equalized images were used and not as in previous chapters images that were contrast normalized. The equalization results in an

enhancement of the contrast between streets and other objects such as houses, leading to stronger responses of the edge detection network.

The results do not vary too much for the different test cases, thus they all will be discussed together. The selected landmark candidate points are distributed all over the images at reasonable distances between each other.

All in all most of the corners and junctions are detected at the junction detection stage. Again, there are some false positives like some detected corners in the Tagus–river in figure 4.3. The most prominent false negatives include the motorway exit close to the park in figure 4.2. It is a general problem of the model that such huge intersections are not detected very well. Since the contrast in the surrounding of these image features is very low, the resulting complex cell response is very low. The same holds for the junctioness values. Thus, these features are only very rarely selected as landmark candidate point. These findings show again the importance to somehow enhance weak but very long contours. This would ensure that all responses passed to the junction detection stage are at approximately the same magnitude.

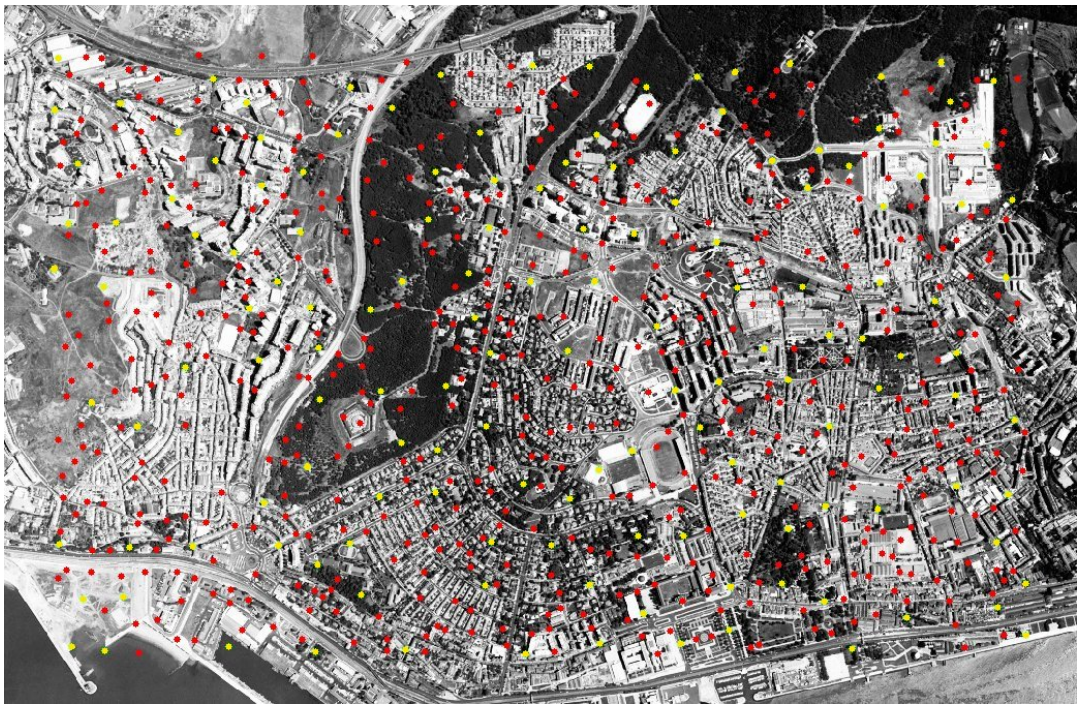
Another problem of the current solution is the response of the edge detection network to small parallel structures like separate lanes of an highway or narrow streets in certain districts. Examples include the highway close to the park in figure 4.2 or some neighborhoods in figures 4.3 and 4.4. In these areas junction points are detected which are due to fine detailed structures. Although almost no landmark candidate points are selected in these neighborhoods, it would be a great improvement of the proposed models if these responses could be suppressed and responses stretching the outlines of these districts could be enhanced. This could help to detect good landmark candidates at places where adjacent neighborhoods meet. An easy approach to suppress these responses would be to smooth the image. However, it is likely that it would be difficult to find an appropriate level of smoothing as tradeoff between the reduction of details and keeping sharp contrast transitions. A more advanced approach could be to add appropriate enhancement mechanisms and regions of inhibition to the grouping stage.

4.1.3. Conclusions

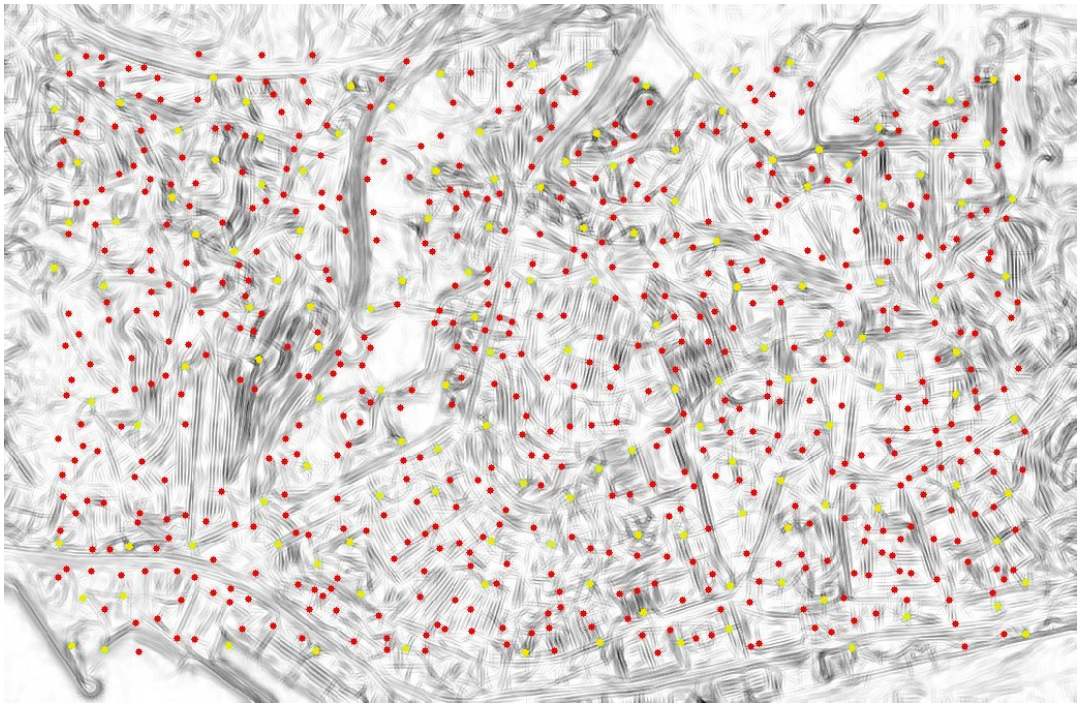
Again, the junction detection stage can be identified as the bottle–neck of the whole processing stages. Since many junction features which are very prominent for an human observer are surrounded by areas of low contrast, they are only detected weakly by the complex cell model and thus receive small junctioness values. Hopefully, these drawbacks can be solved by implementing the contour grouping mechanism as already proposed in chapter 3.

4.2. Landmark Selection

For sake of completeness the two landmark selection algorithms proposed in Gerstmayr et al. (2004a,b) are briefly recapitulated here. They work on the landmark candidates, i.e. image patches around the detected landmark candidate points, and select the most

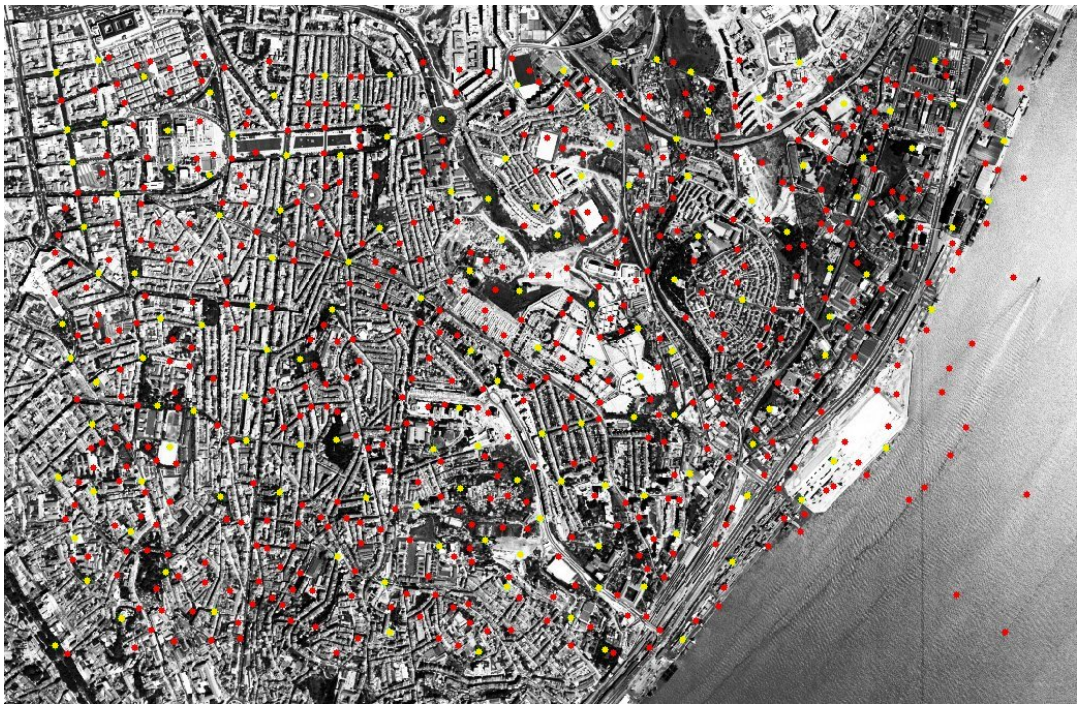


(a) Stimulus

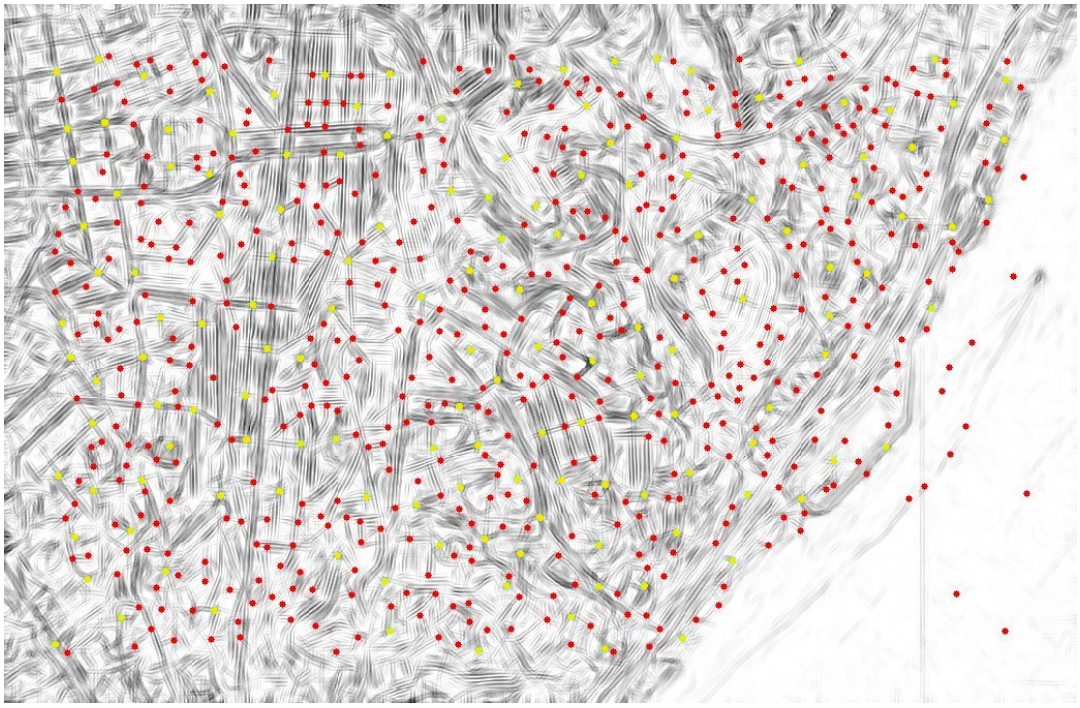


(b) Complex Cell Response

Figure 4.2.: Landmark candidate detection, example 1. Red dots mark junction points, yellow dots mark selected landmark candidates.

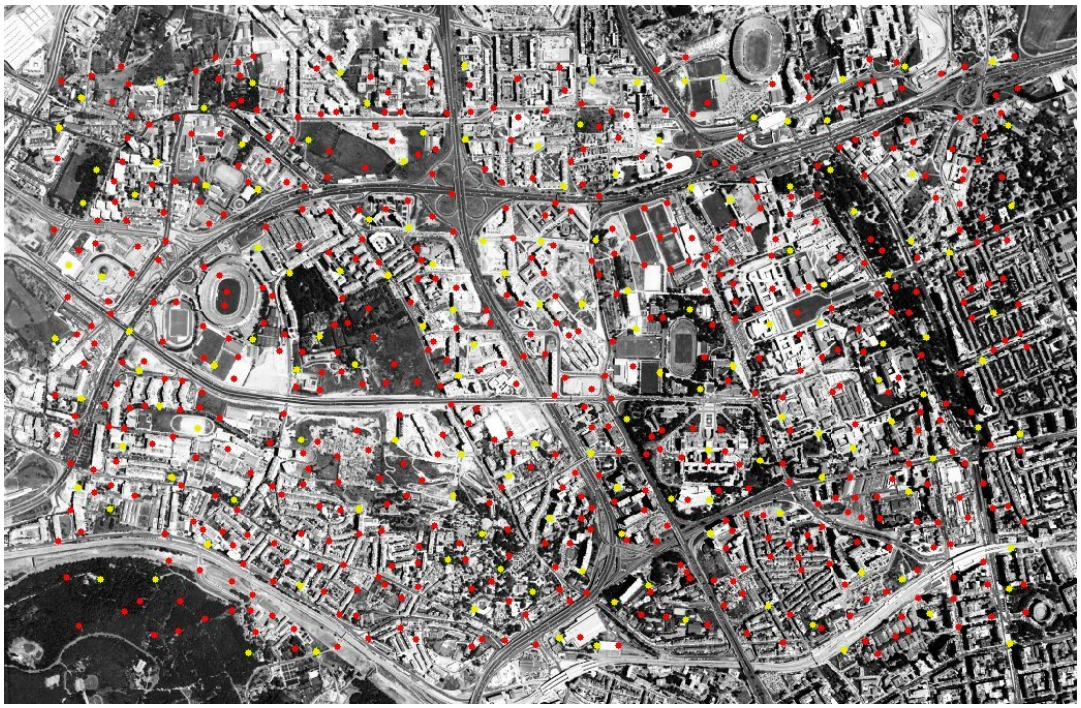


(a) Stimulus

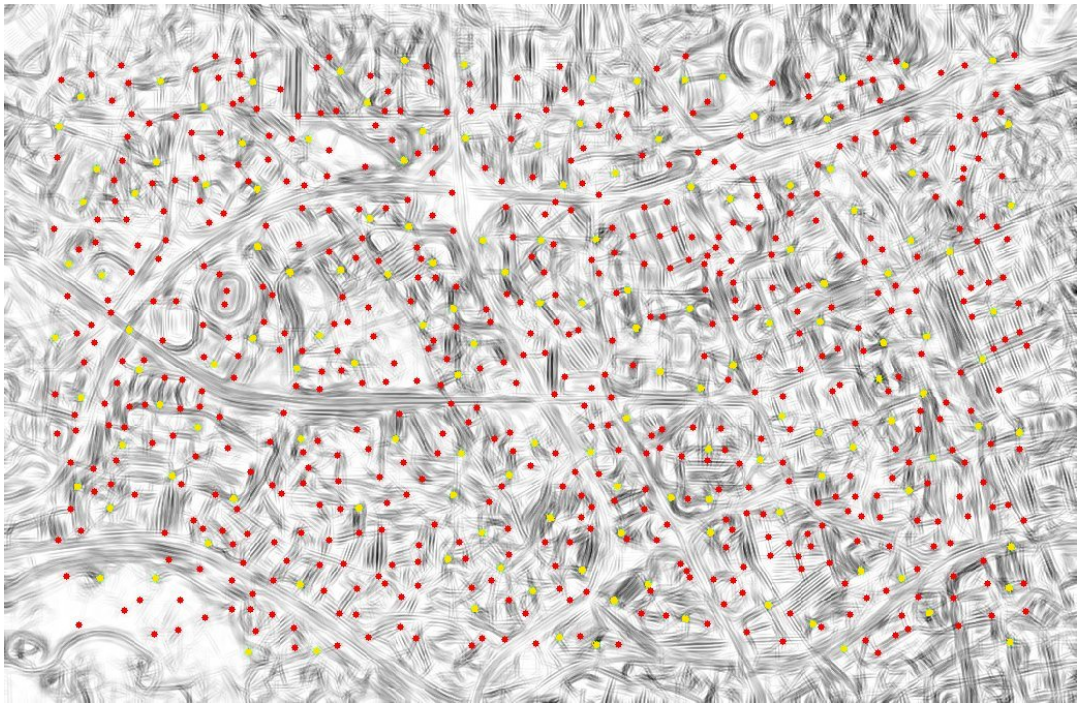


(b) Complex Cell Response

Figure 4.3.: Landmark candidate detection, example 2. Red dots mark junction points, yellow dots mark selected landmark candidates.



(a) Stimulus



(b) Complex Cell Response

Figure 4.4.: Landmark candidate detection, example 3. Red dots mark junction points, yellow dots mark selected landmark candidates.

distinctive candidates. Therefore, they evaluate dissimilarities in a lower–dimensional feature space.

4.2.1. The Algorithms

4.2.1.1. Profile–based Algorithm

The algorithm uses a lower–dimensional representation of the landmarks. This gains a speed up of computation, but also a generalization if information is discarded in the dimensionality reduction step. Here, PCA and, additionally to Gerstmayr et al. (2004a,b), ICA are used as methods to reduce dimensionality. However, as the algorithm only compares pairwise image dissimilarities other dimensionality reduction methods such as Linear Discriminant Analysis (Duda et al., 2001), Factor Analysis (Duda et al., 2001), projection pursuit (Hyvärinen, 1999b), or nonlinear or kernel–based extensions of these methods (Müller et al., 2001) can be used. A very interesting technique is the tensor–rank principle, which exploits the spatial relation of pixels as it does not reshape the sub–view to a vector. Applications to recognition tasks yields a high recognition rate and an high stability against occlusions and image noise (Shashua and Levin, 2001; Rupar et al., 2002)

The profile–based algorithm first extracts sub–windows sized $(\delta_\mathrm{L}, \delta_\mathrm{L})$ around the landmark candidate points $(\mathbf{x}, \mathbf{y})$ and transforms them to vectors. The lower dimensional representation is then obtained as described in sections A.1 and A.2 for PCA and ICA, respectively. The dimensionality k is controlled by the proportional variance τ_L. It is determined as the smallest $k; 1 \leq k \leq n_C$ for which

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n_C} \lambda_i} \geq \tau_L \tag{4.2}$$

holds.

The algorithm computes pairwise image dissimilarities $D_{i,j}$ between all the n_C landmark candidates. As dissimilarity function the normalized SSD between two zero–mean vectors $\mathbf{g}_1, \mathbf{g}_2 \in \mathbb{R}^{k \times 1}$ is used, which gives dissimilarity values in the range $[0, 1]$:

$$\mathrm{dist}\,(\mathbf{g}_1, \mathbf{g}_2) = \frac{\sum_{i=1}^{k} (\mathbf{g}_1(k) - \mathbf{g}_2(k))^2}{2(\|\mathbf{g}_1\|^2 + \|\mathbf{g}_2\|^2)}. \tag{4.3}$$

Then, the so–called distance profile vector $\mathbf{p}$ is computed by averaging over the dissimilarities of a landmark j to all other landmarks:

$$p_j = \frac{1}{n_C - 1} \sum_{i=1, i \neq j}^{n_C} D_{i,j}. \tag{4.4}$$

As $D_{i,j}$ is close to unity if the landmarks $\mathbf{g}_i$ and $\mathbf{g}_j$ are very dissimilar, p_j will be close to unity if landmark $\mathbf{g}_j$ is very dissimilar to all other candidates and therefore is considered to be a good landmark. Finally, the candidates are sorted with respect to the profile values p_i and the n_L best landmarks are selected and returned. The algorithm is summarized in figure 4.5.

Figure 4.5.: Pseudocode notation of profile–based landmark selection

Input: An image $\mathbf{I}$,
the coordinates of landmark candidates $(\mathbf{x}, \mathbf{y})$,
all necessary model parameters
Output: The landmark coordinates $(\mathbf{x}', \mathbf{y}')$
begin
$\mathbf{X} = \text{ExtractCandidates}(\mathbf{I}, \mathbf{x}, \mathbf{y}, \delta_L)$;
$\mathbf{G} = \text{ComputeLowDimRepresentation}(\text{method}, \tau_\text{L})$;
for $i = 1$ ***to*** n_C **do**
for $j = i$ ***to*** n_C **do**
$D_{i,j} = D_{j,i} = \text{dist}(\mathbf{g}_i, \mathbf{g}_j)$;
end
end
end
for $i = 1$ ***to*** n_C **do**
$p_i = \frac{1}{n_C - 1} \sum_{i=1}^{n_C} D_{i,j}$;
end
$[\mathbf{v}, \mathbf{l}] = \text{SortDescending}(\mathbf{p})$;
$\mathbf{x}' = \mathbf{x}(\mathbf{l}(1 : n_\text{L}))$;
$\mathbf{y}' = \mathbf{y}(\mathbf{l}(1 : n_\text{L}))$;

4.2.1.2. IPCA–based Algorithm

This algorithm first computes a small eigenspace of dimensionality n that will in later steps be updated. The computation is done by the standard PCA batch method using the start candidates $(\mathbf{x}_s, \mathbf{y}_s)$. Usually the sub–windows around the strongest four candidate points are used. The candidate detection mechanism overcomes also the drawbacks of Gerstmayr et al. (2004a,b), where it was difficult to determine the start candidates. The next landmarks are selected iteratively. In each step, the norm $q_i = \|\mathbf{r}_i\|$ of the residue $\mathbf{r}_i$ is computed for each of the remaining landmarks giving a measure of how good each remaining landmark can be expressed in the already existing eigenspace. The landmark with the greatest q_i can be expressed worst. Therefore it is as dissimilar as possible to all other already selected landmarks and will be added to the eigenspace in the updating step as described in section A.1.2. The updating step increases the dimensionality n. Of course, in each step the sets of selected and remaining candidates have to be updated, too. The algorithm is outlined in figure 4.6.

4.2.2. Results and Discussion

The results of the landmark selection for the candidates selected in the previous section are shown in figures 4.7 to 4.9. Red, cyan, and yellow circles denote landmarks selected by applying the PCA–profile, the ICA–profile, and the IPCA–based algorithm, respectively. All selections were done for $n_L = 20$, $\delta_L = 50$, $\tau_L = 0.75$, and 4 starting landmarks for the IPCA–based algorithm. The landmark selection algorithms were applied to the

Figure 4.6.: Pseudocode notation of IPCA–based landmark selection

Input: An image $\mathbf{I}$,
the coordinates of landmark candidates $(\mathbf{x}_c, \mathbf{y}_c)$,
a set of landmark candidates $(\mathbf{x}_s, \mathbf{y}_s) \subseteqq (\mathbf{x}, \mathbf{y})$,
all necessary model parameters
Output: The landmark coordinates $(\mathbf{x}', \mathbf{y}')$
begin
$\mathbf{x}_r = \mathbf{x}_c \backslash \mathbf{x}_s$;
$\mathbf{y}_r = \mathbf{y}_c \backslash \mathbf{y}_s$;
$\mathbf{X} = \text{ExtractCandidates}(\mathbf{I}, \mathbf{x}_r, \mathbf{y}_r, \delta_L)$;
$\mathbf{S} = \text{ExtractCandidates}(\mathbf{I}, \mathbf{x}_s, \mathbf{y}_s, \delta_L)$;
$(\bar{\mathbf{x}}, \mathbf{T}, \mathbf{\Lambda}, n) = \text{ComputeEigenspace}(\mathbf{S})$;
while $n \leq n_L$ **do**
$\mathbf{q} = \mathbf{0} \in \mathbb{R}^{n_C - n}$;
for $i = 1$ ***to*** $n_C - n$ **do**
$\mathbf{g} = \mathbf{T}^\top (\mathbf{x}_i - \bar{\mathbf{x}})$;
$q_i = \|(\mathbf{T}\mathbf{g} + \bar{\mathbf{x}}) - \mathbf{x}_i\|$;
end
$m = \arg\max_i(\mathbf{q})$;
$\mathbf{X} = \mathbf{X} \backslash \mathbf{x}_m$;
$\mathbf{x}_s = \mathbf{x}_s \cup \mathbf{x}_r(m); \mathbf{y}_s = \mathbf{y}_s \cup \mathbf{y}_r(m)$;
$\mathbf{x}_r = \mathbf{x}_r \backslash \mathbf{x}_r(m); \mathbf{y}_r = \mathbf{y}_r \backslash \mathbf{y}_r(m)$;
$(\bar{\mathbf{x}}, \mathbf{T}, \mathbf{\Lambda}, n) = \text{UpdateEigenspace}(\mathbf{x}_m, (\bar{\mathbf{x}}, \mathbf{T}, \mathbf{\Lambda}, n))$;
end
$\mathbf{x}' = \mathbf{x}_s; \mathbf{y}' = \mathbf{y}_s$;
end

image as well as to the pooled complex cell responses.

In comparison to Gerstmayr et al. (2004b) only representative results are shown. A detailed analysis how the choice of the different model parameters influences the results of the selection is far beyond the scope of this diploma thesis.

4.2.2.1. Profile–based Algorithm

The selected parameters usually lead to a dimensionality of approximately 50 dimensions for the profile–based algorithm if applied to the image, and of approximately 45 dimensions if applied to the complex cell response. Using PCA as well as ICA as lower dimensional representation, the algorithm detects distinctive landmarks if applied to image data as well as if applied to the pooled complex cell response. The selected areas usually include areas of high contrast which are frequently characterized by structures or patterns that are highly characteristic for that place. These patterns are formed by streets, buildings, parks or forests if the algorithm is used to process image data or by complex edge responses if applied to the pooled complex cell response. Although the

algorithm selects often the same landmarks if the lower–dimensional representation is computed using ICA or PCA, it is not yet clear which method prefers which features. This would give a deeper understanding why a certain landmark was only chosen by one of the methods.

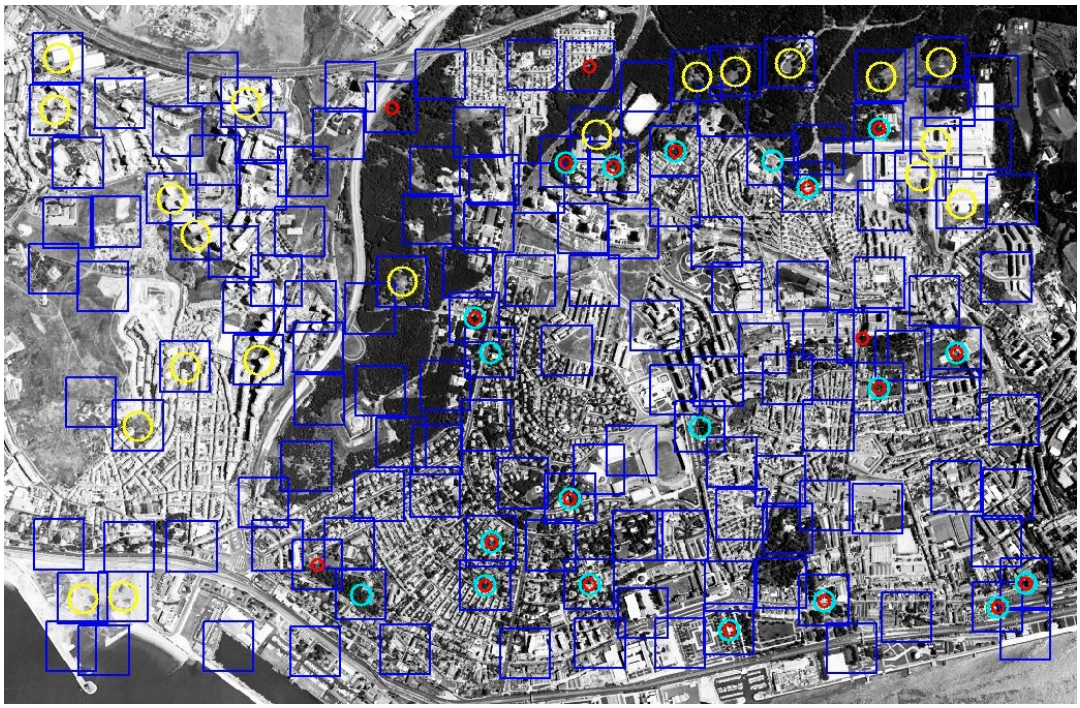
4.2.2.2. IPCA–based Algorithm

It strikes out that the IPCA–based algorithm very often detects other landmarks than the profile–based algorithm. The selected landmarks are also distinctive and are also characterized by a specific pattern. Nevertheless, it seems that the algorithm prefers landmarks in very bright or very dark image regions. Examples include various landmarks in parks that only differ in streets or arrangements of buildings such as in figures 4.7(a) or 4.9(a), or the selected landmarks in the left part of figure 4.7(a).

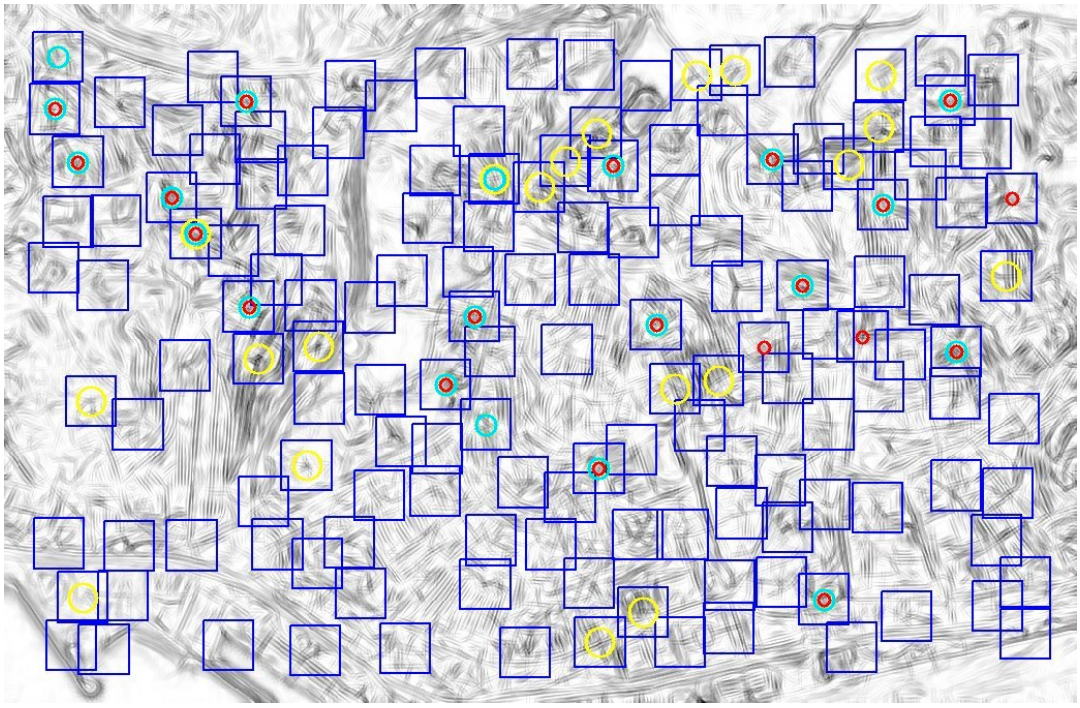
4.2.3. Conclusions

The results reveal that the proposed landmark selection algorithms can select distinctive landmarks from the detected landmark candidates. Between the IPCA–based and the profile–based algorithm there are obvious differences in the selected landmarks: the IPCA–based algorithm detects landmarks in bright or dark areas, whereas the profile–based algorithm selects landmarks in areas of high contrast. For both algorithms the landmarks are characterized by a specific pattern formed by streets or buildings. If the subspace used in the profile–based algorithm is computed using ICA or PCA, there are slight differences in the selected landmarks. However, there is no obvious criterion which method preferably selects which landmarks.

In future work the reliability of the landmarks has to be analyzed. Further it has to be proofen that the selected landmarks can be used to build a topological map allowing robust navigation.

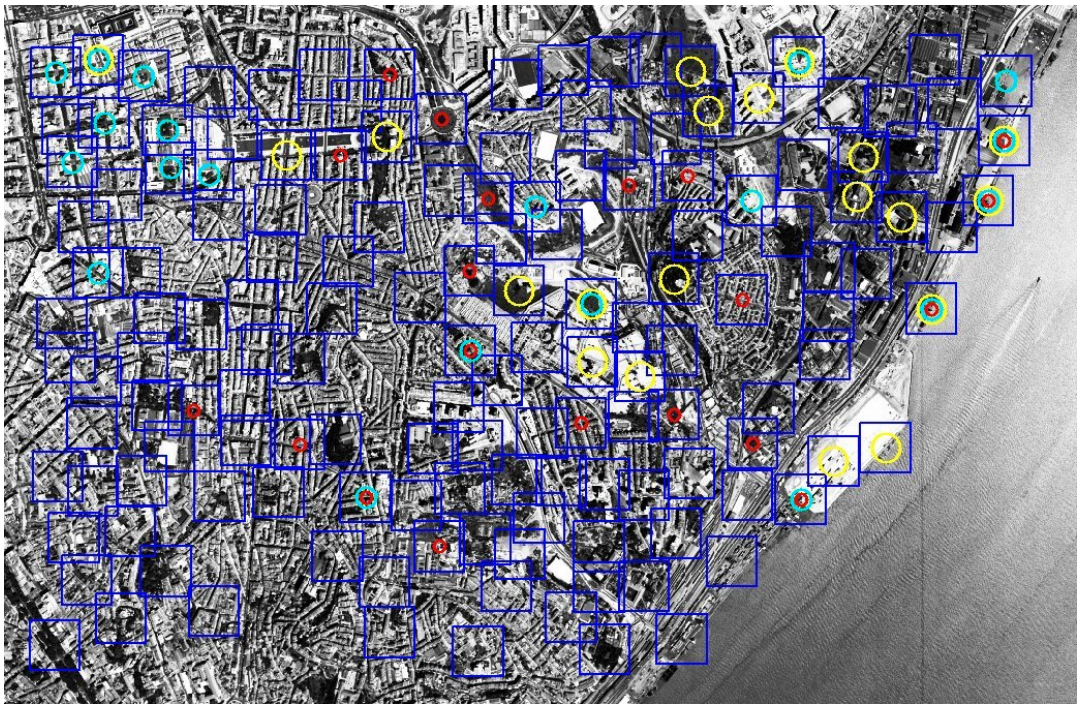


(a) Selection on image data

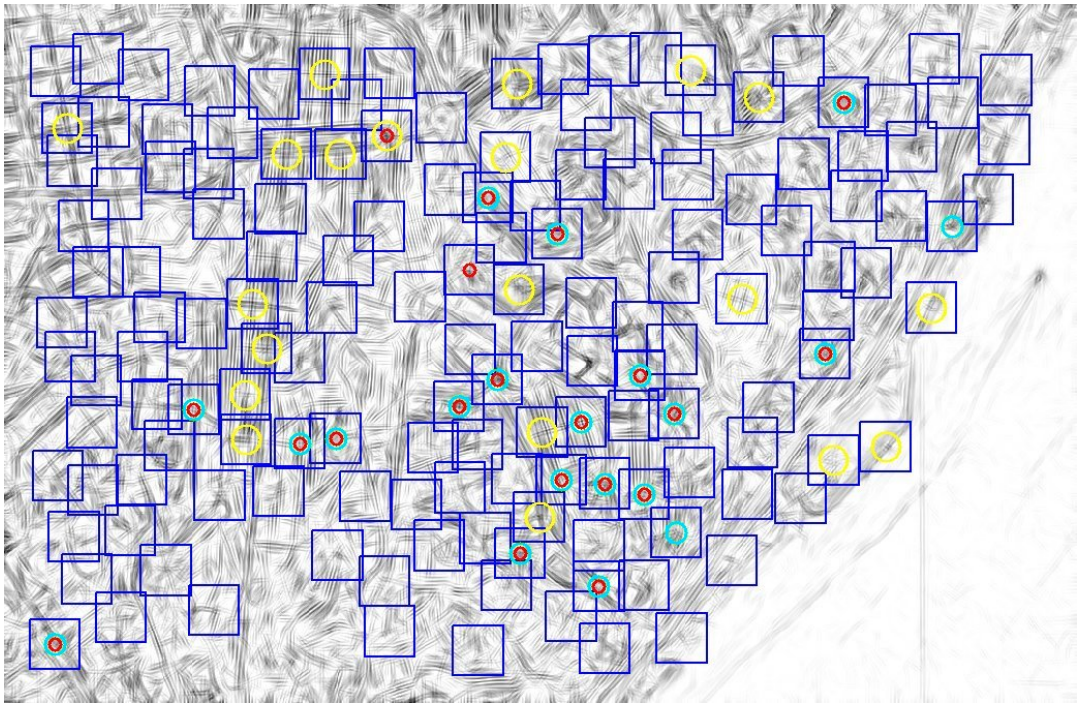


(b) Selection on pooled complex cell response

Figure 4.7.: Landmark selection, example 1. Red, cyan, and yellow circles denote the landmarks selected by the PCA–profile, the ICA–profile, and the IPCA–based method. Empty blue squares denote the landmark candidates.

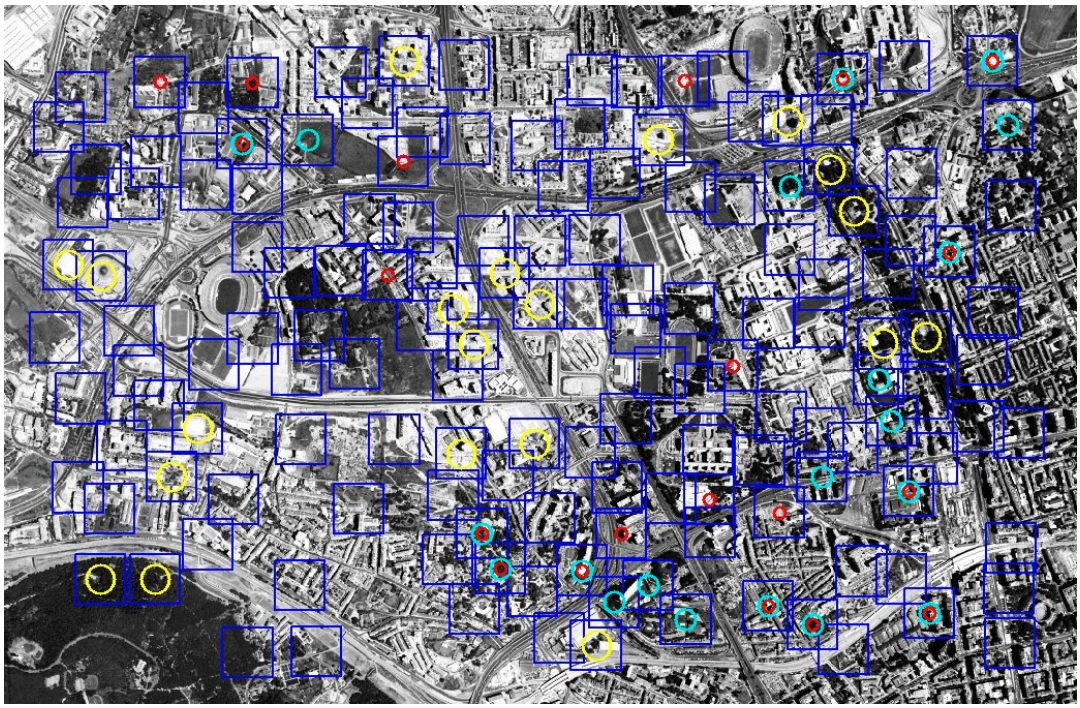


(a) Selection on image data

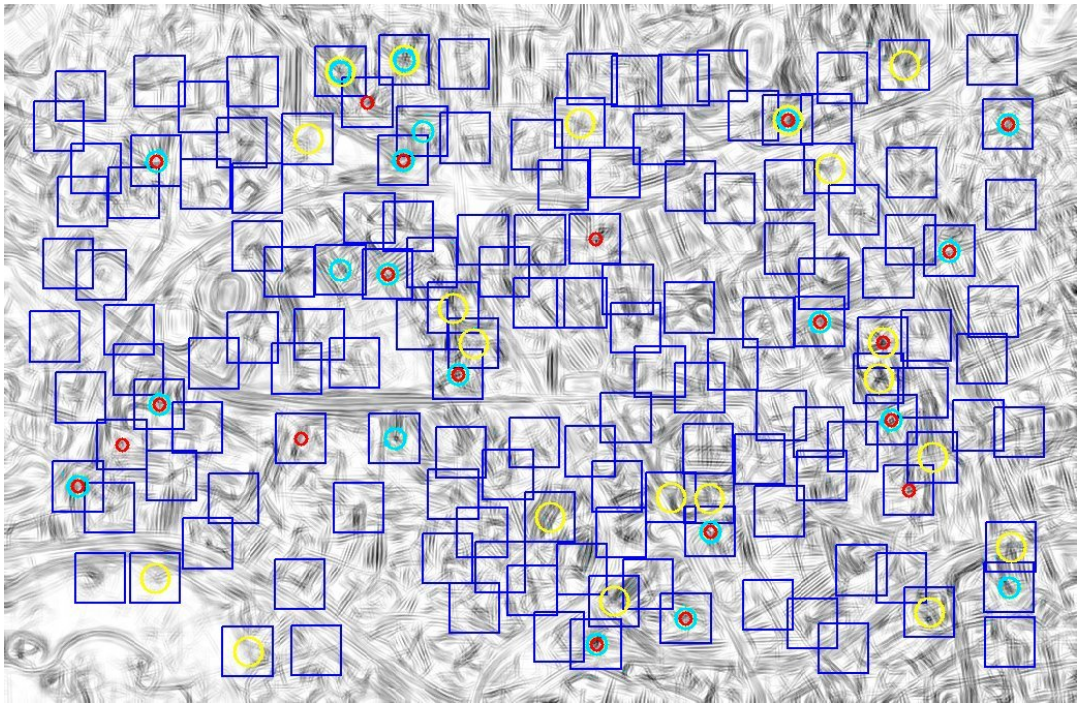


(b) Selection on pooled complex cell response

Figure 4.8.: Landmark selection, example 2. Red, cyan, and yellow circles denote the landmarks selected by the PCA–profile, the ICA–profile, and the IPCA–based method. Empty blue squares denote the landmark candidates.



(a) Selection on image data



(b) Selection on pooled complex cell response

Figure 4.9.: Landmark selection, example 3. Red, cyan, and yellow circles denote the landmarks selected by the PCA–profile, the ICA–profile, and the IPCA–based method. Empty blue squares denote the landmark candidates.

4.3. Chapter Summary and Future Work

The main conclusion from the results is that the candidate as well as the landmark selection stage work together. They now form a complete landmark selection system following the usual two–step approach reported in related literature. This helps to overcome the drawbacks reported in Gerstmayr et al. (2004a,b). Although a direct comparison between the results presented in this work and in prior work is difficult, the results of Gerstmayr et al. (2004a,b) could have been reproduced insofar as the algorithms also selected distinctive landmarks characterized by specific patterns if used in conjunction with the proposed preprocessing system.

Another aspect is that although the landmark candidate detection stage does no longer restrict the candidates to be located at given grid positions, almost none of the selected landmarks is characterized by such outstanding patterns as those formed by highways or roundabouts. Nevertheless, future work has to reveal that the selected landmarks can though be used to achieve robust navigation strategies with the blimp.

4.3.1. Future Work

4.3.1.1. Towards Real Robot Experiments

Although the landmark selection stage works, there is still a long way till the algorithms proposed in this thesis can be used for autonomous navigation. The different model parameters have to be tuned, it has to be decided whether the edge or the image representation will be used for navigation, the real–time demands put strong constraints on the used algorithms and the resolution of the camera image, and the algorithms have to be linked together with the control architecture described in Metelo and Garcia (2003).

The methods described in this thesis only select the nodes of the topographical map, but there is also a mechanism needed to navigate from landmark to landmark. This can be done using some sort of visual odometer based on optic flow. One possible algorithm is sketched in Gerstmayr et al. (2004b). Other methods could include optic flow based approaches such as Iida (2003) or Ruffier and Franceschini (2005). More technically motivated approaches based on homography estimation between consecutive images include Gracias and Santos-Victor (2001) or Ferruz et al. (2004).

The selected topological map could also be used to increase the quality of position estimates in a geometrical navigation approach. If the blimp is supposed to fly from its current position to a given position, it could first approach the closest landmark. There it could reset its position estimate, because the landmark is assumed to be a place that can be approached with high accuracy. Then, it could follow the topological map resetting its odometer and position estimate at each landmark. After reaching the landmark closest to the goal position, it could finally approach it. This method should allow robust geometric navigation, which is e.g. needed for surveillance tasks.

4.3.1.2. Robustness and Reliability Evaluation

Although real robot experiments or at least experiments using a detailed simulator are extremely important, an in–depth analysis of the selected landmarks can be helpful to show that the selection mechanisms can increase the robustness and reliability of the navigation. In Gerstmayr et al. (2004a,b) such a method was introduced.

The idea behind the reliability measure is that the blimp is assumed to be at exactly the same position as when the landmark was selected except for deviations in either the altitude, the orientation, or the position. This results in little deviations between the camera view remembered as landmark and the current camera view. By iteratively increasing the simulated deviation one can compute a limit for which a correct localization is still possible but will fail if the deviation is increased more. The idea is sketched in figure 4.10.

Localization in a topological map is nothing else then a nearest neighbor search in the feature space of selected landmarks. The localization is correct if the current landmark is selected as the current camera view's nearest neighbor, otherwise the localization will fail. However, the method does not take into account whether the analyzed landmark $\mathbf{g}_i$ is very dissimilar to all others or whether it is very similar to at least one other landmark. Therefore, a maximal image dissimilarity ϵ was introduced that is defined as half the minimal dissimilarity between the considered landmark $\mathbf{g}_i$ and all others:

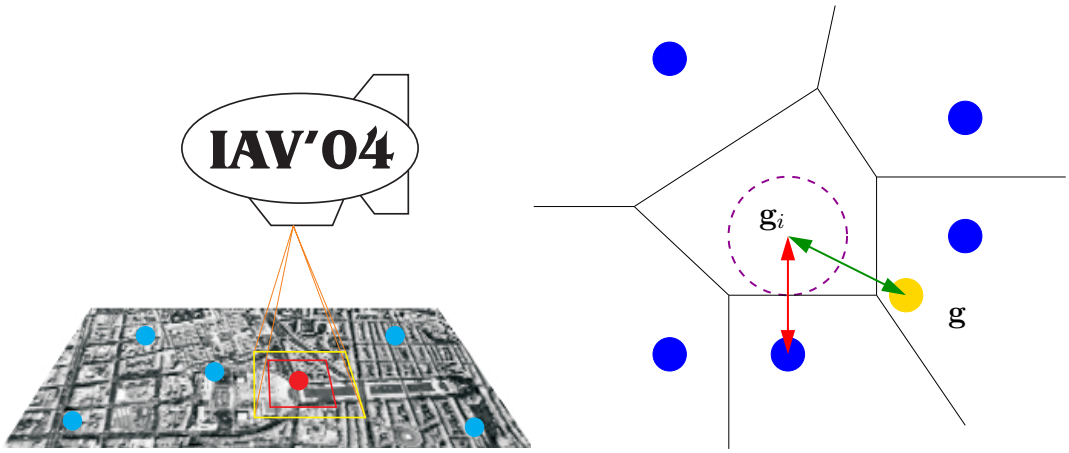
$$\epsilon = \frac{1}{2} \min_{j=1,\ldots,n_L, j\neq i} \operatorname{dist}(\mathbf{g}_i, \mathbf{g}_j) . \tag{4.5}$$

If $\mathbf{g}_i$ is a good landmark ϵ is assumed to be large. If it has at least one close neighbor, ϵ will be small.

Therefore, the method iteratively increases the deviation until the image dissimilarity between the current camera view $\mathbf{g}$ and the considered landmark $\mathbf{g}_i$ exceeds ϵ. An advantage of the method is that it also takes into account how much the image changes for a certain deviation. For the same ϵ, a landmark is rated more reliable if there are huge deviations needed to exceed ϵ. A landmark for which small deviations are sufficient to exceed the threshold will be rated unreliable.

However, the reliability measurement gives one value for each measured parameter and the parameters do not tell how good a landmark is in comparison to an average landmark. In order to overcome this drawback, one can compute the average limits for all the parameters taken into account. For n_C landmark candidates and n_L landmarks to select, there exist $\binom{n_C}{n_L}$ possibilities to select a set of n_L landmarks. As the number is very large, one can randomly sample many selections, measure their reliability, and estimate the mean of the possible deviations. For each parameter one can then use the ratio between the measured limits of deviation and the average limits. These ratios can easily be combined to a single score because they are independent from the dimension of the parameters.

The parameters to measure can include stability against deviations in the blimp's altitude, orientation, yaw– and pitch–movements, position, the size of the catchment



(a) The blimp's altitude deviates slightly from it's altitude when the landmark was selected, thus the current view and the remembered view are different.

(b) In the eigenspace the current camera view $\mathbf{g}$ and the considered landmark $\mathbf{g}_i$ are also different. A correct localization is possible if $\mathbf{g}$ is closer to $\mathbf{g}_i$ then to any other landmark, i.e. if $\mathbf{g}$ is within the Voronoi–cell around $\mathbf{g}_i$. However, a stronger criterion is used for the reliability evaluation: the deviation is step by step increased and the maximal deviation is determined for which the current camera view's projection $\mathbf{g}$ is still within the sphere of radius ϵ around $\mathbf{g}_i$.

Figure 4.10.: Sketch of the reliability measure

area, the size of the zero–vector area,[1] global and local changes in image brightness or contrast, camera noise. This score could also be used for a landmark selection algorithm that searches for the selection with the highest score. In a first approximation the aerial image would be sufficient to run this evaluation. For a more accurate estimation of a landmarks reliability, a more detailed simulation building on a 3D–model of an urban area is needed. Such a 3D–model simulation could for example build on the Open–Source flight simulation environment Flight–Gear.[2] Real–world experiments would take too long and are thus not an option.

[1] defined in Hübner (2005) as the area where the localization method predicts the robot is at a known position although it is not exactly at the known position. The size is influenced by the resolution of the camera system as well as by the the local image variance.

[2] See: `http://www.flightgear.org`

5. Final Conclusions

The estimation of the receptive fields of cortical simple cells in section 2.2 revealed significant differences between receptive fields estimated for different types of environments. From the results of this statistical comparison, class–specific receptive fields were derived that are optimally tuned to the statistical properties of the input images and cover implicit knowledge about the environment.

In section 3.1 the class–specific receptive fields are used in conjunction with a simple cell model for edge detection. The best edge representation is obtained for processing images with the appropriate class–specific receptive fields. In that case the representation looks most natural or contains the least amount of jitter. Based on the complex cell response a corner and junction detection stage is proposed in section 3.2, which, up to now, is the main bottle–neck of the entire system. This is due to the fact that the edge representation provided by the edge detection stage still contains a lot of responses and that responses are contrast dependent. One way to reduce the number of detected junction points would be to restrict them to images of actual street junctions. This could be achieved by including a contour completion algorithm, but the validation of the algorithms requires future work.

The detected junctions are then used in a further processing stage, the landmark candidate detection stage, described in section 4.1. Due to the dense distribution of junction features, additional distance constraints have been added in order not to select two candidates which are too close together. The detected candidates are then passed to a landmark selection algorithm, which selects the most dissimilar landmarks. The results of section 4.2 reveal that all landmarks are distinctive and are often characterized by a unique pattern formed by streets, buildings, or forests.

The whole work is based on a couple of assumptions layed out in the introduction. The first one is the edge representation that was chosen in order to reduce the information contained in aerial images to the key information and to facilitate further processing. This assumption is motivated by findings in neurophysiology that contour detection plays an important role in the early visual system of humans. It is further strengthened by findings in psychology that lines and contours play an important role in the segmentation of a city. Especially if processed with filters tuned to the statistics of aerial images, the underlying edge structure is detected well. However, in some regions such as neighborhoods with very narrow streets, the computed representation still contains too much information that should be reduced to facilitate further processing.

Another key aspect of the work was to always take the statistics of the environment into account and to follow the principles of neural information processing. When filtering the images with kernels derived from their own or a different image class, different results are obtained as can be seen from visual inspection. However, a detailed analysis that

can reveal the advantages of taking the statistical properties of the environment into account, also requires further work.

A further question to discuss concerns the definition of a landmark. Here, landmarks are simply distinctive image patches. They are selected without any object knowledge only according to their dissimilarity in the image space or after edge detection. This definition of a landmark is very close to the snapshot model discussed in the insect literature. It is assumed that insects use a more or less unprocessed snapshot of their surrounding for navigation. In contrast, human observers often suggest to select large structures such as roundabouts or highway exits, which are normally not found by the algorithm. This choice might be influenced by semantical information about these structure or by their relevance for human navigation.

Although many questions still need to be answered, this thesis hopefully can contribute to the general understanding of sensory systems, how these systems are optimized to their particular environment, and how these links between the sensory system and the environment can be used to achieve robust navigation strategies.

A. Mathematical Methods

In this chapter a short introduction to the most important mathematical methods used in this thesis shall be given. First PCA, ICA, and Fourier Transformation will be introduced. These methods are important tools for data analysis because they transform the data to another representation allowing a different view to the same data. Then, a brief outline of the Kolmogorov–Smirnov–Test and on error bounds for classification tasks will be given. Finally, Evolution Strategies as an optimization method will be described.

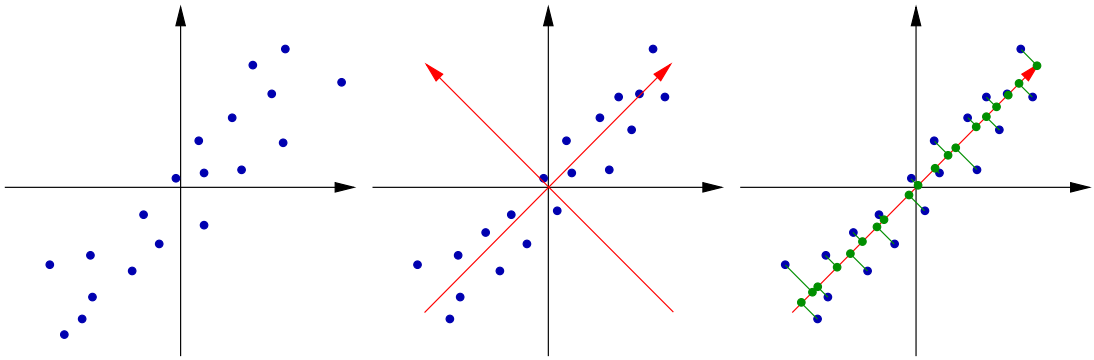
A.1. Principal Component Analysis

PCA got an established method in computer vision as well as for finding patterns or reducing dimensionality in high dimensional datasets. It has been extensively used for face and object recognition (Turk and Pentland, 1991; Murase and Nayar, 1995; Leonardis and Bischof, 2000) and for robot localization and navigation (Jogan and Leonardis, 2000; Gaspar et al., 2000; Winters and Santos-Victor, 2002; Vasallo et al., 2002). The following overview over PCA is based on Murase and Nayar (1995) and Rencher (2002). For a comprehensive overview on the advantages and disadvantages of PCA for computer vision see Gerstmayr et al. (2004b).

The idea behind PCA for a set of observations is to compute an orthogonal basis of eigenvectors for the training set, so that the origin of the new coordinate system equals the average observation and the first axis points into the direction of the dataset's greatest variation, the second axis points into the direction of the second greatest variation, and so on. Additionally, if one of the last axes covers only little information, dimensionality can be reduced by projection to the remaining axes. These basic ideas of PCA are also depicted in figure A.1. Although for an exact representation often a large set of eigenvectors is needed, only few eigenvectors are sufficient to capture significant characteristics of the observation.

A.1.1. Computing the Eigenspace

In case the observations are images, all the n images have to be reshaped to m–dimensional column vectors $x_i, i = 1, \ldots, n$, where m is the number of pixels in each image. All observations are now represented by a $(m \times n)$ matrix $\mathbf{X}$. In the next step, the average observation $\bar{\mathbf{x}}$ is subtracted from each observation leading to the zero–mean



(a) The zero–mean data set with the original coordinate axis

(b) The data set with rotated axis, so that the first axis covers the greatest variance and the second axis is perpendicular covering the second greatest variance.

(c) The data set with reduced dimensionality. The points are projected to the first axis

Figure A.1.: PCA in a nutshell

observation matrix $\tilde{\mathbf{X}}$ with columns

$$\tilde{\mathbf{x}}_i = \mathbf{x}_i - \bar{\mathbf{x}} \tag{A.1}$$

$$= \mathbf{x}_i - \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i. \tag{A.2}$$

Then the covariance matrix of $\tilde{\mathbf{X}}$ is computed:

$$\mathbf{C} = \operatorname{cov}\left(\tilde{\mathbf{X}}\right) \tag{A.3}$$

$$= \frac{1}{n-1}\tilde{\mathbf{X}}\tilde{\mathbf{X}}^{\top}. \tag{A.4}$$

Solving the EVD for $\mathbf{C}$ leads to a diagonal matrix of eigenvalues $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and the associated eigenvectors $\mathbf{e}_i \in \mathbb{R}^{n\times 1}$ with $i = 1, \ldots, n$.

The eigenvectors belonging to the non–zero eigenvalues span a basis with at most $\kappa := \min(m, n)$ dimensions because the rank of $\mathbf{C}$ can be smaller than κ and the matrix has at most κ non–zero eigenvalues. Since $\mathbf{C}$ is positive definite each eigenvalue λ_i is positive and is a measure of the variance in the data set covered by $\mathbf{e}_i$. Up to that point PCA is nothing else than a rotation of the existing basis so that the first axis covers the greatest variation, the second axis covers the second greatest variation, and so on. The axis of the new basis are the eigenvectors.

To reduce the dimensionality of the eigenspace a dimension k with $1 \leq k \leq \kappa$ has to be chosen. Determing the number of dimension to use is a rather critical step in

applying PCA and many different methods have been proposed in the literature. For this work the proportional variance τ_k, defined by

$$\tau_k = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{\kappa} \lambda_i} \tag{A.5}$$

is used. It gives the percentage of the covered information if a k–dimensional PCA–space is used.

At this step, an eigenspace model $\Omega = (\bar{\mathbf{x}}, \mathbf{T}, \Lambda, n)$ has been computed and an observation $\mathbf{y}$ can be transformed to the k–dimensional eigenspace using

$$\mathbf{g} = \mathbf{T}^\top (\mathbf{y} - \bar{\mathbf{x}}) \tag{A.6}$$

with

$$\mathbf{T} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k]\,. \tag{A.7}$$

A transformation of the data to a full–dimensional eigenspace does a decorrelation of the data with the covariance matrix of the decorrelated data being equal to the matrix of eigenvalues Λ.

The retransformation is defined up to a residue vector $\mathbf{r}$, with $\mathbf{r} = \mathbf{0}$ for $k = \kappa$:

$$\mathbf{y} = \mathbf{T}\mathbf{g} + \bar{\mathbf{x}} + \mathbf{r}. \tag{A.8}$$

The great advantages of PCA for computer vision are that it gives very compact representations of images or image features with good generalization capacities, although for a very accurate representation of the images a large number of dimensions might be needed. Since the transformation to the eigenspace is only a rotation, similar images will be transformed to similar points in the eigenspace. Therefore, the image dissimilarity is approximately

$$\|\mathbf{x}_1 - \mathbf{x}_2\|^2 \approx \|\mathbf{g}_1 - \mathbf{g}_2\|^2. \tag{A.9}$$

In case $k = \kappa$ the above equation holds strictly.

A.1.2. Incremental Principal Component Analysis

The batch method to compute the PCA described in the previous section has several drawbacks if the eigenspace has to be enlarged like in exploration tasks for robot localization or for learning new objects in a recognition task. Because of these drawbacks, algorithms for IPCA were proposed e.g. by Hall et al. (1998) and improved by Artac et al. (2002a). Like the batch method, IPCA is widely used in robot localization and navigation (Artac et al., 2002a; Freitas et al., 2003) or object recognition (Artac et al., 2002b). In the following paragraphs, the algorithm will be described.

The description is based on Artac et al. (2002a) and Freitas et al. (2003) and is restricted to updating an existing eigenspace Ω by adding an new observation $\mathbf{x}_{n+1}$. Therefore, the mean has to be updated:

$$\bar{\mathbf{x}}' = \frac{1}{n+1}\left(n\bar{\mathbf{x}} + \mathbf{x}_{n+1}\right). \tag{A.10}$$

Then, the observation is projected into the existing eigenspace

$$\mathbf{g}_{n+1} = \mathbf{T}^{\top}\left(\mathbf{x}_{n+1} - \bar{\mathbf{x}}\right). \tag{A.11}$$

The residual vector $\mathbf{r}$ caused by projecting $\mathbf{x}_{n+1}$ into the existing eigenspace and recovering, can be computed using equation (A.8)

$$\mathbf{r}_{n+1} = \left(\mathbf{T}\mathbf{g}_{n+1} + \bar{\mathbf{x}}\right) - \mathbf{x}_{n+1}. \tag{A.12}$$

As the residue vector is orthogonal to all the basis vectors used so far, its normalized equivalent

$$\hat{\mathbf{r}}_{n+1} = \begin{cases} \frac{\mathbf{r}_{n+1}}{\|\mathbf{r}_{n+1}\|} & \|\mathbf{r}_{n+1}\| \neq 0 \\ \mathbf{0} & \text{otherwise} \end{cases} \tag{A.13}$$

is used as new basis vector in order to enlarge the eigenspace. The new $(m \times k+1)$ sized basis $\mathbf{T}'$ is obtained by appending $\hat{\mathbf{r}}_{n+1}$ to the current basis $\mathbf{T}$ and applying a rotation $\mathbf{R}$:

$$\mathbf{T}' = \left[\mathbf{T}, \hat{\mathbf{r}}_{n+1}\right]\mathbf{R}, \tag{A.14}$$

where $\mathbf{R}$ is of the size $(k+1 \times k+1)$. To update the existing eigenimages $\mathbf{g}_i;\ i = 1...n$ it is necessary to reconstruct each image $\mathbf{g}_i$ using (A.8) and to transform it again to a new low dimensional representation

$$\mathbf{g}'_i = \left(\mathbf{T}'^{\top}\right)\left(\mathbf{x}_i - \bar{\mathbf{x}}'\right);\ i = 1...n+1. \tag{A.15}$$

After the updating step the old basis can be discarded. The $n+1$ images are now represented exactly in a $k+1$ dimensional eigenspace.

In equation (A.14) a rotation has been applied to compute the basis $\mathbf{T}'$. As the derivation of the rotation matrix $\mathbf{R}$ is not essential for the IPCA–based landmark–selection algorithm only a brief description will be given here. Readers interested in the complete derivation are referred to Hall et al. (1998).

The rotation matrix $\mathbf{R}$ is a solution for the eigenproblem

$$\mathbf{D}\mathbf{R} = \mathbf{D}\mathbf{\Lambda}' \tag{A.16}$$

where $\mathbf{\Lambda}'$ is a diagonal Matrix with the new eigenvalues $\lambda'_i;\ i = 1...k+1$ and $\mathbf{D}$ is a matrix consisting of known components of λ and $\mathbf{x}_{n+1}$

$$\mathbf{D} = \frac{n}{n+1}\begin{bmatrix}\mathbf{\Lambda} & \mathbf{0}\\ \mathbf{0} & 0\end{bmatrix} + \frac{n}{(n+1)^2}\begin{bmatrix}\mathbf{g}_{n+1}\mathbf{g}_{n+1}^\top & \gamma\mathbf{g}_{n+1}\\ \gamma\mathbf{g}_{n+1}^\top & \gamma^2\end{bmatrix} \tag{A.17}$$

with

$$\gamma = \hat{\mathbf{h}}_{n+1}^\top\left(\mathbf{x}_{n+1} - \bar{\mathbf{x}}\right). \tag{A.18}$$

Another important aspect not covered by this brief description is the dimensionality of the eigenspace model. As the dimensionality grows with every added observation all observations are represented exactly in the eigenspace. However, if an observation can be described very well in the existing eigenspace it might be desired to not enlarge the dimensionality in order to keep the feature description compact. For that case and for methods on how to decide whether to enlarge the number of dimensions or not, the reader is referred to the references cited above.

A.2. Independent Component Analysis

ICA was developed for applications in signal processing like BSS or the analysis of signals like MEG or EEG data. This approach to ICA is depicted in figure A.2: two signals (2) that are a linear mixture of two unknown independent source signals (1) are recorded. ICA can then be used to recover these source signals (3) from the observations. From a more abstract point of view ICA is a statistical method in which observed data is linearly transformed into a new vector so that the observations are now statistically independent. Independence is a stronger criterion than decorrelation which can be obtained by PCA. This approach to ICA is sketched in figure A.3.

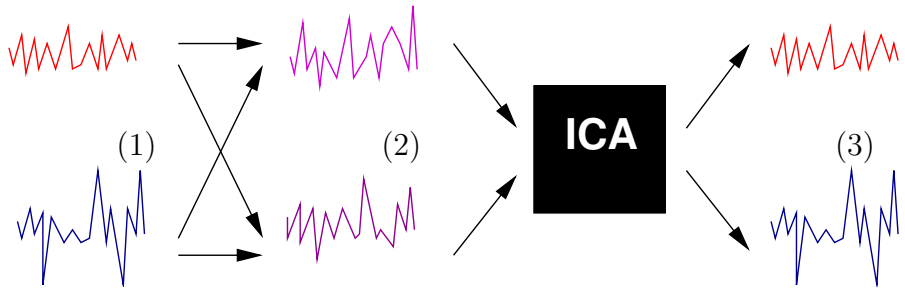


Figure A.2.: ICA for BSS in a nutshell. The signals (2) which are a linear mixture of the unknown source signals (1) are recorded. ICA can recover these source signals (3). After Stone (2004).

Here, a short introduction to ICA based on Hyvärinen (1999b) and Hyvärinen (2000) shall be given. For an in depth discussion the reader is referred to Hyvärinen et al. (2001b) or Stone (2004).

A.2.1. Definition of ICA

Let $\mathbf{x}$ be a random vector representing an observation whose elements $x_1, \ldots, x_n$ are linear mixtures of unknown or latent source signals $s_1, \ldots, s_n$. The linear mixing is represented by a $(n \times n)$–mixing matrix $\mathbf{A}$ with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$ which are also latent. Formally the ICA–problem can be written as

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{A.19}$$

$$= \sum_i \mathbf{a}_i s_i. \tag{A.20}$$

Usually this equation cannot be solved analytically, but neural networks or constrained optimization algorithms can be used to find a solution for the mixing matrix $\mathbf{A}$ from a set of observations $\mathbf{x}_i$. Further on, the model is only identifiable up to a nonzero and possibly negative scaling factor of the components and base vectors and to the permutation of the components. This is in contrast to PCA, where the order of the components plays an important role.

Important are also the socalled independent component filters which are the rows of the separating matrix $\mathbf{W}$ defined as

$$\mathbf{s} = \mathbf{W}\mathbf{x} \tag{A.21}$$

$$= \mathbf{A}^{-1}\mathbf{x} \tag{A.22}$$

A.2.2. ICA and Gaussianity

Another fundamental restriction of ICA is that it does not work with Gaussian distributions. For Gaussian distributions, decorrelation is equivalent to independence. Since the probability density function of a Gaussian with unit variance is completely symmetric, no direction information is available and therefore no direction of the base vectors can be estimated.

An important property of ICA is the statistical independence of the source signals. Statistical independence means that information on one variable $\mathbf{x}_1$ does not give information on any other variable $\mathbf{x}_2$. This means mathematically formulated that the two variables $\mathbf{x}_1$ and $\mathbf{x}_2$ are independent if and only if their joint pdf is factorizable to the product of the marginal distributions:

$$p(\mathbf{x}_1, \mathbf{x}_2) = p(\mathbf{x}_1)p(\mathbf{x}_2). \tag{A.23}$$

Closely related to Gaussianity of a random variable is the Central Limit Theorem. It states that the sum of two independent random variables is closer to a Gaussian than any of the two original variables. Additionally, the sum of the two variables is more dependent than any of the original variables. Therefore, ICA searches a new variable that is more independent, i.e. further away from a Gaussian, than the original variables. ICA algorithms often turn the estimation of the base vectors to an optimization problem

using a measure of non–Gaussianity as a criterion of statistical independence. In the following paragraphs a short overview over different optimization criterions shall be given:

Kurtosis
Kurtosis or the forth–order cumulant is a standard measure for non–Gaussianity. It is defined by

$$\operatorname{kurt}(\mathbf{X}) = E\left(\mathbf{X}^4\right) - 3\left(E\left(\mathbf{X}^2\right)\right)^2 \tag{A.24}$$

or assuming unit variance by

$$= E\left(\mathbf{X}^4\right) - 3. \tag{A.25}$$

For a Gaussian random variable the kurtosis is zero, for most other distributions it is nonzero. So Gaussianity can be measured by the absolute or the squared value of the kurtosis.

Negentropy
Entropy is a very fundamental concept of information theory and can be interpreted as the amount of information that an observation contains. The more random, i.e. unstructured and unpredictable, a variable is, the larger is its entropy. For a discrete random variable $\mathbf{X}$ with realizations $\mathbf{x}_i$ entropy is defined as

$$H(\mathbf{X}) = -\sum_i P(\mathbf{X} = \mathbf{x}_i) \log P(\mathbf{X} = \mathbf{x}_i). \tag{A.26}$$

An important theorem of information theory (Cover and Thomas, 1991) states that a gaussian random variable has the largest entropy among all possible distributions of the same variance. To use entropy as a measure of gaussianity, negentropy J is defined as

$$J(\mathbf{X}) = H(\mathbf{X}_{\text{gauss}}) - H(\mathbf{X}) \tag{A.27}$$

where $\mathbf{X}_{\text{gauss}}$ is a Gaussian random variable with the same variance than $\mathbf{X}$. Due to the mentioned properties J is always positive and zero only if $\mathbf{X}$ is Gaussian.

Mutual Information
Mutual information is an important measure of the dependence or redundancy of random variables. It is defined as the difference between the sum of entropies of a group of random variables and the entropy of their joint distribution

$$I_{\mathbf{X}} = \sum_i H(\mathbf{X}_i) - H(\mathbf{X}) \tag{A.28}$$

If entropy is understood as coding length, $H(\mathbf{X}_i)$ gives the lengths of codes for the $\mathbf{X}_i$ when these are coded separately and $H(\mathbf{X})$ gives the coding length if all the observations are coded together in one vector $\mathbf{X}$. Then, it gets clear that mutual information is only

zero if the observations are statistically independent, i.e. if there is no redundancy in the observations.

All these criteria and many others mentioned in Hyvärinen (2000) for non–Gaussianity can be used to formulate objective functions to maximize the independence of the source signals and to estimate the mixing matrix.

As mentioned in section 1.2.2.5, neural network approaches to sparse–dispersed coding are equivalent to ICA. For the training of these networks, constraints are put on the resulting output code assuring sparseness. As objective functions the above measures of non–Gaussianity can be used.

A.2.3. The FastICA Algorithm

In this work the ICA estimation was done using the FastICA algorithm, which was presented in Hyvärinen (1999a) and has established as standard algorithm for computing the ICA.[1]

The proposed algorithm is – like most neural algorithms – parallel and distributed. The optimization is turned into a fixpoint problem, so a standard Newton–Iteration can be used for the optimization. Therefore, the convergence of the algorithm is cubic for the most cases, and squared for some rate cases. The algorithm maximizes the Mutual Information of the sources, therefore reducing also dependencies higher than of forth order. The other main advantages are that there are no step size parameters to tune and that the components are estimated one by one.

Preprocessing

The first preprocessing step is to transform the observations $\mathbf{x_i}$ to a zero mean observation $\tilde{\mathbf{x}}_\mathbf{i}$ by subtracting the mean observation $\bar{\mathbf{x}}$. The next step is the whitening of the data. This is a linear transformation of the observation so that the observed components have unit variance, i.e. the covariance matrix is the unit matrix:

$$\operatorname{cov}(\mathbf{X}) = \mathbf{1}. \tag{A.29}$$

A method frequently used for whitening is the EVD of the covariance matrix

$$\operatorname{cov}(\mathbf{X}) = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^\top. \tag{A.30}$$

The whitening step then contains

$$\hat{\mathbf{X}} = \boldsymbol{\Lambda}^{-\frac{\mathbf{1}}{\mathbf{2}}}\mathbf{E}^\top\mathbf{X} \tag{A.31}$$

So the whitening is nothing else than a linear transformation to the eigenspace and a multiplication with

$$\boldsymbol{\Lambda}^{-\frac{1}{2}} = \operatorname{diag}\left(\lambda_1^{-\frac{1}{2}}, \lambda_1^{-\frac{1}{2}}, \ldots, \lambda_n^{-\frac{1}{2}}\right). \tag{A.32}$$

[1] A MATLAB package is freely available at `http://www.cis.hut.fi/projects/ica/fastica`.

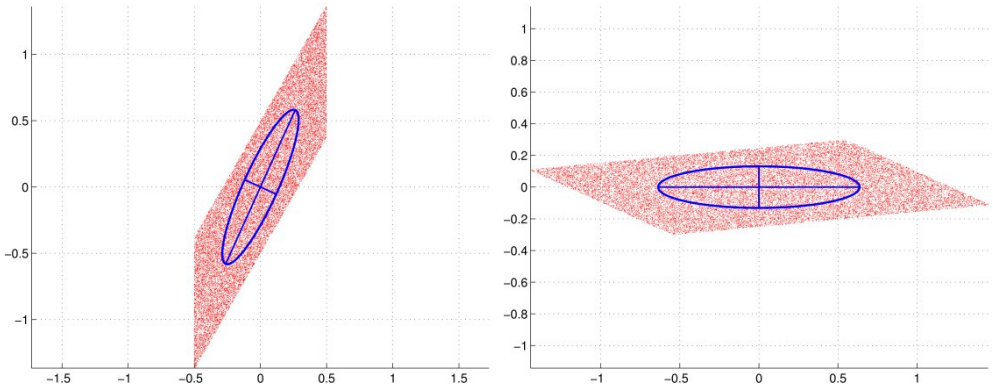
Since the whitening is done using PCA the dimensionality can be effectively reduced as described in A.1. Using (A.20) equation (A.31) can be rewritten

$$\hat{\mathbf{X}} = \mathbf{E}\boldsymbol{\Lambda}^{-\frac{\mathbf{1}}{\mathbf{2}}}\mathbf{A}\mathbf{s} \tag{A.33}$$

$$= \hat{\mathbf{A}}\mathbf{s}, \tag{A.34}$$

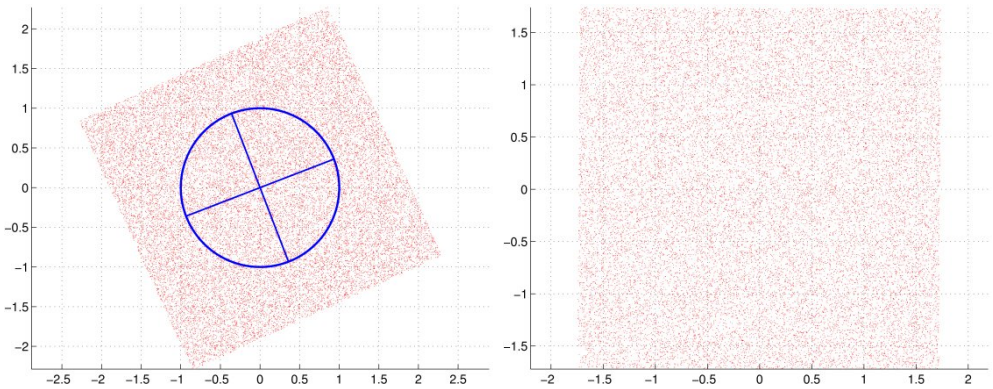
The new mixing matrix $\hat{\mathbf{A}}$ is now orthogonal and further reduces the number of parameters to estimate, because $\hat{\mathbf{A}}$ contains less degrees of freedom.

Additionally, if whitening is done by computing the EVD, an easy and effective reduction of dimensionality can be done by reducing the dimensionality of the eigenspace as described in section A.1.1.



(a) A uniformly distributed data cloud after shearing.

(b) The same data cloud after decorrelation, i.e. after computing PCA. The data is not yet independent.



(c) The whitened data cloud. The error ellipse is now a circle.

(d) After computing ICA. None of the coordinate directions gives information about the other, the distribution is independent.

Figure A.3.: ICA for data mining in a nutshell

A.3. Fourier Transformation

Fourier Analysis is one of the most frequently used mathematical methods and the numerical algorithm to compute the Fourier transformed of a function has become the most frequently used numerical algorithm. By the means of Fourier transformation the data is transformed to another representation, in this case the frequency domain, whose basis consists of orthogonal wave patterns of different frequency. The introduction given here is based on Jähne (2002) and Press et al. (2003), for an in depth introduction the reader is referred to Kammler (2000) or particularly for computer vision and Fourier methods to Bracewell (2003).

A.3.1. Introduction to Discrete Fourier Transformation

For a continuous function $h(x)$ its Fourier transformed $H(f)$, which is no longer a function of space or time but of frequency, is defined as

$$H(f) = \int_{-\infty}^{\infty} h(t) \exp\left(2\pi i f t\right) dt. \tag{A.35}$$

Since for this work only discrete sampled data is used, the focus will be on DFT. For the discrete case a continuous function is sampled or recorded at N evenly spaced intervals Δ in space, resulting in a sequence of sampled values

$$h_n = h(n\Delta) \text{ with } n = \ldots, -2, -1, 0, 1, 2, \ldots \tag{A.36}$$

Instead of estimating the Fourier transform of h_n, only estimates at discrete frequencies

$$f_n = \frac{n}{N\Delta} \text{ with } n = -\frac{N}{2}, \ldots, \frac{N}{2} \tag{A.37}$$

are seeked. The lower and upper frequency limits are exactly the Nyquist critical frequencies that will be introduced in section A.3.2. Then, equation (A.35) can be written as follows:

$$H_n = \sum_{k=0}^{N-1} h_k \exp\left(\frac{2\pi i k n}{N}\right). \tag{A.38}$$

The equation shows that DFT maps the N complex coefficients $h_k; k = 1, \ldots, N-1$ to the coefficients H_n, which are independent from the interval size Δ.

The DFT can be understood as a scalar product of the vector $\mathbf{h} = [h_0, h_1, \ldots, h_{N_1}]^\top$ with a set of N orthonormal basis vectors

$$\mathbf{b}_n = \left[w^0, w^n, w^{2n}, \ldots, w^{(N-1)n}\right]^\top \tag{A.39}$$

with

$$w = \exp\left(\frac{2\pi i}{N}\right). \tag{A.40}$$

because equation (A.38) can be rewritten as

$$H_n = \sum_{k=0}^{N-1} w^{nk} h_k \tag{A.41}$$

$$= \mathbf{b}_n^\top \mathbf{h}. \tag{A.42}$$

From this notation it gets clear that the DFT is a linear transformation method projecting the data to a basis of complex exponential basis functions.

To apply the DFT on images it is necessary to define a 2D–Transformation. The complex function is now sampled at $N_1 N_2$ locations resulting in the function values $h(n_1, n_2)$ which are arranged in the matrix $\mathbf{H}$. The 2D–DFT is then defined as a complex function

$$H(n_1, n_2) = \sum_{k_2=0}^{N_2-1} \sum_{k_1=0}^{N_1-1} \exp\left(\frac{2\pi i k_2 n_2}{N_2}\right) \exp\left(\frac{2\pi i k_1 n_1}{N_1}\right) h(k_1, k_2). \tag{A.43}$$

In the 2D–case one can express the above equation using a set of complex basis matrices $\mathbf{B}_{n_1,n_2}$ defined as an outer product of two basis vectors $\mathbf{b}_{n_1}$ and $\mathbf{b}_{n_2}$:

$$\mathbf{B}_{n_1,n_2} = \mathbf{b}_{n_1} \mathbf{b}_{n_2}^\top \tag{A.44}$$

$$= \begin{bmatrix} w^0 \\ w^{n_1} \\ w^{2n_1} \\ \vdots \\ w^{(N_1-1)n_1} \end{bmatrix} \left[w^0, w^{n_2}, w^{2n_2}, \ldots, w^{(N_2-1)n_2}\right]. \tag{A.45}$$

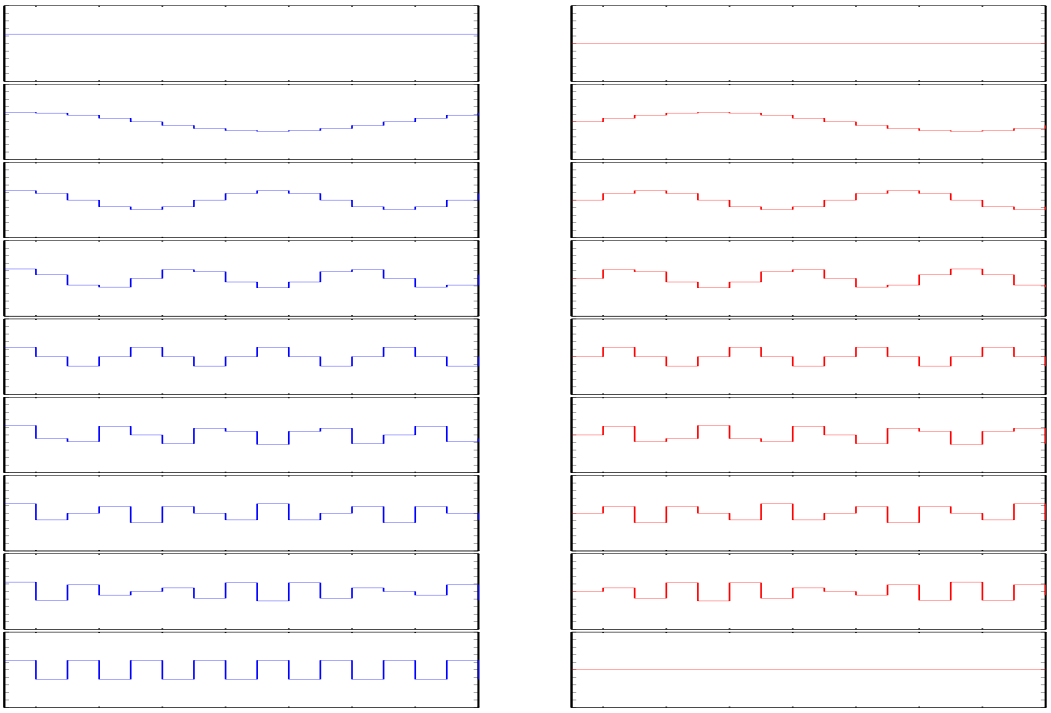
Thus, (A.43) reads

$$H(n_1, n_2) = \sum_{k_2=0}^{N_2-1} \left(\sum_{k_1=0}^{N_1-1} h(k_1, k_2) w^{k_1 n_1} \right) w^{k_2 n_2} \tag{A.46}$$

$$= \langle \mathbf{B}_{n_1,n_2}, \mathbf{H} \rangle, \tag{A.47}$$

with the scalar product between to complex matrices being defined as

$$\langle G, H \rangle = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \bar{g}_{m,n} h_{m,n}. \tag{A.48}$$



(a) Real part (cosine functions) (b) Imaginary part (sine functions)

Figure A.4.: The first nine basis functions for 1D DFT. After Jähne (2002)

Equation (A.46) reveals an important property of DFT, its separability. Thus, the 2D–DFT can be obtained from the one–dimensional case by taking 1D–DFTs sequentially on each index of the original function.

For the experiments described in section 2.1 the power spectrum of the Fourier transformed is important. It gives the energy that is present in a certain frequency and discards amplitude and phase information. The power spectrum of a Fourier transformed $H(f)$ is defined by the squared modulus of the k–th Fourier coefficient

$$S(f_k) = |H(f_k)|^2 \tag{A.49}$$

or for the 2D–case

$$S(f_{x,k}, f_{y,k}) = |H(f_{x,k}, f_{y,k})|^2. \tag{A.50}$$

A.3.2. Discrete Sampling

Although equation (A.37) does not depend on the interval width Δ, the interval width plays an important role for computing the DFT because it influences the accuracy or the resolution of the discretization. The accuracy is measured by the sampling rate and is defined as the reciprocal of the interval size Δ. Directly linked to the sampling rate

is the Nyquist critical frequency defined by

$$f_c = \frac{1}{2\Delta}. \tag{A.51}$$

The Nyquist critical frequency is that frequency for which an exact sampling is still possible. In this case the underlying waves consist of two sampling points per cycle. For computer vision applications each sampling point corresponds to one pixel. Since space is often measured in units of the sampling interval Δ, the critical frequency f_c is often denoted by $\frac{1}{2}$ cycles per pixel.

From Shannon's sampling theorem further follows that a continuous function $h(x)$ is fully determined by its samples h_n if it is *bandwidth limited*. That means that all frequencies are smaller in magnitude than f_c, i.e. $H(f) = 0$ for $\mathrm{abs}\,(f) > f_c$. If a function is not bandwidth limited to a frequency range less than the critical frequency, it is by sampling and computing the power spectrum falsely forced into the frequency range $-f_c < f < f_c$. This phenomenon is called *aliasing*.

To reduce aliasing effects for data where the the sampling rate cannot be influenced, windowing of the data is used. A windowing function is a function raising smoothly from zero to unity and falling back to zero again at the boarder of the range. To reduce the aliasing effects the data is bin by bin multiplied with the window function before the FFT is computed. Since for FFT the input data is assumed to be periodic, there is a large proportion of high frequencies at the boarders of the data, leading to aliasing effects. This proportion of high frequencies is reduced by windowing because the function values at the boarders are than close to zero.

There are many different window functions with subtle differences described in the literature. For the work described here, the choice of the windowing function is not essential. So, the Blackman–Harris window was chosen that is defined by

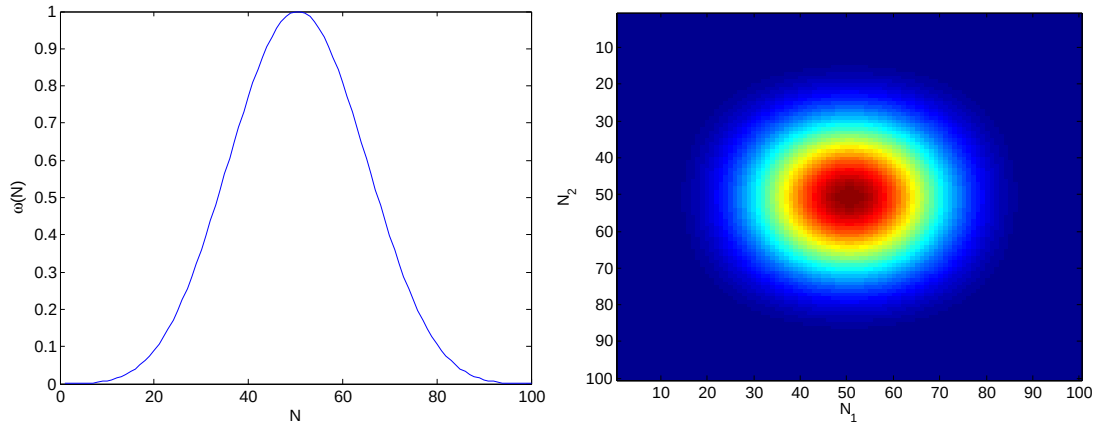
$$\omega_j = 0.3586 - 0.4883 \cos\left(\frac{2\pi j}{N}\right) + 0.1413 \cos\left(\frac{4\pi j}{N}\right) - 0.0117 \cos\left(\frac{6\pi j}{N}\right) \tag{A.52}$$

with $j = 0, \ldots, N-1$. From the resulting vector ω of windowing weights the weighting matrix $\mathbf{W}$ for the 2D case is given by the outer product $\mathbf{W} = \omega\omega^\top$. The Blackman–Harris window function is plotted in figure A.5.

A.4. Kolmogorov–Smirnov–Test

Statistical tests are among the most frequently used statistical techniques. For this work the Kolmogorov–Smirnov–Test (KST) is relevant. It is used to test if two sets of unbinned samples, which are functions of a single independent variable, are likely to be drawn from the same or from different continuous distribution functions. The short introduction given here is based on Press et al. (2003).

Formally spoken it is tested whether the null Hypothesis H_0 that the two sets are equally distributed can be accepted or not. If H_0 is rejected, it is proven that the sets are from different distributions. If H_0 is not rejected, it is likely that the two sets share



(a) 1D Blackman–Harris window (b) 2D Blackman–Harris window

Figure A.5.: Blackman–Harris window

a common distribution, but it cannot be proven that the sets are drawn from identical distributions.

For both sets $\mathcal{S}_1$ and $\mathcal{S}_2$ the cdfs S_{N_1} and S_{N_2} are computed. They are stepwise constant functions jumping at the locations of the samples of each set. Different distribution functions will lead to different estimates of the cdf. To measure the difference between the distributions the KST computes the maximum value of the absolute distance of the two cdfs:

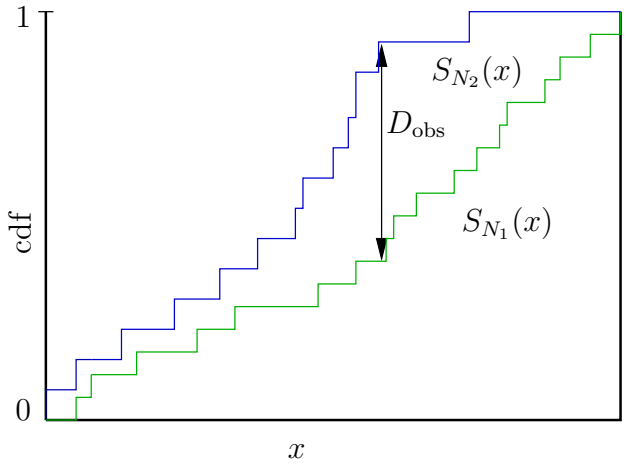


Figure A.6.: Kolmogorov–Smirnov–Test (KST) in a nutshell. It computes the maximum absolute difference D_{obst} between the estimates S_{N_1} and S_{N_2} of the cdf for the two sets of samples. Using this measure the significance value is computed. After Press et al. (2003).

$$D_{\text{obs}} = \max_{-\infty<x<\infty} \text{abs}\left(S_{N_1}(x) - S_{N_2}(x)\right). \tag{A.53}$$

The probability that the maximum of the absolute difference D of the real cdfs is greater than D_{obs} is given by

$$P\left(D > D_{\text{obst}}\right) = Q_{\text{KS}}\left(\frac{\sqrt{N_e} + 0.12 + 0.11}{\sqrt{N_e}} D_{\text{obs}}\right) \tag{A.54}$$

with

$$N_e = \frac{N_1 N_2}{N_1 + N_2} \tag{A.55}$$

and $Q_{KS}(\rho)$ being defined as

$$Q_{\text{KS}}(\rho) = 2\sum_{j=1}^{\infty}(-1)^{j-1} \exp -2j^2\rho^2 \tag{A.56}$$

with limiting values $Q_{\text{KS}}(0) = 1$ and $Q_{\text{KS}}(\infty) = 0$. In case the KST is used to test if an observation is distributed according to a given distribution, S_{N_2} is replaced by the distribution's cdf and $N_e = N_1$.

The hypothesis H_0 is rejected for the chosen significance level α if

$$P\left(D > D_{\text{obst}}\right) > \alpha \tag{A.57}$$

Advantages of the KST are, that it is already reasonable accurate for $N_e > 4$ and that it is invariant under reparametrization, i.e. using a logarithmic scale for the samples will result in the same significance values as using a standard scale.

A.5. Error Bounds for Classification

In classification tasks it is essential to compute the classification error. A classification error occurs if an element x is classified as belonging to class $\mathcal{R}_1$ instead of class $\mathcal{R}_2$ and vice versa. This idea is sketched in figure A.7. Using Bayesian decision theory (Duda et al., 2001) one would decide for x belonging to $\mathcal{R}_1$ if $P\left(\mathcal{R}_1|x\right) > P\left(\mathcal{R}_2|x\right)$ and vice versa for $P\left(\mathcal{R}_1|x\right) < P\left(\mathcal{R}_2|x\right)$. The decision error is then given by

$$P\left(\text{error}|x\right) = \min\left(P\left(\mathcal{R}_1|x\right), P\left(\mathcal{R}_2|x\right)\right). \tag{A.58}$$

As the two sources of error are mutually exclusive the overall probability of error is given by

$$P\left(\text{error}\right) = P\left(x \in \mathcal{R}_1|\mathcal{R}_2\right) + P\left(x \in \mathcal{R}_2|\mathcal{R}_1\right) \tag{A.59}$$

$$= \int_{\mathcal{R}_1} P\left(x|\mathcal{R}_2\right) P\left(R_2\right) dx + \int_{\mathcal{R}_2} P\left(x|\mathcal{R}_1\right) P\left(R_1\right) dx. \tag{A.60}$$

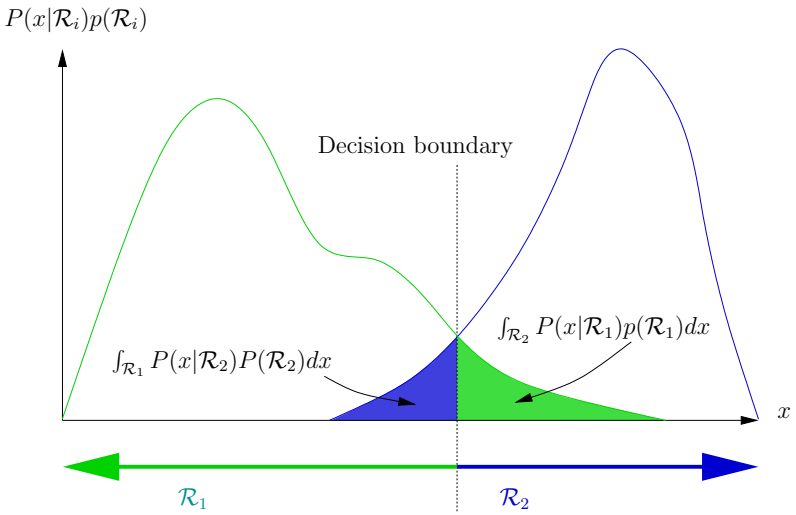


Figure A.7.: Decision error in a nutshell. After Duda et al. (2001)

The classification task is usually not considered for one sample x, but for a large number of samples. The introduction given here will therefore focus on the Log–Likelihood Test and two bounds for its classification error, the Chernoff bound and the Bhattacharyya bound. The introduction is based on Konishi et al. (1999, 2003).

A.5.1. Approach via Log–Likelihood Test

Let $\mathcal{R}_1$ and $\mathcal{R}_2$ be two distinctive classes and $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ a set of N observations randomly drawn from one of these classes. According to the Neyman–Pearson lemma (Cover and Thomas, 1991; Duda et al., 2001) the optimal test for determing whether the samples were drawn from $\mathcal{R}_1$ or $\mathcal{R}_2$ is the log–likelihood ratio

$$r = \log \frac{P\left(\mathcal{R}_1|\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\right)}{P\left(\mathcal{R}_2|\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\right)}. \tag{A.61}$$

Assuming independence gives

$$= \log \frac{\prod_{i=1}^{n} P\left(\mathcal{R}_1|\mathbf{x}_i\right)}{\prod_{i=1}^{n} P\left(\mathcal{R}_2|\mathbf{x}_i\right)} \tag{A.62}$$

$$= \sum_{i=1}^{n} \log \frac{P\left(\mathcal{R}_1|\mathbf{x}_i\right)}{P\left(\mathcal{R}_2|\mathbf{x}_i\right)}. \tag{A.63}$$

In order to estimate the the pdfs for $\mathcal{R}_1$ and $\mathcal{R}_2$ the observations are usually divided into bins $y_j; j = 1 \ldots m$ with $y(x_i)$ indicating the bin of x_i. So the log likelihood can be rewritten:

$$= \sum_{i=1}^{n} \log \frac{P\left(\mathcal{R}_1|y(\mathbf{x}_i)\right)}{P\left(\mathcal{R}_2|y(\mathbf{x}_i)\right)} \tag{A.64}$$

$$= \sum_{j=1}^{m} \log \frac{P\left(\mathcal{R}_1|y_j\right)}{P\left(\mathcal{R}_2|y_j\right)}. \tag{A.65}$$

The larger the log–likelihood ratio r the more likely the sequence of observations was drawn from $\mathcal{R}_1$, for $r = 0$ $\mathcal{R}_1$ and $\mathcal{R}_1$ are equally likely, for $r < 0$ it is more likely that the sample was drawn from $\mathcal{R}_2$. Using the theory of types (Cover and Thomas, 1991; Yuille and Coughlan, 2000) providing a statistical framework for sequences of observations one can show that the expected error rate ϵ for the lok–likelihood test decreases exponentially by

$$\epsilon = \exp\left(-NC\left(P\left(.|\mathcal{R}_1\right), P\left(.|\mathcal{R}_1\right)\right)\right) \tag{A.66}$$

where $C\left(p, q\right)$ is the Chernoff information between $p_{\mathcal{R}_1}$ and $p_{\mathcal{R}_2}$ defined as

$$C\left(p, q\right) = -\min_{0 \leq \lambda \leq 1} \log \sum_{j=1}^{m} p^{\lambda}(y_j) q^{1-\lambda}(y_j). \tag{A.67}$$

The Bhattacharyya bound is given for $\lambda = 0.5$. In general

$$C\left(p, q\right) \geq B\left(p, q\right) \tag{A.68}$$

holds since for the Chernoff bound λ is chosen to minimize $\log \sum_{j=1}^{m} p^{\lambda}(y_j) q^{1-\lambda}(y_j)$, while for the Bhattacharyya coefficient it is just set to $\lambda = 0.5$.

Because of equation (A.66) the classification error gets smaller the larger the Chernoff information or the Bhattacharyya coefficient get. A huge advantage of the Chernoff bound is that the minimization of λ is always done in an one–dimensional space.

A.6. Evolution Strategies

Algorithms from the field of Evolutionary Computation are inspired by the biological evolution and model the evolutionary processes to solve technical optimization problems. By the time, three subfields of evolutionary computation can be distinguished: Evolutionary Algorithms (EA), Genetic Programming (GP), and Evolution Strategies (ES). While the first two are mainly used for integer– or combinatorial optimizations and optimizations of whole program fragments (e.g. for symbolic regression), the latter one was especially designed for real–valued optimization problems. Other advantages of ES include that they are robust, flexible, have little parameters to tune, and that only little function evaluations of the objective function and no computations of derivatives are needed.

A.6.1. Key Aspects of ES

The short overview over ES given here is based on Jacob (2003). ES simulate a collective learning or search process within a population of individuals in the search space. Therefore, ES are often counted to the Stochastic Search Methods. The key aspects of ES include:

Populations and Individuals

A population $\mathcal{G}$ at time t is described by a multi–set of μ individuals:

$$\mathcal{G}(t) = \{\mathcal{I}_1(t), \mathcal{I}_2(t), \ldots, \mathcal{I}_\mu(t)\}. \tag{A.69}$$

Each individual is represented by a tuple of n–dimensional real vectors

$$\mathcal{I}_i(t) = (\mathbf{p}, \mathbf{s}) \tag{A.70}$$

$$= ((p_1, p_2, \ldots, p_n), (s_1, s_2, \ldots, s_n)) \tag{A.71}$$

with $p_i, s_i \in \mathbb{R}$. The vector $\mathbf{p}$ is also called the chromosome of the individual, the p_i are referred to as object parameters. The chromosome specifies at which point of the search space the corresponding individual is located. The s_i are called the strategy parameters and control the range of mutation on the object parameters.

Mutation

Mutation is considered the driving force introducing variability in the chromosomes. Mutation on an individual can be described as follows:

$$\mathcal{I}_{\text{mut}} = (\mathbf{p}_{\text{mut}}, \mathbf{s}_{\text{mut}}) \tag{A.72}$$

$$= (\mathbf{p} + \mathcal{N}_0(\mathbf{s}), \omega_{\text{mut}}(\mathbf{s})). \tag{A.73}$$

This equation also shows the meaning of the strategy parameters: they are the standard deviation of a vector with independent Gaussian random values with zero mean. So the mutation adds Gaussian noise to the individuals, changing slightly the location of the individual in the search space.

The mutation operator ω_{mut} mutates the strategy parameters. Two common mutation operators are the MSA– and the HMB–operator. The first one is defined as

$$\omega_{\text{mut, MSA}} = (s_1\xi_1, s_2\xi_2, \ldots, s_n\xi_n) \tag{A.74}$$

with

$$\xi_i = \begin{cases} \beta & \mathcal{X} \leq 0.5 \\ \beta^{-1} & \text{otherwise} \end{cases}, \tag{A.75}$$

where β is the mutation parameter usually chosen to be approximately 1.5. With $\mathcal{X}$ beeing a uniform distributed random number, the standard deviation of one half of the strategy parameters is enlarged, the other half gets smaller.

The HMB–Operator is defined as

$$\omega_{\text{mut, HMB}} = (s_1\xi_1, s_2\xi_2, \ldots, s_n\xi_n) \tag{A.76}$$

with

$$\xi_i = \exp \mathcal{N}_0(\beta). \tag{A.77}$$

Again, β controls the range of the stepsize adjustment and is commonly chosen to be approximately 2. At the beginning of the optimization, individuals with larger stepsize adjustments are more likely to be selected. At the end of the optimization, when the individuals are close to the maxima, small stepsize adjustments are preferred.

Recombination

The recombination allows the mixing of genetic information between the parents and their offspring or between different subpopulations. Two commonly used recombination operators are the discrete and the intermediate recombination. With $x_{a,i} = (p_{a,i}, s_{a,i})$ being the i–th object and strategy parameter of the individual a, the discrete recombination operator can be written as follows:

$$\omega_{\text{rec,dis}}(x_{a,i}, x_{b,i}) = \begin{cases} x_{a,i} & \mathcal{X} \leq 0.5 \\ x_{b,i} & \text{otherwise} \end{cases}. \tag{A.78}$$

Therefore, the discrete recombination function assigns the recombined individual with 50% probability either the parameters of individual a or those of individual b.

The intermediate recombination operator works on at least two individuals and assigns the recombined individual the mean of the object and strategy parameters:

$$\omega_{\text{rec,int}}(x_{1,i}, x_{2,i}, \ldots, x_{p,1}) = \frac{1}{p} \sum_{j=1}^{p} x_{j,i}. \tag{A.79}$$

Since ES were developed for real–valued optimization problems, the intermediate recombination is often the recombination operator of choice.

The Fitness Function

In terms of ES the objective function is called fitness function $\eta : \mathbb{R}^n \to \mathbb{R}$ and assigns each individual $\mathcal{I}$ a real–valued fitness according to its chromosome $\mathbf{p}$. Like for any other optimization method, the choice of the objective function is crucial to the results of the optimization.

Selection

The selection operator ω_{sel} selects those individuals that will form the next generation by comparing their fitness. Depending on the used strategy the parent individuals are included (plus–strategy) or excluded (comma–strategy).[2] Additional requirements can for example be that from each parent's offspring exactly one offspring makes it to the next generation. The choice of the selection strategy has a large influence on the performance of the optimization: comma–strategies are more tolerant to local maxima while plus–strategies get trapped in local maxima more easily. On the other hand plus–strategies tend to converge faster than comma–strategies.

[2]For sake of simplicity a simpler notation than the standard notation of ES literature is used here.

A.6.2. The ES–Algorithm

After introducing all these key elements of ES the optimization algorithm can be formulated. A pseudo–code notation of the ES–algorithm is depicted in figure A.8. The algorithm starts with the initialization of the initial population $\mathcal{G}(0)$. Then, the fitness of the initial population is evaluated using the fitness function. If the termination criterion is not true, the algorithm starts iterating. Such termination criteria can be the fitness function or the fitness change between two generations falling below a threshold. The next step in the loop is to λ–times identically replicate the parent's genome. Afterwards, the offspring generation is mutated and replication takes place. Then, the fitness of the offspring generation is evaluated and the individuals that form the new generation are selected. If the iteration ends, the last generation of individuals and their fitness are returned.

Figure A.8.: The ES–Algorithm

begin
  $t := 0$;
  Initialization: $\mathcal{G}(0) := \{\mathcal{I}_1(t)\mathcal{I}_2(t), \ldots, \mathcal{I}_\mu(t)\}$;
  Initial fitness evaluation: $\eta\left(\mathcal{G}(0)\right) := \{\eta\left(\mathcal{I}_1(t)\right), \eta\left(\mathcal{I}_2(t)\right), \ldots, \eta\left(\mathcal{I}_\mu(t)\right)\}$;
  while *Termination criterion not true* **do**
    $t := t + 1$;
    Create offspring generation $\mathcal{G}'(t)$: For each individual: $\mathcal{I}_i \rightarrow \{\mathcal{I}_i^1, \mathcal{I}_i^2, \ldots, \mathcal{I}_i^\lambda\}$;
    Mutation: $\mathcal{G}''(t) := \omega_{\text{mut}}\left(\mathcal{G}'(t)\right)$;
    Recombination: $\mathcal{G}'''(t) := \omega_{\text{rec}}\left(\mathcal{G}'(t)\right)$;
    Fitness Evaluation: $\eta\left(\mathcal{G}''(t) \cup \mathcal{G}'''(t)\right)$;
    Selection: $\mathcal{G}(t) := \omega_{\text{sel}}\left(\mathcal{G}(t-1) \cup \mathcal{G}''(t) \cup \mathcal{G}'''(t)\right)$;
  end
  return $\mathcal{G}(t), \eta\left(\mathcal{G}(t)\right)$
end

B. Alternative Simple Cell Models

As mentioned in section 3.1.2 several simple cell models were proposed in Hansen (2002) and Hansen and Neumann (2004b). From these models only the one leading to the most promising results, namely the nonlinear model with DOI, was used and analyzed in detail in section 3.1.2. However, the others have also been implemented. One alternative to the described model does not use the DOI–scheme, i.e. inhibitory and excitatory contributions to the equations (3.11) and (3.12) are weighted equally with $\xi = 1$. Another alternative results from using a linear simple cell circuit as depicted in figure B.1. The simple cell response $\tilde{\mathbf{S}}$ is computed by directly pooling the input activations $\mathbf{R}_{\text{on}}$ and $\mathbf{R}_{\text{off}}$:

$$\tilde{\mathbf{S}} = \mathbf{R}_{\text{on}} + \mathbf{R}_{\text{off}}. \tag{B.1}$$

Due to the different combinations of alternatives four different models were implemented: quasi–linear without DOI, quasi–linear with DOI, nonlinear without DOI, and nonlinear with DOI. The term quasi–linear refers to the linear simple cell circuit which builds on a nonlinear center surround circuit.

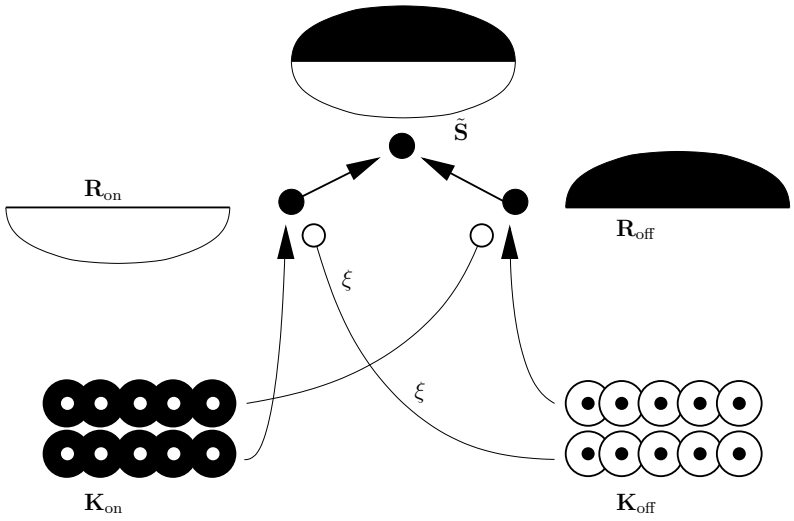


Figure B.1.: Sketch of the linear simple cell model.

For all models the model parameters have been determined as described in section 3.1.2. For the models without DOI the test stimulus used in the optimization step had a contrast of 0.1, all other parameters remained unchanged. The results of the models not described in section 3.1.2 are summarized in the following tables.

Table B.1.: Model parameters for the quasi–linear model without DOI

Parameter	Natural filters	Manmade filters	Aerial filters
σ_c	1.000	1.000	1.000
σ_s	3.000	3.000	3.000
α_{LGN}	14.105	12.092	14.218
β_{LGN}	3.108	0.628	0.800
γ_{LGN}	13.663	0.176	0.806
ξ	1.000	1.000	1.000

Table B.2.: Model parameters for the quasi–linear model with DOI

Parameter	Natural filters	Manmade filters	Aerial filters
σ_c	1.000	1.000	1.000
σ_s	3.000	3.000	3.000
α_{LGN}	4.285	13.810	14.934
β_{LGN}	0.511	0.497	0.434
γ_{LGN}	0.189	0.087	0.370
ξ	3.000	3.000	3.000

Table B.3.: Model parameters for the nonlinear model without DOI

Parameter	Natural filters	Manmade filters	Aerial filters
σ_c	1.000	1.000	1.000
σ_s	3.000	3.000	3.000
α_{LGN}	12.479	13.700	14.941
β_{LGN}	0.013	0.016	0.023
γ_{LGN}	0.053	0.010	0.015
ξ	1.000	1.000	1.000
α_{S}	13.361	14.840	13.875
β_{S}	1079.100	1115.300	1024.500
γ_{S}	1.732	6.822	0.783

C. MATLAB Package Overview

All the necessary computations were done using MATLAB Release 14. The functions related with this work were bundled in different packages. To manage these packages and to make the necessary functions available in the MATLAB search path, a simple package management system was written, which can also resolve dependencies by recursively loading other packages.

The packages which are relevant for the whole department of Cognitive Neuroscience are named with the prefix `cog_`, those that are only relevant for this work have the prefix `lg_`. Packages provided by other people are labeled with the prefix `lib_`. A list of packages is given below and is supposed to be a starting point for people that are going to build on this work. For each package a `Contents.m` file will be made available so that information about the functions contained in the package can be accessed by typing `help <package_name>` on the MATLAB shell or by using the browsable documentation created by m2html.

Table C.1.: List of relevant MATLAB packages

`cog_base`	Package management system
`cog_plot`	Department–specific plotting utilities
`cog_imshow`	Department–specific functions to visualize images
`cog_aux`	Other department–specific auxiliary functions
`cog_es`	A comprehensive library for Evolution Strategies
`cog_demo`	A couple of demos, mainly visualizations for this thesis
`cog_compvis`	Early vision models (chapter 3)
`lg_compvis`	Specific additions for early vision models (chapter 3)
`lg_image_spectrum`	Analysis of Power Spectra (section 2.1)
`lg_imageica`	Estimation of receptive fields (section 2.2)
`lg_contour`	Analysis and optimization of edge detection model (section 3.1)
`lg_lms`	Landmark selection (chapter 4)
`lib_fastica`	The FastICA–algorithm (`http://www.cis.hut.fi/projects/ica/fastica/code/dlcode.shtml`)
`lib_contournet`	Non–negative ICA for contour analysis (`http://www.cs.helsinki.fi/u/phoyer/`)
`lib_nnica`	Non–negative ICA (`http://www.cs.helsinki.fi/u/phoyer/`)
`lib_m2html`	Creates HTML–documentation (`http://www.artefact.tk/software/matlab/m2html/`)

Bibliography

M. Artac, M. Jogan, and A. Leonardis. Mobile robot localization using an incremental eigenspace model. In *Proceedings of the ICRA 2002*, pages 1025–1030, 2002a.

M. Artac, M. Jogan, and A. Leonardis. Incremental PCA for on–line visual learning and recognition. In *Proceedings of the ICPR 2002*, pages 781–784, 2002b.

F. Attneave. Some informational aspects of visual perception. *Psychological Reviews*, 61:183–193, 1954.

T. Bailey and E. Nebot. Localization in large–scale environments. *Robotics and Autonomous Systems*, 37:261–281, 2001.

R. Balboa and N. Gryzwacz. Power spectra and distribution of contrasts of natural images from different habitats. *Vision Research*, 43:2527–2537, 2003.

H. Barlow. *Sensory Communications*, chapter Possible Principles underlying the Transformations of Sensory Messages, pages 217–234. MIT Press, 1961.

H. Barlow. Redundancy reduction revisited. *Network: Computation in Neural Systems*, 12:241–253, 2001.

A. Bell and T. Sejnowski. The independent components of natural images are edge filters. *Vision Research*, 37:3327–3338, 1997.

A. Bernardino, L. Custódio, J. Frazão, T. Krause, P. Lima, M. Ribeiro, and A. V. J. Santos-Victor and. RESCUE – 2nd year technical report. Technical report, Instituto de Sistemas e Robótica, Instituto Superior Técnico, Lisbon, 2003a.

A. Bernardino, L. Custódio, J. Frazão, P. Lima, M. Ribeiro, and A. V. J. Santos-Victor and. RESCUE – 3rd year technical report. Technical report, Instituto de Sistemas e Robótica, Instituto Superior Técnico, Lisbon, 2003b.

A. Bernardino, L. Custódio, J. Frazão, P. Lima, F. Melo, M. Ribeiro, and A. V. J. Santos-Victor and. RESCUE – final technical report. Technical report, Instituto de Sistemas e Robótica, Instituto Superior Técnico, Lisbon, 2004.

C. Blakemore and G. Cooper. Development of the brain depends on the visual environment. *Nature*, 228:477–478, 1970.

E. Bourque, G. Doudek, and P. Ciaravola. Robotic sightseeing – a method for automatically creating virtual environments. In *Proceedings of the ICRA 1998*, pages 3186–3191, 1998.

R. Bracewell. *Fourier Analysis for Imaging.* Prentice Hall, 1st edition, 2003.

G. Buchsbaum and A. Gottschalk. Trichromacy, opponent colours coding, and optimum colour information transmission in the retina. *Proceedings of the Royal Society of London, B*, 220:89–113, 1983.

D. Burschka, J. Geiman, and G. Hager. Optimal landmark configuration for vision–based control of mobile robots. In *Proceedings of the ICRA 2003*, pages 3917–3922, 2003.

J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:679–698, 1986.

M. Carandini. *The Cognitive Neuroscience*, chapter Receptive Fields and Suppressive Fields in the Early Visual System. MIT Press, 2004.

M. Carandini and D. Ringach. Predictions of a recurrent model of orientation selectivity. *Vision Research*, 37:3061–3071, 1997.

M. Carandini, D. Heeger, and J. Movshon. Linearity and normalization in simple cells of macaque primary visual cortex. *Journal of Neuroscience*, 17:8621–8644, 1997.

M. Carandini, D. Heeger, and J. Movshon. *Cerebral Cortex, Volume 13, Cortical Models*, chapter Linearity and Gain Control in V1 Simple Cells, pages 401–443. Plenum Press, 1999.

M. Caywood, B. Willmore, and D. Tolhurst. Independent components of color natural scenes resemble V1 neurons in their spatial and color tuning. *Journal of Neurophysiology*, 91:2859–2873, 2004.

L. Chalupa and J. Weber, editors. *The Visual Neurosciences*, volume 1. MIT Press, 1st edition, 2003a.

L. Chalupa and J. Weber, editors. *The Visual Neurosciences*, volume 2. MIT Press, 1st edition, 2003b.

W. Cochran, H. Mouritsen, and M. Wikelski. Migrating songbirds recalibrate their magnetic compass daily from twilight cues. *Science*, 304:405–408, 2004.

T. Collett and M. Collett. Memory use in insect visual navigation. *Nature Reviews Neuroscience*, 3:542–552, 2002.

T. Cover and J. Thomas. *Elements of Information Theory.* Wiley Interscience, 1st edition, 1991.

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*, 2:1160–1169, 1985.

P. Dayan and L. Abbott. *Theoretical Neuroscience – Computational and Mathematical Modeling of Neural Systems*. MIT Press, 2001.

V. de Verdiere and J. Crowley. Local appearance space for recognition of navigation landmarks. In *Proceedings of the SIRS 1998*, pages 261–269, 1998.

G. DeAngelis and A. Anzai. *The Visual Neurosciences*, chapter A Modern View of the Classical Receptive Field: Linear and Nonlinear Spatiotemporal Processing by V1 Neurons, pages 704–719. MIT Press, 2003.

G. DeAngelis, I. Ohzawa, and R. Freeman. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex I, general characteristics and postnatal development. *Journal of Neurophysiology*, 69:1091–1117, 1993a.

G. DeAngelis, I. Ohzawa, and R. Freeman. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex II, linearity of temporal and spatial summation. *Journal of Neurophysiology*, 69:1118–1135, 1993b.

X. Deng, E. Milios, and A. Mirzaian. Landmark selection strategies for path execution. *Robotics and Autonomous Systems*, 17:171–185, 1996.

E. Doi, T. Inui, T. Lee, T. Wachtler, and T. Sejnowski. Spatiochromatic receptive field properties derived from information–theoretic analyses of cone mosaic responses of natural scenes. *Neural Computation*, 15:397–417, 2003.

R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley–Interscience, 2001.

G. Dudek and M. Jenkin. *Computational Principles of Mobile Robotics*. Cambridge University Press, 2000.

G. Dudek and D. Jugessur. Robust place recognition using local appearance based methods. In *Proceedings of the ICRA 2000*, pages 1030–1035, 2000.

J. Ferruz, S. Hurtado, and A. Ollero. Robot position estimation based on homographies of the ground plane. In *Proceedings of the IAV 2004*, 2004.

M. Franz, B. Schölkopf, H. Mallot, and H. Bülthoff. Learning view graphs for robot navigation. *Autonomous Robots*, 5:111–125, 1998.

R. Freitas, J. Santos-Victor, M. Sarcinelli-Filho, and T. Bastos-Filho. Performance evaluation of incremental eigenspace models for mobile robot localization. In *Proceedings of the ICAR 2003*, pages 417–422, 2003.

J. Gaspar, N. Winters, and J. Santos-Victor. Vision-based navigation and environmental representations with an omnidirectional camera. *IEEE Transactions on Robotics and Automation*, 16:890–898, 2000.

W. Geisler, J. Perry, B. Super, and D. Gallogly. Edge co–occurrence in natural images predicts contour grouping performance. *Vision Research*, 41:711–724, 2001.

D. Geman and B. Jedynak. An active testing model for tracking roads in satellite images. *IEEE Transaction on Pattern Analysis and Machinge Intelligence*, 18:1–14, 1996.

L. Gerstmayr and H. Mallot. How does the environment influence the shape of receptive fields. In *Proceedings of the Eighth Tübinger Perception Conference*, 2005.

L. Gerstmayr, A. Bernardino, and J. Santos-Victor. Appearance based landmark selection and reliability evaluation for topological navigation. In *Proceedings of the IAV 04*, 2004a.

L. Gerstmayr, A. Bernardino, and J. Santos-Victor. Appearance based landmark selection and reliability evaluation for topological navigation. Technical report, Vislab, Instituto de Sistemas e Róbotica, Instituto Superior Técnico, Lisbon, 2004b.

R. Golledge. *Wayfinding Behaviour*, chapter Human Wayfinding and Cognitive Maps. John Hopkins University Press, 1999.

N. Gracias and J. Santos-Victor. Trajectory reconstruction with uncertainty estimation using mosaic registration. *Robotics and Autonomous Systems*, 35:163–177, 2001.

N. Gracias, S. van der Zwaan, A. Bernardino, and J. Santos-Victor. Mosaic based navigation for autonomous underwater vehicles. *IEEE Journal of Ocean Engeneering*, 28:609–634, 2003.

C. Grigourescu, N. Petkov, and A. Westenberg. Contour detection based on nonclassical receptive field inhibition. *IEEE Transactions on Image Processing*, 12:720–739, 2003.

M. Gur, I. Kagan, and D. Snodderly. Orientation and direction selectivity of neurons in V1 of alert monkeys: Functional relationships and laminar distributions. *Cerebral Cortex*, 2004.

A. G.-D. H. Le Borgne. Sparse–dispersed coding and image discrimination with independent component analysis. In *Proceedings of the ICA 2001*, 2001.

H. Haken and J. Portugali. The face of the city is its information. *Journal of Environmental Psychology*, 23:385–408, 2003.

P. Hall, A. Marshall, and R. Martin. Incremental eigenanalysis for classification. In *Proceedings of the BMCV 1998*, pages 286–295, 1998.

T. Hansen. *A Neural Model of Early Vision: Contrast, Contours, Corners and Surfaces*. PhD thesis, University of Ulm, Faculty of Computer Science, 2002.

T. Hansen and H. Neumann. A biologically motivated scheme for robust junction detection. In *Proceedings of the BMCV 2002*, pages 16–26, 2002.

T. Hansen and H. Neumann. Neural mechanisms for the robust representation of junctions. *Neural Computation*, 16:1013–1037, 2004a.

T. Hansen and H. Neumann. A simple cell model with dominating opponent inhibition for robust image processing. *Neural Networks*, 17:647–662, 2004b.

C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, 1988.

F. Heitger, L.Rosenthaler, R. von der Heydt, E. Peterhans, and O. Kübler. Simulation of neural contour mechanisms: From simple to end–stopped cells. *Vision Research*, 32:963–981, 1992.

S. Hinz and A. Baumgartner. Automatic extraction of urban road networks from multi-view aerial imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(1–2): 83–98, 2003.

J. Hirsch, J. Alonso, R. Reid, and L. Martinez. Synaptic integration in striate cortical simple cells. *Journal of Neuroscience*, 18:9517–9528, 1998.

J. Hirsch, L. Martinez, C. Pillai, J. Alonso, Q. Wang, and F. Sommer. Functionally distinct inhibitory neurons at first stage of visual cortical processing. *Nature Neuroscience*, 6:1300–1308, 2003.

P. Hoyer. Modeling receptive fields with non-negative sparse coding. In *Proceedings of the IEEE Workshop on Neural Networks for Signal Processing*, pages 557–565, 2002.

P. Hoyer and A. Hyvärinen. A multi-layer sparse coding network learns contour coding from natural images. *Vision Research*, 42(12):1593–1605, 2002.

P. Hoyer and A. Hyvärinen. Independent component analysis applied to feature extraction from colour and stereo images. *Network: Computation in Neural Systems*, 11(3): 191–210, 2000.

D. Hubel and T. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160:106–154, 1962.

E. Hygounenc, I. Jung, P. Soueres, and S. Lacroix. The autonomous blimp project at LAAS–CNRS: Achievements in flight control and terrain mapping. *The International Journal of Robotics Research*, 23:474–511, 2004.

A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999a.

A. Hyvärinen. Survey on independent component analysis. *Neural Computing Surveys*, 2:94–128, 1999b.

A. Hyvärinen. Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4–5):411–430, 2000.

A. Hyvärinen and P. Hoyer. Emergence of phase– and shift–invariant features by decomposition of natural images into independent feature subspaces. *Neural Computation*, 12(7):1705–1720, 2000.

A. Hyvärinen and P. Hoyer. A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vision Research*, 41(18): 2413–2423, 2001.

A. Hyvärinen, P. Hoyer, and M. Inki. Topographic independent component analysis. *Neural Computation*, 13(7):1527–1558, 2001a.

A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley Interscience, 1st edition, 2001b.

A. Hyvärinen, P. Hoyer, and J. Hurri. Extensions of ICA as models of natural images and visual processing. In *Proceedings of the International Symposium on Independent Component Analysis and Blind Source Separation*, pages 963–974, 2003a.

A. Hyvärinen, J. Hurri, and J. Väyrynen. Bubbles: A unifying framework for low–level statistical properties of natural image sequences. *Journal of the Optical Society of America A*, 20(7):1237–1252, 2003b.

W. Hübner. *From Homing Behaviour to Cognitive Mapping*. PhD thesis, University of Bremen, 2005.

F. Iida. Biologically inspired visual odometer for navigation of a flying robot. *Robotics and Autonomous Systems*, 44:201–208, 2003.

M. Inki. *Extensions of Independent Component Analysis for Natural Image Data*. PhD thesis, Helsinki University of Technology, 2004.

C. Jacob. *Intelligent Data Analysis*, chapter Stochastic Search Methods, pages 351–401. Springer, 2003.

G. Janzen and M. van Turennout. Selective neural representation of objects relevant for navigation. *Nature Neuroscience*, 7:673–677, 2004.

M. Jogan and A. Leonardis. Robust localization using eigenspace of spinning-images. In *In Proceedings of IEEE Workshop on Omnidirectional Vision*, pages 37–44, 2000.

A. Johnson. Surface landmark selection and matching in natural terrain. In *Proceedings of the CVPR 2000*, pages 413–420, 2000.

J. Jones and L. Palmer. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58:1187–1211, 1987a.

J. Jones and L. Palmer. The two-dimensional spectral structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58:1212–1232, 1987b.

J. Jones and L. Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58:1233–1258, 1987c.

D. Jugessur and G. Dudek. Local appearance for robust object recognition. In *Proceedings of the CVPR 2000*, pages 834–839, 2000.

B. Jähne. *Digital Image Processing*. Springer, 5th edition, 2002.

D. Kammler. *A First Course in Fourier Analysis*. Prentice Hall, 1st edition, 2000.

E. Kandel, J. Schwartz, and T. Jessell, editors. *Principles of Neural Science*. McGraw Hill, 4th edition, 2000.

A. Kelly. Mobile robot localization from large scale appearance mosaics. *International Journal of Robotics Research*, 19:1104–1125, 2000.

M. Knapek, R. Oropeza, and D. Kriegmann. Selecting promising landmarks. In *Proceedings of the ICRA 2000*, pages 3771–3777, 2000.

I. Kokkinos, R. Deriche, P. Maragos, and O. Faugeras. A biologically motivated and computationally tractable model of low and mid–level vision tasks. In *Proceedings of the ECCV*, 2004.

M. Kolesnik, A. Barlit, and E. Zubkov. Iterative tuning of simple cells for contrast invariant edge enhancement. In *Proceedings of the BMCV 2002*, pages 27–37, 2002.

S. Konishi, A. Yuille, and J. Coughlan. Fundamental bounds on edge detection: An information theoretic evaluation of different edge cues. In *Proceedings of the CVPS 1999*, 1999.

S. Konishi, A. Yuille, J. Coughlan, and S. Zhu. Statistical edge detection: Learning and evaluating edge cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:57–74, 2003.

T. Labhart and E. Meyer. Neural mechanisms in insect navigation: Polarization compass and odometer. *Current Opinion in Neurobiology*, 12:707–714, 2002.

T. Lee and M. Lewicki. Unsupervisedd image classification, segmentation and enhancement using ICA mixture models. *IEEE Transactions on Image Processing*, 11:270–279, 2002.

T. Lee, M. Lewicki, and T. Sejnowski. ICA mixture models for unsupervised classification of non–gaussian classes and automatic context switching in BSS. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 22:1078–1088, 2000.

J. Leonard and H. Durrant-Whyte. Mobile robot localization by tracking geometric beacons. *IEEE Transactions on Robotics ans Autimation*, 7:376–382, 1991.

A. Leonardis and H. Bischof. Robust recognition using eigenimages. *Computer Vision and Image Understanding*, 78:99–118, 2000.

Z. Li. A neural model of contour integration in the primary visual cortex. *Neural Computation*, 10:903–940, 1998.

Z. Li. Visual segmentation by contextual influences via intra–cortical interactions in the primary visual cortex. *Computation in Neural Systems*, 10:187–212, 1999.

P. Lima, M. Ribeiro, L. Custodio, and J. Santos-Victor. The RESCUE project – cooperative navigation for rescue robots. In *Proceedings of the ASER'03*, 2003.

T. Lindeberg. *Handbook of Computer Vision and Applications*, chapter Principles for Automatic Scale Space Detection, pages 239–274. Academic Press, 1999.

J. Little, J. Lu, and D. Murray. Selecting stable image features for robot localization using stereo. In *Proceedings of the IROS 98*, pages 1072–1077, 1998.

D. Lowe. Distinctive image features from scale invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.

G. Loy and A. Zelinsky. Fast radial symmetry for detecting points of interest. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:959–973, 2003.

K. Lynch. *The Image of the City*. MIT Press, 1960.

H. Lähdesmäki, I. Shmulevich, L. Pezzati, and A. Tozzi. Optimization of edge detectors for topographic maps of cave inscriptions. In *Proceedings of the SCIA 2001*, 2001.

D. Marr. *Vision*. Freeman, 1982.

M. Mata, J. Armingol, A. de la Escalera, and M. Salichs. A visual landmark recognition system for topological navigation of mobile robots. In *Proceedings of the ICRA 2001*, pages 1124–1129, 2001.

G. Medioni, M. Lee, and C. Tang. *A Computational Framework for Segmentation and Grouping*. Elsevier, 2000.

J. Mena. State of the art on automatic road extraction for gis update: a novel classification. *Pattern Recognition Letters*, 24:3037–3058, 2003.

F. Metelo and L. Garcia. Vision–based contraol of an autonomous blimp. Technical report, Vislab, Instituto de Sistemas e Robótica, Instituto Superior Técnico, Lisbon, 2003.

C. Mora, M. Davison, J. Wild, and M. Walker. Magnetoreception and its trigeminal mediation in the homing pigeon. *Nature*, 432:508–511, 2004.

H. Mouritsen. Navigation in birds and other animals. *Image and Vision Computing*, 19: 713–731, 2001.

H. Mouritsen, U. Janssen-Bienhold, M. Liedvogel, G. Feenders, J. Stalleicken, P. Dirks, and R. Weiler. Cryptochromes and neuronal–activity markers colocalize in the retina of migratory birds during magnetic orientation. *PNAS*, 101:14294–14299, 2004.

D. Mumford and B. Gidas. Stochastic models for generic images. *Quarterly of Applied Mathematics*, 1:85–111, 2001.

H. Murase and S. Nayar. Visual learning and recognition of 3–D objects from appearance. *International Journal on Computer Vision*, 14:5–24, 1995.

L. Muratet, S. Doncieux, Y. Briere, and J. Meyer. A contribution to vision–based autonomous helicopter flight in urban environments. *Robotics and Autonomous Systems*, 50:195–209, 2005.

K. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf. An introduction to kernel–based learning algorithms. *IEEE Transactions on Neural Networks*, pages 181–202, 2001.

H. Neumann. Mechanisms of neural architecture for visual contrast and brightness perception. *Neural Networks*, 9:921–936, 1996.

H. Neumann and W. Sepp. Recurrent v1–v2 interactions in early visual boundary processing. *Biological Cybernetics*, 81:425–444, 1999.

H. Neumann, L. Pessoa, and T. Hansen. Interaction of ON and OFF pathways for visual contrast measurement. *Biological Cybernetics*, 81:512–532, 1999.

K. Ohba and K. Ikeuchi. Detectability, uniqueness and reliability of eigen windows for stable verification of partially occluded objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:1043–1048, 1997.

A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001.

B. Olshausen. *The Visual Neurosciences*, chapter Principles of Image Representation in Visual Cortex. MIT Press, 2003.

B. Olshausen. *Sparse Codes and Spikes*, chapter Probabilistic Models of Perception and Brain Function. MIT Press, 2001.

B. Olshausen and D. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37:3311–3325, 1997.

B. Olshausen and D. Field. Natural image statistics and efficient coding. *Network: Computation in Neural Systems*, 7:333–339, 1996.

C. Olson. Selecting landmarks for localization in natural terrain. *Autonomous Robots*, 12:201–210, 2002.

W. Pennebaker and J. Mitchell. *The JPEG Still Image Data Compression Standard*. Van Nostrand-Reinhold, 1993.

W. Press, S.A.Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipes in C++*. Cambridge University Press, 2nd edition, 2003.

D. Reisfeld, H. Wolfson, and Y. Yeshurun. Context free attentional operators: the generalized symmetry transform. *International Journal of Computer Vision*, 14:119–130, 1995.

A. Rencher. *Methods of Multivariate Analysis*. Wiley–Interscience, 2002.

D. Ringach. Mapping receptive fields in primary visual cortex. *Journal of Physiology*, 558:717–728, 2004.

D. Ringach. Spatial structure and symmetry of simple–cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology*, 88:455–463, 2002.

D. Ringach, R. Shapley, and M. Hawken. Orientation selectivity in macaque V1: Diversity and laminar dependence. *Journal of Neuroscience*, 22:5639–5651, 2002.

D. Ruderman. The statistics of natural images. *Network: Computation in Neural Systems*, 5:517–548, 1994.

D. Ruderman, T. Cronin, and C. Chiao. Statistics of cone responses to natural images: Implications for visual coding. *Journal of the Optical Society of America A*, 15(8): 2036–2045, 1998.

F. Ruffier and N. Franceschini. Optic flow regulation: The key to aircraft automatic guidance. *Robotics and Autonomous Systems*, 50:177–194, 2005.

A. Rupar, A. Leonardis, and N. M. Kosta. Robust recognition using the tensor-rank principle. In *Proceedings of the 26th Workshop of the Austrian Association for Pattern Recognition*, pages 45–52, 2002.

J. Santos-Victor, N. Gracias, and S. van der Zwaan. Using vision for underwater robotics: Video mosaics and station keeping. In *Proceedings of the First International Workshop on Underwater robotics for Sea Exploitation and Environmental Monitoring*, 2001.

C. Schmid and R. Mohr. Local grayvalue invariants for image retrival. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:530–535, 1997.

C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *International Journal on Computer Vision*, 37:151–172, 2000.

C. Shannon and W. Weaver, editors. *The Mathematical Theory of Communication*. University of Illinois Press, 1949.

A. Shashua and A. Levin. Linear image coding for regression and classification using the tensor–rank principle. In *Proceedings of the IEEE CVPR*, 2001.

A. Shaw and D. Barnes. Landmark recognition for localisation and navigation of aerial vehicles. In *Proceedings of the ASTRA 2002*, pages 427–434, 2002.

J. Shi and C. Tomasi. Good features to track. In *IEEE Conference on Computer Vision and Pattern Recognition CVPR'94*, pages 593–600, 1994.

A. Si, M. Srinivasan, and S. Zhang. Honeybee navigation: Properties of the visually driven odometer. *Journal of Experimental Biology*, 206:1265–1273, 2003.

R. Siegwart and I. Nourbaksh. *Introduction to Autonomous Mobile Robots*. MIT Press, 2004.

M. Sigman, G. Cecchi, C. Gilbert, and M. Magnasco. On a common circle: Natural scenes and gestalt rules. *Proceedings of the National Academy of Science, USA*, 98: 1935–1940, 2001.

R. Sim, S. Polifroni, and G. Dudek. Comparing attention operators for learning landmarks. Technical Report TR–CIM–03–03, McGill University, Quebec, Canada, 2003.

E. Simoncelli and B. Olshausen. Natural image statistics and neural representation. *Annual Reviews Neuroscience*, 24:1193–1216, 2001.

D. Somers, S. Nelson, and M. Sur. An emergent model of orientation selectivity in cat visual cortical simple cells. *Journal of Neuroscience*, 15:5448–5465, 1995.

L. Spillmann and W. Ehrenstein. *The Visual Neurosciences*, chapter Gestalt Factors in the Visual Neurosciences, pages 1573–1589. MIT Press, 2003.

A. Srivasta, A. Lee, E. Simoncelli, and S. Zhu. On advances in statistical modeling of natural images. *Journal of Mathematical Imaging and Vision*, 18:17–33, 2003.

J. Stone. *Independent Component Analysis*. MIT Press, 1st edition, 2004.

K. Sutherland and W. Thompson. Localizing in unstructured environments: Dealing with errors. *IEEE Transactions on Robotics and Automation*, 10:740–755, 1994.

D. Tailor, L. Finkel, and G. Buchsbaum. Color–oppponent receptive fields derived from independent component analysis of natural images. *Vision Research*, 40:2671–2676, 2000.

S. Thrun. Bayesian landmark learning for mobile robot localization. *Machine Learning*, 33:41–76, 1998.

D. Tolhurst and D. Heeger. Comparison of contrast normalization and threshold models of the responses of simple cells in cat striate cortex. *Visual Neuroscience*, 14:293–309, 1997a.

D. Tolhurst and D. Heeger. Contrast normalization and linear model for the directional selectivity of simple cells in cat striate cortex. *Visual Neuroscience*, 14:19–25, 1997b.

A. Torralba and A. Oliva. Statistics of natural image categories. *Network: Computation in Neural Systems*, 14:391–412, 2003.

B. Triggs. Detecting keypoints with stable position, orientation, and scale under illumination conditions. In *Proceedings of the ECCV 2004*, 2004.

T. Troyer, A. Krukowski, and K. Miller. LGN intput to simple cells and contrast invariant orientation tuning: An analysis. *Journal of Neurophysiology*, 87:2741–2752, 2001.

M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3:71–86, 1991.

I. Ulrich and I. Nourbakhsh. Appearance–based place recognition for topological navigation. In *Proceedings of the ICRA 2000*, pages 1023–1029, 2000.

M. Ursino and G. L. Cara. Comparison of different models of orientation selectivity based on distinct intracortical inhibition rules. *Vision Research*, 44:1641–1658, 2004.

A. Utsugi. Independent components of natural images under variable compression rate. *Neurocomputing*, 49:175–185, 2002.

A. van der Schaaf and J. van Hateren. Modelling the power spectra of natural images: statistics and information. *Vision Research*, 36:2759–2770, 1996.

J. van Hateren and A. van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society London B*, 265:359–366, 1998.

R. Vasallo, J. Santos-Victor, and H. Schneebeli. Using motor commands for topological mapping and navigation. In *Proceedings of the 2002 ICIRS 2002*, pages 478–483, 2002.

T. Wachtler, T. Lee, and T. Sejnowski. Chromatic structure of natural scenes. *Journal of the Optical Society of America, A*, 18:65–77, 2001.

B. Webb. Robots in invertebrate neuroscience. *Nature*, 417:359–363, 2002.

B. Webb. Can robots make good models of biological behaviour? *Behavioral and Brain Sciences*, 24:1033–1050, 2001.

B. Webb. What does robotics offer animal behaviour? *Animal Behaviour*, 60:545–558, 2000.

C. Wei, S. Rafalko, and F. Dyer. Deciding to learn: Modulation of learning flights in honeybees. *Journal of Comparative Physiology, A*, 188:725–737, 2002.

B. Willmore, P. Watters, and D. Tolhurst. A comparison of natural–image–based models of simple–cell–coding. *Perception*, 29:1017–1040, 2000.

R. Wiltschko and W. Wiltschko. Avian navigation: from historical to modern concepts. *Animal Behaviour*, 65:257–272, 2003.

N. Winters and J. Santos-Victor. Information sampling for vision–based robot navigation. *Robotics and Autonomous Systems*, 41:145–159, 2002.

R. Würtz and T. Lourens. Corner detection in color images through a multiscale combination of end–stopped cells. *Image and Vision Computing*, 18:531–541, 2000.

A. Yuille and J. Coughlan. Fundamental limits of bayesian inference: Order parameters and phase transitions for road tracking. *IEEE Transactions on Pattern Analysis and Machine Vision*, 22:160–173, 2000.

C. Zetsche and F. Röhrbein. Nonlinear and extra–classical receptive field properties and the statistics of natural scenes. *Network: Computation in Neural Systems*, 12: 331–350, 2001.

C. Zhang. Towards an operational system of automated updating of road databases by integration of imagery and geodata. *Photogrammetry and Remote Sensing*, 58: 166–186, 2004.

L. Zhang and J. Mei. Shaping up simple cell's receptive fields of animal vision by ICA and its application in navigation system. *Neural Networks*, 16:609–615, 2003.

C. Zieghaus and E. Lang. Independent component analysis of natural and urban image ensembles. *Neural Information Processing*, 1:89–95, 2003.

M. Zigmond, F. Bloom, S. Landis, J. Roberts, and L. Squire. *Fundamental Neuroscience*. Academic Press, 1st edition, 1999.

S. Zucker, A. Dobbins, and L. Iverson. Two stages of curve detection suggest two styles of visual computation. *Neural Computation*, 1:68–81, 1989.